ECONOMIC RESEARCH REPORTS

SOCIAL INTERACTIONS, LOCAL SPILLOVERS AND UNEMPLOYMENT

by
Giorgio Topa

RR# 97-17 **March 1997**

C.V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, NY 10003-6687

Social Interactions, Local Spillovers and Unemployment

Giorgio Topa*
New York University

March 18, 1997

Abstract

This paper investigates whether or not social interactions and information spillovers in an urban context are important in determining one's employment status. I analyze a model which explicitly incorporates local interactions and allows agents to exchange information about job openings within their social networks. Thus agents are more likely to be employed if their social contacts are also employed and can therefore transmit information about potential job opportunities. The model generates a stationary distribution of unemployment that exhibits positive spatial correlations. Simulations of the model allow me to estimate its parameters via an indirect inference procedure. Using geographic distance as a proxy for social distance, I can test the model with Census tract data for the city of Chicago. I find a significantly positive level of social interactions across neighboring areas. This finding is robust to several controls for sorting and unobserved characteristics. The local spillovers are stronger for areas with poorer, younger and less educated workers, and with higher fractions of minorities.

JEL: J64, D31, C21, C63.

Keywords: Local Interactions, Spillovers, Unemployment, Indirect Inference.

*I would especially like to thank José Scheinkman, Lars Hansen and Gary Becker for their encouragement and guidance throughout this project. Comments from Roland Benabou, Tim Conley, Chris Flinn, Jim Heckman, Hide Ichimura, Pat Kehoe, Derek Neal, Tano Santos, and participants at several workshops at Chicago, the NBER Summer Workshop and NYU were very helpful. Financial support from the Reid Fellowship and the C.V. Starr Center at NYU is gratefully acknowledged. All remaining errors are of course mine. *E-mail: giorgio.topa@econ.nyu.edu*

1 Introduction

One of the most striking features of unemployment in Chicago in recent years is its geographic concentration in a few areas, mainly in the South and the West Side. The mean unemployment rate in the City of Chicago was 11.7% in 1980, but there were pockets of unemployment well above 20% in a few well-circumscribed areas, which accounted for about 14% of all Chicago Census tracts. The pattern worsens in 1990: now the mean unemployment rate is 14.9%, and it goes above 20% in *one quarter* of all Chicago tracts. A summary inspection of the map of unemployment suggests that census tracts with high levels of unemployment tend to be clustered together in geographically contiguous areas, rather than being spread around in a random fashion (see Figures (5) and (6)). In addition, the *change* in unemployment rates between 1980 and 1990 is also spatially correlated, as Figure (7) shows.[1] Again, the picture is rather bleak: almost two thirds of all Chicago tracts registered an increase in unemployment between 1980 and 1990; this increase was larger than five percentage points in three out of ten tracts.

The observed geographic "lumping" of unemployment seems consistent with recent work, both in economics and in sociology, that has stressed the role of local spillovers and neighborhood effects in a variety of phenomena, such as joblessness, teenage childbearing, dropping out of school, crime, poverty. Glaeser, Sacerdote and Scheinkman (1996) use a modified version of the "voter model" to explain the high variance of crime rates across U.S. cities. In their model, the agents' propensity to engage in criminal activities is affected by their neighbors' choices. These built-in local interactions generate enough covariance to fit the high variability of crime incidence. The paper is thus able to document empirically the existence of social interactions, to distinguish them from the effects of city characteristics and unobservables, and to give an estimate of the extent of social networks for different types of crime. Durlauf (1996) gives an excellent survey of how models from statistical mechanics can be brought to bear in economic contexts with discrete choice and local interactions. In the economic growth literature, Durlauf (1994) and Benabou (1993) incorporate local spillovers and neighborhood feedbacks into models that can exhibit poverty traps and persistent and widening income inequality.

Several sociological theories also postulate that "one's neighbors matter" in defining one's opportunities and constraints. The notion of neighbors is not necessarily restricted to physical proximity, but refers to closeness in terms of one's social network. Agents are not considered as isolated entities but rather as being part of networks of friends, relatives, acquaintances, neighbors, colleagues, that jointly provide cultural norms, economic opportunities, information flows, social sanctions and so on. Wilson

[1] The extent to which these variables are spatially correlated is made more precise in Section 5.

(1987) focuses on the way adults in a neighborhood influence young people by providing role models in terms of the value of education, steady employment and stable families. Crane (1991) concentrates on the effects of peer influences on teenagers' decisions regarding teenage pregnancy and dropping out of school. Coleman (1988) looks at how social networks can be seen as a sort of "social capital" since they provide valuable information, lower transaction costs, allow monitoring and enforcement of socially optimal outcomes.

An immense body of empirical papers has attempted to test the existence of neighborhood effects and to estimate their size (Jencks and Mayer (1990) give an excellent survey of the existing literature). Most of these papers set out to estimate the impact of neighborhood characteristics on individual outcomes, after controlling for a number of individual traits and for family background (see for example Corcoran et al. (1989), Case and Katz (1991) or Brooks-Gunn et al. (1992)). Crane (1991) also tries to document the existence of non-linearities in the way neighborhood quality affects individual decisions. He maintains that while there is virtually no neighborhood effect for affluent or middle-range neighborhoods, the very worst neighborhoods (in terms of a composite index of their quality) have a severe impact on the likelihood of teenagers dropping out of school or experiencing early parenthood. In recent work, Aaronson (1996) exploits data on siblings that have grown up in different communities to estimate neighborhood effects for educational outcomes. The use of sibling data allows him to better control for possible unobservables in the family background that may bias the estimation. He finds a negative and significant impact of the neighborhood's poverty rate and fraction of school drop-outs on the probability of graduating from high school.

More specifically on labor markets and unemployment, Granovetter (1974) and Corcoran, Datcher and Duncan (1980) (among others) have documented the importance of informal channels in finding jobs. They claim that more than 50 % of all new jobs are found through friends, relatives, neighbors or occupational contacts rather than through formal means. This is especially true for low-skill jobs, for first jobs and for black workers. Montgomery (1991), (1992) has a model in which both workers and firms prefer hiring through referrals rather than through formal channels because of an adverse selection problem. The employers cannot perfectly observe the quality of prospective employees and so rely on referrals from their high-ability workers. The basic assumption is that there exists inbreeding in agents' social networks, so that high-ability workers are more likely to refer individuals like themselves. If there exist different groups with a certain degree of inbreeding, one can observe persistent inequalities in wages and labor force participation across groups.

This paper tries to determine whether the spatial distribution of unemployment in

Chicago is consistent with a model with information spillovers and local interactions (throughout the paper, I will refer to this as the *structural* model). Taking inspiration from Montgomery's work, I assume that agents receive information about job openings from their social contacts. However, these contacts can only transmit information about employment opportunities if they themselves are employed. Thus an agent is more likely to find a job, the higher the number of neighbors who hold jobs.

The model is a version of the *contact process*, which was first studied by Harris (1974) in the context of interacting particle systems. In its simplest form, the contact process is a Markov process defined on a one-dimensional lattice (the integer line $\mathcal{Z}$). Each site i on the lattice is in one of two states, $\{0, 1\}$, and can change state according to the following transition rates:

$$\begin{array}{lll} 1 \rightarrow 0 & \text{at rate} & 1, \\ 0 \rightarrow 1 & \text{at rate} & \lambda(\eta(i+1) + \eta(i-1)), \end{array}$$

where $\lambda > 0$ is an "infection" parameter and $\eta \in \{0,1\}^{\mathcal{Z}}$ is the current configuration of the system. In other words, sites switch to state 1 (which I interpret as the *employed* state) at a rate that is increasing in the number of neighbors in state 1. This process generates positive spatial correlations between nearby sites on the lattice, therefore it seems like an appropriate tool to study the implications of local social interactions on the spatial distribution of unemployment.

In my structural model, the probability of finding a job can also depend on the agent's own characteristics, such as her level of education, independently of the unemployment level of her neighbors. This allows me to deal with the issue of positive sorting. This problem arises from the possibility that agents endogenously sort themselves into different neighborhoods on the basis of their neighbors' characteristics or because they have similar preferences over different consumption bundles (see Becker and Murphy (1994)). For example, more educated people may choose to locate in a given neighborhood because it has better schools; but at the same time, a higher level of education is associated with a higher probability of employment, so the spatial correlation of the education variable alone could drive the spatial correlation of unemployment. Similarly, certain contiguous areas might exhibit very high unemployment rates simply because unemployed people tend to reside in areas with low rent prices or high housing subsidies, and not because of any spillovers among people who hold jobs. So one would like to control for as many characteristics as possible, observed and unobserved, along which people may sort and that may give rise to positive spatial correlation of unemployment even in the absence of any spillovers. In addition, I let the strength of the local interaction (summarized by the λ parameter) be itself a function of agents' characteristics, in order to investigate the importance of local spillovers for agents with different education levels, race, age, etcetera.

The model is tested via the indirect inference method of Gourieroux, Monfort and Renault (1993), since it is not possible to characterize analytically the invariant distribution of the contact process described above, nor is it possible to write directly its likelihood function. The structural parameters are estimated indirectly, by minimizing the distance between the actual data and simulations of the structural model for different parameter values. In particular, one uses the parameters of an *auxiliary* model (more readily estimable than the structural one) to define a criterion function for the indirect estimation.[2]

I use a Spatial Auto-Regression of the sixth order (SAR(6)) as my auxiliary model. This seems to fit the spatial properties of the contact process quite well. In particular, by using a maximum likelihood criterion in the frequency domain, one can estimate the parameters of a SAR that best fit the contact process. This enables me to choose the specific form of SAR. In order to estimate the auxiliary model, I use the spatial GMM setup of Conley (1995). This method is an extension of the familiar time series GMM of Hansen (1982) to cross-sectional data, where the covariance structure is determined by economic distances. I adopt the covariance matrix estimator of Conley (1995) to allow for a general shape of the covariance matrix of the residuals (in space). The auxiliary model also addresses the issue of unobserved characteristics. In addition to controlling for several tract characteristics, I postulate an unobserved, tract-specific fixed effect that can affect employment outcomes in the Census tract. Then one can eliminate this particular fixed effect by first-differencing the data.

I estimate the structural model using Census tract data for the City of Chicago, in 1980 and 1990. Adopting tracts as units of observation still gives me a fine enough grid to be able to look at social interactions *within* neighborhoods, since tracts are smaller units than neighborhoods. The results of the indirect inference estimation support the model with built-in local interactions, and reject the hypothesis that one's employment status only depends on one's own characteristics, independently of the neighbors. The information spillovers are strictly positive both in 1980 and in 1990, but the size of the spillover effect is roughly two to three times as large in 1990 as in 1980 (depending on the specific experiment: see Table (4)). In addition, spillovers are stronger for less educated people and for non-whites.

Furthermore, the auxiliary model yields some interesting results in and of itself. Even controlling for tract characteristics (possibly unobserved), there exists a positive and significant spatial correlation between unemployment in one tract and the average

[2]An excellent introduction to indirect inference methods can be found in Tauchen (1996).

unemployment in the neighboring tracts. I am also able to observe how the interaction effect varies as a function of certain tract characteristics: the local spillovers are stronger for tracts that are younger, poorer, with less skilled workers or with lower education levels, and with a higher fraction of non-whites. These observations corroborate the results of the estimation of the structural model and are consistent with other empirical findings in the literature (Jencks and Mayer (1990), Granovetter (1974) and Corcoran, Datcher and Duncan (1980)).

The paper is organized as follows. Section 2 presents the structural model and its general properties. Section 3 describes the indirect inference methodology, the SAR(6) auxiliary model and the spatial GMM setup used for the auxiliary estimation. Section 4 reports the simulations of the structural model and the results of the indirect inference estimation. Section 5 contains some additional empirical results out of the estimation of the auxiliary model. Finally, Section 6 describes some desired extensions and concludes.

2 The structural model

The starting point of the analysis is that economic agents are embedded into social networks, within which they share information. In the first part of this Section I consider homogeneous agents, in order to isolate the simple information exchange story. The probability of agent i being employed depends positively on the information about job openings that i receives through her social contacts. But agents can transmit information about jobs only if they themselves are currently employed. Thus if i belongs to a network where nobody is employed, she will receive no useful information about employment opportunities through her contacts. In the second part, I introduce agent heterogeneity: this opens up a second channel through which agents can find and lose jobs independently of their contacts, based on their own individual characteristics. Such a framework also allows me to discuss the issue of sorting.

2.1 Homogeneous agents

I assume for the moment that agents are homogeneous in everything *but* their employment status, and that the only process taking place is the information exchange. The basic hypothesis that I would like to test can be formalized in the following way:

$$\text{Prob}(e^i_{t+1} = 1 | e^i_t = 0, \bar{e}_t{}^1) > \text{Prob}(e^i_{t+1} = 1 | e^i_t = 0, \bar{e}_t{}^2), \quad \text{with } \bar{e}_t{}^1 > \bar{e}_t{}^2 \tag{1}$$

where e^i_t is the employment status of agent i, (*1 = employed, 0 = unemployed*), and $\bar{e}_t$ is the employment rate of i's social network (or reference group).

The specific way in which I model the information interactions is through a discrete-time, finite-lattice version of the contact process.[3] Agents are arranged on a two-dimensional lattice and are indexed by a pair of integer coordinates, (i, j). The finite set of agents is $S \subset \mathcal{Z}^2$. Each agent can be in one of two states, *1 = employed, 0 = unemployed.* Thus the state of the whole system is a configuration $\eta \in \mathcal{X} \equiv \{0,1\}^S$. A distance d between any two points $(i,j),(k,l) \in S$ is defined as $|(i,j)-(k,l)| \equiv |i-k|+|j-l|$. I can then define a set of ***neighbors*** of agent (i,j) as:

$$N(i,j) = \{(i \pm 1, j), (i, j \pm 1)\},$$

i.e. the individuals who are located at a unit distance from (i,j). Thus $N(i,j)$ is one way to formalize the social network around agent (i,j). I also define $I_t(i,j)$ as the amount of information transmitted to (i,j) by her neighbors. In general, this can be an increasing function of the employment rate within the set of neighbors $N(i,j)$; for simplicity, I use the employment rate itself:

$$I_t(i,j) = \frac{1}{|N(i,j)|} \sum_{(k,l) \in N(i,j)} \eta_t(k,l).$$

The evolution of the system takes place in discrete time. Each period $t+1$ agents can change their employment state according to the following transition probabilities that depend on the state of the system at t:

$$\Pr(\eta_{t+1}(i,j) = 0 | \eta_t(i,j) = 1) = p, \tag{2}$$
$$\Pr(\eta_{t+1}(i,j) = 1 | \eta_t(i,j) = 0, \eta_t) = p\lambda I_t(i,j). \tag{3}$$

In equation (2) p is an exogenous probability of losing one's job, given that one is

[3]Glaeser et al. (1996) use a particular version of the voter model on a one-dimensional lattice. The basic voter model lets agents change actions occasionally according to their neighbors' choices. However this model exhibits unanimity (eventually either all agents will be engaged in crime, or all agents will not). In order to be able to get a finite variance of crime rates across lattices, the authors introduce as a modeling device some "fixed agents" whose opinions are independent of their neighbors. This allows them to estimate an index of social interactions and to provide some measure of the size of networks.

employed at t. This exogenous shock to employment is i.i.d. across the lattice, but the model could be modified to accommodate economy-wide shocks that are correlated across agents, business-cycle effects and so on. The λ parameter in equation (3) captures the "contagion" aspect of the model and defines how effective the information is in helping one find a job. So equation (3) says that the conditional probability of agent (i, j) being employed next period depends positively on the amount of information $I_t(i, j)$ transmitted by her neighbors, which in turn is increasing in the employment rate among these neighbors.

Notice that I have assumed that an employed agent who knows about a job opportunity will pass the information to her unemployed neighbors with probability one. But even if I set up a game in which agents can choose whether or not to transmit the information they possess, the strategy profile *"transmit"* for all agents is a Nash equilibrium. There is a basic insurance motive why agents will share their information: if I have a job now it is in my interest to tell my neighbors about job opportunities, so that in the future they will in turn help *me* find a job if I get fired.[4]

The behavior of the contact process has been extensively studied. Liggett (1985) gives an excellent treatment of this and other interacting particle processes; Andjel (1992), Liggett (1992), Durrett (1991) and Schonmann (1987) contain more recent results. The standard form of the contact process is a continuous-time, infinite-lattice Markov process on $\mathcal{X} \equiv \{0,1\}^S$, with $S = \mathcal{Z}^2$ (I stick to the two-dimensional case throughout this paper, but in general one can use lattices with any number of dimensions). The state-space $\mathcal{X}$ is still compact and has the product topology. Each point on the integer grid is a "site" that can be infected (state 1) or healthy (state 0). Infected sites recover at a constant exponential rate, that is normalized to one for simplicity. Healthy sites get infected at an exponential rate that is proportional to the number of infected neighbors:

$$\begin{array}{lll} 1 \rightarrow 0 & \text{at rate} & 1, \\ 0 \rightarrow 1 & \text{at rate} & \lambda \sum_{(k,l)\in N(i,j)} \eta(k,l). \end{array}$$

Given these transition rates, one can define a Markov process on $\mathcal{X}$ (details of the construction can be found in Liggett (1985), Ch.1). Let $\mathcal{P}$ be the set of all probability measures on $\mathcal{X}$, with the topology of weak convergence. With respect to this topology

[4]This can be made more precise by setting up a repeated non-cooperative game in which agents transmit information when they are employed, in the expectation of receiving useful information from their neighbors when they are unemployed. Using a framework adapted from Coate and Ravallion (1993), one can show that there exists a subgame-perfect equilibrium in which agents transmit information until someone defects, and then punish defection by withholding any information from then on.

$\mathcal{P}$ is compact since $\mathcal{X}$ is compact. Define also $C(\mathcal{X})$ as the collection of continuous functions on $\mathcal{X}$, regarded as a Banach space with $\|f\| = sup_{\eta \in \mathcal{X}} |f(\eta)|$. Let $\mu_0 \in \mathcal{P}$ be the initial distribution at $t = 0$ of the Markov process on $\mathcal{X}$. Then the distribution of the process $\eta_t, \quad \forall\, t > 0$ is denoted by $\mu_t \in \mathcal{P}$.

Definition 1. A measure $\nu \in \mathcal{P}$ is said to be *invariant* for the Markov process on $\mathcal{X}$ if $\mu_0 = \mu_t = \nu \quad \forall\, t \geq 0$. The class of all invariant measures $\nu \in \mathcal{P}$ will be denoted by $\Im$.

Result 1.
The set of invariant measures $\Im$ is not empty and is a compact convex subset of $\mathcal{P}$.
Proof: see Liggett (1985), p.10.

The set of invariant distributions of the system can be characterized as follows. There is a critical value λ_c such that for any $\lambda \leq \lambda_c$, the infection dies out with probability one as $t \to \infty$, whatever the initial configuration η. (The probability measure that puts all probability mass on the configuration $\eta \equiv 0$ is denoted by δ_0; similarly, δ_1 indicates the pointmass at $\eta \equiv 1$). If, on the other hand, $\lambda > \lambda_c$, then in addition to δ_0 there are other invariant distributions that give positive probability to configurations with at least one site infected. In particular, let ν_λ be the upper invariant measure for the contact process: $\nu_\lambda \equiv \lim_{t \to \infty} \mu_t$ given that $\mu_0 = \delta_1$. If $\lambda > \lambda_c$, then ν_λ is such that an initial infection on the lattice never dies out, with probability one.

The two most interesting results, from my point of view, characterize the correlations between different sites on the lattice under the extremal invariant distribution ν_λ. The first says that the states of any two sites are positively correlated, while the second states that this correlation decays exponentially with distance (proofs for these results are given in Liggett (1985), Ch. 6).[5]

Result 2.
The invariant measure ν_λ has positive correlations, in the sense that

$$E^{\nu_\lambda}\left(f\left(\eta\right)\, g\left(\eta\right)\right) \geq E^{\nu_\lambda}\left(f\left(\eta\right)\right)\, E^{\nu_\lambda}\left(g\left(\eta\right)\right)$$

for any functions $f, g \in M \subset C(\mathcal{X})$, where M is the class of all continuous functions on $\mathcal{X}$ which are monotone in the sense that $f(\eta) \leq f(\zeta)$ whenever $\eta \leq \zeta, \quad \eta, \zeta \in \mathcal{X}$. (Notice that I need to define a partial order on $\mathcal{X}$, compatible with the topology). An example of a function $f \in M$ is the fraction of sites in state one in a finite sub-lattice,

[5] In the particular version of the contact process that I adopt in what follows, there exists a unique stationary distribution where the infection can never die out. That is why I only consider the case of ν_λ.

so one could look at the spatial correlation between two finite subsets of sites (neighborhoods) on the lattice.

Result 3.
If one initializes the system at the configuration $\eta \equiv 1$, then there are constants K, β such that

$$|E^{\mu_t}(f(\eta)\ g(\eta)) - E^{\mu_t}(f(\eta))\ E^{\mu_t}(g(\eta))| \quad \leq \quad K\ e^{-\beta d(R_1,R_2)},$$

where $R_1, R_2 \subset S$, $d(R_1, R_2)$ is the distance between two regions R_1 and R_2 on the lattice, and f, g only depend on the coordinates in R_1 and R_2 respectively.

So the contact process enables me to transform a statement about conditional probabilities of employment (equation (1)) into a statement on spatial correlations. Therefore, I will be interested in studying the spatial correlations of the unemployment distribution. So far the notion of "space" has been the rather abstract one of a social network, where a distance refers to the social distance between any two individuals. Later I will characterize this distance in a geographic sense.[6]

In order to give an idea of the spatial implications of the contact process, Figure (1) contains a simulated outcome of the upper invariant distribution ν_λ defined above. The simulation is run on a lattice (30×30) of 900 sites. The lattice is a two-dimensional torus, so if one views it as a matrix of dimensions ($N \times N$), $N = 30$, then sites in rows 1 and N are adjacent neighbors, and so are sites in columns 1 and N. Figure (1) gives a visual representation of the "density" of employment over the lattice, calculated as a sort of moving average in space of employment.[7]

The plot represents contour lines, with thicker lines indicating a higher density of employment ("peaks"), and thinner lines indicating a lower density ("valleys"). The geographic lumping of unemployment into clusters is very evident. As a term of reference, Figure (2) plots the density of employment for a process that evolves in an i.i.d. fashion over time, with no built-in interactions. The latter model lacks any sort of spatial pattern and unemployment looks very much randomly distributed.

[6]Conley and Tsiang (1994) use a similar notion of "economic distance" to analyze the interrelatedness of local labor markets in Malaysia. They use travel time between locations to estimate economic distances.

[7]For each site (i,j) I plot the value $\eta(i,j) + \sum_{(k,l) \in N(i,j)} \eta(k,l)$, so I get a density that varies between zero and five.

Figure 1: Simulated density of employment over the lattice: contact process

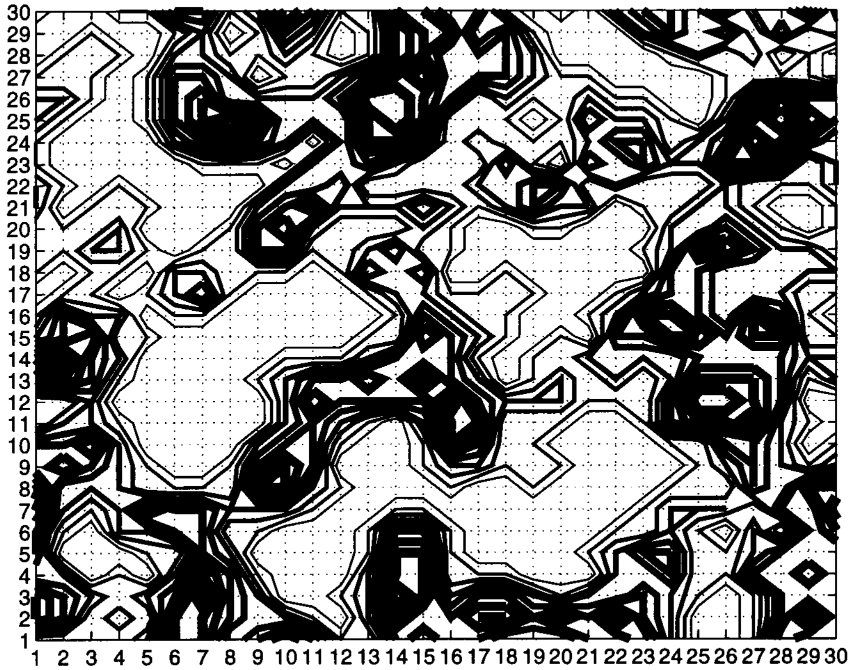
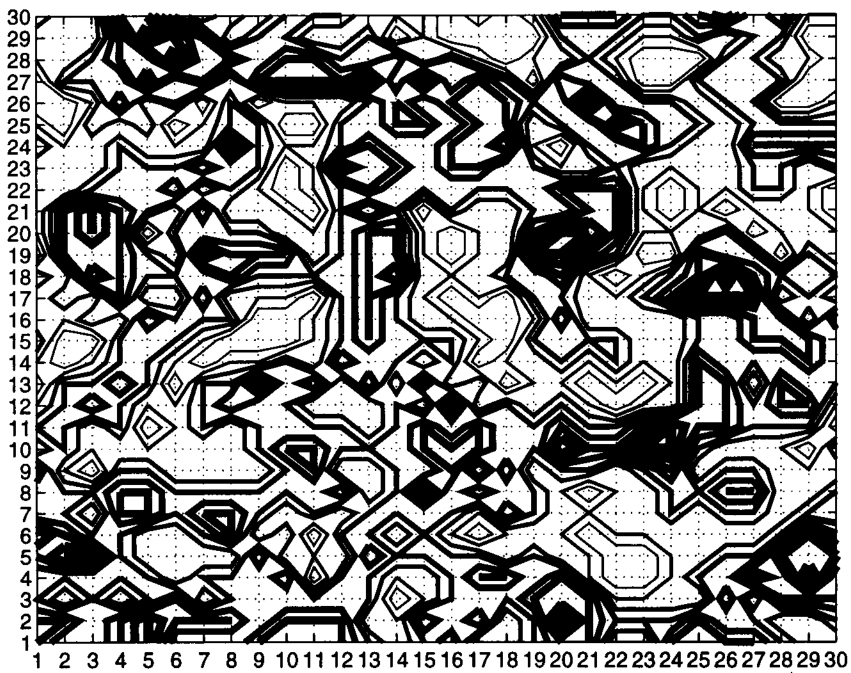


Figure 2: Simulated density of employment over the lattice: i.i.d. process



2.2 Heterogeneous agents

At this point I can complicate the model by letting agents be heterogeneous and by relaxing the assumption that the informal channels are the only means to get information about job openings. In the world of equation (1), I let agent i be defined by a set of characteristics X_i. These characteristics can also affect the individual's ability to find employment, independently of the information being acquired through one's social contacts. Then equation (1) becomes:

$$\begin{aligned} \mathrm{Prob}(e^i_{t+1} = 1 | e^i_t &= 0, \bar{e_t}^1, X_i) > \\ > \mathrm{Prob}(e^i_{t+1} &= 1 | e^i_t = 0, \bar{e_t}^2, X_i), \quad \text{with } \bar{e_t}^1 > \bar{e_t}^2. \end{aligned} \tag{4}$$

Notice that there is no time subscript on the X_i individual characteristics. In fact, I assume that these characteristics are more stable over time than the employment status indicator. This amounts to assuming that the process of losing and finding jobs takes place at a higher frequency than the process that determines the acquisition of these individual characteristics. In terms of the contact process, one way to include individual characteristics in the transition probability of equation (3) is the following:

$$\begin{aligned} \Pr(\eta_{t+1}(i,j) = 1 | \eta_t(i,j) = 0, \eta_t, X(i,j)) &= \\ p\Big(\lambda(X(i,j)) I_t(i,j) &+ \alpha(X(i,j))\Big), \end{aligned} \tag{5}$$

where $\lambda(\cdot)$ and $\alpha(\cdot)$ are both scalar functions of the vector of characteristics $X(i,j)$. Equation (5) implies that there are now two factors that affect one's probability of finding a job: one is the information exchange within one's network, the other one is a direct consequence of one's characteristics and does not depend on any interaction with one's neighbors. It is important to notice that the *strength* λ of the information exchange channel can also be influenced by the agent's characteristics. This allows me to estimate local spillovers in unemployment for different "types" of agents, in terms of their education, skills, race, age, and so on. Thus the estimation results can be compared to existing work in the literature on informal hiring channels.

I also let individual characteristics affect the probability of losing one's job, so I rewrite equation (2) as:

$$\Pr(\eta_{t+1}(i,j) = 0 | \eta_t(i,j) = 1, X(i,j)) = p\gamma(X(i,j)), \tag{6}$$

where $\gamma(\cdot)$ is another scalar function of $X(i,j))$. Notice that the interaction term $\lambda(X(i,j))I_t(i,j)$ appears in the conditional probability of *finding* a job but not in the conditional probability of *losing* a job (equation(6)): this is because I assume that the information interactions can affect employment opportunities, but do not play a role in the transition out of employment.[8]

Now I can discuss the implications of positive sorting. If agents sort into neighborhoods on the basis of their characteristics X, then one will observe a positive spatial correlation of the X's and, since these X's affect the probability of finding jobs, one will also observe a positive spatial correlation of unemployment, even if there are no information spillovers. One good example is education levels. Highly educated people may sort themselves into certain neighborhoods because they enjoy the company of other educated people, or because they attach great importance to school quality so they move to neighborhoods with good schools. On the other hand, one's education level positively affects one's employment opportunities.

In terms of Figure (1), agent heterogeneity implies that now the low-employment clusters are more likely to occur in certain areas of the map, depending on the distribution of characteristics along which people sort. In my simulation and estimation exercise I take as given the spatial distribution of the X's (determined by the sorting process), and conditional on this I run the local interaction process until it converges to the invariant distribution.[9] The structural parameters are estimated off the invariant distribution. The case in which all the spatial correlation is driven by sorting rather than by the local interactions corresponds to the case where $\lambda(\cdot)$ is identically zero for all values of the X's (in equation (5)), so the model delivers a very straightforward way to distinguish the two effects. I postpone the discussion of unobserved characteristics to the next Section.

One last step is necessary to define the structural model used in the indirect inference. The data I am going to use are defined at the Census tract level, whereas the model I have described so far is at the level of the individual agent. Since the focus of this paper is not on aggregation, I will side-step this issue by relabeling the particles on the lattice directly as tracts. There are two assumptions buried in this. One is that the space of one's social network is now interpreted as a geographic space, where one's social contacts are the physical neighbors. Thus I take geographic distance as a proxy for the social distance in terms of people's networks.[10] The second assumption is that the local interactions at the individual level translate into interactions across neighbor-

[8]This asymmetry in the transition probabilities helps towards the identification of the structural parameters, in the indirect inference estimation: more on this in Section 3.1.

[9]This is unique, as Proposition 1 below shows.

[10]One desired extension of this work is to think of different measures of social or economic distance, that track agents' networks more closely and take into account the costs of maintaining one's contacts and of transmitting information within the network.

ing *tracts.* The latter assumption is not very strong, since one can always redefine the radius or the strength of the interaction of the contact process at the individual level to generate a given level of interaction at the tract level.

Therefore the set S of sites on the lattice now refers to tracts. The state of each site needs to be redefined, since the relevant variable is now the employment rate within a tract, which varies continuously between zero and one. For computational reasons, I divide the interval $[0,1]$ into K equally spaced points. So the possible states for each tract (i,j) are now $\{e_1,\ldots,e_K\}$, where $e_1 = 0$, $e_K = 1$ and $e_k - e_{k-1} = 0.1$. The transition probabilities are slightly more complicated but maintain the properties of the two-state model. If the state of tract (i,j) at time t is one, then the only possible transition is to the next lower employment rate:

$$\Pr(\eta_{t+1}(i,j) = e_{K-1}|\eta_t(i,j) = 1, X(i,j)) \;=\; p_d\,\gamma(X(i,j)); \tag{7}$$

similarly, if the state at t is zero, the only possible transition is to the next higher employment rate:

$$\begin{aligned} \Pr(\eta_{t+1}(i,j) = e_2|\eta_t(i,j) = 0, \eta_t, X(i,j)) \;&=\; \\ = p_u\Big(\lambda(X(i,j))I_t(i,j) \;&+\; \alpha(X(i,j))\Big); \end{aligned} \tag{8}$$

finally, if the state at t is in the interior of the unit interval, then the employment rate can go up or down with probability $1/2$, and then the same transition rates defined above apply:

$$\Pr(\eta_{t+1}(i,j) = e_{k-1}|\eta_t(i,j) = e_k, X(i,j)) \;=\; \frac{p_d}{2}\gamma(X(i,j)); \tag{9}$$

$$\begin{aligned} \Pr(\eta_{t+1}(i,j) = e_{k+1}|\eta_t(i,j) = e_k, \eta_t, X(i,j)) \;&=\; \\ = \frac{p_u}{2}\Big(\lambda(X(i,j))I_t(i,j) \;&+\; \alpha(X(i,j))\Big). \end{aligned} \tag{10}$$

In all the above transition probabilities, the amount of information $I_t(i,j)$ now depends on the average employment rate in the neighboring tracts:

$$I_t(i,j) = \frac{1}{|N(i,j)|}\sum_{(m,q)\in N(i,j)} \eta_t(m,q)$$

where $|N(i,j)|$ is the number of tracts in the set of neighbors $N(i,j)$. Also, in each of the above cases, tract (i,j) stays put at the previous employment level with one minus the probability of an up or down transition.

This modified version of the contact process has a unique invariant distribution. This can be easily proved using standard results on Markov chains, since the model has a finite state-space and no longer has an absorbing state at $\eta \equiv 0$. Since the state-space $\mathcal{X} \equiv \{e_1, \ldots, e_K\}^S$ is finite, I can index each state by $h = 1, \ldots, H$ where H is the total number of states. Then a probability measure $\mu \in \mathcal{P}$ on $\mathcal{X}$ is simply a vector of probabilities μ_h, $h = 1, \ldots, H$. In particular, the evolution of the system is governed by the following rule:

$$\mu_{t+1} = P^\top \mu_t, \tag{11}$$

where P is the $(H \times H)$ transition matrix, whose entries p_{rs} denote the transition probabilities from state r to state s. These transition probabilities can in principle be calculated from the conditional transition rules (7) - (10). So then an invariant distribution is a vector ν s.t. $\nu = P^\top \nu$.

Proposition 1.
The finite-lattice, discrete-time contact process described by the transition rules (7) - (10) has a unique stationary distribution $\nu(X)$, for each choice of values of the X characteristics.

Proof: let the X characteristics be fixed. Then the transition probabilities (7) - (10) are given and define a Markov chain with transition matrix P over the finite state-space $\mathcal{X}$. This chain is irreducible and aperiodic, so I can apply Theorem 8.9 in Billingsley (1986) to prove that a stationary distribution $\nu(X)$ exists. Then Theorem 8.6 in Billingsley (1986) ensures that the stationary distribution is unique. Q.E.D.

For simplicity, I pick a linear specification for $\lambda(\cdot)$, $\alpha(\cdot)$, and $\gamma(\cdot)$:

$$\lambda(X(i,j)) = \lambda_0 + \sum_{l=1}^{L} \lambda_l X_l(i,j); \tag{12}$$

$$\alpha(X(i,j)) = \alpha_0 + \sum_{l=1}^{L} \alpha_l X_l(i,j); \tag{13}$$

$$\gamma(X(i,j)) = \gamma_0 + \sum_{l=1}^{L} \gamma_l X_l(i,j). \tag{14}$$

I also define as θ the vector of parameters of the structural model that I estimate through the indirect inference method:

$$\theta = [\lambda_0, \lambda_1, \ldots, \lambda_L, \alpha_0, \alpha_1, \ldots, \alpha_L, \gamma_0, \gamma_1, \ldots, \gamma_L]^\top, \quad \theta \in \Theta \subseteq \Re^p, \tag{15}$$

where p is the dimensionality of θ.

In order to test my hypothesis on the information spillovers I need to test whether or not at least some of the λ's are strictly positive: in other words, I need to test whether the positive spatial correlation of unemployment observed in the data can be completely explained away by the correlation of the X characteristics, through the α's and the γ's, or whether in fact at least some of this spatial correlation can be driven by the local interactions and the information exchange.

One final comment has to do with a possible objection to the model. I focus on the informational externalities that can facilitate a match between the supply and the demand for labor, thus raising the probability of finding employment. One alternative explanation for the clustering of high unemployment areas in the West and the South Side of Chicago is the so-called *spatial mismatch* hypothesis, which focuses on demand-side shocks that can affect the employment opportunities in certain areas. In particular, the argument is that during the 1970's and 1980's employers and therefore jobs have moved out of the city proper, and have relocated in the suburbs. On the other hand, inner-city residents (for example) may face high costs of commuting to jobs, and may have limited opportunities to move where the jobs are. If these conditions hold true, then the spatial mismatch between employers and people looking for jobs would cause high unemployment rates, especially in inner-city neighborhoods.

This hypothesis has been put to test by several authors. Ellwood (1986) uses Census tract data for the Chicago SMSA (Standard Metropolitan Statistical Area) and finds no significant evidence for the spatial mismatch. Inner-city residents appear to have very high mobility in terms of commuting to jobs, and several different measures of proximity to jobs are not significant in terms of explaining the variability of unemployment rates across different tracts. In addition, Chicago provides a convenient natural experiment: residents of the West and the South Side are very similar as far as their observable characteristics go, but the West Side is on average much closer to jobs than the South Side. However, this difference in distances to jobs does not seem to have any effect on the employment outcomes in the two areas.[11] In any case, the results

[11]The evidence on spatial mismatch remains contradictory. Holzer (1991) provides an excellent survey of this literature. To cite but one research, Ihlanfeldt and Sjoquist (1991) find a negative and

of this paper are not significantly affected by the inclusion of variables that measure access to jobs, such as commuting time to work (see Section 4.5). Furthermore, my auxiliary estimation strategy does allow for unobserved characteristics or fixed effects that may be specific to certain neighborhoods: as I explain in more detail in Section 3.3, I can control for these to some extent by first-differencing the data.

3 The indirect inference methodology

As I have discussed in the previous Section, I focus on the spatial, intra-city implications of modeling local interactions and spillovers. The goal is to estimate the parameters θ of the structural model presented above (equation (15)), in order to obtain an empirical measure of the information spillover effect.

I have already mentioned that it is not possible to write the likelihood function for the contact process. That is why, following an indirect inference strategy, I look for an auxiliary model that can fit the data well, but need not necessarily nest the structural model and may in fact even be misspecified. In particular I look for an auxiliary model that best approximates the spatial properties of the invariant distribution out of the contact process. Let ρ be the vector of parameters of the auxiliary model. Then the indirect inference procedure uses the estimates of ρ from the data and from simulations of the structural model to build a GMM-type criterion, that provides minimum chi-square estimators of the underlying parameters θ of the structural model.

The idea is the following. Let $\hat{\rho}$ be the estimated parameters of the auxiliary model based on the actual data. These depend on the *true* values θ_0 of the structural parameters: $\hat{\rho} = \hat{\rho}(\theta_0)$. Then one simulates the structural model for different values of θ in the parameter space Θ. For each θ, one can estimate the auxiliary model using the outcome of the simulations (in this case, a simulated unemployment variable). This estimation yields parameter estimates $\tilde{\rho}(\theta)$ that depend on the given choice of θ used for that particular simulation. The parameters of interest θ are estimated by minimizing the distance between $\hat{\rho}(\theta_0)$ and $\tilde{\rho}(\theta)$. In the remainder of this Section, I first of all present the indirect inference procedure, and then I describe the auxiliary model.

---

significant impact of distance from jobs on the employment probability of youth aged 16-19 years, using data from 43 SMSA's in the US.

3.1 The indirect inference estimation

The contact process that constitutes the structural model provides the following Data Generating Process:

$$y_{t+1} = \phi(y_t^N, x_0, \zeta_{t+1}),$$

where y_t is the outcome variable the researcher is interested in: here y_t is the n-dimensional vector of unemployment rates for all tracts $i = 1, \ldots, n$. x_0 is a $n \times M$ matrix of exogenous variables, and ζ_t is a vector of n i.i.d. shocks (i.i.d. across tracts and over time). As I have mentioned in Section 2, I use the invariant distribution generated by the contact process. In practice, I simulate the contact process for a given initial y_0, fixed exogenous x_0's and a sequence of shocks $\{\zeta_t\}_{t=1}^T$. I let the process run for a large enough T so that I am confident that the invariant distribution has been reached, and then draw a sample $\tilde{y}$ from the invariant distribution (more on this in Section 4.2).

Since I cannot readily write the likelihood function for the contact process, I use a Spatially Auto-Regressive auxiliary model as an approximation to the structural model. I discuss in Section 3.2 a procedure to determine how well the auxiliary model approximates the structural one, and decide which order of auto-regression one should use. The closeness of fit between the structural and the auxiliary models affects the efficiency of the indirect inference estimator (see Tauchen (1996), p.13). For now, I just assume that the auxiliary parameters $\rho \in \mathcal{R} \subset \Re^q$ come out of a spatial GMM estimation of the SAR(6) auxiliary model. When I perform this estimation on the actual data, I obtain:

$$\hat{\rho}_n = \arg\min_{\rho \in \mathcal{R}} J_n(y_n, x_n, \rho) \tag{16}$$

where (y_n, x_n) are the actual data and $J_n(\cdot)$ is the spatial GMM criterion used in the auxiliary estimation.

The auxiliary model is possibly misspecified, since I assume that the structural model is the true model. Tauchen (1996) shows that $\hat{\rho}_n \xrightarrow{a.s.} \bar{\rho}$, where $\bar{\rho}$ is the *pseudo-true* value given by:

$$\bar{\rho} = r(\theta_0). \tag{17}$$

In equation (17), θ_0 is the true value of θ, and $r : \Theta \longrightarrow \mathcal{R}$ is the so-called binding function (Gourieroux, Monfort and Renault (1993)), defined as follows:

$$r(\theta) \;=\; \arg\min_{\rho \in \mathcal{R}} \; J_\infty(G,\, \theta,\, \rho) \tag{18}$$

where $J_\infty(G, \theta, \rho) = \lim_{n\to\infty} J_n(y_n, x_n, \rho)$ and G is the distribution of x_n.[12]

Turning now to the simulations, for each value of $\theta \in \Theta$ I can draw H simulated realizations of y out of the invariant distribution of the structural model, $\tilde{y}_n^h(x_n, \theta)$, $h = 1, \ldots, H$. I then perform the auxiliary estimation on the simulated outcomes, for each value of θ and for fixed x's (from the data); this yields

$$\tilde{\rho}_n^h(\theta) \;=\; \arg\min_{\rho \in \mathcal{R}} \; J_n(\tilde{y}_n^h(x_n, \theta),\, x_n,\, \rho). \tag{19}$$

Again, for the simulated estimator of ρ, one has

$$\tilde{\rho}_n^h(\theta) \;\xrightarrow{\text{a.s.}}\; r(\theta) \quad \forall\, \theta \in \Theta,$$

and in particular for the true value θ_0,

$$\tilde{\rho}_n^h(\theta_0) \;\xrightarrow{\text{a.s.}}\; \bar{\rho} \;\equiv\; r(\theta_0). \tag{20}$$

So the idea of the indirect inference method is simply to evaluate $m_n^h(\theta) \;\equiv\; \hat{\rho} - \tilde{\rho}_n^h(\theta)$ for all $\theta \in \Theta$, and to pick the value $\theta^\star$ that minimizes this distance $m_n^h(\theta)$. The indirect inference estimator of θ is then the solution $\tilde{\theta}_n^H$ to the following minimum distance problem:[13]

$$\min_{\theta \in \Theta} \left[\hat{\rho}_n \;-\; \frac{1}{H}\sum_{h=1}^{H} \tilde{\rho}_n^h(\theta)\right]^\top \Omega_n \left[\hat{\rho}_n \;-\; \frac{1}{H}\sum_{h=1}^{H} \tilde{\rho}_n^h(\theta)\right]. \tag{21}$$

As is the case for a standard GMM procedure, the optimal weighting matrix Ω_n in the quadratic criterion (21) is

$$\hat{\Omega}_n \;=\; \hat{V}_n^{-1}, \tag{22}$$

[12] To be precise, $r(\cdot)$ is also a function of $G(\cdot)$, the distribution of x. Since it does not play a role in what follows, I will omit it for notational simplicity.

[13] This result, and the ones below on asymptotic properties and testing, are taken from Gourieroux, Monfort and Renault (1993).

where $\hat{V}_n$ is the estimator of the covariance matrix of ρ, which will be given in equation (42). The asymptotic distribution of the estimator $\tilde{\theta}_n^H$ is given by:

$$\sqrt{n}(\tilde{\theta}_n^H - \theta_0) \xrightarrow{\mathrm{d}}_{n\to\infty} N(0, Q) \tag{23}$$

where $Q = \frac{1+H}{H}\left\{\frac{\partial r(\theta_0)}{\partial\theta} V^{-1} \frac{\partial r(\theta_0)}{\partial\theta^\top}\right\}^{-1}$.

In addition, the criterion (21) provides a chi-square test for the specification of the structural model. The statistic

$$\kappa_n \equiv \frac{nH}{1+H} \cdot \min_{\theta\in\Theta}\left[\hat{\rho}_n - \frac{1}{H}\sum_{h=1}^{H}\tilde{\rho}_n^h(\theta)\right]^\top \hat{\Omega}_n \left[\hat{\rho}_n - \frac{1}{H}\sum_{h=1}^{H}\tilde{\rho}_n^h(\theta)\right] \tag{24}$$

is distributed as a χ^2 with $(q-p)$ degrees of freedom, where $q = \dim\rho$ and $p = \dim\theta$. Notice, in the definition of Q and in equation (24), the correction term $H/(1+H)$ due to the fact that I use H simulations for each value of θ.

The identification conditions for the structural parameters θ are fairly straightforward. From the definition of $\bar{\rho}$ and of the binding function in equations (17) and (18), one can write the first order conditions for minimization as:

$$\frac{\partial}{\partial\rho} J_\infty\left(\tilde{y}(x,\theta_0), x, \bar{\rho}\right) = 0. \tag{25}$$

In order for the structural parameters to be identified, one needs to assume that the true θ_0 is the only solution to equation (25). This amounts to requiring that the matrix of partial derivatives $\frac{\partial\rho}{\partial\theta}$ be full rank.[14]

Two things help ensure that this condition is satisfied: firstly, the conditional transition probabilities of the contact process are asymmetric, in the sense that the information interaction only affects the probability of an upward transition but not a downward one (see equations (7) - (10)); this helps distinguish between the λ's on the one hand and the α's and γ's on the other hand. Secondly, in the actual simulations I impose a symmetry restriction on the α and the γ parameters that are related to the exogenous tract characteristics (see page 31). Furthermore, I can check whether or not the identification condition is satisfied by running a very long simulation of the structural model at the estimated parameter values $\tilde{\theta}_n^H$ and numerically evaluating the matrix of partial derivatives $\frac{\partial\rho}{\partial\theta}$, calculated at the optimal values $\tilde{\theta}_n^H$. The structural

[14]This is equivalent to the standard rank condition for linear models. The equivalent of the *order* condition is simply that $q = \dim\rho$ be greater or equal than $p = \dim\theta$.

parameters are identified if this matrix is full rank (i.e., the rank must be equal to $p = \dim\theta$). This condition will be tested in Section 4.

Finally, Gourieroux, Monfort and Renault (1993) provide indirect tests of hypotheses on the parameters of interest θ. In particular, let θ be partitioned into $\theta = [\lambda^\top \ \alpha^\top \ \gamma^\top]^\top$, where λ, α, and γ each have dimension $(L+1)$ (see equations (12)-(14)). I can consider the null hypothesis $H_0 = (\lambda = 0)$: this amounts to testing whether the interaction effect in the structural model is identically zero. In order to perform a test I need to define the *constrained* indirect estimator $\tilde{\theta}_n^{0H}$ as the estimator that comes out of the minimization of (21), subject to the constraint $\lambda = 0$. The test statistic is defined as the difference between the constrained and the unconstrained optimum value of (21):

$$\begin{aligned} \kappa_n^C \ \equiv\ & \frac{nH}{1+H}\left[\hat{\rho}_n - \frac{1}{H}\sum_{h=1}^{H}\tilde{\rho}_n^h(\tilde{\theta}_n^{0H})\right]^\top \hat{\Omega}_n \left[\hat{\rho}_n - \frac{1}{H}\sum_{h=1}^{H}\tilde{\rho}_n^h(\tilde{\theta}_n^{0H})\right] - \\ & - \frac{nH}{1+H}\left[\hat{\rho}_n - \frac{1}{H}\sum_{h=1}^{H}\tilde{\rho}_n^h(\tilde{\theta}_n^{H})\right]^\top \hat{\Omega}_n \left[\hat{\rho}_n - \frac{1}{H}\sum_{h=1}^{H}\tilde{\rho}_n^h(\tilde{\theta}_n^{H})\right]. \end{aligned} \tag{26}$$

Gourieroux, Monfort and Renault (1993) prove that the test statistic κ_n^C is distributed as a χ^2 with $(L+1)$ degrees of freedom.

3.2 The choice of an auxiliary model

Since I am particularly interested in the spatial properties of the invariant distribution of the contact process, I need to look for an auxiliary model that well replicates these spatial characteristics. An obvious choice is a Spatially Auto-Regressive model (SAR), since Results 2 and 3 in Section 2.1 indicate that the invariant distribution of the contact process exhibits positive and exponentially decaying spatial covariances. Abstracting from tract heterogeneity in the X characteristics, the most simple auxiliary model one can write in two dimensions is the following SAR(1):

$$y(i,j) = \phi(y(i,j+1) + y(i,j-1) + y(i+1,j) + y(i-1,j)) + \epsilon(i,j) \tag{27}$$

where $y(i,j)$ is affected by the value that y takes at sites at a distance one from it, and the individual error terms $\epsilon(i,j)$ are i.i.d. in space (for higher order SAR's, one includes the realizations of y at distance 2, 3, 4 and so on). The issue then is to see how well this SAR(1) fits the structural model, and to decide what order of auto-regression one should use: for example, is a SAR(2) better than a SAR(1)?[15]

In order to address these questions, I need to find a criterion based on which I can see how close the auxiliary model is to the contact process. One possibility would be to compare Auto-Correlation Functions generated by a contact process and by a SAR(1) model. However, there is no clear sense in which one can assess the closeness of fit between the structural model and the auxiliary one. Therefore I use another approach, following the procedure used by Hansen and Sargent (1993), which involves working in the frequency domain. In particular, I can use Hansen and Sargent (1993)'s approximation criterion to estimate via maximum likelihood the parameters ϕ of the SAR (of a given order D) that best fit the true model, given by the contact process. Then I can repeat the maximum likelihood estimation for SAR's of different orders, to find out which order SAR best fits the original contact process.

I study a Contact Process in two dimensions, on a 30×30 lattice where all tracts are homogeneous.[16] The version of the contact process that I simulate is exactly the one described in Section 2.2, ruled by the transition probabilities of equations (7) - (10). To shut off tract heterogeneity, the X's are set at their city-wide mean values. A realization y drawn from the invariant distribution of this contact process has mean ν and spectral density $F(\omega)$ (at frequency $\omega \in [-\pi, \pi]$). Let $\mu(\phi)$ represent the mean of the approximating model, a SAR(D) of given order D, and let $G(\omega, \phi)$ denote its spectral density. The vector ϕ contains the free parameters of the SAR. Hansen and Sargent (1993) show that the maximum likelihood estimator of ϕ converges a.s. to the minimizer of the following criterion:

$$A(\phi) = A_1(\phi) + A_2(\phi) + A_3(\phi),$$

where

$$\begin{aligned} A_1(\phi) &\equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \, \det G(\omega, \phi) d\omega, \\ A_2(\phi) &\equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{trace} \left[G(\omega, \phi)^{-1} F(\omega) \right] d\omega, \end{aligned}$$

[15] One can also think of more general SARMA models to fit the contact process, but it turns out that a SAR is sufficient to approximate it quite well, so I keep to this class of auxiliary models to avoid the non-linearities involved in a SARMA.

[16] This yields a number of sites on the lattice roughly similar to the number of tracts in the data, 841.

Table 1: Parameter estimates and log likelihood for SAR(D)

ϕ parameters, log lik'd. $A(\phi)$									
Order	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	$A(\phi)$
$D=1$	0.7726	0	0	0	0	0	0	0	-1.1545
$D=2$	0.5292	0.3071	0	0	0	0	0	0	-1.1735
$D=3$	0.5297	0.3051	0.0018	0	0	0	0	0	-1.1735
$D=4$	0.5285	0.3053	0.0060	-0.0036	0	0	0	0	-1.1736
$D=5$	0.5274	0.3137	-0.0044	-0.0012	-0.0002	0	0	0	-1.1736
$D=6$	0.5256	0.2602	0.0821	0.0133	-0.0013	-0.0752	0	0	-1.1748
$D=7$	0.5214	0.2956	0.0267	0.0303	-0.0006	-0.0397	-0.0412	0	-1.1750
$D=8$	0.5283	0.2763	0.0216	0.0542	-0.0027	-0.0332	-0.0187	-0.0392	-1.1753

$$A_3(\phi) \quad \equiv \quad [\nu - \mu(\phi)]^\top G(0,\phi)^{-1}[\nu - \mu(\phi)].$$

To implement this, I simulate the two-dimensional contact process described above for a very long time (in order to make sure that the process has converged to the invariant distribution)[17] and then calculate its spectrum $F(\omega)$ using the sample periodogram. I then estimate the ϕ parameters by minimizing the criterion $A(\phi)$ above. For a given order D of the SAR, the relevant ϕ parameters are the auto-correlation coefficients at each distance d:

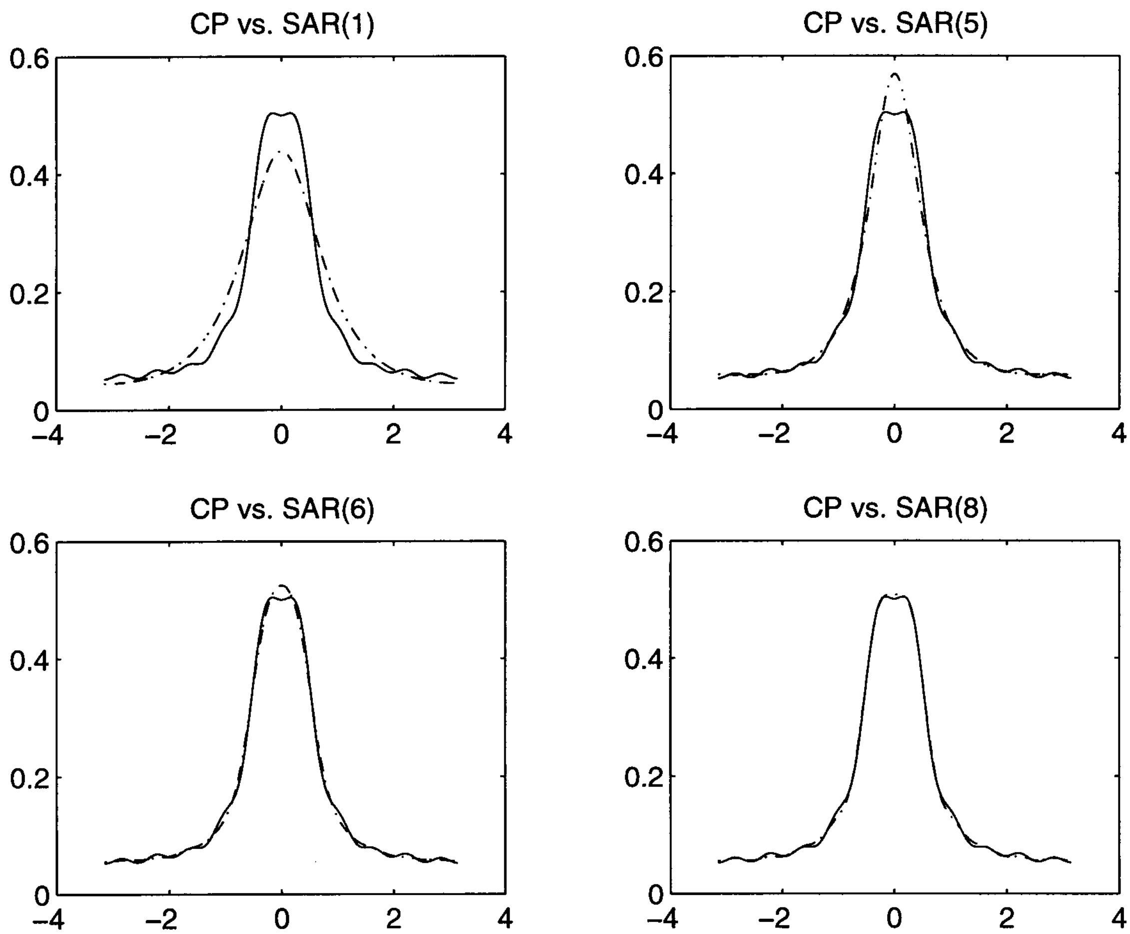
$$y_i = \sum_{d=1}^{D} \phi_d \, y_i^{N_d} + \epsilon_i \tag{28}$$

where $y_i^{N_d}$ is the average of y over tracts at a distance d from tract i.

The results of the estimation are contained in Table (1). It is clear that the fit increases as one raises the order D of the SAR. In this particular simulation, there is a large gain going from a SAR(1) to a SAR(2), then not much gain up to a SAR(5), and then again some improvement beyond a SAR(5). In order to give a visual impression of how well each SAR approximates the contact process, I report in Figure (3) the spectral densities of the same simulation of the contact process vs. SAR's of different orders. Visually, it appears that a SAR(8) fits the original contact process extremely well, but already a SAR(6) performs quite well. This pattern seems to hold (with

[17]See Section 4.2 below.

Figure 3: Spectra from CP (solid) and SAR's (dashed)



varying degrees of improvement as one raises the order of the spatial auto-regression) across different simulations for different values of the structural parameters. Therefore, in the following Indirect Inference estimation, I use a SAR(6) as my auxiliary model. This basic structure will be complicated in Section 3.3 by adding tract-specific characteristics and unobserved fixed effects.

3.3 The estimation of the auxiliary model

The results of Section 3.2 indicate that a SAR(6) is a good approximation to the invariant distribution of the contact process. I now need to complicate the auxiliary model in order to control for tract characteristics that may *both* affect the probability of finding and losing jobs, *and* be dimensions along which people sort into different neighborhoods. In addition, I wish to add unobserved fixed effects in the error terms in order to allow for possible unobservable tract-specific characteristics.

As a preliminary step, I am going to suggest the following regression for my auxiliary model:

$$y_{it} = \sum_{d=1}^{D} \phi_d y_{it}^{N_d} + x_{it}^{\top}\beta + \epsilon_{it} \tag{29}$$

where the superscript N_d refers to the average level of unemployment y in the tracts at a distance d from tract i. This is defined as $y_{it}^{N_d} \equiv W_i(d)y_t$, where $W_i(d)$ is the ith row of a weighting matrix $W(d)$, constructed in the following way. Each matrix $W(d)$ gives equal weights to the tracts that are at a distance d from each tract i. Tracts that share an edge with tract i on the map are considered to be at distance 1; the tracts that are adjacent to these immediate neighbors of i (but are not adjacent to i) are considered to be at distance 2, and so on. The weights in each row i of matrix $W(d)$ add up to one.

Ideally, if one has included all the "true" X characteristics along which people sort and that influence people's ability to find jobs, the parameter vector ϕ should pick up the portion of spatial correlation that is due to local interactions, while the residuals ϵ should exhibit no residual spatial auto-correlation due to unobservables. However, I am not going to make any assumptions at this point on the shape of the covariance matrix of the error terms: the residuals can be correlated across observations, since I am going to estimate the covariance matrix directly via the spatial GMM covariance estimators of Conley (1995).

In the structural model presented in Section 2 I let the local interaction parameters λ be themselves a function of tract characteristics, in order to allow for the possibility that the social network patterns and the intensity of the information exchange differ for agents of different types. In order to capture this in the auxiliary model, I complicate the SAR(6) of equation (29) by adding interaction terms between $y_{it}^{N_1}$ and some X characteristics. Let $\tilde{x}_{it}$ be a $J \times 1$ vector of X variables that are interacted with $y_{it}^{N_1}$ ($\tilde{x}_{it}$ is a subset of x_{it}). Let $\tilde{y}_{it}^N \equiv [y_{it}^{N_1} \; \dots \; y_{it}^{N_6} \quad y_{it}^{N_1} \tilde{x}_{it}^\top]^\top$. Then equation (29) can be rewritten as:

$$y_{it} \;=\; (\tilde{y}_{it}^N)^\top \phi \;+\; (x_{it})^\top \beta \;+\; \epsilon_{it}, \quad i = 1, \dots, n \tag{30}$$

where $\tilde{y}_{it}^N$ is a $(6+J) \times 1$ vector of cross-terms, x_{it} is an $M \times 1$ vector of exogenous variables; ϕ and β are $(6+J) \times 1$ and $M \times 1$ vectors, respectively. The following properties are assumed to hold in terms of the relationship between the RHS variables and the error terms:

$$E(x_{it}\epsilon_{it}) \;=\; 0 \quad \forall i, t; \tag{31}$$

$$E(\tilde{y}_{it}^N \epsilon_{it}) \;\neq\; 0 \quad \forall i, t. \tag{32}$$

Equation (32) implies that one needs to use instruments for those variables: one obvious choice is to use the exogenous variables in the neighboring tracts to i. In the estimation, I use observations in neighbors up to a distance 3 from tract i: so the instruments are $x_{it}^N \equiv [x_{it}^{N_1 \top} \; \dots \; x_{it}^{N_3 \top}]^\top$.

One last refinement of the auxiliary model is in order, before turning to the estimation strategy itself. One would like to allow for unobservable characteristics, on the basis of which agents may sort into different neighborhoods and that may induce positive spatial correlations of unemployment even in the absence of information spillovers. One way to model this is to include a fixed, tract-specific component in the error term:

$$\epsilon_{it} \;=\; \theta_i \;+\; u_{it}. \tag{33}$$

The θ_i term tries to capture features that are unobservable (to the econometrician) about a particular location, that may still attract or turn away people with certain characteristics that may be correlated with the ability to find jobs. The key assumption here is that these features are relatively more stable over time than the other variables (y_t, x_t), so that one can eliminate the θ_i term through first-differencing. Taking first differences of equation (30) one obtains:

$$\Delta y_{it} \;=\; (\Delta \tilde{y}_{it}^N)^\top \phi \;+\; (\Delta x_{it})^\top \beta \;+\; \Delta u_{it}, \quad i = 1, \dots, n \tag{34}$$

where the notation is $\Delta x_t \equiv x_t - x_{t-1}$, for any variable x in the model. In the actual estimation I only have two time periods, 1990 and 1980; therefore, for notational simplicity, I am going to drop the Δ and the time subscript in what follows. The model is to be taken in its first-difference specification from now on, unless otherwise specified.

Now in equation (34) the $\tilde{y}_i^N$ variates are still correlated with the error terms, so one needs to find instruments for them to estimate ϕ and β consistently. As I have already mentioned, I use the exogenous variables in the neighboring tracts to i, x_{it}^N. Therefore one can define the complete set of instruments as $z_i \equiv [x_i^\top \; x_i^{N\top}]^\top$. These instruments are assumed to be uncorrelated with the error terms:[18]

$$E(z_i u_i) \;=\; 0. \tag{35}$$

Equation (35) provides moment conditions that I can use to estimate the auxiliary model of equation (34) via a GMM procedure. Let π be the vector of parameters $[\phi^\top \, \beta^\top]^\top$. Then I can express the error terms explicitly in terms of the π's:

$$u_i(\pi) \;=\; y_i \;-\; (\tilde{y}_i^N)^\top \phi \;-\; x_i^\top \beta.$$

Finally, let $\xi_i \equiv [y_i^\top \; \tilde{y}_i^{N\top} \; x_i^\top \; x_i^{N\top}]^\top$ and $g(\xi_i, \pi) \equiv z_i u_i(\pi)$. Then the spatial GMM criterion for the estimation of the reduced-form model is:

$$J_n(\pi) \;=\; \left[\frac{1}{n}\sum_{i=1}^{n} g(\xi_i, \pi)\right]^\top C_n \left[\frac{1}{n}\sum_{i=1}^{n} g(\xi_i, \pi)\right]. \tag{36}$$

The GMM estimator of π is obtained via minimization of the quadratic criterion $J_n(\pi)$ as usual. Conley (1995) provides results for the asymptotic distribution of π_n^{GMM}, and gives conditions for consistency and asymptotic efficiency in this spatial context. In addition, he provides a way to estimate the covariance matrix of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(\xi_i, \pi)$ and hence to optimally choose the weighting matrix C_n. He proves that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(\xi_i, \pi) \;\rightarrow\; N(0, S)$$

[18] This assumption is a bit stronger than the one of equation (31): in addition to having current exogenous variables uncorrelated with current errors, it also requires that $E(x_{it} u_{it-1}) = 0$. This is perhaps not too strong an assumption because I will use observations ten years apart, so presumably the effect of a shock in a given year will have died out ten years later.

where the covariance matrix S can be consistently estimated by:

$$\hat{S}_n = \frac{1}{n} \sum_{d=0}^{\tilde{D}} K(d) \sum_{i=1}^{n} g(\xi_i, \pi_n^{GMM}) \cdot g(\xi_{i+d}, \pi_n^{GMM})^{\top}. \tag{37}$$

In equation (37), the subscript d refers to tracts that are at a distance d from tract i: again, tracts that are immediately adjacent to i (share an edge with i on the physical map) are at a distance $d = 1$. The neighbors of the neighbors of tract i are at a distance $d = 2$, and so on.[19] $\tilde{D}$ is a cutoff value for the spatial autocorrelations that needs to be fixed by the econometrician. $K(d)$ is a weighting function that varies with distance. The estimator in (37) is based on the assumption that the spatial process I am studying is *isotropic*: that is, on a two-dimensional lattice, the covariance between one site and its nearest neighbor to the North is the same as the covariance between that site and its nearest neighbor to the East, South or West.

The optimal choice of a weighting matrix is thus $C_n = \hat{S}_n^{-1}$. In this case, one obtains the result that

$$\sqrt{n}(\pi_n^{GMM} - \pi) \xrightarrow{\text{d}} N(0, V_n),$$

where V_n can be estimated by

$$\hat{V}_n = \left\{ \left[\frac{1}{n} \sum_{i=1}^{n} Dg(\xi_i, \pi_n^{GMM}) \right]^{\top} \hat{S}_n^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} Dg(\xi_i, \pi_n^{GMM}) \right] \right\}^{-1} \tag{38}$$

and $Dg(\cdot, \cdot)$ denotes the derivative of g with respect to π. So the asymptotic variance of the estimator π_n^{GMM} can be approximated by $\hat{V}_n/n$.[20]

One final complication is in order. As I mentioned earlier, this method estimates θ (the parameters of the structural model) by minimizing the distance between the pa-

[19] This is only one possible choice of distance: by changing the weighting matrix W one can accommodate different notions of distance, based on economic or social considerations.

[20] Operationally, the estimation technique is very simple: since the moment conditions g are linear in π, one can solve analytically for $\pi_n^{GMM}(C_n)$, for a given weighting matrix C_n. Thus the procedure is the following. One can start with any matrix, for example $C_n = I$, the identity matrix. So one can calculate $\pi_n^{GMM}(I)$; given this, one calculates $\hat{S}_n$, replaces $C_n = I$ with $C_n = \hat{S}_n^{-1}$, and iterates till convergence.

rameter estimates of the auxiliary model from the data, $\hat{\rho}$, and the auxiliary parameters estimated using the simulations for different values of θ, $\tilde{\rho}(\theta)$.

So far ρ consists only of the parameters π of the reduced form of the auxiliary model. Now I augment ρ by a certain number of moments of the unemployment variable. The mean and the variance are obvious choices. But in particular, since the structural model delivers implications in terms of the spatial correlations of unemployment, I also consider the spatial covariances, for different distances d, as additional parameters ρ for the indirect inference estimation.

These additional auxiliary parameters can be estimated together with π using the same GMM framework. Let ψ be the vector of moments to be estimated:

$$\psi = [\mu,\, \sigma^2,\, c_1, \ldots, c_D]^\top, \tag{39}$$

where $c_d = Cov(y_i,\, y_i^d)$, $\forall d = 1, \ldots, D$. Here y_i^d indicates the unemployment rate in tracts that are at a distance d from tract i. So the auxiliary parameters are now $\rho = [\pi^\top\, \psi^\top]^\top$. I also need to redefine $g(\cdot)$ as follows:

$$g(\xi_i, \rho) \equiv \begin{bmatrix} u_i(\pi) \cdot z_i \\ y_i - \mu \\ y_i^2 - \mu^2 - \sigma^2 \\ y_i\, y_i^1 - \mu^2 - c_1 \\ \cdots \\ y_i\, y_i^D - \mu^2 - c_D \end{bmatrix} \tag{40}$$

where ξ_i is now: $\xi_i \equiv [y_i\; \tilde{y}_i^{N\top}\; y_i^1\; \ldots\; y_i^D\; x_i^\top\; x_i^{N\top}]^\top$.

Then all the auxiliary parameters ρ can be estimated via a modified version of the GMM criterion of equation (36):

$$J_n(\rho) = \left[\frac{1}{n}\sum_{i=1}^{n} g(\xi_i, \rho)\right]^\top C_n \left[\frac{1}{n}\sum_{i=1}^{n} g(\xi_i, \rho)\right]. \tag{41}$$

Now the covariance matrix of ρ, V_n, includes the cross-covariances of the π and of the ψ parameters. This will be useful for the indirect inference. Equations (37) and (38) have to be rewritten accordingly in terms of the newly defined $g(\xi_i, \rho)$, augmented by the raw moments of y: in particular,

$$\hat{V}_n = \left\{ \left[\frac{1}{n}\sum_{i=1}^{n} Dg(\xi_i, \rho_n^{GMM})\right]^\top \hat{S}_n^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} Dg(\xi_i, \rho_n^{GMM})\right] \right\}^{-1}. \tag{42}$$

4 Indirect Inference results

In this section I return to the structural model presented in Section 2, to describe the data, the simulations and the results of the indirect inference estimation of the structural parameters θ.

4.1 The data

The data that I use come from the Summary Tape Files 3a of the 1980 and 1990 Census. These are summary statistics at the tract level for a wide set of variables, based on the 100% count data. Each Census tract constitutes a single observation.[21] Looking at data at the tract level might seem a very coarse approximation to the kind of interactions at the level of individual agents' networks that I am trying to detect (tracts in the City of Chicago had a mean population of $\sim$ 3400, with a std.deviation of $\sim$ 2500, in 1980). However, the tract-level grid is fine enough to look at interactions *within* and *across* neighborhoods. In fact, Census tracts are grouped into "Community Areas", which are supposed to have a distinctive identity as a neighborhood. In particular, a certain geographic space is considered a Community Area if it has "a history of its own as a community, a name, an awareness on the part of its inhabitants of common interests, and a set of local businesses and organizations oriented to the local community" (Erbe et al. (1980), p.xix). These criteria seem to hint at certain social interactions or networks within each area, so by looking at individual tracts I should still be able to pick up the kind of local interactions I am interested in, and to estimate their range.

There are 77 Community Areas in the City of Chicago, sub-divided into 863 Census tracts. So each Community Area has about 11 tracts on average. I concentrate on the City, rather than on the whole Standard Metropolitan Statistical Area (SMSA), because the city is more "dense" in population, so the approximation of social networks by spatial proximity should have more validity. Some tracts were dropped from the sample because of their zero (or near-zero) population and labor force. Even so, the sample contains 841 observations.

[21] A list of the names of the variables used and their content is given in the Appendix.

4.2 The simulations

The numerical simulations of the structural model were performed in order to deliver realizations of $\tilde{y}_n^h(x_n, \theta)$ to use in the estimation of θ. The simulations use the same transition probabilities that I have defined in Section 2, equations (7) - (10). In the simulations, the outcome variable y takes values in the interval $[0,1]$ and stands for the *employment* rate. Therefore, to make the simulated $\tilde{y}_n^h(x_n, \theta)$ comparable with $\hat{y}_n$ from the data, I operate the transformation $\tilde{y}_n^h(x_n, \theta) = 100 \cdot (1-y)$ on the outcome y of each simulation. I choose as a starting value for y the vector with all unit entries; i.e., I start from a configuration with full employment.[22] I also take the exogenous tract characteristics x from the actual data; these variables stay fixed at a given level throughout the simulation (either at their 1980 or their 1990 level, as I will explain in section 4.3).

Computational limitations imply that the dimensionality of the parameter space Θ has to be small, otherwise it becomes very hard to compute simulated y's for a grid of values of θ. Therefore I choose only two tract characteristics as my X's, namely the percentage of people with at least high school diplomas (*pchigh*) and the fraction of minorities in the tract (*pcnowhi*).[23] In addition, I impose symmetry restrictions on α and γ:

$$\begin{aligned}
\gamma_0 &= \alpha_0; \\
\gamma_{\text{pchigh}} &= -\alpha_{\text{pchigh}}; \\
\gamma_{\text{pcnowhi}} &= -\alpha_{\text{pcnowhi}}.
\end{aligned}$$

This means that θ is a 12-dimensional vector of parameters (if one allows them to be different in 1980 than in 1990):

$$\theta = [\lambda_0^{80}\ \lambda_0^{90}\ \lambda_{\text{pchigh}}^{80}\ \lambda_{\text{pchigh}}^{90}\ \lambda_{\text{pcnowhi}}^{80}\ \lambda_{\text{pcnowhi}}^{90}\ \alpha_0^{80}\ \alpha_0^{90}\ \alpha_{\text{pchigh}}^{80}\ \alpha_{\text{pchigh}}^{90}\ \alpha_{\text{pcnowhi}}^{80}\ \alpha_{\text{pcnowhi}}^{90}]^\top. \tag{43}$$

Starting from the configuration $y \equiv 1$, I let the process evolve according to the transition rules (7) through (10), keeping the X characteristics fixed at their initial values. The implicit assumption here is that the process through which agents exchange information and enter and exit unemployment takes place at a higher frequency than

[22]The initial value does not really matter, since there exists a unique stationary distribution.

[23]The spatial distributions of *pchigh* and *pcnowhi*, in levels as well as in first differences, are plotted in Figures (8) through (13).

the process that governs people's locational choices and hence the distribution of the X's across tracts.

At each iteration t, the amount of information $I_t(i,j)$ available to tract (i,j) is calculated using the employment rate in the tracts physically adjacent to (i,j) on the map of Chicago. I let the simulated process run for 1500 iterations. This seems a high enough number to let the contact process converge to its stationary distribution (to check this, I look at the behavior over time of the mean, the variance and the first six spatial auto-covariances of the simulated y's, and I let the simulation run till these moments reach "stationary" values). Then I run the simulation for 20 additional iterations and average y over these 20 "snapshots": this is the one realization $\tilde{y}_n^h(x_n, \theta)$ used for the indirect inference. It is essential of course to use the same sequence of shocks for all the different simulations with different values of θ. For computational reasons I run only one simulation for each value of θ, therefore $H = 1$.

In order to search over the parameter space Θ, I follow a multi-step procedure. In step one, I simulate the model over a coarse grid of parameter values for θ. I compute the value of the criterion in equation (21) for each point θ on the grid and choose the two points that yield the two lowest values of the criterion. As a second step, I conduct a finer grid search around each of the two previous minimizers, and discard the one "area" on the grid that performs more poorly. I then iterate on these two steps choosing finer and finer grids. As a final step, I pick a few of the best-performing values of θ and I minimize the criterion (21) *locally* around them, using a minimization routine in MATLAB that is based on the simplex algorithm. This procedure should ensure (at least to a certain extent) that I do not choose a local minimum of the criterion but I look over a broader set of values that are candidates for a global minimum.

The grid of values used for the first step of minimization of the indirect inference criterion (21) is:

$$
\begin{aligned}
\lambda_0^{80} &= [0 \quad .1 \quad .2 \quad .3] \\
\lambda_0^{90} &= [0 \quad .1 \quad .2 \quad .3] \\
\lambda_{\text{pchigh}}^{80} &= [-.15 \quad -.1 \quad -.05 \quad 0 \quad .1 \quad .2] \\
\lambda_{\text{pchigh}}^{90} &= [-.15 \quad -.1 \quad -.05 \quad 0 \quad .1 \quad .2] \\
\lambda_{\text{pcnowhi}}^{80} &= [-.1 \quad -.05 \quad 0 \quad .1 \quad .2] \\
\lambda_{\text{pcnowhi}}^{90} &= [-.1 \quad -.05 \quad 0 \quad .1 \quad .2] \\
\alpha_0^{80} &= [.6 \quad .7 \quad .8 \quad .9] \\
\alpha_0^{90} &= [.6 \quad .7 \quad .8 \quad .9] \\
\alpha_{\text{pchigh}}^{80} &= [.2 \quad .3 \quad .4 \quad .5]
\end{aligned}
$$

$$\begin{aligned}\alpha^{90}_{\text{pchigh}} &= [.2 \quad .3 \quad .4 \quad .5] \\ \alpha^{80}_{\text{pcnowhi}} &= [-.5 \quad -.4 \quad -.3 \quad -.2] \\ \alpha^{90}_{\text{pcnowhi}} &= [-.5 \quad -.4 \quad -.3 \quad -.2]\end{aligned}$$

.

This is admittedly a very coarse and limited grid, but it is the outcome of many preliminary estimations to try to define a reasonable region of the parameter space Θ. These parameter values still allow me to test the hypothesis that the interaction effect is strictly positive, since I consider the case in which all the λ parameters are zero. Notice that α_{pchigh} and α_{pcnowhi} lie in a subset of $\Re^+$ and $\Re^-$ respectively, because the outcome variable of the simulations is the *employment* rate, not unemployment.

4.3 Results of the indirect inference

The indirect inference procedure involves minimizing the distance between the auxiliary parameters $\hat{\rho}_n$ from the data and the auxiliary parameters $\tilde{\rho}^h_n(\theta)$ from the simulations. Therefore, as a first step, I report the results of the auxiliary estimation on the data in Table (2). This Table comes from estimating equation (34) via spatial GMM, where the local interaction variables $\tilde{y}^N_{it}$ are [*unempr(nbs-1), ... , unempr(nbs-6), unempr(nbs-1)* $\times$ *pcnowhi, unempr(nbs-1)* $\times$ *pchigh*], all in first differences (1990-1980). For the indirect inference I use the parameters marked with an asterisk, since I only consider *pchigh* and *pcnowhi* as my X's for the simulations of the structural model. A couple of comments on the auxiliary estimation itself are in order. The first thing to notice is that, even after controlling for tract characteristics, the unemployment rate of neighboring tracts has a positive and significant coefficient associated with it, at least for neighbors up to a distance two from tract i (the coefficients are not significantly different than zero for larger distances). Even though the parameters of the auxiliary model are not the actual parameters of interest, they still give an indication that unemployment is characterized by positive spatial correlations: this in turn is consistent with the model of local interactions.

Secondly, the auxiliary model passes the chi-square specification test (of the spatial GMM estimation). The minimized value of the quadratic criterion, properly scaled by n, is distributed as a χ^2 with $(m - q)$ degrees of freedom, where m is the number of moment conditions and q is the number of parameters to estimate. In this case there are 40 degrees of freedom, so the test-statistic is well below the rejection value, at the 95% confidence level (the p value of the test is actually .54!). This means that the

Table 2: Auxiliary regression on the data

Dep. Variable: unempl. rate, 1990-80			
Variable Name	ρ_{GMM}	S.E.	included?
unempr(nbs-1)	0.5300	0.1684	*
unempr(nbs-2)	0.4373	0.1280	*
unempr(nbs-3)	-0.0601	0.1586	*
unempr(nbs-4)	-0.4430	0.2533	*
unempr(nbs-5)	0.2564	0.4094	*
unempr(nbs-6)	0.0022	0.3560	*
u(nbs-1)× pcnowhi	0.0177	0.0109	*
u(nbs-1)× pchigh	-0.0234	0.0099	*
constant	0.5530	0.8442	
pc1824	-0.5124	0.1856	
segr	0.0083	0.0212	
pcnowhi	-0.0257	0.0509	*
pchigh	0.0768	0.0571	*
pperhh	0.0481	1.2603	
pcvac	0.1150	0.0511	
pcmnger	-0.2070	0.0631	
mgroren	-0.0089	0.0076	
hgvalue	0.0207	0.0132	
pcolf(m)	-0.0542	0.0304	
pcolf(f)	-0.1271	0.0369	
pcclge	-0.0898	0.0548	
pcfem	-0.1206	0.0730	
pc018	-0.4980	0.2185	
pc024	0.6251	0.2581	
pchisp	-0.0629	0.0258	
mean	2.9525	0.2609	*
variance	55.2852	8.4616	*
sp.cov(1)	16.2293	3.3922	*
sp.cov(2)	12.4444	2.1832	*
sp.cov(3)	9.3488	1.8389	*
sp.cov(4)	6.4059	1.5324	*
sp.cov(5)	5.9999	1.2941	*
sp.cov(6)	4.1570	1.3029	*
χ^2 test (40 d.f.) = 38.4375 (p-value = 0.54)			
adj. R^2 = 0.2002			
Box-Ljung test on $\hat{u}$ = 5.3776 (p-value = 0.50)			

moment conditions are satisfied and that the instruments being used are valid ones. The overall pattern of these results is robust to different specifications of the set of instruments.[24]

I can now turn to the actual results of the indirect inference procedure. The auxiliary estimation is based on the first-difference specification of equation (34). To make the simulations of y consistent with this, I use the following strategy. For each value of θ, I first run the simulation fixing the X characteristics at their 1980 levels. This yields a simulated unemployment variable for 1980, $\tilde{y}(\theta)_{80}$. Then I repeat the simulation for the same θ (and the same sequence of shocks) using the 1990 values of the X's. This delivers $\tilde{y}(\theta)_{90}$. I then compute $\tilde{y}(\theta)_{90-80} = \tilde{y}(\theta)_{90} - \tilde{y}(\theta)_{80}$: this is the simulated counterpart to the LHS variable of the auxiliary regression in Table 2.

As I have mentioned in Section 4.2, I initially let all the parameters differ between 1980 and 1990. However, from the results of the estimation over finer and finer grids, it seems that for at least *some* parameters there is no gain in allowing differences between years. Therefore, in the end I only let $\lambda_{\text{pchigh}}(80) \neq \lambda_{\text{pchigh}}(90)$, and $\alpha_{\text{pcnowhi}}(80) \neq \alpha_{\text{pcnowhi}}(90)$, imposing equality across decades for the other parameters: this helps reduce the "curse of dimensionality" problem. The estimated structural parameters $\tilde{\theta}$, as well as the estimated auxiliary parameters used in the indirect inference (from the data and from the simulations), are reported in Table (3). The chi-square global specification test statistic and the test on the null hypothesis ($\lambda \equiv 0$) are also reported. The geographic distribution of the *simulated* unemployment variable is plotted in Figure (14): one can visually compare this map with the actual unemployment map in Figure (7).

Table (3) shows that the structural model cannot be rejected for the first-differenced data 1990-1980, not even at the 60 % confidence level (the p value is .41). The structural parameters λ, associated with the local interaction term in the transition probabilities of the contact process, are significantly different than zero. Furthermore, when I perform the test on the null hypothesis $H_0 = (\lambda = 0)$, I can reject the null at the 98 % confidence level (using the test statistic κ_n^C, defined in equation (26)). So the estimation of the structural model supports the hypothesis that a portion of the high spatial autocorrelation of unemployment observed in the data can be attributed to the local interactions that take place across neighboring tracts. Tract characteristics alone (operating through the α parameters) are not sufficient to fit the observed spatial distribution of unemployment.

In addition, I can characterize how the information spillovers vary with tract characteristics: $\lambda(X(i,j))$ has a positive intercept, is decreasing in the level of education

[24] Section 5.1 contains further comments on the auxiliary estimation used here for the indirect inference.

Table 3: Indirect inference: 1990-80

Structural parameters: 1980			
	constant	pchigh	pcnowhi
λ	0.0733	-0.1286	0.0724
	(0.0182)	(0.0200)	(0.0180)
α	0.8954	0.2681	-0.2402
	(0.0208)	(0.0215)	(0.0184)
Structural parameters: 1990			
	constant	pchigh	pcnowhi
λ	0.0733	-0.0440	0.0724
	(0.0182)	(0.0162)	(0.0180)
α	0.8954	0.2681	-0.3519
	(0.0208)	(0.0215)	(0.0131)
χ^2 test (10 d.f.): $\kappa_n = 10.3576$ (p-value = 0.41)			
χ^2 test on $H_0 = (\lambda = 0)$: $\kappa_n^C = 10.0825$ (p-value = 0.018)			
Auxiliary parameters: 1990-80			
Variable Name	ρ_{GMM}(DATA)	ρ_{GMM}(SIM)	S.E.
unempr(nbs-1)	0.5300	0.5002	0.1684
unempr(nbs-2)	0.4373	0.2685	0.1280
unempr(nbs-3)	-0.0601	-0.1284	0.1586
unempr(nbs-4)	-0.4430	-0.1914	0.2533
unempr(nbs-5)	0.2564	0.2870	0.4094
unempr(nbs-6)	0.0022	0.0240	0.3560
u(nbs-1)×pcnowhi	0.0177	0.0052	0.0109
u(nbs-1)×pchigh	-0.0234	-0.0141	0.0099
pcnowhi	-0.0257	0.0647	0.0509
pchigh	0.0768	-0.0164	0.0571
mean	2.9525	3.0183	0.2609
variance	55.2852	33.8374	8.4616
sp.cov(1)	16.2293	12.8885	3.3922
sp.cov(2)	12.4444	9.3696	2.1832
sp.cov(3)	9.3488	8.1583	1.8389
sp.cov(4)	6.4059	7.2098	1.5324
sp.cov(5)	5.9999	5.9641	1.2941
sp.cov(6)	4.1570	5.2908	1.3029

in the tract, and is increasing in the percentage of minorities. Thus the spillover effects are stronger for areas with lower education levels and with a higher percentage of non-white residents. This is consistent with empirical results contained in Corcoran, Datcher and Duncan (1980) and in Granovetter (1974). This literature concentrates on the nature of informal hiring channels and is based on rather detailed data on employees' work history (in particular, how they were hired). The authors find that informal contacts used to obtain jobs are more important for younger workers, low-skilled jobs, less educated workers and minorities. This is therefore a very interesting result, since it corroborates some independent estimates from the existing literature on informal hiring. An alternative explanation for the reported signs of $(\lambda_{\mathbf{pchigh}}, \lambda_{\mathbf{pcnowhi}})$ has to do with the choice of geographic distance as a proxy for social distance, in my structural model. As social networks of poorer, less educated agents tend to be more geographically concentrated, it is quite intuitive that the local interaction effect is stronger for tracts with these characteristics, since I am focusing on the geographic component of social networks. The signs of the α parameters are also intuitive: higher education levels raise the probability of increasing employment in the tract, whereas a higher fraction of minorities is associated with a lower probability of employment.

Finally, I can test the identification condition of page 20 by fixing θ at the estimated values reported in Table (3), running a very long simulation (150,000 iterations) and numerically calculating the matrix of partial derivatives $\frac{\partial \rho}{\partial \theta}$. This matrix turns out to be full rank, both if I take the average over 100 different "snapshots" and if I look at each individual snapshot. Thus the identification condition seems to be satisfied.

It is worthwhile at this point to get an idea of the *magnitude* of the information spillovers: in other words, I can determine the expected impact on the employment rate of tract (i, j) of an increase in the employment rate of the neighboring tracts. This is done by using the transition probabilities (7) - (10) of Section 2 to compute the expected effect of a change in the amount of information transmitted by the neighboring tracts, $I_t(i, j)$. Of course this statement can only be made *conditional* on a given level of the tract characteristics $X(i, j)$, and needs to be made separately for 1980 and 1990, since the structural parameters and the X's differ in the two decades. Table (4) summarizes the results of several experiments. First, I consider an average neighborhood (setting the X's at their city-wide average level) and look at the effect of raising $I_t(i, j)$ by one standard deviation (about 8 percentage points in 1980, and 12 points in 1990). The effect is very small: the expected unemployment rate for this average neighborhood would decrease by a quarter of a percent in 1980, and by three quarters of a percent in 1990.[25]

[25] In Table (4) the bold-face numbers are the effects using the point estimates for λ and α, whereas the numbers in brackets above and below the bold-face ones are the effects when one uses the point estimates of the structural parameters *plus* or *minus* one standard error, respectively.

Table 4: Magnitude of spillover effects (1990-80).

Change in Unemployment rate (% points)			
Neighborhood	**1980**	**1990**	**experiment**
	(-0.02)	(-0.45)	
Average Nbd.	**-0.23**	**-0.75**	Raise Info. by one s.d.
	(-0.41)	(-1.00)	
	(-0.44)	(-0.93)	
Grand Boulevard	**-0.73**	**-1.28**	Raise Info. by one s.d.
	(-0.96)	(-1.58)	
	(-0.94)	(-2.48)	
Grand Boulevard	**-1.56**	**-3.43**	Same Info. as Lake View
	(-2.06)	(-4.21)	
	(—)	(0.16)	
Lake View	**0.03**	**0.51**	Lower Info. by one s.d.
	(0.28)	(0.82)	
	(—)	(0.42)	
Lake View	**0.07**	**1.37**	Same Info. as G. Blvd.
	(0.61)	(2.20)	

Secondly, I repeat the experiment for tracts in two polar Community Areas in Chicago: Grand Boulevard is a poor neighborhood on the South Side, which had a 23.7 % unemployment rate in 1980 and 37.5 % in 1990;[26] Lake View is a rich neighborhood close to Lincoln Park (the "yuppie" part of town), with unemployment rates of 6.1 % in 1980 and 5.2 % in 1990. Now the effect of raising information by one standard deviation in Grand Boulevard is stronger: expected unemployment decreases by 3/4 of a percent for 1980 and by 1.3 percentage points in 1990. This is because this neighborhood has lower education levels and is almost entirely non-white, and we have seen that the local interaction channel is stronger for such areas. Interestingly, the polar case of decreasing information in Lake View is very asymmetric: the expected effect on unemployment is almost zero in 1980, and only half of a percent in 1990. In other words, Lake View would not suffer from a decrease in information levels as much as Grand Boulevard would gain from an increase in information of the same size. This is again because of the different tract characteristics that affect the strength of the local interactions. Finally, I conduct the following mental experiment: take a tract in Grand Boulevard and give it the same amount of information $I_t(i,j)$ as a tract in Lake View, and viceversa. Now the effect for Grand Boulevard is quite large (1.5 points in 1980, 3.4 in 1990), whereas again the opposite effect for Lake View is not nearly as large. For all these different experiments, the spillover effect is roughly two to three times larger in 1990 than in 1980, indicating that informal hiring channels seem to have acquired more importance during the 1980's.

4.4 Results for 1980 and 1990 separately

At this point, I want to perform the indirect inference estimation for 1980 and 1990 separately, to see how crucial it is for the results of Section 4.3 to allow for unobservable characteristics in each tract. In terms of the discussion of Section 3, this strategy amounts to dropping the fixed effect θ_i in the error terms ϵ_{it} and performing the estimation in levels rather than in first differences. The results of the estimations are reported in Tables (5) and (6) for 1980 and 1990, respectively. The geographic distributions of the *simulated* unemployment variables are plotted in Figures (15) and (16): one can visually compare them with the actual unemployment maps in Figures (5) and (6).

The main difference between the two years is that the structural model cannot be rejected for 1980 (the p-value of the specification test is .17), whereas it can be rejected, at the 99% confidence level, for 1990. In terms of the actual estimates, the

[26]Roughly 15 % of all Chicago tracts had similar or higher unemployment rates in the two decades.

Table 5: Indirect inference: 1980

Structural parameters			
	constant	pchigh	pcnowhi
λ	0.2108	-0.0821	0.2223
	(0.0179)	(0.0134)	(0.0187)
α	0.6006	0.1602	-0.3220
	(0.0060)	(0.0157)	(0.0107)
χ^2 test (12 d.f.): $\kappa_n = 16.5587$ (p-value = 0.17)			
χ^2 test on $H_0 = (\lambda = 0)$: $\kappa_n^C = 26.5632$ (p-value = 0.0000)			

Auxiliary parameters			
Variable Name	ρ_{GMM}(DATA)	ρ_{GMM}(SIM)	S.E.
unempr(nbs-1)	1.6374	0.4392	0.3911
unempr(nbs-2)	-0.1394	0.1221	0.1115
unempr(nbs-3)	-0.3077	-0.6136	0.1741
unempr(nbs-4)	0.5011	0.4904	0.2014
unempr(nbs-5)	-0.2139	0.0143	0.2877
unempr(nbs-6)	-0.0547	0.1104	0.2199
u(nbs-1)×pcnowhi	-0.0021	0.0068	0.0024
u(nbs-1)×pchigh	-0.0230	-0.0135	0.0051
pcnowhi	0.0958	0.0096	0.0254
pchigh	0.1591	0.0689	0.0783
mean	11.4826	11.7423	0.2523
variance	57.1889	62.5664	8.7160
sp.cov(1)	35.3763	36.3965	6.2622
sp.cov(2)	26.9788	31.2106	5.2328
sp.cov(3)	22.8999	29.4153	4.7935
sp.cov(4)	19.9133	26.4182	4.4975
sp.cov(5)	17.5247	22.4109	4.1703
sp.cov(6)	15.4298	20.4994	3.9952

Table 6: Indirect inference: 1990

Structural parameters			
	constant	pchigh	pcnowhi
λ	0.1039	-0.0813	0.2089
	(0.0194)	(0.0141)	(0.0266)
α	0.8734	0.2292	-0.3524
	(0.0116)	(0.0147)	(0.0078)
χ^2 test (12 d.f.): $\kappa_n = 35.9070$ (p-value $= 0.0003$)			
χ^2 test on $H_0 = (\lambda = 0)$: $\kappa_n^C = 9.8408$ (p-value $= 0.02$)			

Auxiliary parameters			
Variable Name	ρ_{GMM}(DATA)	ρ_{GMM}(SIM)	S.E.
unempr(nbs-1)	-0.0963	-0.0343	0.3699
unempr(nbs-2)	-0.0550	0.1668	0.0910
unempr(nbs-3)	0.0262	-0.1780	0.1363
unempr(nbs-4)	-0.0012	0.1066	0.1508
unempr(nbs-5)	0.0876	0.0811	0.2678
unempr(nbs-6)	-0.2143	0.0410	0.2012
u(nbs-1)×pcnowhi	0.0044	0.0063	0.0023
u(nbs-1)×pchigh	0.0005	-0.0055	0.0037
pcnowhi	-0.0068	0.0180	0.0267
pchigh	-0.1148	0.0212	0.0655
mean	14.3406	13.4661	0.3710
variance	116.3633	65.6205	17.9766
sp.cov(1)	80.2717	34.5092	12.5436
sp.cov(2)	67.6376	29.7573	11.0424
sp.cov(3)	56.6507	28.9329	9.7855
sp.cov(4)	46.9895	25.9565	8.8554
sp.cov(5)	42.8450	22.9111	8.3329
sp.cov(6)	37.9107	21.8697	7.7969

Table 7: Magnitude of spillover effects (1980 and 1990).

Change in Unemployment rate (% points)			
Neighborhood	**1980**	**1990**	**experiment**
	(-1.83)	(-1.16)	
Average Nbd.	**-1.95**	**-1.44**	Raise Info. by one s.d.
	(-2.05)	(-1.69)	
	(-3.80)	(-2.30)	
Grand Boulevard	**-3.88**	**-2.63**	Raise Info. by one s.d.
	(-3.95)	(-2.91)	
	(-8.15)	(-6.14)	
Grand Boulevard	**-8.33**	**-7.01**	Same Info. as Lake View
	(-8.48)	(-7.76)	
	(1.78)	(0.42)	
Lake View	**1.97**	**0.79**	Lower Info. by one s.d.
	(2.13)	(1.11)	
	(3.81)	(1.13)	
Lake View	**4.22**	**2.11**	Same Info. as G. Blvd.
	(4.58)	(2.98)	

pattern of results is similar to that of Table (3). The structural parameters λ are again significantly different than zero, and the null hypothesis $H_0 = (\lambda = 0)$ can be rejected at least at the 98 % confidence level, in both years.[27] So again, tract characteristics alone (through α) are not sufficient to fit the observed spatial properties of the empirical unemployment distribution: I need to include the local interaction term that affects the probability of finding jobs. Furthermore, the way in which the information spillovers vary with tract characteristics is very similar to what was found using 1990-80 data: the information exchange channel is stronger for tracts with lower education levels and more non-whites. Thus the same remarks apply, in terms of the similarity between these results and the ones of the empirical literature on informal hiring channels, mentioned above.

However, the *magnitude* of the spillovers is much larger for the two estimations in levels rather than in first differences (see Table (7)). In addition, the ranking is reversed: now the size of the effects for all the different experiments is larger in 1980 than in 1990. This shows the importance of including a fixed effect term in the estimation, to try to capture possible unobservables. Failing to do so produces larger estimates of the spillover effects, as it is to be expected. Therefore, Table (7) can be seen as a consistency check: the fact that the size of the spillovers decreases with the introduction of unobservables gives one more confidence in the validity of the estimation strategy.

4.5 Access to jobs

I mentioned in Section 2.2 that an alternative hypothesis to explain the high unemployment levels in inner-city Chicago neighborhoods has been the spatial mismatch hypothesis. Therefore, following Ellwood (1986) and Holzer (1991), I repeat the indirect inference estimation including two additional tract characteristics that can be seen as proxies for proximity to jobs. In particular, from the Census data one can construct a measure of median commuting time to work for residents in each tract, and a measure of the fraction of people who go to work in the same county (for Chicago, Cook County includes most of the central city). I add these variables to the set of X characteristics that I use in the auxiliary regression, to control for tract-specific conditions that can produce spatial correlation of unemployment that is not due to information spillovers. In this case the aim is to control for local labor market conditions.

As Tables (8) and (9) indicate, the indirect inference results do not change much,

[27] The identification condition for the structural parameters is still satisfied, both for 1980 and 1990: the matrix of partial derivatives of the auxiliary parameters with respect to the structural ones is full rank.

Table 8: Indirect inference (1990-80): access to jobs

Structural parameters: 1980			
	constant	pchigh	pcnowhi
λ	0.0676	-0.1150	0.0708
	(0.0147)	(0.0273)	(0.0157)
α	0.9008	0.2192	-0.2232
	(0.0275)	(0.0104)	(0.0130)
Structural parameters: 1990			
	constant	pchigh	pcnowhi
λ	0.0676	-0.0340	0.0708
	(0.0147)	(0.0142)	(0.0157)
α	0.9008	0.2192	-0.3276
	(0.0275)	(0.0104)	(0.0081)
χ^2 test (10 d.f.): $\kappa_n = 10.5719$ (p-value = 0.39)			
χ^2 test on $H_0 = (\lambda = 0)$: $\kappa_n^C = 11.0408$ (p-value = 0.012)			
Auxiliary parameters: 1990-80			
Variable Name	ρ_{GMM}(DATA)	ρ_{GMM}(SIM)	S.E.
unempr(nbs-1)	0.4479	0.4900	0.1504
unempr(nbs-2)	0.3949	0.3250	0.1334
unempr(nbs-3)	-0.1828	-0.1145	0.1498
unempr(nbs-4)	-0.1857	-0.1468	0.2424
unempr(nbs-5)	0.3767	0.1840	0.3508
unempr(nbs-6)	-0.2877	0.0278	0.2939
u(nbs-1)×pcnowhi	0.0198	0.0008	0.0104
u(nbs-1)×pchigh	-0.0186	-0.0153	0.0091
pcnowhi	-0.0373	0.0895	0.0487
pchigh	0.0470	-0.0247	0.0532
mean	2.9655	2.8637	0.2595
variance	51.1996	30.6334	8.1528
sp.cov(1)	14.7763	11.4340	3.2842
sp.cov(2)	12.1237	8.5549	2.1436
sp.cov(3)	8.9674	7.3349	1.8173
sp.cov(4)	6.1727	6.1695	1.5198
sp.cov(5)	5.8471	5.1543	1.2766
sp.cov(6)	4.1718	4.5416	1.2670

Table 9: Magnitude of spillovers (1990-80): access to jobs.

Change in Unemployment rate (% points)			
Neighborhood	**1980**	**1990**	**experiment**
	(-0.03)	(-0.51)	
Average Nbd.	**-0.24**	**-0.75**	Raise Info. by one s.d.
	(-0.42)	(-0.97)	
	(-0.44)	(-0.95)	
Grand Boulevard	**-0.71**	**-1.24**	Raise Info. by one s.d.
	(-0.93)	(-1.49)	
	(-0.95)	(-2.55)	
Grand Boulevard	**-1.52**	**-3.32**	Same Info. as Lake View
	(-2.00)	(-3.98)	
	(—)	(0.25)	
Lake View	**0.06**	**0.55**	Lower Info. by one s.d.
	(0.32)	(0.82)	
	(—)	(0.68)	
Lake View	**0.12**	**1.48**	Same Info. as G. Blvd.
	(0.68)	(2.19)	

especially in terms of the magnitude of the information spillovers. The λ parameters are still significantly different than zero, and the test on the hypothesis that they be jointly equal to zero is rejected at the 99 % confidence level. Therefore, it seems that at least for the proxies of access to jobs employed here, the spatial mismatch hypothesis is not sufficient to eliminate the local interaction effect examined in this paper. Of course this remains a preliminary result in this respect, as availability of jobs needs to be measured more directly than simply through proximity measures.

5 Empirical results of the auxiliary model

In this Section I report some additional results from the estimation of the auxiliary model. Firstly, I look at the Auto-Correlation Function of unemployment, in order to decompose its spatial covariances into the portion that can be attributed to unemployment in the neighboring tracts, as opposed to the portion that is due to auto-correlation in the X tract characteristics. Secondly, I repeat the auxiliary estimation of Table (2) with different cross-terms (the $\tilde{y}_{it}^{N}$ variables of equation (34)) to look for additional points of contact with the existing literature on informal hiring channels mentioned in Section 4.3.

5.1 Spatial autocorrelations

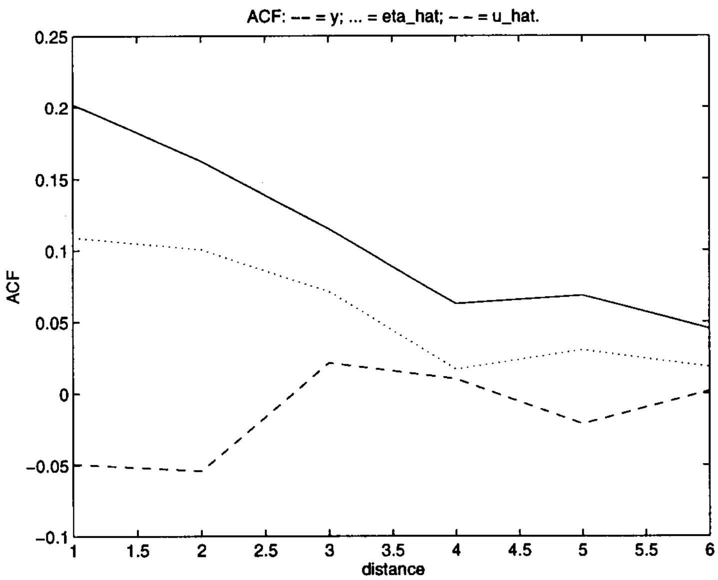
The local interactions built into the structural model, the contact process, generate positive spatial correlations in the stationary distribution of unemployment. It is therefore interesting to look at the spatial properties of the empirical distribution of unemployment in Chicago, and further analyze the results of the auxiliary regression of Table (2), in order to measure the portion of spatial correlation that can be directly attributed to unemployment in the neighboring tracts.

Table (10) contains the autocorrelation coefficients, as a function of distance, for the following variables. The first is the unemployment rate (in levels) in 1980. This is included as a term of reference. The second is the dependent variable of the SAR(6) auxiliary regression in Table (2): the unemployment rate in first differences. The third variable is the vector of fitted residuals $\hat{u}$ of the regression (including *all* RHS variables). The fourth variable is a vector of fitted residuals $\hat{\eta}$ that are calculated excluding from the RHS variables the direct effect of unemployment in the neighboring tracts (by itself

Table 10: Correlation coefficients as a function of distance

correlations and Q-tests						
Distance	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$
y_{80}	0.5261	0.3841	0.3001	0.2533	0.1949	0.1767
Q-test = 546.7418 (p-value = 0.0000)						
y_{90-80}	0.2022	0.1624	0.1151	0.0625	0.0684	0.0452
Q-test = 76.9320 (p-value = 0.0000)						
$\hat{\eta}$	0.1091	0.1007	0.0715	0.0166	0.0301	0.0182
Q-test = 24.1924 (p-value = 0.0005)						
$\hat{u}$	-0.0493	-0.0543	0.0211	0.0100	-0.0213	0.0017
Q-test = 5.3776 (p-value = 0.496)						

Figure 4: Auto-Correlation function



and interacted with $\tilde{x}$):

$$\hat{\eta}_i = \hat{u}_i + (\tilde{y}_i^N)^\top \phi.$$

The idea is to determine what portion of the spatial correlation is due to the direct effect of unemployment itself. In addition, Table (10) reports the value of a Box-Ljung-type test for each variable. This is a test on the null hypothesis that the variable in question exhibits no spatial correlation. Under the null, the test behaves as a χ^2 with D degrees of freedom, where D is the number of correlation coefficients used in the test.

Table (10) indicates that the unemployment variable, in 1980 levels, exhibits a very high degree of spatial correlation. The null hypothesis of no spatial autocorrelation is rejected at the 99% confidence level. The same holds true for y_{90-80}. More interestingly, though, the fitted residuals $\hat{\eta}$ still display positive spatial correlations, and the null hypothesis is still rejected at the 99% confidence level. It is only when one controls for the direct effect of unemployment of the neighbors that most residual spatial correlation disappears: for the "true" fitted residuals $\hat{u}$ one can no longer reject the null hypothesis of no spatial autocorrelation, even for a very low confidence level (the p-value of the test is .496). This result is by no means conclusive, but it still lends support to the idea that the mere existence of unemployment in contiguous areas generates a positive spatial correlation that cannot be explained away by other tract characteristics. This is again consistent with a model of local interactions and information spillovers. Figure (4) displays the Auto-Correlation Function of y_{90-80}, $\hat{\eta}$, and $\hat{u}$, as a function of distance.

Furthermore, the fact that the fitted residuals $\hat{u}$ do not display any significant spatial autocorrelation allows me to simplify the estimator of the covariance matrix $\hat{S}_n$ of equation (37):

$$\hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} g(\xi_i, \pi_n^{GMM}) \cdot g(\xi_i, \pi_n^{GMM})^\top. \tag{44}$$

In other words, I can assume that the error terms are not spatially correlated across tracts, so I can just consider the "contemporaneous" term in the sum over different distances d. The estimated coefficients of the auxiliary model do not change significantly with this new estimator of the covariance matrix. Finally, the lack of spatial auto-correlation of the residuals also confirms that the auxiliary model is a fairly good approximation to the spatial properties of the structural model, thus raising the efficiency of the indirect inference estimator of θ.

5.2 Cross-effects of tract characteristics with unemployment

At this point I would like to repeat the basic auxiliary estimation of Table (2) with different sets of cross-terms, in order to see whether the spillover effects of unemployment in neighboring tracts are stronger or weaker for different tract characteristics, such as education, age, ethnic composition, income and so on. To determine this, I simply look at the coefficients associated with the cross-terms *unempr(nbs-1)* $\times X^k$, for a given set of characteristics: *pc024, pcnowhi, pchigh.* The results of this regression are presented in Table (11).

The following observations are in order.

(1) The spatial correlation of unemployment is stronger for tracts with younger people.[28]
(2) The correlation is weaker for tracts with a higher fraction of people with at least a high school diploma; a similar result holds if we replace this variable with the fraction of professional and managerial workers.
(3) The correlation is stronger for tracts with a higher fraction of non-whites.

These observations are again consistent with empirical results contained in Corcoran, Datcher and Duncan (1980) and in Granovetter (1974): they find that informal contacts used to find jobs are more important for non-whites, younger workers, first jobs and low-skilled jobs. Here lower skill levels correspond to my measures of managerial/professional jobs and of education.

Observations (1) and (2) could be explained by an economic model in which building and maintaining one's social contacts is time-costly. Then younger and less educated/skilled agents might prefer to rely on informal channels to find a job rather than on formal qualifications because they may have a lower opportunity cost of time. In addition, low-skill workers are less specialized and can fit more diverse types of jobs, so information about generic job openings gathered through neighbors can be more useful for them. An alternative interpretation is that people with professional or managerial jobs tend to have business networks that do not follow a geographic pattern; since my model picks up the *spatial* component of social interactions, this will not be as strong for this kind of agents. The third observation, on non-whites, could also be related to the issue of skills and occupation, or it could be explained in terms of income levels, which is the object of the next Table.

Table (12) reports the estimates for a different set of cross-terms, in which I add the percentage of males out of the labor force (*pcolf(m)*) and an income dummy variable defined as follows:

[28] This result becomes statistically less significant if I use the fraction of people between 18 and 24 years of age, instead of between 0 and 24.

Table 11: Cross-terms, I

Dep. Variable: unemployment rate			
Variable Name	π_{GMM}	S.E.	cross-term
unempr(nbs-1)	0.6856	0.2030	
unempr(nbs-2)	0.3446	0.1386	
unempr(nbs-3)	0.0309	0.1619	
unempr(nbs-4)	-0.5285	0.2500	
unempr(nbs-5)	0.2167	0.4254	
unempr(nbs-6)	0.0110	0.3610	
unempr(nbs-1)	0.0433	0.0197	pc024
unempr(nbs-1)	0.0247	0.0117	pcnowhi
unempr(nbs-1)	-0.0225	0.0107	pchigh
constant	0.5202	0.8112	
pc1824	-0.5771	0.1852	
segr	0.0106	0.0226	
pcnowhi	-0.0682	0.0532	
pchigh	0.0716	0.0588	
pperhh	-0.9910	1.2985	
pcvac	0.1468	0.0544	
pcmnger	-0.1999	0.0587	
mgroren	-0.0081	0.0078	
hgvalue	0.0275	0.0123	
pcolf(m)	-0.0494	0.0275	
pcolf(f)	-0.1234	0.0364	
pcclge	-0.1414	0.0575	
pcfem	-0.1289	0.0705	
pc018	-0.5435	0.2187	
pc024	0.5905	0.2527	
pchisp	-0.0431	0.0297	
χ^2 test (39 d.f.) = 34.5078 (p-value = 0.67)			
adj. R^2 = 0.0421			
Box-Ljung test on $\hat{u}$ = 3.6416 (p-value = 0.72)			

Table 12: Cross-terms, II

Dep. Variable: unemployment rate			
Variable Name	π_{GMM}	S.E.	cross-term
unempr(nbs-1)	-0.3065	0.4026	
unempr(nbs-2)	0.1490	0.1408	
unempr(nbs-3)	-0.1723	0.1511	
unempr(nbs-4)	-0.3430	0.2744	
unempr(nbs-5)	0.7112	0.4598	
unempr(nbs-6)	-0.1848	0.3593	
unempr(nbs-1)	0.0543	0.0188	pc024
unempr(nbs-1)	0.0205	0.0110	pcnowhi
unempr(nbs-1)	0.0074	0.0099	pchigh
unempr(nbs-1)	0.0439	0.0080	pcolf(m)
unempr(nbs-1)	0.7308	0.3842	d50inc
constant	2.7173	0.8322	
pc1824	-0.8144	0.2025	
segr	0.0163	0.0242	
pcnowhi	-0.0156	0.0508	
pchigh	-0.0944	0.0555	
pperhh	-1.0419	1.5535	
pcvac	0.1587	0.0673	
pcmnger	-0.2212	0.0650	
mgroren	0.0014	0.0076	
hgvalue	0.0257	0.0156	
pcolf(m)	-0.3131	0.0523	
pcolf(f)	-0.1159	0.0360	
pcclge	-0.2174	0.0667	
pcfem	-0.1299	0.0769	
pc018	-0.7903	0.2498	
pc024	0.8418	0.2809	
pchisp	-0.0826	0.0267	
χ^2 test (37 d.f.) = 30.9421 (p-value = 0.75)			
adj. R^2 = 0.0608			
Box-Ljung test on $\hat{u}$ = 7.7087 (p-value = 0.26)			

d50inc: this dummy is 1 if tract i had a median income in 1980 that was lower than the median of *mhhinc* in 1980 across tracts in the city of Chicago; 0 otherwise.

The main result is that the spatial correlation of unemployment is much stronger for tracts that are below the median of the income distribution across tracts. This again can be interpreted with a model in which social networks are costly to maintain in terms of the time involved, and poorer people tend to have a lower opportunity cost of time. This is also consistent with observations in the sociological literature (see Jencks and Mayer (1990), p.124, Granovetter (1974) or Fischer (1982)), according to which the social networks of poorer families are in general more geographically restricted than those of affluent families. Then my empirical approximation of *social* distance in one's network with *spatial* distance is more accurate especially for poorer agents. In addition, the interaction term associated to males out of the labor force also has a positive and significant coefficient. In general, lower levels of participation into the labor market are associated with tracts with lower-skill workers and poorer households, so this result confirms the observation that local spillovers of unemployment are stronger for tracts with these characteristics.

6 Conclusion

In this paper I have looked for empirical evidence of local information spillovers in an urban labor market. A structural model has been presented, that exhibits information exchanges and local interactions. The information transmitted by one's neighbors affects one's likelihood to find a job. The model also allows agents to find employment independently of the informal channels, on the basis of their own characteristics. This enables me to take into account the effects of positive sorting. The parameters of the underlying structural model have been estimated through an indirect inference procedure, since it is not possible to write down explicitly the likelihood function associated to the invariant distribution out of the local interaction model. The results indicate that information spillovers are strictly positive both in 1980 and in 1990. The magnitude of the spillover effect is roughly two to three times as large in 1990 as in 1980. In particular, if we consider a high-unemployment neighborhood, increasing the information transmitted by neighboring tracts by one standard deviation would raise the expected employment rate in a given tract by about three quarters of a percentage point in 1980 and by 1.3 percentage points in 1990. Other experiments indicate even stronger effects. The social interaction effect is stronger for tracts with lower education levels, and with a higher fraction of non-whites.

Furthermore, the estimation of the auxiliary model gives additional information on the nature of the interactions. The spillovers are larger for areas with younger, less educated people, with lower median income, lower labor force participation, higher percentage of minorities, and fewer skilled workers. These results are consistent with the existing literature on informal hiring channels in the labor market.

The current work can be extended in several directions. First of all, I would like to repeat the empirical analysis using different distance metrics, such as travel times between locations or ethnic and religious composition. Thus two tracts would be considered closer if they have very similar ethnic, linguistic or religious profiles: in fact, social networks tend to follow rather closely these lines. Better still, one should work with individual-level data, trying to trace more accurately the social networks around individual agents. This would allow me to look at spillover effects within agents' networks instead of using geographic distance as a proxy for social distance. It would also be useful to repeat the analysis for different cities, to see how general these results are. Secondly, I plan to incorporate sorting more explicitly into the model, by considering agents' decisions to locate in a given neighborhood. These decisions should be based on characteristics of the area and on the expected benefit from any local interactions within that neighborhood. Finally, I would like to look more closely at the dynamic aspects of the rise and fall of neighborhoods. The structural model contains predictions about the behavior of the system over time, that could be tested by using data from several time periods. One could then derive impulse response functions to see how exogenous shocks propagate through the system both in time and in space.

References

[1] **Aaronson, Daniel (1996),** "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes", Working Paper, Federal Reserve Bank of Chicago.

[2] **Andjel, Enrique D. (1992),** "Survival of Multidimensional Contact Process in Random Environments", *Boletim da Sociedade Brasileira de Matematica,* vol. 23, nos. 1-2, 109-119.

[3] **Becker, Gary S. and Kevin M. Murphy (1994),** "The Sorting of Individuals into Categories when Tastes and Productivity Depend on the Composition of Members", unpublished manuscript, University of Chicago.

[4] **Benabou, Roland (1993),** "Workings of a City: Location, Education, and Production", *The Quarterly Journal of Economics,* 619-652.

[5] **Billingsley, Patrick (1986),** *Probability and Measure,* New York: John Wiley and Sons.

[6] **Brooks-Gunn, Jeanne, Greg Duncan, Pamela Klebanov and Naomi Sealand (1992),** "Do Neighborhoods Influence Child and Adolescent Behavior?", unpublished manuscript, University of Michigan.

[7] **Case, Anne C. (1991),** "Spatial Patterns in Household Demand", *Econometrica,* vol. 59, no. 4, 953-965.

[8] **Case, Anne C. and Lawrence F. Katz (1991),** "The Company You Keep: the Effects of Family and Neighborhood on Disadvantaged Youths", NBER Working Paper no. 3705.

[9] **Coate, Stephen and Martin Ravallion (1993),** "Reciprocity without Commitment", *Journal of Development Economics,* vol. 40, 1-24.

[10] **Coleman, James S. (1988),** "Social Capital in the Creation of Human Capital", *The American Journal of Sociology,* vol. 94, Suppl., S95-S120.

[11] **Conley, Timothy G. (1995),** "Econometric Modelling of Cross-Sectional Dependence", unpublished Ph.D. Dissertation, University of Chicago.

[12] **Conley, Timothy G. and Grace R. Tsiang (1994),** "Spatial Patterns in Labor Markets: Malaysian Development", unpublished manuscript, Dept. of Economics, University of Chicago.

[13] **Corcoran, Mary, Linda Datcher and Greg Duncan (1980),** "Information and Influence Networks in Labor Markets", in *Five Thousand American Families: Patterns of Economic Progress,* edited by Greg Duncan and James Morgan, vol. 7, 1-37, Ann Arbor, MI: Institute For Social Research.

[14] **Corcoran, Mary, Roger Gordon, Deborah Laren and Gary Solon (1989),** "Effects of Family and Community Background on Men's Economic Status", NBER Working Paper no. 2896.

[15] **Crane, Jonathan (1991),** "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing", *The American Journal of Sociology,* vol. 96, no. 5, 1226-1259.

[16] **Cressie, Noel A.C. (1993),** *Statistics for Spatial Data,* New York: John Wiley and Sons.

[17] **Durlauf, Steven N. (1994),** "Neighborhood Feedbacks, Endogenous Stratification, and Income Inequality", unpublished manuscript, Dept. of Economics, University of Wisconsin at Madison.

[18] ________ **(1996),** "Statistical Mechanics Approaches to Socioeconomic Behavior", unpublished manuscript, University of Wisconsin at Madison.

[19] **Durrett, Richard (1991),** "The Contact Process, 1974-1989", in *Mathematics of Random Media, Lectures in Applied Mathematics,* Blacksburg, VA: American Mathematical Society.

[20] **Ellwood, David T. (1986),** "The Spatial Mismatch Hypothesis: Are There Teenage Jobs Missing in the Ghetto?", in *The Black Youth Employment Crisis,* edited by R. Freeman and H. Holzer, Chicago: University of Chicago Press, 147-190.

[21] **Erbe, William et al. (1984),** *Local Community Fact Book: Chicago Metropolitan Area,* Chicago: Chicago Review Press.

[22] **Fischer, Claude S. (1982),** *To Dwell Among Friends. Personal Networks in Town and City,* Chicago: The University of Chicago Press.

[23] **Glaeser, Edward L., Bruce Sacerdote, and José A. Scheinkman (1996),** "Crime and Social Interactions", *Quarterly Journal of Economics,* vol. 111, 507-548.

[24] **Gouriéroux, C., A. Monfort, and E. Renault (1993),** "Indirect Inference", *Journal of Applied Econometrics,* vol. 8, S85-S118.

[25] **Granovetter, Mark S. (1974),** *Getting a Job: A Study of Contacts and Careers,* Cambridge, MA: Harvard University Press.

[26] **Hansen, Lars P. (1982),** "Large sample properties of generalized method of moments estimators", *Econometrica,* vol. 50, 1029-1054.

[27] **Hansen, Lars P. and Thomas J. Sargent (1993),** "Seasonality and Approximation Errors in Rational Expectations Models", *Journal of Econometrics,* vol. 55, 21-55.

[28] **Harris, Theodore E. (1974),** "Contact Interactions on a Lattice", *The Annals of Probability,* vol. 2, 969-988.

[29] **Holzer, Harry J. (1991),** "The Spatial Mismatch Hypothesis: What Has The Evidence Shown?", *Urban Studies,* vol. 28, 105-122.

[30] **Ihlanfeldt, Keith R. and David L. Sjoquist (1991),** "The Effect of Job Access on Black and White Youth Employment: a Cross-Sectional Analysis", *Urban Studies,* vol. 28, 255-265.

[31] **Jencks, Christopher and Susan E. Mayer (1990),** "The Social Consequences of Growing Up in a Poor Neighborhood", in *Inner-City Poverty in the United States,* edited by L. Lynn and M. McGeary, Washington, D.C.: National Academy Press, 111-186.

[32] **Liggett, Thomas M. (1985),** *Interacting Particle Systems,* New York: Springer Verlag.

[33] ———— **(1992),** "The Survival of One-Dimensional Contact Processes in Random Environments", *The Annals of Probability,* vol. 20, no. 2, 696-723.

[34] **Montgomery, James D. (1991),** "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis", *The American Economic Review,* vol. 81, no. 5, 1408-1418.

[35] ———— **(1992),** "Social Networks and Persistent Inequality in the Labor Market", unpublished manuscript, Northwestern University.

[36] **Schonmann, Roberto H. (1987),** "A New Look at Contact Processes in Several Dimensions", in *Percolation Theory and Ergodic Theory of Infinite Particle Systems,* edited by Harry Kesten, New York: Springer Verlag.

[37] **Tauchen, George (1996),** "New Minimum Chi-Square Methods in Empirical Finance", unpublished manuscript, Duke University.

[38] **Wilson, William J. (1987),** *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy,* Chicago: University of Chicago Press.

Appendix

The following are the labels of the variables used in the paper.

pc1824: the % of persons between the ages of 18 and 24 in the tract;
pcnowhi: the % of non-white persons in the tract;
segr: this is an index of segregation built as the Euclidean distance of a specific tract from the city-wide mix of ethnicities: if (w, b, s, o) are the city-wide proportions of whites, blacks, Hispanics and other ethnicities (respectively) on the whole population, then the index for tract k is defined as:

$$segr_k = \sqrt{(w_k - w)^2 + (b_k - b)^2 + (s_k - s)^2 + (o_k - o)^2}$$

pchigh: the % of persons over 16 years old who have a high school diploma or more;
pcclge: the % of persons over 16 years old who have a college degree or more;
pperhh: the average number of persons per household;
pcwelf: the % of average household income coming from public assistance;
pcvac: the % of housing units that are vacant;
hgvalue: average housing value in the tract;
mgroren: median gross rent in the tract;
mhhinc: median household income;
pcmnger: the % of employed persons 16 years old and over with professional or managerial jobs;
pcolf: the % of persons 16 years old and over who are out of the labor force; as I break it down by gender, I label this variable as *pcolf(m)* and *pcolf(f)* for males and females, respectively;
pcfem: the % of persons who are females;
pc018: the % of persons between the ages of 0 and 18 in the tract;
pc024: the % of persons between the ages of 0 and 24 in the tract;
pchisp: the percentage of Hispanic persons in the tract;
d50inc: a dummy variable that takes value 1 if tract i has a median income in 1980 below the median of *mhhinc* in 1980 across all tracts in the City of Chicago; 0 otherwise.

Figure 5: Map of unemployment, 1980

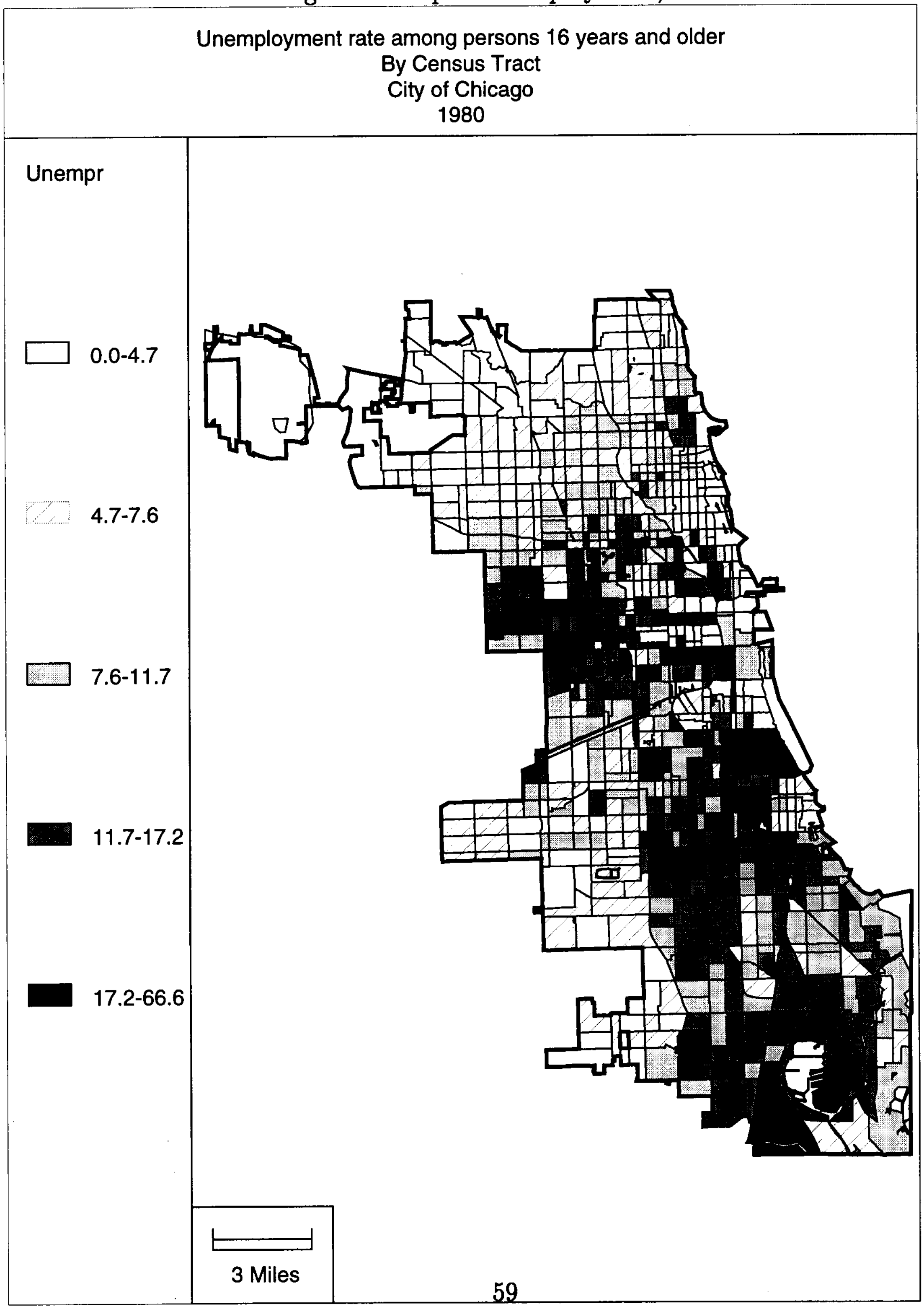


Figure 6: Map of unemployment, 1990

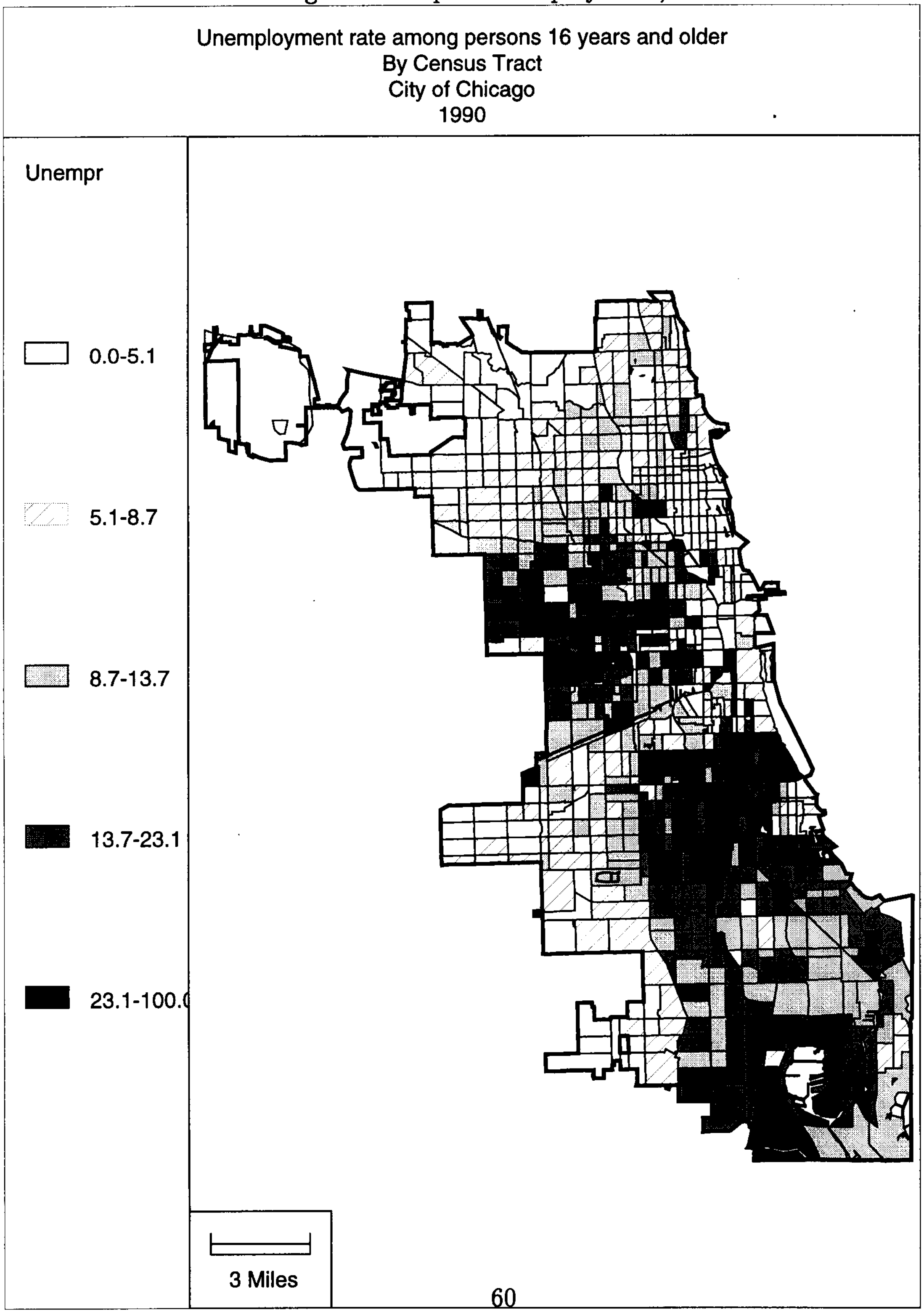


Figure 7: Map of unemployment, 1990-1980

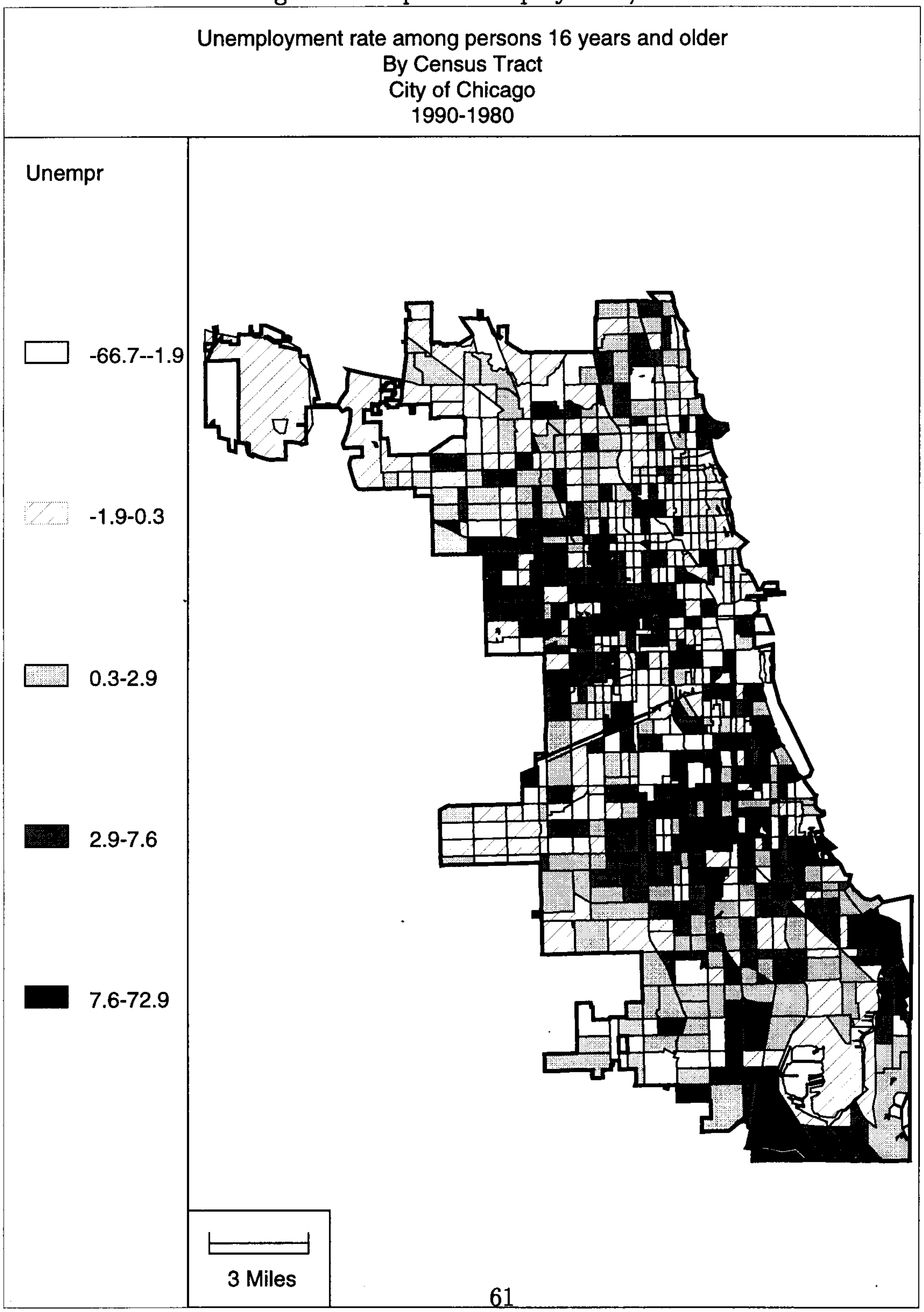


Figure 8: Map of *pchigh*, 1980

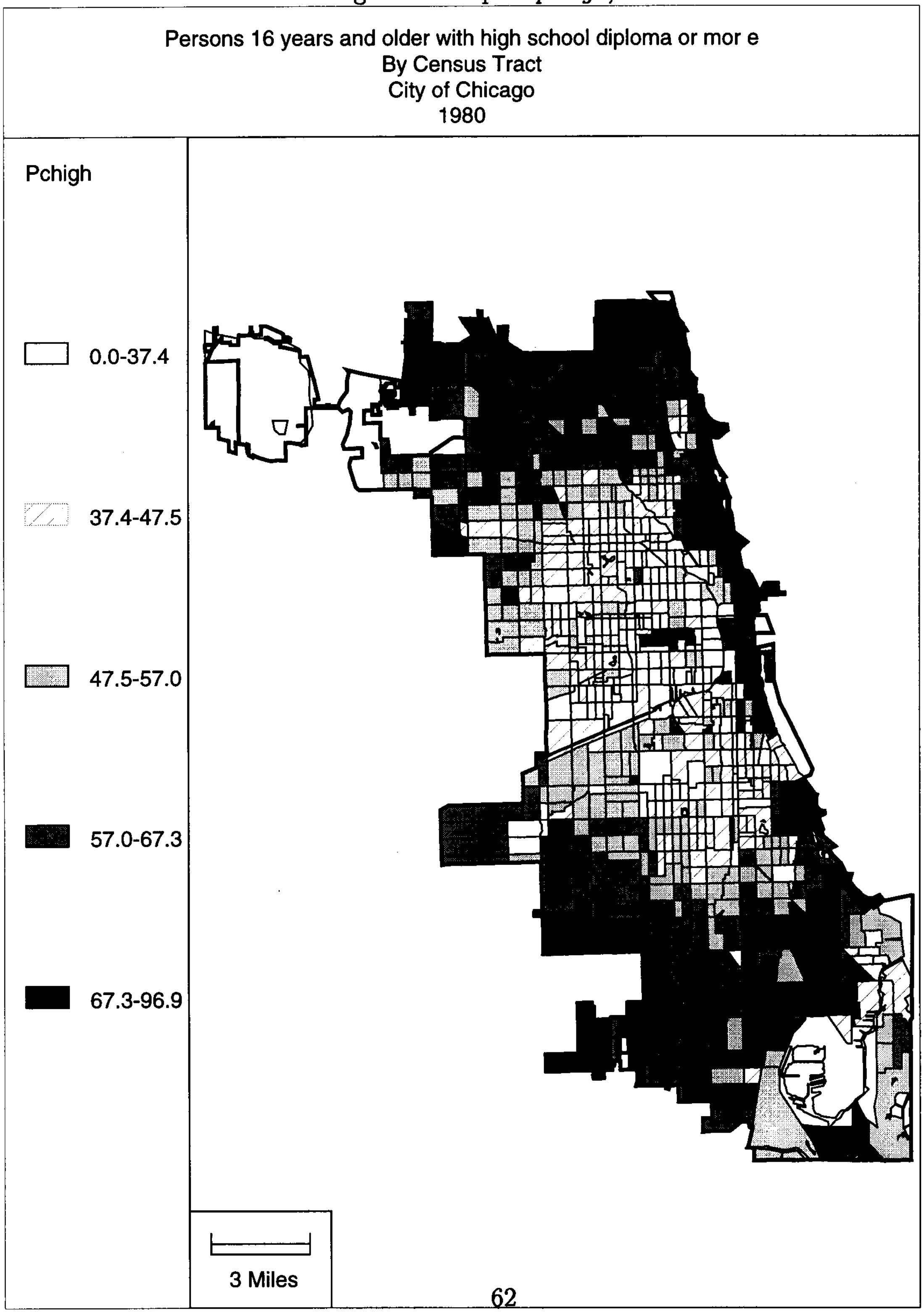


Figure 9: Map of *pchigh*, 1990

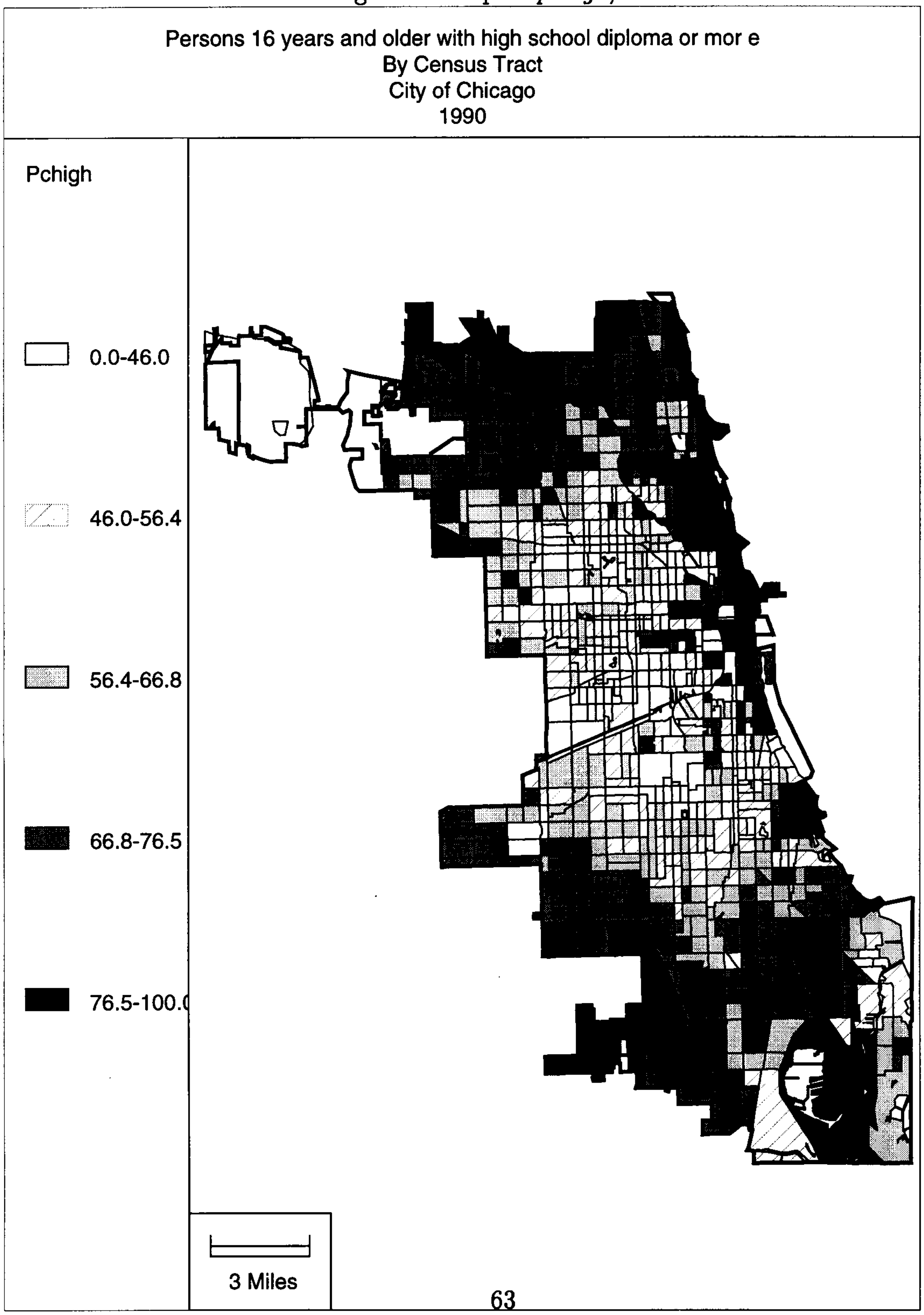


Figure 10: Map of *pchigh*, 1990-1980

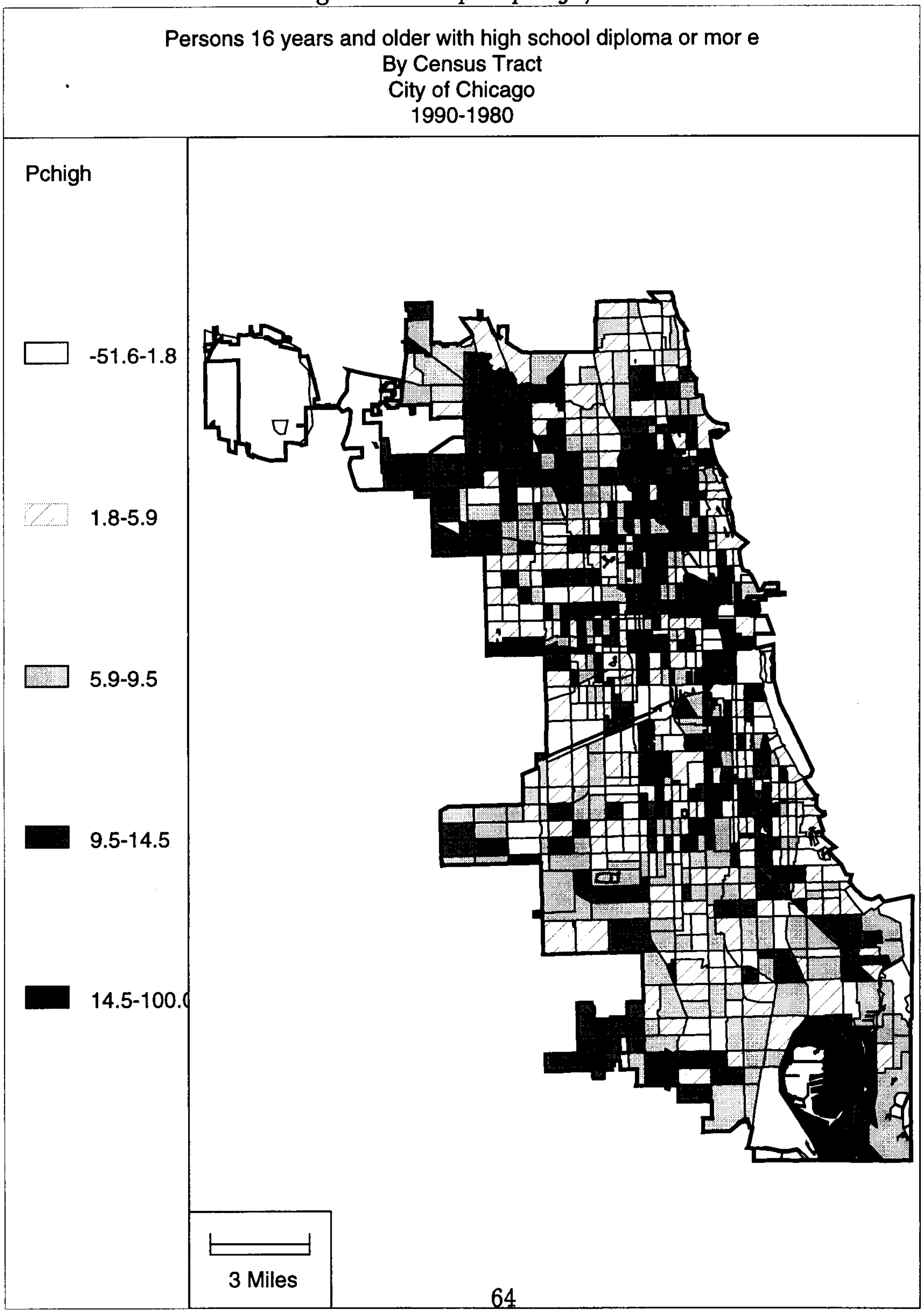


Figure 11: Map of *pcnowhi*, 1980

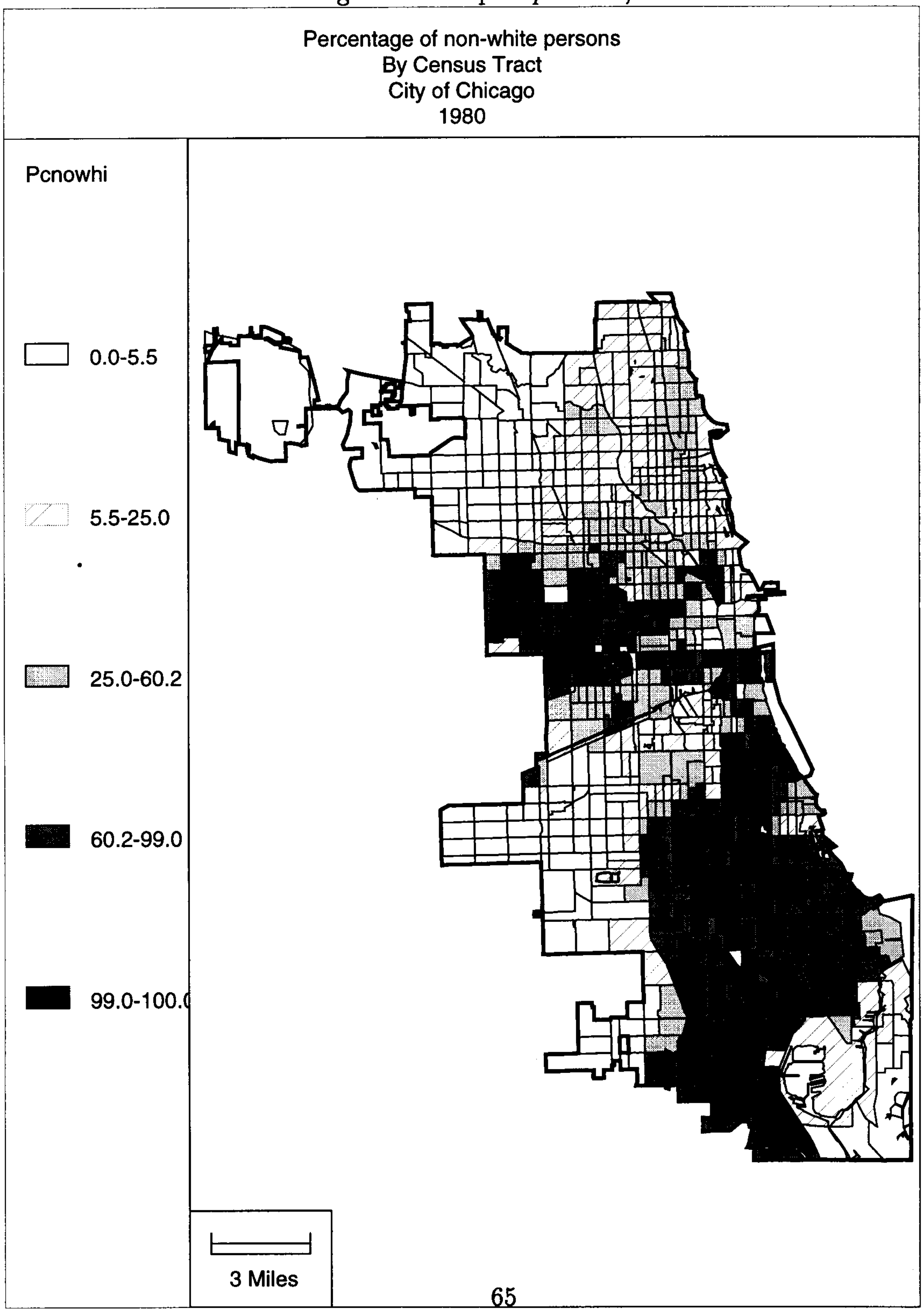


Figure 12: Map of *pcnowhi*, 1990

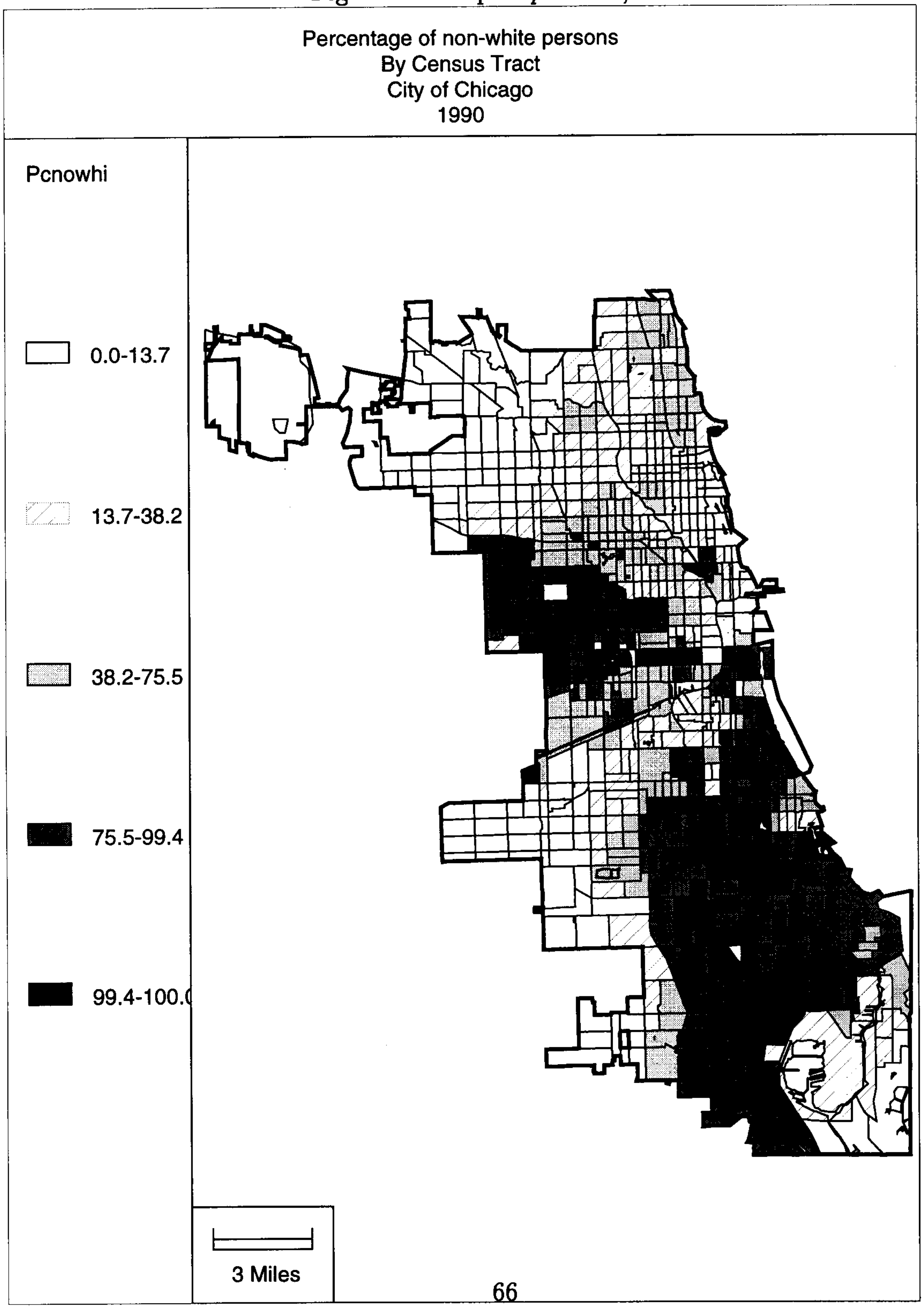


Figure 13: Map of *pcnowhi*, 1990-1980

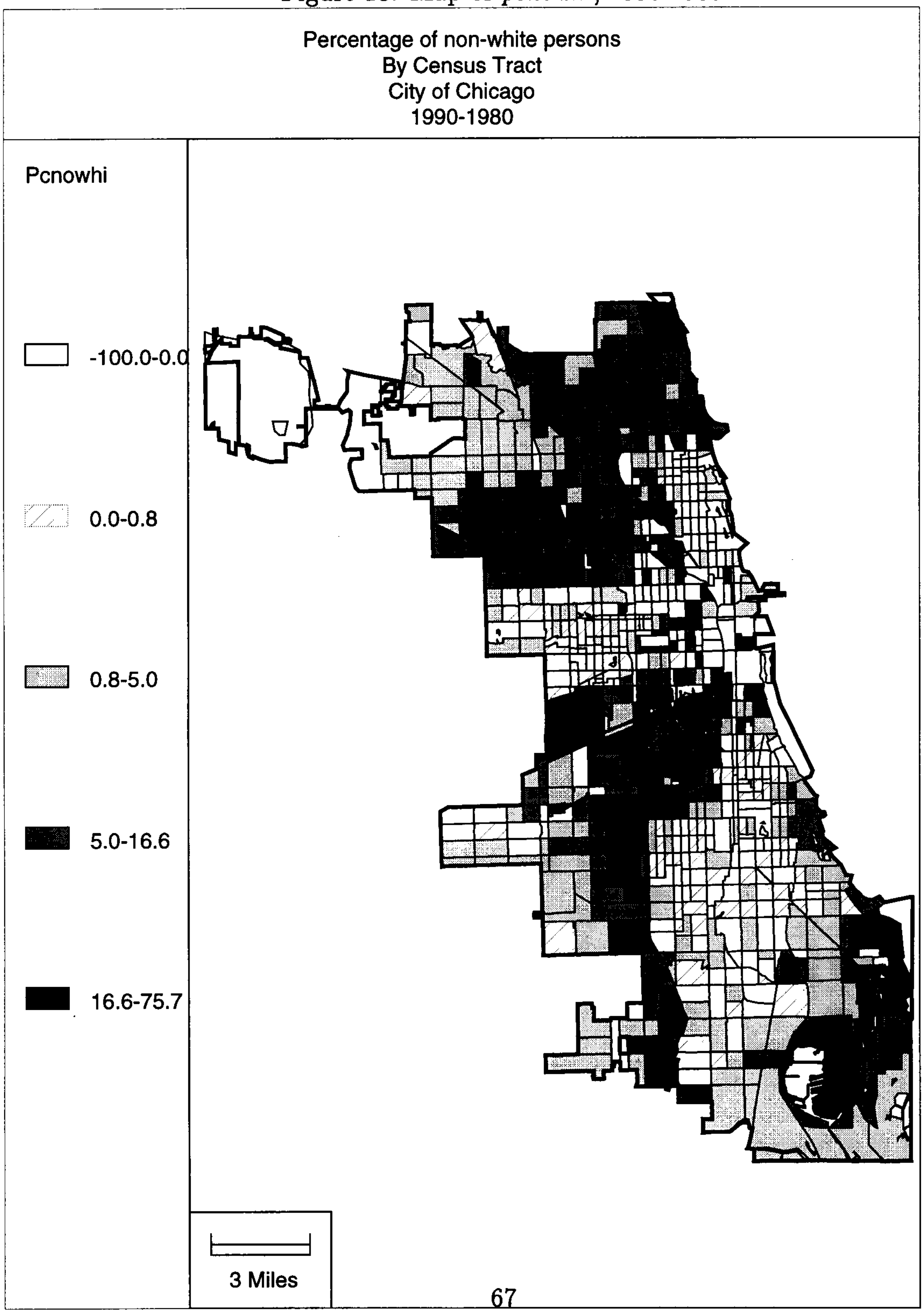


Figure 14: Map of SIMULATED unemployment, 1990-1980

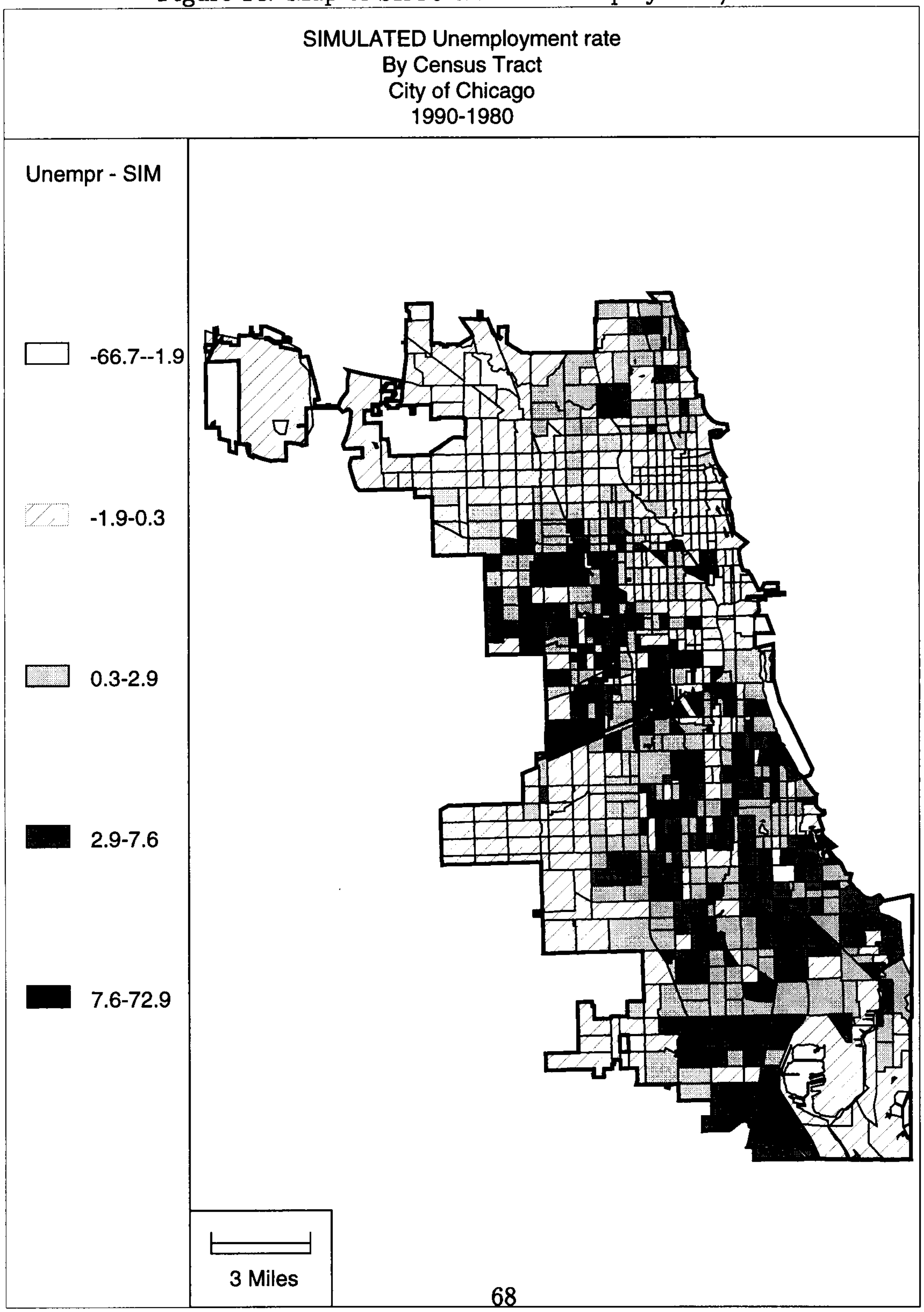


Figure 15: Map of SIMULATED unemployment, 1980

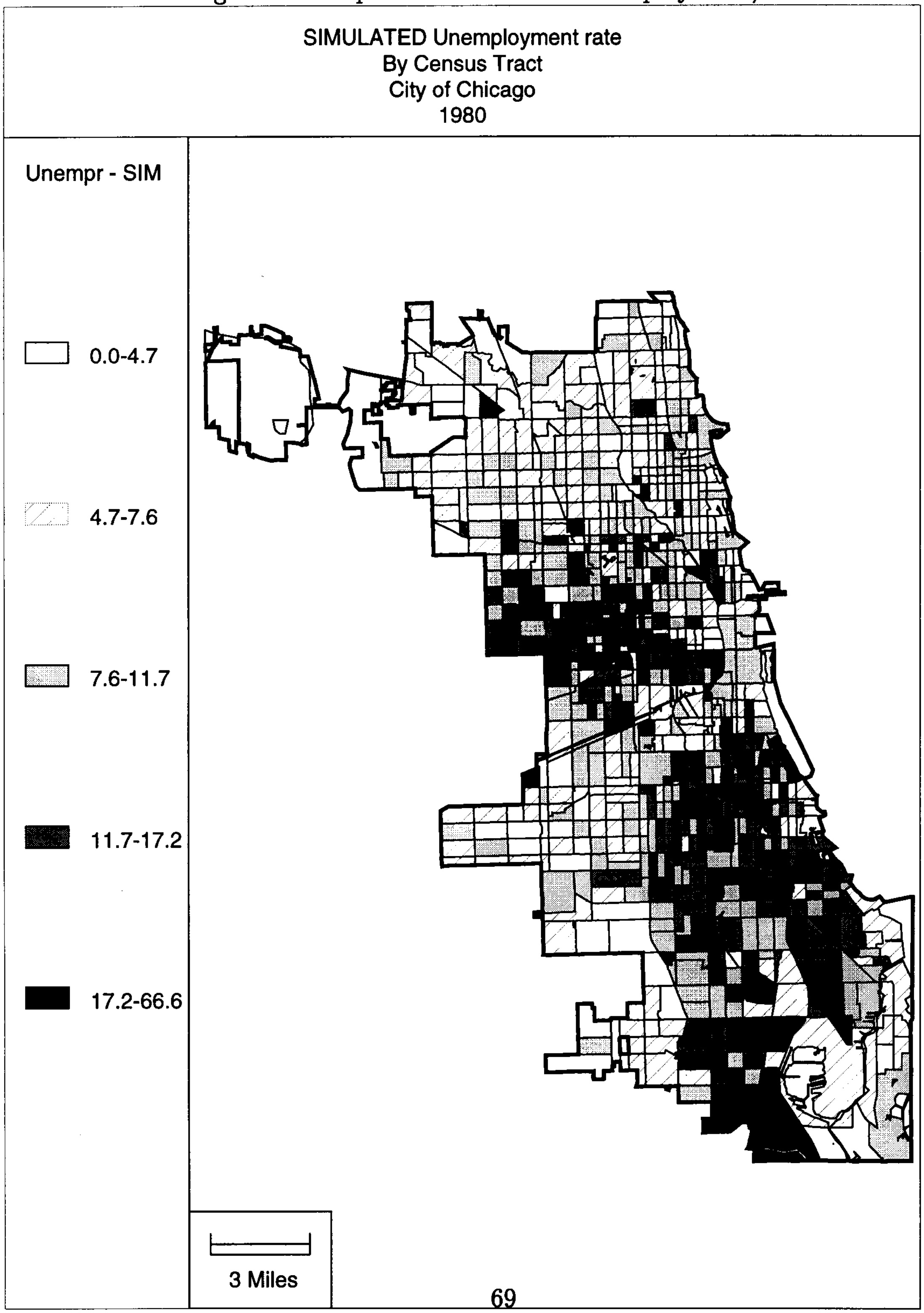


Figure 16: Map of SIMULATED unemployment, 1990