UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Essays in Microeconometrics

Permalink
https://escholarship.org/uc/item/6bg4r29s

Author
Kasy, Maximilian

Publication Date
2011

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Essays in Microeconometrics

by

Maximilian Kasy

A dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor James Powell, Chair
Professor Bryan Graham
Professor Mark van der Laan

Spring 2011

Essays in Microeconometrics

Copyright © 2011

by

Maximilian Kasy

Abstract

Essays in Microeconometrics

by

Maximilian Kasy

Doctor of Philosophy in Economics

University of California, Berkeley

Professor James Powell, Chair

This dissertation consists of four chapters contributing to the development of microeconometric methodology, with a particular emphasis on questions of identification. The methodological problems discussed are motivated by substantive questions about the causes of urban segregation and of long term unemployment.

Chapter 1 develops static and dynamic models of sorting in which location choices depend on the location choices of other agents as well as prices and exogenous location characteristics. In these models, demand slopes and hence preferences are not identifiable without further restrictions because of the absence of independent variation of endogenous composition and exogenous location characteristics. Four solutions of this problem are presented and applied to data on neighborhoods in US cities: The first three use exclusion restrictions, based on either subgroup demand shifters, the spatial structure of externalities, or the dynamics of prices and composition in response to an amenity shock. The fourth tests for multiplicity of equilibria in the dynamics of composition, using the test proposed in chapter 2. The empirical results consistently suggest the presence of strong social externalities, that is, a dependence of location choices on neighborhood composition.

Chapter 2 proposes an estimator and develops an inference procedure for the number of roots of functions which are nonparametrically identified by conditional moment restrictions. The estimator is superconsistent, and the inference procedure is based on non-standard asymptotics. This procedure is used to construct confidence sets for the number of equilibria of static games of incomplete information and of stochastic difference equations. In an application to panel data on neighborhood composition in the United States, no evidence of multiple equilibria is found.

Chapter 3 proposes a test for path dependence in discrete panel data based on a characterization of stochastic processes that are mixtures of Markov Chains. This test is applied to European Community Household Panel data on employment histories. The data allow to reject the null of no path dependence in all subsamples considered.

Chapter 4 discusses identification in nonparametric, continuous triangular systems. It provides conditions which are both necessary and sufficient for the existence of control functions satisfying conditional independence and support requirements. Confirming a commonly noticed pattern, these conditions restrict the admissible dimensionality of unobserved heterogeneity in the first stage structural relation, or more generally the dimensionality of the family of conditional distributions of second stage heterogeneity given explanatory variables and instruments. These conditions imply that no such control function exists without assumptions that seem hard to justify in most applications. In particular, none exists in the context of a generic random coefficient model.

I thank Kehinde Ajayi, David Card, Giuseppe Cavaliere, Kiril Datchev, Bryan Graham, Helmut Hofer, Michael Jansson, Susanne Kimm, Patrick Kline, Zachary Liscow, Enrico Moretti, Denis Nekipelov, James Powell, Álvaro Ramos-Chaves, Steven Raphael, Geert Ridder, Alexander Rothenberg, Jesse Rothstein, Paul Ruud, Sarath Sanga, Estefania Santacreu-Vasut, Mark van der Laan, and several anonymous referees for many valuable discussions and comments. I also thank my family and friends, both in Berkeley and in Vienna, for having been there through the years.

This work was supported by a DOC fellowship of the Austrian Academy of Sciences at the Department of Economics, UC Berkeley. Access to the NCDB data used in chapter 2, as well as their scripts, was provided by David Card, Alexander Más, and Jesse Rothstein. Access to the ECHP data used in chapter 3 was provided by the Jubiläumsfonds of the Austrian Nationalbank grant no 9632.

To Susi, my family and friends

Contents

Chapter 1

Identification in models of sorting with social externalities

Abstract

This paper develops static and dynamic models of sorting in which location choices depend on the location choices of other agents as well as prices and exogenous location characteristics. In these models, demand slopes and hence preferences are not identifiable without further restrictions because of the absence of independent variation of endogenous composition and exogenous location characteristics. Four solutions of this problem are presented and applied to data on neighborhoods in US cities: The first three use exclusion restrictions, based on either subgroup demand shifters, the spatial structure of externalities, or the dynamics of prices and composition in response to an amenity shock. The fourth tests for multiplicity of equilibria in the dynamics of composition. The empirical results consistently suggest the presence of strong social externalities, that is, a dependence of location choices on neighborhood composition.

1.1 Introduction

Urban areas in the United States and across the world show large degrees of social segregation across neighborhoods. For instance, as documented by Cutler, Glaeser, and Vigdor (2008), the average dissimilarity index of immigrant groups' distribution across neighborhoods in US cities has risen continuously since 1920 from an initial value of 0.34 to a high point of 0.56 in 2000. The dissimilarity index of Hispanic distribution across neighborhoods within Metropolitan Areas for the sample used in this paper also rose slightly, from .42 in 1980 to .44 in 2000.[1] Similar degrees of segregation can be observed along many dimensions. This

[1] The dissimilarity index is defined as $0.5 \sum_i |h^i/H - nh^i/NH|$, where the sum is taken over census tracts within a city, h^i denotes number of Hispanic residents in the tract, H the number of Hispanic residents in

is of concern if the social environment in neighborhoods is an important determinant of life outcomes.

There are two polar explanations of segregation. Households might sort across locations because of different willingness to pay for exogenous location characteristics, which may be due to differences in income or differences in preferences. This is the explanation emphasized by accounts of sorting such as the classic Tiebout (1956) and Rosen (1974). Alternatively, households might care about who their neighbors are, and hence choose their neighborhood based on demographic composition. This possibility was discussed by Schelling (1971) and Becker and Murphy (2000).

The present paper allows for both possibilities. Consider the following setup: Households have to choose whether or not to locate in a given neighborhood. This choice depends upon exogenous neighborhood characteristics, which will be denoted X, a vector describing the endogenous composition of the residents of a neighborhood, M, as well as endogenous rental prices, P. The local housing market is in equilibrium if the composition of households that *want* to locate in a neighborhood, D, equals the composition of those that *are* in the neighborhood, i.e., $D(X, M, P) = M$, and if total housing demand E equals housing supply at the given price, $E(X, M, P) = S(X, P)$.

Suppose we want to distinguish between the causes of segregation. In the setup just described, identification problems arise similar to those in models of endogenous peer effects, termed the "reflection problem" by Manski (1993). The reason is that both composition and rental prices are endogenous functions of X, preventing the separate identification of the effects of X and (M, P) on demand, or X and M on equilibrium prices. This support problem is of a different kind than that of omitted variable bias in hedonic or choice regressions, which is due to unobservability of relevant regressors. Solutions to the omitted variable problem have been proposed by Black (1999), who controls for border fixed effects, and by Chay and Greenstone (2005), who use exogenous variation in amenities. Bayer, Ferreira, and McMillan (2007) estimate hedonic and discrete choice models of sorting, and explicitly recognize the possibility of a preference for neighborhood composition. However, in light of the issues raised here, their identification strategy of using controls for composition may be problematic.

The first goal of this paper is to provide a general framework in which this identification problem becomes obvious. The object that we will be particularly interested to identify is the presence and degree of social externalities, defined here as a dependence of demand of various subgroups on the composition of neighborhood residents. In terms of the model, social externalities exist when $D_M \neq 0$.

The presence of social externalities in sorting is of relevance for several reasons. First, it poses a methodological problem in the estimation of willingness-to-pay parameters, which in turn are often used for cost-benefit analyses of policies. Knowing the magnitude of the

the city, and similarly for nh^i, NH. The number reported is a population weighted average across cities. The dataset will be described in detail in section 1.6.

dependence of demand on composition allows us to assess the magnitude of biases in hedonic slopes and choice regressions. Second, externalities matter for understanding the causes of social segregation across locations and imply multipliers on policies affecting segregation. Third, if externalities are strong, multiple equilibria in population composition at a given location arise. Multiple equilibria in turn can imply discontinuous and large effects of demand shifting policies due to bifurcations, as emphasized by Schelling (1971) and Card, Mas, and Rothstein (2008). Such bifurcations might explain phenomena of rapid gentrification or the reverse.

Finally, it is interesting to contrast the importance households attach to neighborhood composition in their location choice with the available evidence on the effect of neighborhood environment on observable outcomes. Evidence on the latter is mixed. See, for example, Katz, Kling, and Liebman (2007). The present paper, on the other hand, finds strong effects of composition on location choice.

It is important to recognize the differences between the setup developed here and the models of peer effects discussed in the literature. First, in sorting models a location is matched with an endogenous set of agents with fixed characteristics, whereas in models such as those discussed by Manski (1993) or Moffitt (2001), there is a fixed set of agents with endogenous outcomes. Second, the reflection problem in models of peer effects is the problem of distinguishing endogenous from exogenous peer effects, *not* the problem of distinguishing peer effects from non-random matching, whereas in the sorting model developed here the fundamental problem is to identify whether there are social externalities at all. Third, in the setup discussed here, there is a price mechanism allocating households to neighborhoods, which is absent from peer-effects models. Finally, in peer effects models, endogenous sorting might be a cause of identification problems, and as such is a nuisance, whereas here it is the object of interest.

Four possible solutions to the identification problem are discussed in this paper. The first three are based on assuming exclusion restrictions. The first approach uses exogenous shifters of demand of certain subgroups that are excluded from the demand of other subgroups. This builds on the idea of randomized subgroup treatment used for the identification of peer effects, as recommended in Moffitt (2001) and applied for instance by Duflo and Saez (2003). The second approach exploits the spatial structure of cities in an extension of the baseline model, allowing for interactions across adjacent neighborhoods. Identification comes from the assumption that exogenous demand shifters for neighborhoods beyond a certain distance are excluded from local demand. This idea is analogous to the use of social network structures to identify endogenous versus exogenous peer effects, as in Bramoullé, Djebbari, and Fortin (2009) and De Giorgi, Pellizzari, and Redaelli (2009).

The third and fourth approaches are based on a dynamic extension of the baseline model. This dynamic extension assumes search frictions in moving from one neighborhood to another. This dynamic model is similar to search models of the labor market as surveyed in Pissarides (2000). It builds upon search models of the housing market such as Wheaton

(1990). The third approach is based on the result that, under certain conditions, past amenity shocks are excluded from future price changes, as the value of amenities is immediately reflected in rental prices. Composition, however, does adjust with delay due to search frictions, and hence prices adjust to this composition change with the same delay. The fourth and last approach tests for multiplicity of equilibria in the dynamics of composition, as implied by sufficiently strong social externalities in the search model. This approach is a modification and theoretical foundation of the test for neighborhood tipping in Card, Mas, and Rothstein (2008). It uses techniques developed in chapter 2, where we discuss inference on the number of roots of functions nonparametrically identified via conditional moment restrictions.

These approaches are applied to data from the Neighborhood Change Data Base (NCDB), which aggregates US Census data to the level of census tracts. The composition variable considered is Hispanic share. The first approach uses a synthetic instrument based on prior composition and national immigration. This instrument is a composition shifter that is excluded from the demand of non-Hispanics. A similar instrument has been used by Card (2001). In the second approach, the spatial structure of cities is exploited, assuming cross-neighborhood spillovers. In particular, the identifying assumption is made that predicted immigration for neighborhoods more than three kilometers away from a given neighborhood is excluded from demand conditional on predicted immigration in the given neighborhood. This approach uses variation in actual composition orthogonal to the first approach, as the instrument of the first approach is included as a control variable. The third approach uses past composition change as as an instrument for future composition change conditional on present composition, as justified by the search model.

All three instruments yield surprisingly consistent estimates. They suggest that a 1% increase in the Hispanic share of neighborhood population results in a 6 to 10% decline in non-Hispanics' demand, and a 3 to 4% rise in Hispanics' demand. Housing prices appear to decline by around 0.5% to 1% for a 1% increase in Hispanic share. Inference on the number of equilibria in the dynamics of composition mostly allows us to reject the null hypothesis of multiple equilibria. These results are also consistent with the conclusions of Cutler, Glaeser, and Vigdor (2008), who use variation in segregation across time, city, and immigrant groups in trying to disentangle the causes of segregation.

The models in this paper are described in terms of households choosing a neighborhood and paying rents. However, most of the insights should apply to other contexts of sorting. Examples include sorting of workers across firms, students across schools, customers across mobile-phone network providers, faculty across universities, or the spatial agglomeration and dispersion of firms.

Some further relevant contributions in the recent literature have to be mentioned before proceeding. Nesheim (2001) and Graham (2008) discuss identification issues in specific models of sorting where peer composition enters an educational production function. Heckman, Matzkin, and Nesheim (2002) and Ekeland, Heckman, and Nesheim (2004) derive identification of preferences from cross-sectional price data based on functional form restrictions (separability). Chiappori, McCann, and Nesheim (2009) show the equivalence of hedonic sorting, matching and optimal transport problems and derive existence results for equilibria in these models. There is a literature of discrete choice estimation of sorting models, emphasizing that hedonic regressions only identify the preferences of marginal households. Examples that allow for preferences for neighbors are Bayer, Ferreira, and McMillan (2007) and Caetano (2009).

The test for multiplicity of equilibria in the dynamics of composition applied here uses an inference procedure which is proposed in chapter 2. The development of the asymptotic theory for this inference procedure applies results from Kong, Linton, and Xia (2010) on Bahadur expansions for local polynomial m-regression, and uses somewhat similar arguments as Horváth (1991), who discusses the asymptotic distribution of L_p-norms of multivariate density estimators. The notion of sequences of experiments is discussed in van der Vaart (1998). The argument on bootstrap-based inference draws on the review of Horowitz (2001).

The rest of the paper is structured as follows: Section 1.2 develops a static model of locational sorting. The model of subsection 1.2.1 provides the general framework in which the identification problem becomes immediate. Subsection 1.2.2 restricts to a special case that allows for graphical illustration. Section 1.3 discusses the identification problem in the static model as well as solutions based on subgroup shifters and spatial structure. Section 1.4 provides a search model of the housing market that allows us to characterize the dynamics of composition and prices. Section 1.5 provides further routes to identification based on this dynamic extension, and discusses inference on the number of equilibria. Section 1.6 applies four estimators of the degree of social externalities to the NCDB data. Section 1.7 summarizes and concludes. Appendix 1.A states a series of results decomposing linear IV coefficients into weighted average structural slopes, where the weights are identifiable. All proofs are relegated to Appendix 1.B, all figures and tables can be found in appendix 1.C.

1.2 The static model of sorting with social externalities

This section presents the baseline model of sorting discussed in this paper. The model generalizes both discrete choice and hedonic sorting models. As this paper is about identification, functional forms and heterogeneity of utility are left unrestricted. The central feature of the model is that it allows for social externalities, in the sense that demand for housing at a location, and household utility, are allowed to depend on the composition of residents at that location. This dependence can reflect a direct preference over neighbors' types. It can also reflect a preference over amenities or production processes affected by neighbors' types, such as peer effects in education, crime etc., as in Nesheim (2001) or Graham (2008). A number of specializations and extensions of this basic model will be presented later, allowing in particular for cross-neighborhood externalities and for search frictions. Most of the analysis will be partial-equilibrium, in the sense that cross-neighborhood feedback is ignored.

1.2.1 General static model of location choice with social externalities

The baseline model, which will be formally stated below, can be summarized as follows: Consider one neighborhood among many. This neighborhood is characterized by exogenous properties, the composition of residents, and rental prices. Household utility for choosing to locate in this neighborhood depends on these exogenous characteristics, composition and rents. Households choose to live in the neighborhood if and only if the utility provided by living there is higher than utility for their best outside option. An example of exogenous neighborhood characteristics would be geographic location, an example of composition would be the share of various ethnic groups living in the neighborhood.

Households can be one of several types. Arbitrary heterogeneity of utility within and across types is allowed. Composition is restricted to enter household utility only in terms of the number of households of each type present in the neighborhood. Utility maximizing choices imply demand schedules for each type of household as a function of exogenous characteristics, composition, and prices. The observable prices and composition are assumed to be in equilibrium given exogenous neighborhood characteristics.

The baseline model consists of three assumptions. The structure of the model is illustrated in figure 1.1. As this figure shows, the assumptions map primitives to higher level objects, going from preferences to demand functions (assumption 1.2.3), from demand functions to equilibrium comparative statics (assumption 1.2.1 and definition 1), and from equilibrium comparative statics to the observable data distribution (assumption 1.2.2). The problem of identification is essentially the problem of inverting these maps. In this paper, the inversion of the map from demand functions to equilibrium comparative statics is of particular interest.

Assumption 1.2.1 describes the basic setup in terms of demand functions D. In combination with definition 1 of partial sorting equilibrium, it provides a mapping from demand functions to equilibrium schedules (M^*, P^*). These equilibrium schedules give population composition and rental prices in a neighborhood as a function of exogenous neighborhood characteristics and any other determinants of location choices.

Assumption 1.2.2 describes how the equilibrium schedules translate into a distribution of observable data $\mathbb{P}$. It restricts observable composition and prices to be in equilibrium given exogenous location characteristics. If we additionally assume full observability of all relevant location characteristics X, the resulting model provides an "upper bound" for identifiability. In particular, we can at most identify features of the model that can be written as a function of the equilibrium schedules. This is a useful point of departure for negative identification results. If on the other hand we have only partial observability of location characteristics but exogenous variation of some of the observed components of X, we get a useful setup for positive identification results.

Assumption 1.2.3, finally, connects household demand functions to underlying household utility u. Housing demand for a given type of household is defined as the number of households in the population for which the utility of living in the neighborhood is greater than the exogenously given utility of their outside option. This assumption characterizes the model as being partial equilibrium because the utility of outside options is assumed not to change as a function of the characteristics of the given neighborhood.

Throughout, superscripts will denote indices (for instance M^c is the cth component of M) and subscripts will denote partial derivatives (for instance $D_M = \partial D / \partial M$). Probabilities will be denoted by $\mathbb{P}$ in order to distinguish them from prices P.

Assumption 1.2.1 (The local economy)**.**

- *There are $\mathscr{C}$ types of households, $c = 1, \ldots, \mathscr{C}$.*
- *A neighborhood is characterized by*
  1. *the number of households of each type, $M = (M^1, \ldots, M^{\mathscr{C}})$*
  2. *a (rental) price P*
  3. *an exogenous vector X of* all *other location characteristics and factors influencing demand or supply.*
- *Demand for housing in a neighborhood, for each type, is a bounded continuously differentiable function of these. Denote by D^c the demand of households of type c, and let $D = (D^1, \ldots, D^{\mathscr{C}})$, then $D = D(X, M, P)$. Total demand is given by $E := \sum_c D^c$.*
- *Housing supply is a bounded continuously differentiable function of P and X, $S = S(P, X)$.*

It should be emphasized that X is defined inclusively as comprising *all* exogenous demand and supply shifters, including random fluctuations. Locations in this model only differ if X is different, all other variables will be endogenously determined given the exogenous X. The vector X includes determinants of demand such as preference and income distributions and demographic composition of the population, as well as local labor demand. In particular, the empirical application at the end of the paper will use changing demographic composition due to immigration as a demand shifter.

The influence of M on D is a reduced-form social externality. It includes both direct preferences over neighborhood composition and preferences over neighborhood properties influenced by composition, such as school quality. The assumption allows only for a finite number of types, as far as neighbors are concerned. This greatly simplifies exposition. Allowing for more general type-sets does not alter most conclusions.

The following definition of *partial sorting equilibrium* requires that the neighborhood composition is consistent with the demand of each of the different types, and that housing demand equals housing supply.

Definition 1 (Partial Sorting Equilibrium). *A partial sorting equilibrium (M^*, P^*) given X solves the $C+1$ equations*

$$D(X, M^*, P^*) = M^* \tag{1.1}$$

$$S(P^*, X) = \sum_c M^{*c}. \tag{1.2}$$

Let $(M^*(X), P^*(X))$ denote the correspondence mapping X into the partial sorting equilibria given X.

Equilibrium existence is guaranteed by

Proposition 1. *Under assumption 1.2.1, there exists at least one Partial Sorting Equilibrium (M^*, P^*) given X.*

The proof of this proposition, and of all further results, can be found in Appendix 1.B.

The following assumption states that M and P are in equilibrium given the exogenous determinants of demand and supply, X. Furthermore, $X = (X^1, \epsilon)$ where ϵ is a vector of unobserved determinants. In the case of unique partial equilibrium, all conditional randomness in M and P given X^1 comes from ϵ, i.e., $M = M^*(X^1, \epsilon)$ and $P = P^*(X^1, \epsilon)$. Two cases will be considered. The full observability case is useful to discuss the central identification problems, as it assumes away the problem of mapping from the data distribution to equilibrium schedules, and allows to focus on the problem of mapping equilibrium schedules to demand functions. The partial observability case with exogenous variation then allows to generalize the results obtained in the full observability case to more realistic settings.

Assumption 1.2.2 (Observable data).

- *The observable data consist of repeated observations of (X^1, M, P) where $X = (X^1, \epsilon)$ for vectors X^1 and ϵ.*
- *M and P are in equilibrium given X for all observations, i.e., $(M, P) \in (M^*(X), P^*(X))$.*
- *X is continuously distributed on its support in $\mathbb{R}^{dim(X)}$.*
- ***Full observability case:*** *$X = X^1$ and (M, P) have full support on $(M^*(X), P^*(X))$, so that $(M^*(X), P^*(X))$ is identified on the support of X.*
- ***Partial observability with exogenous variation case:*** *X^1 is statistically independent of ϵ and the equilibrium selection mechanism. Therefore, conditional average slopes of equilibrium schedules are identified by $E\left[\frac{\partial}{\partial x^1} M^*(x^1, \epsilon)\right] = \frac{\partial}{\partial x^1} E\left[M | X^1 = x^1\right]$. The first expectation in this expression is taken over the unconditional distribution of ϵ and the equilibrium selection mechanism. Similarly for P^*.*

The next assumption gives demand as the outcome of utility maximizing household choices. Utility is indirect in the sense that it is given as a function of neighborhood characteristics and rents, where all other household choices are "concentrated out." Households choose between locating in the neighborhood or elsewhere. The indirect utility of choosing to live elsewhere is not modeled, but assumed to be exogenously given. Unrestricted heterogeneity of utility is allowed. This assumption can be understood as describing a nonparametric discrete choice setup. If we have a priori knowledge that some factors are excluded from the location choices of some group, this will in fact yield a route to identification.

Assumption 1.2.3 (Household utility maximization)**.**

- *Households are characterized by the triple $(u(X, M, P), u^o, c)$, where u is their continuously differentiable indirect utility dependent on neighborhood characteristics, u^o is the utility of their best outside option and c is their type.*

- *Households locate in the given neighborhood iff $u(X, M, P) \geq u^o$.*

- *u^o is exogenously determined, i.e., constant in (X, M, P).*

- *There is a continuum of households of total mass M^{tot} in the economy. The vector (u, u_X, u_M, u_P, u^o), evaluated at any (X, M, P), has a continuous joint distribution.*

- *D^c is the mass of households that want to locate in the given neighborhood,*

$$D^c = M^{tot} \cdot \mathbb{P}(u \geq u^o, c).$$

  Similarly $E = M^{tot} \cdot \mathbb{P}(u \geq u^o)$.

The assumption of a continuous joint distribution is necessary for well defined differentials of the demand functions D^c.

In the subsequent discussion of identification, different combinations of these assumptions will be used, where the identification results discuss how to invert the mappings provided by these assumptions, possibly under additional restrictions. In the next subsection, assumption 1.2.1 will be specialized to the case of two household types, which allows for easy illustration of some important features of the model. Assumption 1.2.3 will also provide the connection to the dynamic model discussed later in this paper, which is more naturally stated in terms of utilities rather than resulting choices. As it will turn out, the model defined by assumptions 1.2.1 and 1.2.3 is a limit case of the dynamic model in the absence of search frictions, and describes steady state comparative statics of the dynamic model more generally.

1.2.2 A specialization to the case of two types

To provide some intuition for the implications of this model, let us consider a special case with only two types of households.

Assumption 1.2.4 (Two type model)**.**

- *There are only* $\mathscr{C} = 2$ *types.*
- *The price elasticity of demand of the two types is the same:*

$$\frac{D^1_P}{D^1} = \frac{D^2_P}{D^2}.$$

- *Both types have the same demand elasticity with respect to the scale of the neighborhood:*

$$\frac{1}{D^1}(D^1_{M^1}M^1 + D^1_{M^2}M^2) = \frac{1}{D^2}(D^2_{M^1}M^1 + D^2_{M^2}M^2).$$

Define d as the share of type 1 households among those who *want* to live in the neighborhood, i.e., $d = D^1/(D^1 + D^2)$. Similarly, let m be the share of type 1 households among those who *do* live in the neighborhood, $m = M^1/(M^1 + M^2)$. Recall finally that E is the total demand for housing in the neighborhood, $E = D^1 + D^2$. Under assumption 1.2.4, the demand share of type 1, d, can be written as a function of m and X alone, where X is exogenous. Put differently, the relative demand of the two types is not affected by prices or population density. This implies that (partial) equilibrium can be defined by the conditions

$$\begin{aligned} d(m^*, X) &= m^* & (1.3)\\ E(P^*, m^*, X) &= S(P^*, X), & (1.4)\end{aligned}$$

which have a recursive form that we can easily analyze, both graphically and analytically. The share of either type is a solution to the first equation. Given this equilibrium share, the second condition is a conventional partial-equilibrium supply and demand equation. Figure 1.2 represents these two equilibrium conditions as well as the comparative statics of the model.

Formally, consider a small change in X that does not affect housing supply, $S_X = 0$. Make assumptions 1.2.1 and 1.2.4 and assume that social externalities are not too strong, so that $d_m < 1$. Assume furthermore that partial sorting equilibrium is unique, or let (m^*, P^*) denote a differentiable selection from the set of partial equilibria. Then

$$m^*_X = \frac{d_X}{1 - d_m} \tag{1.5}$$

and

$$P^*_X = \frac{E_X + E_m m^*_X}{S_P - E_P}. \tag{1.6}$$

This follows immediately from equations 1.3 and 1.4.

Equation 1.6 gives the response of rents to amenity shifts. Below we will compare this to the hedonic slope with inelastic supply in the absence of externalities, $P_X^+ = -\frac{E_X}{E_P}$. P_X^+ is the response in prices that would hold housing demand constant if composition m did not change in response to changes in amenities, and corresponds to the slope that hedonic regressions try to estimate. The price schedule P^+, which takes composition as exogenously given, will be referred to as the counterfactual partial equilibrium price. Relative to P_X^+, there is an additional term in P_X^* if there are social externalities, i.e., $E_m \neq 0$, and equilibrium composition does depend on exogenous location characteristics, i.e., $m_X^* \neq 0$.

This leads us to the identification problem that will be discussed formally in the general case below: Knowledge about equilibrium schedules M^* and P^* does not allow us to identify the demand functions D^c, nor the slopes D_m^c, D_X^c, P_m^+ or P_X^+. The reason is that m^*, in the two type case, is functionally dependent on X. There never is independent variation of the two. Therefore, the slopes D_X^c and D_m^c can not be identified separately. If the partial equilibrium is unique, any equilibrium schedule (M^*, P^*) can be rationalized by a version of the model without social externalities, for instance by setting $D^c(X, m, P) = M^{c*}(X, m^*(X), P^*(X))$.

Equation 1.5 implies a "multiplier" effect in the sense that any immediate causal effect of amenities on composition, d_X, is amplified by a factor $\frac{1}{1-d_m}$. This factor is bigger than one iff $d_m > 0$, that is iff

$$\frac{D_m^1}{D^1} > \frac{D_m^2}{D^2}.$$

In particular, this is the case if the left hand side is positive and the right hand side is negative. This case could be described as homophilia, following Currarini, Jackson, and Pin (2009). Note however, that this could also hold under "hierarchical preferences" in the sense that both groups prefer, say, a higher share of group 1 but this preference is stronger among group 1 members. Conversely, in the case of "heterophilia," or

$$\frac{D_m^1}{D^1} < \frac{D_m^2}{D^2},$$

social externalities have a dampening effect on amenity variation. In this case they lead to a more integrated residential distribution. Finally, if social preferences are strong enough, it is also quite possible that there are unstable equilibria with $d_m > 1$, in which case there must be at least two more stable equilibria. This case is the one emphasized in discussions of "tipping" such as Card, Mas, and Rothstein (2008). In section 1.5.2, a test for multiplicity of equilibria is discussed. If $d_m > 1$, amenity or population changes might lead to a bifurcation where two equilibria merge and then disappear, the so called tipping of a neighborhood. Figure 1.3 illustrates this case.

1.3 The identification problem and solutions in the static model

This paper is about identification in sorting models with social externalities.[2] Parameters of interest in the model introduced in the last section are the demand schedule $D(X, M, P)$, its slopes (D_X, D_M, D_P), and in particular whether demand exhibits social externalities.

Definition 2 (Social externalities). *Demand is said to exhibit social externalities if $D_M \neq 0$.*

Related questions are whether landowners would be able to ask for different rents depending on neighborhood composition, holding amenities fixed, or what the marginal willingness to pay for neighborhood composition is.

It is useful to decompose the identification problem into several steps, going from the observable data distribution and regression slopes to equilibrium comparative statics to demand functions to preferences, as illustrated in figure 1.1, which also provides a roadmap through the following discussion. Objects further down in this figure are increasingly hard to identify, and not identifiable without identification of the previous steps. The problem of identification is one of mapping from observable data distributions into objects of interest, thus inverting the mappings from primitives into observables provided by the model assumptions. The structure of the problem displayed in figure 1.1 is similar for any setup with optimizing agents where outcomes are determined by some form of equilibrium.

The next subsection explores the relation between the demand and price schedules and the underlying distribution of household utility. It also introduces the notion of counterfactual partial equilibrium. The corresponding counterfactual price schedule, P^+, equates demand and supply of housing while not requiring equality of actual composition and composition entering location decisions, thus treating composition like an exogenous amenity.

Subsection 1.3.2 discusses the relationship between equilibrium comparative statics and demand functions. It exposes the fundamental identification problem in our model, which is due to the endogeneity of equilibrium $(M^*(X), P^*(X))$, and the implied degeneracy of the joint support of (X, M, P). Without further restrictions, demand slopes with respect to (X, M, P) are not identified, and neither are the slopes of the counterfactual price schedule P^+ with respect to X and M. In particular, both cross-section and panel data, even with experimental variation in location characteristics X, are uninformative about the presence of social externalities. The cause of the identification problem is the lack of independent variation of composition and other demand shifters. Any test for social externalities will have to "drive a wedge" between these two.

Subsections 1.3.3 and 1.3.4 develop positive identification results based on exclusion restrictions. Subsection 1.3.3 extends the idea of randomized subgroup treatment familiar

[2] A parameter is said to be identified if it can be written as a function of the distribution of the observable data.

from the peer effects literature. Subsection 1.3.4 suggests exclusion restrictions based on the spatial structure of interactions across neighborhoods, in an extended version of the model that allows for externalities across adjacent neighborhoods. The main results in these subsections map equilibrium schedules into demand functions under such exclusion restrictions. Corollaries to these results combine them with a mapping from regression slopes to equilibrium comparative statics, using the LATE representations of appendix 1.A.

1.3.1 Counterfactual partial equilibrium, price slopes and preferences

In this subsection, the relationship between demand slopes, slopes of equilibrium prices and household utility is explored. Since Rosen (1974), price slopes of the form P^*_X have often been used as estimates of household willingness to pay for X, which equals $-u_X/u_P$ in the notation of the present paper. In the context of discrete choice models, it becomes evident that price slopes are not necessarily equal to willingness to pay for infra-marginal households. Lemma 1 and its corollary 1 show this in the present, nonparametric, setup. They represent the price gradient as an appropriately weighted average of willingness to pay of marginal households. If, however, there is a continuum of similar outside options, all households in the neighborhood are marginal and have identical willingness to pay, as illustrated by lemma 2.

Due to the possible presence of social externalities, $u_M \neq 0$, price slopes may also deviate from willingness to pay for X for marginal households. Price changes P^*_X must compensate for the change in composition, M^*_X. This is made apparent by defining a notion of *counterfactual partial equilibrium*, as a point of reference. Counterfactual partial equilibrium gives the price P^+ that would prevail if M were determined exogenously and need not equal D. In Lemma 1, the comparison of P^*_X and P^+_X shows the bias in P^*_X relative to average marginal willingness to pay for X, P^+_X, due to externalities. This also suggests P^+_X and P^+_M as empirical objects of interest in their own right.

In section 1.2.1, partial sorting equilibrium was defined as the solution to equating total housing demand and supply as well as composition and type specific demand. Counterfactual equilibrium $(M^+(M, X), P^+(M, X))$ is defined as the solution to equating housing demand and supply, while the argument M to the demand functions is exogenously given and not necessarily equal to D:

Definition 3 (Counterfactual Partial Equilibrium)**.** *A counterfactual partial equilibrium (M^+, P^+) given X and M solves the $C+1$ equations*

$$M^+ = D(X, M, P^+) \tag{1.7}$$

$$S(P^+, X) = \sum_c M^{+c}. \tag{1.8}$$

Let $(M^+(X,M), P^+(X,M))$ denote the function mapping (X,M) into the counterfactual partial equilibrium given (X,M).

Lemma 1 (Price gradients and marginal households' utility)**.** *Under assumptions 1.2.1 and 1.2.3, the slope of total housing demand with respect to X is given by*

$$E_X = M^{tot} \cdot f^{u-u^o}(0) \cdot E[u_X | u = u^o],$$

where f^{u-u^o} denotes the density of $u - u^o$. Similarly for E_M, E_P, D^c_X, D^c_M, and D^c_P. Assume additionally that partial sorting equilibrium is unique or assume (M^*, P^*) is a differentiable selection from the set of partial equilibria, and assume $S_P = S_X = 0$. Then

$$P^+_X = -\frac{E[u_X | u = u^o]}{E[u_P | u = u^o]},$$

$$P^+_M = -\frac{E[u_M | u = u^o]}{E[u_P | u = u^o]},$$

and

$$P^*_X = P^+_X + P^+_M M^*_X = -\frac{E[u_X + u_M M^*_X | u = u^o]}{E[u_P | u = u^o]}.$$

Lemma 1 expresses the price gradients as ratios of average marginal utilities among marginal households. Since one can rewrite any ratio of averages as weighted average of ratios, the following corollary expresses the gradients as average willingness to pay for X, where the average is taken with respect to a reweighted distribution. The reweighting can be interpreted as a re-normalization of household utility to a constant marginal disutility of P, which implies a rescaling of the conditional density of marginal utilities among marginal households.

Corollary 1 (Price gradient as weighted average willingness to pay)**.** *Under the assumptions of lemma 1, if $u_P < 0$ for all households,*

$$P^+_X = \widetilde{E}\left[-\frac{u_X}{u_P} \middle| u = u^o \right],$$

where the expectation $\widetilde{E}$ is taken with respect to the density

$$f^{u_X, u_P | u - u^0}(u_X, u_P | 0) \cdot \frac{u_P}{E[u_P | u = u^0]}.$$

Similarly for P^+_M and P^*_X.

If, relative to lemma 1, we assume additionally that there is a continuum of alternative location choices, as in hedonic models, tighter characterizations of equilibrium prices and sorting follow. All households in a neighborhood become marginal and have the same marginal willingness to pay.

Lemma 2 (Hedonic gradient given continuum of outside options)**.** *Make assumptions 1.2.1 and 1.2.3. Assume additionally that u^o is bounded by the supremum of $u(X, M, P)$ over a set of outside options including an ϵ ball around X, and the corresponding equilibria $(M, P) \in (M^*(x), P^*(x))$:*

$$u^o \geq \sup_{\substack{x:||x-X||<\epsilon \\ (M,P)\in(M*(x),P^*(x))}} u(x, M, P)$$

Then

$$P_X^* = -\frac{u_X + u_M M_X^*}{u_P} \tag{1.9}$$

for all households choosing a neighborhood with given X.

This subsection concludes with a lemma characterizing local comparative statics of average reservation prices among households in the neighborhood in terms of average marginal willingness to pay of *all* households in the neighborhood. This lemma will be central in the characterization of prices in the dynamic model presented later in this paper. In this dynamic model, landowners will extract all surplus value generated by a match to a tenant, so that rents are equal to household reservation prices. Formally, define the reservation price of a household for living in the neighborhood, given X and M, as

$$P^{res} = P^{res}(X, M) := \sup\{P : u(X, M, P) \geq u^o\}.$$

Lemma 3 characterizes the dependence on X of average reservation prices, *conditional* on locating in the neighborhood, i.e., conditional on $P^{res} \geq P^*$. Changes in X can, in principle, influence average reservation prices in three ways: directly, through their effect on M, and through a reshuffling of residents. The last may matter if, under the new X, households with higher reservation prices crowd out the initial residents. The central message of lemma 3 is that this effect is not of first-order importance if housing supply is inelastic, so that the number of households in the neighborhood is constant, or if housing demand is elastic, so that all households in the neighborhood have reservation prices equal to P^*.

Lemma 3 (Comparative statics of average reservation prices)**.** *Assume partial sorting equilibrium is unique or assume (M^*, P^*) is a differentiable selection from the set of partial equilibria. Under assumptions 1.2.1 and 1.2.3*

$$\begin{aligned} \frac{\partial}{\partial X} E\left[P^{res}|P^{res} \geq P^*\right] &= E\left[P_X^{res} + P_M^{res} M_X^* | P^{res} \geq P^*\right] \\ &- \left(\frac{\partial}{\partial X} \log \mathbb{P}\left(P^{res} \geq P^*\right)\right) \cdot \left(E\left[P^{res}|P^{res} \geq P^*\right] - P^*\right) \end{aligned} \tag{1.10}$$

where $P_X^{res} = -u_X/u_P$ and $P_M^{res} = -u_M/u_P$. In particular, if housing supply is price inelastic and constant in X, i.e., $S_P = S_X = 0$, or if all households in the neighborhood are marginal, i.e., $E[P^{res}|P^{res} \geq P^*] = P^*$, then

$$\frac{\partial}{\partial X} E\left[P^{res}|P^{res} \geq P^*\right] = E\left[P_X^{res} + P_M^{res} M_X^* | P^{res} \geq P^*\right]. \tag{1.11}$$

1.3.2 Degenerate support and non-identification

Under assumptions 1.2.1 and 1.2.2, even in the full observability case, any cross-section or panel dataset can at best identify the joint distribution of (X, M, P). This joint distribution can be decomposed into the exogenous distribution of X and the conditional distribution of M and P given X. By assumption, the latter has its support on the set of partial sorting equilibria given X, $(M^*(X), P^*(X))$. The present section discusses what knowledge of $(M^*(X), P^*(X))$ allows us to learn about the demand schedule $D(X, M, P)$, its derivatives D_X, D_M and D_P, the counterfactual price schedule $P^+(X, M)$ defined in the previous subsection, and underlying utilities $u(X, M, P)$.

This subsection will abstract from any identification problems due to partial observability of X or the types c which might obstruct identification of $(M^*(X), P^*(X))$ itself. Such lack of observability underlies omitted variable bias problems and necessitates the search for exogenous variation of X. The positive identification results in the following sections will first assume knowledge of $(M^*(X), P^*(X))$, and express parameters of interest in terms of derivatives of these equilibrium schedules. They will then extend these results to the partial observability case with exogenous variation of components of X, where (weighted averages of) these derivatives can be recovered by regression.

The following relationships hold, by definition, between partial sorting equilibrium, counterfactual partial equilibrium and the demand schedule:

$$(M^*(X), P^*(X)) = (M^+(X, M^*(X)), P^+(X, M^*(X))) \tag{1.12}$$
$$D(X, M, P^+(X, M)) = M^+(X, M) \tag{1.13}$$
$$D(X, M^*(X), P^*(X)) = M^*(X) \tag{1.14}$$

Proposition 2 ((Non)identification)**.** *Make assumptions 1.2.1 and 1.2.2, and consider the full observability case. Then:*
$D(X, M, P)$ is not identified for $(M, P) \notin (M^*(X), P^*(X))$.
$(M^+(X, M), P^+(X, M))$ is not identified for $M \notin M^*(X)$.
But:
$D(X, M, P)$ is identified on the joint support of (X, M, P).
$(M^+(X, M), P^+(X, M))$ is identified on the joint support of (X, M).

The proof of this proposition is quite straightforward. The focus on the full observability case makes evident that the underlying reason of the problem is lack of support, not lack of observability or randomized variation in X.

There is a parallel between this identification problem and the classic simultaneity problem in identifying price elasticities, as well as the reflection problem, as Manski (1993) named it, in the identification of models with endogenous peer effects. In all these problems, an endogenous equilibrium outcome serves as argument to some structural relationship. There is no (continuous) variation of the equilibrium outcome conditional on the other arguments of the same relationship, at least without further exclusion restrictions.

The problem can be restated in terms of the slopes of demand and the counterfactual price schedule, as in the following lemma. Lemma 4 can be understood as showing that the identification problem is a "multicollinearity" problem. There is no independent variation of X and $(M^*(X), P^*(X))$.

Lemma 4 (Identification of slopes)**.** *Make assumptions 1.2.1 and 1.2.2 and consider the full observability case. Assume partial sorting equilibrium is unique or assume (M^*, P^*) is a differentiable selection from the set of partial equilibria. Linear combinations of the demand slopes are identified as*

$$D_X + D_M M^*_X + D_P P^*_X = M^*_X. \tag{1.15}$$

Linear combinations of the counterfactual price gradient are identified as

$$P^+_X + P^+_M M^*_X = P^*_X. \tag{1.16}$$

No other linear combinations of (D_X, D_M, D_P) and (P^+_X, P^+_M) are identified.

This subsection concludes with a discussion of two apparent solutions to the identification problem. First, there might be a temptation to "break the multicollinearity problem" by assuming functional form restrictions, such that slopes are identified using the curvature or higher order properties of the equilibrium schedules (M^*, P^*). However, *any* result on social externalities can be rationalized with any dataset using the appropriate functional form assumption, as the following lemma shows. In the proof of the lemma, the derivatives D_M and D_P are chosen in a data-independent way. Arbitrary generalizations of the counterexample used can be constructed by choosing them as a function of the data.

Lemma 5 (Spurious identification by functional form assumptions)**.** *Make assumptions 1.2.1 and 1.2.2 and consider the full observability case, and assume that partial equilibrium is unique.*

Fix an arbitrary $\mathscr{C} \times \mathscr{C}$ matrix A and a $\mathscr{C}$ vector B. Then there exists a just-identified model for $D(X, M, P)$ such that $D_M \equiv A$ and $D_P \equiv B$ for the unique D in the model such that $D(X, M^*(X), P^*(X)) = M^*(X)$ for all X.

Similarly, fixing again a $\mathscr{C}$ vector B, there exists a just-identified model for P^+, such that $P^+_M = B$ for the unique P^+ in the model such that $P^+(X, M^*(X)) = P^*(X)$.

Another temptation might be to search for exogenous variation in X and control for composition M in order to estimate the counterfactual price gradient P_X^+, as in Bayer, Ferreira, and McMillan (2007). However, by endogeneity of M, such conditioning on M introduces dependence between unconditionally independent components of X. Formally, make assumptions 1.2.1 and 1.2.2 and consider the partial observability case, where $X^1 \perp \epsilon$, and assume that partial sorting equilibrium is unique. Suppose we were to use the regression slope

$$\frac{\partial}{\partial x^1} E\left[P | X^1 = x^1, M\right]$$

as an estimator for

$$E\left[P_{X^1}^+ | X^1 = x^1, M\right].$$

This would be valid if $X^1 \perp \epsilon$ *conditional* on M. However, the conditional expectation is taken over the distribution of ϵ on the subspace defined by $M^*(X^1, \epsilon) = M$, which changes as a function of X^1. In particular, if $dim(\epsilon)$ is equal to $dim(M)$, then ϵ is function of X^1 conditional on M, and the bias in the regression conditioning on M relative to $P_{X^1}^+$ is equal to

$$-P_\epsilon^+ \cdot \frac{M_{X^1}^*}{M_\epsilon^*}.$$

1.3.3 Exclusion restrictions based on subgroup demand shifters

In the last subsection, it was argued that D, and in particular the slopes (D_X, D_M, D_P), are unidentifiable due to functional dependence of (M^*, P^*) and X. Holding X constant, there is no variation in M that allows us to identify the effect of M on D. After these negative results, the rest of the theoretical development in this paper is dedicated to positive identification results based on additional model restrictions and extensions.

The first identification result is based on the assumption that some components of X are *excluded* from the demand of some subgroup, or from demand of all groups but not from supply. Under such exclusion restrictions, variation in the components of X that are excluded generates variation in M and P that is not functionally dependent on the relevant arguments of demand. This idea underlies proposition 3, which expresses demand slopes in terms of equilibrium slopes under exclusion restrictions. If there additionally is a source of variation of these observed components of X that is statistically independent of variation in the unobserved components ϵ, equilibrium slopes, and thereby demand slopes, can be identified by regression, as in corollary 2.

Exclusion restrictions of the form $D_{X^1}^1 = 0$ are the natural analogon to using (randomized) subgroup treatment as a source of identification of peer effects. Compare for instance the general discussion in Moffitt (2001), and Duflo and Saez (2003). The latter provided information about pension plans to a random subset of employees in a random subset of departments of a university, and studied the effect on the behavior of other employees of the same departments. Exclusion restrictions of the form $D_{X^2} = 0$ but $S_{X^2} \neq 0$ correspond to the classic use of supply side instruments to identify the price elasticity of demand.

Exposition will be simplified from now on by considering the two-type model of section 1.2.2. In that case $M^{*1}_X = D^1_X + D^1_m m^*_X + D^1_P P^*_X$. Similar arguments hold more generally, however.

Proposition 3 (Subgroup identification)**.** *Make assumptions 1.2.1 and 1.2.4. Assume that $D^1_{X^1} = 0$ but $D^2_{X^1} \neq 0$ for some component X^1 of X, and $D^1 \neq 0$. Then*

$$D^1_m = \frac{1}{m^*_{X^1}} \left(M^{*1}_{X^1} - D^1_P P^*_{X^1} \right). \tag{1.17}$$

Assume additionally $D^1_{X^2} = D^2_{X^2} = 0$ but $S_{X^2} \neq 0$. Then

$$D^1_m = \frac{1}{m^*_{X^1}} \left(M^{*1}_{X^1} - \frac{M^{*1}_{X^2}}{P^*_{X^2}} P^*_{X^1} \right). \tag{1.18}$$

Proposition 3 expresses slopes of demand D in terms of equilibrium slopes. The latter are not observable in the realistic partial observability case, however. To translate this result into one that can be used in practice, we have to substitute partial derivatives by estimable slopes. This is what corollary 2 does below.

Without any restrictions on functional form, assuming smoothness and appropriate exogeneity, linear OLS and IV regressions recover weighted average derivatives of the structural functions of interest, as shown in appendix 1.A. If the setup is restricted to linear random coefficients, weighted averages of the random slopes are identified by IV. If, finally, functional forms are restricted to be linear in the arguments of interest, the corresponding partial derivatives are constant.

Corollary 2. *Make assumptions 1.2.1, 1.2.2 in the partial observability case, and 1.2.4. Assume we observe a two period panel of locations, with changes in exogenous demand shifters, composition, and prices, (dX, dm, dP).*
Assume $\log D^1$ is linear in m and $\log P$. Assume dX^1 is uncorrelated with changes in D^1 induced by the other components of dX, and similarly for dX^2.
Denote by $\beta^{\log M^1, X^1}$ the expectation of the OLS regression coefficient of $d \log M^1$ on dX^1 etc.
If $D^1_{X^1} = 0$ and $D^2_{X^1} \neq 0$, and $\partial \log D^1 / \partial \log P \in [\eta^{min}, \eta^{max}]$, then

$$\frac{\partial \log D^1}{\partial m} \in \frac{1}{\beta^{m,X^1}} \left(\beta^{\log M^1, X^1} - [\eta^{min}, \eta^{max}] \cdot \beta^{\log P, X^1} \right). \tag{1.19}$$

If additionally $D^1_{X^2} = D^2_{X^2} = 0$ but $S_{X^2} \neq 0$,

$$\frac{\partial \log D^1}{\partial m} = \frac{1}{\beta^{m,X^1}} \left(\beta^{\log M^1, X^1} - \frac{\beta^{\log M^1, X^2}}{\beta^{\log P, X^2}} \cdot \beta^{\log P, X^1} \right). \tag{1.20}$$

If the assumption on linearity of $\log D^1$ is dropped, equation 1.19 still holds if we replace $\frac{\partial \log D^1}{\partial m}$ by the weighted average $E\left[\frac{\partial \log D^1}{\partial m} \cdot \omega\right]$, where the weight is given by

$$\omega = \frac{m_t(dX^1 - E[dX^1])}{E[m_t(dX^1 - E[dX^1])]},$$

and the expectations are taken over the product distribution of the crosssectional distribution over the i and the uniform distribution over the time interval $[1, 2]$.

1.3.4 Exclusion restrictions based on spatial structure

In the model considered so far, in the absence of exclusion restrictions there is no independent variation of X and $(M^*(X), P^*(X))$. Put differently, the arguments determining demand and supply and the arguments determining equilibrium outcomes are exactly the same. The identification results in this subsection, as well as those using dynamics which will be developed later, are based on model extensions which generate variation in equilibrium composition conditional on *all* relevant exogenous arguments of demand and supply, X.

In assumption 1.2.1, the possibility of cross-neighborhood externalities has been ignored. This subsection extends the baseline model of assumption 1.2.1 by adding a spatial structure. It is assumed that the relevant composition variable $\widetilde{m}$ affecting demand is a weighted average of the composition of adjacent neighborhoods, m, and similarly $\widetilde{X}$ is a weighted average of the Xs of adjacent neighborhoods. The weights are given by a matrix $\mathbf{G}$, where the (k, l)th entry of $\mathbf{G}$ describes the strength of externalities from neighborhood l to neighborhood k. Under these conditions, too, variation in the X of adjacent neighborhoods always induces variation in $\widetilde{m}$, thus not allowing for separate identification of the effects of the two. However, there is a "propagation" effect of composition along chains of adjacent neighborhoods. A change in X in one location changes $\widetilde{m}$ in adjacent locations, which in turn affects demand and composition in their adjacent neighborhoods etc. Thus, variation in X in non-adjacent neighborhoods is excluded from demand in a location, yet might generate variation in m and $\widetilde{m}$. Proposition 4 formalizes this idea.

The simplifications of the two-type model of assumption 1.2.4 are made again, but the results generalize.

Assumption 1.3.1 (Cross neighborhood interactions)**.** *There are $\mathcal{N}$ neighborhoods. $\mathbf{G}$ is a $\mathcal{N} \times \mathcal{N}$ matrix with non-negative entries summing to one in each row and with positive*

diagonal entries.
Let $\mathbf{m}$ *be the* $\mathcal{N}$ *vector of* m *for all neighborhoods,* $\widetilde{\mathbf{m}} = \mathbf{G}\mathbf{m}$ *the vector of* $\mathbf{G}$ *weighted averages of* m*, similarly for* $\mathbf{X}$ *and* $\widetilde{\mathbf{X}}$*.*
Then, for each neighborhood, with $\widetilde{X}, \widetilde{m}$ *being the neighborhood specific entries of the corresponding vectors,*

$$d(\widetilde{m}, \widetilde{X}) = m \tag{1.21}$$
$$E(P^*, \widetilde{m}, \widetilde{X}) = S(P^*, X). \tag{1.22}$$

Note that d is constant in components of $(\mathbf{m}, \mathbf{X})$ with a corresponding zero entry in $\mathbf{G}$. In this sense, this is still a partial equilibrium model which does not consider the effect on outside options of changes in remote locations.

In immediate generalization of definition 3 , let $P^+(\widetilde{m}, \widetilde{X}, X)$ denote the equilibrium price that would prevail if $(\widetilde{m}, \widetilde{X}, X)$ were to characterize a neighborhood.

Proposition 4 (Spatial identification)**.** *Make assumption 1.3.1, and assume* $S_X = 0$ *and* $0 < d_{\widetilde{m}} < 1$ *as well as* $d_{\widetilde{X}} \neq 0$*, for all neighborhoods.*
Fix two neighborhoods k *and* l*. If the* k,l*th entry of* $\mathbf{G}$ *equals* 0 *and there exists a power* $j > 1$ *of* $\mathbf{G}$*, such that the* k,l*th entry of* $\mathbf{G}^j$ *is not equal to zero*[3]*, then:*

$$d_{\widetilde{m}}\left(\widetilde{m}^k, \widetilde{X}^k\right) = \frac{m^k_{X^l}}{\widetilde{m}^k_{X^l}}, \tag{1.23}$$

$$P^+_{\widetilde{m}}\left(\widetilde{m}^k, \widetilde{X}^k, X^k\right) = \frac{P^k_{X^l}}{\widetilde{m}^k_{X^l}}, \tag{1.24}$$

and
$$D^c_{\widetilde{m}}\left(\widetilde{m}^k, \widetilde{X}^k, P^k\right) = \frac{1}{\widetilde{m}^k_{X^l}}\left(M^{*c,k}_{X^l} - D^{c,k}_P P^k_{X^l}\right). \tag{1.25}$$

Again, this is a theoretical identification result in terms of equilibrium slopes. It is translated by the following corollary into an implementable one, using OLS slopes. As before, the results are first stated imposing linearity assumptions in order to facilitate exposition, and then the general case follows. In this corollary, an instrument X^f is constructed as an average of changes in X in non-adjacent neighborhoods.

Corollary 3. *Make assumptions 1.3.1 and 1.2.2, and assume* $S_{X^1} = 0$ *and* $0 < d_{\widetilde{m}} < 1$ *as well as* $d_{\widetilde{X^1}} \neq 0$*, for all neighborhoods.*
Assume we observe a two period panel of locations, with changes in exogenous demand shifters, composition, and prices, (dX, dm, dP)*.*
For each neighborhood k *let* dX^f *be an average of* dX^1 *over a set of neighborhoods* l*, such*

[3]Note that $\mathbf{G}^j$ here denotes $\mathbf{G} \cdot \ldots \cdot \mathbf{G}$, i.e., j is a power, not an index.

that the k, lth entry of $\mathbf{G}$ equals 0 and there exists a power $j > 1$ of $\mathbf{G}$, such that the k, lth entry of $\mathbf{G}^j$ is not equal to zero.
Assume that d, $\log P^+$ and $\log D$ are linear in $\widetilde{m}$. Assume that dX^f is uncorrelated with changes in D and P^+ which are induced by the other components of dX.
Denote by β^{m,X^f} the expectation of the OLS regression coefficient of dm on dX^f etc.
Then

$$d_{\widetilde{m}} = \frac{\beta^{m,X^f}}{\beta^{\widetilde{m},X^f}}, \tag{1.26}$$

$$(\log P^+)_{\widetilde{m}} = \frac{\beta^{\log P,X^f}}{\beta^{\widetilde{m},X^f}}, \tag{1.27}$$

and

$$(\log D^c)_{\widetilde{m}} \in \frac{1}{\beta^{\widetilde{m},X^f}} \left(\beta^{\log M^c,X^f} - [\eta^{min}, \eta^{max}] \cdot \beta^{\log P,X^f} \right). \tag{1.28}$$

In the last equation, $[\eta^{min}, \eta^{max}]$ are bounds on $D_P^{c,k}$ as in corollary 2. If the linearity assumptions are dropped, these equations still hold if we replace the partial derivatives by weighted averages thereof, as in corollary 2.

1.4 A dynamic extension of the static model with search frictions

The model discussed so far is static. We can think of it as describing an economy with negligible search frictions in which equilibrium is instantaneously achieved. Alternatively, it could be considered as describing the long run steady state of an economy with frictions. However, explicitly considering dynamics and frictions reveals additional sources of identification.

A well established literature in labor economics discusses the dynamics and comparative statics of unemployment and wages in models with search frictions. Its central presumption is that finding a job or an employee takes time and unemployment is due to this search time. Pissarides (2000) provides an extensive overview of this literature. Wheaton (1990) applies the insights of this literature to the housing market. The focus of either of these is the relationship between vacancies (unemployment) and prices. Wheaton (1990) in particular models housing vacancies as corresponding to the search time of households who decided to move due to lifecylce events (shocks), found another place and now attempt to sell their old home. The present section extends the basic sorting model of section 1.2.1 using similar techniques as these papers.

Relative to the static model described by assumptions 1.2.1 and 1.2.3, the main extensions in the dynamic model are as follows. There is an explicit, continuous time dimension, and exogenous location characteristics X can change over time. Households that would like to move to a different neighborhood are subject to search frictions. If they decide to search for a new home, offers arrive at Poisson rate λ. Similarly, owners of vacant units have to search for tenants and find them at rate μ. Households are maximizing expected discounted utility, and make their search decisions in a forward looking way. Due to search frictions, composition M changes continuously over time and only reacts with delay to shocks in X. Finally, once a match is formed between homeowner and household, they are in a situation of bilateral monopoly: By breaking the match they both would have to search again, and thereby incur a loss of utility. Therefore, they have to negotiate over the division of the surplus, and rents are match-specific.

The purpose of this extension is twofold. First, the delayed adjustment of composition M to changes in exogenous characteristics X generates independent variation between X and M, contrary to the static case where M is essentially a function of X. This allows, under certain conditions, to separately identify household willingness to pay for X and for M. Second, the dynamic structure provides a connection between multiplicity of equilibria in the static sense, as mentioned in section 1.2.2, and multiplicity of equilibria in a dynamic sense. A test for the latter will be constructed below.

This section presents a search model of the *rental* market for housing. Considering homeownership would add the additional complication of housing being an asset in addition to being a consumption good. Under complete financial markets, the results derived for the rental market of housing immediately extend to the more general case however, as will be discussed briefly at the end of this section.

For simplicity of notation, household and time superscripts are mostly dropped. As before, we consider one fixed neighborhood.

Assumption 1.4.1 (The local economy, dynamic setup)**.**

- *There are $\mathscr{C}$ types of households, $c = 1, \ldots, \mathscr{C}$.*
- *Households can be in one of four states: Living in the neighborhood and not searching, living in the neighborhood and searching for a place outside, living outside and searching for a place in the neighborhood, living outside and not searching for a place in the neighborhood.*
- *Housing units can be in one of two states, vacant or occupied by one household.*
- *A neighborhood, at each point of time $t \in \mathbb{R}$, is characterized by*
    1. *the number of households of each type living in the neighborhood, $M = (M^1, \ldots, M^{\mathscr{C}})$*
    2. *an exogenous vector X of* all *other location characteristics and factors influencing demand or supply.*
- *The time paths of X and M are piecewise differentiable.*

- *There is a match specific rental price P for each match between a unit and a household.*
- *Households living outside searching for a place in the neighborhood, or living in the neighborhood and searching for a place outside, find a match at rate λ. Vacancies are matched to a household at rate μ. These rates can vary over time but are constant across households and units.*
- *Vacant units and searching households are matched uniformly at random.*

Assumption 1.4.2 (The household problem)**.**

- *Households are characterized by their type c, their flow utility $u(X, M, P)$ of living in the given neighborhood, their discount rate r and the value of their outside option V^o. Except for type, all of these may depend on time t. V^o does not depend on X, M.*
- *Households have the choice between searching or not. They do so to maximize their expected discounted utility.*
- *There are no costs of search.*
- *There is a continuum of households of total mass M^{tot} in the economy.*

Denote the value of living in the given neighborhood by $V = \max(V^s, V^{ns})$, where V^s and V^{ns} are the values of searching and not searching, respectively. Denote the time derivative of V by $\dot{V}$. The value functions are to be understood as conditional expectations, given the information set at time t, as are their time derivatives. Assumptions 1.4.1 and 1.4.2 imply

$$rV^s = u(X, M, P) + \lambda(V^o - V) + \dot{V} \tag{1.29}$$

and

$$rV^{ns} = u(X, M, P) + \dot{V}. \tag{1.30}$$

A household living in the neighborhood wants to search for a place outside if and only if $V^o > V$, and V satisfies

$$(r + \lambda)V = u(X, M, P) + \lambda \max(V^o, V) + \dot{V}. \tag{1.31}$$

Let us now turn to the landowners.

Assumption 1.4.3 (The landowner's problem)**.**

- *Landowners are risk neutral, maximize their discounted stream of incomes and are otherwise indifferent about the residents of their units. Their discount rate is denoted by r.*

- *Owners of vacant units can and do search for renters among the pool of households that search for a home in the given neighborhood.*

Denote the value of an occupied unit by $W = \max(W^s, W^{ns})$ where W^s and W^{ns} are the values of the unit when the renting household is searching and not searching, respectively. Denote the value of a vacant unit by W^v. Under assumptions 1.4.1 and 1.4.3, the value of an occupied unit where the renter is not searching for a new place is characterized by

$$rW^{ns} = P + \dot{W}. \tag{1.32}$$

The value of an occupied unit with a searching renter is

$$rW^s = P + \lambda(W^v - W) + \dot{W}. \tag{1.33}$$

The value of a vacant unit satisfies, finally,

$$rW^v = \mu(W^{new} - W^v) + \dot{W^v}. \tag{1.34}$$

Note that the value of a match to the landowner is household specific, and therefore W^{new}, the expected value of a match with a new renter, is in general different from the value of the current match, W. These values describe the expected discounted revenue for a given unit.

Once a potential renter and a landowner holding a vacant unit meet, they have to negotiate a rental contract.

Assumption 1.4.4 (Rent determination)**.**

- *The contract specifies rental payments. Contracts can be continuously renegotiated.*
- *Each of the contract parties can unilaterally decide at any time to end the contract and initiate search of the renting household, where this decision is reversible. The renter can not be evicted before she has found a new place, but can be committed to search.*
- *Rents are determined by Nash bargaining over the division of the surplus relative to the outside option of searching (not searching)*[4]*, that is, current rents maximize $(V^{ns} - V^s)^\beta (W^{ns} - W^s)^{(1-\beta)}$, where $\beta \in [0, 1]$ is the relative bargaining power of tenants.*

We have $(V^{ns} - V^s)^\beta (W^{ns} - W^s)^{(1-\beta)} = \lambda (V - V^o)^\beta (W - W^v)^{(1-\beta)}$, and the first order condition for the solution to Nash bargaining is

$$(V - V^o) = \beta \left[-u_P (W - W^v) + (V - V^o) \right]. \tag{1.35}$$

[4]Note that the potential outside option of searching for a different home in the same neighborhood is always strictly dominated. It leaves the household indifferent and the landowner strictly worse off, since she foregos rents while searching for a new tenant.

If there exists a price P such that both $V > V^o$ and $W > W^v$, then these conditions must hold under the bargaining solution, no matter what β is. Search happens if and only if there is no such P, and the search decision is always consensual. This is a feature of any privately efficient contract. For households living outside the neighborhood, the decision to search in this given neighborhood is independent of β. Our last assumption pins down bargaining power β:

Assumption 1.4.5 (Bargaining power). *All bargaining power lies with the landowners, i.e., $\beta = 0$.*

This assumption allows for a clean characterization of price dynamics, since all changes in household utility will be compensated by price changes. If we were to drop assumption 1.4.5, only a fraction $1 - \beta$ of utility changes would be compensated by price changes.

Readers familiar with the literature on search models of unemployment will notice a central feature of these models missing in the assumptions just stated. Neither λ and μ, nor housing supply, are explicitly modeled. Common practice in the literature, for instance in the models reviewed in Pissarides (2000), is to assume a matching technology where the rates λ and μ are a function of the ratio of searching workers (households) to vacancies, and there is free entry of firms (landlords). This is crucial in the context of search models of the labor market that attempt to explain unemployment and vacancy rates. It also has important implications for the speed of adjustment following shocks. As neither vacancies nor variation in the speed of adjustment are of central concern in the present context, exposition is simplified by not explicitly modeling intertemporal variation in λ and μ.

1.4.1 Implications of the model

The rest of this section develops some central properties of the model described by assumptions 1.4.1 through 1.4.5. First, the dynamics of composition are shown to follow a differential equation of the form $\dot{M} = \lambda \cdot (D - M)$, where D is a dynamic generalization of the demand function. If we specialize this to the two type case and consider discrete intervals of time, then changes in composition m follow the difference equation $\Delta m := m^1 - m^0 = \kappa \cdot (d - m^0)$, where κ is a rate derived from λ.

Next, the reaction of prices to shocks in X is studied. In the short run, because of search frictions, M is not affected by such shocks. Under assumptions 1.4.4 and 1.4.5, rents immediately adjust so that all surplus of the match is appropriated by landlords, and changes in rents correspond to household willingness to pay for changes in X. In the long run, M does change however. Rental changes occurring with delay correspond to household willingness to pay for this change in M.

Finally, the relationship between this dynamic model and the static model studied so far is clarified. First, the long run comparative statics of M, as a function of X, are the same as those of an appropriately defined corresponding static model. Second, if search frictions are low and/or discount rates high, then the dynamic model is approximated by the static model in a sense made precise below.

These results will be synthesized in proposition 5 in the next section. This proposition describes how the dynamic structure of responses in prices and composition to changes in exogenous characteristics allows us to identify average willingness to pay for both X and m among households living in the neighborhood. Section 1.5 will also characterize the relationship between quantile regressions of Δm on m and the function d, and will discuss how tests for multiplicity of roots of $d - m$ in m can be constructed.

The dynamics of composition

Under assumptions 1.4.1 and 1.4.2 we have, at each point in time, a set of households of each type c that want to move out of the neighborhood, because for them $V^o > V$, and a corresponding outflow. Similarly, at each point in time there is a set of households of each type c that want to move into the neighborhood, because for them $V \geq V^o$, and a corresponding inflow. The net flow will equal λ times the difference between the number of households that *want* to live in the neighborhood, i.e., for which $V \geq V^o$, and those that *do* live in the neighborhood, M. This motivates the following definition of demand D in the dynamic model.

Definition 4 (Demand in the dynamic model)**.** *Denote by D^c the mass of households of type c for which $V \geq V^o$:*

$$D^c = M^{tot} \cdot \mathbb{P}(V \geq V^o, c)$$

Lemma 6 (Dynamics of composition)**.** *Make assumptions 1.4.1, 1.4.2, 1.4.3 and 1.4.4. Then*

$$\dot{M} = \lambda \cdot (D - M), \tag{1.36}$$

where $\dot{M}$ is the expected time derivative of M.

The following result specializes to a two-type model and describes changes of composition m over discrete time intervals.

Lemma 7 (Dynamics of composition in the two-type model)**.** *Make the assumptions of lemma 6, as well as assumption 1.2.4. Then the change in m from time 0 to time 1 is given by*

$$\Delta m := m^1 - m^0 = \kappa \cdot (d(m, X) - m^0) \tag{1.37}$$

for some m, X at a time in $[0, 1]$, where

$$\kappa = 1 - \exp\left(- \int_0^1 \lambda \cdot \frac{D^1 + D^2}{M^1 + M^2} ds \right) > 0. \tag{1.38}$$

The determination of prices

Under assumption 1.4.5, the landowner appropriates all surplus from the match, and equation 1.35 implies that the participation constraint for the renter is binding at all times, i.e., $V = V^o$. By equation 1.30, the household specific rental price is then determined by

$$u(X, M, P) = rV^o - \dot{V}^o, \tag{1.39}$$

where X, M, and V^o are given to the household and landowner. This implies in turn that changes in rents must directly reflect willingness to pay for changes in X and M, for any household that lives in the neighborhood. This is reflected in the following lemmas 8 and 9. Lemma 8 additionally uses the fact that composition M is constant in the short run.

Lemma 8 (Short run comparative statics of prices)**.** *Make assumptions 1.4.1, 1.4.2, 1.4.3, 1.4.4, specifying the dynamic model, and 1.4.5 on bargaining power.*
Assume that $X = x$ before time 0, $X = x + \xi$ for a jump ξ after time 0, and (u, V^o) is continuously differentiable with respect to time for all households.
Then $\frac{\partial}{\partial \xi} \lim_{t \to 0^+} E[P] = E\left[-\frac{u_X}{u_P}\right]$ where the expectation is taken over the set of households living in the neighborhood at time $t = 0$.

We recover short-run comparative statics of prices in response to changes in X and M which look similar to the ones in the static model in the absence of social externalities and with inelastic housing supply. In the static model, the neighborhood rental gradient P_X^+ equals the average marginal willingness to pay of marginal households, according to corollary 1, whereas here the match specific rent gradient P_X equals the marginal willingness to pay of any given household. In the static case, marginal households had to be kept indifferent by changes in X and P^* for demand to be constant. In the present case, all households have to be kept indifferent by changes in X and P, since by the assumption on bargaining power all households are marginal, in the sense that their utility is equal to their reservation utility.

As households re-sort across neighborhoods, however, prices adjust further for two reasons. First, holding outside options as well as X and M constant, some households will want to move in which have a willingness to pay for the given bundle (X, M) which is higher than the willingness to pay of the current residents. Second, composition M will adjust over time, and influence the households' valuation of the neighborhood. If housing supply is constant or all households are marginal, the first reason can be ignored to first order, however. This follows from lemma 3. As a consequence, long run effects of changes in X on rents P reflect the sum of the willingness to pay for X and of the willingness to pay for the change in M induced by X.

Lemma 9 (Long run comparative statics of prices)**.** *Make assumptions 1.4.1, 1.4.2, 1.4.3, 1.4.4, specifying the dynamic model, and 1.4.5 on bargaining power. Assume that housing supply is constant or all households are marginal.*

Assume that $X = x$ before time 0, $X = x + \xi$ for a jump ξ after time 0, and (u, V^o) is constant for all households. Denote $M^{lr} = lim_{t\to\infty} M$, where it is assumed that this limit exists.
Then $\frac{\partial}{\partial \xi} \lim_{t\to\infty} E[P] = E\left[-\frac{u_X + u_M M^{lr}_\xi}{u_P}\right]$, where the expectation is taken over the set of households living in the neighborhood at time $t = 0$.

Demand in the dynamic and the static model
D, as given by definition 4, equals the number of households for which $V \geq V^o$. In the static model, under assumption 1.2.3, D was equal to the number of households for which $u \geq u^o$. How do these notions of demand relate to each other? To connect the dynamic model to our discussion of the static model, the following definition is useful. It derives a static model from the given dynamic model. Equilibrium prices in the static model correspond to cut-off prices, below which landlords do not accept a tenant in the dynamic model in steady state. The utility of households' outside option, u^o, is implicitly given by V^o. Corresponding static demand, finally, is equal to the mass of households for which flow utility u is bigger than outside utility u^o. As shown in lemma 10, the static model defined in this way describes the long run comparative statics of composition in the dynamic model. Lemma 11 implies that it also approximates the short run behavior of the dynamic model in the case of low search frictions or high discount rates.

Definition 5 (The corresponding static model)**.** *Under assumptions 1.4.1, 1.4.2, 1.4.3, 1.4.4 and 1.4.5, the corresponding static model is defined as follows: Let $u^o := rV^o - \dot{V}^o$. Let $P^{res} = \max\{P : u(X, M, P) \geq u^o\}$ be the reservation price for each household. Let P^* be the "cut-off" price below which landowners in the dynamic model would not accept a tenant in steady state. As will be shown, this cut-off is given by*

$$P^* = \frac{r\mu}{r + \mu} E^s[P^{res} | P^{res} > P^*], \tag{1.40}$$

where the expectation E^s is taken over the set of households searching for a place in the neighborhood. This equation implicitly defines the corresponding static housing supply.
The corresponding static demand of type c is equal to

$$\widetilde{D}^c = M^{tot} \cdot \mathbb{P}(u(X, M, P^*) > u^o | C = c) = M^{tot} \cdot \mathbb{P}(P^{res} > P^* | C = c) \tag{1.41}$$

for all c. Let $\widetilde{E} = \sum_c \widetilde{D}^c$. Denote the equilibrium (set) of this corresponding static model by (M^*, P^*).

Lemma 10 (Long run comparative statics of composition)**.** *Make assumptions 1.4.1, 1.4.2, 1.4.3, 1.4.4, specifying the dynamic model, and 1.4.5 on bargaining power.*
Assume that λ is uniformly bounded away from zero for $t > 0$, and that X and (u, u^o) is constant over time for all households.

If $M^{lr} := \lim_{t\to\infty} M$ exists, then $M^{lr} \in M^*$, i.e., composition converges to an equilibrium composition of the corresponding static model.

Lemma 11 (Low-friction limit of the dynamic model)**.** *Make assumptions 1.4.1, 1.4.2, 1.4.3 and 1.4.4. Define u^o as $u^o = (r+\lambda)V^o - \dot{V}^o - \lambda \max(V, V^o)$. Assume u and u^o are continuous in time and bounded. Then, for V, V^o, u and u^o evaluated at time t^0,*

$$\lim \frac{V - V^o}{\int_{t^0}^{\infty} e^{-\int_{t^0}^{t}(r+\lambda)ds}dt} = u - u^o$$

as $r + \lambda \to \infty$ uniformly in a neighborhood of t^0, if $r + \lambda$ remains bounded away from 0 uniformly on $[t^0, \infty)$.

Lemma 11 says that, if discount rates are large or search frictions low, then relative values are approximately equal to relative flow utilities. Similarly, if u and u^o are constant over time, relative values equal relative flow utilities. If either of these is the case uniformly across households, then dynamic demand D is approximately equal to demand in the corresponding static model $\widetilde{D}$. It is in this sense that the static model can be regarded as an approximation to the dynamic model in the cases of either "myopic" behavior (high discount rates), low search frictions (high λ) or steady state (constant u).

Home ownership

So far we have been discussing the market for housing rentals. What about home ownership? Under an assumption of perfect financial markets a no-arbitrage condition between either renting and holding financial assets or home-ownership must hold. In particular, we could extend the above model assuming that at each point in time a landowner can decide to sell her unit to the tenant or to another potential landowner, if the latter agrees. The price for such a (potential) sale into ownership is $\mathbf{P}$. Agreement on such a sale requires that each party is indifferent between holding financial assets and home ownership. Such indifference implies the asset equation

$$r\mathbf{P} = P + \dot{\mathbf{P}}, \tag{1.42}$$

where r is now a market rate of return. Tenant households could at the same time be landowners, for instance for units previously occupied. The interest rate r implicitly already incorporates a risk premium and a compensation for depreciation.

The focus of the present paper is identification of the determinants of the consumption value u. This consumption value conceptually maps more closely to rental prices P rather than home values $\mathbf{P}$, since decisions about homeownership reflect both consumption and investment considerations. This explains our focus on rental prices.

1.5 Identification in the dynamic model

In this section, we discuss the observable implications of social externalities for price and composition dynamics. Section 1.5.1 synthesizes into one proposition the characterization of price setting that was developed in the lemmas 6 through 10. This proposition in particular allows us to identify the average willingness to pay for composition, both for all households and for arbitrarily defined subgroups, by using the delayed response in average prices to exogenous shocks in X. If we add an assumption on the cross-sectional data generating process, this result allows to construct estimators of average willingness to pay.

In section 1.5.2, the implications of multiplicity of equilibria in composition m, as discussed in section 1.2.2, for cross-sectional quantile regressions of Δm on m will be analyzed. Under plausible restrictions on the data generating process, structural multiplicity of equilibria implies multiplicity of roots of such quantile regressions, although the reverse does not hold true. Section 1.5.2.1 reviews inference on the number of roots of nonparametric regressions (such as quantile regressions), which is discussed in more detail in chapter 2.

1.5.1 Exclusion restrictions based on the dynamic structure of price and composition responses

The following proposition 5 characterizes what happens to average rental prices in the neighborhood after a shock of size ξ to exogenous neighborhood characteristics X. As illustrated in figure 1.4, immediately following the shock prices jump by $P^{sr}_x\xi$. This is because households' reservation prices are a function of X and landowners extract all the surplus generated by the contract. This price jump reflects households' valuation of X. Due to search frictions, M evolves continuously in time and therefore remains unchanged in a vicinity of time 0. As time progresses, composition M converges to its new equilibrium, and prices change according to households' average valuation of M, yielding a total change in prices of $P^{lr}_x\xi$. Formally:

Proposition 5 (Dynamic identification of hedonic slopes)**.** *Make assumptions 1.4.1, 1.4.2, 1.4.3, 1.4.4, specifying the dynamic model, and 1.4.5 on bargaining power. Assume that housing supply is constant or all households are marginal.*
Assume that $X = x$ before time 0, $X = x + \xi$ for a jump ξ after time 0, and (u, V^o) is constant for all households.
Denote $P^b = \lim_{t\to 0^-} E[P]$ the average price before the jump, $P^{sr} = \lim_{t\to 0^+} E[P]$ the price right after the jump, i.e., "in the short run," and $P^{lr} = \lim_{t\to\infty} E[P]$ the price in the long run, similarly $M^{lr} = lim_{t\to\infty}M$, where it is assumed that these limits exists..
Then the following claims hold, where all the derivatives are evaluated at $\xi = 0$:

1. $P^{sr}_\xi = E\left[-\frac{u_X}{u_P}\right]$.

2. $P^{lr}_\xi = E\left[-\frac{u_X + u_M M^{lr}_\xi}{u_P}\right]$ *and* $M^{lr}_\xi = M^*_X$.
3. *If additionally assumption 1.2.4 holds,*

$$E\left[-\frac{u_m}{u_P}\right] = \frac{P^{lr}_\xi - P^{sr}_\xi}{m^{lr}_\xi}. \tag{1.43}$$

4. *More generally, for times* $t^2 > t^1 > 0$, *making assumption 1.2.4 again and taking* P^{t^1}, P^{t^2} *as the time specific averages,*

$$E\left[-\frac{u_m}{u_P}\right] = \frac{P^{t^2}_\xi - P^{t^1}_\xi}{m^{t^2}_\xi - m^{t^1}_\xi}. \tag{1.44}$$

Completely analogous claims hold for any subgroup, i.e., replacing $E[P]$ *by* $E[P|C]$, $E\left[-\frac{u_X}{u_P}\right]$ *by* $E\left[-\frac{u_X}{u_P}|C\right]$, *etc., if either all households in the subgroup are marginal,* $P = P^{co}$, *or we are looking at an upper tail of* P *of constant mass for this subgroup,* $\mathbb{P}(C) = const.$

So far, the discussion of the dynamic model was restricted to one neighborhood, there was no data generating process as in assumption 1.2.2. Proposition 5 is stated in terms of all-else-equal comparative statics. In order to use proposition 5 for identification of average willingness to pay parameters, we have to add an assumption on exogenous shocks to X in a cross-sectional dataset, as done in the following corollary. This corollary is stated in the fully nonparametric case, i.e., without imposing partial linearity assumptions.

Corollary 4. *Make assumptions 1.4.1, 1.4.2, 1.4.3, 1.4.4, specifying the dynamic model, and 1.4.5 on bargaining power. Assume that housing supply is constant or all households are marginal.*
Assume that we observe a cross-sectional dataset including $P^{t^1}, P^{t^2}, m^{t^1}, m^{t^2}$ *and* ξ, *where* ξ *is a random change of* X *taking place before time* t^1.
Let $\Delta \log P = E[\log P|t^2] - E[\log P|t^1]$ *and* $\Delta m = m^{t^2} - m^{t^1}$. *Denote by* $\beta^{\log P, \xi}$ *the expectation of the OLS regression coefficient of* $\Delta \log P$ *on* ξ *etc.*
Assume that ξ *is uncorrelated with changes in* P *in the time interval* $[t^1, t^2]$ *induced by changes in* X *or the distribution of* u, V^o *in this interval.*
Then

$$\frac{\beta^{\log P, \xi}}{\beta^{m, \xi}} = E\left[E\left[-\frac{u_m}{u_{\log P}}\right] \cdot \omega\right], \tag{1.45}$$

where

$$\omega = \frac{m_t(\xi - E[\xi])}{E[m_t(\xi - E[\xi])]}.$$

The inner expectation in equation 1.45 is taken over all households of a given neighborhood, and the outer expectation is taken over the product distribution of the cross-sectional distribution of the neighborhoods and the uniform distribution over the time interval $[t^1, t^2]$.

1.5.2 Social externalities, static and dynamic multiplicity of equilibria

Under the assumptions of the dynamic model of section 1.4 and assumption 1.2.4, according to lemma 7, the change of m over a time interval $[0, 1]$ is approximately given by

$$\Delta m = m^1 - m^0 \approx \kappa \cdot (d(m^0, X) - m^0).$$

In section 1.2.2, we saw that strong social externalities can imply multiplicity of equilibrium compositions m^*, that is, solutions to the equation $d(m^*, X) = m^*$. What are the empirical implications of multiple equilibria? Multiple equilibria, like path dependence more generally, imply a positive *causal* relation between past values of m and current values. However, time invariant unobserved heterogeneity (in X) also implies positive *correlation* between past and present m.

To illustrate, consider the following linear autoregressive panel model with fixed effects:

$$m^{i,t} = \alpha m^{i,t-1} + \beta X^i + \epsilon^{i,t}.$$

The coefficient α reflects path dependence, βX reflects time invariant heterogeneity. Below, we will study what happens around unstable equilibria of nonlinear difference equations. Locally, they are similar to this linear model with $\alpha > 1$. Unobserved heterogeneity biases cross-sectional regression estimates of α upward, as we will show now for this linear example, and then discuss in a nonparametric context. This upward bias implies that we will see unstable roots in regressions if there are multiple equilibria, but the reverse does not hold true.

Assuming that the ϵ are i.i.d. and uncorrelated with X, cross-sectional OLS regression of $\Delta m = m^t - m^{t-1}$ on $m = m^{t-1}$, using a two period panel, estimates the slope coefficient

$$\alpha - 1 + \beta \cdot \frac{Cov(m, X)}{Var(m)}.$$

The bias in this expression relative to $\alpha - 1$ is positive if $Cov(m, \beta X)$ is positive. This is in particular the case if the m^{t-1} are generated from the stationary distribution of this panel model, which exists if $|\alpha| < 1$:

$$\beta \cdot \frac{Cov(m^{t-1}, X)}{Var(m^{t-1})} = \frac{\beta^2 \sigma_X^2/(1-\alpha)}{(\beta^2 \sigma_X^2/(1-\alpha) + \sigma_\epsilon^2)/(1-\alpha^2)} = (1+\alpha) \cdot \frac{\beta^2 \sigma_X^2}{\beta^2 \sigma_X^2/(1-\alpha) + \sigma_\epsilon^2}.$$

The bias term here is positive if α is bigger than -1. A positive bias implies that cross-sectional regression of Δm on m will yield a positive slope if $\alpha > 1$, but the reverse is not true. The following discussion generalizes this idea to a nonparametric context, using the specification of Δm derived from the dynamic model in section 1.4.

Let us now return to the nonlinear equation $\Delta m = m^1 - m^0 \approx \kappa \cdot (d(m^0, X) - m^0)$. Suppose we observe a two-period panel of m across neighborhoods, or equivalently a cross-sectional distribution of $(\Delta m, m)$. By endogeneity of m, if there is serial dependence in X then m will not be independent of X and the conditional distribution of Δm given m does not permit direct inference on d. However, by a generalization of the argument just made, structural multiplicity of equilibria implies multiplicity of roots in such cross-sectional regressions, under certain assumptions.

Denote the τth conditional quantile of Δm given m by $Q^{\Delta m|m}(\tau|m)$, the conditional cumulative distribution function at Q by $F^{\Delta m|m}(Q|m)$, and the conditional probability density by $f^{\Delta m|m}(Q|m)$. The following lemma shows that quantile regressions of Δm on m yield biased slopes relative to the structural slope $\kappa \cdot (d_m - 1)$.

Lemma 12 (Bias in quantile regression slopes)**.** *If $\Delta m = \kappa \cdot (d(m, X) - m)$, and Q and F are differentiable with respect to the conditioning argument m, then*

$$\begin{aligned} \frac{\partial}{\partial m} Q^{\Delta m|m}(\tau|m) &= E\left[\kappa \cdot (d_m - 1) | \Delta m = Q, m\right] \\ &- \frac{1}{f^{\Delta m|m}(Q|m)} \cdot \frac{\partial}{\partial m} \mathbb{P}\left(\kappa \cdot (d(m', X) - m') \leq Q|m\right)\bigg|_{m'=m}. \end{aligned}$$

In the linear example discussed above, unobserved heterogeneity X was constant over time, which implied a positive correlation between βX at time t and m at time $t-1$. The bias term in the linear example continues to be positive under the much weaker condition that X is not *negatively* correlated across time. In the nonparametric quantile regression case, assumption 1.5.1 provides the natural analogon of such non-negative correlation. It states that there is no negative dependence between current d, evaluated at fixed m, and current m. Violation of this assumption would require some underlying cyclical dynamics, in continuous time, with a frequency close enough to half the frequency of observation, or more generally with a ratio of frequencies that is an odd number divided by two. It seems safe to discard this possibility in most applications. This assumption might not hold, for instance, if outcomes were influenced by seasonal factors and observations were semi-annual.

Assumption 1.5.1 (First order stochastic dominance)**.** *$\mathbb{P}((d(m', X) - m') \leq Q|m)$ is non-increasing as a function of X, holding x' constant.*

We can now formally state the claim that if there are unstable equilibria structurally, then quantile regressions should exhibit multiple roots.

Proposition 6 (Unstable equilibria in dynamics and quantile regressions)**.** *Assume $\Delta m = \kappa \cdot (d(m, X) - m)$ and assumption 1.5.1 holds.*
Then, if $Q^{\Delta m|m}(\tau|m)$ has only one root[5] *m for all τ, then the conditional average structural functions $E[\kappa \cdot (d(m', X) - m')|d(X, m) = m, m]$, as functions of m', are "stable" at the roots m:*

$$E[\kappa \cdot (d_m - 1)|\Delta m = 0, m] \leq 0$$

for all m, where $(0, m)$ is in the support of $(\Delta m, m)$.

1.5.2.1 Testing for the number of roots of nonparametrically identified functions

Chapter 2 develops an inferential procedure to construct (integer valued) confidence sets for the number of roots of some function g in a certain range, where g is a function identified by conditional moment restrictions such as a conditional mean or quantile. In what follows, some of the central results of chapter 2 are summarized. For more detail the interested reader is referred to that chapter.

Assume we are interested in the number of roots $Z(g)$ of some function g on the range $[0, 1]$:

Definition 6. *For g continuously differentiable*

$$Z(g) := |\{m \in [0, 1] : g(m) = 0\}|.$$

The inference procedure proposed is based on a smoothed version of Z, Z_ρ:

Definition 7.

$$Z_\rho(g(.), g'(.)) := \int_0^1 L_\rho(g(m))|g'(m)|dm, \tag{1.46}$$

where L_ρ is a Lipschitz continuous, positive symmetric kernel integrating to 1 with bandwidth ρ and support $[-\rho, \rho]$

If ρ is small enough, then for generic g it can be shown that $Z(g) = Z_\rho(g)$. The function g is assumed to be identified by the conditional moment restriction $g(m) = \text{argmin}_d E_{\Delta m|m}[\rho(\Delta m - d)|m]$, where ρ is some loss function. This includes in particular the case $\rho_q(\delta) := \delta(q\mathbf{1}(\delta > 0) - (1 - q)\mathbf{1}(\delta < 0))$ for conditional qth quantile regression. Local linear m-regression estimates g and g' by analogy, fitting a line to the data by essentially minimizing a kernel estimate of the conditional expectation of ρ:

$$(\hat{g}(m), \hat{g}'(m)) = \text{argmin}_{a,b} \sum_k K_\tau(m_k - m)\rho(\Delta m_k - a - b(m_k - m)). \tag{1.47}$$

[5] i.e., one solution m to the equation $Q^{\Delta m|m}(\tau|m) = 0$

An "estimator" $\hat{Z}$ of Z can be formed by plugging $(\hat{g}, \hat{g'})$ into the functional Z_ρ, i.e., $\hat{Z} := Z_\rho(\hat{g}, \hat{g'})$. As it turns out, uniform convergence of $(\hat{g}, \hat{g'})$ implies a degenerate asymptotic distribution for $\hat{Z}$, and any recentered/rescaled version of it. Define the following functional norm:

Definition 8 ($\mathscr{C}^1$ norm)**.** *Let $\mathscr{C}^1([0,1])$ denote the space of continuously differentiable functions on the interval $[0,1]$. The norm $||.||$ on $\mathscr{C}^1([0,1])$ is defined by*

$$||g|| := \sup_{m\in[0,1]} |g(m)| + \sup_{m\in[0,1]} |g'(x)|.$$

Then we have

Proposition 7. *$Z(.)$ is constant in a $||.||$ neighborhood of any generic function $g \in C^1$ with $Z(g) < \infty$, and so is Z_ρ for ρ small enough.*

This implies the corrolary that, under standard i.i.d. sequences of experiments, no non-degenerate asymptotic distribution can be obtained for $\widehat{Z}$. More precisely, if $(\widehat{g}, \widehat{g'})$ converges uniformly in probability towards (g, g'), if g is generic, and if $\alpha_n \to \infty$ is some arbitrary diverging sequence, then

$$\alpha_n(Z(\widehat{g}) - Z(g)) \to^p 0.$$

Furthermore, if ρ is small enough for $Z_\rho(g, g') = Z(g)$, then

$$\alpha_n(Z_\rho(\widehat{g}, \widehat{g'}) - Z(g)) \to^p 0.$$

Given this result, no useful asymptotic theory for inference on $Z(g)$ can be found using standard i.i.d. sequences of experiments. The central result of chapter 2 therefore gives the asymptotic distribution of $\widehat{Z}$ for a non-standard sequence of experiments, which is defined by

$$\begin{aligned} m_{i,n} &\sim^{iid} f_m(.) &(1.48)\\ \gamma_{i,n}|m_{i,n} &\sim f_{\gamma|m} &(1.49)\\ \Delta m_{i,n} &= g(m_{i,n}) + r_n\gamma_{i,n}, &(1.50)\end{aligned}$$

where $0 = \text{argmin}_x E[\rho(\gamma - x)|m] = \text{argmin}_x E[\rho(\delta - x)|m]$. This sequence increases the variance of the "residual," $r_n\gamma_{i,n} = \Delta m_{i,n} - g(m_{i,n})$, as sample size n increases, thereby reducing the "signal to noise ratio," leaving the model otherwise unchanged.

Theorem 1 (Asymptotic normality)**.** *Under the above model assumptions and regularity conditions stated in chapter 2, and if $r_n = (n\tau^5)^{1/2}$, $n\tau \to \infty$, $\rho \to 0$ and $\tau/\rho^2 \to 0$, then* *there exist $\mu > 0$ and V such that*

$$\sqrt{\frac{\rho}{\tau}}\left(\widehat{Z} - \mu - Z\right) \to N(0, V)$$

for $\widehat{Z} = Z_\rho\left(\widehat{g}, \widehat{g'}\right)$. Both μ and V depend on the data generating process only via the asymptotic mean and variance of $\widehat{g'}$ at the roots of g.

This result allows the construction of integer-valued confidence sets using t-statistics with bootstrapped standard error and bias.

1.6 Application to data on cities in the United States

In this section the identification results of the previous sections are applied to data on neighborhood composition in cities in the United States. In particular, we estimate the extent to which the Hispanic share in a given neighborhood affects the decision of Hispanics and non-Hispanics to move to that neighborhood. The next subsection provides a description of the Neighborhood Change Data Base (NCDB), which aggregates data of the US census to the level of census tracts, and discusses sample selection as well as variable construction. The sample is restricted to larger urban areas and outliers are omitted. Imputed rents have to be calculated from observed rents and house values.

Subsection 1.6.2 provides estimates of demand slopes and hedonic slopes based on the various exclusion restrictions that were discussed in sections 1.3 and 1.5. These estimates consistently suggest large positive dependence of demand of Hispanics on Hispanic share and similarly large positive dependence of demand of non-Hispanics on non-Hispanic share. The estimates of price slopes with respect to Hispanic share imply moderately negative dependence. Subsection 1.6.3 checks the robustness of these results by applying the estimators to various subsamples and different housing cost variables and by decomposing the linear IV coefficients using the results of appendix 1.A. In section 1.6.4, finally, we look at the dynamics of neighborhood composition over time for the largest metropolitan areas in the United States. The evidence suggests the absence of multiple equilibria in the dynamics of composition. Social externalities do not seem to be strong enough to cause tipping behavior.

1.6.1 The data

The data set used is an extract from the Neighborhood Change Database (NCDB) which aggregates US census variables to the level of census tracts. Tract definitions are changing between census waves but the NCDB matches observations from the same geographic area over time, thus allowing to observe the development over several decades of the universe of US neighborhoods.

The sample is selected in a manner similar to Card, Mas, and Rothstein (2008), who use the same database. In particular, all rural tracts are dropped, as well as all metropolitan standard areas (MSA) with fewer than 100 tracts, all tracts with population below 200 and tracts that grew by more than 5 standard deviations above and beyond the MSA mean. The definition of MSA used is the MSAPMA from the NCDB, which is equal to "Primary Metropolitan Statistical Area" if the tract lies in one of those, and equal to the MSA it lies in otherwise.

Three measures of housing prices are used, median reported rents, median reported values, and an "imputed rent" variable created by myself. The latter imputes rents based on housing values and takes share weighted averages of imputed and reported rents as follows. By the arbitrage condition between owning and renting discussed in section 1.4.1, we have $P = r\mathbf{P} - \dot{\mathbf{P}}$. Under the assumption that the expected value appreciation $\dot{\mathbf{P}}$ is uncorrelated with baseline value $\mathbf{P}$, the appropriate interest rate can then be determined by a cross-sectional linear regression of P on $\mathbf{P}$. Rents are imputed from housing values as the predicted rents from such cross-sectional regressions in each decade. The imputed rent variable used in regressions is a weighted mean of average observed rents and average rents predicted from house values, where the weights correspond to the respective share of rental and non-rental units in a tract.

1.6.2 Exclusion restrictions based estimates

The parameters of interest are the dependence of the demand of Hispanics and non-Hispanics on Hispanic share, everything else equal, as well as the (weighted average) willingness to pay for composition, which is reflected in the counterfactual hedonic slope P_m^+ of housing costs with respect to Hispanic share.

Table 1.1 shows a number of "naive" hedonic and demand regressions that ignore problems of omitted variables and the endogeneity of composition in the presence of social externalities. Clearly, problems of omitted variable bias are severe as, throughout the demand regressions shown, demand is increasing in prices, suggesting price variation is driven by fluctuations in demand due to variation in omitted factors X. Taken at face value, these regressions would furthermore suggest a negative preference of non-Hispanics for Hispanic share, and a strong positive preference of Hispanics for Hispanic share, as well as an average willingness to pay for Hispanic share of around 0.

In this subsection, several instrumental variables for neighborhood composition are constructed, as suggested by the theoretical results of sections 1.3 and 1.5. Throughout this section, the restrictive assumption is made that $\mathscr{C} = 2$ and the relevant type variable is Hispanic origin. Hispanics are denoted by $c = 1$ and non-Hispanics by $c = 2$. The preferred specifications are run in decadal differences on the dataset pooling changes over the 80s and 90s, controlling for MSA $\times$ time fixed effects. They furthermore control for neighborhood and decade specific initial conditions, as described below. Let us now define the instruments used and discuss the conditions of their validity. Then, the empirical results of the preferred specifications, as shown in table 1.2, will be discussed and put in the context of the theoretical models. The theoretical interpretations of the entries of table 1.2 are summarized in table 1.3.

1.6.2.0.1 Subgroup shifters

Let $\widetilde{c}$ denote "subtypes" of Hispanics which correspond to country of origin (Mexico, Puerto Rico and Cuba). $M^{\widetilde{c}}$ is the initial population of type $\widetilde{c}$ in a neighborhood and $dM^{\widetilde{c},nat}/M^{\widetilde{c},nat}$ is the the total change of population of type $\widetilde{c}$, summed over all neighborhoods in the dataset, divided by total initial population of type $\widetilde{c}$. The instrument dX^I is defined as

$$\mathbf{dX^I} = \frac{1}{M^1 + M^2} \sum_{\widetilde{c}} M^{\widetilde{c}} \cdot \frac{dM^{\widetilde{c},tot}}{M^{\widetilde{c},tot}}. \tag{1.51}$$

This is a synthetic instrument similar to the one used by Card (2001) on the MSA level as a predictor of changes in labor supply. It is predictive for local changes in Hispanic share if new immigrants from the same source countries have a similar distribution of preferences as prior migrants, whether the preference is for exogenous location characteristics or for the presence of their compatriots.

Formally, in order to use dX^I as an instrument for composition m in estimation of D^2_m, we need dX^I to satisfy the conditions $D^2_{X^I} = 0$, $D^1_{X^I} \neq 0$ and dX^I has to be uncorrelated with counterfactual changes in M^2. The instrument dX^I is excluded from the demand of type $c = 2$ if there is no causal effect of total immigration on demand of type 2, i.e., non-Hispanics. This might not hold if the set of outside options is affected by immigration. All regression control for MSA $\times$ time fixed effects, however, absorbing any city-wide shifts in outside options. They furthermore control flexibly for initial Hispanic share, so that identification is driven by variation in the composition of initial Hispanic population of a neighborhood in terms of country of origin, conditional on Hispanic share. We also need the instrument dX^I to be independent of changes in ϵ affecting demand of type 2. A potential threat to validity here would be some delayed adjustment due to frictions, which could imply a correlation between current composition and future adjustment of type 2 population. The fact that shifting immigrant demand will also shift rental prices for non-Hispanics will be explicitly taken into account below.

1.6.2.0.2 Spatial structure and exclusion restrictions

Spatial structure is extracted from the data as follows: For each census tract, the Neighborhood Change Data Base reports latitude α and longitude β of an interior point. Distance between neighborhoods is defined, based upon these coordinates, as the euclidean distance between the corresponding coordinates in $\mathbb{R}^3$, which are given by

$$6371 \cdot (\cos(\alpha) \cdot \cos(\beta), \cos(\alpha) \cdot \sin(\beta), \sin(\alpha)) \,. \tag{1.52}$$

Here 6371 is taken to be the radius of earth in kilometers. Denote the average predicted shift of Hispanic demand in neighborhoods that are at least 3 km away, but among the 15 closest neighborhoods, by $\mathbf{dX^{>3}}$, where predicted shift is the synthetic instrument dX^I defined by equation 1.51. Denote $\widetilde{m}$ the average Hispanic share in the given neighborhood and its 4 closest adjacent tracts.

In order to use $dX^{>3}$ as an instrument for composition $\widetilde{m}$ in estimation of $D^2_{\widetilde{m}}$, in the context of the spatial model of section 1.3.4, we need $dX^{>3}$ to satisfy the conditions $D^2_{X^{>3}} = 0$, $\widetilde{m}_{X^{>3}} \neq 0$, and $dX^{>3}$ has to be uncorrelated with counterfactual changes in M and P. Furthermore, we need to assume that the composition variable which *does* matter for households' location choices is $\widetilde{m}$. The regressions using $dX^{>3}$ as an instrument for $\widetilde{m}$ will control for dX^I, and thus use variation in composition orthogonal to the one used in the subgroup approach. While we might expect some amenities to be relevant for neighborhoods further than 3km away, it seems plausible that the exclusion restrictions are satisfied in the case of predicted immigration for neighborhoods at a certain distance conditional on local predicted immigration. The use of $\widetilde{m}$, the average of m for 5 adjacent neighborhoods, as relevant composition variable is quite arbitrary. The results are robust to different specifications of $\widetilde{m}$, however.

1.6.2.0.3 The dynamic structure of price responses and exclusion restrictions

Let $\mathbf{dX^L}$ be the decadal change in m, lagged by a decade. The variable dX^L is used as an instrument for Δm in regressions of ΔP on Δm, controlling for m. In the context of the dynamic model, past changes in m are predictive of future changes if they reflect incomplete adjustments to past shocks in X.

Proposition 5 stated that, under the appropriate assumptions, any shocks to X are immediately incorporated into prices P according to household willingness to pay. Due to search frictions, however, composition m only adjusts with delay, with prices following accordingly. Past changes in m are a valid instrument for future changes in m iff they are uncorrelated with future changes in X. We shall make the strong identifying assumption that this holds true, conditional on current Hispanic share m. The main threat to the validity of this assumption would be anticipated changes in amenities X that are reflected in past composition changes.

1.6.2.0.4 Discussion and interpretation of estimates

Table 1.2 shows the central empirical results, using these instruments. Instrumental variable regressions which are not theoretically meaningful are omitted from the table. Table 1.3 shows the corresponding interpretations of these results under the assumptions of the respective models. The first thing to note is that the instrument is a highly significant predictor of the change in local composition for all three specifications. Strength of the instruments is therefore not an issue.

For both the subgroup instrument and the spatial instrument, the results suggest a strong negative dependence of the demand of non-Hispanics on Hispanic share. We have to take care, however, to correct for the price effect of changing Hispanic share in order to obtain structural slopes of demand (compare corollaries 2 and 3). This is reflected in the bias terms of the form $D_P^2 \frac{P_X^*}{m_X}$ in table 1.3. If we assume that the elasticity of non-Hispanic demand with respect to rents is between 0 and 2, and taking into account that the IV regressions of P on m yield coefficients of around -0.5, this implies a bias of around 0 to 1. Subtracting this bias yields estimates of $\mathbf{D_m^2}$ of **-6.3** to **-9.4**. For Hispanics, the estimate based on the spatial instrument implies a positive dependence of demand on Hispanic share. Correcting again for the rent-bias, we get an estimate of $\mathbf{D_m^1}$ of around **2.4** to **3.4**. The IV regressions of prices on Hispanic share, using the spatial and dynamic instruments, yield moderately negative estimates of $\mathbf{P_m^+}$ of **-0.75** and **-0.52**. This implies a moderately negative average marginal willingness to pay for Hispanic share.

These results are remarkably consistent across instruments. While we might have doubts about the validity of each of the instruments, they do rely on different assumptions and use orthogonal variation in the data, so that this consistency might add to the credibility of the results. Finally, let us compare these results to those using "naive" regressions, as shown in table 1.1. Consistently across specifications, it seems that the naive estimates of D_m^1, D_m^2 are strongly upward biased, and the estimates of P_m^+ are moderately upward biased. This also holds true for the specifications in differences controlling for initial Hispanic share and MSA $\times$ time fixed effects. One interpretation of this result might be that Hispanic location decisions were more "pro-cyclical" relative to non-Hispanics, i.e., Hispanic demand reacted more strongly to unobserved shocks in X.

1.6.3 Robustness - Subsamples, different housing cost variables, and decomposition of the LATE

The regressions of the previous subsection used the full sample of the 114 largest MSAs in the United States, pooling the data for changes in the 80s and in the 90s. In this subsection, the robustness of the results is checked by replicating the regressions on subsamples. In particular, table 1.4 presents estimates for the subset of MSAs with Hispanic shares larger than 8% in 2000. This corresponds to roughly 50% of the sample. Furthermore, there might be concerns about the effect of rent controls. Table 1.5 replicates the regressions on the sample of MSAs excluding California and the state of New York, where rent controls might play some role. Table 1.6 finally shows estimates for the 80s and for the 90s separately. This table also uses median rents and median reported house values as alternative housing cost variables.

The results are largely consistent with those obtained previously, with a few exceptions. First, in the sample of MSAs with large Hispanic shares, Hispanics seem less responsive in their location decision to the Hispanic share of a neighborhood. Second, in the sample excluding California and New York, price responses seem somewhat stronger. This might indicate a certain role of rent controls. Finally, in this sample the subgroup instrument is quite weak, and the corresponding estimate of D_m^2 very high with a very large standard error. The different housing cost variables behave in a roughly similar way.

Finally, instrumental variables estimates describe the local average treatment effect (LATE) for the subpopulations for which the instruments do affect the treatment. What are the characteristics of these subpopulations of neighborhoods for our instruments? The decomposition results of appendix 1.A can be used to shed some light on this question. In particular, the IV coefficient controlling for covariates can be decomposed as a weighted average of structural slopes over the sampling population, where the weights ω are identifiable and are given by

$$\omega = \frac{\Delta m \cdot e}{E[\Delta m \cdot e]}.$$

In this expression, Δm is the regressor of interest and e is the residual of a regression of the instrument on the controls. Panels 1.5 through 1.7 show the unweighted density of the initial Hispanic share across neighborhoods for the sample used, as well as this density reweighted by ω, for weights ω corresponding to the various instruments. They furthermore show plots of estimates of the "conditional first stage" and "conditional reduced form," $E[\omega|m]$ and $E[\Delta Y \cdot e|m]$, where ΔY corresponds to the change of various outcomes of interest. The plots of the reweighted densities are particularly instructive. They show that the specification using the subgroup instrument estimates a LATE for neighborhoods with a medium to high initial Hispanic share, using the spatial instrument yields a LATE for neighborhoods with lower Hispanic shares (although still upweighting higher shares relative to the population), and the dynamic instrument estimates a LATE for neighborhoods somewhere in between. The conditional expectation estimates for higher values of m should be interpreted with caution, as they are quite imprecisely estimated due to limited support of Hispanic share in the right tail.

The graphs of the conditional reduced form of price responses, $E[\Delta P \cdot e|m]$ for the spatial and dynamic instrument, when compared to the conditional first stage, $E[\omega|m]$, are somewhat worrisome. They suggest significant variation of the conditional IV coefficient given m over the range of m. This does not imply invalidity of the instrument, but it cautions to be careful when extrapolating the willingness-to-pay results to different populations.

1.6.4 Multiple equilibria

Finally, we shall now look at the dynamics of neighborhood composition in the largest metropolitan areas of the United States, and discuss what they reveal about underlying multiplicity of equilibria in composition. According to lemma 10, if the corresponding static model has multiple equilibria, then the dynamic model has multiple equilibria. Under assumption 1.5.1, if the dynamic model has multiple equilibria, then cross sectional quantile regressions of Δm on m have multiple roots. Section 1.5.2.1 reviewed inference on the number of roots of nonparametrically identified functions as proposed in chapter 2.

Panel 1.8 shows kernel density plots of the distribution of Hispanic share across census tracts in the three largest Metropolitan Areas of the United states, in 1980 and 1990. Panel 1.9 displays quantile regressions of the change in Hispanic share on initial Hispanic share for the 1980s and 1990s for the same cities, where the plots show the 0.2, 0.5 and 0.8 conditional quantile.

Visual inspection of these quantile regressions suggests a pattern of stability with mean-reversion for New York, where all neighborhoods appear to be converging to a medium level of Hispanic share, although at a rather slow rate. Los Angeles, on the other hand, experienced growth of Hispanic shares across neighborhoods, with the Hispanic share in intermediate neighborhoods growing the fastest. The pattern for Chicago is less clear. In interpreting these regressions, we have to be careful about the initial support of m. As Chicago had very few neighborhoods with high Hispanic shares, the estimates for high initial m are largely based on extrapolation. That said, these pictures do not seem to suggests a pattern of unstable equilibria, which would imply regressions crossing the horizontal axis from below.

Table 1.7 formalizes this visual intuition, testing for the number of roots of these quantile regressions in the interval $[0, 1]$. Each city and decade is considered separately. As can be seen from this table, in nearly all cases considered, multiplicity of roots can be rejected at the 5% level. Two things are important to note. First, the smoothing parameter ρ was chosen to equal 0.04. For regressions that stay within the interval $[-.04, .04]$ over extended ranges this might imply an upward bias, if estimated regressions are "wiggly" due to estimation noise, and a downward bias, if the true regressions peak within this range and intersect the axis on both sides of the peak. Second, the range of integration was chosen to equal $[0, 1]$. This implies, in particular, that roots lying right at the boundary of this interval might only be "counted half". It implies furthermore that the regressions might be extrapolating in ranges of initial Hispanic share for which no observations are available. Hence, in interpreting these estimates, the corresponding graphs should always be considered.

1.7 Summary and conclusion

In this paper we presented models of sorting in which location choices depend on the location choices of other agents, as well as exogenous location characteristics. In such a setup, the composition of agents at a location is an endogenous equilibrium outcome with generically degenerate support given exogenous location characteristics. This leads to an identification problem similar to the "simultaneity problem" and the "reflection problem" discussed in the literature: the effects of endogenous composition and exogenous characteristics on agents' location choices and prices are not separately identified.

A series of approaches to overcome this problem was proposed here. The first is based on assuming that some exogenous, location specific demand shifters are excluded from the choices of a subgroup of agents. If that is the case, random variation in such exogenous characteristics can serve as an instrument for endogenous composition. The second is based on assuming a spatial structure with externalities across adjacent locations. Given such a spatial structure, variation in exogenous characteristics at a location generates variation in composition propagating across adjacent neighborhoods, and can serve as an instrument for composition in neighborhoods not immediately adjacent. The third is based on a dynamic search-model extension. In this extension prices adjust immediately but location composition reacts only with delay to changes in exogenous characteristics, because of search frictions. Past shocks in exogenous characteristics can therefore serve as instruments for future composition changes. Finally, the testable implications of multiplicity of equilibria in composition, as implied by strong social externalities, were discussed.

In an application of these approaches, the impact of the share of Hispanics in neighborhoods in the United States on housing demand of Hispanics and non-Hispanics as well as rental prices was studied. The results consistently suggest a strong impact of composition on location choices, in the form of an own-group preference. This contrasts with the rather weak evidence on the impact of neighborhood composition on observable outcomes of residents, as in Katz, Kling, and Liebman (2007). It remains a task for future research to further disentangle the nature of the social externalities that were found here. For instance, we could think of the reduced form demand functions $D(X, M, P)$ as reflecting preferences over endogenous amenities $W(X, M)$, $D(X, M, P) = D(X, W(X, M), P)$, where $\dim(W) = \dim(M)$. Under this assumption, $D_M = D_W W_M$. Given identification of D_M, one could attempt to identify either D_W or W_M, and then invert to get for instance $D_W = D_M W_M^{-1}$. Identification of W_M could come from shocks to X which are excluded from W but do affect composition M. This approach would require full observability of W.

Application of the methods developed here to a number of different problems seems interesting. For instance, in the field of economic geography firm location choices are studied which depend on exogenously given geographic factors and the location choices of other firms (and households). One central question of this field is to understand the mechanisms determining the agglomeration or dispersion of economic activity, see for instance Krugman (1991) and Ellison and Glaeser (1999). It seems that the problem of firm location choice has a very similar structure to the problem of household neighborhood choice within a city, which motivated this paper. Another interesting application might be the academic job market: In choosing among job-offers, academics will generally make their decision based not only on exogenous characteristics (location, facilities...) and pay, but based also on who else is working at a given university.

Appendix 1.A Decomposition representations of linear IV coefficients

In this appendix, a series of representations of linear IV coefficients in terms of weighted average slopes is developed. These results resemble closely the LATE representations introduced by Imbens and Angrist (1994). The distinguishing feature of the results presented here is that all weights are defined in terms of observable and identifiable quantities, as opposed to first stage structural slopes (in the binary case, compliance versus noncompliance). This allows to describe the distribution of any observable covariates for the population over which structural slopes are averaged to obtain the linear IV coefficients. In the terminology of Imbens and Angrist (1994), we don't know who the compliers are but we do know how they look like. Results similar in spirit were used by Kling (2001).

The first set of results is stated for a random coefficient, cross-sectional setup. These results suggest to plot densities of covariates with respect to a reweighted distribution, and to plot conditional IV coefficients in the case of linear IV with controls. Then, these results are generalized to the fully non-parametric panel difference case, which is the setup relevant for the present paper.

Lemma 13 (Crossectional IV, random coefficient case)**.** *Assume that*

$$Y^i = \alpha^i + \beta^i X^i \tag{1.53}$$

and assume $Cov(Z, \alpha) = 0$*. Then*

$$\beta^{IV} = \frac{Cov(Y, Z)}{Cov(X, Z)} = E\left[\beta^i \cdot \omega\right]$$

for a weighting function

$$\omega = \frac{X(Z - E[Z])}{E[X(Z - E[Z])]}.$$

Lemma 14 (Crossectional OLS with controls, random coefficient case)**.** *Assume that*

$$Y^i = X^{1,i}\beta^{1,i} + X^{2,i}\beta^{2,i} + \epsilon \tag{1.54}$$

for a scalar X^1 *and a vector* X^2*. Assume* $X^1 \perp (\beta, \epsilon)|X^2$*, and* $E[X^{2,i}\beta^{2,i} + \epsilon|X^2]$ *is linear* *in X^2. Then the coefficient on X^1 in OLS regression of Y on X is in expectation equal to*

$$\beta^{1,OLS} = E\left[E[\beta^{1,i}|X^2]\frac{E[X^1e|X^2]}{E[X^1e]}\right],$$

where e is the residual from OLS regression of X^1 on X^2.

Lemma 15 (Crossectional IV with controls, random coefficient case)**.** *Assume that*

$$Y^i = X^{1,i}\beta^{1,i} + X^{2,i}\beta^{2,i} + \epsilon \tag{1.55}$$

for a scalar X^1 and a vector X^2. Assume $Z \perp (\beta^2, \epsilon)|X^2$ for a scalar instrument Z, and $E[X^{2,i}\beta^{2,i} + \epsilon|X^2]$ is linear in X^2. Denote by e the residual of OLS regression of Z on X^2.

Then the coefficient on X^1 in IV regression of Y on X, instrumented by (Z, X^2), is in expectation equal to

$$\beta^{1,IV} = \frac{E[Ye]}{E[X^1e]} = E\left[\frac{E[\beta^{1,i}X^1e|X^2]}{E[X^1e]}\right] = E\left[\beta^{1,i}\cdot\omega\right]$$

for a weighting function

$$\omega = \frac{X^1e}{E[X^1e]}.$$

These lemmas give a LATE representation of IV coefficients. In the setup of lemma 15, the following two exercises seem instructive:

Suggestion 1: Plot the distribution of covariates (in particular of components of X^2), reweighted by ω. In the terminology of Imbens and Angrist (1994), this gives the distribution of covariates for the set of compliers.

Suggestion 2: Calculate conditional IV given (components of) X^2: Let $\widehat{E}$ denote some flexible ("nonparametric") estimator of the conditional expectation. For components of X^2, plot (nonparametric) regressions of

$$\widehat{\beta}^{IV}(X^2) := \frac{\widehat{E}\left[Ye|X^2\right]}{\widehat{E}\left[X^1e|X^2\right]}$$

on these components. The estimator $\widehat{\beta}^{IV}(X^2)$ converges to a conditional weighted average of the structural slope β^2,

$$E\left[\beta^1\frac{X^1e}{E[X^1e|X^2]}\middle| X^2\right].$$

In, practice, however, such estimates of $\beta^{IV}(X^2)$ might be poorly behaved. If the denominator, $\widehat{E}\left[X^1e|X^2\right]$, has positive mass around 0, then $\widehat{\beta}^{IV}(X^2)$ might not have a finite expectation. In that case, it can still be insightful to plot the "conditional reduced form" estimator

$\widehat{E}[Ye|X^2]$.

The following lemmas extend the previous results to the panel-difference case.

Lemma 16 (Panel difference IV, random coefficient case)**.** *Assume that*

$$Y^{it} = \alpha^{it} + \beta^{it} X^{it} \tag{1.56}$$

for $t \in \{0, 1\}$*, and assume* $\Delta Z \perp (\Delta\alpha + \Delta\beta \cdot X^{i,1})$.[6] *Then*

$$\beta^{IV,\Delta} := \frac{Cov(\Delta Y, \Delta Z)}{Cov(\Delta X, \Delta Z)} = E\left[\beta^{i,0} \cdot \omega\right]$$

for a weighting function

$$\omega = \frac{\Delta X(\Delta Z - E[\Delta Z])}{E[\Delta X(\Delta Z - E[\Delta Z])]}.$$

Lemma 17 (Panel difference IV, nonparametric case)**.** *Assume that*

$$Y^{it} = g(X^{it}, \epsilon^{it}) \tag{1.57}$$

for $t \in [0, 1]$*, and assume*

$$\Delta Z \perp \int_0^1 g_\epsilon(X^{it}, \epsilon^{it}) \cdot \epsilon_t \, dt.$$

Then

$$\beta^{IV,\Delta} := \frac{Cov(\Delta Y, \Delta Z)}{Cov(\Delta X, \Delta Z)} = E\left[g_X \cdot \omega\right]$$

for a weighting function

$$\omega = \frac{X_t(\Delta Z - E[\Delta Z])}{E[X_t(\Delta Z - E[\Delta Z])]}.$$

All expectations here are taken over the product distribution of the crosssectional distribution over the i *and the uniform distribution over the time interval* $[0, 1]$.

Lemma 18 (Panel difference IV, nonparametric case, if exclusion is violated)**.** *Assume that*

$$Y^{it} = g(X^{1,it}, X^{2,it}, \epsilon^{it}) \tag{1.58}$$

for $t \in [0, 1]$*, and assume*

$$\Delta Z \perp \int_0^1 g_\epsilon(X^{it}, \epsilon^{it}) \cdot \epsilon_t \, dt.$$

[6] That is, ΔZ is uncorrelated with the counterfactual change in Y which would have occurred if X had stayed constant.

Then

$$\beta^{Y1,IV,\Delta} := \frac{Cov(\Delta Y, \Delta Z)}{Cov(\Delta X^1, \Delta Z)} = E\left[g_{X^1} \cdot \omega^1\right] + R$$

for weighting functions ($j = 1, 2$)

$$\omega^j = \frac{X_t^j(\Delta Z - E[\Delta Z])}{E[X_t^j(\Delta Z - E[\Delta Z])]}$$

and an error term

$$R = E\left[g_{X^2} \cdot \frac{X_t^2}{X_t^1} \cdot \omega\right] = E[g_{X^2}\omega^2] \cdot \beta^{21,IV,\Delta} = E[g_{X^2}\omega^2] \cdot \frac{Cov(\Delta X^2, \Delta Z)}{Cov(\Delta X^1, \Delta Z)}.$$

All expectations here are taken over the product distribution of the crosssectional distribution over the i and the uniform distribution over the time interval $[0, 1]$.

Suggestion 3: Bound the error term by making a-priori assumptions giving bounds on $E[g_{X^2}\omega^2]$. Estimate $\beta^{21,IV,\Delta} = Cov(\Delta X^2, \Delta Z)/Cov(\Delta X^1, \Delta Z)$ directly from the data.

This appendix concludes with a characterization of cross-sectional linear IV in a triangular system, where the weights in this lemma are now expressed in terms of first stage structural slopes.

Lemma 19 (Cross-sectional linear IV in nonparametric triangular systems)**.** *Consider the triangular system $Y = g(X, \epsilon)$, $X = h(Z, \eta)$, $Z \perp (\epsilon, \eta)$, where all variables are continuously distributed and g, h are continuously differentiable. Then*

$$\beta^{IV} = \frac{Cov(Y, Z)}{Cov(X, Z)} = E\left[g_x(X, \epsilon)\omega(Z, \eta)\right]$$

for a weighting function ω which is given, up to normalization, by

$$\omega(z, \eta) = const. \cdot \frac{h_z(z, \eta)}{f(z)} \cdot \left(E[Z|Z > z] - E[Z|Z \leq z]\right) \mathbb{P}(Z > z)\mathbb{P}(Z \leq z).$$

The constant is such that $E[\omega] = 1$.

Appendix 1.B Proofs

Section 1.2

Proof of proposition 1:

This follows from applying Brouwer's fixed point theorem to the following bounded continuous mapping with convex domain:

$$(M, P) \to \left(D(X, M, P), P - \left(S(P, X) - \sum_c M^c \right) \right). \tag{1.59}$$

The fixed points of this mapping are exactly the partial sorting equilibria. $\square$

Section 1.3

Proof of Lemma 1 :
Plugging 1.7 into 1.8 and differentiating w.r.t. X gives

$$S_X + S_P P_X^+ = \sum_c M_X^{+c} = \sum_c (D_X^c + D_P^c P_X^+) = E_X + E_P P_X^+.$$

Inelastic supply $S_P = 0$ and constancy $S_X = 0$ imply

$$P_X^+ = -\frac{E_X}{E_P}.$$

Analogously,

$$P_M^+ = -\frac{E_M}{E_P}$$

and

$$P_X^* = -\frac{E_X + E_M M_X^*}{E_P}.$$

By assumption 1.2.3 and iterated expectations, we can write $E = M^{tot} \cdot E[\mathbb{P}(u \geq u^o | u_X)]$. Denote $f^{u-u^o|u_X}$ the conditional density of $u - u^o$ given u_X, which exists according to assumption 1.2.3. We get

$$\frac{1}{M^{tot}} E_X = E\left[\frac{\partial}{\partial X} \mathbb{P}(u - u^o \geq 0 | u_X) \right] = E\left[u_X f^{u-u^o|u_X}(0|u_X) \right]$$

$$= \int u_X \frac{f^{u-u^o, u_X}(0, u_X)}{f(u_X)} f(u_X) du_X = f^{u-u^o}(0) E[u_X | u = u^o].$$

Similarly for E_M and E_P and for D. $\square$

Proof of Corollary 1:
Immediate from lemma 1, once we check that this density integrates to one and is non-

negative. $\square$

Proof of Lemma 2:
$u(X, M^*(X), P^*(X)) \geq u^o$ implies the first order condition

$$u_X + u_m M^*_X + u_P P^*_X = 0$$

for all the households choosing the given neighborhood. $\square$

Proof of Lemma 3:
For simplicity of notation, the superscript res will be omitted from reservation prices in this proof. Furthermore, assume for the moment that there are no social externalities, i.e., $u_M = 0$. The general case is completely analogous. By iterated expectations we can write

$$E[P|P \geq P^*] = E[E[P \cdot \mathbf{1}(P \geq P^*)|P_X]]/\mathbb{P}(P \geq P^*).$$

In integral notation, the conditional expectation is given by

$$E[P \cdot \mathbf{1}(P \geq P^*)|P_X] = \int_{P^*}^{\infty} P f(P|P_X) dP.$$

Differentiating this conditional expectation gives

$$\frac{\partial}{\partial X} E[P \cdot \mathbf{1}(P \geq P^*)|P_X] = P_X \cdot \mathbb{P}(P \geq P^*|P_X) + P^* \cdot (P_X - P^*_X) \cdot f^{P-P^*|P_X}(0|P_X).$$

The second term is due to the change in the boundaries of integration. Hence

$$\frac{\partial}{\partial X} E[P \cdot \mathbf{1}(P \geq P^*)] = E[P_X \mathbf{1}(P \geq P^*)] + P^* \cdot E[P_X - P^*_X|P = P^*] \cdot f^{P-P^*}(0).$$

Similarly

$$\frac{\partial}{\partial X} \mathbb{P}(P \geq P^*) = E[P_X - P^*_X|P = P^*] \cdot f^{P-P^*}(0).$$

Collecting terms then gives

$$\begin{aligned}
\frac{\partial}{\partial X} E[P|P \geq P^*] &= \left(E[P_X \mathbf{1}(P \geq P^*)] + P^* \frac{\partial}{\partial X} \mathbb{P}(P \geq P^*) \right) / \mathbb{P}(P \geq P^*) \\
&- E[P|P \geq P^*] \frac{\partial}{\partial X} \mathbb{P}(P \geq P^*) / \mathbb{P}(P \geq P^*) \\
&= E[P_X|P \geq P^*] \\
&- \left(\frac{\partial}{\partial X} \log \mathbb{P}(P \geq P^*) \right) \cdot \left[E\left[P^{res}|P \geq P^*\right] - P^* \right].
\end{aligned}$$

Finally, inelastic housing supply implies that, in equilibrium, the number of households must be constant, i.e., $\mathbb{P}(P \geq P^*)$ does not depend on X. $\square$

Proof of Proposition 2:
Identification follows from identification of supp $p(M,P|X)$, the fact that by assumption 1.2.2 supp $p(M,P|X) = (M^*(X), P^*(X))$ and equations 1.12 and 1.14. Non-identification is a corollary of lemma 5 below. $\square$

Proof of Lemma 4:
Identification follows from identification of $(M^*(X), P^*(X))$ and equations 1.12 and 1.14. Non-identification again follows from lemma 5 below. $\square$

Proof of Lemma 5:
Take the family of demand functions

$$\{\widetilde{D}(X,M,P) = f(X) + AM + BP : f \ arbitrary\}.$$

In the absence of multiple equilibria, this model is just identified, where we get $f(X) = (1-A)M^*(X) - BP^*(X)$, and an "estimate" of D of $\hat{D}(X,M,P) = (1-A)M^*(X) - BP^*(X) + AM + BP$, with $D_M = A$, $D_P = B$.
The proof for P^+ is completely analogous. $\square$

Proof of Proposition 3:
1.17 is immediate from $M_X^{*1} = D_X^1 + D_m^1 m_X^* + D_P^1 P_X^*$, once we have shown $m_{X^1}^* \neq 0$. Under assumption 1.2.4, $m_{X^1}^* = d_{X^1}/(1-d_m)$. Since $d = D^1/(D^1+D^2)$,

$$d_{X^1} = \frac{-D^1 D_{X^1}^2}{(D^1+D^2)^2} \neq 0$$

by assumption.
Equation 1.18 follows from 1.17 if we can show $\frac{M_{X^2}^{*1}}{P_{X^2}^*} = D_P^1$. Under assumption 1.2.4, $d_P = 0$, and by assumption of this lemma $d_{X^2} = 0$, hence $m_{X^2}^* = 0$. It follows that $M_{X^2}^{*1} = D_P^1 P_{X^2}^*$. Finally,

$$P_{X^2}^* = \frac{S_{X^2}}{E_P - S_P} \neq 0$$

again by assumption. $\square$

Proof of Corollary 2:
This follows from proposition 3 and lemma 18. $\square$

Proof of Proposition 4:
By equation 1.21 it is immediate that

$$m^k_{X^l} = d^k_{\widetilde{m}^k}\widetilde{m}^k_{X^l},$$

since $\widetilde{X}^k_{X^l} = 0$ by the assumption that the k,lth entry of $\mathbf{G}$ equals 0. Similarly

$$P^{+,k}_{X^l} = d^k_{\widetilde{m}^k}\widetilde{m}^k_{X^l}$$

and

$$M^{*c,k}_{X^l} = D^{c,k}_{\widetilde{m}}\widetilde{m}^k_{X^l} + D^{c,k}_P P^k_{X^l}.$$

To proof the claim, it remains to show that the denominator $\widetilde{m}^k_{X^l}$ does not equal zero. Differentiating equation 1.21 in its vector form, i.e., stacking up the equations for all neighborhoods, gives

$$\mathbf{d}_{\widetilde{\mathbf{m}}}\mathbf{G}\mathbf{m}_{\mathbf{X}} + \mathbf{d}_{\widetilde{\mathbf{X}}}\mathbf{G} = \mathbf{m}_{\mathbf{X}}$$

and hence

$$\mathbf{m}_{\mathbf{X}} = (I - \mathbf{d}_{\widetilde{\mathbf{m}}}\mathbf{G})^{-1}\mathbf{d}_{\widetilde{\mathbf{X}}}\mathbf{G}, \tag{1.60}$$

where I is the $\mathscr{N} \times \mathscr{N}$ identity matrix and $\mathbf{d}_{\widetilde{\mathbf{m}}}$ is a diagonal matrix with positive diagonal entries, by assumption. Invertibility of $(I - \mathbf{d}_{\widetilde{\mathbf{m}}}\mathbf{G})$ follows from the normalization of rows of $\mathbf{G}$ to sum to one, and $d_{\widetilde{m}} < 1$. We can expand equation 1.60 as a geometric series,

$$\mathbf{m}_{\mathbf{X}} = \left(\sum_{j\geq 0} (\mathbf{d}_{\widetilde{\mathbf{m}}}\mathbf{G})^j\right)\mathbf{d}_{\widetilde{\mathbf{X}}}\mathbf{G}. \tag{1.61}$$

All of the terms in the series have non-negative entries, the k,lth entry of $\mathbf{G}^j$ is not equal 0 for some power j by assumption, the same holds for $(\mathbf{d}_{\widetilde{\mathbf{m}}}\mathbf{G})^j$ by $\mathbf{d}_{\widetilde{\mathbf{m}}}$ being a diagonal matrix with positive diagonal entries, and finally $\mathbf{d}_{\widetilde{\mathbf{X}}}\mathbf{G}$ has non-zero diagonal entries. □

Proof of Corollary 3:
This follows from proposition 3, lemma 17 and lemma 18. □

Section 1.4

Proof of Lemma 6:
We can divide households of type c into four classes, depending on whether or not they live in the neighborhood (indexed by 1 and o) and depending on whether the want to stay (s) or to move (m) into or out of the neighborhood. Denote these classes by $D^{c,1,s}, \ldots, D^{c,o,m}$. By definition $M^c = D^{c,1,s} + D^{c,1,m}$ and $D^c = D^{c,1,s} + D^{c,o,m}$. A fraction λ of those that want

to move will be successful per time unit, giving

$$\begin{aligned}\dot{M}^c &= \lambda\left(D^{c,o,m} - D^{c,1,m}\right) = \lambda\left(\left(D^{c,1,s} + D^{c,o,m}\right) - \left(D^{c,1,s} + D^{c,1,m}\right)\right) \\ &= \lambda\left(D^c - M^c\right).\end{aligned}$$

□

Proof of Lemma 7:
Recalling the definitions $m = M^1/(M^1+M^2)$ and $d = D^1/(D^1+D^2)$, and using the result of the previous lemma,

$$\dot{m} = \frac{\partial m}{\partial M}\dot{M} = \lambda \cdot \frac{\partial m}{\partial M} \cdot (D - M) = \breve{\lambda} \cdot (d - m),$$

where

$$\breve{\lambda} := \lambda \cdot \frac{\frac{\partial m}{\partial M} \cdot (D - M)}{d - m} = \lambda \cdot \frac{D^1 + D^2}{M^1 + M^2}.$$

The second equality in this expression follows from

$$\frac{\frac{\partial m}{\partial M} \cdot (D - M)}{d - m} = \frac{\frac{1}{(M^1+M^2)^2}\left(M^2, -M^1\right) \cdot \left(D^1 - M^1, D^2 - M^2\right)'}{\frac{D^1}{D^1+D^2} - \frac{M^1}{M^1-M^2}} = \frac{D^1 + D^2}{M^1 + M^2}.$$

By assumption 1.2.4, the price and scale elasticities of both types are identical and hence $d = d(X, m)$. Therefore $\dot{m} = \breve{\lambda} \cdot (d(X, m) - m)$.
Taking the time path of d and $\breve{\lambda}$ as given, the solution to this differential equation can be written as

$$m^t = m^0 e^{-\int_0^t \breve{\lambda} ds} + \int_0^t \breve{\lambda} d e^{-\int_s^t \breve{\lambda} du} ds.$$

This gives m^t as a weighted average of initial m^0 and d in the time interval from 0 to t. Letting $\kappa = 1 - e^{-\int_0^1 \breve{\lambda} ds}$ and (m, X) some appropriate intermediate values in the time interval $[0, 1]$ the claim follows. □

Proof of Lemma 8:
From equation 1.39 it is immediate that, for any given household, $P_{X,m} = \frac{(u_X, u_M)}{-u_P}$. By assumption, due to search frictions, M has a smooth time path and in particular $\frac{\partial}{\partial \xi} \lim_{t \to 0^+} M = 0$. □

Proof of Lemma 9: For any given household, it can be shown as in lemma 8 that $P_{X,m} = -\frac{u_X + u_M M_\xi^{lr}}{u_P}$. To prove the claim we have to show, that resorting of households according to willingness to pay has no first order effect on the average reservation price within the neighborhood. But this follows immediately from lemma 3. □

Proof of Lemma 10:
First, $M \in M^*$ are the only constant solutions of the differential equation 1.36: Any stable solution must imply $M = D$. By constancy of X, M and u, u^o, $\dot{V} = 0$ and $V = u/r$, $V^o = u^o/r$. Hence $V > V^o$ if and only if $u > u^o$, and D is equal to demand $\widetilde{D}$ in the corresponding static model. A landlord accepts a tenant if and only if W for this tenant is greater than W^v, i.e., if

$$P = rW \geq rW^v = \frac{r\mu}{r+\mu} E\left[P^{new}\right].$$

By random matching $E\left[P^{new}\right] = E^s[P^{res}|P^{res} > P^*]$, and hence D equals to demand $\widetilde{D}$ of the corresponding static model.
The claim follows, since any limit of a converging path must satisfy $\dot{M} = 0$. □

Proof of Lemma 11:
Let w.l.o.g. $t^0 = 0$. If we denote $V^{\max} = \max(V^o, V)$ and impose a transversality condition, we can solve equation 1.31 for V and get

$$V = \int_0^\infty e^{-\int_0^t (r+\lambda) ds} \left[u(X, M, P) + \lambda V^{\max}\right] dt. \tag{1.62}$$

This is again to be understood as a conditional expectation given the information set at time 0. A similar equation holds for V^o.
Equation 1.62 implies

$$V - V^o = \int_0^\infty e^{-\int_0^t (r+\lambda) ds} \left[u - u^o\right] dt$$

and hence

$$\frac{V - V^o}{\int_0^\infty e^{-\int_0^t (r+\lambda) ds} dt} = \int_0^\infty w^t \left[u - u^o\right] dt, \tag{1.63}$$

where we denote

$$w^t := \frac{\phi^t}{\int_0^\infty \phi^t dt}$$

for $\phi^t = e^{-\int_0^t (r+\lambda) ds}$ The weights w^t integrate to one. Let $\epsilon > 0$ be such that $r + \lambda > C^1$ and $|(u^{1,t} - u^{2,t}) - (u^{1,0} - u^{2,0})| < \delta$ on the interval $[0, \epsilon]$, and assume $r + \lambda > C^2$ on $[0, \infty)$. We get

$$\left|(u^{1,0} - u^{2,0}) - \int_0^\infty w^t \left[u - u^o\right] dt\right| < C^3 \delta + (1 - C^3) \sup_t (u - u^o) \tag{1.64}$$

for $C^3 = \int_0^\epsilon w^t dt$. Some algebraic manipulation yields

$$C^3 = \frac{1}{1 + \frac{\phi^\epsilon}{\int_0^\epsilon \phi^t dt} \frac{\int_\epsilon^\infty \phi^t dt}{\phi^\epsilon}}.$$

By $r + \lambda > C^1$ on $[0, \epsilon]$

$$\int_0^\epsilon \phi^t dt > \phi^\epsilon \int_0^\epsilon e^{C^1[\epsilon - t]} dt = \frac{\phi^\epsilon}{C^1} \left[e^{C^1 \epsilon} - 1 \right].$$

By $r + \lambda > C^2$ on $[\epsilon, \infty)$

$$\int_\epsilon^\infty \phi^t dt < \phi^\epsilon \int_\epsilon^\infty e^{-C^2[t-\epsilon]} dt = \frac{\phi^\epsilon}{C^2}.$$

Hence, as $C^1\epsilon \to \infty$

$$C^3 > \frac{1}{1 + \frac{C^1 C^2}{e^{C^1\epsilon} - 1}} \to 1.$$

The claim now follows from equation 1.64. □

Section 1.5

Proof of Proposition 5:
Claims 1 through 3 follow from the lemmata 6 through 10.
Claim 4 follows from lemma 3, because the resorting of marginal households according to willingness to pay has no first order effect on the average willingness to pay if housing supply is constant. □

Proof of Corollary 4: This follows from proposition 5 and lemma 17, once we note that under the given assumptions

$$\frac{\partial \log P}{\partial m} = E\left[-\frac{u_m}{u_{\log P}} \right]$$

for all neighborhoods and times. □.

Proof of Lemma 20:
By definition of conditional quantiles, $F^{\Delta m|m}\left(Q^{\Delta m|m}(\tau|m)|m\right) = \tau$. Differentiating this with respect to m gives

$$\frac{\partial}{\partial m} Q^{\Delta m|m}(\tau|m) = -\frac{\frac{\partial}{\partial m} F^{\Delta m|m}(Q|m)}{f^{\Delta m|m}(Q|m)}. \tag{1.65}$$

The differential in the numerator has two components, one due to the structural relation between Δm and m, i.e., the derivative with respect to the argument m of $d(m, X) - m$, and one due to the stochastic dependence of m and X.

$$\begin{aligned}\frac{\partial}{\partial m} F^{\Delta m|m}(Q|m) &= E\left[\kappa \cdot (d_m - 1) \cdot f^{\Delta m|d_m,m,\kappa}(Q|d_m, m, \kappa)\big| m\right] \\ &+ \frac{\partial}{\partial m} \mathbb{P}\left(\kappa \cdot (d(m', X) - m') \leq Q|m\right)\Big|_{m'=m}.\end{aligned}$$

This can be seen as follows: We can decompose the derivative according to

$$\frac{\partial}{\partial m} F^{\Delta m|m}(Q|m) = \left[\frac{\partial}{\partial m'} + \frac{\partial}{\partial m}\right] \mathbb{P}\left(\kappa \cdot (d(m', X) - m') \leq Q|m\right)\Big|_{m'=m}.$$

To simplify the first derivative, note that by iterated expectations

$$\mathbb{P}\left(\kappa \cdot (d(m', X) - m') \leq Q|m\right) = E[F(\kappa \cdot (d(m', X) - m')|m, \kappa, d_m)|m].$$

Differentiating this with respect to m' gives

$$E\left[\kappa \cdot (d_m - 1) \cdot f^{\Delta m|d_m,m}(Q|d_m, m)|m\right].$$

The claim now is immediate. □

Proof of Proposition 10:
Since m and $m+\Delta m$ have their support in the interval $[0, 1]$, $Q^{\Delta m|m}(\tau|0) \geq 0$ and $Q^{\Delta m|m}(\tau|1) \leq 0$. Therefore the uniqe root m of $Q^{\Delta m|m}(\tau|m)$ must be stable, $\frac{\partial}{\partial m} Q^{\Delta m|m}(\tau|m) \leq 0$.
By lemma 20 and assumption 1.5.1, this implies that $E\left[\kappa \cdot (d_m - 1)|\Delta m = Q, m\right] \leq 0$.
Finally, note that for all m where $(0, m)$ is in the support of $(\Delta m, m)$, there exists a τ such that $Q^{\Delta m|m}(\tau|m) = 0$. □

Appendix 1.A

Proof of lemma 13:
Since $Cov(\alpha^i, Z) = 0$, we have $Cov(Y, Z) = Cov(\beta^i X, Z) = E[\beta^i X(Z - E[Z])]$ □

Proof of lemma 14:
By the Frisch-Waugh theorem, $\beta^{1,OLS} = \frac{E[Ye]}{E[X^1 e]}$, where e is the residual from OLS regression of X^1 on X^2. By linearity of $E[X^{2,i}\beta^{2,i} + \epsilon|X^2]$ and independence of $\beta^{2,i}, \epsilon^i$ and $(X^{1,i}, X^{2,i})$, $E[Ye] = E[\beta^{1,i}X^{1,i}e]$. By independence of $\beta^{1,i}$ and $(X^{1,i}, X^{2,i})$, $E[\beta^{1,i}X^{1,i}e|X^{2,i}] = E[\beta^{1,i}|X^2]E[X^1e|X^2]$.

The claim then follows from iterated expectations. □

Proof of lemma 15:
$\beta^{1,IV} = \frac{E[Ye]}{E[X^1 e]}$ follows again from the Frisch-Waugh theorem, applied to the two-stage least-squares representation of $\beta^{1,IV}$, and $E[Ye] = E[\beta^{1,i}X^{1,i}e]$ from linearity of $E[X^{2,i}\beta^{2,i} + \epsilon|X^2]$ and conditional independence $Z \perp (\beta^2, \epsilon)|X^2$. □

Proof of lemma 16:
Immediate from lemma 13, with differences replacing levels. □

Proof of lemma 17:
Under appropriate smoothness assumptions, we can write

$$\Delta Y = \int_0^1 \left(g_X(X^{it}, \epsilon^{it}) \cdot X_t + g_\epsilon(X^{it}, \epsilon^{it}) \cdot \epsilon_t\right) dt.$$

By exogeneity of the instrument, we then get

$$Cov(\Delta Y, \Delta Z) = E\left[\int_0^1 g_X(X^{it}, \epsilon^{it}) \cdot X_t \, dt(\Delta Z - E[\Delta Z])\right] == E\left[g_X \cdot \omega\right].$$

□

Proof of lemma 18:
This is an immediate extension of lemma 17. □

Proof of lemma 19:
First, consider the covariance of X and Z. Denote $\mu(Z) := E[X|Z] = E[h(Z,\eta)|Z]$. Then

$$Cov(X, Z) = E[\mu(Z)(Z - E[Z])] = \int_{-\infty}^{\infty}\int_{-\infty}^{z} \mu_z(\tilde{z})(z - E[Z]) f(z) d\tilde{z}dz =$$

$$= \int_{-\infty}^{\infty} \mu_z(\tilde{z}) \int_{\tilde{z}}^{\infty} (z - E[Z]) f(z) dz d\tilde{z} = \int_{-\infty}^{\infty} \mu_z(\tilde{z}) \left(E[Z|Z > z] - E[Z|Z \le z]\right) \mathbb{P}(Z > z)\mathbb{P}(Z \le z) d\tilde{z}dz =$$
$$= E\left[h_z(Z, \eta)\tilde{\omega}(Z)\right],$$

where

$$\tilde{\omega}(z) := \frac{1}{f(z)}\left(E[Z|Z > z] - E[Z|Z \le z]\right) \mathbb{P}(Z > z)\mathbb{P}(Z \le z).$$

Similarly,

$$Cov(Y, Z) = E\left[g_x(X, \epsilon) h_z(Z, \eta)\tilde{\omega}(Z)\right].$$

The assertion follows from $\beta^{IV} = \frac{Cov(Y,Z)}{Cov(X,Z)}$. □

Appendix 1.C Figures and tables

Table 1.1: Naive hedonic and demand regressions

		log non-Hisp pop	log Hisp pop	log mean imputed rent
Cross-section	Hisp shr	-1.815	5.616	-0.476
		(0.023)	(0.039)	(0.005)
	log mean	0.117	0.198	
	imputed rent	(0.014)	(0.031)	
Differences	Hisp shr	-1.674	5.946	-0.321
		(0.025)	(0.064)	(0.008)
	log mean	0.398	-0.278	
	imputed rent	(0.014)	(0.015)	
Differences	Hisp shr	-1.681	7.433	0.293
w. controls		(0.027)	(0.076)	(0.009)
	log mean	0.378	0.555	
	imputed rent	(0.014)	(0.037)	

Notes: This table shows demand regressions of log non-Hispanic population and log Hispanic population on Hispanic share and log mean imputed rents, as well as hedonic regressions of log mean imputed rents on Hispanic share. The first specification is a pooled cross-sectional regression using data from 1980, 1990 and 2000, the second and third are regressions in decadal differences for the 80s and 90s. All regressions control for MSA × time fixed effects, the third specification additionally for initial Hispanic share and its square.

Table 1.2: INSTRUMENTAL VARIABLE ESTIMATES, DECADAL CHANGES IN THE 80S AND 90S

	first stage	IV regressions		
Instrument		log non-Hisp pop	log Hisp pop	log mean imputed rent
subgroup	0.146	-8.360	–	–
	(0.016)	(0.740)		
spatial	0.119	-6.251	3.437	-0.758
	(0.007)	(0.620)	(0.733)	(0.119)
dynamic	0.198	–	–	-0.516
	(0.011)			(0.049)

Notes: This table shows instrumental variables regressions of the change in log non-Hispanic population, log Hispanic population, and mean imputed rent on the change in Hispanic share using the instruments discussed in the text. All regressions pool data for the 80s and the 90s and control for time × MSA fixed effects. The subgroup and dynamic instrument regressions control for initial Hispanic share and its square, the spatial instrument regressions control for predicted immigration.

Table 1.3: THEORETICAL INTERPRETATION OF THE ENTRIES OF TABLE 1.2

	first stage	IV regressions		
Instrument		log non-Hisp pop	log Hisp pop	log mean imputed rent
subgroup	$m^*_{X^I}$	$\frac{M^{*2}_{X^I}}{m^*_{X^I}} =$	–	–
		$\mathbf{D}^{\mathbf{2}}_{\mathbf{m}} + D^2_P \frac{P^*_{X^I}}{m^*_{X^I}}$		
spatial	$\widetilde{m}_{X^{>3}}$	$\frac{M^{*2}_{X^{>3}}}{\widetilde{m}_{X^{>3}}} =$	$\frac{M^{*1}_{X^{>3}}}{\widetilde{m}_{X^{>3}}} =$	$\frac{P_{X^{>3}}}{\widetilde{m}_{X^{>3}}} = \mathbf{P}^{+}_{\widetilde{\mathbf{m}}} =$
		$\mathbf{D}^{\mathbf{2}}_{\widetilde{\mathbf{m}}} + D^2_P \frac{P^*_{X^{>3}}}{\widetilde{m}_{X^{>3}}}$	$\mathbf{D}^{\mathbf{1}}_{\widetilde{\mathbf{m}}} + D^1_P \frac{P^*_{X^{>3}}}{\widetilde{m}_{X^{>3}}}$	$\widetilde{E}\left[-\frac{\mathbf{u}_{\widetilde{\mathbf{m}}}}{\mathbf{u}_{\mathbf{P}}} \middle\| u = u^o\right]$
dynamic	Δm_{X^L}	–	–	$\frac{\Delta P_{X^L}}{\Delta m_{X^L}} =$
				$E\left[-\frac{\mathbf{u}_{\mathbf{m}}}{\mathbf{u}_{\mathbf{P}}}\right]$

Notes: This table shows the theoretical interpretations of the first stage and instrumental variable coefficients displayed in table 1.2. The regression coefficients estimate weighted averages of the slopes shown here.

Table 1.4: Subsample of MSAs with large Hispanic share - Instrumental Variable estimates

	first stage		IV regressions	
Instrument		log non-Hisp pop	log Hisp pop	log mean imputed rent
subgroup	0.146	-8.262	–	–
	(0.016)	(0.742)		
spatial	0.114	-6.419	0.651	-0.760
	(0.007)	(0.677)	(0.762)	(0.128)
dynamic	0.210	–	–	0.210
	(0.011)			(0.011)

Notes: This table replicates table 1.2 for the subset of cities with Hispanic shares larger than 8% in 2000, which corresponds to roughly 50% of the neighborhoods in the full sample.

Table 1.5: Subsample excluding California and New York - Instrumental Variable estimates

	first stage		IV regressions	
Instrument		log non-Hisp pop	log Hisp pop	log mean imputed rent
subgroup	0.043	-33.575	–	–
	(0.024)	(17.116)		
spatial	0.122	-8.257	5.513	-1.021
	(0.010)	(0.891)	(1.046)	(0.163)
dynamic	0.171	–	–	-0.981
	(0.016)			(0.092)

Notes: This table replicates table 1.2 for the subset of cities outside the states of California and New York.

Table 1.6: Decades separately, different housing price variables - Instrumental Variable estimates

		first stage	IV regressions, dependent variable is log -				
sample	Instrument		non-Hisp pop	Hisp pop	mean imputed rent	median rent	median house value
80s	subgroup	0.146	-9.515	–	–	–	–
		(0.027)	(1.171)				
	spatial	0.093	-8.673	3.181	-1.395	0.293	NA
		(0.008)	(1.147)	(1.218)	(0.221)	(0.331)	
	dynamic	0.181	–	–	-0.092	0.008	NA
		(0.015)			(0.085)	(0.130)	
90s	subgroup	0.285	-4.67055	–	–	–	–
		(0.024)	(0.475)				
	spatial	0.16	-4.053	3.665	-0.134	-0.429	-0.757
		(0.012)	(0.639)	(0.853)	(0.108)	(0.236)	(0.207)
	dynamic	0.254	–	–	-0.343	-0.606	-0.575
		(0.016)			(0.049)	(0.098)	(0.147)

Notes: This table replicates table 1.2 for the the 1980s and 1990s separately, and includes alternative measures of housing costs as dependent variables. For 1980, median house value is not available.

Table 1.7: .95 CONFIDENCE SETS FOR $Z(g)$ FOR THE LARGEST MSAS BY DECADE AND QUANTILE

place	**80s**			**90s**		
	$q = .2$	$q = .5$	$q = .8$	$q = .2$	$q = .5$	$q = .8$
New York	[1,1]	[0,0]	[0,0]	[0,0]	[1,1]	[0,0]
Los Angeles	[0,0]	[1,1]	[1,1]	[1,1]	[1,1]	[1,1]
Chicago	[1,1]	[1,1]	[0,0]	[0,0]	[0,0]	[0,0]
Houston	[0,1]	[0,0]	[0,0]	[0,1]	[1,1]	[0,0]
Phoenix	[1,3]	[0,0]	[0,0]	[1,1]	[0,0]	[0,0]
Philadelphia	[1,3]	[0,0]	[0,1]	[1,1]	[0,1]	[0,0]
San Antonio	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]
Dallas	[1,1]	[0,0]	[0,0]	[1,2]	[0,0]	[0,0]
San Diego	[1,1]	[0,0]	[0,0]	[1,1]	[1,1]	[1,1]
San Jose	[0,1]	[0,0]	[0,0]	[1,1]	[1,1]	[0,0]
San Francisco	[1,1]	[1,1]	[0,0]	[2,2]	[1,1]	[0,0]

Notes: The table shows confidence intervals in the integers for $Z(g)$ for the ten largest MSAs of the Unites States, and for San Francisco, ordered by population size, where g is estimated by local linear quantile regression of the change in Hispanic share over a decade on the initial Hispanic share for the quantiles $q =$ .2, .5 and .8. Regression bandwidth τ is $n^{-.2}$, ρ is chosen as .04. Confidence sets are based on t-statistics using bootstrapped bias and standard errors.

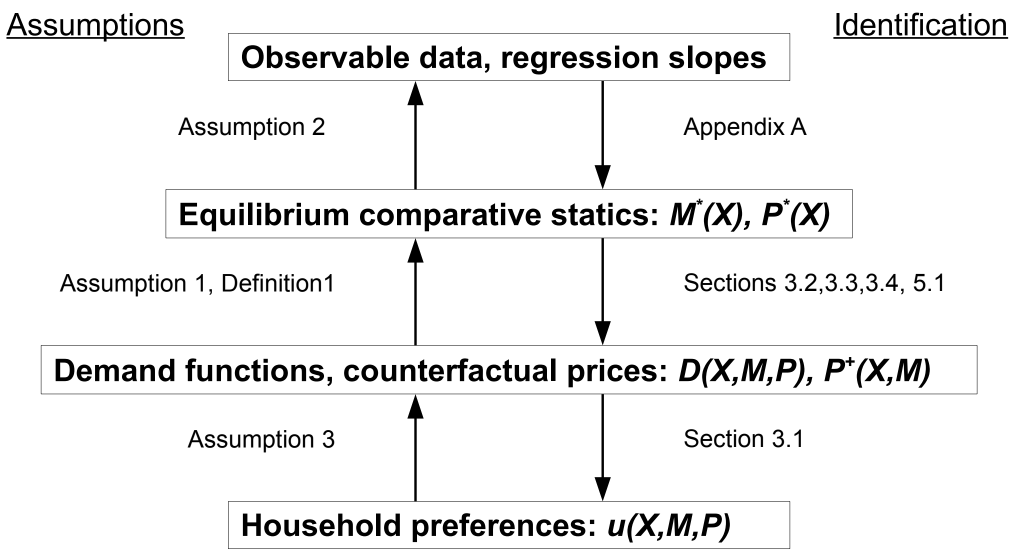


Figure 1.1: Assumptions and steps of the identification problem

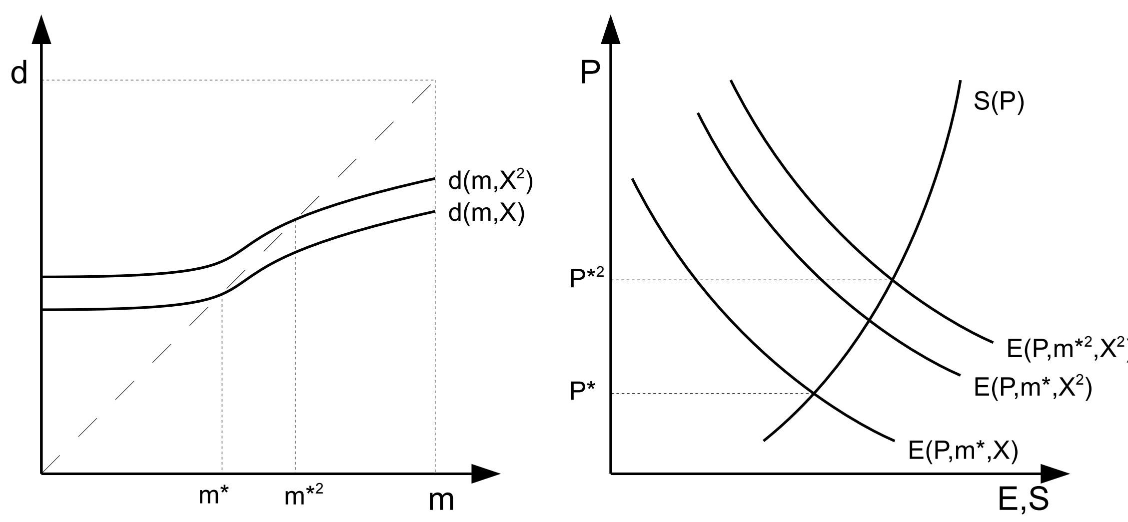


Figure 1.2: Comparative statics in the simplified $\mathscr{C} = 2$ model

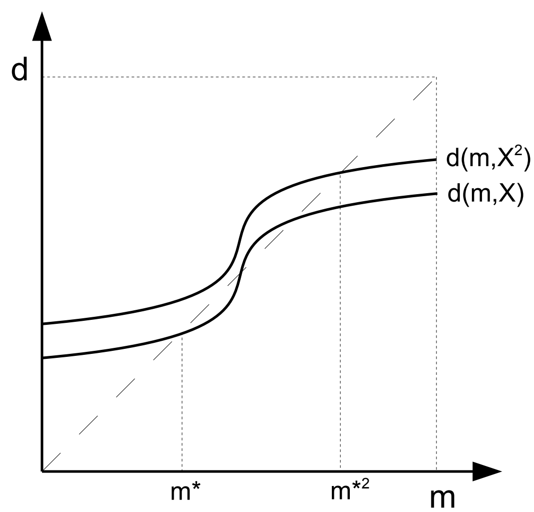


Figure 1.3: Multiple equilibria and tipping

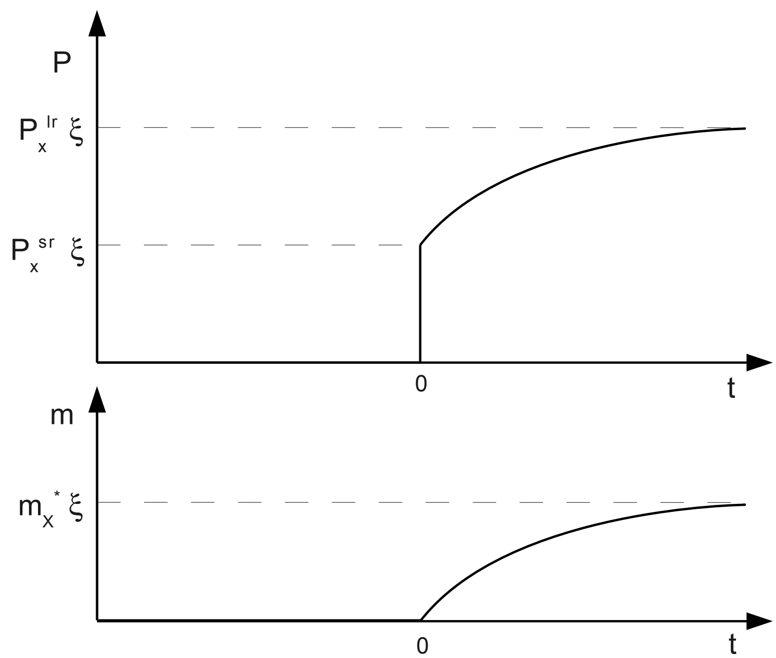
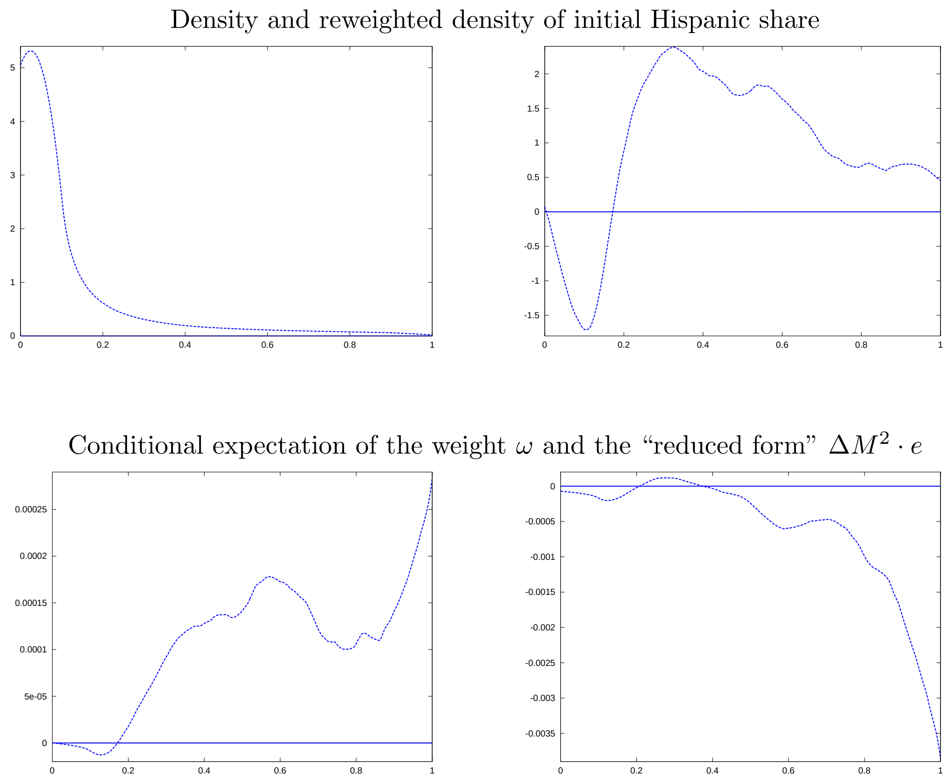


Figure 1.4: Dynamic response to shock in X

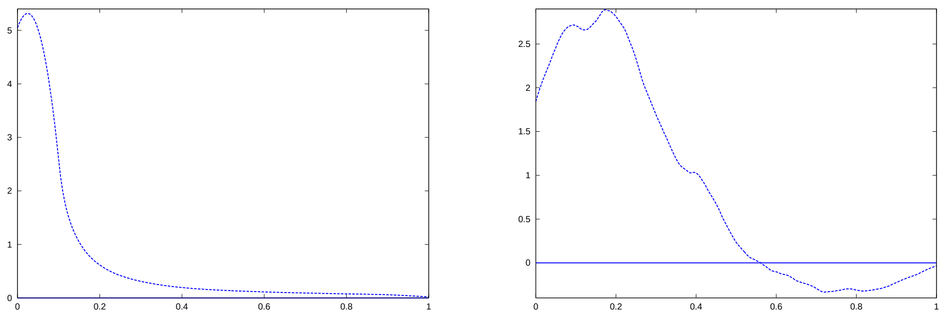
Figure 1.5: Decomposition of the subgroup instrumental variable estimate



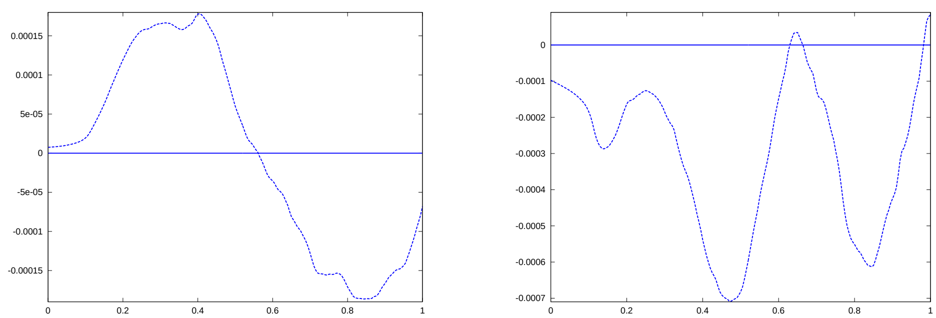
Notes: These graphs decompose the IV estimate based the subgroup instrument shown in table 1.2, according to lemma 15. The top row shows a kernel estimate of the density of initial Hispanic share across neighborhoods in the sample, as well as this density reweighted to give the distribution among the population for which the LATE is estimated. The bottom row shows kernel estimates of the conditional expectation of the weight ω, as well as the "conditional reduced form", $\Delta M^2 \cdot e$.

Figure 1.6: DECOMPOSITION OF THE SPATIAL INSTRUMENTAL VARIABLE ESTIMATE

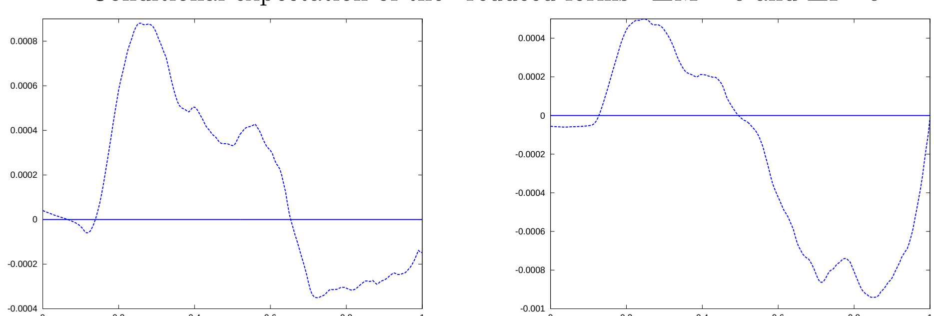
Density and reweighted density of initial Hispanic share



Conditional expectation of the weight ω and the "reduced form" $\Delta M^2 \cdot e$



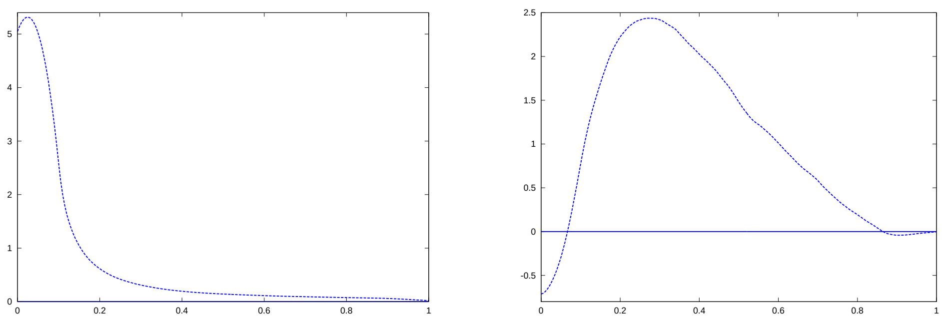
Conditional expectation of the "reduced forms" $\Delta M^1 \cdot e$ and $\Delta P \cdot e$



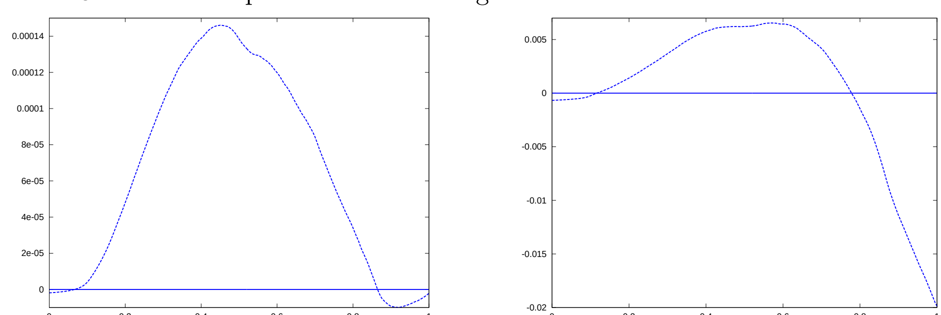
Notes: These graphs replicate those of figure 1.5 for the spatial instrument, and display conditional reduced forms for the additional outcome variables M^1 and P.

Figure 1.7: DECOMPOSITION OF THE DYNAMIC INSTRUMENTAL VARIABLE ESTIMATE

Density and reweighted density of initial Hispanic share

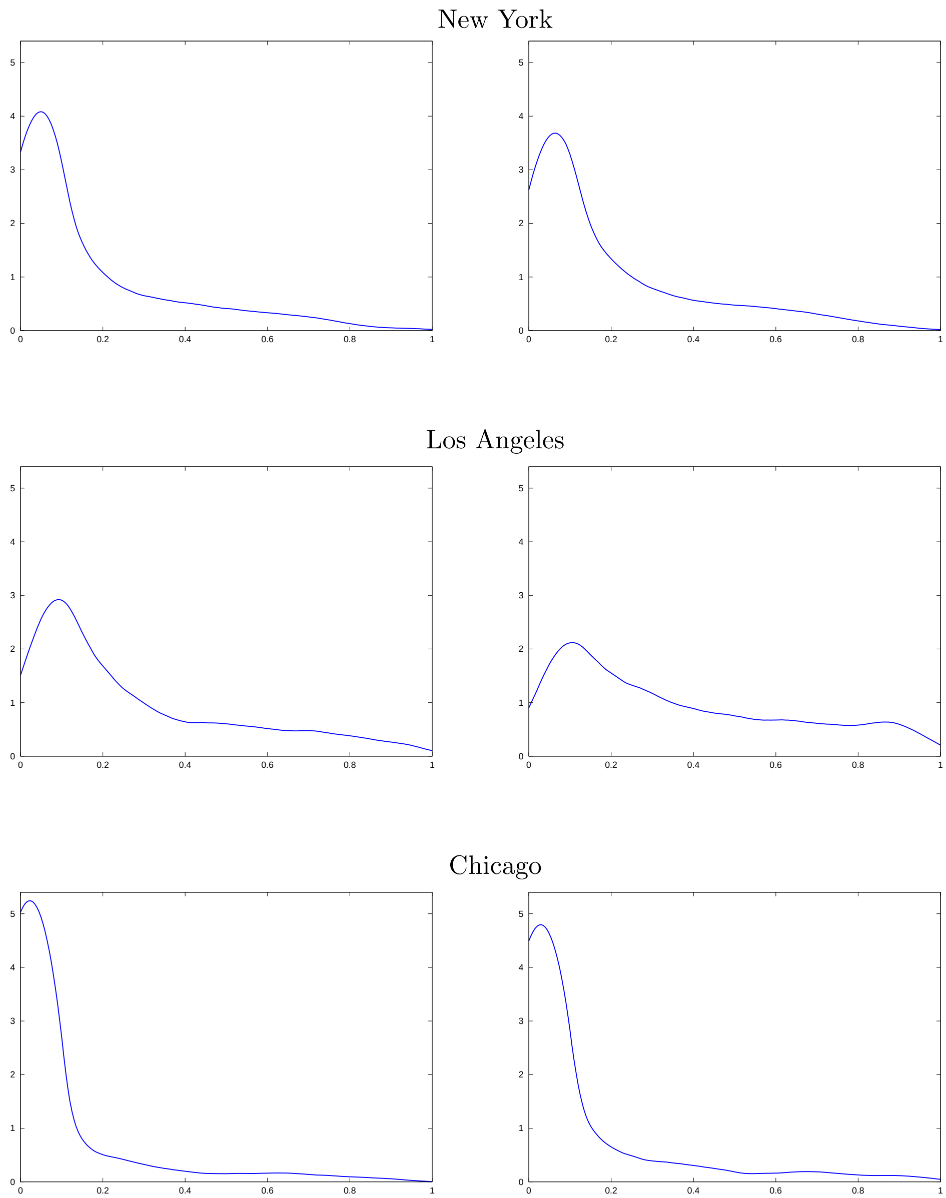


Conditional expectation of the weight ω and the "reduced form" $\Delta P \cdot e$



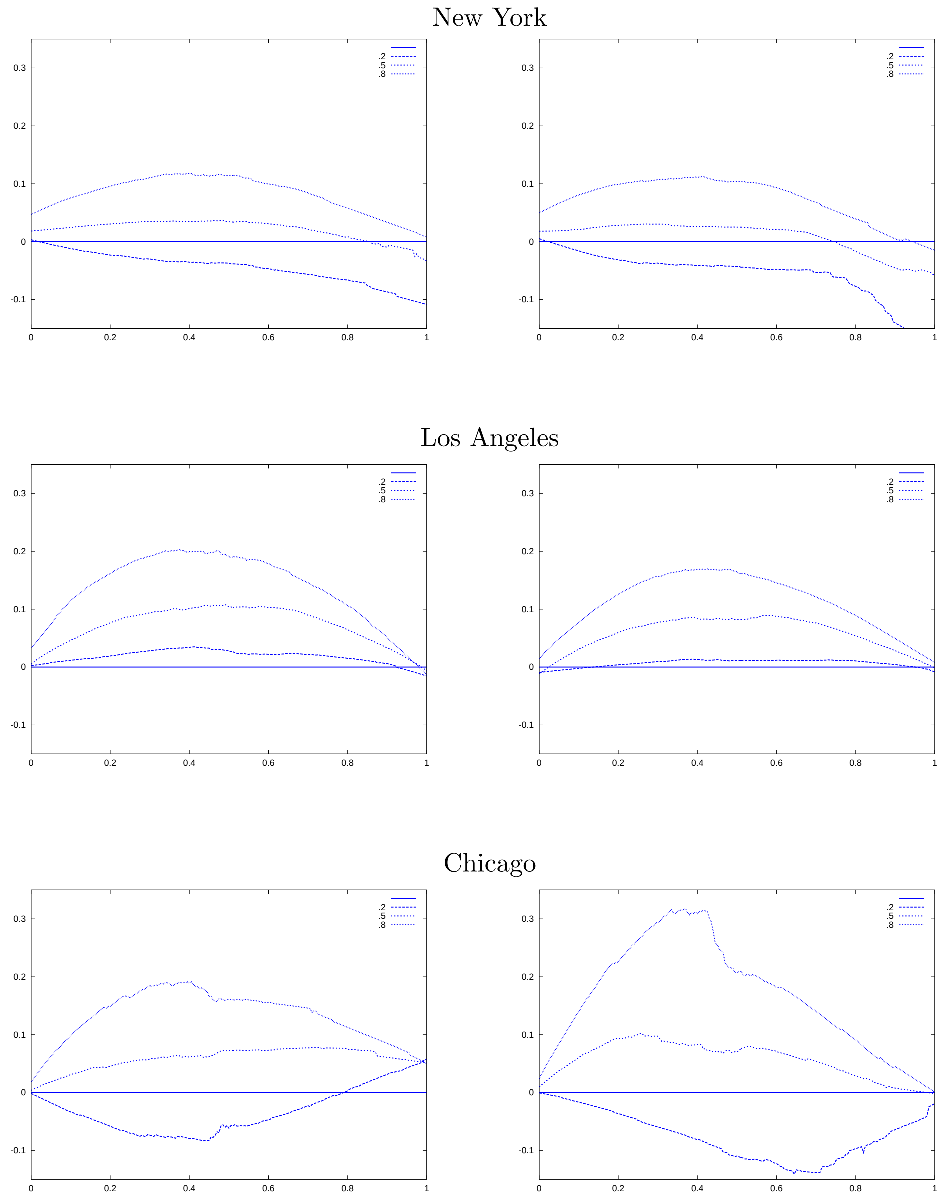
Notes: These graphs replicate those of figure 1.5 for the dynamic instrument, where the conditional reduced form is for the outcome variable P.

Figure 1.8: DENSITY OF INITIAL HISPANIC SHARE IN 1980 AND 1990



Notes: These graphs show kernel density estimates of the distribution of initial Hispanic share across neighborhoods for the years 1980 (left column) and 1990 (right column).

Figure 1.9: Quantile regressions of the change in Hispanic share on initial Hispanic share over the 1980s and 1990s



Notes: These graphs show local linear quantile regressions of the change in Hispanic share and on initial Hispanic share, .2, .5 and .8th conditional quantile, for the 80s (left column) and 90s (right column). The graphs do not show confidence bands.

Chapter 2

Nonparametric inference on the number of equilibria

Abstract

This paper proposes an estimator and develops an inference procedure for the number of roots of functions which are nonparametrically identified by conditional moment restrictions. The estimator is superconsistent, and the inference procedure is based on non-standard asymptotics. This procedure is used to construct confidence sets for the number of equilibria of static games of incomplete information and of stochastic difference equations. In an application to panel data on neighborhood composition in the United States, no evidence of multiple equilibria is found.

2.1 Introduction

Some economic systems show large and persistent differences in outcomes even though the observable exogenous factors influencing these systems differ little.[1] One explanation for such persistent differences in outcomes is multiplicity of equilibria. If a system indeed has multiple equilibria, temporary, large interventions might have a permanent effect, by shifting the equilibrium attained, while long-lasting, small interventions might not have a permanent effect. This paper develops a superconsistent estimator[2] and an inference procedure on the number of equilibria of economic systems. Suppose the equilibria of a system can be represented as solutions to the equation $g(x) = 0$, and suppose g can be identified by some conditional moment restriction. The procedure proposed here provides confidence sets for the number $Z(g)$ of solutions to the equation $g(x) = 0$.

[1] "System" might refer to households, firms, urban neighborhoods, national economies, etc.

[2] An estimator is called superconsistent if it converges at a rate faster than the usual parametric rate, which equals the square root of the sample size.

Some examples help motivate interest in the number of equilibria. Multiple equilibria and poverty traps are discussed by Dasgupta and Ray (1986), Azariadis and Stachurski (2005), and Bowles, Durlauf, and Hoff (2006). Poverty traps can arise, for instance, if an individual's productivity is a function of her income and if wage income reflects productivity, as in models of efficiency wages. Productivity might depend on wages because nutrition and health are improving with income. If this feedback mechanism is strong enough, there might be multiple equilibria, and extreme poverty might be self-perpetuating. In that case, public investments in nutrition and health can permanently lift families out of poverty.

Multiple equilibria and urban segregation are discussed by Becker and Murphy (2000) and Card, Mas, and Rothstein (2008). Urban segregation, along ethnic or sociodemographic dimensions, might arise because households' location choices reflect a preference over neighborhood composition. If this preference is strong enough, different compositions of a neighborhood can be stable, given constant exogenous neighborhood properties. Transition between different stable compositions might lead to rapid composition change, or "tipping," as in the case of gentrification of a neighborhood. Interest in such tipping behavior motivated Card, Mas, and Rothstein (2008), and is the focus of the application discussed in section 2.4 of this paper.

Multiple equilibria and the market entry of firms are discussed by Bresnahan and Reiss (1991) and Berry (1992). Entering a market might only be profitable for a firm if its competitors do not enter that same market. As a consequence, different configurations of which firms serve which markets might be stable.

In sociology, finally, multiple equilibria are of interest in the context of social norms. If the incentives to conform to prevailing behaviors are strong enough, different behavioral patterns might be stable norms, i.e., equilibria, see Young (2008). Transitions between such stable norms correspond to social change. One instance where this has been discussed is the assimilation of immigrant communities into the mainstream culture of a country.

We will discuss two general setups that allow us to translate the hypothesis of multiple equilibria into a hypothesis on the number of roots of some identifiable function g; these setups are (i) static games of incomplete information and (ii) stochastic difference equations. In section 2.3.4, a nonparametric model of static games of incomplete information, similar to the one analyzed in Bajari, Hong, Krainer, and Nekipelov (2006), will be discussed.[3] Assume there are two players $i = 1, 2$, who both have to choose between one of two actions, $a = 0, 1$. Player i makes her choice based on public information s, as well as private information ϵ_i. The public information s is observed by the econometrician, and ϵ_i is, assumed to be independent of s.[4] Denote the probability that player i plays strategy $a = 1$ given the public information s by $\sigma_i(s)$. Player i's expected utility given her information, and hence her optimal action a_i, depend on s and ϵ_i, as well as player $-i$'s probability of choosing $a = 1$,

[3] Note that this paper does *not* contribute to the literature discussing estimation problems in games of complete information with multiple equilibria.

[4] Note that this excludes, for instance, correlated value auctions.

$\sigma_{-i}(s)$. Let us denote the average best response of player i, integrating over the marginal distribution of ϵ_i, by

$$g_i\left(\sigma_{-i}, s\right) = E\left[a_i|\sigma_{-i}, s\right]. \tag{2.1}$$

Figure 2.1 illustrates, by plotting the response functions g_i for given s. In Bayesian Nash Equilibrium, the probability of player i choosing $a = 1$, σ_i, equals the average best response of player i, g_i. This implies the two equilibrium conditions

$$\sigma_i(s) = g_i\left(\sigma_{-i}(s), s\right),$$

for $i = 1, 2$. In figure 2.1, the Bayesian Nash Equilibria correspond to the intersections of the graphs of the two g_i. If we impose exclusion restrictions, the response functions g_i are identifiable from the equilibrium probabilities $\sigma_i(s)$, and this in turn allows to identify the equilibria which are not directly observable. Note that no functional form restrictions are needed for identification of the choice functions g_i. This stands in contrast to Bajari, Hong, Krainer, and Nekipelov (2006), who need to impose such restrictions in order to be able to identify the underlying preferences. Bayesian Nash Equilibrium in this game requires $g(\sigma_1, s) = 0$, where

$$g(\sigma_1, s) = g_1(g_2(\sigma_1, s), s) - \sigma_1. \tag{2.2}$$

The number of roots of $g(\sigma_1, s)$ in σ_1 is the number of Bayesian Nash Equilibria in this game, given s.

As a second general setup, in section 2.3.5 we will consider data generated by the difference equation

$$\Delta X_{i,t+1} = X_{i,t+1} - X_{i,t} = g(X_{i,t}, \epsilon_{i,t}). \tag{2.3}$$

Holding ϵ constant, the number of roots of g in X is the number of equilibria of this difference equation. If ϵ is stochastic, the number of roots can still serve to characterize qualitative dynamics in terms of "equilibrium regions", as will be discussed below and is illustrated in figure 2.2. In this figure there are ranges of X in which the sign of ΔX does not depend on ϵ, so that in these ranges X moves towards the equilibrium regions, which are the regions in which the roots of $g(., \epsilon)$ lie.

We cannot hope to identify g outside the support of X in the data. Identification of g is further complicated by the fact that positive statistical dependence of $\Delta X_{i,t+1}$ and $X_{i,t}$ can be due to either a positive causal relationship, or due to unobservable exogenous determinants of X which are positively related over time. Therefore, cross-sectional nonparametric quantile regressions of $\Delta X_{i,t+1}$ on $X_{i,t}$ will in general estimate slopes that are upward-biased relative to the slopes of g, as will be shown. As a consequence of such upward-bias, structural functions g that only exhibit one stable root might generate quantile regressions $\widehat{g}^q$ with more than one root, as will be discussed in section 2.3.5. However, if the functions $\widehat{g}^q$ have only one stable root, the dynamics of the system are stable, i.e., we can reject the hypothesis of multiple equilibria in the support of X.

Before we move on to formally discuss the inference procedure on the number of roots of g, let us give a heuristic preview of some of the main ideas. The procedure is based upon first stage estimation of g using local linear m-regression. Local linear m-regression replaces the conditional expectation in the moment condition identifying g with a kernel estimate thereof, and minimizes this kernel estimate in order to estimate g. Based on this estimate $\widehat{g}$ of g, the most intuitive estimator of the number of roots of g, $Z(g)$, would be the plug-in estimator $Z(\widehat{g})$. Instead of using this plug-in estimator, we propose an alternative that allows for inference that is easily implemented in practice and that dominates the plug-in estimator $Z(\widehat{g})$ in terms of asymptotic efficiency. We estimate $Z(g)$ by $Z_\rho(\widehat{g})$, where $Z_\rho(g)$ is a smooth functional which approximates the number of roots of g and is defined by

$$Z_\rho := \int_{\mathscr{X}} L_\rho\big(g(x)\big)\big|g'(x)\big|dx.$$

In this expression, L_ρ is a kernel function with bandwidth ρ which integrates to 1 and $\mathscr{X}$ is the support of X. The functional $Z_\rho(g)$ equals the integer valued $Z(g)$ for most functions g. It differs from $Z(g)$ in the neighborhood of functions $\tilde{g}$ where $Z(g)$ jumps. In these neighborhoods, $Z_\rho(g)$ varies continuously with g.

The functional $Z_\rho(g)$ has some peculiar features, which make the statistical theory of $Z_\rho(\widehat{g})$ mathematically interesting. In particular, since $Z_\rho(g)$ is equal to the discrete-valued functional $Z(g)$, and hence constant, in the neighborhood of any generic function $\tilde{g}$, any "delta-method"-type approximations of the distribution of $Z_\rho(\widehat{g})$ only yield a degenerate limiting distribution. Under standard i.i.d. asymptotics and regularity conditions, $\widehat{g}$ and $\widehat{g'}$ converge uniformly to (g, g'). As a consequence of this and the local constancy of $Z_\rho(g)$, $Z_\rho(\widehat{g})$ converges to $Z(g)$ at an infinite rate. Meaningful asymptotic approximations can therefore only be obtained using non-standard asymptotics. As is well known, estimates of g' converge at a slower rate than estimates of g. If variation in $\widehat{g}$ is negligible relative to variation in $\widehat{g'}$, variation in $Z_\rho(\widehat{g})$ (and $Z(\widehat{g})$) is driven by "wiggles" in $\widehat{g}$ in the neighborhood of the true roots of g.

The central mathematical result of this paper states that a rescaled version of $Z_\rho(\widehat{g})$ converges to a normal distribution under a non-standard sequence of experiments using increasing levels of noise and shrinking bandwidth as sample size increases. This sequence is chosen such that $\widehat{g}$ converges uniformly to g, while $\widehat{g'}$ converges to a non-degenerate limit. Under the same sequence of experiments, the bootstrap provides consistent estimates of the bias and standard deviation of $Z_\rho(\widehat{g})$ relative to $Z(g)$. Inference on $Z(g)$ can therefore be performed by using t-statistics which are based on $Z_\rho(\widehat{g})$, as well as bootstrapped standard errors and bias. Monte Carlo evidence largely conforms to the asymptotic results, although the inference procedure appears to be conservative in the range of experiments simulated.

The statistical theory developed in this paper is based on results from Kong, Linton, and Xia (2010) on Bahadur expansions for local polynomial m-regression, and uses somewhat similar arguments as Horváth (1991), who discusses the asymptotic distribution of L_p-norms of multivariate density estimators. The argument on bootstrap-based inference draws on the review by Horowitz (2001). The generalizations of the inference procedure discussed in section 2.3 use results on the asymptotic theory of partial means estimation developed in Newey (1994). The applications were motivated, in particular, by Bajari, Hong, Krainer, and Nekipelov (2006) and by Card, Mas, and Rothstein (2008).

The rest of this paper is structured as follows. Section 2.2 presents the inference procedure, as well as its asymptotic justification, for the baseline case. Section 2.3 discusses generalizations, as well as identification and inference in static games of incomplete information and in stochastic difference equations. Section 2.4 applies the inference procedure to the data on neighborhood composition studied by Card, Mas, and Rothstein (2008). In contrast to their results, no evidence of "tipping" (equilibrium multiplicity) is found here. Section 2.5 concludes. Appendix 2.A presents some Monte Carlo evidence. All proofs are relegated to appendix 2.B, all figures and tables can be found in appendix 2.C. Additional figures and tables are in the web appendix, Kasy (2010). This web appendix also contains a second application of the inference procedure to data on economic growth, similar to those discussed by Azariadis and Stachurski (2005), section 4.1, and by Quah (1996).

2.2 Inference in the baseline case

Throughout this paper, the paramter of interest is the number of roots Z of some function g on its support $\mathscr{X}$:

$$Z(g) := |\{x \in \mathscr{X} : g(x) = 0\}|. \tag{2.4}$$

The identification of this parameter follows from identification of g on $\mathscr{X}$. In this section, inference on $Z(g)$ is discussed for functions g with one dimensional and compact domain and range. Throughout, the following assumptions will be maintained. The observable data are i.i.d. draws of (Y_i, X_i). The density of X is bounded away from 0 on $\mathscr{X}$. The function g is identified by a conditional moment restriction of the form

$$g(x) = \text{argmin}_y E_{Y|X}[m(Y-y)|X=x]. \tag{2.5}$$

This in particular covers the cases $m(\delta) = \delta^2$ for conditional mean regression and $m_q(\delta) = \delta(q - \mathbf{1}(\delta < 0))$ for conditional qth quantile regression. Furthermore, g is assumed to be continuously differentiable and generic in the following sense.

Definition 9 (Genericity). *A continuously differentiable function g is called generic if*

$$\{x : g(x) = 0 \text{ and } g'(x) = 0\} = \varnothing$$

and if all roots of g are in the interior of $\mathscr{X}$.

Genericity of g implies that g has only a finite number of roots.

We propose the following inference procedure for the number of roots of g, $Z(g)$: First, estimate $g(.)$ and $g'(.)$ using local linear m-regression:

$$\left(\widehat{g}(x), \widehat{g'}(x)\right) = \text{argmin}_{a,b} \sum_i K_\tau(X_i - x) m(Y_i - a - b(X_i - x)), \tag{2.6}$$

where $K_\tau(\delta) = \frac{1}{\tau} K(\frac{\delta}{\tau})$ for some (symmetric, positive) kernel function K integrating to one with bandwidth τ. Equation (2.6) is a sample analog of equation (2.5), where a kernel weighted local average is replacing the conditional expectation. Next, estimate $Z(g)$ by $\widehat{Z} = Z_\rho\left(\widehat{g}(.), \widehat{g'}(.)\right)$, where Z_ρ is defined as

$$Z_\rho\left(g(.), g'(.)\right) := \int_{\mathscr{X}} L_\rho\big(g(x)\big)\big|g'(x)\big|dx. \tag{2.7}$$

In this expression, L_ρ is a Lipschitz continuous, positive symmetric kernel integrating to 1 with bandwidth ρ and support $[-\rho, \rho]$. Estimate the variance and bias of $\widehat{Z}$ relative to Z using bootstrap. Finally, construct integer valued confidence sets for Z using t-statistics based on $\widehat{Z}$ and the bootstrapped variance and bias.

The rest of this section will motivate and justify this procedure. First, we will see that $\widehat{Z}$ is a superconsistent estimator of Z, in the sense that $\alpha_n(\widehat{Z} - Z) \to^p 0$ for any diverging sequence $\alpha_n \to \infty$, under i.i.d. sampling and conditions to be stated. Then we will present the central result of this paper, which establishes asymptotic normality of $\widehat{Z}$ under a non-standard sequence of experiments. From this result it follows that inference based on t-statistics, using bootstrapped standard errors and bias corrections, provides asymptotically valid confidence sets for Z. The theorem also suggests that $\widehat{Z}$ is an efficient estimator relative to the simple plugin estimator $Z(\widehat{g})$.

The following proposition states that $Z(g) = Z_\rho(g)$ for generic g and ρ small enough. The two functionals only differ around non-generic g, or "bifurcation points," that is g where Z jumps. The functional Z_ρ is a smooth approximation of Z which varies continuously around such jumps.

Proposition 8. *For g continuously differentiable and generic, if $\rho > 0$ is small enough, then*

$$Z_\rho(g(.), g'(.)) = Z(g(.)).$$

The proof proceeds as follows: Given a generic function g, consider the subset of $\mathscr{X}$ where $L_\rho(g)$ is not zero. If ρ is small enough, this subset is partitioned into disjoint neighborhoods of the roots of g, and g is monotonic in each of these neighborhoods. A change of variables, setting $y = g(x)$, shows that the integral over each of these neighborhoods equals one. Figure 2.3 illustrates the relationship between Z and Z_ρ. The two functionals are equal, if g does not peak within the range $[-\rho, \rho]$. If g does peak within the range $[-\rho, \rho]$, they are different and Z_ρ is not integer valued.

It is useful to equip the space of continuously differentiable functions on $\mathscr{X}$ with the following norm:

Definition 10 ($\mathscr{C}^1$ norm)**.** *Let $\mathscr{C}^1(\mathscr{X})$ denote the space of continuously differentiable functions on the compact domain $\mathscr{X}$. The norm $||.||$ on $\mathscr{C}^1(\mathscr{X})$ is defined by*

$$||g|| := \sup_{x \in \mathscr{X}} |g(x)| + \sup_{x \in \mathscr{X}} |g'(x)|.$$

Given this norm, we have the following proposition:

Proposition 9 (Local constancy)**.** *$Z(.)$ is constant in a neighborhood, with respect to the norm $||.||$, of any generic function $g \in C^1$, and so is Z_ρ if ρ is small enough.*

Using a neighborhood of g with respect to the sup norm in levels only, instead of $||.||$, is not enough for the assertion of proposition 9 to hold. For any function g_1 that has at least one root, we can find a function g_2 arbitrarily close to g_1 in the uniform sense, which has more roots than g_1, by adding a "wiggle" around a root of g_1. Figure 2.4 illustrates by showing two functions which are uniformly close in levels but not in derivatives, and which have different numbers of roots. If one, however, additionally restricts the first derivative of g_2 to be uniformly close to the the derivative of g_1, additional wiggles are precluded around generic roots, since around these g_1 has a non-zero derivative. Since derivatives are "harder" to estimate than levels, variation in the estimated derivatives dominates the asymptotic distribution of estimators for $Z(g)$, as will be shown. Proposition 9 implies the following corollary, which states that the plugin estimator $\widehat{Z} = Z_\rho\left(\widehat{g}(.), \widehat{g'}(.)\right)$ converges to a degenerate limiting distribution at an "infinite" rate, if $\widehat{g}$ converges with respect to the norm $||.||$.

Corollary 5 (Superconsistency)**.** *If $\left(\widehat{g}, \widehat{g'}\right)$ converges uniformly in probability to (g, g'), if g is generic and if $\alpha_n \to \infty$ is some arbitrary diverging sequence, then*

$$\alpha_n(Z(\widehat{g}) - Z(g)) \to^p 0.$$

Furthermore, if ρ is small enough so that $Z_\rho(g, g') = Z(g)$ holds, then

$$\alpha_n \left(Z_\rho\left(\widehat{g}, \widehat{g'}\right) - Z(g)\right) \to^p 0.$$

This corollary implies that $\alpha_n \left(Z_\rho \left(\widehat{g}, \widehat{g'} \right) - Z(g) \right) \to^p 0$ if $\rho \to 0$ as $n \to \infty$.

To further characterize the asymptotic distribution of $\widehat{Z}$, we need a suitable approximation for the distribution of the first stage estimator $\left(\widehat{g}(.), \widehat{g'}(.) \right)$. Kong, Linton, and Xia (2010) provide uniform Bahadur representations for local polynomial estimators of m-regressions. We state their result, for the special case of local linear m-regression, as an assumption.

Assumption 2.2.1 (Bahadur expansion)**.** *The estimation error of the estimator* $\left(\widehat{g}(x), \widehat{g'}(x) \right)$ *defined by equation* (2.6) *can be approximated by a local average as follows:*

$$\left(\widehat{g}(x), \widehat{g'}(x) \right) - (g(x), g'(x)) = R-$$

$$-f_x^{-1}(x) s^{-1}(x) I_n(x) \frac{1}{n} \sum_i K_\tau(X_i - x) \phi(Y_i - g(x) - g'(x)(X_i - x)) \left(\frac{1}{\tau}, \frac{X_i - x}{\nu_2 \tau^3} \right), \quad (2.8)$$

where f_x *is the density of* x, $\nu_2 := \int K(x) x^2 dx$, $\phi := m'$ *(in a piecewise derivative sense),* $s(x) = \frac{\partial}{\partial g(x)} E[\phi(Y - g(x)) | X = x]$, *and* $I_n(x)$ *is a non-random matrix converging uniformly to the identity matrix, and where*

$$R = o_p \left(\left(\widehat{g}(x), \widehat{g'}(x) \right) - (g(x), g'(x)) \right)$$

uniformly in x.

Kong, Linton, and Xia (2010) provide regularity conditions under which

$$R = \left(1, \frac{1}{\tau} \right) O_p \left(\frac{log(n)}{n\tau} \right)^\lambda$$

uniformly in X, for some $\lambda \in (0, 1)$ as $n \to \infty$. In the case of qth quantile regression, $\phi(\delta) = q - \mathbf{1}(\delta < 0)$ and $s(x) = -f_{y|x}(g(x)|x)$. In the case of mean regression, $\phi(\delta) = -2\delta$ and $s(x) = -2$.

The asymptotic results in the remainder of this section depend on the availability of an expansion in the form of expansion (2.8) and the relative negligibility of the remainder, but not on any other specifics of local linear m-regression. This will allow for fairly straightforward generalizations of the baseline case considered here to the cases discussed in section 2.3 as well as to other cases which are beyond the scope of this paper, once we have appropriate expansions for the first stage estimators.

By proposition 9, consistency of any plugin estimator follows from uniform convergence of $\left(\widehat{g}(.), \widehat{g'}(.)\right)$. Such uniform convergence follows from assumption 2.2.1, combined with a Glivenko Cantelli-theorem on uniform convergence of averages, assuming i.i.d. draws from the joint distribution of (Y, X) as $n \to \infty$, see van der Vaart (1998), chapter 19. Superconsistency of $\widehat{Z}$ therefore follows, which implies that standard i.i.d. asymptotics with rescaling of the estimator yield only degenerate distributional approximations. This is because Z_ρ and Z are constant in a C^1 neighborhood of any generic g, even though they jump at "bifurcation points", i.e., non-generic g. As a consequence, all terms in a functional Taylor expansion of Z_ρ, as a function of g, vanish, except for the remainder. The application of "delta method" type arguments, as in Newey (1994), gives only the degenerate limit distribution.

In finite samples, however, the sampling variation of $\widehat{Z}$ is in general not negligible, as the simulations of appendix A confirm, which makes the distributional approximation of the degenerate limit useless for inference. Asymptotic statistical theory approximates the finite sample distribution of interest by a limiting distribution of a sequence of experiments, of which our actual experiment is an element. The choice of sequence, such as i.i.d. sampling, is to some extent arbitrary. In econometrics, non-standard asymptotics are used for instance in the literature on weak instruments (e.g., Imbens and Wooldridge (2007)). In the present setup, a non-degenerate distributional limit of $\widehat{Z}$ can only be obtained under a sequence of experiments which yields a non-degenerate limiting distribution of the first stage estimator $\left(\widehat{g}(.), \widehat{g'}(.)\right)$. We will now consider asymptotics under such a sequence of experiments. The sequence we consider has increasing amounts of "noise" relative to "signal" as sample size increases. Assume that for the nth experiment we observe $(Y_{i,n}, X_{i,n})$ for $i = 1, \ldots, n$, and assume

$$
\begin{aligned}
X_{i,n} &\sim^{iid} f_x(.) && (2.9)\\
\gamma_{i,n}|X_{i,n} &\sim f_{\gamma|X} && (2.10)\\
Y_{i,n} &= g(X_{i,n}) + r_n \gamma_{i,n}, && (2.11)
\end{aligned}
$$

where $\{r_n\}$ is a real-valued sequence and $0 = \operatorname{argmin}_a E[m(\gamma - a)|X] = \operatorname{argmin}_a E[m(r_n\gamma - a)|X]$. The last equality requires the criterion function m to be "scale neutral". For a given sample size n, this is the same model as before. As n changes, the function g identified by equation (2.5) is held constant. If r_n grows in n, the estimation problem in this sequence of models becomes increasingly difficult relative to i.i.d. sampling. Note that equation (2.11) does *not* describe an additive structural model, which would allow to predict counterfactual outcomes. Instead, $r_n \gamma_{i,n}$ is simply the statistical residual, given by the difference of Y and $g(X)$, which is also well-defined for non-additive structural models.

By corollary 5, a necessary condition for a non-degenerate limit of $\widehat{Z}$ is that $\left(\widehat{g}, \widehat{g'}\right)$ converges to a non-degenerate limiting distribution. As is well known, and also follows from assumption 2.2.1, $\widehat{g'}$ converges at a slower rate than $\widehat{g}$, so that asymptotically variation in $\widehat{g'}$ will dominate, namely by adding "wiggles" around the actual roots. If $r_n = (nh^5)^{1/2}$ in the sequence of experiments just defined, $\widehat{g}$ converges uniformly in probability to g, whereas $\widehat{g'}$ converges point-wise (and indeed functionally) to a non degenerate limit. This is the basis for the following theorem.

Theorem 2 (Asymptotic normality)**.** *Under the above model assumptions and assumption 2.2.1, and if $r_n = (n\tau^5)^{1/2}$, $n\tau \to \infty$, $\rho \to 0$ and $\tau/\rho^2 \to 0$, then there exist $\mu > 0$ and V such that*

$$\sqrt{\frac{\rho}{\tau}}\left(\widehat{Z} - \mu - Z\right) \to N(0, V)$$

for $\widehat{Z} = Z_\rho\left(\widehat{g}, \widehat{g'}\right)$. Both μ and V depend on the data generating process only via the asymptotic mean and variance of $\widehat{g'}$ at the roots of g, which in turn depend upon f_X, g', s and $Var(\phi|X)$ evaluated at the roots of g.

This result justifies the use of t-tests based on $\widehat{Z}$ for null hypotheses of the form $Z(g) = Z_\rho(g) = z_0$. The construction of a t-statistic requires a consistent estimator of V and an estimator of μ converging at a rate faster than $\sqrt{\rho/\tau}$. The last part of theorem 2 suggests a way to obtain those. Any plug-in estimator that consistently estimates the (co)variances of $\widehat{g'}$ under the given sequence of experiments consistently estimates μ and V. One such plug-in estimator is standard bootstrap, that is resampling from the empirical distribution function. The Bahadur expansion in assumption 2.2.1, which approximates $\widehat{g'}$ by sample averages, implies that the bootstrap gives a resampling distribution with the asymptotically correct covariance structure for $\widehat{g'}$. From this and theorem 2 it then follows that the bootstrap gives consistent variance and bias estimates for Z_ρ, where the bias is estimated from the difference of the resampling estimates relative to $Z_\rho(\widehat{g})$. If sample size grows fast enough relative to $\sqrt{\rho/\tau}$ and τ, the asymptotic validity of a standard normal approximation for the pivot follows.

In principle, higher order bootstrapping could now be applied to obtain distributional refinements for this statistic, as discussed in detail by Horowitz (2001). However, higher order bootstrapping might be very computationally demanding in the present case, in particular if criteria like quantile regression are used to identify g.

Theorem 2 also implies that increasing the bandwidth parameter ρ reduces the variance without affecting the bias in the limiting normal distribution. Asymptotically, the difficulty in estimating Z is driven entirely by fluctuations in $\widehat{g'}$. These fluctuations lead both to upward bias and to variance in plug-in estimators. When ρ is larger, these fluctuations are averaged over a larger range of X, thereby reducing variance. Theorem 2 implies that Z_{ρ_1} is asymptotically inefficient relative to Z_{ρ_2} for $\rho_1 < \rho_2$. Furthermore, by proposition 8, $Z(g) = \lim_{\rho\to 0} Z_\rho(g)$ for all generic g. If the relative inefficiency carries over to the limit as

$\rho \to 0$, it follows that the simple plug-in estimator $Z(\widehat{g})$ is asymptotically inefficient relative to $\widehat{Z}$. Note, however, that this is only a heuristic argument. We can not exchange the limits with respect to ρ and with respect to n to obtain the limit distribution of $Z(\widehat{g})$.

2.3 Extensions and applications

In this section, several extensions and applications of the results of section 2.2 are presented. Subsections 2.3.1 through 2.3.3 discuss, in turn, inference on Z if g is identified by more general moment conditions, inference on Z if the domain and range of g are multidimensional, and inference on the number of stable and unstable roots. Subsections 2.3.4 and 2.3.5 discuss identification and inference for the two applications mentioned in the introduction, static games of incomplete information and stochastic difference equations.

2.3.1 Conditioning on covariates

In the previous section, inference on $Z(g)$ was discussed for functions g identified by the moment condition

$$g(x) = \text{argmin}_y E_{Y|X}[m(Y-y)|X=x].$$

This subsection generalizes to functions g identified by

$$g(x, w_1) = \text{argmin}_y E_{W_2}\left[E_{Y|X,W}[m(Y-y)|X=x, W_1=w_1, W_2]\right], \tag{2.12}$$

where the parameter of interest now is $Z(g(., w_1))$, the number of roots of g in x given w_1. The conditional moment restriction (2.12) can be rationalized by a structural model of the form $Y = h(X, W_1, \epsilon)$, where $\epsilon \perp (X, W_1)|W_2$ and g is defined by

$$g(x, w_1) := \text{argmin}_y E_\epsilon\left[m(h(x, w_1, \epsilon) - y)\right].$$

We will assume that the joint density of X, W is bounded away from zero on the set $supp(X, W_1) \times supp(W_2)$, where $supp$ denotes the compact support of either random vector.

The vector W_2 serves as a vector of control variables. The conditional independence assumption $\epsilon \perp (X, W_1)|W_2$ is also known as "selection on observables." The function g is equal to the average structural function if $m(\delta) = \delta^2$, and equal to a quantile structural function if $m_q(\delta) = \delta(q - \mathbf{1}(\delta < 0))$. The average structural function will be of importance in the context of games of incomplete information, as discussed in section 2.3.4, quantile structural functions will be used to characterize stochastic difference equations in section 2.3.5. When games of incomplete information are discussed in section 2.3.4, $W = W_1$ will correspond to the component of public information which is not excluded from either player's response function.

The inference procedure proposed in the previous section is based upon two steps. First, the function g and its derivative are estimated using local linear m-regression. In the second step, the estimator $\left(\widehat{g}, \widehat{g'}\right)$ is plugged into the functional $Z_\rho(.,.)$, which is a smooth approximation of the functional $Z(.)$. We can generalize this approach by maintaining the same second step while using more general first stage estimators $\left(\widehat{g}, \widehat{g'}\right)$. Equation (2.12) suggests estimating g by a nonparametric sample analog, replacing the conditional expectation with a local linear kernel estimator of it, and the expectation over W_2 with a sample average. Formally, let

$$\left(\widehat{g}(x, w_1), \widehat{g'}(x, w_1)\right) = \mathrm{argmin}_{a,b} M(a, b, x, w_1),$$

where

$$M(a, b, x, w_1) = \frac{1}{n} \sum_j \frac{\sum_i K_\tau(X_i - x, W_{1i} - w_1, W_{2i} - W_{2j}) m(Y_i - a - b(X_i - x))}{\sum_i K_\tau(X_i - x, W_{1i} - w_1, W_{2i} - W_{2j})}. \quad (2.13)$$

An asymptotic normality result can be shown in this context which generalizes theorem 2. In light of the proof of theorem 2, the crucial step is to obtain a sequence of experiments such that $\widehat{g}$ converges uniformly to g while $\widehat{g'}$ has a non degenerate limiting distribution. If we obtain an approximation of $\widehat{g'}$ equivalent to the approximation in assumption 2.2.1, all further steps of the proof apply immediately. This can be done, using the results of Newey (1994), for the following sequence of experiments.

$$\begin{aligned}
(X_{i,n}, W_{i,n}) &\sim^{iid} f_{x,w}(.) &(2.14)\\
\gamma_{i,n} | (X_{i,n}, W_{i,n}) &\sim f_{\gamma|X,W} &(2.15)\\
Y_{i,n} &= g(X_{i,n}, W_{1i,n}) + r_n \gamma_{i,n}. &(2.16)
\end{aligned}$$

Theorem 3 (Asymptotic normality, with control variables)**.** *Under the assumptions of section 2.2, but with g identified by equation 2.12 and the data generated by the model given by equation 2.14 through 2.16, if $R = o_p\left(\left(\widehat{g}, \widehat{g'}\right) - (g, g')\right)$ uniformly as $n \to \infty$ in the Bahadur expansion of $\left(\widehat{g}, \widehat{g'}\right)$, and if $r_n = \left(n\tau^{4+d}\right)^{1/2}$, where $d = dim(X) + dim(W_1)$, $n\tau^d \to \infty$, $\rho \to 0$ and $\tau/\rho^2 \to 0$, then there exist $\mu > 0$ and V such that*

$$\sqrt{\frac{\rho}{\tau}} \left(\widehat{Z} - \mu - Z\right) \to N(0, V).$$

2.3.2 Higher dimensional systems

Thus far, only one-dimensional arguments x and one-dimensional ranges for the function g were considered, where x is the argument over which Z_ρ integrates. All results of section 2.2

are easily extended to a higher dimensional setup. In particular, assume we are interested in the number of roots of a function g from $\mathbb{R}^d$ to $\mathbb{R}^d$. Generalizing equation (2.7), we can define $\widehat{Z}$ as

$$\widehat{Z} := \int L_\rho(\widehat{g}) \left| \det \widehat{g'} \right|, \tag{2.17}$$

where $\left(\widehat{g}(.), \widehat{g'}(.)\right)$ are again estimated by local linear m regression, L_ρ is a kernel with support $[-\rho, \rho]^d$, and the integral is taken over the set $\mathscr{X} \subset \mathbb{R}^d$ in the support of g.
As in the one dimensional case, superconsistency follows from uniform convergence of $(\widehat{g}, \widehat{g'})$.
The following theorem, generalizing theorem 2, holds for arbitrary d:

Theorem 4 (Asymptotic normality, multidimensional systems)**.** *Under the assumptions of section 2.2, but with $g : \mathbb{R}^d \to \mathbb{R}^d$ and $\widehat{Z}$ defined by equation (2.17), if $R = o_p\left(\left(\widehat{g}, \widehat{g'}\right) - (g, g')\right)$ uniformly as $n \to \infty$ in the Bahadur expansion of $\left(\widehat{g}, \widehat{g'}\right)$, and if $r_n = (n\tau^{4+d})^{1/2}$, $n\tau^d \to \infty$, $\rho \to 0$ and $\tau/\rho^{d+1} \to 0$, then there exist $\mu > 0, V$ such that*

$$\left(\frac{\rho}{\tau}\right)^{d/2} \left(\widehat{Z} - \mu - Z\right) \to N(0, V).$$

2.3.3 Stable and unstable roots

Instead of testing for the total number of roots, one might be interested in the number of "stable" and "unstable" roots, Z^s and Z^u. Stable roots are those where g' is negative, unstable roots those where g' is positive.

Definition 11. *For g continuously differentiable, let*

$$Z^s(g) := |\{x \in \mathscr{X} : g(x) = 0 \text{ and } g'(x) < 0\}|$$

and

$$Z^u(g) := |\{x \in \mathscr{X} : g(x) = 0 \text{ and } g'(x) > 0\}|.$$

In the multidimensional case, we could more generally consider roots with a given number of positive and negative eigenvalues of g'. We can define smooth approximations of the parameters Z^s and Z^u as follows:

$$\begin{aligned} Z^s_\rho(g(.), g'(.)) &:= \int_{\mathscr{X}} L_\rho\big(g(x)\big) \left|g'(x)\right| \mathbf{1}\big(g'(x) < 0\big) dx \\ Z^u_\rho(g(.), g'(.)) &:= \int_{\mathscr{X}} L_\rho\big(g(x)\big) \left|g'(x)\right| \mathbf{1}\big(g'(x) > 0\big) du. \end{aligned} \tag{2.18}$$

Again, all arguments of section 2.2 go through essentially unchanged for these parameters. In particular, theorem 2 applies literally, replacing Z with Z^s or Z^u.

More generally, functionals which are smooth approximations of the number of roots with various stability properties can be constructed in the multidimensional case by multiplying the integrand with an indicator function depending on the signs of the eigenvalues of $\widehat{g'}$.

2.3.4 Static games of incomplete information

Consider the two player, two action static game of incomplete information discussed in the introduction. In this subsection, identification and inference on the number of Bayesian Nash Equilibria of this game, given the public information s, will be discussed. Assume we observe an i.i.d. sample of $(a_{1,j}, a_{2,j}, s_j)$, the players' realized actions and the public information of the game, where $a_{i,j} \in \{0,1\}$ for $i = 1,2$ and $s \in \mathbb{R}^k$. In this subsection, i indexes players and j indexes observations. Rational expectation beliefs of player $-i$ about the expected action of player i are given by $\sigma_i(s) = E[a_i|s]$. The following two-stage estimation procedure is a nonparametric variant of the procedure proposed by Bajari, Hong, Krainer, and Nekipelov (2006). We can get an estimate of the beliefs, $\widehat{\sigma}_i(s) = \widehat{E}[a_i|s]$, by local linear mean regression.

$$(\widehat{\sigma}_i(s), \widehat{\sigma}'_i(s)) = \operatorname{argmin}_{b,c} \sum_j K_\tau(s_j - s)\,(a_{i,j} - b - c(s_j - s))^2 \tag{2.19}$$

Average best responses of players are given by $g_i(\sigma_{-i}, s) = E[a_i|\sigma_{-i}, s]$, since we assumed that the private information of players is independent, conditional on s. Without further restrictions, g_i is not identified, since by definition σ is functionally dependent on s. If, however, exclusion restrictions of the form

$$g_i(\sigma_{-i}, s) = g_i(\sigma_{-i}, s_i) \tag{2.20}$$

are imposed, the g_i can be identified. In particular, assume that exclusion restriction (2.20) holds, with $\dim(s_i) = \dim(s) - 1 = k - 1$. There is one excluded component of s for each player, the remaining $k-2$ components are not excluded from either response function g_i. Assume furthermore that $\sigma_i(s)$ has full support $[0,1]$ given s_{-i}, for $i = 1,2$. Under these assumptions, we can estimate the best response functions, $\widehat{g}_i(\bar{\sigma}_{-i}, s_i) = \widehat{E}[a_i|\widehat{\sigma}_{-i} = \bar{\sigma}_{-i}, s_i]$, again using local linear mean regression:

$$\left(\widehat{g}_i(\bar{\sigma}_{-i}, s_i), \widehat{g'}_i(\bar{\sigma}_{-i}, s_i)\right) =$$

$$\operatorname{argmin}_{b,c} \sum_j K_\tau(\widehat{\sigma}_{-i,j} - \bar{\sigma}_{-i}, s_{i,j} - s_i)\,(a_{i,j} - b - c\,(\widehat{\sigma}_{-i,j} - \bar{\sigma}_{-i}, s_{i,j} - s_i))^2 \tag{2.21}$$

The condition for Bayesian Nash Equilibrium in this game is given by

$$g(\bar{\sigma}_1, s) = g_1(g_2(\bar{\sigma}_1, s_2), s_1) - \bar{\sigma}_1. \tag{2.22}$$

Inserting $\widehat{g}_2$ into $\widehat{g}_1$, both estimated by (2.21), yields an estimator of g which can be written as

$$\widehat{g}(\bar{\sigma}_1, s) = \widehat{E}\left[a_1 \middle| \widehat{\sigma}_2 = \widehat{E}\left[a_2 | \widehat{\sigma}_1 = \bar{\sigma}_1, s_2\right], s_1\right] - \bar{\sigma}_1. \tag{2.23}$$

Based on this estimator, we can perform inference on the number of Bayesian Nash Equilibria given s, $Z(g(., s))$. In particular, let

$$\widehat{Z} = Z_\rho\left(\widehat{g}(., s), \widehat{g'^1}(., s)\right), \tag{2.24}$$

where $\widehat{g}(., s)$ is given by (2.23). The term $\widehat{g'^1}(., s)$ refers to the estimated derivative of g w.r.t. $\bar{\sigma}_1$, and similarly for $\widehat{g_1'^1}$ and $\widehat{g_2'^1}$, so that

$$\widehat{g'^1}(\bar{\sigma}_1, s) = \widehat{g_1'^1}(\widehat{g}_2(\bar{\sigma}_1, s_2), s_1) \cdot \widehat{g_2'^1}(\bar{\sigma}_1, s_1). \tag{2.25}$$

Inference on $Z(g(., s))$ can now proceed as before, if an asymptotic normality result similar to theorem 2 can be shown. In the proof of theorem 2, three properties of $\left((\widehat{g}(.), \widehat{g'}(.)\right)$ needed to be proven for the statement of the theorem to follow: First, under the given sequence of experiments, $\widehat{g}(.)$ converges uniformly in probability to a degenerate limit. Second, $\widehat{g'}(.)$ converges in distribution to a non-degenerate limit. Third, $\widehat{g'}(x_1)$ and $\widehat{g'}(x_2)$ are asymptotically independent for $|x_1 - x_2| > const \cdot \tau$. These properties can be shown for $r_n \cdot \left((\widehat{g}(., s), \widehat{g'^1}(., s)\right)$ in the present case, with $\bar{\sigma}_1$ replacing x, for an appropriate choice of sequence of experiments, where r_n is a scale parameter as before.

The choice of sequence of experiments may seem to be more complicated here than in the baseline case, since the dependent variable a is naturally bounded by $[0, 1]$, so that increasing the residual variance would be inconsistent with the structural model. This is not a problem, however, if we note that the distribution of $\widehat{Z}$, in the baseline model, is invariant to a proportional rescaling of Y, g and ρ. We can therefore define a sequence of experiments which is equivalent to the one defined by equations (2.9) through (2.11) if we replace equation (2.11) by

$$Y_{i,n} = \frac{1}{r_n} g(X_{i,n}) + \gamma_{i,n} \tag{2.11'}$$

and ρ by ρ/r_n. Intuitively, shrinking the "signal" g is equivalent to increasing the "noise" $r_n \gamma_{i,n}$.

Returning to games of incomplete information, consider the following sequence of experiments indexed by n. Assume that, for $i = 1, 2$, $g_{i,0}$ is continuously differentiable and monotonic in σ_{-i}, and let $g_{i,n}^{-1}$ denote the inverse of $g_{i,n}$ with respect to the $\sigma_{i,n}$ argument, given s_i. Assume also

$$\begin{aligned}
s_{j,n} &\sim^{iid} f_s(.) && (2.26)\\
a_{i,j,n}|s_{j,n} &\sim \mathrm{Bin}(\sigma_{i,n}(s_{j,n})) && (2.27)\\
\sigma_{i,n}(s) &= g_{i,n}(\sigma_{-i,n}(s), s_i) && (2.28)\\
g_{1,n}(\sigma_2, s_1) &= \frac{1}{r_n} g_{1,0}(\sigma_2, s_1) + \left(1 - \frac{1}{r_n}\right)\sigma_2 && (2.29)\\
g_{2,n}^{-1}(\sigma_2, s_2) &= \frac{1}{r_n} g_{2,0}^{-1}(\sigma_2, s_2) + \left(1 - \frac{1}{r_n}\right)\sigma_2. && (2.30)
\end{aligned}$$

Equations (2.26) to (2.28) are the same as in the model we have been discussing so far. Equations (2.29) and (2.30) shrink the graphs of the best response functions $g_i(., s_i)$ towards the $\sigma_1 = \sigma_2$ line (compare figure 2.1), parallel to the σ_1 axis. Denote $\sigma_{2,n} = g_{2,n}(\sigma_1, s_2)$. We get

$$\begin{aligned}
g_n(\sigma_1, s) = g_{1,n}(g_{2,n}(\sigma_1, s_2), s_1) - \sigma_1 &= g_{1,n}(\sigma_{2,n}, s_1) - g_{2,n}^{-1}(\sigma_{2,n}, s_2)\\
&= \frac{1}{r_n}\left[g_{1,0}(\sigma_{2,n}, s_1) - g_{2,0}^{-1}(\sigma_{2,n}, s_2)\right].
\end{aligned}$$

By equation (2.30), if $r_n \to \infty$, then $\sigma_{2,n} \to \sigma_1$, and hence

$$r_n g_n(\sigma_1, s) \to g_{1,0}(\sigma_1, s_1) - g_{2,0}^{-1}(\sigma_1, s_2). \qquad (2.31)$$

Using this sequence of experiments, we can now state an asymptotic normality result, similar to theorem 2, for static games of incomplete information. The statement of the theorem differs in two respects from the baseline case. First, ρ is replaced by $r_n\rho$ in all expressions. Since this sequence of experiments shrinks g rather than expanding the error, the bandwidth ρ must also shrink correspondingly. Second, the rate of growth of r_n is smaller. Since all regressions are controlling for s_1 or s_2, rates of convergence are slower. In particular, $r_n \cdot \widehat{g_i'^1}$ converges to a non-degenerate limit iff $r_n = O((n\tau^{4+k})^{1/2})$, where k is the dimensionality of the support of the response functions g_i, $k = \dim(s)$.

Theorem 5 (Asymptotic normality, static games of incomplete information)**.** *Under the sequence of experiments defined by equation (2.26) to (2.30), if $R = o_p\left(\left(\widehat{g}, \widehat{g'}\right) - (g, g')\right)$ uniformly in the Bahadur expansions as $n \to \infty$, and if $r_n = (n\tau^{4+k})^{1/2}$, $n\tau \to \infty$, $r_n\rho \to 0$ and $\tau/(r_n\rho)^2 \to 0$, then there exist $\mu > 0$ and V such that*

$$\sqrt{\frac{r_n\rho}{\tau}}\left(\widehat{Z} - \mu - Z\right) \to N(0, V).$$

2.3.5 Stochastic difference equations

In this subsection, identification and interpretation of the number of roots of g for stochastic difference equations of the form

$$\Delta X_{i,t+1} = X_{i,t+1} - X_{i,t} = g(X_{i,t}, \epsilon_{i,t}) \tag{2.32}$$

is discussed. This discussion will form the basis of the empirical application in section 2.4. First, it will be shown that, under plausible assumptions, finding only one root in cross-sectional quantile regressions of ΔX on X implies that there is only one stable root for every member of a family of conditional average structural functions. Second, it will be argued that the number of roots of g allows to characterize of the qualitative dynamics of the stochastic difference equation in terms of equilibrium regions.

The first claim is based on the fact that unobserved heterogeneity which is positively related over time leads to an upward bias in quantile regression slopes relative to the corresponding structural slopes. To show this, denote the qth conditional quantile of ΔX given X by $Q_{\Delta X|X}(q|X)$, the conditional cumulative distribution function at Q by $F_{\Delta X|X}(Q|X)$, and the conditional probability density by $f_{\Delta X|X}(Q|X)$. The following lemma shows that quantile regressions of ΔX on X yield biased slopes relative to the structural slope $\frac{\partial}{\partial X}g$, if X is not exogenous. The second term in equation 2.33 reflects the bias due to statistical dependence between X and ϵ.

Lemma 20 (Bias in quantile regression slopes)**.** *If $\Delta X = g(X, \epsilon)$, and if Q and F are differentiable with respect to the conditioning argument X, then*

$$\begin{aligned} \frac{\partial}{\partial X} Q_{\Delta X|X}(\tau|X) &= E\left[\frac{\partial}{\partial X} g(X,\epsilon) \middle| \Delta X = Q, X \right] \\ &- \frac{1}{f_{\Delta X|X}(Q|X)} \cdot \frac{\partial}{\partial X} \mathbb{P}\left(g(X',\epsilon) \le Q|X\right)\Big|_{X'=X}. \end{aligned} \tag{2.33}$$

The following assumption of *first order stochastic dominance* states that there is no negative dependence between current $g(x', \epsilon)$, evaluated at fixed x', and current X:

Assumption 2.3.1 (First order stochastic dominance)**.** $\mathbb{P}(g(x', \epsilon) \le Q|X)$ *is non-increasing as a function of X, holding x' constant.*

Violation of this assumption would require some underlying cyclical dynamics, in continuous time, with a frequency close enough to half the frequency of observation, or more generally with a ratio of frequencies that is an odd number divided by two. It seems safe to discard this possibility in most applications. This assumption might not hold, for instance, if outcomes were influenced by seasonal factors and observations were semi-annual.

We can now formally state the claim that, if there are unstable equilibria structurally, then quantile regressions should exhibit multiple roots.

Proposition 10 (Unstable equilibria in dynamics and quantile regressions)**.** *Assume that $\Delta X = g(X, \epsilon)$ and that $g(\inf \mathscr{X}, \epsilon) > 0$, and $g(\sup \mathscr{X}, \epsilon) < 0$ for all ϵ. If assumption 2.3.1 holds and $Q_{\Delta X|X}(q|X)$ has only one root X for all q, then the conditional average structural functions $E\left[g(x', \epsilon)|g(X, \epsilon) = 0, X\right]$, as functions of x', are "stable" at the roots m:*

$$E\left[\frac{\partial}{\partial X} g(X, \epsilon)\middle| \Delta X = 0, X\right] \leq 0$$

for all X, where $(0, X)$ is in the support of $(\Delta X, X)$.

This proposition assumes "global stability" of g, i.e., X does not diverge to infinity. Under such global stability, if there is only one root of g, then this root is stable. According to this proposition, if quantile regressions only have one stable root, then the same is true for the conditional average structural functions. This is not conclusive, but it is suggestive that the $g(., \epsilon)$ themselves have only one root.

Let us now turn to the implications of the number of roots of g for the qualitative dynamics of the stochastic difference equation (2.32). Let $\tilde{g}(x, \epsilon) := g(x, \epsilon) + x$. If g describes a structural relationship, the counterfactual time path under "manipulated" initial condition $X_{i,0} = x'$ is given by

$$\begin{aligned} X_{i,1} &= \tilde{g}(x', \epsilon_{i,0}) \\ X_{i,2} &= \tilde{g}(X_{i,1}, \epsilon_{i,1}) \\ &\vdots \\ X_{i,t} &= \tilde{g}(X_{i,t-1}, \epsilon_{i,t-1}). \end{aligned} \tag{2.34}$$

Given the initial condition $X_{i,1}$ and shocks $\epsilon_{i,1}, \ldots, \epsilon_{i,t}$, equation (2.32) describes a time inhomogenous deterministic difference equation. The following argument makes statements about the qualitative behavior of this difference equation based on properties of the function g, in particular based on the number of roots in x of $g(x, \epsilon)$ for given unobservables $\epsilon_{i,1}, \ldots, \epsilon_{i,t}$. Consider figure 2.2, which shows g^U and g^L defined by

$$g^U_{i,t}(x) = \max_{0 \leq s < t} g(x, \epsilon_{i,s}) \tag{2.35}$$

$$g^L_{i,t}(x) = \min_{0 \leq s < t} g(x, \epsilon_{i,s}). \tag{2.36}$$

The functions $g^U_{i,t}$ and $g^L_{i,t}$ are the upper and lower envelope of the family of functions $g(x, \epsilon_{i,s})$ for $s = 1, \ldots, t$. The direction of movement of X over time does not depend on s in the ranges where $g^U_{i,t} < 0$ or $g^L_{i,t} > 0$ (which is where the horizontal axis is drawn solid in figure 2.2), since the sign of $g(x, \epsilon_{i,s})$ does not depend on s in these ranges. In other words, suppose we start off with an initial value below x_1 in the picture. If that is the case, $X_{i,s}$ will converge monotonically toward the left-hand dashed range and then remain within that range for all $s \leq t$. Similarly, for $X_{i,0}$ in the upper "basin of attraction" beyond x_2, $X_{i,s}$ will converge to the upper "equilibrium range" given by the right hand dashed range. Hence small changes of initial conditions (from x_1 to x_2) can have large and persistent effects on X in this case, in contrast to the case where $g(., \epsilon)$ only has only one stable root for all ϵ. These arguments are summarized in the following proposition.

Proposition 11 (Characterizing dynamics of stochastic difference equations)**.** *Assume that $g^U_{i,t}$ and $g^L_{i,t}$, defined by equation (2.35) and (2.36), are smooth and generic, positive for sufficiently small x and negative for sufficiently large x, and have the same number z of roots, $x^U_1 < \ldots < x^U_z$ and $x^L_1 < \ldots < x^L_z$, and let $x^L_0 = -\infty$, $x^U_{z+1} = \infty$. Define the following mutually disjoint ranges:*

$$
\begin{aligned}
N_c &= [x^U_c, x^U_{c+1}] \; for \; c = 1, 3, \ldots, z \\
P_c &= [x^L_c, x^L_{c+1}] \; for \; c = 0, 2, \ldots, z-1 \\
S_c &= [x^L_c, x^U_c] \; for \; c = 1, 3, \ldots, z \\
U_c &= [x^U_c, x^L_c] \; for \; c = 2, 4, \ldots, z-1
\end{aligned}
$$

Then all $g(x, \epsilon_{i,s})$ are negative on the N_c, and positive on the P_c. Furthermore, all $g(x, \epsilon_{i,s})$ are negative in a neighborhood to the right of the maximum of the S_c and positive to the left of the minimum, and the reverse holds for the U_c. Therefore, if $X_{i,0} \in N_c$ and $S_c \neq \varnothing$, then $X_{i,s}$ will converge monotonically toward S_c and then remain within S_c. If $X_{i,0} \in P_c$ and $S_{c+1} \neq \varnothing$, then $X_{i,s}$ will converge monotonically toward S_{c+1} and then remain within S_{c+1}.

Assuming nonemptiness of these ranges, the interval $P_{c-1} \cup S_c \cup N_c$ is a "basin of attraction" for S_c, i.e., X in this interval converges monotonically to S_c and then remains there. The main difference relative to the deterministic, time homogenous case is the "blurring" of the stable equilibrium to a stable set S_c.

We did not make any assumptions on the joint distribution of the unobserved factors $\epsilon_{i,1}, \ldots, \epsilon_{i,t}$. The whole argument of the preceding theorem is *conditional* on these factors. However, the predictions of the theorem will be sharper (given g) if serial dependence of unobserved factors is stronger, increasing the number of units i to which the assertion is applicable and reducing the size of the intervals S_c and U_c, since $g^U_{i,t} - g^L_{i,t}$ is going to be smaller on average.

In summary, proposition 10 implies that, if we do not find multiple roots in quantile regressions, then the conditional average structural functions $E\left[g(x', \epsilon) | g(X, \epsilon) = 0, X\right]$ do not have multiple roots. Proposition 11 implies that, if upper and lower envelopes of $g(., \epsilon_{i,s})$

do not have multiple roots, then the dynamics of the system are stable and initial conditions do not matter in the long run.

2.4 Application to the dynamics of neighborhood composition

This section analyzes the dynamics of minority share in a neighborhood, applying the methods developed in the last two sections to the data used for analysis of neighborhood composition dynamics by Card, Mas, and Rothstein (2008). Card, Mas, and Rothstein (2008) study whether preferences over neighborhood composition lead to a "white flight", once the minority share in a neighborhood exceeds a certain level. They argue that such "tipping" behavior implies discontinuities in the change of neighborhood composition over time as a function of initial composition, and test for the presence of such discontinuities in cross-sectional regressions over different neighborhoods in a given city. The authors provided full access to their datasets, which allows us to use identical samples and variable definitions as in their work.

The data set is an extract from the Neighborhood Change Database, or NCDB, which aggregates US census variables to the level of census tracts. Tract definitions are changing between census waves but the NCDB matches observations from the same geographic area over time, thus allowing observation of the development over several decades of the universe of US neighborhoods. In the dataset used by Card, Mas, and Rothstein (2008), all rural tracts are dropped, as well as all tracts with population below 200 and tracts that grew by more than 5 standard deviations above the MSA mean. The definition of MSA used is the MSAPMA from the NCDB, which is equal to Primary Metropolitan Statistical Area if the tract lies in one of those, and equal to the MSA it lies in otherwise. For further details on sample selection and variable definition, see Card, Mas, and Rothstein (2008).

The graphs and tables to be discussed are constructed as follows. For each of the MSAs and each of the decades separately, we run local linear quantile regressions of the change in minority share of a neighborhood (tract) on minority share at the beginning of the decade. This is done for the quantiles 0.2, 0.5 and 0.8, with a bandwidth τ of $n^{-.2}$, where n is the sample size.[5] The left column of graphs in figure 2.5 shows these quantile regressions for the three largest MSAs.

For each of the regressions, Z_ρ is calculated, where ρ is chosen as 0.04. The integral in the expression for Z_ρ is taken over the interval $[0, 1]$, intersected with the support of initial minority share if the latter is smaller. Note that it is possible to find no (stable) equilibrium for an MSA, i.e. $Z_\rho < 1$, if high initial minority shares do not occur in that MSA and most neighborhoods experienced growing minority shares. Figure 2.6 shows kernel density plots of

[5] The implementation of local linear quantile regression uses code downloaded from Koenker (2009).

the regressor, initial minority share, which suggest that support problems are not an issue, at least for the largest MSAs. For each Z_ρ, bootstrap standard errors and bias are calculated, as well as the corresponding t-test statistics for the null hypothesis $Z_\rho = 0, 1, 2, 3, \ldots$, implying an integer-valued confidence set (of level .05) for z. By the results of section 2.2, these confidence sets have an asymptotic coverage probability of 95%. By the Monte Carlo evidence of appendix A, they are likely to be conservative, i.e., have a larger coverage probability. If the confidence sets thus obtained are empty, the two neighboring integers of $\widehat{Z}$ are included in the intervals shown. This makes inference even more conservative. Table 2.1 shows the resulting confidence sets for the twelve largest MSAs in the United States (by 2009 population), for all quantiles and decades under consideration.[6]

As can be seen from the table, in very few cases there is evidence of Z exceeding 1. In all cases shown, except for the .2 quantile for Atlanta in the 1980s, we can reject the null $Z \geq 3$. Similar patterns hold for almost all of the 118 cities in the dataset. Rather than exhibiting multiple equilibria, the data indicate a general rise in minority share that is largest for neighborhoods with intermediate initial share, but not to the extent of leading to tipping behavior. Proposition 10 in section 2.3.5 suggests that, if we do not find multiple roots in quantile regressions, we can reject multiple equilibria in the underlying structural relationship. I take these results as indicative that tipping is not a widespread phenomenon in US ethnic neighborhood composition over the decades under consideration. This stands in contrast to the conclusion of Card, Mas, and Rothstein (2008), who do find evidence of tipping.

The approach used here differs from the main analysis in Card, Mas, and Rothstein (2008) in a number of ways. Card, Mas, and Rothstein (2008) (i) use polynomial least squares regression with a discontinuity. They (ii) use a split sample method to test for the presence of a discontinuity, and they (iii) regress the change in the non-Hispanic, white population, divided by initial neighborhood population, on initial minority share. We (i) use local linear quantile regression without a discontinuity, we (ii) run the regressions on full samples for each MSA and test for the number of roots, and we (iii) regress the change in minority share on initial minority share.

To check whether the differing results are due to variable choice (iii) rather than testing procedure, the figures and tables that were just discussed are replicated using the change in the non-Hispanic, white population relative to initial population as the dependent variable, as did Card, Mas, and Rothstein (2008). The right column of figure 2.5 shows such quantile regressions. These figures correspond to the ones in Card, Mas, and Rothstein (2008), p.190, using the same variables but a different regression method and the full samples. Table 2.2 shows confidence sets for the number of roots of these regressions for the 12 largest MSAs. In comparing tables 2.1 and 2.2, note that there is a correspondence between the lower quantiles of the first (low increase in minority share) and the upper quantiles of the latter

[6] The full set of results for all 115 MSAs in the dataset can be found in the web-appendix, Kasy (2010).

(higher increase/lower decrease of white population). The two tables show fairly similar results. Again, no systematic evidence of multiple roots is found.

Some factors might lead to a bias in the estimated number of equilibria, using the methods developed here. First, the test might be sensitive to the chosen range of integration if there are roots near the boundary. If a root lies right on the boundary of the chosen range of integration, it enters Z_ρ as $1/2$ only. Extending the range of integration beyond the unit interval, however, might also lead to an upward bias in the estimated number of roots, if extrapolated regression functions intersect with the horizontal axis. Second, choosing a bandwidth parameter ρ that is too large might bias the estimated number of equilibria downwards, if the function g peaks within the range $[-\rho, \rho]$. Third, there might be roots of g in the unit interval but beyond the support of the data.

2.5 Summary and conclusion

This paper proposes an inference procedure for the number of roots of functions nonparametrically identified using conditional moment restrictions, and develops the corresponding asymptotic theory. In particular, it is shown that a smoothed plug-in estimator of the number of roots is super-consistent under i.i.d. asymptotics, but asymptotically normal under non-standard asymptotics, and asymptotically efficient relative to a simple plug-in estimator. In section 2.3, these results are extended to cover various more general cases, allowing for covariates as controls, higher dimensional domain and range, and for inference on the number of equilibria with various stability properties. This section also discusses how to apply the results to static games of incomplete information and to stochastic difference equations. In an application of the methods developed here to data on neighborhood composition dynamics in the United States, no evidence of equilibrium multiplicity is found.

The inference procedure can also be used to test for bifurcations, i.e., (dis)appearing equilibria as a function of changing exogenous covariates. It is easy to test the hypothesis $Z(g(., W_1)) = Z(g(., W_2))$, since the corresponding estimators $\widehat{Z}(g(., W_i))$ are independent for W_1 and W_2 further apart than twice the bandwidth τ. If there are bifurcations, small exogenous shifts might have a large (discontinuous) effect on the equilibrium attained, if the "old" equilibrium disappears.

In the dynamic setup, one might furthermore consider to apply the procedure to detrended data, for instance by demeaning ΔY. It seems likely that regressions of detrended data have a higher number of roots. The rationale of such an approach could be found in underlying models in which the dynamics of a detrended variable are stationary. This is in particular the case in Solow-type growth models, in which GDP or capital stock is stationary after normalizing by a technological growth factor.

Finally, it might also be interesting to extend the results obtained here to cover further cases where g can not be directly estimated using conditional moment restrictions. The crucial step for such extensions, as illustrated by the various cases discussed in section 2.3, is to find a sequence of experiments such that the first stage estimator $\widehat{g}$ converges in probability to a degenerate limit whereas $\widehat{g'}$ converges in distribution to a non-degenerate limit. Furthermore, $\widehat{g'}(x_1)$ needs to be asymptotically independent of $\widehat{g'}(x_2)$ for all $|x_1 - x_2| > const. \cdot \tau$.

There are many potential applications of the results obtained here, where it might be interesting to know whether the underlying dynamics or strategic interactions imply multiple equilibria. Examples include household level poverty traps, intergenerational mobility, efficiency wages, macro models of economic growth (as analyzed in the web appendix), financial market bubbles (herding), market entry, and social norms.[7]

Appendix 2.A Monte Carlo evidence

This section presents simulation results to check the accuracy in finite samples of the asymptotic approximations obtained in theorem 2. In all simulations, the X are i.i.d. draws of Uni$[0,1]$ random variables, and the additive errors γ are either uniformly or normally distributed:

$$\begin{aligned} X_i &\sim^{iid} \text{Uni}[0,1] \\ \gamma_i|X_i &\sim f_{\gamma|X} \\ Y_i &= g^j(X_i) + \gamma_i, \end{aligned} \tag{2.37}$$

where $f_{\gamma|X}$ is an appropriately centered and scaled uniform or normal distribution. Two functions g^j are considered, the first with one root and the second with three roots:

$$\begin{aligned} g^1(x) &= 0.5 - x \\ g^2(x) &= 0.5 - 5x + 12x^2 - 8x^3. \end{aligned}$$

The function g is estimated by median regression, mean regression and .9 quantile regression, where the γ in the simulations are shifted appropriately to have median, mean or .9 quantile at the respective g. The figures and tables show sequences of four experiments with 400, 800, 1600 and 3200 observations. The variance of γ in each experiment is chosen to yield the same variance for $\widehat{g'}$, as implied by the asymptotic approximation of the Bahadur expansion, in all experiments for a given g. By the proof of theorem 2, we should therefore get similar simulation results across all setups. Furthermore, the variance of $\widehat{Z}$ should be constant up to a factor τ/ρ. The parameters of these simulations are chosen to lie in an intermediate range where variation in $\widehat{g'}$ is existent but moderate.

[7] The Matlab/Octave code written for this paper is available upon request.

Figure 2.7 shows density plots for $\widehat{Z}$ for the sequences of experiments with uniform errors and median regressions; in the web-appendix, Kasy (2010), similar figures are presented for the other experiments. Visual inspection does not reveal obvious failures of the approximation by normal densities. Table 2.3 shows the results of simulations using bootstrapped standard deviations and biases, for mean regression with uniform errors. The results show, for the range of experiments considered, that rejection frequencies are lower than the 0.05 value implied by asymptotic theory. If this pattern generalizes, inference based upon the t-statistic proposed in this paper is conservative in finite samples. In particular, it seems that bootstrapped standard errors are too large.

Appendix 2.B Proofs

Proof of proposition 8: By continuity of g' as well as genericity of g we can choose ρ small enough such that $sgn(g'(x))$ is constantly equal to $sgn(g'(x_c)) \neq 0$ in each of the neighborhoods of the $c = 1, \dots, z$ roots of g, $\{x_c\}$, defined by $L_\rho(g(x)) \neq 0$. Hence we can write the integral $\int_{\mathscr{X}} L_\rho(g(x))|g'(x)|dx$ as a sum of integrals over these neighborhoods, in each of which there is exactly one root. Assume w.l.o.g. that $z = 1$ and g' is constant in the range of x where $L_\rho(g(x)) \neq 0$. Then, by a change of variables setting $y = g(x)$,

$$\int_{\mathscr{X}} L_\rho(g(x))|g'(x)|dx = \int_{g(\mathscr{X})} L_\rho(y)|g'(g^{-1}(y))|\frac{1}{|g'(g^{-1}(y))|}dy = 1$$

□

Proof of proposition 9: We need to find ϵ such that $||g - \tilde{g}|| < \epsilon$ implies $Z(\tilde{g}) = Z(g)$. By genericity of g, each root x_c of g is such that $sgn(g'(x_c)) \neq 0$. By continuous first derivatives we can then find δ such that $sgn(g'(.))$ is constant in the neighbourhood $NH_c := (x_c - \delta, x_c + \delta)$ of each of the finitely many roots x_c and the NH_c are mutually disjoint. By continuity of g,

$$\epsilon_1 := \inf_{x \notin \bigcup_c NH_c} g(x) > 0 \tag{2.38}$$

and

$$\epsilon_2 := \inf_{x \in \bigcup_c \bar{NH}_c} |g'(x)| > 0, \tag{2.39}$$

where $\bar{NH}_c$ is the closure of NH_c. Choosing $\epsilon = \frac{1}{2}\min(\epsilon_1, \epsilon_2)$ fulfills our purpose. To see this choose a $\tilde{g}$ such that $||g - \tilde{g}|| < \epsilon$. For $x \notin \bigcup_c NH_c$ $\tilde{g}$ is bounded away from zero by equation (2.38). In NH_c there must be exactly one x such that $\tilde{g}(x) = 0$: Since the NH are mutually disjoint, $sgn(g(x_c - \delta)) \neq sgn(g(x_c + \delta))$, by (2.38) again $sgn(g(x_c - \delta)) = sgn(\tilde{g}(x_c - \delta))$ and $sgn(g(x_c + \delta)) = sgn(\tilde{g}(x_c + \delta))$, and finally the sign of $\tilde{g}'$ is constantly equal to $sgn(g'(x_c))$

in NH_c by equation (2.39).

The assertion for Z_ρ follows now from the first part of this proof, combined with proposition 8, if we can choose a ρ independent of $\tilde{g}$ such that proposition 8 applies. Sufficient for this is a ρ that separates roots. Choosing $\rho = \epsilon$ accomplishes this. By equation (2.38), L_ρ will separate the NH_c, and by the previous argument each of the NH_c will contain exactly one root of $\tilde{g}$. □

Proof of theorem 2:
Write $Z^1 =^A Z^2$, if $a_n Z^1 - b_n$ and $a_n Z^2 - b_n$ have the same non-degenerate distributional limit for some non-random sequences a_n and b_n. In particular, as long as such sequences exist that guarantee convergence to a non-degenerate limit, this is implied by equality up to a remainder which is asymptotically negligible under the given sequence of experiments, i.e., $Z^1 =^A Z^2$ if $Z^1 - Z^2 = o_p(Z^2)$. We will use Z^1, Z^2, Z^3 to denote a sequence of approximations to $\widehat{Z}$.

1) Approximation of $\widehat{g}$ with g:

$$\widehat{Z} =^A Z_\rho(g, \widehat{g'})$$

The remainder of this approximation is given by

$$\int (L_\rho(g) - L_\rho(\widehat{g}))|\widehat{g'}|$$

Negligibility of this remainder follows from uniform convergence of $\widehat{g}$ under our sequence of experiments at a rate faster than ρ, which is a consequence of Bahadur expansion (2.8) and of $\rho/\tau \to \infty$. Assuming that L_ρ is Lipschitz with constant C/ρ, this in turn implies uniform convergence of $(L_\rho(g) - L_\rho(\widehat{g}))$ to 0. This, combined with the arguments proving distributional convergence of $\int \widehat{g'}$ over neighborhoods of the roots of g, given below, proves that the remainder is $o_p(\widehat{Z})$.

2) Approximation of $\widehat{g'}$ by the Bahadur expansion:

$$\begin{aligned} Z_\rho(g, \widehat{g'}) =^A \int L_\rho(g(x)) \Big| g'(x) - f^{-1}(x) s_n^{-1}(x) I_n(x) \cdot \\ \cdot \frac{1}{n} \sum_i K_\tau(X_i - x) \phi(Y_i - g(x) - g'(x)(X_i - x)) \left(\frac{X_i - x}{\nu_2 \tau^3} \right) \Big| \, dx =: Z^1 \end{aligned}$$

The absolute value of the remainder of this approximation is less than or equal to

$$\int L_\rho(g)|R|,$$

where R is the remainder of the Bahadur expansion. Negligibility of the remainder of the approximation is a consequence of the assumption that the remainder of the Bahadur expansion is negligible, i.e., $R = o_p\left(\left(\widehat{g}, \widehat{g'}\right) - (g, g')\right)$ uniformly in x.

3) Restriction to one root at 0 and Taylor approximations:
Assume that $g(0) = 0$ and $g(x) \neq 0$ for $x \neq 0$ (i.e., $Z = 1$). This is without loss of generality, since the integral for the general case is simply a sum of the independent integrals in a neighborhood of each root.

Now define $c = g'(0)$, $w = -f^{-1}(0)s^{-1}(0)\frac{1}{\nu_2}$, $\phi_i = \phi(e_i)$ and $\tilde{K}_\tau(d) = K_\tau(d)\frac{d}{\tau}$.

By replacing g with $g'(0)x$ in $L_\rho(g(x))$ and replacing $-f^{-1}(x)s^{-1}(x)\frac{1}{\nu_2}$ with w, both justified by smoothness and $\rho \to 0$, as well as $I_n(x) \to 1$ uniformly, we get

$$Z^1 =^A \int L_\rho(cx) \left| g'(0) - f^{-1}(0)s^{-1}(0)\frac{1}{\nu_2 \tau^3}\frac{1}{n}\sum_i K_\tau(X_i - x)(X_i - x)\phi(\epsilon_i)\right| dx$$

$$= \int L_\rho(cx) \left| c + w\frac{r_n}{\tau^2}E_n\left[\tilde{K}_\tau(X_i - x)\phi_i\right]\right| dx = Z^2$$

The absolute value of the remainder of this approximation is less than or equal to

$$\int |L_\rho(g) - L_\rho(cx)| \left| g' - \sum \right| + \int L_\rho(cx)\left| f^{-1}(x)s^{-1}(x)I_n(x)\frac{1}{\nu_2\tau^2} - w\right| |E_n| .$$

Both terms in this expression go to 0 as $\rho \to 0$. We can assume furthermore that

$$X_i \sim^{iid} Uni([-\rho/c, \rho/c])$$

conditional on falling in this interval and that

$$\phi_i \sim^{iid} \phi(e)|X = 0$$

These assumptions are justified by another Taylor approximation, this time of the distribution functions $F_X(x) = F_X(0) + f_X(0)X + o(X)$ and $F_{\phi|X}(\phi|X) = F_{\phi|X}(\phi|0) + O(X)$, assuming both distribution functions to be C^1. To see that this approximation is justified, note that distributional convergence to the same limit is equivalent to convergence of the expectations of any Lipschitz continuous bounded function of the statistics to the same limit.

The difference in expectations between a function h of Z^2 and of its approximation using conditionally uniform X and i.i.d. ϕ is given by

$$\int h(Z^2) \prod_i \left(f_X(X^i) f_{\phi|X}(\phi^i|X^i) - f_X(0) f_{\phi|X}(\phi^i|0) \right)$$

This integral goes to 0 because the support of $h(Z^2)$ in X is a neighborhood of 0 shrinking to 0.

4) Partitioning the range of integration:
Partition $[-\rho/c, \rho/c]$ into subintervals $[t_j, t_{j+1}]$, $j = 1, \ldots, \lfloor \rho/\tau \rfloor$ with $t_{i+1} - t_i = 2\tau$. Then

$$Z^2 =^A \sum_{j=1}^{\lfloor \rho/c\tau \rfloor} L_\rho(ct_j)\xi_j = Z^3$$

with

$$\xi_j = \int_{t_j}^{t_{j+1}} \left| c + w\frac{r_n}{h^2} E_n \left[\tilde{K}_\tau(X_i - x)\phi_i \right] \right| dx$$

The remainder of this approximation is given by

$$\int \left(L_\rho(cx) - L_\rho \left(c \left(\max_{t_j < x} t_j \right) \right) \right) \left| c + w\frac{r_n}{h^2} E_n \right|$$

This approximation is warranted by Lipschitz continuity of L_ρ with a Lipschitz constant of order $1/\rho^2$, and by $\tau/\rho^2 \to 0$.

5) Poisson approximation:
The following argument essentially replaces the number of X falling into the interval $[-\rho/c, \rho/c]$, which is approximately distributed $Bin(n, 2f(0)\rho/c)$, with a Poisson random variable with parameter $2nf(0)\rho/c$; the distribution of everything else conditional on this number remains the same.

Let n_j be distributed i.i.d. $Poisson(2n\tau f(0))$ for $j = 1, \ldots, \lfloor \rho/\tau \rfloor$. This is an approximation to the number of X falling into the bin $[t_j, t_{j+1}]$.
Draw $X_{jl} \sim^{iid} Uni([t_j, t_{j+1}])$ and $\phi_{jl} \sim^{iid} \phi(e)|X = 0$ for $j = 1, \ldots, \lfloor \rho/\tau \rfloor$ and $l = 1, \ldots, n_j$. Now define

$$\pi_j = \int_{t_j}^{t_{j+1}} \left| c + w\frac{r_n}{n\tau^2} \sum_{k=j-1}^{j+1} \sum_{l=1}^{n_k} \left[\tilde{K}_\tau(X_{jl} - x)\phi_{jl} \right] \right| dx.$$

Then

$$Z^3 =^A \sum_{j=1}^{\lfloor \rho/c\tau \rfloor} L_\rho(ct_j)\pi_j$$

where the π_j are identically distributed and π_j is independent of π_k for $|j-k| \geq 2$.

Conditional on $\tilde{n} := \sum_j n_j$, the equality is exact. The exact distribution of the number of observations falling in the interval $[-\rho/c, \rho/c]$, corresponding to $\tilde{n}$, would be given by

$$\frac{(2n(\rho/c)f(0))^{\tilde{n}}}{\tilde{n}!} \frac{n!}{n^{\tilde{n}}(n-\tilde{n})!}(1-2(\rho/c)f(0))^{(n-\tilde{n})}.$$

The Poisson approximation sets the latter part of this expression to a constant in $\tilde{n}$. This is justified by the usual arguments deriving the Poisson distribution as a limit of Binomial distributions. The approximation of Z^3 follows by an argument similar to the one of point 3, second part, once we note that the multinomial p.m.f. converges uniformly.

6) Moments of the integrals over the subintervals:

- $E[\pi_j] = \tau\mu_1 + o(\tau)$
- $E[\pi_j^2] = \tau^2\mu_2 + o(\tau^2)$
- $E[\pi_j\pi_{j+1}] = \tau^2\mu_{11} + o(\tau^2)$
- $E[\pi_j^3] = \tau^3\mu_2 + o(\tau^3)$

These equations follow from noting first pointwise convergence to normality of

$$\Gamma(x) = w\frac{r_n}{n\tau^2}\sum_{k=j-1}^{j+1}\sum_{l=1}^{n_k}\left[\tilde{K}_\tau(X_{jl}-x)\phi_{jl}\right] \to N(0,v)$$

under our sequence of experiments. This is the point where the rate r_n matters:

$$\Gamma(x) = w\frac{\tau^{1/2}}{n^{1/2}}\sum_{k=j-1}^{j+1}\sum_{l=1}^{n_k}\left[\tilde{K}_\tau(X_{jl}-x)\phi_{jl}\right]$$

$$\sim w\frac{1}{(n\tau)^{1/2}}\sum_{l=1}^{(n_{j-1}+n_j+n_{j+1})}[K(\zeta_l)\zeta_l\phi_l] =$$

$$= w\left(\frac{n_{j-1}+n_j+n_{j+1}}{n\tau}\right)^{1/2}\left(\frac{1}{n_{j-1}+n_j+n_{j+1}}\right)^{1/2}\sum_{l=1}^{(n_{j-1}+n_j+n_{j+1})}[K(\zeta_l)\zeta_l\phi_l],$$

where the ζ_j are i.i.d. $Uni[-3,3]$. Now asymptotic normality follows by noting $\left(\frac{n_{j-1}+n_j+n_{j+1}}{n\tau}\right) \to^p 6f(0)$, $(n_{j-1}+n_j+n_{j+1}) \to^p \infty$ and $E[\phi_l|X_l]=0$. Similarly

$$\left(\begin{array}{c}\Gamma(x_1)\\ \Gamma(x_1+\tau\delta)\end{array}\right) \to N\left(\left(\begin{array}{c}0\\0\end{array}\right),\left(\begin{array}{cc}v & corr(|\delta|)\cdot v\\ corr(|\delta|)\cdot v & v\end{array}\right)\right).$$

Second, a change of the order of integration and the limit in n delivers the claims, where this change of order is justifiable by the dominated convergence theorem. For instance,

$$\lim(E[\pi_j^2]/\tau^2) = 4\lim E\left[\int_{[0,1]^2}|(c+\Gamma(t_j+2\tau\delta_1))(c+\Gamma(t_j+2\tau\delta_2))|\,d\delta_1 d\delta_2\right]$$

$$= 4\int_{[0,1]^2}\lim E\left[|(c+\Gamma(t_j+2\tau\delta_1))(c+\Gamma(t_j+2\tau\delta_2))|\right]d\delta_1 d\delta_2$$

7) Central limit theorem applied to the sum of integrals over the subintervals: Now apply a central limit theorem for m-dependent sequences to the sum of integrals. For a definition of m-dependence, see Hoeffding and Robbins (1994). Note that $L_\rho(ct_j)\pi_j$ is an m-dependent sequence with $m=1$. We have

$$Var\left(\sqrt{\frac{\rho}{\tau}}\sum_{j=1}^{\lfloor\rho/c\tau\rfloor}L_\rho(ct_j)\pi_j\right) =$$

$$\frac{\rho}{\tau}\left(\sum_j L_\rho^2(ct_j)Var(\pi_j) + \sum_j L_\rho(ct_j)\left(L_\rho(ct_{j-1})+L_\rho(ct_{j+1})\right)Cov(\pi_j,\pi_{j+1})\right)$$

$$\approx \left(\frac{\rho}{\tau}\right)\left(\frac{c}{\tau}\right)\int_{-\rho/c}^{\rho/c}L_\rho^2(cu)\tau^2(\mu_2+2\mu_{11}-3\mu_1^2)du$$

$$= c(\mu_2+2\mu_{11}-3\mu_1^2)\int_{-1}^{1}L_1^2(cu)du$$

Asymptotic normality for $\sqrt{\frac{\rho}{\tau}}\left(Z^3-E[Z^3]\right)$ follows, and by $\widehat{Z} =^A Z^3$, the same holds for $\sqrt{\frac{\rho}{\tau}}\left(\widehat{Z}-E[\widehat{Z}]\right)$. Furthermore, $E\left[Z^3\right]=O(1)$, and hence so is $E\left[\widehat{Z}\right]$. $\square$

Proof of theorem 3 (Sketch):
We will approximate $M(a,b,x,w_1)$ by a criterion function that has the form of equation (2.6), i.e., a local weighted average over the empirical distribution of some objective function. Based on this approximation we can then again apply the results of Kong, Linton, and Xia (2010). Newey (1994) provides a set of results that facilitate such approximations of

partial means. In particular, lemma 5.4 in Newey (1994) allows derivation of the required approximation by replacing the outer sum over j in equation (2.13) with an expectation, and by linearizing the fraction inside. The first replacement is asymptotically warranted since the variation created by averaging over the empirical distribution is of order $1/\sqrt{n}$ and hence dominated by the variation in the nonparametric component. The second replacement follows from differentiability and requires in particular that the denominator of the fraction be asymptotically bounded away from zero. This is guaranteed by the requirement that W_2 has full conditional support given (X, W_1). Formally, lemma 5.4 in Newey (1994) gives

$$M(a,b,x,w_1) - E_{W_2}[E_{m|X,W}[m|X=x, W_1=w_1, W_2]] = \tilde{M}(a,b,x,w_1) + o_p(\tilde{M}(a,b,x,w_1)),$$

where $\tilde{M}(a,b,x,w_1) :=$

$$\frac{1}{n}\sum_j \left(K_\tau(X_j - x, W_{1j} - w_1) \frac{m(Y_j - a - b(X_j - x)) - E[m(Y_j - a - b(X_j - x))|X_j, W_j]}{f_{X,W_1|W_2}(X_j, W_{1j}|W_{2j})} \right). \tag{2.40}$$

This approximation of the objective function has the general form assumed in Kong, Linton, and Xia (2010) if we set

$$\tilde{m}(Y,X,W,a,b,x) := \frac{m(Y - a - b(X-x)) - E[m(Y - a - b(X-x))|X,W]}{f_{X,W_1|W_2}(X, W_1|W_2)}, \tag{2.41}$$

providing us with the desired Bahadur expansion. Choosing the appropriate sequence of experiments, from here on the entire proof and result of theorem 2 go through unchanged. If $W_1 \neq const.$, the rates have to be adapted as follows. The number of observations within each rectangle of size τ^d goes to ∞ if $n\tau^d \to \infty$. Finally, the variance of $\widehat{g}'$ converges iff $r_n = O\left(n\tau^{(4+d)}\right)^{1/2}$. □

Proof of theorem 4: The proof requires the following modifications relative to the one-dimensional case: Assumption 2.2.1 is still applicable, where the only difference in the d-dimensional case is that (2.8) has to be multiplied by $1/\tau^{d-1}$. For $\widehat{g}'$ to have a point-wise non-degenerate distributional limit, we have to choose the rate r_n to equal $(n\tau^{4+d})^{1/2}$, which is slower for higher d. To see this note that $Var(\widehat{g}') = O\left(\frac{r_n^2}{n\tau^{4+d}}\right)$. L_ρ is Lipschitz continuous of order $\rho^{-(1+d)}$, so that we require $\tau/\rho^{d+1} \to 0$ for step 4 of the proof. The range of integration has to be partitioned into rectangular subranges of area τ^d instead of intervals of length τ. There will be approximately $const \cdot (\rho/\tau)^d$ such subintegrals. The variance of the integral of $|\widehat{g}'|$ over each of these subranges will be of order τ^{2d}, similarly for expectations and covariances. This yields a variance of $\widehat{Z}$ of $O((\tau/\rho)^d)$; see point 7 of the proof. □

Proof of theorem 5: By equations (2.23) and (2.25), it is sufficient to show that $r_n \cdot \widehat{g_1'^1}(g_{2,n}(\bar{\sigma}_1, s_2), s_1)$ and $r_n \cdot \widehat{g_2'^1}(\bar{\sigma}_1, s_1))$ converge jointly in distribution, while $r_n \cdot \widehat{g}(\bar{\sigma}_1, s)$, as well as $\widehat{\sigma}$, converge in probability. These claims follow as before if we combine the convergence of $r_n g_n$ from display (2.31) with Bahadur expansion (2.8) for $\widehat{g_2'^1}$ and $\widehat{g_1'^1}$, where the latter are evaluated at $\bar{\sigma}_{2,n}$, which is not constant but converges. $\square$

Proof of lemma 20:
By definition of conditional quantiles, $F^{\Delta X|X}\left(Q^{\Delta X|X}(q|X)|X\right) = q$. Differentiating this with respect to X gives

$$\frac{\partial}{\partial X} Q^{\Delta X|X}(q|X) = -\frac{\frac{\partial}{\partial X} F^{\Delta X|X}(Q|X)}{f^{\Delta X|X}(Q|X)}. \tag{2.42}$$

The differential in the numerator has two components, one due to the structural relation between ΔX and X, i.e., the derivative with respect to the argument X of $d(X, \epsilon)$, and one due to the stochastic dependence of X and ϵ.

$$\begin{aligned} \frac{\partial}{\partial X} F^{\Delta X|X}(Q|X) &= E\left[g_X \cdot f^{\Delta X|g_X, X}(Q|g_X, X)\middle|X\right] \\ &+ \frac{\partial}{\partial X} \mathbb{P}\left(g(X', \epsilon) \leq Q|X\right)\Big|_{X'=X}. \end{aligned}$$

This can be seen as follows: We can decompose the derivative according to

$$\frac{\partial}{\partial X} F^{\Delta X|X}(Q|X) = \left[\frac{\partial}{\partial X'} + \frac{\partial}{\partial X}\right] \mathbb{P}\left(g(X', \epsilon) \leq Q|X\right)\Big|_{X'=X}.$$

To simplify the first derivative, note that by iterated expectations

$$\mathbb{P}\left(g(X', \epsilon) \leq Q|X\right) = E[F(g(X', \epsilon)|X, g_X)|X].$$

Differentiating this with respect to X' gives

$$E\left[g_X \cdot f^{\Delta X|g_X, X}(Q|g_X, X)|X\right].$$

The claim now is immediate. $\square$

Proof of proposition 10:
Since X and $X+\Delta X$ have their support in the interval $[0,1]$, $Q^{\Delta X|X}(q|0) \geq 0$ and $Q^{\Delta X|X}(q|1) \leq 0$. Therefore the unique root X of $Q^{\Delta X|X}(q|X)$ must be stable, $\frac{\partial}{\partial X} Q^{\Delta X|X}(q|X) \leq 0$.
By lemma 20 and assumption (2.3.1), this implies that $E\left[g_X|\Delta X = Q, X\right] \leq 0$.
Finally, note that for all X where $(0, X)$ is in the support of $(\Delta X, X)$, there exists a q

such that $Q^{\Delta X|X}(q|X) = 0$. $\square$

Proof of proposition 11: The claims are immediate, noting that $N_c = \bigcap_s [x_c^s, x_{c+1}^s]$ and similarly for P_c. Furthermore, $x_c^s \in S_c$ for all s, $c = 1, 3, \ldots$ and $x_c^s \in U_c$ for all s, $c = 2, 4, \ldots$. Next, $g(., e_{i,s}) < 0$ on $[x_c^s, x_{c+1}^s]$, $c = 1, 3, \ldots$ from which negativity on N_c follows, similarly for P_c.

Finally, under monotonicity of potential outcomes, assuming for simplicity differentiability of g,

$$\frac{\partial}{\partial e} x_c = -\frac{\frac{\partial}{\partial e} g}{\frac{\partial}{\partial x} g}.$$

The numerator is always positive by assumption, the denominator is negative for $c = 1, 3, \ldots$ and positive for $c = 2, 4, \ldots$ since we had assumed g positive for sufficiently small x, hence $\frac{\partial}{\partial e} x_c$ is positive for $c = 1, 3, \ldots$ and negative for $c = 2, 4, \ldots$. $\square$

Appendix 2.C Figures and tables

Table 2.1: .95 CONFIDENCE SETS FOR $Z(g)$ FOR THE 12 LARGEST MSAS OF THE UNITED STATES BY DECADE AND QUANTILE, CHANGE IN MINORITY SHARE

MSA	**70s**			**80s**			**90s**		
	$q=.2$	$q=.5$	$q=.8$	$q=.2$	$q=.5$	$q=.8$	$q=.2$	$q=.5$	$q=.8$
New York, NY PMSA	[0,1]	[0,1]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]
Los Angeles-Long Beach, CA PMSA	[1,1]	[1,1]	[0,1]	[0,1]	[0,1]	[0,1]	[1,1]	[1,1]	[0,0]
Chicago, IL PMSA	[0,1]	[0,1]	[0,1]	[2,2]	[0,1]	[0,1]	[1,1]	[0,1]	[0,0]
Dallas, TX PMSA	[1,2]	[1,1]	[0,0]	[0,1]	[0,0]	[0,0]	[0,1]	[0,1]	[0,0]
Philadelphia, PA-NJ PMSA	[1,2]	[0,1]	[0,1]	[1,1]	[0,1]	[0,1]	[1,1]	[0,1]	[0,0]
Houston, TX PMSA	[1,1]	[0,0]	[0,0]	[1,2]	[0,1]	[0,0]	[0,1]	[0,0]	[0,0]
Miami, FL PMSA	[0,1]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]
Washington, DC-MD-VA-WV PMSA	[0,1]	[0,0]	[0,0]	[1,1]	[0,1]	[0,0]	[1,1]	[0,1]	[0,0]
Atlanta, GA MSA	[1,1]	[1,1]	[0,0]	[2,3]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]
Boston, MA-NH PMSA	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,0]	[1,1]	[0,0]	[0,1]
Detroit, MI PMSA	[1,2]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,0]
Phoenix-Mesa, AZ MSA	[1,1]	[0,0]	[0,0]	[1,1]	[0,1]	[0,0]	[1,1]	[0,1]	[0,0]
San Francisco, CA PMSA	[1,1]	[0,1]	[0,1]	[0,0]	[0,1]	[0,0]	[1,1]	[0,0]	[0,0]

Notes: The table shows confidence intervals in the integers for $Z(g)$ for the 12 largest MSAs of the United States, ordered by population, where g is estimated by quantile regression of the change in minority share over a decade on the initial minority share for the quantiles .2,.5 and .8. Regression bandwidth τ is $n^{-.2}$, σ is chosen as .04. Confidence sets are based on t-statistics using bootstrapped bias and standard errors.

Table 2.2: .95 confidence sets for $Z(g)$ for the 12 largest MSAs of the United States by decade and quantile, change in white population

MSA	**70s**			**80s**			**90s**		
	$q = .2$	$q = .5$	$q = .8$	$q = .2$	$q = .5$	$q = .8$	$q = .2$	$q = .5$	$q = .8$
New York, NY PMSA	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]
Los Angeles-Long Beach, CA PMSA	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]
Chicago, IL PMSA	[0,1]	[0,1]	[0,1]	[0,0]	[0,1]	[1,1]	[0,1]	[0,1]	[0,1]
Dallas, TX PMSA	[0,1]	[0,1]	[0,1]	[0,0]	[1,1]	[0,2]	[0,1]	[1,1]	[0,1]
Philadelphia, PA-NJ PMSA	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[1,1]
Houston, TX PMSA	[0,1]	[0,1]	[0,1]	[1,1]	[1,1]	[1,1]	[0,1]	[0,1]	[0,1]
Miami, FL PMSA	[0,1]	[0,1]	[0,1]	[0,0]	[0,0]	[1,1]	[1,1]	[1,1]	[1,1]
Washington, DC-MD-VA-WV PMSA	[0,1]	[0,0]	[0,1]	[0,0]	[1,1]	[0,0]	[0,1]	[0,1]	[0,1]
Atlanta, GA MSA	[0,1]	[1,1]	[0,1]	[1,1]	[1,1]	[1,1]	[1,1]	[1,2]	[0,1]
Boston, MA-NH PMSA	[0,1]	[0,1]	[0,1]	[0,0]	[0,0]	[1,1]	[0,0]	[0,1]	[0,1]
Detroit, MI PMSA	[0,1]	[0,1]	[0,1]	[0,0]	[0,0]	[1,1]	[0,1]	[0,1]	[0,1]
Phoenix-Mesa, AZ MSA	[0,1]	[0,1]	[0,1]	[0,0]	[1,1]	[0,0]	[0,1]	[0,1]	[0,1]
San Francisco, CA PMSA	[0,1]	[0,1]	[0,1]	[0,0]	[0,0]	[0,0]	[0,0]	[1,1]	[0,0]

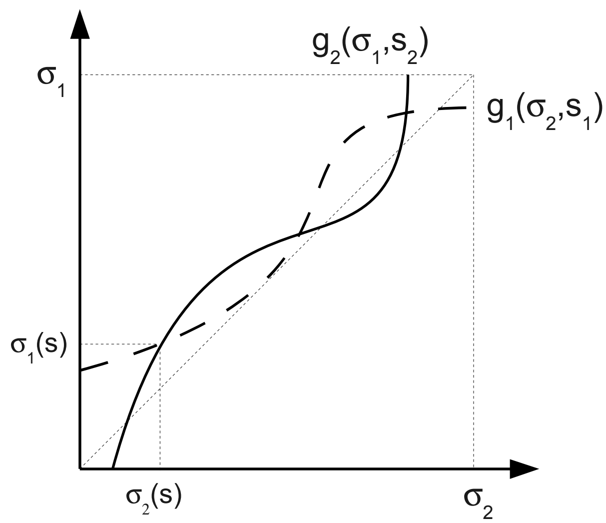
Notes: The table shows confidence intervals in the integers for $Z(g)$ for the 12 largest MSAs of the United States, ordered by population, where g is estimated by quantile regression of the change in the non hispanic, white population over a decade, divided by initial total population, on the initial minority share for the quantiles .2,.5 and .8. Regression bandwidth τ is $n^{-.2}$, σ is chosen as .05 times the maximal change. Confidence sets are based on t-statistics using bootstrapped bias and standard errors.

Table 2.3: Montecarlo rejection probabilities

n	τ	**r**	$\widehat{P}(\zeta > z_\alpha)$	$\widehat{P}(\zeta < -z_\alpha)$
400	0.065	0.179	0.05	0.01
800	0.059	0.194	0.03	0.02
1600	0.055	0.231	0.02	0.01
3200	0.052	0.290	0.02	0.01
400	0.065	0.268	0.03	0.02
800	0.059	0.292	0.01	0.02
1600	0.055	0.347	0.01	0.01
3200	0.052	0.434	0.01	0.02

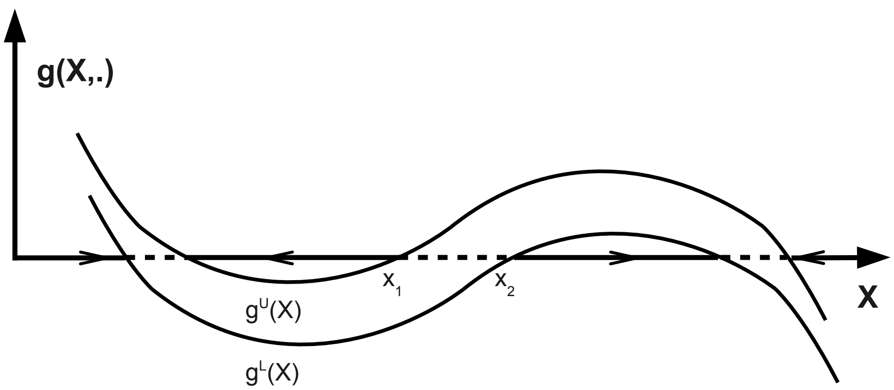
Notes: This table shows the frequency of rejection of the null under a test of asymptotic level 5%, for the sequences of Monte Carlo experiments described in appendix A. The g are estimated by mean regression, the errors are uniformly distributed, and the first four experiments are generated using g^1 with one root, the next four using g^2 with three roots. The columns show in turn sample size, regression bandwidth, error standard deviation, and the rejection probabilities of one-sided tests.

Figure 2.1: Response functions and Bayesian Nash Equilibria



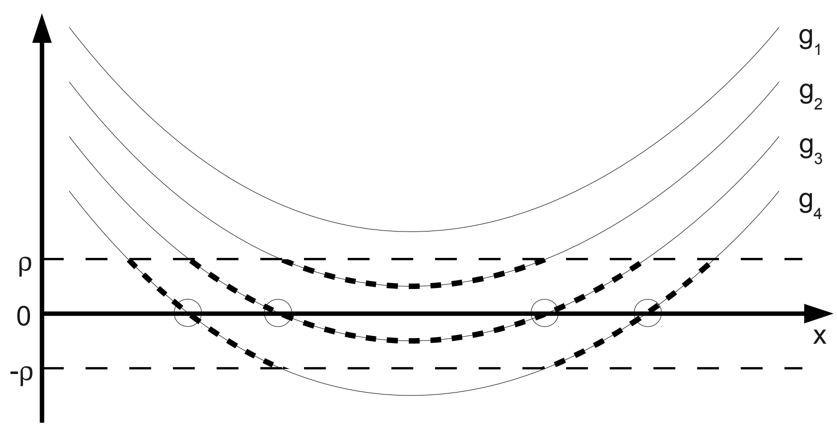
Notes: This figure illustrates the two player, two action static game of incomplete information discussed in section 2.3.4. The functions g_i are the (average) best response functions, Bayesian Nash Equilibrium requires $g(\bar{\sigma}_1, s) := g_1(g_2(\bar{\sigma}_1, s_2), s_1) - \bar{\sigma}_1 = 0$, and we observe one equilibrium $(\sigma_1(s), \sigma_2(s))$ in the data. In this graph, there are two further equilibria which are not directly observable.

Figure 2.2: QUALITATIVE DYNAMICS OF STOCHASTIC DIFFERENCE EQUATIONS



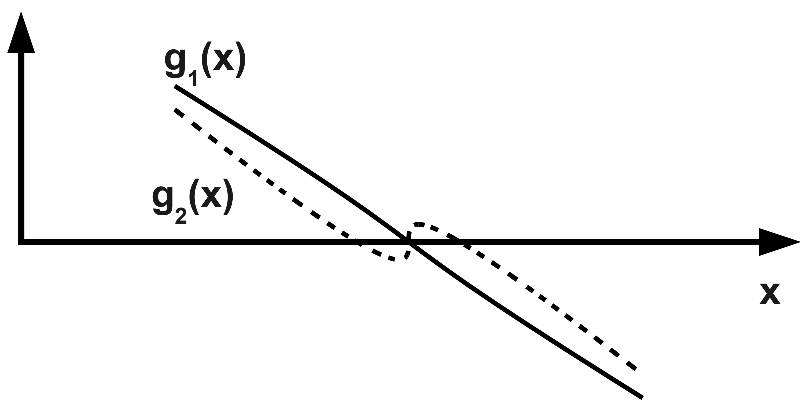
Notes: This figure illustrates proposition 11, where g^U is the upper envelope of $g(., \epsilon_s)$ for $s \leq t$, and g^L is the lower envelope. In the graph, equilibrium regions correspond to the dashed segments of the X axis, the basin of attraction of the lower equilibrium region is given by $(-\infty, x_1]$, and the basin of attraction of the upper equilibrium region is $[x_2, \infty)$.

Figure 2.3: Z AND Z_ρ



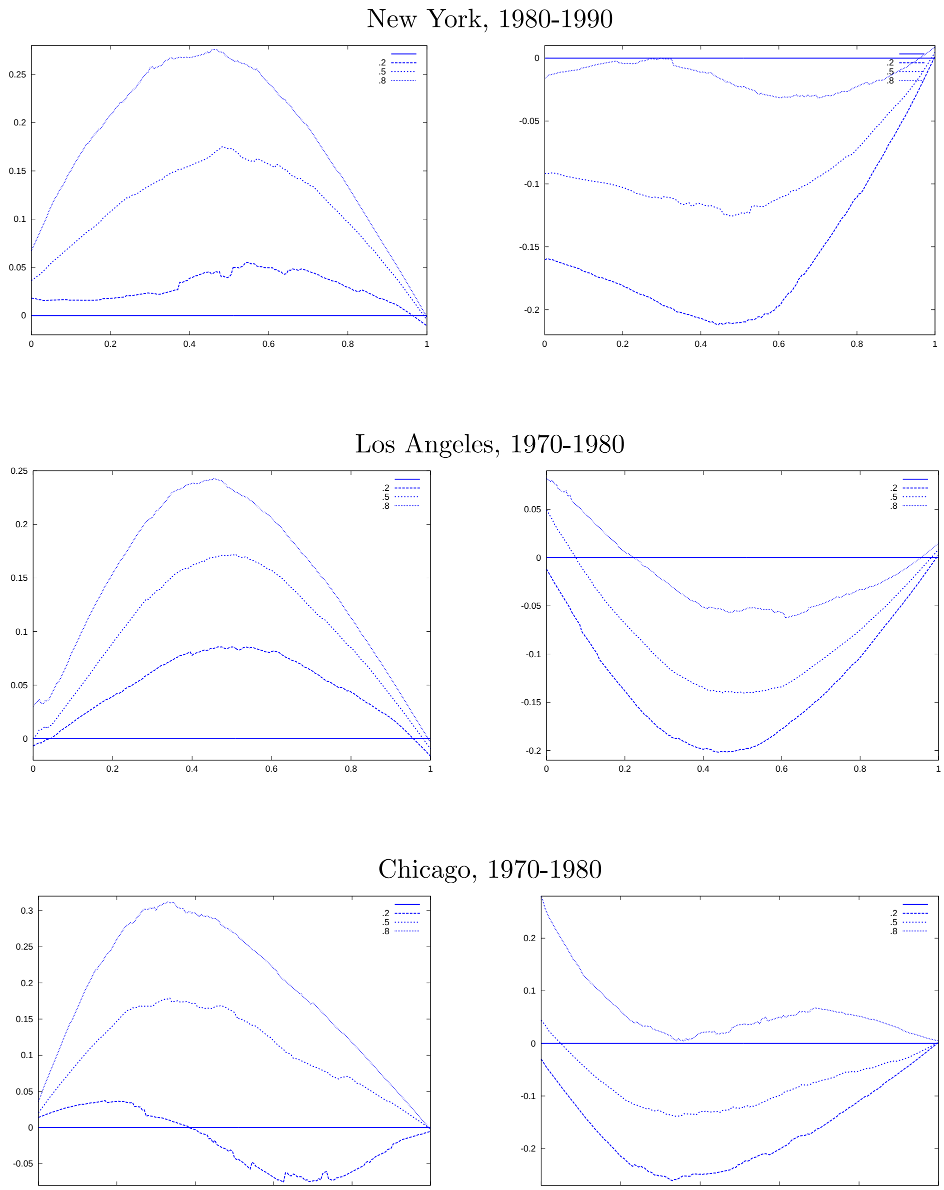
Notes: This figure illustrates the relationship between Z and Z_ρ. For the functions g depicted, $Z(g_1) = Z_\rho(g_1) = 0$, $Z(g_2) = 0 < Z_\rho(g_2) < 1$, $Z(g_3) = 2 > Z_\rho(g_3) > 1$, and $Z(g_4) = Z_\rho(g_4) = 2$.

Figure 2.4: On the importance of wiggles



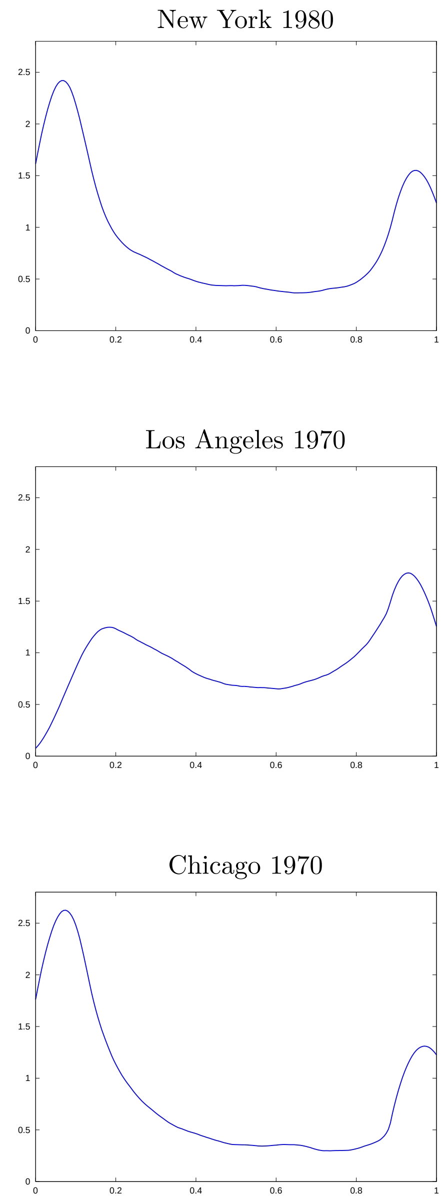
Notes: This figure illustrates how functions that are uniformly close can have different numbers of roots.

Figure 2.5: Quantile regressions of the change in minority share and of the change in white population on initial minority share



Notes: These graphs show local linear quantile regressions of the change in minority share (left column) and of the change in white population relative to initial population (right column) on initial minority share for the quantiles .2, .5 and .8. The graphs do not show confidence bands.

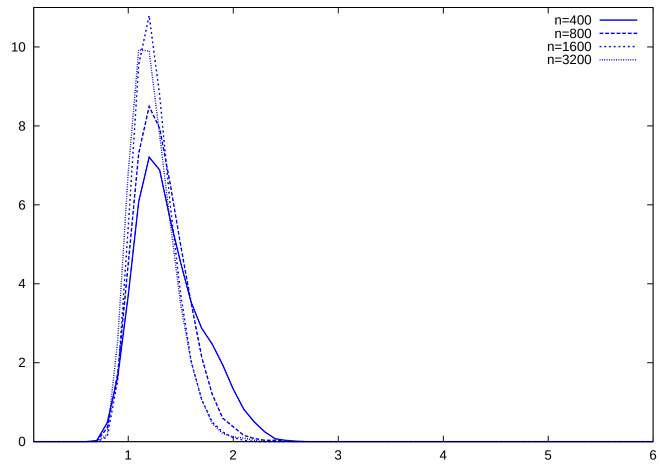
Figure 2.6: Density of minority share across neighborhoods



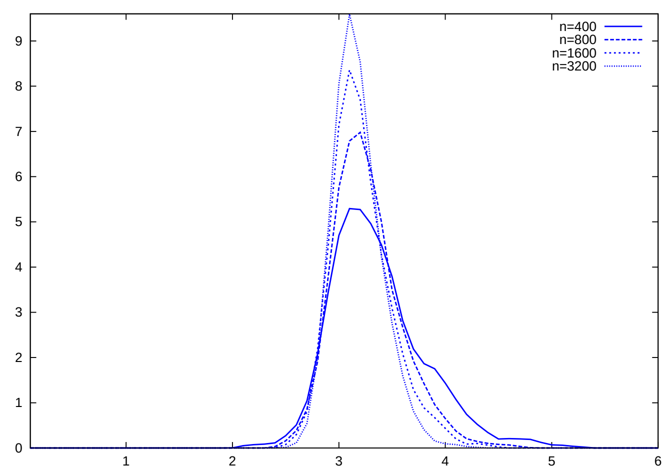
Notes: These graphs show kernel density estimates of the distribution of minority share across neighborhoods.

Figure 2.7: Density of $\widehat{Z}$ in Monte Carlo experiments

$g^1(x) = 0.5 - x,\ Z(g^1) = 1$



$g^2(x) = 0.5 - 5x + 12x^2 - 8x^3,\ Z(g^2) = 3$



Notes: This figure shows density plots of $\widehat{Z}$ from Monte Carlo experiments with uniform errors and g identified by median regression, as described in appendix A. The upper graph shows the distribution from four experiments with increasing samplesize n and correspondingly growing variance of the residual γ, where the true parameter Z equals one. The same holds for the lower graph, except that $Z = 3$.

Chapter 3

A nonparametric test for path dependence in discrete panel data

Abstract

This paper proposes a test for path dependence in discrete panel data based on a characterization of stochastic processes that are mixtures of Markov Chains. This test is applied to European Community Household Panel data on employment histories. The data allow to reject the null of no path dependence in all subsamples considered.

3.1 Introduction

Path dependence is of potential relevance in many areas of economics and the social sciences more generally. Path dependence here is understood to signify a causal impact of past states of some system on the future of that system, holding constant the present state. For instance, the employment history of an individual i might have a causal impact on that individual's chance of finding a job, given present unemployment. This is suggested by the empirical observation that past employment status $Y_{i,0}$ is predictive of future status $Y_{i,2}$, conditional on present status $Y_{i,1}$, in panel data on individual employment histories. However, if there is unobserved and exogenous heterogeneity across individuals that is serially dependent, and influences employment prospects, a similar implication for observable data follows.

Several different approaches can be taken to identify the nature of path dependence in the presence of unobserved heterogeneity. Experimental variation of initial $Y_{i,0}$ identifies path dependence as the excess causal impact of $Y_{i,0}$ on $Y_{i,2}$, beyond the effect mediated through $Y_{i,1}$. The latter is identified by compounding the effect of $Y_{i,0}$ on $Y_{i,1}$ and the effect of $Y_{i,1}$ on $Y_{i,2}$.

Functional form assumptions underly popular models of panel data as well as duration data. For instance, additive separability of heterogeneity is required in fixed effects models (see Chamberlain (1985)), and multiplicative separability of heterogeneity is imposed in the mixed proportional hazards model (see Heckman and Singer (1985), Van den Berg (2001)).

Without either experimental variation or functional form restrictions, models with arbitrary unobserved heterogeneity but no path dependence are still testable. In the case of spell durations, Heckman, Robb, and Walker (1990) devise tests based on characterizations of mixtures of exponential distributions. In the case of discrete panel data, Lee (1987) discusses restrictions on the coefficients of log linear probability models implied by mixture assumptions. The present paper is based on a characterization of mixtures of Markov Chains, proven by Diaconis and Freedman (1980).

3.2 The test for path dependence

The time path $(Y_{i,t})$ of an individual's status is described by a Markov chain, conditional on time invariant individual specific heterogeneity α_i, if two assumptions hold: First, the conditional distribution of future status given the individual's history and time invariant exogenous characteristics does not depend on the individual's history: $P(Y_{i,t+1}|\alpha_i, Y_{i,t}, Y_{i,t-1}, \ldots) = P(Y_{i,t+1}|\alpha_i, Y_{i,t})$. This is implied by the absence of both path dependence and time varying heterogeneity. Second, this conditional distribution does not depend on time t. This paper proposes a test for the hypothesis that individuals' histories follow a Markov chain, conditional on time invariant individual specific heterogeneity. This implies that the population distribution of histories can be represented as a mixture of Markov Chains.

Throughout we consider discrete panel data with finite support, $Y_{i,t} \in \{y^1, \ldots, y^m\}$. The event $\{Y_t = \sigma_t : t = 0, \ldots, T\}$ is denoted A_σ. The null hypothesis for which a test statistic is developed is the hypothesis that the data are generated from a mixture of Markov Chains:

Definition 12 (Mixture of Markov Chains)**.**
A process (Y_t) is called a mixture of Markov Chains if its law can be represented by

$$P(A_\sigma) = \int_{\mathscr{P}} \prod_{t=0}^{T-1} p(\sigma_t, \sigma_{t+1}) \mu(dp)$$

for some μ on the set of stochastic matrices $\mathscr{P}$, where w.l.o.g. $y_0 = 1$.

In this definition, $p(\sigma_t, \sigma_{t+1}) = P(Y_{i,t+1} = \sigma_{t+1}|\alpha_i, Y_{i,t} = \sigma_t)$. Conditional on α_i (that is, p), the probability of a given sequence $(y_0, \ldots, y_T)$ is the product of the probabilities of transitions from y_t to y_{t+1}, where these transition probabilities are statistically independent and constant over time. Given the initial state, the probability of such a sequence only depends on the number of transitions between any pair of states. It does not depend on the order of these transitions. Two sequences with the same initial state and number of transitions have the same probability. This equality is preserved under mixing. This motivates the following definition.

Definition 13 (Partial Exchangeability)**.**
Two finite sequences of states, σ and τ, are called equivalent if they start with the same state and they have the same number of transition counts from p to q for every pair of states p and q, that is they contain the ordered tuple pq the same number of times.
A process is called partially exchangeable, iff for all equivalent strings σ and τ, $P(A_\sigma) = P(A_\tau)$.

Consider the sequences 1011 and 1101. Partial exchangeability implies those two to be equally likely. If there was negative duration dependence in state 1, the second sequence might be less likely.

By the above argument, any process that is a mixture of Markov chains is partially exchangeable. That the reverse also holds true was proven by Diaconis and Freedman (1980), in an extension of the classic de Finetti's theorem.

Definition 14 (Recurrence)**.**
A process (Y_t) is called recurrent, if it returns with probability one to its initial state.

Diaconis and Freedman (1980) prove:

Theorem 6. *Let (Y_t) be recurrent. Then it is partially exchangeable iff it is a mixture of Markov chains.*

For a balanced panel, testing partial exchangeability amounts to testing equality restrictions on the multinomial distribution of state *sequences* in the population[1]. This can be done, in principle, using a generalized likelihood ratio test for equality restrictions on a multinomial distribution:

$$X^2 := 2\sum_{\sigma} N_\sigma \log\left(\frac{N_\sigma}{np_\sigma}\right) \xrightarrow{d} \chi^2_k$$

under partial exchangeability as $n \to \infty$, where n is the number of cross-sectional units, σ is an index ranging over all possible sequences, N_σ is the number of observations of type σ, p_σ are the maximum likelihood probabilities of sequences σ subject to the equality restrictions implied by partial exchangeability, and k are the number of linearly independent restrictions. The null can be rejected for large test statistics.

The number of restrictions k implied by partial exchangeability is shown in table 3.1. As can be seen, greater length of the panel increases the ratio of restrictions to possible sequences, k/m^{T+1}. Many states (large m) might be problematic since the number of possible sequences, m^{T+1}, explodes, thus making the probability of observing any particular equivalence class low.

[1] The assumption of recurrence in theorem 6 is only needed for the implication from partial exchangeability to representability as a mixture of Markov Chains. Hence rejection of partial exchangeability implies rejection of a mixture of Markov Chains even without recurrence.

	states				
periods	2	3	4	5	6
3	0	0	0	0	0
	(8)	(27)	(64)	(125)	(216)
4	2	6	12	20	49
	(16)	(81)	(256)	(625)	(1296)
5	10	57	168	370	2925
	(32)	(243)	(1024)	(3125)	(7776)
6	32	315	1368	4742	29720
	(64)	(729)	(4096)	(15625)	(46656)
7	84	1347	8492	44247	234602
	(128)	(2187)	(16384)	(78125)	(279936)
8	198	4983	44660	304587	1572245
	(256)	(6561)	(65536)	(390625)	(1679616)
9	438	16899	211124	1759026	9829528
	(512)	(19683)	(262144)	(1953125)	(10077696)
10	932	54387	932168	9353166	59932614
	(1024)	(59049)	(1048576)	(9765625)	(60466176)

Table 3.1: Number of linearly independent restrictions implied by partial exchangeability, in brackets number of different possible sequences

The asymptotic χ^2 approximation might fail in practice for two related reasons. First, some equivalence classes have actual probability 0. Second, some equivalence classes have very few observations. According to van der Vaart (1998) chapter 17, the χ^2 approximation under the null is "good" if there are, in expectation, at least 5 observations per possible sequence σ.

The following modification of the test is asymptotically valid and gives significant finite sample improvements: Count the number of observed sequences falling into each equivalence class, and discard all classes that contain less than "5 times the number of cells in the class" observations. Calculate the generalized likelihood ratio test statistic of the restrictions on this subsample. Reject the null if the statistic exceeds the critical value of a χ^2 distribution with degrees of freedom corresponding to the number of implied restrictions in the classes retained in the sample.

The asymptotic validity of this approach can be seen as follows: Condition on the distribution across equivalence classes. Calculate the generalized likelihood ratio test statistic for the null of a uniform distribution within each equivalence class. Note that for each of these test statistics standard χ^2 asymptotics apply, conditional on the distribution across classes. Note, finally, that these statistics are conditionally independent across equivalence classes, and that the sum of independent χ^2 variables is χ^2 itself.

3.3 Application to employment data

We shall now apply this test to panel data on employment histories in Western European countries. The data set used is the ECHP household panel for the years 1995 to 2001. Individuals are coded to be in one of three states, employed, unemployed or unobserved. After discarding all individuals where first period employment status is unobserved, we get 156060 sequences of 7 periods (years) and 3 states.

The results of applying the modified χ^2 test for partial exchangeability to these data are shown in table 3.2. In all cases but the UK we get p-values far below 1 percent, and even for the UK we are below 5 percent. Only a small fraction of the observations have to be discarded for the modified test. The null of a mixture of Markov Chains can be rejected in all subsamples considered.

3.4 Discussion and Conclusion

We have tested and rejected the null that individual employment histories are generated from a mixture of Markov Chains. This null, and the notion of path dependence more generally, are relative to the coding of status. If the causal effect of $Y_{i,0}$ on $Y_{i,2}$ is mediated through $Y_{i,1}$ and $Z_{i,1}$, including Z in the coding of status eliminates path dependence. In particular, even if data are generated from a Markov process, aggregation of states leads to violation of the Markovian property and hence of partial exchangeability.

The null also implies time homogeneity. This cannot be relaxed fully, since any distribution of sequences can be generated from a mixture of time inhomogeneous processes without path dependence. We could allow for aggregate structural breaks in an extension of the test. For instance, choose a breakpoint and calculate the previous test statistic for either part of the time window, take the sum and reject for the appropriate critical value of a χ^2 distribution with degrees of freedom equal to the sum of the number of restrictions from both parts. In another generalization, one can test for higher order Markovian behavior conditional on time invariant heterogeneity. By redefining states as $Z_t = (Y_t, Y_{t-1})$ for instance, second order Markov behavior of Y is equivalent to first order Markov behavior of Z. Applying either extension to the ECHP data again allows to reject the null hypotheses.

To conclude, it should be emphasized that the test proposed is a complement rather than a substitute for inference procedures relying on stronger assumptions. It is not able to disentangle the nature of path dependence (duration dependence in which state?), nor can we allow for time varying exogenous covariates, as D'Addio and Honoré (2006) do. It is attractive, however, because it requires neither functional form restrictions nor exogenous sources of variation.

χ^2 TEST FOR PARTIAL EXCHANGEABILITY

	(1) observations	(2) observations after discarding	(3) unrestricted likelihood	(4) restricted likelihood	(5) χ^2 statistic	(6) df after discarding	(7) p-value
Germany	21,544	21,040	-60,696	-60,890	387.45	97	0.0000
Denmark	5,503	5,043	-15,511	-15,636	248.26	37	0.0000
The Netherlands	9,151	8,663	-27,148	-27,276	256.18	52	0.0000
Belgium	6,454	6,093	-17,834	-17,887	107.40	31	0.0000
Luxembourg	8,754	8,444	-23,253	-23,283	59.89	19	0.0000
France	13,306	12,710	-41,317	-41,453	271.10	66	0.0000
United-Kingdom	17,211	16,731	-45,570	-45,602	65.20	48	0.0498
Ireland	8,531	8,127	-26,102	-26,137	71.34	29	0.0000
Italy	17,780	17,172	-56,928	-57,100	344.09	88	0.0000
Greece	12,271	11,741	-38,270	-38,522	504.23	73	0.0000
Spain	16,263	15,528	-55,144	-55,435	582.11	122	0.0000
Portugal	11,858	11,243	-33,905	-33,984	158.36	59	0.0000
Austria	7,437	7,006	-21,442	-21,545	207.19	35	0.0000
male, all countries	75,081	74,132	-243,400	-244,180	1576.20	194	0.0000
female, all countries	80,982	80,185	-273,530	-274,390	1727.90	225	0.0000

Table 3.2: Column (1) shows the number of observations in the subsample of each row, column (2) shows the number of remaining observations after discarding all sequences that lie in equivalence classes with too few observations, column (3) shows the fully unrestricted multinomial likelihood, column (4) shows the maximal likelihood under the restriction of partial exchangeability, column (5) shows the corresponding χ^2 statistic, $\chi^2 = 2\sum_{\sigma} N_{\sigma} \log\left(\frac{N_{\sigma}}{np_{\sigma}}\right)$ where the p_{σ} are the restricted estimates, column (6) contains the number of equality restrictions implied by partial exchangeability in the subsample after discarding, and column (7) contains the p-values from a χ^2 distribution corresponding to this number of restrictions.

Chapter 4

Identification in Triangular Systems using Control Functions

Abstract

This paper discusses identification in nonparametric, continuous triangular systems. It provides conditions which are both necessary and sufficient for the existence of control functions satisfying conditional independence and support requirements. Confirming a commonly noticed pattern, these conditions restrict the admissible dimensionality of unobserved heterogeneity in the first stage structural relation, or more generally the dimensionality of the family of conditional distributions of second stage heterogeneity given explanatory variables and instruments. These conditions imply that no such control function exists without assumptions that seem hard to justify in most applications. In particular, none exists in the context of a generic random coefficient model.

4.1 Introduction

In a recent paper, Imbens and Newey (2009) develop nonparametric identification results in triangular systems for models with a first stage that is monotonic in unobservables. Based on these results, they construct inference procedures. In related work, Imbens (2007) surveys control functions in triangular systems more generally. This note elaborates on their analysis by providing conditions which are both necessary and sufficient for the existence of control functions. The nonparametric, continuous triangular system setup considered is given by:

$$Y = g(X, \epsilon) \tag{4.1}$$
$$X = h(Z, \eta) \tag{4.2}$$

where we assume

$$Z \perp (\epsilon, \eta) \tag{4.3}$$

with Z, X, Y each continuously distributed in $\mathbb{R}$. This is the general setup in which the application of instrumental variable methods is usually discussed, where Z is the exogenous instrument, X is the treatment and Y is the outcome variable. The object of interest is the structural function g.

Recent contributions to the literature have generalized identification in parametric (linear) triangular models to nonparametric setups. The idea of nonparametric identification using the control function approach is to find a function C of X and Z such that, for $V = C(X, Z)$,

$$X \perp \epsilon | V. \tag{4.4}$$

If C is a one-dimensional, strictly monotonic function of both X and Z, then there exists a one to one mapping between (X, Z), (X, V) and (Z, V). Existence of an invertible mapping between (X, V) and (Z, V) implies that conditional independence 4.4 is equivalent to

$$Z \perp \epsilon | V. \tag{4.5}$$

In contrast to the literature providing mostly sufficient conditions for identification, this note provides conditions which are both necessary and sufficient for the existence of control functions that satisfy conditional independence and support requirements. These conditions impose restrictions on the dimensionality of unobserved heterogeneity. While the importance of dimensionality restrictions has been noted repeatedly, among others by Imbens (2007), this note is, to the best of our knowledge, the first to formally show that they are both necessary and sufficient.

The central object of interest in the control function literature is the average structural function (ASF). Let E_ϵ denote the expectation taken over the marginal distribution of ϵ, and similarly for E_V. The ASF was defined by Blundell and Powell (2003) as $ASF(x) := E_\epsilon[g(x, \epsilon)]$. Given a control function, the ASF is identified by

$$ASF(x) = E_V[E[g(X, \epsilon)|V, X = x]] = E_V[E[Y|V, X = x]]. \tag{4.6}$$

The first equality requires conditional independence 4.4. Identification of the conditional expectation given X, V requires full support of V given X. Under the same conditions, one can identify the quantile structural function (QSF) $g^\tau(x)$, which is given by the τth quantile of $g(x, \epsilon)$ over the marginal distribution of ϵ, as well as functions defined by more general conditional moment restrictions.[1]

Various choices for C have been proposed in the literature. Newey, Powell, and Vella

[1] Suppose the object of interest is $\widetilde{g}(x) = argmin_{\breve{g}} E_\epsilon[\rho(g(x, \epsilon), \breve{g})]$ for a loss function $\rho(Y, \widetilde{g})$. The function $\widetilde{g}$ is identified under the same conditions as the ASF. To show this, replace Y with ρ in equation 4.6.

(1999) suggest using the residual of a first stage mean regression:

$$V = C(X, Z) = X - E[X|Z]. \tag{4.7}$$

This is justified by an additive model for h, i.e. $h(Z, \eta) = \tilde{h}(Z) + \eta$ and $V = \eta$. Such additivity would not hold, for example, in models with heteroskedastic residuals.

Imbens and Newey (2009) propose, more generally, to use the conditional cumulative distribution function F of X given Z:

$$V = C(X, Z) = F[X|Z]. \tag{4.8}$$

This is justified by a first stage h that is strictly monotonic in a one-dimensional η, implying that $V = F(\eta)$.

In either of these cases, the fact that V is a function of η alone immediately implies that conditional independence 4.5 holds. Under strict monotonicity of the conditional expectation or distribution function in Z, this in turn implies conditional independence 4.4.

The next section provides an example of failure of conditional independence 4.4 using the control function 4.8. Sections 4.3.1 and 4.3.2 provide the central results of this note. Section 4.3.1 gives a condition which is both necessary and sufficient for the existence of a control function that is constant in the instrument given unobserved heterogeneity. Section 4.3.2 does the same for the more general case of control functions satisfying conditional independence 4.4. Section 4.3.2 also shows that no control function can exist in the case of the random coefficient model. Section 4.3.3 states extensions to the case of higher dimensional X and Z. Section 4.4 concludes. All proofs are relegated to appendix 4.A.

4.2 Counterexample

Consider the following random coefficient model:

$$X = \eta_1 + \eta_2 Z = \eta \cdot (1, Z) \tag{4.9}$$

$$(\eta_1, \eta_2, \epsilon) \sim N(\mu, \Sigma) \tag{4.10}$$

$$Z \perp (\eta, \epsilon) \tag{4.11}$$

This model will serve as a counterexample for identification attempts using control functions. Imbens (2007) subsection 5.2 uses the same example of failure of the control function of Imbens and Newey (2009). As will be shown in section 4.3, no control function exists in this specification because first-stage heterogeneity η is more than one-dimensional.

For an economic example of this[2], suppose the following: We are interested in the production function relating output of firms to a single variable input, e.g., labor l. Production technology is Cobb-Douglas, i.e., log output is $y_i = A_i + \alpha_i l_i$, where A and α are unobserved heterogeneity in firm technology or endowment with other factors. Prices for the output good vary exogenously, wages are constant, and firms maximize profits. Then both the first stage relationship, that is, firm specific labor demand as a function of prices, and the second stage production function exhibit a linear random coefficient structure.

A similar example is the problem of estimating returns to schooling when returns are heterogeneous, schooling depends on returns, and we observe an independent cost variable affecting school choice that can serve as an instrument.

Now we will show why the control function proposed by Imbens and Newey (2009), $V = F(X|Z)$, fails in this random coefficient model. For jointly normally distributed variables, the conditional expectation is given by the best linear predictor. Hence we get, by ordinary least squares regression of ϵ on X given Z,

$$E[\epsilon|X,Z] = \mu_\epsilon + (X - E[X|Z]) \cdot \frac{Cov(X,\epsilon|Z)}{Var(X|Z)}.$$

The assumptions imply that X and ϵ are jointly normal given Z, with

$$\begin{aligned} Cov(X,\epsilon|Z) &= \Sigma_{\eta_1,\epsilon} + Z\Sigma_{\eta_2,\epsilon} \\ Var(X|Z) &= \Sigma_{\eta_1,\eta_1} + 2Z\Sigma_{\eta_1,\eta_2} + Z^2\Sigma_{\eta_2,\eta_2} \\ E[X|Z] &= \mu_{\eta_1} + Z\mu_{\eta_2}. \end{aligned}$$

This gives

$$E[\epsilon|X,Z] = \mu_\epsilon + (X - \mu_{\eta_1} - Z\mu_{\eta_2}) \cdot \frac{\Sigma_{\eta_1,\epsilon} + Z\Sigma_{\eta_2,\epsilon}}{\Sigma_{\eta_1,\eta_1} + 2Z\Sigma_{\eta_1,\eta_2} + Z^2\Sigma_{\eta_2,\eta_2}}. \tag{4.12}$$

The control function proposed by Imbens and Newey (2009),

$$V = F(X|Z) = \Phi\left(\frac{(X - \mu_{\eta_1} - Z\mu_{\eta_2})}{\sqrt{Var(X|Z)}}\right), \tag{4.13}$$

is monotonic in X. If the support of Z is restricted to an appropriate range[3], it is also monotonic in Z. Hence, for at least a subrange of V, the following equalities hold:

$$E[\epsilon|V,X] = E[\epsilon|V,Z] = E[\epsilon|X,Z] =$$

[2] I thank Bryan Graham for this motivation.

[3] If $\mu_{\eta_2} > 0$, then $Z \leq -\Sigma_{\eta_1,\eta_2}/\Sigma_{\eta_2,\eta_2}$ is sufficient, though not necessary.

$$= \mu_\epsilon + \Phi^{-1}(V) \frac{\Sigma_{\eta_1,\epsilon} + Z\Sigma_{\eta_2,\epsilon}}{\sqrt{\Sigma_{\eta_1,\eta_1} + 2Z\Sigma_{\eta_1,\eta_2} + Z^2\Sigma_{\eta_2,\eta_2}}}. \tag{4.14}$$

From this it follows that conditional independence 4.4 is violated: By invertibility of C in both X and Z, conditional independence 4.4 and 4.5 are equivalent. Conditional independence 4.5 requires conditional mean independence, i.e., that $E[\epsilon|V, Z]$ is constant in Z given V. By equation 4.14, this holds if and only if $\Sigma_{\eta_2,\eta_2} = 0$, that is, if the slope of the first stage is constant. If the slope η_2 has positive variance, conditional independence 4.4 does not hold. A similar argument can be made about the conditional variance of ϵ given V, X.

4.3 Characterization of models in which control functions exist

The next three subsections will present the general results characterizing triangular systems for which control functions exist. Subsection 4.3.1 shows that control functions that do not depend on Z given η exist if and only if η is one-dimensional. This requirement is in particular violated by the random coefficient model of the previous section. Subsection 4.3.2 shows that control functions that satisfy conditional independence 4.4 exist if and only if the family of conditional distributions $P(\epsilon|X, Z)$ is one-dimensional. This dimensionality requirement is again violated by the random coefficient model. Section 4.3.3 finally generalizes the previous results to setups with higher dimensional X and Z.

4.3.1 Control functions that do not depend on Z given η

The following proposition covers all variants of the control function approach that we are aware of, in particular Newey, Powell, and Vella (1999) and Imbens and Newey (2009):

Proposition 12. *If $V = C(h(Z, \eta), Z)$ does not depend on Z given η, then conditional independence 4.5 holds.*

As mentioned in the introduction, conditional independence 4.5 is equivalent to 4.4 if there exists a mapping $(Z, V) \to (X, V)$, which is true if C is invertible. Conditional independence 4.4 is necessary for the use of V as a control. The condition of proposition 12, however, comes at the price of restricting the first stage structural function, h:

Proposition 13. *If $V = C(h(Z, \eta), Z)$ does not depend on Z given η for a $C(X, Z)$ that is smooth and almost surely invertible in X, then $\{h(\cdot, \eta)\}$ is a one-dimensional family of functions in Z*

Remark: Identification of average structural functions or quantile structural functions for a given $X = x$ requires, in addition to conditional independence 4.4, that V has full support given $X = x$. In other words, the range of $C(X, Z)$ must be independent of X.

Remark: Assume that almost surely $h(Z, \eta_1) \neq h(Z, \eta_2)$ for independent draws Z, η_1, η_2 from the respective distributions of Z and η. Then the family of functions $\{h(., \eta)\}$ is one-dimensional if and only if it is possible to predict the counterfactual X under manipulation of Z from knowledge of X and Z. This possibility is a much stronger requirement than the possibility of identifying the ASF or QSF for the first stage relationship, which follows immediately from exogeneity of Z. The counterfactual outcome setting $Z = z_0$, $h(z_0, \eta)$, is used as a control function in the proof of proposition 14 below.

Remark: If invertibility is dropped from the assumptions of proposition 13, one-dimensionality of the family $\{h(\cdot, \eta)\}$ does not necessarily follow, but neither does conditional independence 4.4. For example, if $C = const.$, then conditional independence 4.5 holds, but 4.4 does not necessarily hold.

The reverse of proposition 13 is also true:

Proposition 14. *If $\{h(., \eta)\}$ is a one-dimensional family of functions in Z and almost surely $h(Z, \eta_1) \neq h(Z, \eta_2)$ for independent draws Z, η_1, η_2 from the respective distributions of Z and η, then there exists a control function $V = C(h(Z, \eta), Z)$ which does not depend on Z given η.*

Remark: If the family $\{h(., \eta)\}$ is not only one-dimensional but also monotonic in unobserved heterogeneity, that is

$$h(z_1, \eta_1) > h(z_1, \eta_2) \Leftrightarrow h(z_2, \eta_1) > h(z_2, \eta_2) \; \forall \; z_1, z_2, \eta_1, \eta_2, \tag{4.15}$$

then $C(X, Z) = F(X|Z)$ is the same control function as the one constructed in the proof of proposition 14, in that there is an invertible mapping between the two. If monotonicity fails, however, $C(X, Z) = F(X|Z)$ cannot satisfy the sufficient condition of proposition 12.

Remark: It follows from proposition 13 that in the random coefficient example of section 4.2, no control function satisfying the sufficient condition of proposition 12 and invertibility in X can exist. The family of functions

$$h(Z, \eta_1, \eta_2) = \eta_1 + \eta_2 Z \tag{4.16}$$

assumed in the random coefficient model is two-dimensional, which implies that we cannot predict the counterfactual X under a manipulation setting $Z = z$, $h(z, \eta)$, for a given observational unit from X and Z alone.

4.3.2 Control functions satisfying conditional independence

Next we will consider the more general case of control functions satisfying conditional independence 4.4, which is required to identify $E_{\epsilon|V}[g(x,\epsilon)|V]$ by $E[Y|X=x,V]$.

Proposition 15. *There exists a control function $V = C(X,Z)$ such that conditional independence 4.4 holds and which is invertible in Z if and only if $P(\epsilon|X,Z)$ is an at most one-dimensional family of distributions that is not constant in Z if it is not constant.*

Remark: If C is not invertible in Z the following situation is theoretically possible: The family of conditional distributions $p(\epsilon|Z,X)$ is two-dimensional. The conditional support of (X,Z) given V, $\mathbb{XZ}(V)$, is comprised of a discrete set of points given X. Hence $p(\epsilon|X,V)$ is a mixture of $p(\epsilon|Z,X)$ over a discrete set of points Z. None of the components of this mixture is constant as X varies and Z covaries to remain within the manifold $\mathbb{XZ}(V)$. Nevertheless, the changes in the components cancel exactly, implying that $p(\epsilon|X,V)$ is constant in X.

Intuitively, such canceling seems a highly non-generic phenomenon and of little practical relevance. We do not have results, however, precluding this possibility in the absence of invertibility of C in Z.

Remark: The theorem only characterizes conditions for the existence of a control function. It does not give conditions for identifiability of C itself.

In the random coefficient model of section 4.2, the necessary condition of proposition 15 is not fulfilled in general. We have

$$\epsilon|X,Z \sim N\left(\mu_\epsilon + (X - \mu_{\eta_1} - \mu_{\eta_2} Z)\frac{Cov(X,\epsilon|Z)}{Var(X|Z)}, Var(\epsilon) - \frac{Cov^2(X,\epsilon|Z)}{Var(X|Z)}\right), \tag{4.17}$$

which is a two-dimensional family as long as $Cov(X,\epsilon|Z)$ is not identical 0 and $\frac{Cov^2(X,\epsilon|Z)}{Var(X|Z)}$ is not constant in Z, i.e., so long as

$$\frac{(\Sigma_{\eta_1,\epsilon} + Z\Sigma_{\eta_2,\epsilon})^2}{\Sigma_{\eta_1,\eta_1} + 2Z\Sigma_{\eta_1,\eta_2} + Z^2\Sigma_{\eta_2,\eta_2}}$$

depends on Z. Since this is the case for generic Σ, the following corollary holds:

Corollary 6. *There exists no control function invertible in Z in the generic random coefficient model of section 4.2 such that conditional independence 4.4 holds.*

4.3.3 Higher dimensional X and Z

The results of the previous sections extend to the case of higher dimensional X and Z. In particular, if we allow $X \in \mathbf{R}^k$ and $Z \in \mathbf{R}^l$ with $l \geq k$, the following generalization of proposition 13 holds:

Proposition 16. *If $V = C(h(Z,\eta), Z)$ does not depend on Z given η for a $C(X,Z)$ that is smooth and almost surely invertible in X, then $\{h(.,\eta)\}$ is a k-dimensional family of functions in Z.*

The proof is analogous to the one-dimensional case. Similarly for proposition 14:

Proposition 17. *If $\{h(.,\eta)\}$ is a k-dimensional family of functions in Z and almost surely $h(Z,\eta_1) \neq h(Z,\eta_2)$ for independent draws Z, η_1, η_2 from the respective distributions of Z and η, then there exists a control function $V = C(h(Z,\eta), Z)$ which does not depend on Z given η.*

Finally, since none of the arguments leading to proposition 15 depended on the dimensionality of X or Y, we get the following generalization:

Proposition 18. *There exists a control function $V = C(X,Z)$ such that conditional independence 4.4 holds and which is invertible in Z if and only if $P(\epsilon|X,Z)$ is an at most l-dimensional family of distributions that is not constant in Z if it is not constant.*

4.4 Conclusion

This note characterizes triangular models for which control functions satisfying conditional independence and support requirements exist. These characterizations seem restrictive and will generally not be fulfilled. In particular, proposition 13 states that having a control function that is a function of unobserved heterogeneity η requires a one-dimensional first stage family of structural functions.

Examples of such one-dimensional families include (i) families that are monotonic in unobserved heterogeneity, as in Imbens and Newey (2009), (ii) models with $X = h(|Z-\eta|)$, which could describe the loss from missing an unknown target η, and (iii) multiplicative families of the form $X = h(Z)\cdot\eta$, where h is of non-constant sign. An economic example of (iii) is an income equation where X is income, $h(Z)$ is the (possibly negative) amount of some asset that an individual owns, and η is the rate of return. The characterizations proven in this paper show, however, that while such alternative families could be considered, no less restrictive family will allow construction of a control function. That is, there is no scope for generalization in conditions beyond Imbens and Newey (2009).

Appendix 4.A Proofs

Proof of Proposition 12: This is immediate from independence of Z and (η, ϵ). By assumption we can write V as a function of η, and therefore

$$Z|(V(\eta), \epsilon) \sim Z. \tag{4.18}$$

Hence, the conditional distribution of Z given $V(\eta), \epsilon$ does not depend on ϵ, implying conditional independence 4.5. $\square$

Proof of Proposition 13: Invertibility in X and smoothness of C implies that the range of $V = C(X, Z)$ is a one-dimensional, smooth manifold for a given Z. Since V is a function of η only, its range is independent of Z. Hence the range of V is a one-dimensional, smooth manifold. Invertibility and smoothness of C imply further that we can define a function $\tilde{h}$ such that $X = \tilde{h}(Z, V)$.

These assertions are true for many "reduced form" representations of the first stage such as regression residuals or conditional quantiles. However, the assumption that V does not depend on Z given η makes the first stage "structural" in the sense that we can write

$$h(Z, \eta) = \tilde{h}(Z, V(\eta)). \tag{4.19}$$

Because V is one-dimensional, this is a one-dimensional family of functions in Z. $\square$

Proof of Proposition 14: Since $\{h(., \eta)\}$ is a one-dimensional family of functions, we can assume without loss of generality that η has its support in $\mathbf{R}$. Pick a generic z_0 from the distribution of Z, and define $C(X, Z) = h(z_0, h^{-1}(Z, X))$, where the inverse is understood with respect to the η argument of h holding Z fixed. This inverse is well defined by the non-constancy of h in η and the one-dimensionality of η. By definition, $C(h(Z, \eta), Z) = h(z_0, \eta)$, which is a function of η alone. $\square$

Proof of Proposition 15: Consider the family of conditional distributions of ϵ given Z, X. This is an at most two-dimensional family, indexed by a parameter which shall be denoted $\theta(Z, X)$, that is, $p(\epsilon|Z, X) =: p(\epsilon, \theta(Z, X))$. The distribution of ϵ given X, V is in general a mixture over Z of $p(., \theta(Z, X))$ for Z such that $C(X, Z) = V$. If C is invertible in Z, no mixing takes place and this reduces to $p(., \theta(Z, X))$ for $Z = C^{-1}(X, V)$. In this case, conditional independence 4.4 is equivalent to constancy of $\theta(Z, X)$ on the manifold

$$\mathbb{XZ}(V) := \{(x, z) : C(x, z) = V, (x, z) \in supp(X, Z)\}. \tag{4.20}$$

Hence θ could be written as a function of C, which implies that the dimensionality of the range of θ is no higher than the dimensionality of the range of C. The range of C, however, cannot be of dimensionality larger than the dimension of Z if C has full range *given* X, as required for identification of the ASF and implied by the invertibility of C. This implies that the range of θ is at most one-dimensional. $\square$

Bibliography

Azariadis, C., and J. Stachurski (2005): "Poverty Traps," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf, vol. 1 of *Handbook of Economic Growth*, chap. 5. Elsevier.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2006): "Estimating static models of strategic interaction," *NBER working paper*.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A unified framework for measuring preferences for schools and neighborhoods," *Journal of Political Economy*, 115(4).

Becker, G. S., and K. M. Murphy (2000): *Social Economics: Market Behavior in a Social Environment*. Harvard University Press.

Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60(4), 889–917.

Black, S. E. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114(2), 577–599.

Blundell, R., and J. Powell (2003): "Endogeneity in nonparametric and semiparametric regression models," in *Advances in economics and econometrics: Theory and applications, eighth world congress*, vol. 2, pp. 655–679.

Bowles, S., S. N. Durlauf, and K. Hoff (2006): *Poverty traps*. Princeton University Press.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150(1), 41–55.

Bresnahan, T. F., and P. C. Reiss (1991): "Entry and competition in concentrated markets," *Journal of Political Economy*, 99(5), 977–1009.

Caetano, G. (2009): "Estimation of Parental Valuation of School Quality in the U.S.," *Dissertation, UC Berkeley*.

Card, D. (2001): "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 19(1), 22–64.

Card, D., A. Mas, and J. Rothstein (2008): "Tipping and the Dynamics of Segregation," *Quarterly Journal of Economics*, 123(1), 177–218.

Chamberlain, G. (1985): "Heterogeneity, omitted variable bias, and duration dependence," *Longitudinal analysis of labor market data*, pp. 3–38.

Chay, K. Y., and M. Greenstone (2005): "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 113(2), 376–424.

Chiappori, P. A., R. J. McCann, and L. P. Nesheim (2009): "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, pp. 1–38.

Currarini, S., M. O. Jackson, and P. Pin (2009): "An economic model of friendship: Homophily, minorities and segregation," *Econometrica*, 77(4), 1003–1045.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (2008): "Is the melting pot still hot? Explaining the resurgence of immigrant segregation," *The Review of Economics and Statistics*, 90(3), 478–497.

D'Addio, A. C., and B. E. Honoré (2006): "Duration dependence and timevarying variables in discrete time duration models," *Princeton University*.

Dasgupta, P., and D. Ray (1986): "Inequality as a determinant of malnutrition and unemployment: Theory," *The Economic Journal*, 96(384), 1011–1034.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2009): "Identification of Social Interactions through Partially Overlapping Peer Groups," .

Diaconis, P., and D. Freedman (1980): "de Finetti's theorem for Markov chains," *The Annals of Probability*, 8(1), 115–130.

Duflo, E., and E. Saez (2003): "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence From a Randomized Experiment," *Quarterly Journal of Economics*, 118(3), 815–842.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(S1), 60–109.

Ellison, G., and E. L. Glaeser (1999): "The geographic concentration of industry: Does natural advantage explain agglomeration?," *American Economic Review*, 89(2), 311–316.

Graham, B. (2008): "On the Identification of Neighborhood Externalities in the Presence of Endogenous Neighborhood Selection," *Workingpaper, UC Berkeley*.

Heckman, J., R. Matzkin, and L. Nesheim (2002): "Nonparametric estimation of nonadditive hedonic models," *Manuscript, Univ. Chicago*.

Heckman, J. J., R. Robb, and J. R. Walker (1990): "Testing the mixture of exponentials hypothesis and estimating the mixing distribution by the methods of moments," *Journal of the American Statistical Association*, 85(410), 582–589.

Heckman, J. J., and B. Singer (1985): "Social science duration analysis," *Longitudinal analysis of labor market data*, pp. 39–110.

Hoeffding, W., and H. Robbins (1994): "The central limit theorem for dependent random variables," *The collected works of Wassily Hoeffding*, p. 205.

Horowitz, J. L. (2001): "The Bootstrap," *Handbook of Econometrics*, 5, 3159–3228.

Horváth, L. (1991): "On L_p-norms of multivariate density estimators," *The Annals of Statistics*, 19(4), 1933–1949.

Imbens, G. W. (2007): "Nonadditive Models with Endogenous Regressors," *Econometric Society World Congress Series*, 75, 17–46.

Imbens, G. W., and J. D. Angrist (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62(2), 467–475.

Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

Imbens, G. W., and J. M. Wooldridge (2007): "What's new in econometrics? Weak instruments and many instruments," *NBER Lecture Notes 13, Summer 2007*.

Kasy, M. (2010): "Nonparametric inference on the number of equilibria, web appendix," https://sites.google.com/site/maxkasywp/Home/wps/Appendixtestingmulteq.pdf.

Katz, L., J. Kling, and J. Liebman (2007): "Experimental analysis of neighborhood effects," *Econometrica*, 75(1), 83–119.

Kling, J. R. (2001): "Interpreting instrumental variables estimates of the returns to schooling," *Journal of Business and Economic Statistics*, 19(3), 358–364.

Koenker, R. (2009): "Quantile Regression," http://www.econ.uiuc.edu/~roger/research/rq/rq.html, Accessed January 30, 2009.

Kong, E., O. Linton, and Y. Xia (2010): “Uniform bahadur representation for local polynomial estimates of M-regression and its application to the additive model,” *Econometric Theory*, pp. 1–36.

Krugman, P. (1991): “Increasing returns and economic geography,” *Journal of political economy*, 99(3), 483–499.

Lee, L. F. (1987): “Non-parametric testing of discrete panel data models,” *Journal of Econometrics*, 34(1-2), 147–177.

Manski, C. F. (1993): “Identification of endogenous social effects: The reflection problem,” *The Review of Economic Studies*, 60(3), 531–542.

Moffitt, R. A. (2001): “Policy Interventions, Low-Level Equilibria, and Social Interactions,” *Social Dynamics*, pp. 45–82.

Nesheim, L. (2001): “Equilibrium Sorting of Heterogeneous Consumers Across Locations: Theory and Empirical Implications,” *Dissertation, University of Chicago.*

Newey, W. (1994): “The asymptotic variance of semiparametric estimators,” *Econometrica: Journal of the Econometric Society*, 62(6), 1349–1382.

Newey, W. K., J. L. Powell, and F. Vella (1999): “Nonparametric Estimation of Triangular Simultaneous Equations Models,” *Econometrica*, 67(3), 565–603.

Pissarides, C. A. (2000): *Equilibrium unemployment theory*. The MIT press.

Quah, D. T. (1996): “Empirics for economic growth and convergence,” *European Economic Review*, 40(6), 1353–1375.

Rosen, S. (1974): “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,” *Journal of Political Economy*, 82(1), 34.

Schelling, T. (1971): “Dynamic Models of Segregation,” *Journal of Mathematical Sociology*, 1, 143–186.

Tiebout, C. M. (1956): “A Pure Theory of Local Expenditures,” *The Journal of Political Economy*, 64(5), 416.

Van den Berg, G. (2001): “Duration models: specification, identification and multiple durations,” *Handbook of econometrics*, pp. 3381–3460.

van der Vaart, A. (1998): *Asymptotic statistics*. Cambridge University Press.

Wheaton, W. C. (1990): “Vacancy, search, and prices in a housing market matching model,” *The Journal of Political Economy*, 98(6), 1270–1292.

Young, H. P. (2008): “Social norms,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume, vol. 2.