WORKING PAPER

Neighborhood Income Distributions

Yannis M. Ioannides
Department of Economics
Tufts University

Discussion Paper 2001 - 03

Department of Economics
Tufts University
Medford, MA 02155
(617) 627 – 3560
http://ase.tufts.edu/econ

NEIGHBORHOOD INCOME DISTRIBUTIONS

Yannis M. Ioannides[1]

Department of Economics

Tufts University

Medford, MA 02155

Email: Yannis.Ioannides@tufts.edu Voice: 617-627-3294 Fax: 617-627-3917

June 1, 2001

Abstract

This paper studies the distribution of income within neighborhoods and contrasts it with the national income distribution. It relies on a unique feature of the American Housing Survey, whose 1985, 1989 and 1993 waves of interviews provide data for small residential neighborhoods. These consist of a dwelling unit and up to ten of its nearest neighbors. Most previous work on neighborhoods has used information for much larger groupings of the population, such as census tracts. The paper employs a variety of parametric and nonparametric econometric tools to study income sorting across US residential neighborhoods. It documents the patterns of dependence among neighbors' income and imperfect sorting, with moderate but very significant correlation among incomes of neighbors and of considerable income mixing in U.S. neighborhoods.

JEL classification codes: D31, C14.
Keywords: income distribution, neighborhood effects, neighborhood sorting, nonlinear kernel estimation.

neinequa.tex

[1] I thank Luis Ahn, Marcelo Coca-Peraillon, Pamela Jia, Steve McDermid, and Annie K. Thompson for their generous assistance and Henry G. Overman for his help with tSrF. Anna Hardman, Danny T. Quah, Stuart Rosenthal, Jeff Zabel and especially Dennis Epple provided insightful comments at various stages of this work. I also thank Noel Cressie for a key reference. I am grateful to the National Science Foundation, the John D. and Catherine T. MacArthur Foundation and Tufts University for research support.

An earlier version of this paper, dated March 17, 2000 was presented at the Econometric Society Winter Meeting, New Orleans, January 2001. I am grateful to Arie Kapteyn for his careful comments.

1 Introduction

Research in economics and sociology has examined whether or not U.S. residential areas exhibit much more sorting by income, relative to what would occur if households were randomly drawn from the national income distribution and assigned to communities. This question has motivated a vast amount of research, especially starting with Tiebout (1956), who sought to explain the endogenous formation of communities in economies where local government use taxes to provide for public services. Individuals are seen to sort themselves into communities according to their willingness to pay for local public services. More recently, attention has been directed to the role of sorting in the reproduction of economic inequality. Prominent theoretical contributions in this area are by Durlauf (1996) and Benabou (1996), who study community formation jointly with human capital in the presence of social spillovers. Among empirical contributions, a notable one is by Kremer (1997), who shows that although parents' education and the mean education of residents of census tracts where their children grew up are important determinants of children's education, the estimated parameters are not large enough for residential sorting by education (and income) to contribute substantially to the dispersion of education within the population.

While the literature on Tiebout sorting, or at least the theoretical one, has been able to explain stratification according to income across communities, it has not provided a natural benchmark against which to evaluate the extent of stratification and has predicted, in a sense, too much stratification. Recent attempts to remedy this include Epple and Platt (1998) and Epple and Sieg (1999). Epple and Platt show that if individuals differ with respect to income and to a preference parameter, then the resulting sorting at equilibrium is partial but incomplete. Individuals with identical incomes may be found in different communities at equilibrium, which broadly accords with the facts. Epple and Sieg estimate a model somewhat along the lines of Epple and Platt. They show that when income and a preference characteristic are assumed to vary across the population, then individuals sort themselves across communities, which are characterized by an *ascending bundles property*: communities may be ranked in terms of incomes, housing prices and community-specific public services. The most recent empirical work on Tiebout sorting by Hoyt and Rosenthal (1997) rejects the strict implications of Tiebout sorting that all individuals in a particular community would derive the same marginal benefit from local public goods.

Imperfect sorting across communities is consistent with a variety of motives. The extent in which it rests on preferences for public services is important for designing public policy. For example, local communities, state governments and even the U.S. government have staked out positions on the desirability of income (and ethnic) mixing in residential patterns and adopted policies to promote them. It matters for policy whether or not individuals want to be near others with particular characteristics. Communities are made up of smaller neighborhoods, whose characteristics (like patterns of building density and availability of amenities, etc.) and appeal are essential to determining the character of a community. This paper is also motivated by the fact that very little is known about income distributions at the microscopic level of residential neighborhoods.

Hardman and Ioannides (1998) provide a qualitative description of sorting and mixing by income in the immediate neighborhoods surrounding a randomly chosen sample of urban dwellers. They use the neighborhood clusters data of the American Housing Survey for 1985 and 1993, which include information on the characteristics of ten closest neighbors of the basic random sample. Using U.S. government definitions of income categories, Hardman and Ioannides find that low income households are widely represented within these micro-neighborhoods: very low and extremely low income households are present, for example, in almost nine- tenths of all US neighborhoods. In three out of five neighborhoods sampled, the poorest two or three households (out of ten) come from the poorest 30 % of the population. High income households (the richest 30 %) are widely distributed. They are present in almost three fourths of all neighborhoods, and are represented by at least two or three households out of ten in about two in about two fifths of all neighborhoods. Mayer (1996) contrasts changes in inequality in urban areas using micro data from the March CPS.

The importance of such a microscopic scale of analysis is highlighted by the work of Thomas Schelling [Schelling (1971; 1978)]. Schelling studies spatial social structure, that is, the spatial outcomes that are possible where different individuals differ with respect to their preference for the characteristics of their *immediate* neighbors. Schelling's theory explains how individuals' interaction in their immediate neighborhoods can be responsible for key features of entire urban communities. Neighbor-to-neighbor interactions can have large-scale consequences, because they can lead to chain reactions. One household moves for its own reasons, but if its move tips some balance, it may cause others to move, too. As a result, a variety of stable and unstable spatial outcomes are possible. In contrast, Tiebout sorting of individuals into communities rests on preferences with respect to local public goods.

Schelling's model provides valuable theoretical underpinnings for the intuitive insight and growing body of evidence that immediate proximity is an important element of the social fabric of U.S. cities. To understand patterns of mixing at larger scales, and to assess the feasibility and potential impact of deliberate policies aimed at income mixing, we clearly need also to examine mixing at smaller scales.

This paper aims at a better understanding of the distribution of income within neighborhoods and its relationship with the national income distribution. Neighborhoods in this paper are loosely defined to the smallest reasonable scale for which data are available: they consist of a dwelling unit and up to its ten nearest neighbors. Most previous work on neighborhoods has relied on census tracts as a concept of neighborhood. Census tracts comprise much larger population groups, that is 3500 to 5000 inhabitants. If individuals differ with respect to many characteristics, observed imperfect sorting would be consistent with different behavioral patterns. Sorting according to income is particularly interesting, as income is a central ingredient of economic models. Yet, economics research to date provides little or no direct guidance regarding sorting into *small* neighborhoods. Therefore, additional structure and, possibly, specific assumptions must be made about individuals's preferences over the characteristics of their neighbors. Differences among households with respect to income may simply be due differences in life cycle stages different households are found. Clearly, neighborhoods may be mixed in terms of people of different ages, whose current incomes differ because they happen to be on different points in their life cycles and but whose permanent incomes might differ by less. Still, such differences are factors that affect the character of neighborhoods.

Our results reveal the importance of accounting for neighbor selection bias. The samples of kernels and neighbors are not very different, when considered separately. If income data for kernels and neighbors are used simultaneously, then the extent of income mixing in neighborhoods is obscured and confused with dispersion of incomes across the entire economy. Our results show that treating kernels and neighbors alike, even while allowing for stochastic dependence across kernels and their neighbors, exaggerates national income inequality and confuses it with neighborhood income mixing. By allowing for incomes of neighbors to have different dispersion characteristics than the national income distribution, while controlling for the income of a typical neighbor, we may measure the outcome of neighborhood selection, i.e. sorting, in a reduced form fashion. We find, for example, that the correlation coefficient between incomes of a randomly chosen individual and her neighbors is, at around 0.3, moderate but statistically very significant. When the sets of neighbors' incomes are defined as conditional on those of their neighbor in common, then the data do support the notion that neighbors' incomes are dependent on those of their common neighbor, and very significantly, although weakly, correlated.

The remainder of this paper is organized as follows. Section 2 presents a basic model of neighborhood sorting and Section 3 the econometric models. Section 4 discusses the data. Section 5 presents the results and 6 concludes.

2 The Sorting Model

Let $\mathcal{I}$ denote a set comprising of all individual members of the economy at some point in time, possibly very large and therefore represented by a continuum, and let I denote the total population of the economy at the corresponding time, $I = |\mathcal{I}|$. Time is not essential in our analysis, because we work with repeated cross sections and do not attempt to link the observations over time. We fix ideas and set notation as follows.

Let $F(y)$ denote the distribution function of household income y in the entire economy. Suppose the population is distributed into K different geographical areas, to be referred to as neighborhoods, $\mathcal{I}_k,\ k = 1, \ldots, K$, each with population $I_k = |\mathcal{I}_k|$, and neighborhood-specific income distribution function F_k, $k = 1, \ldots, K$, with support $\mathcal{I}_k(y)$. By definition, the decomposition of the population into neighborhoods ie exhaustive, so that $\mathcal{I} = \bigcup_{k=1}^{K} \mathcal{I}_k$. In that case, the economy-wide, national, distribution of income among individuals, *the national income distribution,* is given by:

$$F(y) = \sum_{k=1}^{K} \frac{I_k}{I} F_k(y), \tag{1}$$

from which the density function follows by differentiation.

We say that the national income distribution exhibits *perfect mixing*, if all neighborhood income distributions are identical to the national one, that is, if Equ. (1) implies:

$$F_k(y) \equiv F(y), \ \forall k \text{ and } \forall y. \tag{2}$$

It exhibits *perfect sorting*, if the supports of the neighborhood income distributions do not overlap: $\mathcal{I}_{k_1}(y) \bigcap \mathcal{I}_{k_2}(y) = \emptyset$. E.g., in the simplest possible case where the $F_k(y)$'s are degenerate, then:

$$F_k(y) = 1, \text{if } \ y = y_k; \ F_k(y) = 0, \text{if } y \neq y_k, \forall k, \tag{3}$$

and where all the y_k's are all different. In that case, the national income distribution will be described by the discrete frequency distribution $\{(y_k, \frac{I_k}{I}) : k = 1, \ldots, K.\}$.

2.1 The Epple–Sieg Model of Sorting

We use the basic model in Epple and Sieg (1999) to seek key characteristics of neighborhood income distributions that are associated with the general case of imperfect sorting of individuals into neighborhoods. The sorting model may be analyzed in terms of the standard tools of sample selection bias. Yet, it is the associated income distributions that have not really been analyzed in full detail. This is in part due to the fact that the econometric literature has sought to obtain suitable estimators for problems that are subject to sample selection bias rather than to identify the statistical characteristics of the associated income distributions. In a sense, this paper looks at estimates of magnitudes which sample selection correction considers auxiliary to the main task of correcting for selection bias.

We assume that all individuals make decisions at the same time about where to locate among $k = 1, \ldots, K$ neighborhoods. Individuals' preferences are defined in terms of their indirect utility functions, as functions of individual income y, of the price of housing in neighborhood k, P_k, an observable variable in principle, and of a quality attribute of the neighborhood, the state of the neighborhood, g_k. Following Epple and Sieg (1999), we assume an indirect utility function for a household with income y residing in neighborhood k, as a function of (y, P_k, g_k) of the form:

$$V(y, P_k, g_k; \epsilon) \equiv \left[\epsilon g_k^{\psi} + \left[\exp\left(\frac{y^{1-\nu} - 1}{1 - \nu} \right) \exp\left(-B \frac{P_k^{\eta+1} - 1}{1 + \eta} \right) \right]^{\psi} \right]^{\frac{1}{\psi}}, \tag{4}$$

where $\epsilon > 0$, is an individual characteristic, that is distributed across the population, possibly jointly with income y; $\psi < 0,\ \eta < 0,\ \nu > 0$,and $B > 0$ are parameters which are constant across all agents.

To see how the assumption about preferences according to (4) serves to clarify sorting across neighborhoods, we may consider "indirect indifference" curves in (P_k, g_k) space. Their slopes given by:

$$\frac{\partial P}{\partial g}|_{V=\text{const}} = \frac{\epsilon g^{\psi-1}\left[\exp\left(\frac{y^{1-\nu}-1}{1-\nu}\right)\exp\left(-B\frac{P_k^{\eta+1}-1}{1+\eta}\right)\right]^{-\psi}}{Bp^{\eta}} > 0. \tag{5}$$

These curves are essential in characterizing neighborhood sorting for the following reason. Since they are monotonic in y and ϵ, they satisfy the single-crossing property with respect to income, y, and to the taste parameter, ϵ, given y. As Epple and Sieg, *op. cit.,* show, this property is crucial for obtaining separating equilibria, with respect to both income, y and the taste parameter, ϵ. To see this intuitively, consider the indirect indifference curves for two values of income, $y',\ y'',\ y' < y''$, with the same value of ϵ. As the indifference curve for y'' cuts the one for y' from below, individuals with incomes equal to y'' are willing to bid a higher value to locate in any particular community, *cet. par.*.[2] Therefore, a neighborhood with higher value of P_k, holding g_k constant, would be populated by households with higher incomes.

We index the K neighborhoods in individuals' opportunity sets, so that: $d_k < d_{k+1},\ P_k < P_{k+1},\ k = 0, \ldots, K-1$. According to Epple and Sieg, *op. cit.* p. 651, there must be an ordering of communities that must be confirmed at equilibrium. [3] We assume that this indexing coincides with the equilibrium ordering. We work out the specifics of selection which is likely to emerge under preferences (4). Next we seek to characterize the marginal density function for neighborhood k, $f_k(\ell ny)$ in terms of the joint distribution of preferences characteristics and income, (y, ϵ), across the population.

The set of individuals $j \in \mathcal{I}_k$ who reside in neighborhood k are characterized by the set of values (y_j, ϵ_j) such that:

$$V(y_j, P_{k-1}; g_{k-1}; \epsilon_j) < V(y_j, P_k; g_k; \epsilon_j) \leq V(y_j, P_{k+1}; g_{k+1}; \epsilon_j). \tag{6}$$

We follow Epple and Sieg (1999) but simplify by setting $\nu = 1$ in which case $\lim_{\nu=1} : \left(\frac{y^{1-\nu}-1}{1-\nu}\right) =$

[2] The single-crossing property in the context of residential sorting was introduced by Ellickson (1991).

[3] This ordering must satisfy *boundary indifference, income stratification, and ascending bundles.* That is, if $P_i > P_j$, then $g_i > g_j$, iff neighborhood i is populated by higher income people than neighborhood j.

ℓny. It turns out that the boundary of neighborhoods k and $k+1$ in (y, ϵ) space is the straight line given by $\ell n\epsilon - \psi \ell ny = C_k$. Conditions (6) are transformed into:

$$C_{k-1} + \psi \ell ny < \ell n\epsilon \leq C_k + \psi \ell ny, \tag{7}$$

where the C_k's are auxiliary variables defined by

$$C_k \equiv \ell n \left(\frac{\exp\left[\frac{-\psi}{\eta+1}(BP_{k+1}^{\eta+1} - 1)\right] - \exp\left[\frac{-\psi}{\eta+1}(BP_k^{\eta+1} - 1)\right]}{g_k^{\psi} - g_{k+1}^{\psi}} \right), k = 1, \ldots, K-1. \tag{8}$$

Note that C_k is increasing in g_k and P_k. Our assumptions about the ranking of the K neighborhoods imply that the C_k's, which are functions of the $(g_k, P_k; g_{k+1}, P_{k+1})$'s, satisfy $C_{k+1} > C_k$. For completeness, we define $C_0 = -\infty$, and $C_K = \infty$. The proposition that follows summarizes the implications of this model for the neighborhood income distributions. The proof of this proposition is rather straightforward and has been relegated to Appendix A.

Proposition 1. *Let the distribution of $(\ell ny, \ell n\epsilon)$ be bivariate normal with parameters $(\mu_y, \mu_\epsilon, \lambda, \sigma_y, \sigma_\epsilon)$.*

Part A. *The distribution of income in neighborhood k, $f_k(\ell ny)$, is given by the marginal with respect to ℓny, multiplied by the probability that $\ell n\epsilon$ lies within the bounds given in (7) conditional on ℓny. That is,*

$$f_k(\ell ny) = \frac{1}{G_k} \frac{1}{\sqrt{2\pi}\sigma_y} exp\left(-\frac{1}{2} \frac{(\ell ny - \mu_y)^2}{\sigma_y^2} \right) \cdot \left[\Phi\left(Z_k(y)\right) - \Phi\left(Z_{k-1}(y)\right)\right], \tag{9}$$

where $\Phi(\cdot)$ denotes the cumulative standardized normal distribution function, the auxiliary variables Z_{k-1}, Z_k are defined as

$$Z_k(y) \equiv \Omega_j + \omega \ell ny,$$

$$\Omega_k \equiv \frac{C_k - \mu_\epsilon + \lambda \frac{\sigma_\epsilon \mu_y}{\sigma_y}}{(1-\lambda^2)^{\frac{1}{2}} \sigma_\epsilon}, \tag{10}$$

$$\omega \equiv \frac{1}{(1-\lambda^2)^{\frac{1}{2}}} \left(\frac{\psi}{\sigma_\epsilon} - \frac{\lambda}{\sigma_y} \right), \tag{11}$$

and G_k is a normalizing constant, which is equal to the unconditional probability that an individual be in neighborhood k:

$$G_k = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_y} exp\left(-\frac{1}{2} \frac{(\ell ny - \mu_y)^2}{\sigma_y^2} \right) \cdot \left[\Phi\left(Z_k(y)\right) - \Phi\left(Z_{k-1}(y)\right)\right] d\ell ny.$$

Part B. *The income distribution in neighborhood k, $f_k(\ell ny)$, may under certain conditions be bimodal.*

This result is interpreted as follows. Individual choices result in sorting into neighborhoods consisting of individuals with income distributed within the neighborhood according to the unconditional national income distribution, scaled by the difference between two suitable defined cumulative normal distribution functions, $\Phi^*(\ell ny) \equiv \Phi(Z_k(\ell ny)) - \Phi(Z_{k-1}(\ell ny))$. These functions, if viewed as functions of ℓny, may be plotted over the horizontal axis as two cumulative normal distribution functions corresponding to standardized normal distributions centered at $\frac{\Omega_k}{-\omega}$ and $\frac{\Omega_{k-1}}{-\omega}$. The properties of $\Phi^*(\ell ny)$ as a function of ℓny depend upon the sign of ω. If the taste parameter and income are positively correlated, $\lambda > 0$, then $\omega < 0$, and the Z_k's are decreasing functions of ℓny. If, on the other hand, income and the taste parameter are negatively correlated, $\lambda < 0$, then the sign of ω depends on the relative magnitudes of all parameters, and the $Z_k(\ell ny)$'s can become increasing functions of ℓny.

If individuals do not differ with respect to their evaluation of g_k, then the sorting conditions are simplified. Under the normalizing assumption $\epsilon = 1$, the sorting conditions (7) become:

$$\frac{1}{-\psi} C_{k-1} < \ell ny \leq \frac{1}{-\psi} C_k. \tag{12}$$

These inequalities define bounds in terms of ℓny, and no integration with respect to ϵ is called for. The resulting distributions may be described concicely by:

$$f_k^S(\ell ny) = \frac{f(\ell ny)}{\int_{C_{k-1}/-\psi}^{C_k/-\psi} f(\ell ny) d(\ell ny)}, \ k = 1, \ldots, K-1, \tag{13}$$

where ℓny lies in the interval defined by (12). The neighborhood income distributions are simply doubly truncated segments of the national distribution. They describe perfect sorting. If, on the other hand, individuals' evaluations of g differ, the neighborhood income distributions extend over the entire support of the national income distribution, even if income and the individual preference characteristic are uncorrelated. As long as taste heterogeneity is present, and even if it is uncorrelated with income, sorting according to income would be imperfect.

Careful inspection of the difference in the second term in the r.h.s. of (9) hints to the possibility that neighborhood income distributions could be bimodal. The weighting of the national distribution is accentuated between C_{k-1} and C_k, with the no taste heterogeneity case being the extreme case, where the distribution has mass only between the two bounds. However, the clustering of incomes around the national mean suggests that a second mode may survive even after the national income distribution has been scaled down. This is clearly more likely to happen the further the bounds are away from the national mean. We note that a straightforward consequence of bimodality in this context would be that neighborhood distribution might have larger variance than the national distribution.

The fact that neighborhood income distributions depend upon the C_k's only is not a special property of the model in the absence of individual taste heterogeneity. In fact, as Lemmata 1 and 2, Epple and Sieg, *op. cit.,* p. 654 show, the C_k's may be recursively defined in terms of the unconditional probabilities that individuals locate in neighborhoods $k = 1, \ldots, K$. In our notation, if we normalize I to one, then

$$C_k = \mathcal{C}(C_{k-1}, G_k | \mu_y, \mu_\epsilon, \lambda, \sigma_y, \sigma_\epsilon), \tag{14}$$

which along with $C_0 = -\infty$, $C_K = \infty$, define fully the C_k's. These auxiliary variables may be rightly referred to as *community-specific intercepts,* and serve as sufficient statistics.

The analysis so far has taken prices as given. Ultimately, of course, prices are endogenous. Given individual valuations of neighborhood conditions, individuals bid for the privilege in locating in different neighborhoods. A complete analysis would require a more precise definition of a neighborhood in terms of the housing stock. Other interesting aspects of the neighborhood location problem, such as individuals' taking into consideration the characteristics of others who also choose to locate in the same neighborhood, also complicate the problem considerably. In the remainder of the discussion in this paper, we take prices as given. In an effort to account for variations in housing prices across the US, we use the full detail of geographic information which is available in our data in addition to incomes.

We note that the model of heterogeneous preferences introduced in (4) above may not be fully identified from income distribution data alone. Nonetheless, the model serves to structure the empirical investigation below.

3 Econometric Models

Suppose now that data are available in the form of a finite random sample of neighborhoods in the economy. Let different observations of neighborhoods be indexed by k, $k = 1, \ldots, K$. The sample is made up of observations on incomes of a random subsample of the entire national sample, $\mathbf{Y}_{ke} = \{y_k : k = 1, \ldots, K\}$, which may be treated as a representative sample drawn from the national income distribution. Each observation k is known as the *kernel* of neighborhood k, and ke is mnemonic for *ke*rnel. The AHS data provide information on the neighbors of each kernel. We use $j = 1, \ldots, n$ to index the sample of *nearest neighbors* to kernel k in her neighborhood, $\mathcal{I}_k$. The set of nearest neighbors of k in the data will be referred to as *neighborhood cluster k*. In practice, the number of neighbors differs across kernels because of missing values and changes in sampling procedures. We return to this issue below.

This sampling procedure samples K neighborhoods, and conditional on a randomly chosen member in each neighborhood, their n nearest neighbors. We shall, for simplicity, consider that neighborhoods are spatially homogeneous groups, so that the samples $\mathbf{Y}_k = \{y_{k1}, \ldots, y_{kn}\},\ k = 1, \ldots, K$, that is sets of random vectors drawn from the distribution function of incomes in neighborhood clusters $k = 1, \ldots, K$. That is, $\mathbf{Y}_k$ is a random vector of size n whose components are i.i.d., conditional on y_k, drawn from probability density function $f_k(y_{kj}|y_k), j = 1, \ldots, n$. It will be useful to define, for every neighborhood cluster k, the sample maximum and minimum, $\text{MAX}_k = \max_{j \in \mathcal{I}_k}\{y_{kj}\}$, $\text{MIN}_k = \min_{j \in \mathcal{I}_k}\{Y_{kj}\}$.

The distribution of income within each neighborhood, according to Proposition 1, and given housing prices and the states of neighborhoods, depends on the parameters of the national income distribution, of the distribution of the taste parameter evaluating the state of each neighborhood, and on the remainder of preference characteristics. However, as mentioned above, the community-specific intercepts help simplify the estimation problem. In principle, it should be possible to handle the problem of estimating the neighborhood income distributions, that were defined in Section 2 above, by means of maximum likelihood methods. Unfortunately, this direct approach did not work.[4]

[4] I am truly grateful to Dennis Epple for his generosity in working out the following detailed suggestions. Since a neighborhood is a very small share of the metropolitan population, the integral between the upper and lower neighborhood boundaries for a given value of income can be written in terms of the density. Then, assuming all neighborhoods are of the same size provides a normalization that can be exploited by completing the square of the exponents of the likelihood function and then integrating over $(-\infty, \infty)$. In this fashion, a density of the sample of income observations for a given neighborhood can be written, conditional on the unobserved community-specific intercept. When neighborhoods are small and the sample of neighborhoods is randomly drawn, the community-specific intercepts are normally distributed. This distribution permits integrating out the community-specific intercepts to yield a likelihood function written in terms of observables and parameters to be estimated. It is very unfortunate that maximum likelihood estimation along these lines did not work out for computational reasons, in spite of a relatively

We therefore resort to use of class of parametric models as reduced-form approximations of the true model. This approach can handle some important hypotheses derived from the theoretical analysis and generally helps to further clarify the properties of neighborhood income distributions. A particularly interesting hypothesis, derived from the theoretical analysis, is that neighborhoods are imperfectly segregated in a manner that implies higher variance than the national income distribution.

Sample design suggests that we could take as the maintained hypothesis that the incomes of kernels in the data are a random sample from the national income distribution. We shall use upper case Y from now on to denote log incomes. Let the distribution of log incomes among kernels, $Y = \ell n y$, be normally distributed with mean and variance denoted, respectively, by μ_{ke} and ς^2. Let the distribution of log incomes of neighbors also be normal but correlated with that of their kernel, that is, let μ_{ne}, σ^2, ρ denote, respectively, the mean income of neighbors (mnemonic *ne*, from *ne*ighbors), its variance, and the correlation coefficient between incomes of neighbors and kernels. We think such a specification of the distribution of income of neighbors conditional on that of their kernels follows the intuition that emanates from Proposition 1.

Conditional on the log income of the kernel of neighborhood k, Y_k, the incomes of its neighbors are normal with parameters $(\mu_{ne}+\rho\frac{\sigma}{\varsigma}(Y_k-\mu_{ke}), \sigma^2(1-\rho^2))$. This allows us to write the log likelihood function for a set of observations from neighborhood cluster k, $(Y_k; Y_{k1}, \dots, Y_{kn})$, as follows:

$$\mathrm{LLF}(Y_k;\mathbf{Y}_k) = -\frac{1}{2}\left[\frac{(Y_k-\mu_{ke})^2}{\varsigma^2} + \ell n[2\pi\varsigma^2]\right]$$
$$-\frac{1}{2}\sum_{j=1}^{|n_k|}\left[\frac{(Y_{kj}-\mu_{ne}-\rho\frac{\sigma}{\varsigma}(Y_k-\mu_{ke}))^2}{\sigma^2(1-\rho^2)} + \ell n[2\pi\sigma^2(1-\rho^2)]\right] \qquad (15)$$
$$-\sum_{j=1}^{|n_k|}\ell n\left[\Phi\left(\frac{(\mathrm{MAX}_k-\mu_{ne}-\rho\frac{\sigma}{\varsigma}(Y_k-\mu_{ke}))}{\sigma(1-\rho^2)^{\frac{1}{2}}}\right) - \Phi\left(\frac{(\mathrm{MIN}_k-\mu_{ne}-\rho\frac{\sigma}{\varsigma}(Y_k-\mu_{ke}))}{\sigma(1-\rho^2)^{\frac{1}{2}}}\right)\right].$$

We note that maximum likelihood estimations with this model allows for data for both kernels and neighbors to contribute to the estimated parameters of the national income distribution.

In the remainder of the paper, we refer to the model whose log likelihood function is given by Equ. (15) as Model 5. It serves as the most general case in our setting that nests a number of simpler models that we may define. So, Model 0 is the case with equal means and variances

small number of parameters.

across kernels and neighbors, $\mu_{ke} = \mu_{ne}, \sigma^2 = \varsigma^2$, no correlation, $\rho = 0$, and with the $\Phi(\cdot)$'s not present. Model 1 allows for correlation between kernels and neighbors, $\rho \neq 0$, but keeps all other assumptions of Model 0. Model 2 allows for different means and variances and nonzero correlation between incomes of kernels and neighbors, $\mu_{ke} \neq \mu_{ne}, \sigma^2 \neq \varsigma^2, \rho \neq 0$, but the $\Phi(\cdot)$'s are not present. Model 3 corresponds to Model 0, except that the $\Phi(\cdot)$'s are present, and Model 4 to Model 1, except that $\Phi(\cdot)$'s are present. Even though the $\Phi(\cdot)$'s do not include additional unknowns, they do bring into the estimation additional information associated with the observed lower and upper bounds within each cluster and for each observation. The exact expressions for the likelihood functions for each of the models 1—4 are given in Appendix B.

The generality of Model 5 is meant to account for possible biases inherent in using the clusters data, that is to account for neighborhood selection and the dependence it implies among incomes of kernels and neighbors. For example, if income data for kernels and neighbors are used simultaneously and there is dependence among neighbors' incomes, then the extent of income mixing in neighborhoods is obscured and confused with dispersion of incomes across the economy. Comparison of Models 1 and 2 suggests that treating kernels and neighbors alike, even while allowing for stochastic dependence across kernels and their neighbors exaggerates national income inequality and confuses it with neighborhood income mixing. By allowing for incomes of neighbors to have different dispersion characteristics than the national income distribution we may measure the outcome of neighborhood selection, i.e. sorting, in a reduced form fashion.

Next, after a brief discussion of the data, we present a nonparametric analysis of the neighborhood cluster structure of the data and then follow up with parametric econometric analyses, obtained with a variety of estimation methods. The first part of our analysis of neighborhood income distributions involves nonlinear estimations of neighbors' incomes, conditional those of the kernels. The second part is parametric estimation of that same pattern among incomes of neighbors and their kernel when we allow for general dependence among all neighbors' incomes, while conditioning on their kernels' incomes, by using seemingly unrelated regressions. The third part involves parametric maximum likelihood estimations along the lines of Equ. (15). As we mentioned above, such estimations adopt but simplify the neighborhood income distribution model articulated by Proposition 1.

4 The Data

The AHS is a panel of housing units, which involves in its entirety more than 50,000 dwelling units that are interviewed each two years, and serves the basis for US housing statistics. We explore a somewhat neglected aspect of the data, data on neighborhood clusters, which are available only for years 1985, 1989, and 1993. In those years only, a random sample of, respectively 670, 805, and 1014 urban units were selected and for each one of them originally up to ten and later on more neighbor units were interviewed. Each neighborhood cluster is defined by its kernel and includes all neighbors.[5] Restricting attention to regular interviews only (that is those with an actual household head) reduces the number of available points. [6] Also, additional units in existing clusters were included in 1989 to reflect additional units that had been added within the perimeter of the "neighborhood." By 1993, a maximum of 20 neighboring units were allowed per cluster.[7] In view of the change in the number of observations over time, and after considerable experimentations, we decided to conduct separate parametric estimations with data for neighborhood clusters that include the same number of observations, that is $n = 10$ neighbors, for 1985, 1989 and 1993, which is of course a subset of the data, as well as the full sample, in which the number of observations may vary across neighborhoods. The nonparametric estimations, however, have been conducted with the full data.

We should also note that by full data we mean the sample of owner occupants. We have restricted this study to that sample for two main reasons. First, we wish to maintain comparability among several studies of neighborhood effects, such Ioannides (1999) and Ioannides and Zabel (2000), for example. Second, renters and owners could employ very different approaches to the evaluation of neighborhoods where they wish to reside, in part because of different time horizons, with renters generally being younger and much more mobile. In spite of these limitations, our data are quite representative of both the entire AHS data and national statistics for the US. This

[5]Only a handful of papers, that is, de Bartolome and Rosenthal (1995), Gabriel and Rosenthal (1995; 1996), Hoyt and Rosenthal (1995), Hardman and Ioannides (1998), Ioannides (1999; 2000), Ioannides and Seslen (1999), Ioannides and Zabel (2000), and Kiel and Zabel (1998) have utilized the AHS clusters data to date. Kiel and Zabel compare the performance of clusters data against census tract level attributes by means of privileged access to census-tract coding of the data. Ioannides (1999) aims at understanding neighborhood interactions in housing consumption and maintenance decisions. Ioannides and Seslen (1999) merge AHS clusters data and PSID micro data to compare neighborhood income and wealth distributions.

[6]See the Appendix in Ioannides (1999) for further details on the count of observations.

[7]I am grateful to Barbara T. Williams, US Bureau of the Census, for this clarification.

established by Appendix C.

5 Results

We present in this part of the paper results according to the three stages of estimations outlined above of statistical and econometric procedures to discern patterns in neighborhood income distributions. We order randomly the observations for each cluster into a vector of ten observations and associate them with their kernel. We start by an examination of the importance of the cluster structure of the data.

5.1 The Neighborhood Cluster Structure of the Data

We report first estimated correlation coefficient among neighbors, for each of the three waves of data, 1985, 1989 and 1993. The estimated correlation coefficients, respectively, among the elements of $\mathbf{Y}_k$ across the data, $\mathbf{Y}_k = \{Y_{k1}, \ldots, Y_{k10}\},\ k = 1, \ldots, K$ and between Y_k, and $\{Y_{k1}, \ldots, Y_{k10}\},\ k = 1, \ldots, K$ range uniformly between .3301 and .5268, and between .3499 and .4897, for 1985, between .2465 and .4735, and between .3373 and .4528, for 1989, and between .2793 and .4940, and between .3501 and .5192, for 1993.

Decomposing the variance of log incomes, separately for 1985, 1989 and 1993, in terms of cluster-specific effects, we find that the cluster structure yields the following. For 1985, cluster fixed effects explain 31% of variance, and cluster random effects 25.9%; for 1989, cluster fixed effects explain 30% of variance, and cluster random effects 24.8%; and for 1993, cluster fixed effects explain 27.1% of variance, and cluster random effects 21.4%. In view of the fact that these estimations are conducted with data from the *entire* US, it is quite surprising, at least to me, that only a small portion of the variance is explained by cluster-specific effects.

In this connection, let us recall the finding by Epple and Sieg (1999), p. 671, that 89% of the total variance of income in the Boston metropolitan area is accounted for by within-community variance. This leaves 11% of the variance to be explained by inter-community variance, the counterpart to our cluster-specific effects. Our finding is, *mutatis mutandis,* [8] compatible with theirs, and especially if one were to consider that a single metropolitan area (Boston) is most likely more homogeneous than the entire US.

While the decomposition of variance gives a sense of the importance of income sorting, it would be interesting to know how closely correlated incomes of neighbors are in a group sense. Such a group correlation coefficient has been used in the literature as a measure of segregation. See Kremer (1997) and Kremer and Maskin (1996), who show that it may be computed as the R^2 from a regression of incomes against a set of dummies for all clusters in the data.[9] Our results show that this correlation coefficient is estimated at .3318, for 1985, .3220 for 1989 and .2915 for 1993. We repeated this estimation by using data from the metropolitan sample of the AHS for 1985, 1989 and 1993.[10] The correlation coefficient among inhabitants of the same census tract was estimated for the following metro areas: Boston, MA; Dallas, TX; Detroit, MI; Forth Worth-Arlington, TX; Los Angeles, CA; Minneapolis–St. Paul, MN; Philadelphia, PA; Phoenix, AZ; San Francisco–Oakland, CA; Tampa–St. Petersburg, FL; and Washington, DC. The data here come from the metropolitan sample of the AHS, which is completely different from the national sample used in the remainder of this analysis. For 1985, these estimated correlation coefficients range from a maximum of .3716, for Boston, to a minimum of .2493, for Tampa-St. Petersburg. For 1989, these correlation coefficients range from a maximum of .4436, for Washington DC, to a minimum of .2504, for Tampa-St. Petersburg.

[8] The Epple–Sieg variance decomposition is defined, in our notation and context, as:

$$\text{var}(Y) = \sum_{k=1}^{K} G_k \text{var}_{\mathcal{I}_k}(Y) + \sum_{k=1}^{K} G_k [E_{\mathcal{I}_k}(Y) - E_{\mathcal{I}}(Y)]^2.$$

With perfect sorting, the first component would be relative small. The larger is the first component relatively to the total variance the more imperfect sorting is. Actually, the Epple–Sieg is cast in terms of the importance of taste heterogeneity relative to income heterogeneity. If taste heterogeneity is more important relative to income heterogeneity, community income distributions would be quite similar. Therefore, the finding that the first component is large relative to the second suggests that taste heterogeneity is more important than income as a determinant of sorting across communities.

[9] The expression for this correlation coefficient is:

$$\text{Corr} = \frac{\sum_{k=1}^{K} \sum_{j=1}^{|n_k|} (Y_{kj} - \bar{Y}) \sum_{i=1}^{|n_k|} (Y_{ki} - \bar{Y}) \frac{1}{n_k}}{\sum_{k=1}^{K} \sum_{i=1}^{|n_k|} (Y_{kj} - \bar{Y})^2}.$$

[10] As an aside, this is a completely different sample than the national one, from which the clusters data are drawn. The metropolitan sample identifies observations that belong to the same census tract but neither reveals the respective census tract nor provides information on clusters, a feature of the national sample.

We suggest that the general similarity of the estimates for the correlation of incomes within tracts and within clusters implies that this analysis of this paper is not subject to the so-called *modifiable areal unit problem* (MAUP) from geostatistics and statistical geography. That is, the MAUP problem in spatial statistics concerns the influence of the level of aggregation (scale) within which a spatial variable is sampled on relationships between any two (or more) spatial variables. The recent statistical geography literature on MAUP emphasizes three different ways to address the different scales problem, that is distinguishing between subsetting, stratification and aggregation [Cressie (1996), p. 167]. Working with samples of different clusters within a census tract involves stratification, except that the "strata" have been randomly selected. On that score, it is encouraging that the estimated correlation coefficients from those very different samples are so close to one another, suggesting that neighborhood clusters are indeed representative random samples of census tracts.

5.2 Nonparametric Estimations

We plotted the data and estimated nonparametric models for the univariate densities. We found no evidence of bimodality in the data.

Figures 2, 3 and 4 report nonparametric estimates for the distributions of neighbors' log incomes, conditional on their kernel's log income, $f\left(Y_{k1}, \ldots, Y_{k10} | Y_k\right)$, respectively for 1985, 1989, and 1993. The descriptive statistics for the data used in these estimations are given in Table 1.

Each of the figures reports three-dimensional pictures of estimated stochastic kernels as well as their contours. Both are obtained by using Danny T. Quah's tSrF package.[11] The three-dimensional figures, Figures 3, 5 and 7, are to be read as follows. The intersection of the surface drawn by a plane perpendicular to the axis marked kernels yields the neighborhood income distribution, conditional on the respective value of income for the kernel. The contour pictures, Figures 2, 4, and 6, are to be read in the standard fashion for map contours.

Both, the three-dimensional kernels and the two-dimensional contours help clarify the process of neighborhood sorting in US neighborhoods. To appreciate what they show, consider the two extreme cases of, respectively, perfect mixing and perfect sorting, as defined above. If US neigh-

[11] This is available from Danny T. Quah's LSE web page at http://econ.lse.ac.uk/~dquah/tsrf.html. See Quah (1996) for a standard application.

borhood income distributions reflected perfect mixing, then the stochastic kernels would have their peaks lined up parallel to the kernel income axis. If US neighborhood income distributions reflected perfect sorting, then the stochastic kernels would have their peaks lined up along the 45^o line. As the contour pictures clarify, the peaks of the conditional distributions line up along a line which is steeper than 45^o. This suggests imperfect sorting. The mode of the distribution of neighbors' incomes increases less than proportionately with those of their kernels'. Careful observations of the contour maps suggests that, generally, the variance of the income of neighbors conditional on the income of kernels, declines with kernel income. It is interesting to consider adapting the estimation so that to each kernel one associate the entire sample of neighbors' incomes. This is accomplished in a linear model by means of the regressions which follow.

5.3 Neighbors' Incomes Conditional on Kernels' Incomes

We report estimation results for seemingly unrelated regressions with a model of neighbors' incomes as a function of the respective kernel's income, separately for each of the three waves of the data:

$$\boldsymbol{Y}_k = \boldsymbol{\alpha} \ + \boldsymbol{\beta} \, Y_k + \boldsymbol{\epsilon}_k \tag{16}$$

where $\boldsymbol{\alpha}$ is a 10-vector of intercepts and $\boldsymbol{\beta}$ a 10-vector of coefficients. We allow for general correlation matrix for the 10-vector of stochastic shocks $\boldsymbol{\epsilon}_k$.

The estimates of the components of β, with t statistics given in parentheses and ranging roughly within similar bounds, range from .3809 (6.02) to .5276 (8.56), for 1985, from .3163 (5.15) to .4947 (9.33), for 1989, and from .2457 (4.13) to .4997 (8.72), for 1993. Interestingly, allowing for a variable number of neighbors and thus extending the size of the data set does not reveal any noticeable differences. The estimated correlation matrices for the ϵ_k's are generally smaller than the raw ones, reported above, rather uniform and still different from 0. They range from .2179 to .3992, for 1985, from .1595 to .3570, for 1989, and from .1392 to .4041, for 1993. These results are in broad agreement with the nonparametric estimation results reported in subsection 5.2 just above. As we saw there, the mode of the distributions of neighbors's incomes, conditional on those of their kernels' incomes, increases with the kernel's income but by less than proportionately to it.

5.4 Maximum Likelihood Estimations of Neighborhood Income Distributions

An important feature of the parametric results with neighbors' incomes conditional on those of their kernels is that the correlations among all neighbors' incomes do not differ very much. We report estimations with models that would hopefully provide additional intuition about the phenomenon of the neighborhood income sorting. Specifically, Model 5, which is described in Section 3 above, imposes a restricted structure of correlation relative to the model in section 5.3 above.

We report in Tables 2 – 7 maximum likelihood estimation results for the family of models, whose loglikelihood functions are nested in Equ. (15). This stochastic structure allows for the incomes of kernels and neighbors to have different means and variances and the incomes of neighbors may be correlated with those of their kernels.

We report in detail below in Table 2 the results for Models 0 – 5 for 1985. Model 0 below imposes the same means and variances across kernels and neighbors. Model 1 allows, in addition, for correlation between neighbors' and kernels' incomes. Model 2 allows the means to vary and also allows for correlation between neighbors' and kernels' incomes. Models 3, 4 and 5 are like Models 0, 1, and 2, respectively, except that in addition the requirement is imposed that all data lie between the observed minimum and maximum among the incomes in each neighborhood.

For each of these models, the second column in Table 2 reports, in addition, estimations for an extended model where dummies are introduced to allow for the means of kernels and of neighbors to vary according to four US Census regions, climatic regions, and categories of urbanicity. Such an extended list of regressors is included in order to account for the fact that the determinants of neighborhood sorting, according to the model presented in Section 2, may in general vary across different regions and geographic areas with different density characteristics.

The results for 1989 are quite similar to those of 1985, but those for 1993 show some important differences. The estimate of the correlation coefficient is much smaller, at .163 (9.15), and the variances of kernels and neighbors are much nearer one another, 1.511 (13.53) and 1.590 (104.31), respectively. However, when the data are grouped by region and, alternatively, by category of urbanicity the decrease in the correlation coefficient is hard to explain.

Table 3 compares the results obtained with Model 5 for 1985, as reported in the last column of Table 2, with those for 1989 and 1993. Comparison of the first and third columns of that table suggests the coefficient of variation of kernels' incomes increases from .109 in 1985 to .126 in 1993, while at the same time, income mixing decreases in neighborhoods, as the coefficient of variation of neighbors' incomes decreases from .137 to .126. This pattern is also present when the stochastic structure is allowed to be region-specific. Generally, means for kernels and neighbors are quite close, but variances differ considerably, except for 1993.

Tables 4 and 5 report the results obtained with Model 5 estimated separately for each of the four regions in the data, for 1985 and 1993, respectively. From Tables 4 and 5, the coefficient of variation of log incomes of kernels increases from 1985 to 1993, for the Northeast and the South, and income mixing in neighborhood decreases for all regions except the Midwest.

Tables 6 and 7 report the results obtained with Model 5 estimated separately for each of the four groupings of the data according to degree of urbanicity, for 1985 and 1993, respectively. According to these results, income inequality in central cities of metropolitan areas decreased from 1985 to 1993, with the respective coefficient of variation of log incomes going from .138 to .103, but it increased in the suburbs, with the respective coefficient of variation going from .083 to .144. And income mixing within neighborhoods decreased over the same period of time, in both central cities and suburbs, with the respective coefficient of variation going from .152 to .141, and from .102 to .098, respectively.

The availability of separate estimations when we allow for the stochastic structure to vary by region, as in Tables 4 and 5, and by urbanicity, as in Tables 6 and 7, prompts a number of remarks. The differences between the variances of kernel and neighbor incomes appear to be eliminated when we estimate separate models for the four regions, but remain quite pronounced when we estimate separate models for the four types of urban areas. It would be interesting to reconcile our findings with evidence of widening income segregation in urban areas along with a slight narrowing of racial segregation and an increase in income inequality, as reported by Mayer (1996) on the basis of different data.[12]

[12] Mayer (1996) compares the central city concentration of families in each income quintile to the central city concentration of the average family using CPS data for MSAs from 1964 to 1994. He shows that the ratio of the central city percentage of families in the bottom quintile to the central city percentage of all families exhibits an upward trend and the respective ratio for the top quintile exhibits an downward trend.

6 Conclusions

Our results reveal the importance of accounting for neighbor selection bias. The samples of kernels and neighbors are not very different, when considered separately. If income data for kernels and neighbors are used simultaneously, then the extent of income mixing in neighborhoods is obscured and confused with dispersion of incomes across the entire economy. Our results show that treating kernels and neighbors alike, even while allowing for stochastic dependence across kernels and their neighbors, exaggerates national income inequality and confuses it with neighborhood income mixing. By allowing for incomes of neighbors to have different dispersion characteristics than the national income distribution, while controlling for the income of a typical neighbor, we may measure the outcome of neighborhood selection, i.e. sorting, in a reduced form fashion. We find, for example, that the correlation coefficient between incomes of a randomly chosen individual and her neighbors is, at around 0.3, moderate but statistically very significant. When the sets of neighbors' incomes are defined as conditional on those of their neighbor in common, then the data do support the notion that neighbors' incomes are dependent on those of their common neighbor, and very significantly, although weakly, correlated. The estimated dispersion of incomes within neighborhoods is significantly larger than the dispersion of incomes of a randomly drawn sample of the national population. In other words, income mixing, another way to express imperfect sorting, is a real phenomenon.

While this possibility was anticipated by our theoretical discussion, which draws on a model by Epple and Sieg, we believe that the findings reported here are the first measurements of important characteristics of income distributions within small US neighborhoods. It would be interesting to see what these findings imply for the effectiveness of programs deliberately designed to bring about income mixing, such as Moving to Opportunity and others.

TABLE 1

Descriptive Statistics AHS Neighborhood Clusters Data: 1985, 1989, 1993

Year	1985	1989	1993
Number of clusters	232	244	310
Mean, Income of Kernels (current $)	30,610	40,779	41,530
Standard Deviation, Income of Kernels	24,113	40,091	33,705
Mean, Log Income of Kernels	9.943	10.221	10.220
Standard Deviation, Log Income of Kernels	1.141	.965	1.251
Mean, Maximum Income in Cluster (current $)	61,938	77,248	88,540
Mean, Minimum Income in Cluster (current $)	8,773	12,140	12,771
Per Cent in Northeast	25.4	25.4	28.1
Per Cent in Northwest	22.8	22.1	21.9
Per Cent in South	28.0	23.4	25.5
Per Cent in West	23.8	28.1	24.5
Per Cent in Central Cities	44.8	43.4	41.0
Per Cent in Suburbs (Urbanized Areas)	40.5	40.6	44.2
Per Cent in Suburbs (Other Urban Areas)	3.9	6.6	5.2
Per Cent in Rural and Non-metropolitan Areas	10.8	9.4	9.5

TABLE 2: Neighborhood Sorting: 1985 AHS

Models	Model 0		Model 1		Model 2		Model 3		Model 4		Model 5	
Obs	232	232	232	232	232	232	232	232	232	232	232	232
LLF	-4274	-4203	-4121	-4088	-4116	-4077	-4103	-4039	-3968	-3938	-3961	-3925
Geo. Dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
mean	9.893	9.565	9.870	9.466	9.943	9.812	9.900	9.581	9.881	9.488	9.943	9.812
kernels	(636.89)	(114.00)	(330.95)	(66.38)	(100.30)	(15.78)	(561.55)	(101.00)	(293.22)	(59.32)	(100.29)	(15.75)
variance	1.668	1.578	1.722	1.623	1.297	1.154	1.767	1.672	1.821	1.719	1.297	1.154
kernels	(172.74)	(158.60)	(66.81)	(65.66)	(24.67)	(19.23)	(161.09)	(147.37)	(60.50)	(59.58)	(24.37)	(19.12)
mean					9.888	9.540					9.896	9.55
neighbors					(233.69)	(41.17)					(228.12)	(40.55)
variance					1.705	1.617					1.808	1.718
neighbors					(74.08)	(70.87)					(70.10)	(66.70)
correlation			.391	.349	.353	.311			.387	.345	.342	.302
coefficient			(28.90)	(21.94)	(23.18)	(17.21)			(25.24)	(19.18)	(20.81)	(15.45)

TABLE 3

Neighborhood Sorting, Model 5: 1985, 1989 and 1993 AHS

Year	1985	1989	1993
Obs	232	244	309
LLF	-3925	-3939	-5257
mean kernels	9.812 (15.75)	9.859 (29.05)	9.768 (14.48)
variance kernels	1.154 (19.12)	.860 (11.26)	1.511 (13.53)
mean neighbors	9.553 (40.55)	9.897 (67.26)	10.017 (84.40)
variance neighbors	1.718 (66.70)	1.503 (67.54)	1.590 (104.31)
correlation coefficient	.302 (15.48)	.300 (16.52)	.163 (9.15)

Regressions include dummy variables in the means for kernels and neighbors. The dummy variables are for: US Census regions (Northeast, Midwest and South), climatic regions (five categories), and degree of urbanicity (central city of metropolitan area, suburb of metropolitan area, and other suburb).

TABLE 4

Neighborhood Sorting, Model 5, 1985: Northeast, Midwest, South and West

Region	Northeast	Midwest	South	West
Obs	59	53	65	55
LLF	-1135	-756	-1100	-875
mean	9.730	9.816	10.182	9.696
kernels	(27.79)	(62.11)	94.48	(48.28)
variance	2.543	1.042	.677	1.807
kernels	(7.38)	(4.26)	(4.60)	(14.78)
mean	9.848	9.881	9.941	9.662
neighbors	(66.55)	(145.39)	(116.12)	(127.80)
variance	2.738	1.050	1.956	2.034
neighbors	(19.60)	(21.83)	(23.03)	(47.94)
correlation	.378	.389	.389	.350
coefficient	(6.50)	(8.77)	(8.96)	(14.48)

TABLE 5

Neighborhood Sorting, Model 5, 1993: Northeast, Midwest, South and West

Region	Northeast	Midwest	South	West
Obs	87	68	79	76
LLF	-1697	-1173	-1124	-1099
mean	10.052	10.318	10.357	10.182
kernels	(20.88)	(91.57)	(88.34)	(70.31)
variance	3.333	.752	.750	1.033
kernels	(6.72)	(4.85)	(6.01)	(6.28)
mean	10.288	10.233	10.286	10.313
neighbors	(184.75)	(178.07)	(122.63)	(165.78)
variance	2.229	1.810	1.193	1.099
neighbors	(55.40)	(54.00)	(20.23)	(27.40)
correlation	.030	.172	.482	.385
coefficient	(1.08)	(4.83)	(12.46)	(10.28)

TABLE 6

Neighborhood Sorting, Model 5, 1985, by Urbanicity

Urbanicity	Central City (MSA)	Suburb (MSA)	Other Suburb	Rural and Non-metro
Obs	104	94	9	25
LLF	-1880	-1363	-106	-489
mean	9.696	10.242	10.311	9.710
kernels	(48.24)	(91.27)	(61.17)	(46.35)
variance	1.807	.730	.180	1.015
kernels	(14.65)	(7.77)	(1.00)	(2.74)
mean	9.667	10.179	10.074	9.717
neighbors	(126.10)	(208.43)	(83.38)	(65.11)
variance	2.157	1.087	.657	2.901
neighbors	(45.05)	(40.75)	(4.90)	(22.42)
correlation	.338	.319	.258	.213
coefficient	(13.02)	(10.59)	(1.57)	(3.19)

TABLE 7

Neighborhood Sorting, Model 5, 1993, by Urbanicity

Urbanicity	Central City (MSA)	Suburb (MSA)	Other Suburb	Rural and Non-metro
Obs	127	137	16	30
LLF	-2255	-2109	-251	-557
mean	10.132	10.348	10.127	10.056
kernels	(92.37)	(37.37)	(39.00)	(58.62)
variance	1.091	2.233	.674	.805
kernels	(8.96)	(10.42)	(2.06)	(3.12)
mean	10.072	10.523	10.111	10.156
neighbors	(191.96)	(459.39)	(77.97)	(79.79)
variance	2.004	1.057	1.338	2.468
neighbors	(49.19)	(81.28)	(15.24)	(19.73)
correlation	.270	.076	.241	.223
coefficient	(10.65)	(3.24)	(.03))	(2.70)

7 Appendix A: Proof of Proposition 1

Proof: Part A. The probability density function of income in neighborhood k is obtained as the density of y, times the probability that $\ell n\epsilon$ satisfies (7), given ℓny. By assumption, the distribution of $\ell n\epsilon$ conditional on ℓny, is normal with mean $\mu_\epsilon + \frac{\sigma_\epsilon}{\sigma_y}\lambda(\ell ny - \mu_y)$, and variance $\sigma_\epsilon^2(1-\lambda^2)$. Then the probability distribution is given by the marginal distribution of ℓny, weighted by the mass of the doubly truncated distribution function of $\ell n\epsilon$, with lower and upper bounds defined by Equ. (7). The auxiliary variables Ω_k and ω transform the truncation bounds in (7) in terms of the standardized variate $\frac{\ell n\epsilon}{(1-\lambda^2)^{\frac{1}{2}}\sigma_\epsilon}$. The probability that $\ell n\epsilon$ falls in that interval is given by $\Phi\left(Z_k(y)\right) - \Phi\left(Z_{k-1}(y)\right)$, which is a function of ℓny. For the resulting expression to be a density function it must integrate to 1, and thus the need for a normalizing constant G_k. Q.E.D. □

Proof: Part B. (Sketch) The possibly multiple modes of $f_k(\ell ny)$ are identified by the zeroes of the first derivative of $\ell n[f_k(\ell ny)]$. Taking logs of both sides of Equ. (9), differentiating the R.H.S. of Equ. (9) with respect to ℓny, and setting equal to 0 yields:

$$-\frac{\ell ny - \mu_y}{\sigma_y^2} + \frac{1}{\sqrt{2\pi}}\frac{e^{-\frac{1}{2}(Z_k)^2} - e^{-\frac{1}{2}(Z_{k-1})^2}}{\Phi\left(Z_k\right) - \Phi\left(Z_{k-1}\right)}\omega = 0.$$

We sketch the main intuition of the proof as follows. The first term contributes to the derivative a linear term with a negative slope:$-\frac{\ell nY - \mu_y}{\sigma_y^2}$. By inspection of the second term, we see that it contributes a complicated expression whose numerator is proportional to the difference of two standardized normal densities, centered at $\frac{\Omega_k}{-\omega}$ and $\frac{\Omega_{k-1}}{-\omega}$, respectively. Referring to Figure 1, this difference is magnified between the modes of those standardized densities. Therefore, it is possible that there are either one or three zeroes, of which the smallest and the largest correspond to local maxima of the density function $f_k(\ell ny)$, which would characterize the two modes of the distribution. Referring to Figure 1, we note that if $\omega < 0$, which holds if $\lambda > 0$ (income and the preference parameter ϵ are positively correlated), then the modes are indicated by the zeroes of the derivative associated with points M_1^- and M_2^-. Similarly, if $\omega > 0$, which requires that λ be negative and absolutely larger than ψ, then the modes are indicated by M_1^+ and M_2^+. If parameter values eliminate any of the roots, it would be the two smallest roots. That is, if there is a single root it is the one close to the mode of the national distribution. Q.E.D. □

8 Appendix B: Maximum Likelihood Estimations

Let $k = 1, \ldots, K$ denote neighborhoods, Y_k, log income of kernel, and Y_{kj}, $j = 1, \ldots, |I_k|$, the log incomes of neighbors of kernel k. For every neighborhood, define

$$\text{MAX}_k = \max_j \{Y_{kj}\},$$

$$\text{MIN}_k = \min_j \{Y_{kj}\}.$$

8.1 Model 0

The LLF defined for each kernel and its neighbors:

$$\text{LLF} = -\frac{1}{2}\left[\frac{(Y_k - \mu)^2}{\sigma^2} + \ell n[2\pi\sigma^2]\right]$$
$$-\frac{1}{2}\sum_{j=1}^{|I_k|}\left[\frac{(Y_{kj} - \mu)^2}{\sigma^2} + \ell n[2\pi\sigma^2]\right] \tag{17}$$

where μ, σ^2, are the unknown mean and variance, common for neighbors and kernel, to be estimated.

8.2 Model 1

The LLF defined for each kernel and its neighbors:

$$\text{LLF} = -\frac{1}{2}\left[\frac{(Y_k - \mu)^2}{\sigma^2} + \ell n[2\pi\sigma^2]\right]$$
$$-\frac{1}{2}\sum_{j=1}^{|I_k|}\left[\frac{(Y_{kj} - \mu - \rho(Y_k - \mu))^2}{\sigma^2(1-\rho^2)} + \ell n[2\pi\sigma^2(1-\rho^2)]\right] \tag{18}$$

where μ, σ^2, ρ are the unknown mean and variance, and correlation coefficient between neighbors and kernel to be estimated.

8.3 Model 2

The LLF defined for each kernel and its neighbors:

$$\text{LLF} = -\frac{1}{2}\left[\frac{(Y_k - \mu_{ke})^2}{\varsigma^2} + \ell n[2\pi\varsigma^2]\right]$$
$$-\frac{1}{2}\sum_{j=1}^{|I_k|}\left[\frac{(Y_{kj} - \mu_{ne} - \rho\frac{\sigma}{\varsigma}(Y_k - \mu_{ke}))^2}{\sigma^2(1-\rho^2)} + \ell n[2\pi\sigma^2(1-\rho^2)]\right] \tag{19}$$

where $\mu_{ke}, \mu_{ne}, \varsigma^2, \sigma^2, \rho$ are the unknown mean for kernels and for neighbors, and variance for kernels and for neighbors, and correlation coefficient between neighbors and kernel to be estimated.

8.4 Model 3

The LLF defined for each kernel and its neighbors:

$$\begin{aligned} \text{LLF} = &-\frac{1}{2}\left[\frac{(Y_k-\mu)^2}{\sigma^2}+\ell n[2\pi\sigma^2]\right] \\ &-\frac{1}{2}\sum_{j=1}^{|I_k|}\left[\frac{(Y_{kj}-\mu)^2}{\sigma^2}+\ell n[2\pi\sigma^2]\right] \\ &-\sum_{j=1}^{|I_k|}\ell n\left[\Phi\left(\frac{(\text{MAX}_k-\mu)}{\sigma}\right)-\Phi\left(\frac{(\text{MIN}_k-\mu)}{\sigma}\right)\right] \end{aligned} \tag{20}$$

where $\Phi(\cdot)$ denotes the standardized normal cumulative function, and μ, σ^2, are the unknown mean and variance, common for neighbors and kernel, to be estimated.

8.5 Model 4

The LLF defined for each kernel and its neighbors:

$$\begin{aligned} \text{LLF} = &-\frac{1}{2}\left[\frac{(Y_k-\mu)^2}{\sigma^2}+\ell n[2\pi\sigma^2]\right] \\ &-\frac{1}{2}\sum_{j=1}^{|I_k|}\left[\frac{(Y_{kj}-\mu-\rho(Y_k-\mu))^2}{\sigma^2(1-\rho^2)}+\ell n[2\pi\sigma^2]\right] \\ &-\sum_{j=1}^{|I_k|}\ell n\left[\Phi\left(\frac{(\text{MAX}_k-\mu-\rho(Y_k-\mu))}{\sigma(1-\rho^2)^{\frac{1}{2}}}\right)-\Phi\left(\frac{(\text{MIN}_k-\mu-\rho(Y_k-\mu))}{\sigma(1-\rho^2)^{\frac{1}{2}}}\right)\right] \end{aligned} \tag{21}$$

where $\Phi(\cdot)$ denotes the standardized normal cumulative function, and μ, σ^2, ρ are the unknown mean, variance, common for neighbors and kernel, and correlation coefficient between neighbors and kernel, to be estimated.

8.6 Model 5

The LLF defined for each kernel and its neighbors:

$$\text{LLF} = -\frac{1}{2}\left[\frac{(Y_k-\mu_{ke})^2}{\varsigma^2}+\ell n[2\pi\varsigma^2]\right]$$

$$-\frac{1}{2}\sum_{j=1}^{|I_k|}\left[\frac{(Y_{kj}-\mu_{ne}-\rho\frac{\sigma}{\varsigma}(Y_k-\mu_{ke}))^2}{\sigma^2(1-\rho^2)}+\ell n[2\pi\sigma^2(1-\rho^2)]\right] \tag{22}$$

$$-\sum_{j=1}^{|I_k|}\ell n\left[\Phi\left(\frac{(\mathrm{MAX}_k-\mu_{ne}-\rho\frac{\sigma}{\varsigma}(Y_k-\mu_{ke}))}{\sigma(1-\rho^2)^{\frac{1}{2}}}\right)-\Phi\left(\frac{(\mathrm{MIN}_k-\mu_{ne}-\rho\frac{\sigma}{\varsigma}(Y_k-\mu_{ke}))}{\sigma(1-\rho^2)^{\frac{1}{2}}}\right)\right]$$

where $\Phi(\cdot)$ denotes the standardized normal cumulative function, and $\mu_{ke}, \mu_{ne}, \varsigma^2, \sigma^2, \rho$ are the unknown means for kernels and for neighbors, and variance for kernels and for neighbors, and correlation coefficient between neighbors and kernel to be estimated.

9 Appendix C:

Comparison of Incomes between American Housing Survey and National Data by Regions, 1985 and 1993

Sample: Kernels and Neighbors

U.S.- designated statistics are obtained from the *Statistical Abstract of the United States* [U.S. Bureau of the Census (1987; 1995)] and apply to the entire U.S. and regions, as appropriate, and not just urban areas. U.S. median housing costs and property values also apply to the entire U.S. and regions and are obtained from the AHS [U.S. Bureau of the Census (1985; 1993)]. All other statistics are based on the author's own processing of the American Housing Survey data [U.S. Bureau of the Census (1996)].

Year	1985					1993				
Regions	All SMSAs	Mid West	North East	South	West	All SMSAs	Mid West	North East	South	West
Summary Statistics										
Mean Income ($)	29410	26658	31140	28934	30928	37490	34085	41001	35893	39470
CV Income	.846	.818	.858	.859	.827	.854	.849	.850	.874	.819
Median Income ($)	23000	21700	24000	22145	24565	28248	26000	30075	26312	30336
U.S. Mean Income ($)	29066	28149	31146	27044	31475	41428	39442	45319	38249	45284
U.S. Median Income ($)	23618	23551	25485	21397	25782	31241	31400	33747	28441	33739

10 References

Cressie, Noel (1996), "Change of Support and the Modifiable Areal Unit Problem," *Geographical Systems,* 3, 159–180.

Ellickson, Bryan D. (1971), "Jurisdictional Fragmentation and Residential Choice," *American Economic Association Papers and Proceedings,* 61, 334–339.

Epple, Dennis, and Glenn J. Platt (1998), "Equilibrium and Local Redistribution in an Urban Economy When Households Differ in Both Preferences and Incomes," *Journal of Urban Economics,* 43, January, 23–51.

Epple, Dennis, and Holger Sieg (1999), "Estimating Equilibrium Models of Local Jurisdictions," *Journal of Political Economy,* 107, 4, 645–681.

Hardman, Anna M. and Yannis M. Ioannides (1998), "Income Mixing and Housing in U.S. Cities: Evidence from Neighborhood Clusters of the American Housing Survey," presented at AREUEA meetings, Chicago, January, working paper, Tufts University.

Hoyt, William, and Stuart S. Rosenthal (1997), "Household Location and Tiebout: Do Families Sort according to Preferences for Locational Amenities?" *Journal of Urban Economics,* 42, 159–178.

Ioannides, Yannis M. (1999) "Residential Neighborhood Effects," working paper, Tufts University, May.

Ioannides, Yannis M. (2000) "Interactive Property Valuations," Tufts University working paper, December.

Ioannides, Yannis M., and Tracey N. Seslen (1999) "Neighborhood Income and Wealth Distributions," May, in progress.

Ioannides, Yannis M., and Jeffrey E. Zabel (2000) "Neighborhood Effects and Housing Demand," presented at the Brookings Conference on Empirics of Social Interactions, January; working paper, April.

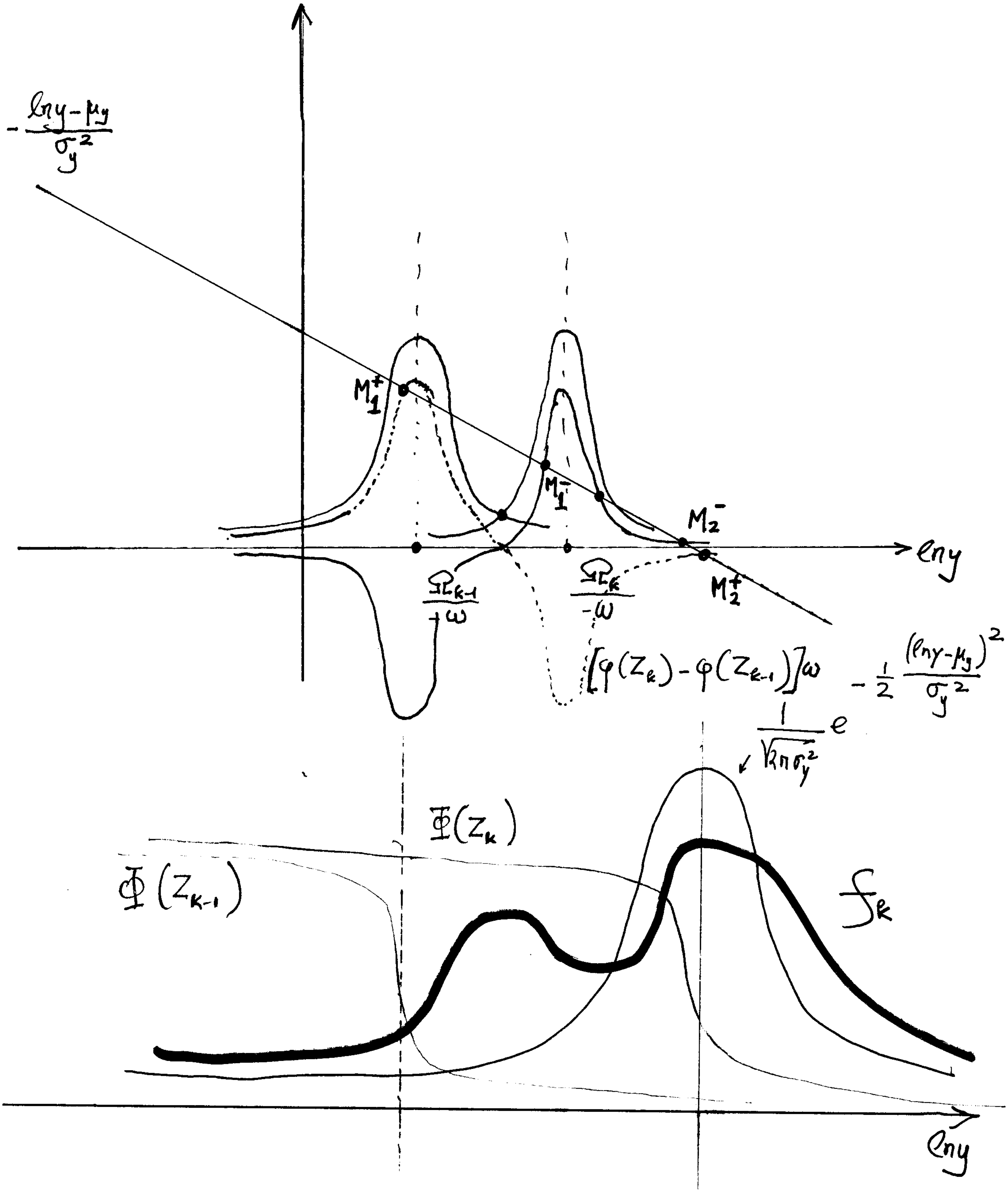
Kiel, Katherine, and Jeffrey Zabel (1998), "The Impact of Neighborhood Characteristics on House Prices: What Geographic Area Constitutes a Neighborhood?" working paper 98-04, Wellesley College, presented at the AREUEA meetings, Chicago, January.

Kremer, Michael (1997), "How Much Does Sorting Increase Inequality?" *Quarterly Journal of Economics,* February, 115–139.

Kremer, Michael, and Eric Maskin (1996), "Wage Inequality and Segregation by Skill," NBER working paper No. 5718, August.

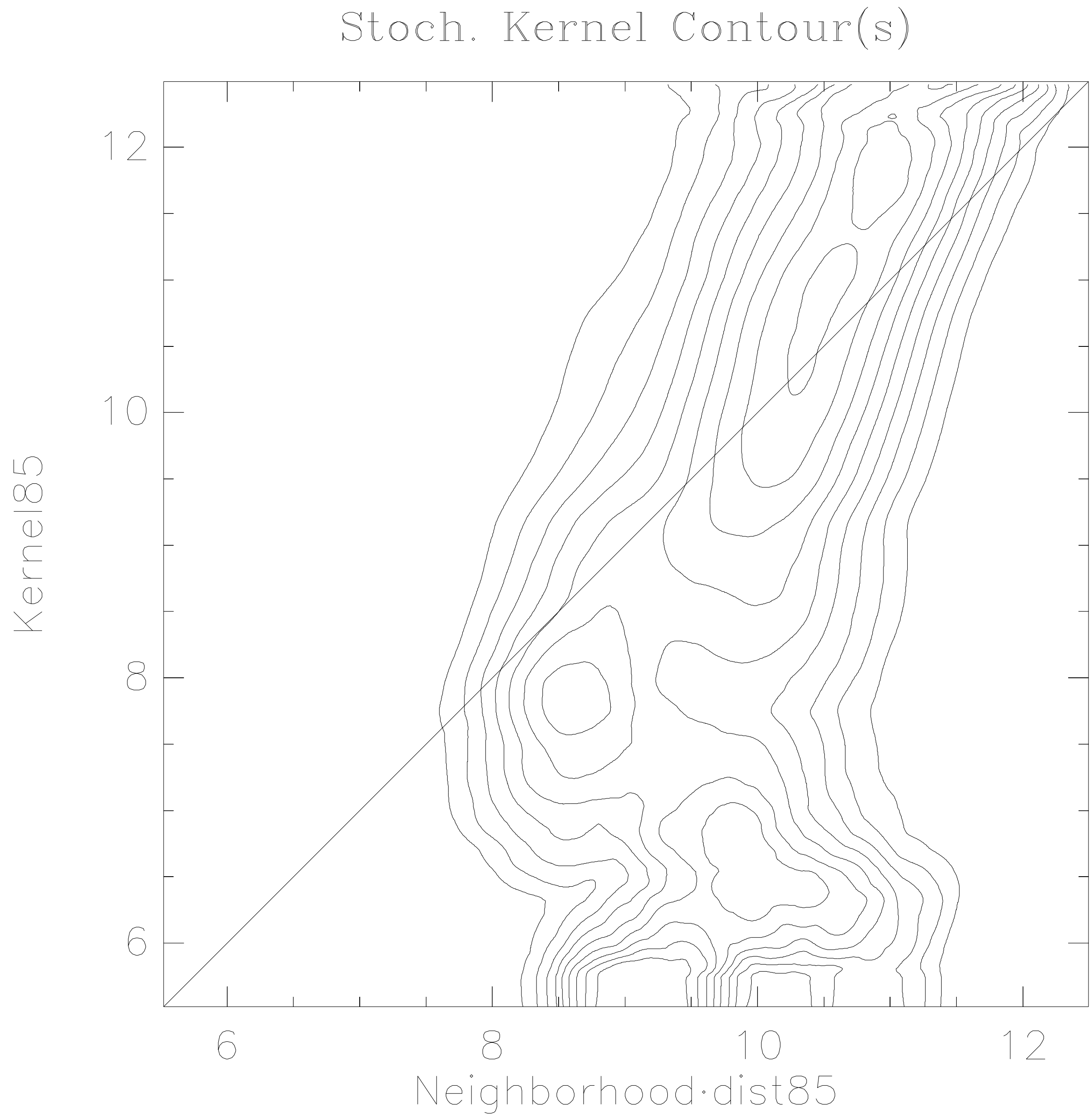
Mayer, Christopher J. (1996), "Does Location Matter?" *New England Economic Review,* May-June, 26-40.

Quah, Danny T. (1996) "Twin Peaks: Growth and Convergence in Models of Distribution Dynamics," *Economic Journal,* 106, 1045–1055.



$-\frac{\ln y - \mu_y}{\sigma_y^2}$
M_1^+
M_1^-
M_2^-
M_2^+
$\ln y$
$\frac{\Omega_{k-1}}{+\omega}$
$\frac{\Omega_k}{-\omega}$
$[\varphi(Z_k) - \varphi(Z_{k-1})]\omega$
$\frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{1}{2}\frac{(\ln y - \mu_y)^2}{\sigma_y^2}}$
$\Phi(Z_k)$
$\Phi(Z_{k-1})$
f_k
$\ln y$

FIGURE 1



Stoch. Kernel Contour(s)
12
10
8
6
Kernel85
6
8
10
12
Neighborhood·dist85

Figure 2

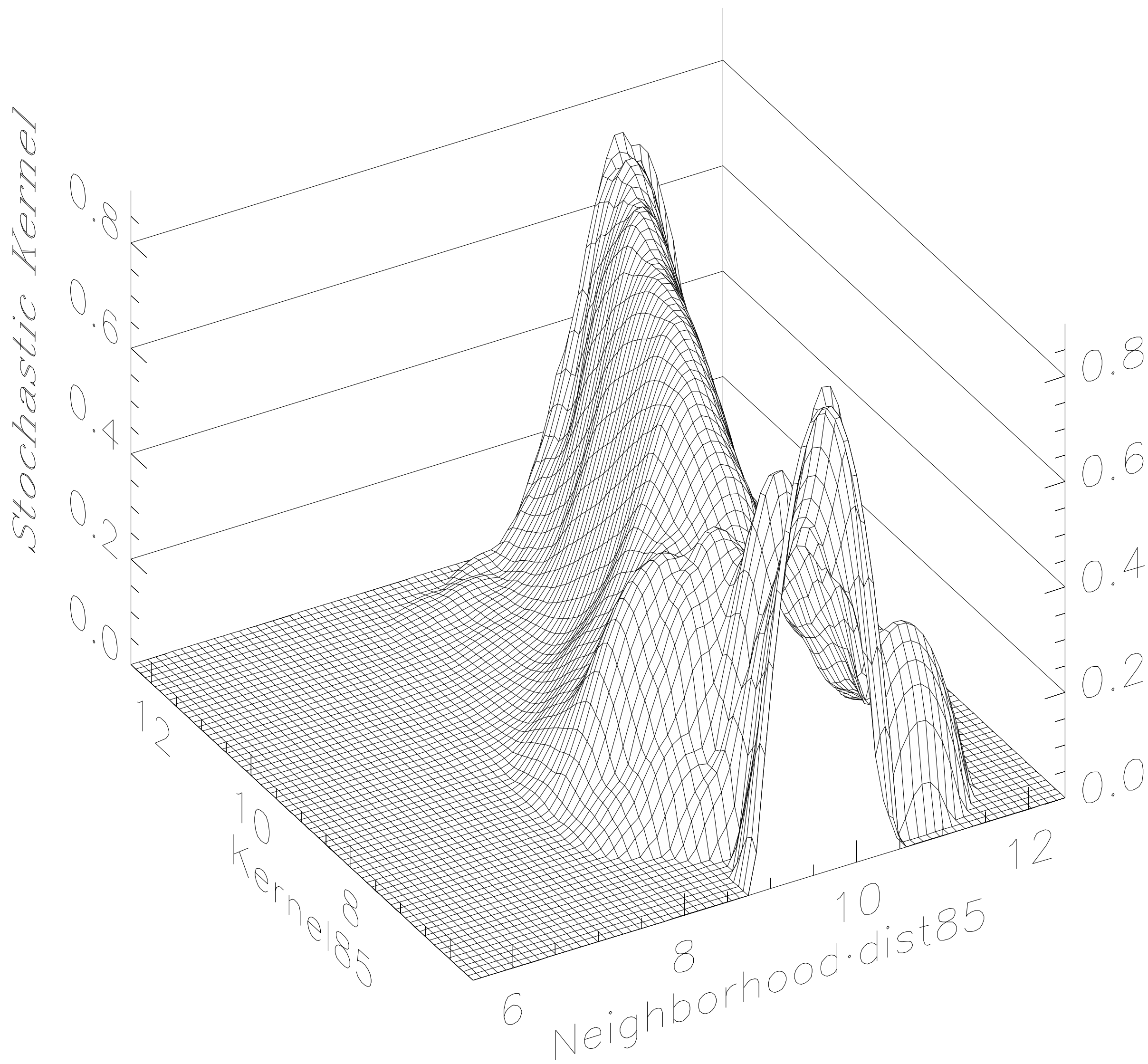

Stochastic Kernel
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
12
10
8
Kernel85
6
8
10
12
Neighborhood·dist85

Figure 3

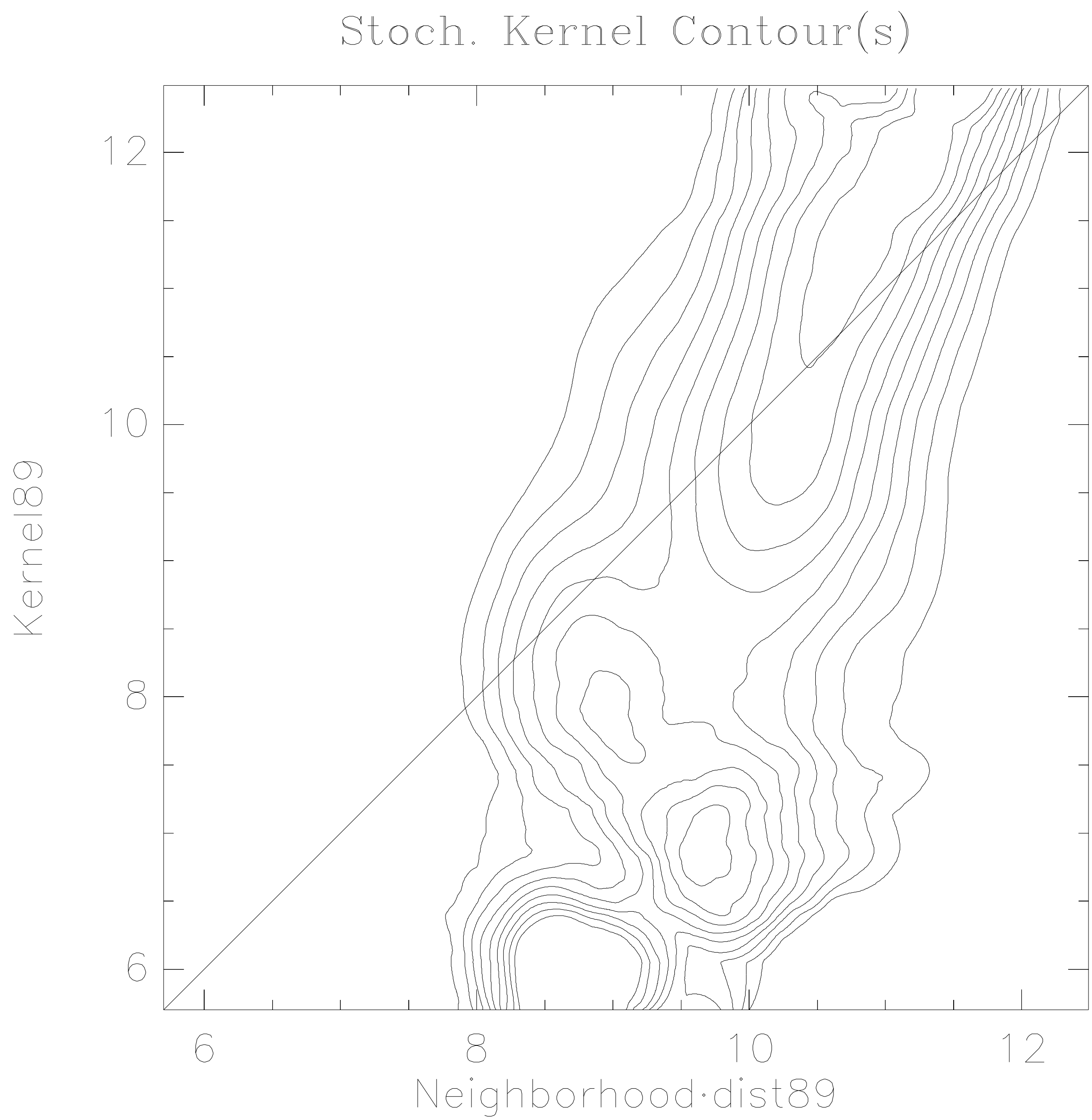

Stoch. Kernel Contour(s)
12
10
8
6
Kernel89
6
8
10
12
Neighborhood·dist89

Figure 4

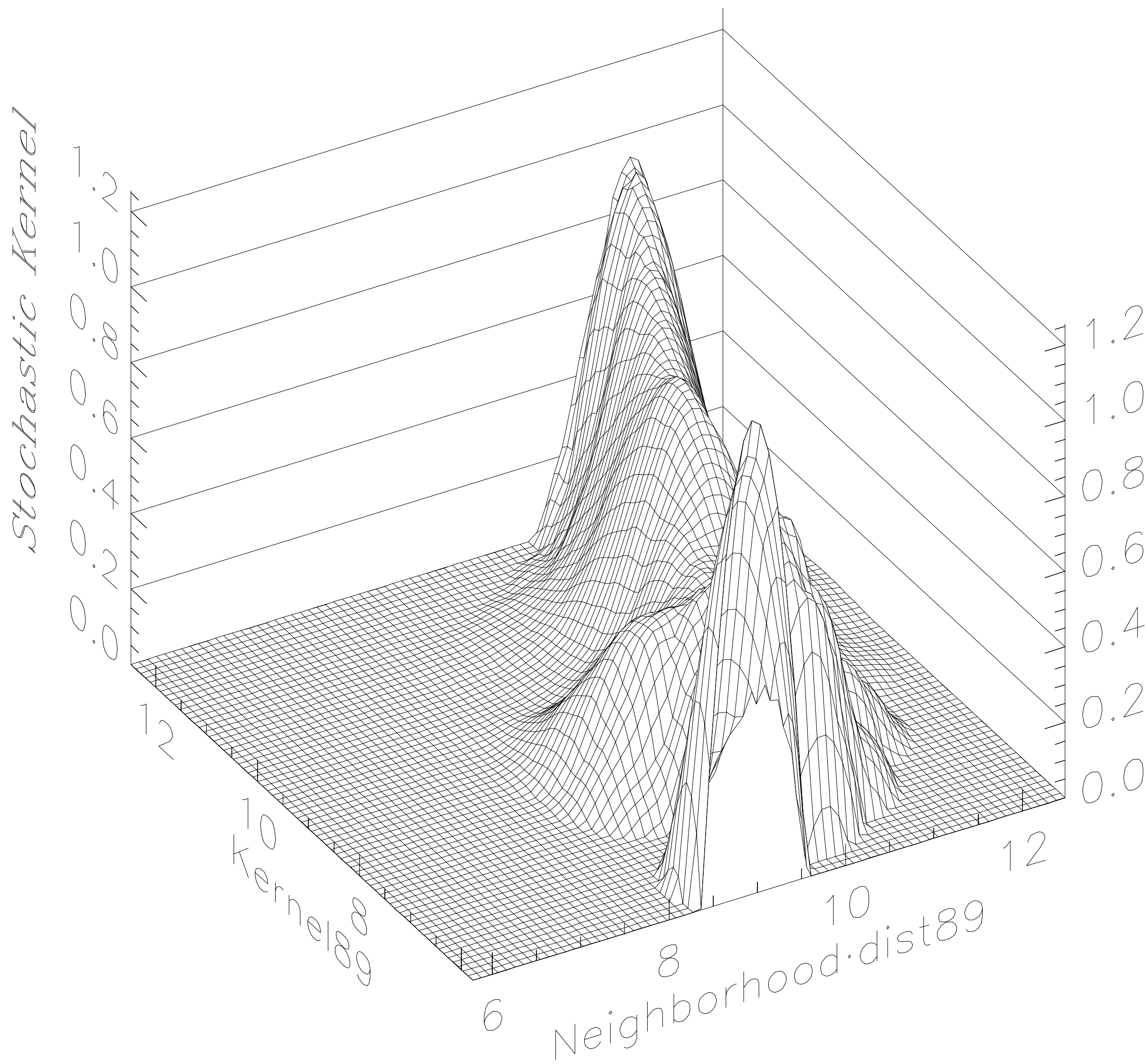

Stochastic Kernel
0.0
0.2
0.4
0.6
0.8
1.0
1.2
12
10
8
Kernel89
6
8
10
12
Neighborhood·dist89

Figure 5

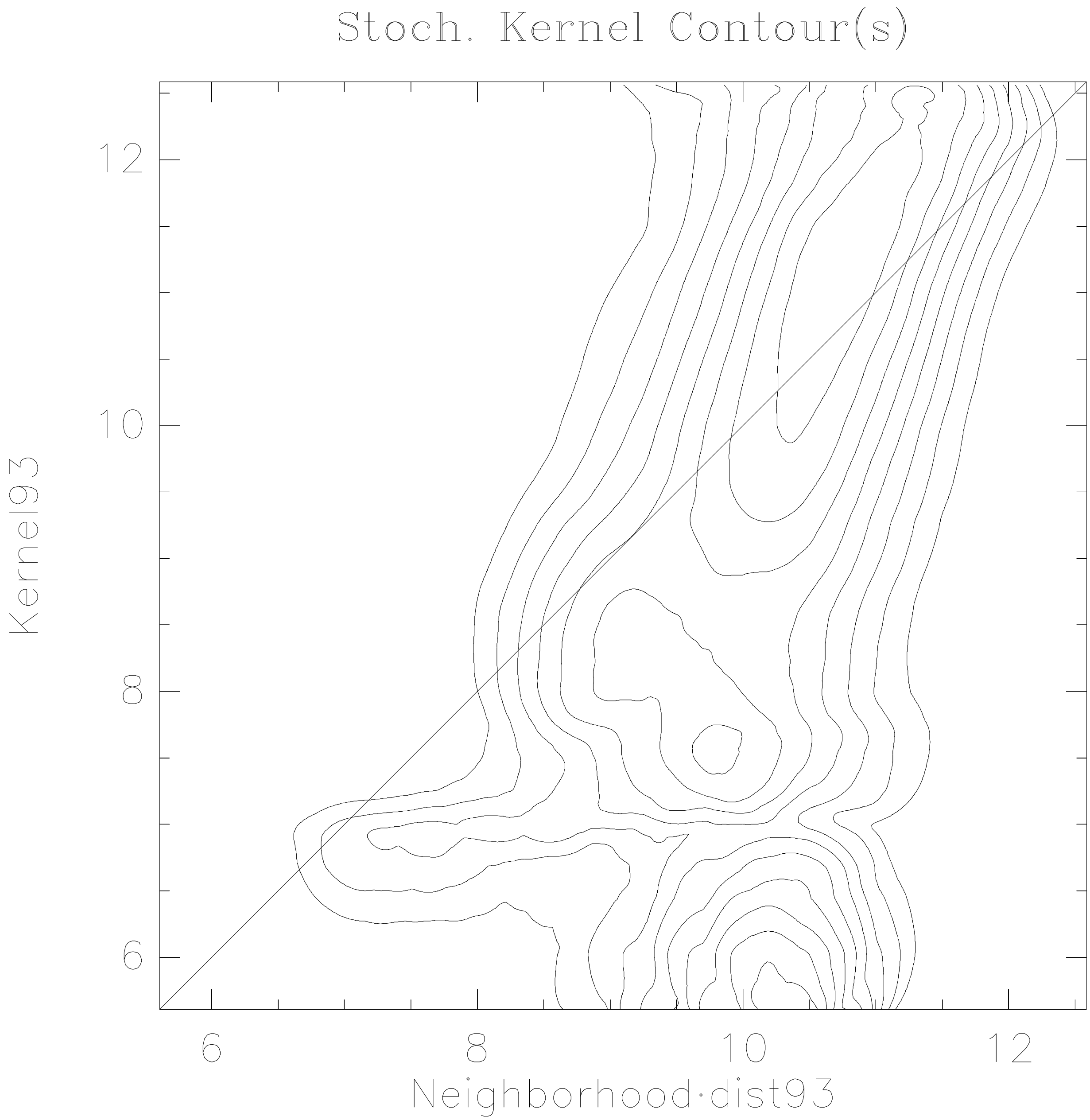

Stoch. Kernel Contour(s)
12
10
8
6
Kernel93
6
8
10
12
Neighborhood·dist93

Figure 6

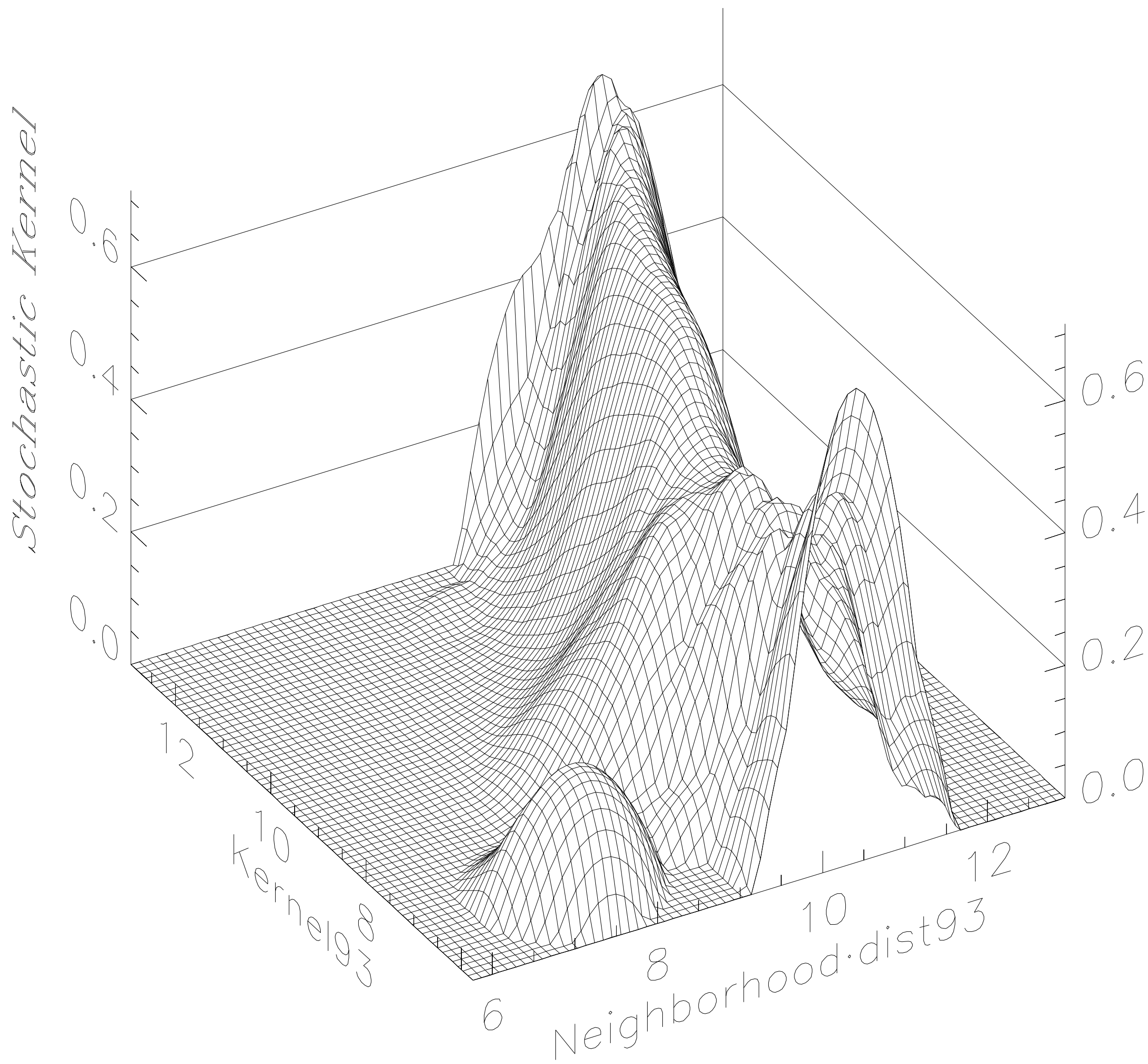

Stochastic Kernel
0.0
0.2
0.4
0.6
12
10
8
Kernel93
6
8
10
12
Neighborhood·dist93

Figure 7

WORKING PAPER SERIES 2001

ase.tufts.edu/econ/papers

2001-01 EGGLESTON, Karen. Multitasking, Competition and Provider Payment.

2001-02 IOANNIDES, Yannis M. Interactive Property Valuations.

2001-03 IOANNIDES, Yannis M. Neighborhood Income Distributions.

2001-04 IOANNIDES, Yannis M. Topologies of Social Interactions.