ECONOMIC RESEARCH REPORTS

Socio-Economic Distance and Spatial Patterns in Unemployment

By

Timothy G. Conley
&
Giorgio Topa

RR# 99-04 **February 1999**

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

Socio-Economic Distance and Spatial Patterns in Unemployment

Timothy G. Conley, Northwestern University
Giorgio Topa, New York University*

February 10, 1999

Abstract

This paper examines the spatial patterns of unemployment in Chicago between 1980 and 1990. We study unemployment clustering with respect to different social and economic distance metrics that reflect the structure of agents' social networks. Specifically, we use physical distance, travel time, and differences in ethnic and occupational distribution between locations. Our goal is to determine whether our estimates of spatial dependence are consistent with models in which agents' employment status is affected by information exchanged locally within their social networks. We present non-parametric estimates of correlation across Census tracts as a function of each distance metric as well as pairs of metrics, both for unemployment rate itself and after conditioning on a set of tract characteristics. Our results indicate that there is a strong positive and statistically significant degree of spatial dependence in the distribution of raw unemployment rates, for all our metrics. However, once we condition on a set of covariates, most of the spatial autocorrelation is eliminated, except for physical and occupational distance.

JEL: J64, R12, C21.

Keywords: Social networks, economic distance, spatial econometrics, unemployment.

---

*Corresponding author: Giorgio Topa, Dept. of Economics, New York University, 269 Mercer St. - NY, NY 10003. Giorgio.Topa@econ.nyu.edu. The authors are grateful to Alberto Bisin, Caterina Musatti, Chris Taber, Wilbert van der Klaauw, and Frank Vella for helpful comments. Aron Betru and Margaret Burke provided excellent research assistance. Giorgio Topa gratefully acknowledges financial support from the C.V. Starr Center for Applied Economics at New York University. The authors are of course responsible for all errors.

1 Introduction

In this paper we examine the spatial patterns of unemployment in Chicago over two decades, 1980 and 1990. We study unemployment clustering with respect to different economic distance metrics that reflect the structure of agents' social networks. Our goal is to characterize spatial patterns of unemployment with respect to these metrics, in order to determine whether they are consistent with models in which agents' employment status may be affected by information exchanged locally within their social networks.

There is considerable evidence that social networks are important for job search. A vast body of research in economics and sociology has shown that at least 50% of all jobs are found through informal channels, such as talking to one's friends, family, neighbors, and social contacts in general. In a study of 282 male professional and technical workers in the Boston area, Granovetter [6] finds that about 57% of current jobs were found through personal contacts or referrals. Occupational contacts and weak ties were especially important.[1] Corcoran et al. [3] find very similar results using a much larger data set from the 1978 wave of the PSID. Montgomery [17] reports additional evidence and develops an adverse selection model in which informal hiring (referrals) coexists in equilibrium with more formal hiring channels.

Such information exchange processes occurring within agents' networks can generate observable implications if the network structure is at least partially observable to the econometrician. For example, if social networks are geographic in nature, i.e. individuals talk mostly to those who are physically nearby, then individuals' outcomes will be related to their physical distance. A job acquisition process that operated through such a network would generate clustering of unemployment with respect to physical distance. Of course, social networks need not be strictly geographic. Networks develop along other dimensions, such as race or ethnicity, religious affiliation, education. In addition, the information exchange between any pair of agents is likely to be more productive (in terms of generating job offers) if the two agents are close in terms of their respective occupations. If these non-physical metrics are important, they will be systematically related to agents' outcomes.

In this paper we investigate the relationship of patterns in unemployment to several metrics. Our goal is to identify the most appropriate metrics for describing clustering of unemployment. We can then use the relevant measures of dependence according to these metrics to calibrate or estimate structural models of behaviour. For example, a local interaction model of information exchange can be estimated by choosing parameters to match spatial autocorrelation patterns of unemployment as a function of a given metric, via a Simulated Method of Moments or an Indirect Inference procedure. Topa [23] provides an example of this estimation technique. Thus, this paper is intended as a first exercise in a wider research agenda, aimed at

[1] Two persons A and B have weak ties if agent A's social network has very little overlap with agent B's set of contacts. They have strong ties if they talk to roughly the same group of people.

bringing formal models of local interactions to the data.

We construct several different metrics over Chicago Census tracts that attempt to capture the non-geographic dimensions of social networks and job information exchanges. These metrics are measures of the travel time between tracts, difference in ethnic and occupation distributions, and physical distance between tracts. Such metrics are often described by the sociological literature as likely dimensions along which networks develop. To illustrate the differences between metrics and how they have changed over time, we present a graphical representation of each metric using a method called multidimensional scaling.[2] Comparisons of tracts' relative distances under each metric are straightforward to make. Changes in relative distances from 1980 to 1990 under ethnic and occupation metrics are also easily visible.

Upon construction of our metrics, we present nonparametric estimates of correlation across tracts as a function of each distance metric. First, we estimate these spatial Auto-Correlation Functions (ACFs) for unemployment rates themselves. Then, we estimate ACFs for residuals from a regression of unemployment on a set of observable tract characteristics that are likely to affect the unemployment rate of the area and to be spatially correlated due to the sorting of individuals across areas. We compare these two estimates to get an idea of whether the clustering with respect to a given metric can be 'explained' by conditioning on these variables.

We also present nonparametric estimates of correlation as functions of combinations of metrics. For example, we estimate the auto-correlation of unemployment as a function of both physical and ethnic distance and analyze the resulting two-dimensional auto-correlation surface. We estimate two-distance ACFs for several combinations of metrics, allowing us to examine the clustering patterns of unemployment according to the first metric, for the set of census tracts that are at a given distance with respect to the second.

Finally, we provide a simple test of the hypothesis of spatial independence for a given metric based on our ACF estimates. We use a bootstrap method to generate an acceptance region for ACF estimates under the hypothesis of spatial independence. In addition to being easy to implement, these tests overcome the problem that usual distribution approximations for local average estimates tend to overstate their precision when the data is dependent. Our tests are also robust to measurement error in our metrics. In so far as our measurements of distance are just proxies for the extent of connections between agents, they will certainly contain error.

Our main results indicate that there is a strong positive and statistically significant degree of spatial dependence in the distribution of raw unemployment rates, at distances close to zero, for all our metrics. The correlation decays roughly monotonically with distance. The two-metric ACF estimates offer some additional insights. When the physical or travel time metric is coupled with the ethnic metric, the latter drives most of the variation in spatial clustering: tracts that are at a given ethnic

[2] Mardia et. al. [14] provide an excellent description of this method. In Appendix A, we provide a brief description.

distance exhibit a roughly constant degree of auto-correlation no matter what the physical distance is between them. On the other hand, when physical or travel time metrics are combined with distance in occupations, the estimated ACF surface is roughly decreasing in both distances.

The ACF estimates for the residuals from the regression of unemployment on a set of observable tract characteristics are quite flat. Once we condition on covariates, most of the spatial clustering is eliminated. The only exceptions are the one metric ACF estimated using physical distance, and the two metric ACF surfaces based on a combination of physical and occupational metric. Therefore, it seems that most of the spatial clustering observed in the data may be driven by sorting of heterogeneous agents across locations.

We also make a start at trying to identify the specific conditioning variables that are most important in eliminating the observed spatial dependence of unemployment. Our preliminary results indicate that the racial and ethnic composition within each tract contributes the most to 'explaining' the spatial correlation present in the raw data. Measures of human capital have a limited impact, whereas so-called spatial mismatch variables do not seem to play a role.

The methods in this paper may prove useful for data exploration and description of stylized facts in many other contexts. Our approach could be applied to other socio-economic outcomes of interest, such as participation in welfare programs, crime, dropping out of school, or teenage childbearing. All that is required is measures of the relevant metric(s) describing the relationship between units of observation – households, individuals, or Census tracts.

The remainder of this paper is organized as follows. Section 2 describes the data. Section 3 discusses the different distance metrics used in our analysis, as well a comparison of the configurations implied by each of the metrics via multidimensional scaling. The spatial econometric model is presented in Section 4. Section 5 reports the results of our ACF estimates for the different metrics. Finally, Section 6 presents preliminary conclusions and discusses extensions of the current paper.

2 Data

Most of the data come from the Bureau of the Census[3] for the city of Chicago, at the tract level, for 1980 and 1990.[4] There are 866 Census tracts in the city of Chicago, and they are grouped into 75 Community Areas, which are considered to have a distinctive identity as a neighborhood.[5] Our unit of observation is the Census tract.

[3] Summary Tape Files 3A.

[4] Travel times between locations were calculated using published CTA documentation.

[5] A set of tracts is defined as 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. [4], p. xix).

In this paper we examine the clustering patterns of unemployment rates across Census tracts. Our outcome variable is defined as the percentage of unemployed persons over the civilian labor force (16 years and older). Unemployed persons are people who were neither "at work" nor "with a job but not at work" during the reference week used by the Census, and who were actively looking for work during the last four weeks prior to the reference week.

As we mentioned above, we want to examine the spatial patterns of unemployment both *unconditionally* and *net* of the clustering effect that may be simply due to sorting of individuals into locations. Therefore, we control for a rather long list of observable neighborhood attributes, that may be correlated with the probability of being employed on the one hand, and may be dimensions along which people sort when deciding where to reside on the other. We use the following three sets of covariates.

First of all, we define a set of *sorting variables,* i.e., variables that may affect the decisions by different types of individuals to locate in a given area. These include average housing values in the Census tract, median gross rents, the fraction of vacant housing units in the area, the fraction of persons with managerial or professional jobs, the percentage of non-white persons, the percentage of Hispanic persons, a segregation index, and the number of persons per household.

Secondly, we consider variables that may be linked more directly to the probability of being employed. These include the percentage of persons with at least a high school diploma, the fraction of persons with at least a college degree, the age composition in the tract (to proxy for potential experience), the fraction of females 16 years and older, and the percentage of males and females out of the labor force in the tract.

Finally, there is a relatively large literature in urban and labor economics that discusses the *spatial mismatch* hypothesis. The literature aims at explaining the high unemployment levels in mostly black, inner city neighborhoods by local labor market conditions. The basic idea is that during the 1970's and 1980's many jobs (especially low-skill ones in the service industry) have moved from central city areas to the suburbs. In addition, the contention is that there is low residential mobility and a certain degree of housing segregation for inner-city blacks. For example, it may be very costly for a black household to relocate to the suburbs in a mostly white neighborhood, where the social capital provided by a black community would be missing. Several authors, such as Holzer [8] and Ihlanfeldt and Sjoquist [10], [11] have analyzed this issue empirically. By now, there is a certain consensus that physical proximity to jobs explains a portion of black/white unemployment differences, for instance. Therefore, we include the median commuting time to work for workers who reside in each tract as our measure of proximity to jobs.

3 Distance metrics

In this Section we introduce several different distance metrics that we use in the remainder of the paper to estimate spatial ACFs. In each instance, we try to motivate the particular choice of metric. It is important to keep in mind that our analysis is at the Census tract level, and not at the individual agent level. Therefore, we do not trace out individual agents' networks, but rather we define distances between pairs of tracts that attempt to reflect the dimensions along which social networks are stratified. In so doing, we refer to a rich sociological literature that documents the patterns of relations among agents. One unifying theme in the literature is that networks appear to be fairly homogeneous with regard to certain socio-demographic attributes.

3.1 Costs of interaction

The first obvious choice for a distance metric is physical distance. The underlying idea is that agents exchange information more frequently with people who live physically close. In fact, the development and maintenance of social contacts may be limited by physical distance and by transportation opportunities. There may be monetary and time costs to maintaining active ties with persons who reside far away. In addition, local organizations such as churches, local businesses, neighborhood clubs, or daycare centers, play an important role in fostering social ties and facilitating information exchanges at the local neighborhood level.

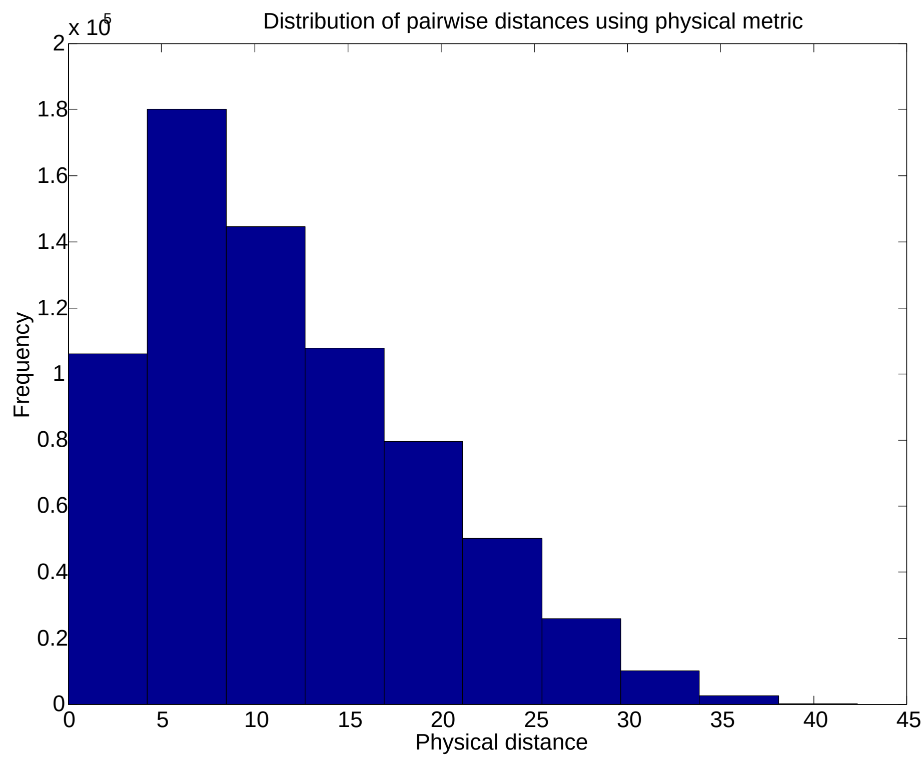
Of course, there is a legitimate concern that physical distance may have become less and less important in shaping social networks, as many interactions take place between people who reside in different neighborhoods or even cities. High residential mobility and the availability of communication tools such as the telephone or the Internet weaken the constraints that geographic space imposes on communities.

However, there is some evidence that seems to suggest that physical distance still plays an important role. In a study of Toronto inhabitants in the 1980's, Wellman [25] finds that a surprisingly high fraction of interactions takes place among people who live less than 5 miles apart. The study asks respondents (egos) to name a set of persons (alters) with whom they have active social ties, and records where respondents and alters reside, as well as the frequency of interactions within each ego-alter pair. The data show that about 38% of yearly contacts in all networks take place between ego-alter pairs that live less than 1 mile away. Roughly 64% of all contacts take place between agents who live at most 5 miles away.

One important qualification is that such studies do not tell us anything about the *content* of these contacts: ideally, we would like to restrict our attention to networks whose primary content is the information exchange about job openings. But one aspect reported by Wellman is encouraging. He observes that ties with one's neighbors are weaker (in the sense specified in the Introduction) than ties with friends or kin. It is precisely this kind of ties that is more conducive to generating useful information about jobs (see Granovetter [6]). Finally, there is some evidence reported in Fischer [5], showing that networks of lower-income, as well as younger, people are especially limited by physical space.

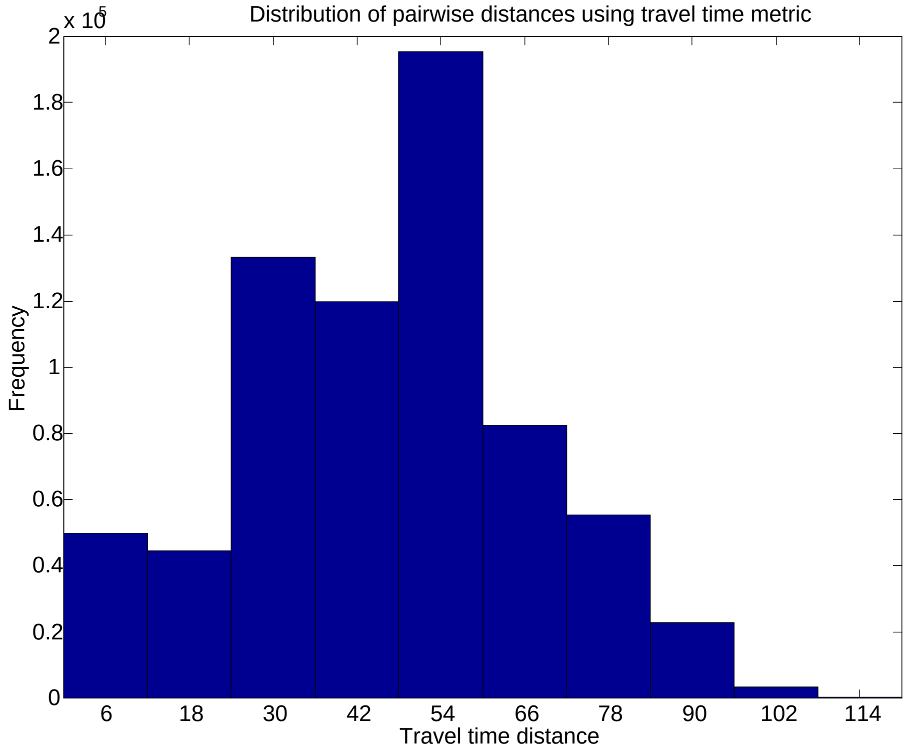
We use two different metrics to represent costs of interaction: physical distance and travel time.

Physical distance. PD_{ij} is the 'as the bird flies' distance in km. between the centroids of tract i and j. This metric is rather rough, as it does not take into account physical barriers, such as rivers or highways. It may also be worthwhile to examine variations of this metric that take into account the population size or density in each census tract. The distribution of pairwise distances across tracts using this metric is depicted below.



Travel Time distance. TD_{ij} is the travel time distance (in minutes), using public transportation (CTA), between the centers of the Community Areas in which tract i and j are located. Travel times were calculated from CTA timetables published in 1997 and reflect best-case travel times.[6]

[6] We hope that travel times during the 1980's are not very different from current ones. We tried as much as possible not to consider new lines that were opened in the mid-1990's.



3.2 Race and Ethnicity

Even casual observation suggests that personal networks may be stratified along specific socio-demographic attributes, such as race, ethnicity, religious affiliation, language, age, gender, education levels. In other words, agents are likely to draw a disproportionate share of their social contacts among sets of people that are very similar to themselves. This tendency is denoted as inbreeding, or homophily, among sociologists. Economists, on the other hand, refer to this phenomenon as positive sorting, or assortative matching.[7]

But how exactly widespread is inbreeding among agents' personal networks? There is some evidence that inbreeding is very strong among immigrant communities (see, e.g., Light et al. [12]). The strongest evidence, however, comes from the 1985 General Social Survey. This study, begun in 1972, is an annual[8] survey of the attitudes and behaviors of Americans on a wide variety of topics. The 1985 edition included a module on social networks of 1534 individuals, drawn as a nationally representative sample. Respondents were asked to name people with whom they "discussed important matters". Several characteristics of these alters were then collected, among which their age, sex, education, race, ethnicity, and religious affiliation.

[7] See Becker and Murphy [1], for example.

[8] The GSS produced surveys annually in 1973-78 and 1983-1993. Since 1994, it has been conducted every two years.

Marsden [15], [16] has used the GSS data in order to analyze the question of inbreeding. The results indicate that personal networks are quite homogeneous along several dimensions. In particular, network homogeneity with respect to race and ethnicity is very high. Only 8% of the respondents reported alters with *any* racial or ethnic diversity (both between ego and alters, and among alters). In addition, racial and ethnic heterogeneity of alters is only 13% of the total racial and ethnic heterogeneity among respondents.

Marsden [16] then looks specifically at ego-alter pairs, and decomposes the cell frequencies (e.g., the relative frequency of black-black pairs over all the ego-alter pairs) into a portion that is due to purely random matching (based on the marginal distributions of the race/ethnic categories among respondents and alters), and a portion that is due to inbreeding. It turns out that the strongest level of association, over and above random matching, takes place for the race/ethnicity attribute.[9] For example, the chance of observing a black-black tie is 4.2 times higher than that generated by pure random matching, given the relative proportions of the different racial and ethnic categories in the population.[10]

Since our study is based on aggregated data and not on individual network data, we would like to incorporate the inbreeding feature of social networks into a distance metric, that considers two tracts with very similar ethnic compositions to be close. The objective is to track more closely this important dimension along which personal networks are structured. We propose the following metric to take into account racial and ethnic attributes.

Race and Ethnicity distance. ED_{ij} is the euclidean distance between the vector e_i of percentages of nine races and ethnicities[11] present in tract i and the corresponding vector e_j in tract j:

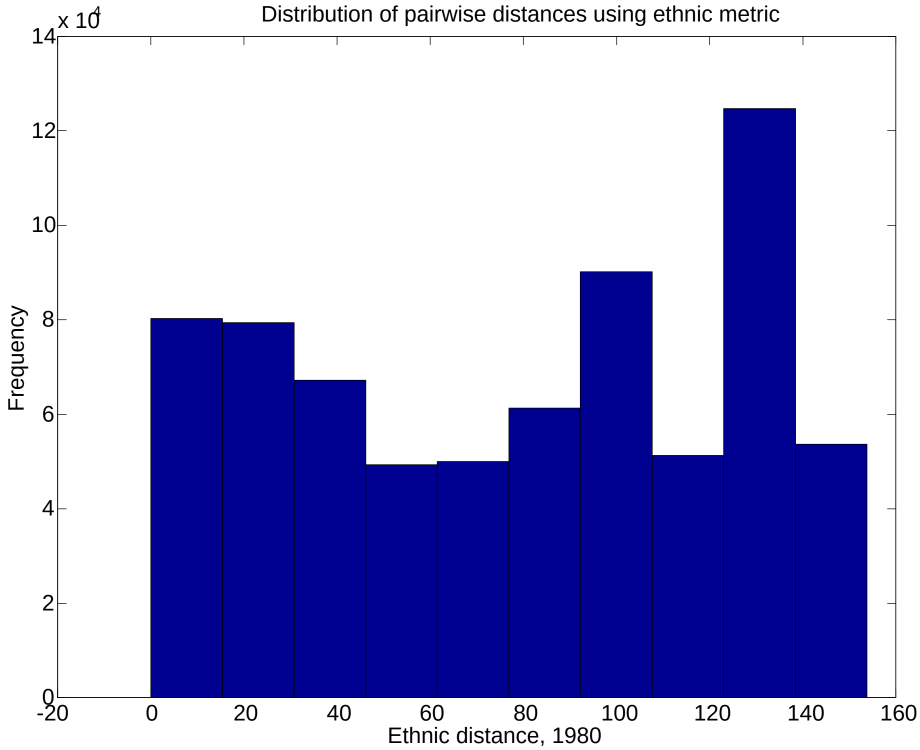
$$ED_{ij} = \sqrt{\sum_{k=1}^{9} (e_{ik} - e_{jk})^2}.$$

Thus two tracts with exactly the same racial and ethnic composition will be considered to be at racial and ethnic distance zero. Two tracts with extreme racial and ethnic compositions (e.g., one is 100% Italian whereas the other is 100% Polish) will have a maximal r/e distance of $100\sqrt{2}$. We display the distribution of ethnic distances for 1980: as is well known from even casual observation, Chicago is indeed quite segregated, with the modal distance being between 120 and 140.

[9] Marsden also considered age, gender, education, and religion. All these attributes exhibit a statistically significant positive degree of assortative matching, over and above random matching.

[10] The same statistic is 2.9 for hispanics, 3.1 for asians, 2.6 for whites.

[11] We use the percentage of Black, Native American, Asian and Pacific, Hispanic, White, German, Irish, Italian, and Polish persons 16 years and older in each tract.



3.3 Occupations

The last distance metric that we propose focuses on the informational content of social interactions. From our unemployment perspective, not all contacts are meaningful. We would like to keep track of those social ties that are more likely to convey useful information about job openings, or to generate referrals. For example, if agent A is a graphic designer and agent B is a doctor, even if they appear in each other's personal network it is unlikely that they would communicate any useful information about jobs to each other.

Again, there exists a certain amount of evidence on this. For example, Schrader [22] uses a survey of about 300 middle level managers in the context of the U.S. specialty steel industry. He finds strong support for the hypothesis that information flows quite freely through professional networks. Granovetter [6] reports that a significant fraction of tips on job openings comes from business acquaintances and contacts with similar occupations. Finally, a study by O'Regan and Quigley [18] finds that a youth's industry affiliation is positively affected by whether his/her parents work in the same industry. In other words, a job opening mentioned by agent A to agent B is likely to be in the same industry as that in which agent A works.[12]

Therefore, we think it may be relevant to construct a distance metric between pairs of tracts that is based on the within-tract distribution of occupations, in order

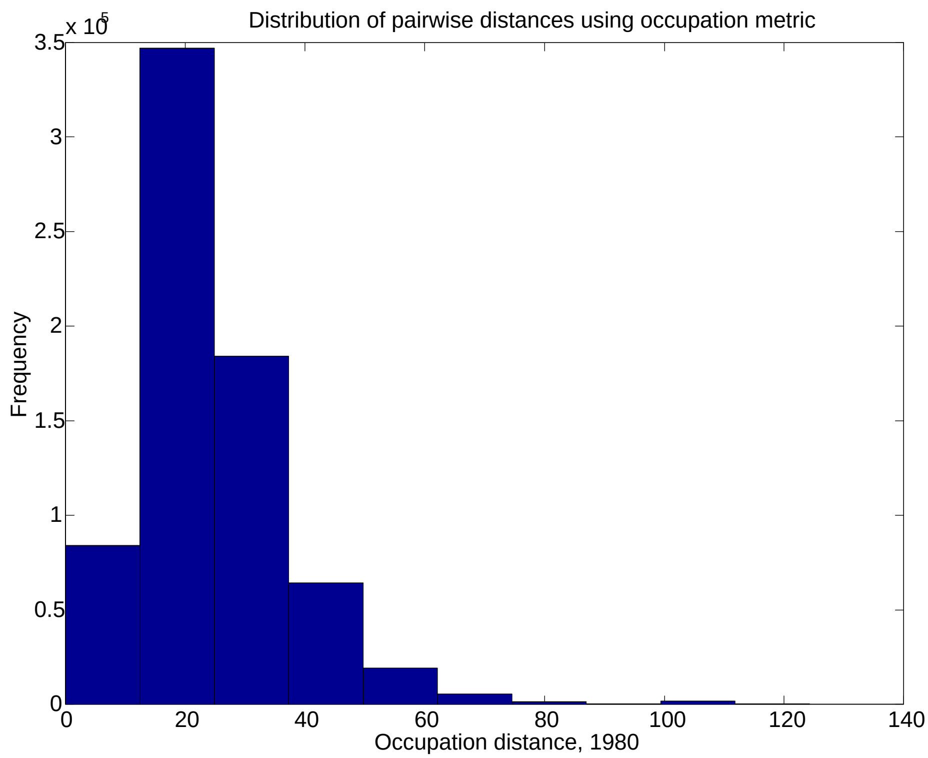
[12] Although this observed correlation may also be due to demand shocks.

to take into account the potential usefulness of the informational content of social networks. We propose the following.

Occupational distance. OD_{ij} is the euclidean distance between the vector o_i of percentages of workers in 13 different occupations in tract i and the corresponding vector o_j in j:[13]

$$OD_{ij} = \sqrt{\sum_{k=1}^{13}(o_{ik} - o_{jk})^2}.$$

The interpretation of this metric analogous to that of the race/ethnicity one, tracts with similar occupation proportions are close. However as can be seen below, the distribution of occupational distances is quite different than that for race/ethnicity distance.



3.4 Combinations of distance metrics

According to the definition of the racial/ethnic metric, or the occupational metric, two tracts are going to be at zero distance if they have the same racial/ethnic (or

[13] The occupations are: executive, administrative, and managerial; professional specialty; technicians; sales; administrative support; private housheold; protective service; other service; farming, forestry, and fishing; precision production, craft, and repair; machine operators, assemblers, and inspectors; transportation and material moving; handlers, equipment cleaners, helpers, and laborers.

occupational) composition, even though they are located at the opposite physical ends of the city. This may not be entirely satisfactory. One might think that the appropriate social distance metric, in order to identify who is more likely to talk to whom, is some combination of physical distance (or travel time) and racial/ethnic or occupational distance. Thus areas with exactly the same ethnic composition but at the opposite ends of the city may interact less than areas with slightly different ethnic compositions but much closer geographically.

In order to take this possibility into account, we also estimate ACFs of unemployment as a function of pairs of metrics. In this case, the estimated correlation between any two tracts will be a function of their distances with respect to both metrics. We can then examine how unemployment clustering depends on different combinations of the two distances. Consider, for example, physical and racial/ethnic distance. Fixing racial/ethnic distance at a level close to zero (say 20), and tracing the ACF as a function of physical distance alone, one can get an idea of how spatial correlation varies with physical distance, *conditional* on ethnic distance being equal to 20. Several possibilities may arise. The ACF may decay monotonically in both directions, or be flat with respect to one metric and only vary with the other.

We consider the following combinations of different metrics: physical and racial/ethnic distance; physical and occupational distance; travel time and racial/ethnic distance; travel time and occupational distance. In other words, we take combinations of the two types of geographic distance with racial/ethnic metric on the one hand, and with the occupational metric on the other hand.

3.5 Comparison of Economic Distance Metrics

This subsection describes differences in relative locations of tracts under each of our metrics. We use a method to represent each of our constructed metrics as a configuration of points on the plane, and compare this configuration to a map of tracts' physical locations. Specifically, we use a method called multidimensional scaling (MDS) to construct a configuration of points in two dimensions whose interpoint distances approximate those for each of the metrics. Essentially, our algorithm constructs a configuration using the first two principal components of a standardized version of a distance matrix. Of course the MDS configuration is unique only up to a choice of location and orientation, so we can only compare relative distances under each metric.[14] A goodness of fit statistic for the MDS configuration is available that can be roughly interpreted as the percentage of the variation in original distances captured by the fitted configuration. It is roughly analogous to the percentage of variance explained by the first two principal components of a covariance matrix. A brief description of MDS and this goodness of fit statistic is contained in Appendix

[14] To facilitate comparisons, we have translated and rotated each fitted configuration to line up (to the extent possible) with that based on physical distance using a method called Procustes rotation. See Mardia et. al. (1979) for a complete explanation of this procedure.

A (see, e.g., Mardia et. al. (1979) for a thorough exposition).

We plot the first tract in each of the 75 community areas to give an idea of these configurations under various metrics. Fifteen of these tracts are labeled with their community name so that changes in relative positions and clustering of these tracts can be examined. These areas are Armour Square, Austin, Bridgeport, Clearing, Dunning, Englewood, Gage Park, Hyde Park, Lincoln Park, Loop, Morgan Park, Rogers Park, South Chicago, South Shore, and Uptown.

The physical locations of the tracts are depicted in Figure 1. The origin on this map is centered at the geographic center of the points, near the Bridgeport neighborhood. The vertical and horizontal axis represent deviations in kilometers from the center. The axes are not labeled, however, in order to emphasize that relative distances between tracts are the objects of comparison across metrics, units will vary.

Figure 2 reports the MDS configuration for the CTA travel time metric. The goodness of fit statistic for this metric is 36%. The reason for this relatively poor fit is that there are many locations across Chicago that are close to being equidistant in terms of travel time via public transportation. Thus many points are equidistant under this metric, making it difficult to represent in a low-dimensional Euclidean space. Nevertheless the main features of the travel time metric are apparent in Figure 2. Neighborhoods that lie on elevated train lines that radiate from the center city are close in travel time to the Loop. This is why Rogers Park is relatively close to the Loop in Figure 2 despite being at the physical edge of the city. The largest distances in travel times are between those tracts that are physically far apart and do not have train lines connecting them like Dunning and Morgan Park.

The MDS configurations for our measure of ethnic distance in 1980 and 1990 are presented in Figures 3 and 4, respectively. These MDS configurations capture almost all the variation in the ethnic metrics, having goodness of fit statistics of 90% for 1980 and 96% for 1990. The clustering of community areas is striking. Predominantly minority areas of South Shore, Englewood, and Morgan Park clustered at the bottom of the Figures in both years. These three Community Areas have a proportion of black persons that ranges between 94% and 100% in 1980, between 97% and 100% in 1990. On the other hand, the cluster composed by Clearing, Dunning, Lincoln Park, and the Loop is predominantly white: in these Areas, whites make up between 89% and 98% of the population in 1980, between 87% and 98% in 1990. The ethnic composition in these clusters is remarkably stable across the two Census years.[15]

On the other hand, one can notice some interesting ethnic dynamics in the areas called Austin and Gage Park. In 1980, roughly 30% of the population in both these two areas was Hispanic, and about 80% was white; thus they were relatively close to the mainly white cluster mentioned earlier. By 1990, Hispanics made up between 45% and 60% of the population, and the fraction of whites had decreased to less

[15] Both Clearing and Dunning have a significant presence of persons of Polish origin. This presence is very stable at around 15% in both years.

than 50% in both areas. This is reflected in the MDS configuration for 1990: Austin and Gage Park are still close together because they have a similarly strong Hispanic component, but have moved away from the predominantly white cluster.

Figures 5 and 6 report the configurations for occupation distance for 1980 and 1990, respectively. The goodness of fit statistics for 1980 and 1990 are 54% and 60%, respectively. As with the ethnic metric, one can identify certain clusters in the MDS configurations. For example, South Chicago, Englewood, and Morgan Park are quite close in terms of their occupation distributions in both years. In fact, they all have between 25% and 30% of their workers in administrative support occupations. The other most popular occupation in all three areas is other service. One can also notice a shift between 1980 and 1990 from manufacturing to service sectors: precision production, repair, and machine operators become less present, and sales personnel or protection services increase.

The other most noticeable cluster is formed by Lincoln Park, Hyde Park and the Loop. Here the most popular occupations are executive, managerial, and professional specialty. Sales and administrative support are also strong.[16] One can notice that the South Chicago, Englewood, Morgan Park cluster has gotten closer to this latter cluster between 1980 and 1990, probably due to the shift out of manufacturing and into services.

4 Spatial Econometric Model

This section contains a brief description of our spatial econometric model.[17] We model observation i as being located at a point s_i in a Euclidean space. The basic model of dependence is that the distance between observations' positions, corresponding to their economic distances, characterizes the dependence between their random variables. If observations i and j are close, then their random variables, say X_{s_i} and X_{s_j}, may be very highly correlated. As the distance between s_i and s_j grows large, X_{s_i} and X_{s_j} become closer to being independent.

Formally, we assume that our vector of variables X_s is stationary and satisfies regularity conditions in Conley [2].[18] Stationarity means that the joint distribution of X_s for any collection of locations $\{s_i\}_{i=1}^m$ (i.e., $\{X_{s_1}, X_{s_2}, ..., X_{s_m}\}$) is invariant to shifts in the entire set of locations $\{s_i\}_{i=1}^m$. So, for example, the covariance of X_{s_i} and X_{s_j} is a function of $s_i - s_j$. Furthermore we assume that this covariance is a function

[16] It is interesting to note that Hyde Park has an increasing fraction of administrative support personnel over time, perhaps due to the expansion of the University of Chicago.

[17] A more complete description of this model can be found in Conley[2].

[18] The assumption of stationarity can be relaxed to allow non-explosive processes that have covariances that vary over space. In this case our covariance function can be interpreted as an average of non-stationary covariances.

The foremost regularity condition is that the process is mixing, that X_s and X_r become asymptotically independent as the distance between s and r goes to infinity.

of distance, not the direction of the vector $s_i - s_j$:

$$cov(X_{s_i}, X_{s_j}) = f(\|s_i - s_j\|). \quad (1)$$

We will use estimates of this spatial covariance function to describe the comovement of variables as a function of distances.

To estimate the spatial autocovariance function in Equation 1, we use a nonparametric estimator of the spatial autocovariance function. The estimator is essentially that proposed by Hall et. al. [7]. The autocovariance at distance δ is estimated by a local average of cross-products of de-meaned observations that are close to δ units apart. Letting $D_{ij} = \|s_i - s_j\|$, we estimate $f(\delta)$ with:

$$\hat{f}(\delta) = \sum_{i=1}^{N} \sum_{j=1}^{N} W_N[|\delta - D_{ij}|](X_{s_i} - \bar{X})(X_{s_j} - \bar{X}).$$

Where $\bar{X}$ is the sample mean of X, and the weight function $W_N(\cdot)$ is normalized to sum to one. In other words, we run a kernel regression of $(X_{s_i} - \bar{X})(X_{s_j} - \bar{X})$ on D_{ij}. We require $W_N(\cdot)$ to be a function of sample size that will concentrate its mass at zero as the sample becomes arbitrarily large at an appropriate rate. Thus, in large samples, the spatial covariance at distance δ will be estimated by an average of cross-products of only those observations that arbitrarily close to δ units apart and $\hat{f}$ will be consistent.

We will also estimate a generalization of this model that effectively allows us to consider two different distance metrics. To do this, we can simply interpret each observation's position as reflecting two metrics. For example if $s_{1,i}$ corresponds to the physical location of observation i and $s_{2,i}$ describes its ethnic composition, then we could index this observation by: $s_i = \begin{bmatrix} s_{1,i} \\ s_{2,i} \end{bmatrix}$. Now, rather than restricting covariances to depend on the distance between s_i and s_j we can allow them to depend on distances between these two components:

$$cov(X_{s_i}, X_{s_j}) = f(\|s_{1,i} - s_{1,j}\|, \|s_{2,i} - s_{2,j}\|). \quad (2)$$

Covariances here can depend on distance according to both metrics. We estimate this more general covariance function with a nonparametric regression as above. Estimates of 2 will be kernel regressions of $(X_{s_i} - \bar{X})(X_{s_j} - \bar{X})$ on two different measures of distance between observations i and j.

4.1 Testing Spatial Independence

We take a slightly unusual approach to conducting a test of whether there is spatial independence. Instead of using a limiting distribution of $\hat{f}$ to test the implication that there is zero spatial correlation, $f(\delta) = 0$, we plot an acceptance region for the specific null hypothesis of spatial independence. Then our hypothesis test can be done by simply observing whether our point estimate of f lies inside the acceptance region.

To compute an acceptance region for the hypothesis of spatial independence we employ a simple bootstrap techinque. We hold the sample locations fixed and simulate draws from a distribution with the same stationary (marginal) distribution as our data but with spatial independence. To do this simulation, we just sample with replacement from the empirical marginal distribution of our variables. For each of these bootstrap samples, which by construction are spatially independent, we can calculate an bootstrap estimate of f exactly as we had done for the original data. For each value of δ we take an envelope containing say 95% of our bootstrap estimates to give us an approximate acceptance region for the hypothesis of spatial independence.

We prefer this bootstrap method to tests based on limiting distributions for reasons beyond its simplicity. Tests based on the limiting distributions for our local average estimates of f will tend to be unreliable in the presence of spatial dependence. Because the estimator relies on local averages, estimates of asymptotic variances will be the same as if the data were spatially independent. There is evidence that such asymptotic approximations can be very misleading in time series applications of local average methods to data with a high degree of dependence (see e.g. Robinson [20], Pritsker[19]). Since, *a priori*, we expect there to be a significant dependence between observations we want to avoid overstating the information in our sample by using these estimators to deliver pointwise standard errors. Furthermore, we want to entertain the possibility that our economic distances are measured with error. In this case, our estimator $\hat{f}$ will recover a weighted average of the true autocovariances. The positive weights will be an unknown function of measurement errors so we will be unable to compute standard errors based on a limiting distribution. However, our bootstrap still computes a valid acceptance region for the null of spatial independence for the resulting statistic.

5 ACF Estimates

5.1 One Metric Spatial ACF estimates

We now report the results of our spatial ACF estimates for each metric separately. These ACF estimates are formed using the kernel regression described above to estimate f and then normalizing it by dividing by the sample second moment to form a correlation function estimate.[19] In each plot, we represent both the estimated ACF for the raw unemployment rate (as a solid line), and the ACF for the residuals from

[19]The kernel used was a normal kernel in all cases with standard deviations of: 0.3 for physical, 3 for travel time, 10 for ethnic, and 2.5 for occupation distance.

the regression of unemployment on our covariates (as a dashed line). In addition, the portions of the ACFs that lie outside the 95% acceptance bands for the null hypothesis of spatial independence are marked with asterisks and circles, respectively.[20]

Figures 7 through 9 contain ACF estimates for the physical distance metric. The estimation is conducted for 1980, 1990, and for the change in unemployment rate between 1980 and 1990.[21] Figures 10-12 contain similar plots for the travel time metric; Figures 13-15 use the racial/ethnic distance metric, and Figures 16-18 complete this set of plots for the occupation metric.

The first result to be noted is that the spatial ACF of unemployment is strongly and significantly positive at distances close to zero, and decreases roughly monotonically with distance for all metrics, in both years and in the first-difference case. Therefore, spatial clustering of unemployment is quite robust to the different choices of metrics. The second interesting result that is common across metrics is that clustering increases over time: The ACF estimates for the change in unemployment rate over the decade indicate a positive and statistically significant auto-correlation, that again decays with distance. These patterns are roughly consistent with a model of local interactions, in which the evolution of the state of each area is affected by the current state of the neighboring areas.

An important difference exists between the ACF plots using physical and travel time metrics on the one hand, and ethnic or occupation metrics on the other hand. For the former ones, the ACF is positive for small distances and decays to zero at large distances, whereas for the latter metrics the ACF actually becomes strongly and significantly negative at large distances. This is especially true for the ethnic metric. Thus, two census tracts with very different racial/ethnic compositions are likely to experience divergent patterns of employment. Typically, a mostly white tract in, say, Lincoln Park, is likely to experience very low unemployment, whereas a tract with a high proportion of minorities in, say, Englewood, is likely to exhibit rather high unemployment rates.

Turning now to the ACF plots for the residuals of unemployment, one can notice important differences across metrics. For the physical metric, unemployment residuals still display a positive and statistically significant level of spatial auto-correlation at distances close to zero. Interestingly, this result holds true even for the regression in first differences.[22] A more general point can be made here about cross sectional dependence in panel data. Individual intercept effects are often used to try and account for, among other things, cross-sectional dependence in panel data. Figure 9 shows quite clearly that data can still be spatially correlated even after individual effects have been differenced out.

The degree of spatial dependence still present in the residuals of unemployment

[20] 240 bootstrap draws were used to create this acceptance region.

[21] In the latter case, the covariates are first-differenced as well.

[22] This is the type of spatial dependence that was used in Topa [23] to estimate a measure of local spillovers.

is weaker for the travel time metric, and disappears altogether for the ethnic and occupation metrics, in the level regressions. For the change regressions, the auto-correlation of unemployment residuals is still significantly different than zero, but the point estimates are very small. Thus it seems that after one controls for covariates that may affect or reflect the sorting decisions of individual agents, there is very little spatial clustering left. This is especially true as one moves towards metrics that are thought to represent the dimensions along which agents' networks develop better than mere physical distance.

Therefore, there appears to be little support in the data for a local interaction story based on local information exchanges among agents, once we use metrics that should better capture such spillovers. It seems that sorting of individuals of different types in different areas may suffice to explain the clustering observed in the raw data. However, it may be ultimately very hard to distinguish empirically between a sorting and a local interaction interpretation. The observed sorting of individuals may already reflect and incorporate the possibility of local spillovers. Agents may anticipate that they can benefit from local information exchanges and thus choose locations accordingly. The gross rent and housing value differentials observed across census tracts may also incorporate, among other things, the monetary gain from any such local spillovers. Thus, when one controls for a set of sorting variables, this may already eliminate a portion of spatial dependence that is due to local interactions.

5.2 Two Metric Spatial ACF estimates

The estimates of spatial auto-correlations as a function of pairs of metrics are reported in Figures 19 through 38. The x-axis represents physical or travel time distance, whereas the y-axis represents ethnic or occupational distance. As in the one metric case, the estimation is carried out both for unemployment itself and for its residuals from a regression on the covariates described in Section 2; again, unemployment distributions in 1980, in 1990, and in first-differences are considered for each metric. For the plots that are based on raw unemployment rates, only the point estimates are reported. Contour lines are also included, to help identify the gradient of the function. For the plots involving residuals, the ACF surfaces are represented as a mesh. The portions that lie outside the 95% level acceptance bands for the spatial independence hypothesis, are filled in and shaded.[23]

Figure 19 reports the ACF of raw unemployment in 1980, when physical and ethnic metrics are employed. If one abstracts from the peak at $(PD \approx 0, ED \approx 100)$, which is imprecisely estimated due to a paucity of tracts at those distances, the pattern is very clear. Conditional on any given physical distance, there is a very strong positive auto-correlation at low ethnic distances. The ACF is quite flat with

[23] We used a product kernel comprised of univariate normal kernels having the following standard deviations: .6 for physical, 10 for ethnic, and 2.5 for occupation, and 3 for travel time distances. The bootstrap acceptance regions were formed with 80 draws.

respect to physical distance, whereas it decreases roughly monotonically with ethnic distance. In other words, conditional on any fixed ethnic distance, physical distance does not affect spatial clustering much. This pattern is even more striking in 1990 (Figure 21), and is present also in the first differenced data (Figure 23). Thus it seems that most of the spatial clustering of unemployment is driven by the racial and ethnic distribution over the city of Chicago, and once we condition on it physical distance has little, if any, explanatory power. This may be consistent with a local interaction model of information exchange, given the evidence we have reviewed[24] on the strong inbreeding of social networks along racial or ethnic lines. However, it may be difficult to distinguish such a model from a competing assortative matching explanation.

The spatial ACF estimates for the unemployment residuals with respect to physical and ethnic metrics are reported in Figures 20, 22, and 24. There is a small significant area at approximately ($PD \approx 0$, $ED \approx 0$) in all three plots, but it is rather tiny. There is also a significant portion for 1980 at around ($PD \approx 0$, $ED \approx 120$), but it is also imprecisely estimated due to a small number of observations in that area. As for the one metric case, once we control for a set of observable characteristics in each tract, the spatial distribution of unemployment loses most of its spatial clustering.

The spatial ACF estimates for raw unemployment with respect to physical and occupational metrics reveal a different pattern (Figures 25, 27, 29). Here both physical and occupational distance do matter. The degree of spatial auto-correlation is strongest at about ($PD \approx 0$, $OD \approx 0$) and is decreasing in both PD and OD. By looking at the contour lines, it seems that the physical metric plays a larger role at short distances, whereas the occupation metric becomes more important at medium to large physical distances.

As for the residuals, the ACF surface is again rather flat (Figures 26, 28, 30). However, there is a larger statistically significant area of the ACF surface at around ($PD \approx 0$, $OD \in [0, 10]$) than we observed for the physical and ethnic metric combination. This is true both in 1980 and 1990, and in the change regression. Therefore, there is some residual spatial dependence of unemployment, even after controlling for covariates that should reflect sorting decisions by agents. Furthermore, this significant portion of the ACF surface occurs in the range of physical and occupational distance that we expect to be most conducive to useful information exchanges about jobs within agents' social networks. In fact, we have reviewed some evidence from the sociological literature according to which professional contacts and social ties with persons in similar occupations are very important in generating referrals or simply information about job openings that increase one's employment opportunities.

Finally, Figures 31 to 38 report ACF plots when we use travel time instead of physical distance as a metric. Again, we look at combinations of this metric with ethnic and occupational metrics. Similar patterns arise to the ones we have discussed above, for the estimates that involve unemployment itself. When combined with ethnic distance, travel time does not have much impact on the spatial auto-correlation of

[24] See Section 3.2.

unemployment: most of the spatial clustering is driven by the racial and ethnic distribution over census tracts. On the other hand, physical distance plays the larger role when coupled with the occupational metric. As for the unemployment residuals, most of the ACF surfaces lie within the 95% acceptance bands for the null hypothesis of no spatial dependence. Only for the travel time and occupational metric combination in 1980 (Figure 36) there is a small statistically significant region at approximately ($TD \approx 0$, $OD \approx 10$).

5.3 Covariance Decompositions in the One Metric Case

We have seen in Section 5.1 that the spatial correlation patterns of the residuals of unemployment display little, if any, significant spatial dependence, except for the physical distance metric. Thus it seems that the set of observable characteristics, that we have considered to account for agent heterogeneity as well as sorting across locations, eliminates most of the spatial dependence in unemployment rates.

It seems worthwhile then to look more closely at these covariates, to try to identify the single characteristics that contribute the most to 'explaining' the strong clustering that appears in the raw unemployment data. From experimentation with sets of conditioning variables, it seems that the most important sets of regressors in terms of marginal increases in adjusted R^2 are racial composition and human capital measures.[25] These regressors are likely to be the most influential in capturing spatial correlation as well. An issue arises here on how to best decompose spatial correlation. An orthogonal decomposition of our spatial correlation estimates into components that could be attributed to say measures of human capital, ethnic composition and other groups of covariates would be ideal. However, obtaining such a decomposition is complicated by the fact that our covariates are certainly not independent and there are multiple ways to orthogonalize them.

As a first step, we perform the following simple experiment. We re-estimate the one-metric spatial ACFs reported in Figures 7 through 18 by taking out of the set of regressors three separate types of variables one at a time (with replacement), to see how each particular characteristic affects the estimated ACF. We first exclude the spatial mismatch variable (the median commuting distance to jobs) and include all remaining variables as regressors. We then exclude the racial and ethnic composition variables (the percentage of non-whites and Hispanics in the Census tract) from the original set of regressors (including spatial mismatch). Finally, we eliminate the education variables from the original set of covariates: i.e., the fraction of people with at least a high school or college degree in the tract. The resulting ACF estimates are reported in Figures 39 through 46.

In the first exercise, it appears that excluding the spatial mismatch variable does not alter the ACF estimates in a significant way. Figures 39 and 40 report these estimates for employment rates in 1980, using physical and ethnic distance metrics.

[25] Table 1 reports the changes in adjusted R^2 following the exclusion of one regressor at a time.

There is no noticeable difference between these plots and the corresponding Figures 7 and 13 with the full set of covariates.[26] Therefore, there is little support for the spatial mismatch hypothesis playing an important role in explaining the observed patterns of spatial dependence, at least given our proxy for access to jobs.[27]

As it could be expected, the racial and ethnic composition variables do have an impact, for all metrics proposed here. Figures 41 through 44 report the revised ACF estimates for all metrics, using 1990 data.[28] Excluding the fraction of non-whites and Hispanics from the original set of covariates leaves a statistically significant positive amount of spatial correlation at distances close to zero. The maximum amount of autocorrelation ranges from about 0.2 in the physical metric case to about 0.05 in the case of occupational distance. Compared to the original set of characteristics, the point estimates of spatial autocorrelation increase substantially. However, the residual spatial dependence is still quite small relative to the autocorrelation in the raw unemployment data.

Finally, the education variables have a limited impact on our ACF estimates. The only cases in which the ACF changes significantly are for 1990, using either the physical or the occupation metric. Figure 45 and 46 report the new ACFs and can be compared with Figures 8 and 17, respectively. In both cases leaving out our human capital measures brings about a statistically significant amount of spatial dependence, at low distances. However, the amount of spatial correlation is still quite small if compared to the raw autocorrelation.

Overall, it seems that specific types of conditioning variables, taken one at a time, do not contribute much to eliminating the observed spatial dependence in unemployment. Rather, it is the combined impact of a wide range of tract characteristics that jointly 'explains' unemployment's autocorrelation patterns. Therefore, using a single index, such as the poverty level or median income, to summarize the characteristics of a given neighborhood may be misleading.

Keeping the above qualification in mind, it seems that the racial and ethnic composition within each census tract is the single most important factor in explaining the observed degree of spatial correlation of unemployment. Several explanations could be given for this result. One possibility is that the racial and ethnic distribution over the map of Chicago may be itself very strongly spatially correlated, because of sorting. We explore this aspect in Figures 47 and 48, that report spatial ACF estimates for the fraction of non-whites and Hispanics in 1980, using physical distance. Both dependent variables exhibit a large positive degree of spatial correlation, that decays with distance.[29] The same pattern holds in 1990 and using first differences. In

[26] The same lack of noticeable change holds for all metrics and for all three time periods considered, 1980, 1990 and 1990-80. The full set of ACF plots is available upon request from the authors.

[27] A similar finding was reported in Topa [23].

[28] Again, the Figures that refer to the 1980 and change regressions indicate similar results and are not reported in the paper for brevity of exposition.

[29] It is interesting to note that the spatial correlation for non-whites reaches zero at about 6 km, whereas for Hispanics it reaches zero faster, at roughly 3 km. This indicates that clusters of

addition, race and ethnicity may be important explanatory variables for unemployment, either because they may proxy for unobserved heterogeneity in skills or human capital, or because they may reflect differential access to the labor market: a certain degree of discrimination may be involved.[30]

The other possibility is that the racial and ethnic composition within each tract may be a sufficient statistic for the amount of local interactions and information exchanges about jobs. In other words, information exchanges may take place not only *across* nearby tracts but also *within* tracts themselves. This, coupled with the observation that social networks follow racial and ethnic lines quite closely, would explain why any residual spatial correlation of unemployment would be eliminated once we condition on ethnic composition variables. Of course, these are but conjectures that require further examination of the data.

6 Conclusion

This paper has tried to characterize spatial patterns of unemployment in the city of Chicago. We defined several distance metrics that, following economic and sociological considerations, we expected to closely track the dimensions along which networks are constructed. In particular we used physical distance, travel time, and the difference between the ethnic or occupational distribution within any two areas. We presented MDS representations of these metrics that illustrated some of their differences. We then presented nonparametric estimates of the auto-correlation function with respect to each metric and pairs of metrics, both for unemployment and for residuals from its regression upon tract characteristics.

Our results are mixed. For the one metric case, when the variable is raw unemployment, we find a strong and positive level of auto-correlation of unemployment at distances close to zero, for all the metrics proposed here. This spatial correlation decays roughly monotonically with distance. However, when we look at the residuals from a regression of unemployment on a set of observable tract characteristics, most of the spatial dependence is eliminated, especially when we consider ethnic or occupational metrics.

In the two metric case, some additional patterns emerge. When combinations of physical or travel time distance are used together with ethnic distance, the latter seems to drive most of the spatial dependence of raw unemployment data. The ACF are roughly constant with respect to physical distance, once we condition on any given ethnic distance. On the other hand, when one considers pairs of physical or travel time metric and occupational metric, the degree of spatial dependence of raw unemployment varies the most with physical distance (or travel time), but

non-whites (predominantly blacks) are larger geographically than those of Hispanics.

[30] Holzer [9] reports that employers may avoid hiring people of a certain race or ethnicity, or who come from specific neighborhoods.

occupational distance plays a role too. As in the one metric case, conditioning on sorting variables eliminates most of the spatial dependence of unemployment. The lone exception occurs in the case of physical and occupation metric combinations.

Finally, we report some preliminary results to address the question of which regressors are most important to eliminate the spatial correlation present in the raw data. It seems that our racial and ethnic composition variables are the single most important factor in reducing the amount of spatial dependence present in the raw data, for all years and under all metrics. On the other hand, education variables play a limited role, only in 1990 and using physical and occupation distance metrics. The spatial mismatch variable does not change our initial results in any noticeable way.

7 Appendix A

This Appendix offers a brief description of classical Multidimensional Scaling (MDS). It borrows heavily from the presentation in Mardia et. al. [14] who offer a much more complete description. Our goal is to obtain a configuration of T points in $\Re^2$ with distance matrix D of distances that is close to the raw economic distances for a given metric. Of course the configuration is unique only up to translations and rotations. For the sake of exposition, consider a general distance matrix R that has zero elements on the main diagonal and whose other elements R_{ij} are positive. Define a centering matrix $H \equiv I - \frac{1}{T}\mathbf{11}'$, a matrix A whose elements $A_{ij} \equiv -\frac{1}{2}R_{ij}^2$, and $B \equiv HAH$.

The basic approach of MDS can be made clear by considering what the matrix B would look like for a Euclidean distance matrix. As was first proved by Schoenberg [21] R can be a distance matrix for a configuration of points in a k-dimensional Euclidean space not lying in a proper subspace if and only if B is positive semi-definite and rank k. Consider a sketch of the proof of this proposition for $k = 2$.

First take the "if" part, if R were a matrix of Euclidean interpoint distances for a configuration $Z = (z_1, z_2, ..., z_T)'$ in $\Re^2$ then

$$R_{ij}^2 = -2A_{ij} = (z_i - z_j)'(z_i - z_j) \tag{3}$$

The matrix B can be written as:

$$B = HAH = A - \frac{1}{T}A(\mathbf{11}') - \frac{1}{T}(\mathbf{11}')A + \frac{1}{T^2}(\mathbf{11}')A(\mathbf{11}')$$

Define:

$$\overline{A}_{i\cdot} = \frac{1}{T}\sum_{j=1}^{T} A_{ij},\ \overline{A}_{\cdot j} = \frac{1}{T}\sum_{i=1}^{T} A_{ij},\ \overline{A}_{\cdot\cdot} = \frac{1}{T^2}\sum_{i=1}^{T}\sum_{j=1}^{T} A_{ij} \tag{4}$$

Then the elements of B can be expressed as:

$$B_{ij} = A_{ij} - \overline{A}_{i\cdot} - \overline{A}_{\cdot j} + \overline{A}_{\cdot\cdot} \tag{5}$$

Using equations 3 and 4 to rewrite equation 5 yields an expression for the elements of B:

$$B_{ij} = (z_i - \bar{z})' \left(z_j - \bar{z}\right)$$
$$\text{or in matrix form: } B = (HZ)(HZ)'$$

Thus the matrix B, referred to as a centered inner product matrix, is positive semi-definite and rank two (assuming the points in Z are not on a line).

Now consider the "only if" part. Since B is positive semi-definite and rank 2, let its nonzero eigenvalues be denoted λ and μ. Take the eigenvectors ℓ and m associated with these nonzero eigenvalues and normalize them so that $\ell'\ell = \lambda$ and $m'm = \mu$. Let a configuration of points $Z = (z_1, z_2, ..., z_T)'$ in $\Re^2$ have as first coordinates ℓ_i and second coordinates given by m_i, i.e. $Z = \left[\begin{array}{cc} \ell & m \end{array}\right]$. The spectral decomposition theorem implies that:

$$B = \left[\begin{array}{cc} \gamma_1 & \gamma_2 \end{array}\right] \left[\begin{array}{cc} \lambda & 0 \\ 0 & \mu \end{array}\right] \left[\begin{array}{cc} \gamma_1 & \gamma_2 \end{array}\right]'$$

for $\gamma_1'\gamma_1 = \gamma_2'\gamma_2 = 1$. The definitions of ℓ and m imply that $\ell = \lambda^{-1/2}\gamma_1$ and $m = \mu^{-1/2}\gamma_1$, therefore:

$$B = \left[\begin{array}{cc} \ell & m \end{array}\right] \left[\begin{array}{cc} \ell & m \end{array}\right]' = ZZ'$$

So $B_{ij} = z_i'z_j$. Now 5 can be used to show that the interpoint distances of this configuration equal R.

$$\begin{aligned} (z_i - z_j)'(z_i - z_j) &= z_i'z_i - 2z_i'z_j + z_j'z_j \\ &= B_{ii} - 2B_{ij} + B_{jj} \end{aligned}$$

Equation 5 implies:

$$B_{ii} - 2B_{ij} + B_{jj} = A_{ii} - 2A_{ij} + A_{jj}$$

$A_{ii} = 0$ so:

$$A_{ii} - 2A_{ij} + A_{jj} = -2A_{ij} = R_{ij}^2$$

Therefore the configuration Z has interpoint distances given by R.

When R is not Euclidean or is a Euclidean distance matrix of a configuration in greater than k dimensions, the above result offers the basis of the MDS fitted configuration (Torgerson [24]). Since the normalized eigenvectors of the constructed matrix B produce a configuration with distance matrix equal to R when it is a Euclidean distance matrix in k dimensions, it is reasonable to suspect that the k eigenvectors with the largest eigenvalues might, when normalized as above, give a configuration in k dimensions whose distances are close to R. This is in fact the MDS solution to finding a representation in k-dimensional Euclidean space of points with interpoint distances R : take the eigenvectors (normalized as above) corresponding to the k largest eigenvalues of the constructed matrix B as the points' coordinates. Let the MDS solution in k dimensions be summarized by a distance matrix D. The MDS solution is optimal in the sense that the configuration summarized by D minimizes the quantity: $tr\left(B-\hat{B}\right)^2$ over all $\hat{B}$ where $\hat{B}$ is the centered inner product matrix of a configuration in $\Re^k$ (Mardia [13]). Furthermore, if the matrix R corresponds to a configuration in $\Re^n$ where $n > k$, the MDS solution is an optimal projection that minimizes $\sum_i \sum_j \left(R_{ij}^2 - \hat{D}_{ij}^2\right)$for any projection of the configuration R onto k dimensional subspaces of $\Re^n$. [31]

A measure of the goodness of fit of a configuration in k dimensions to the distances in the matrix R can be obtained by looking at the ratio of the sum of the first k eigenvalues of B to the sum of the absolute value of all of its eigenvalues. This ratio provides a measure of the distance information in the matrix R captured by the MDS configuration.

[31] Proofs of both of these optimal properties can be found in Mardia et.al. [14] Sec. 14.4.

References

[1] 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.

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

[3] 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, pp. 1-37, Ann Arbor, MI: Institute For Social Research.

[4] Erbe, William et al. (1984), *Local Community Fact Book: Chicago Metropolitan Area,* Chicago: The University of Chicago Press.

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

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

[7] Hall, P., N.I. Fisher, and B. Hoffman (1992), "On the Non-Parametric Estimation of Covariance Functions", unpublished manuscript, Australian National University.

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

[9] Holzer, Harry J. (1996), *What Employers Want: Job Prospects for Less-Educated Workers,* New York: Russell Sage Foundation.

[10] Ihlanfeldt, Keith R., and David L. Sjoquist (1990), "Job Accessibility and Racial Differences in Youth Employment Rates", *American Economic Review,* Vol. 80, pp. 267-276.

[11] 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, pp. 255-265.

[12] Light, Ivan, Parminder Bhachu, and Stavros Karageorgis (1993), "Migration Networks and Immigrant Entrepreneurship", in *Immigration and Entrepreneurship,* edited by I. Light and P. Bhachu, New Brunswick, NJ: Transaction.

[13] Mardia, K.V. (1978), "Some Properties of Classical Multi-dimensional Scaling" *Communications in Statistics; A: Theory and Methods.* 7:13, 1233-1241.

[14] Mardia, K.V., Kent, J.T., and Bibby, J.M.(1979), *Multivariate Analysis.* Academic Press, London.

[15] Marsden, Peter V. (1987), "Core Discussion Networks of Americans", *American Sociological Review*, Vol. 52, pp. 122-131.

[16] Marsden, Peter V. (1988), "Homogeneity in Confiding Relations", *Social Networks,* Vol. 10, pp.57-76.

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

[18] O'Regan, Katherine M., and John M. Quigley (1993), "Family Networks and Youth Access to Jobs", *Journal of Urban Economics,* Vol. 34, pp. 230-248.

[19] Pritsker, M. (1996) "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models" *Review of Financial Studies*, Forthcoming.

[20] Robinson, P.M. (1983) "Nonparametric Estimators for Time Series" *Journal of Time Series Analysis*

[21] Schoenberg, I.J. (1935), "Remarks to Maurice Frechet's Article 'Sur la Definition Axiomatique d'Une Class d'Espace Distances Vectoriellement applicable sur l'espace Hilbert"' *Annals of Mathematics* 36:724-732.

[22] Schrader, Stephan (1991), "Informal Technology Transfer between Firms: Cooperation through Information Trading", *Research Policy,* Vol. 20, pp. 153-170.

[23] Topa, Giorgio (1997), "Social Interactions, Local Spillovers, and Unemployment", unpublished manuscript, New York University.

[24] Torgerson, Warren S. (1958), *Theory and Methods of Scaling,* New York: Wiley.

[25] Wellman, Barry (1996), "Are Personal Communities Local? A Dumptarian Reconsideration", *Social Networks,* Vol. 18, pp. 347-354.

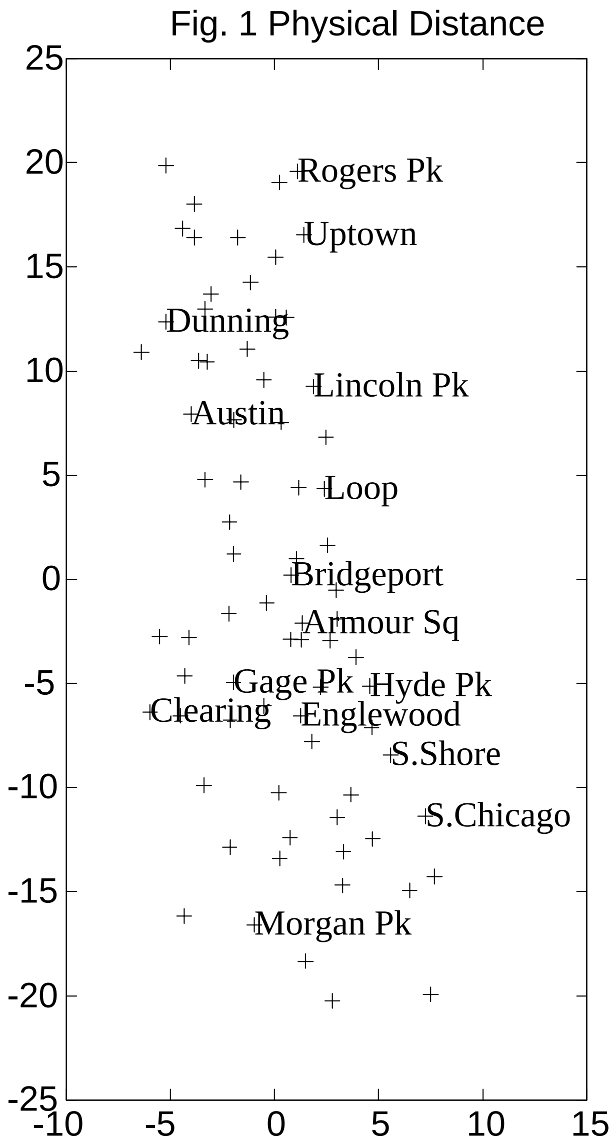


Figure 1: MDS configuration using physical distance.

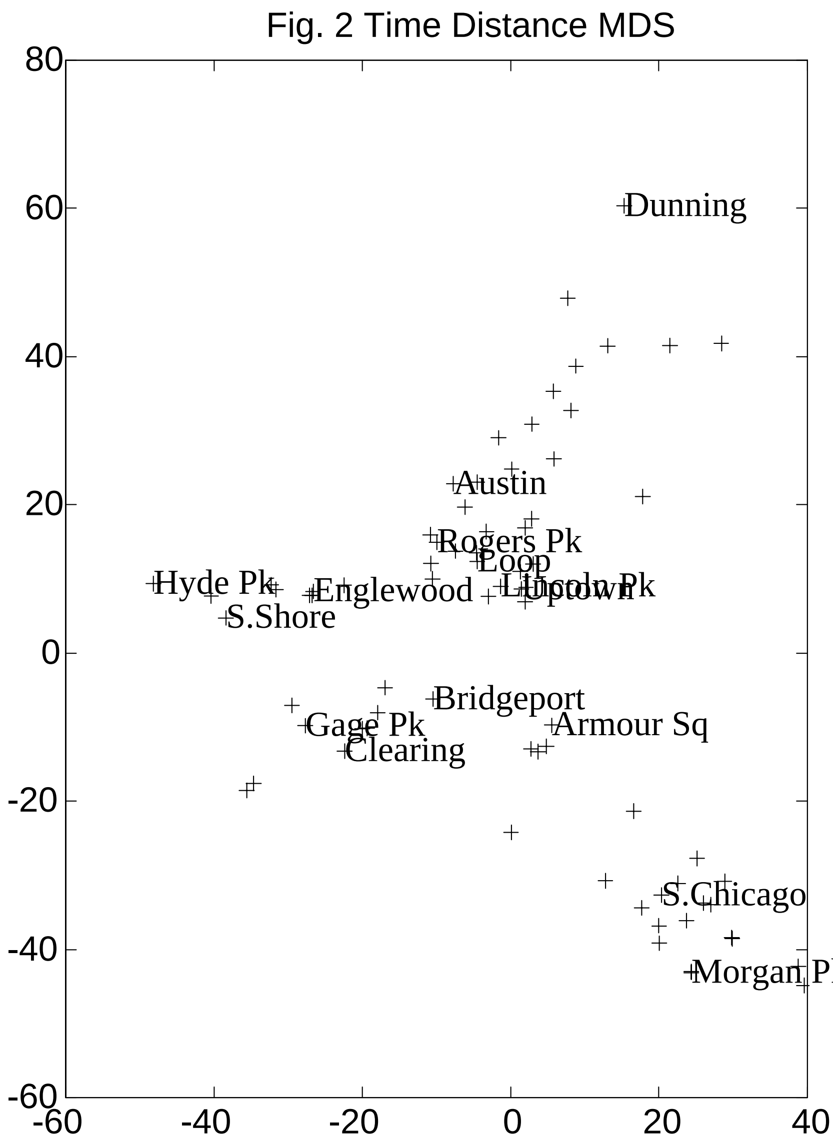


Figure 2: MDS configuration using travel time distance.

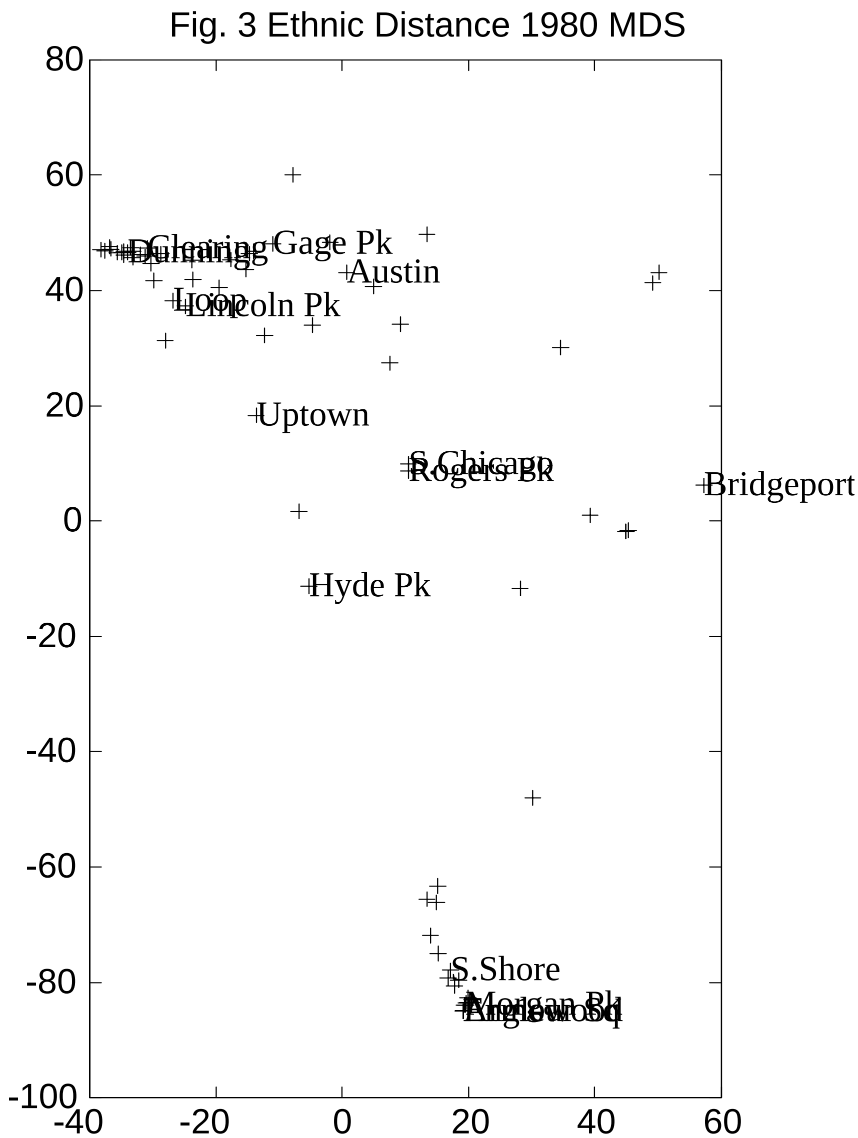


Figure 3: MDS configuration using ethnic distance in 1980.

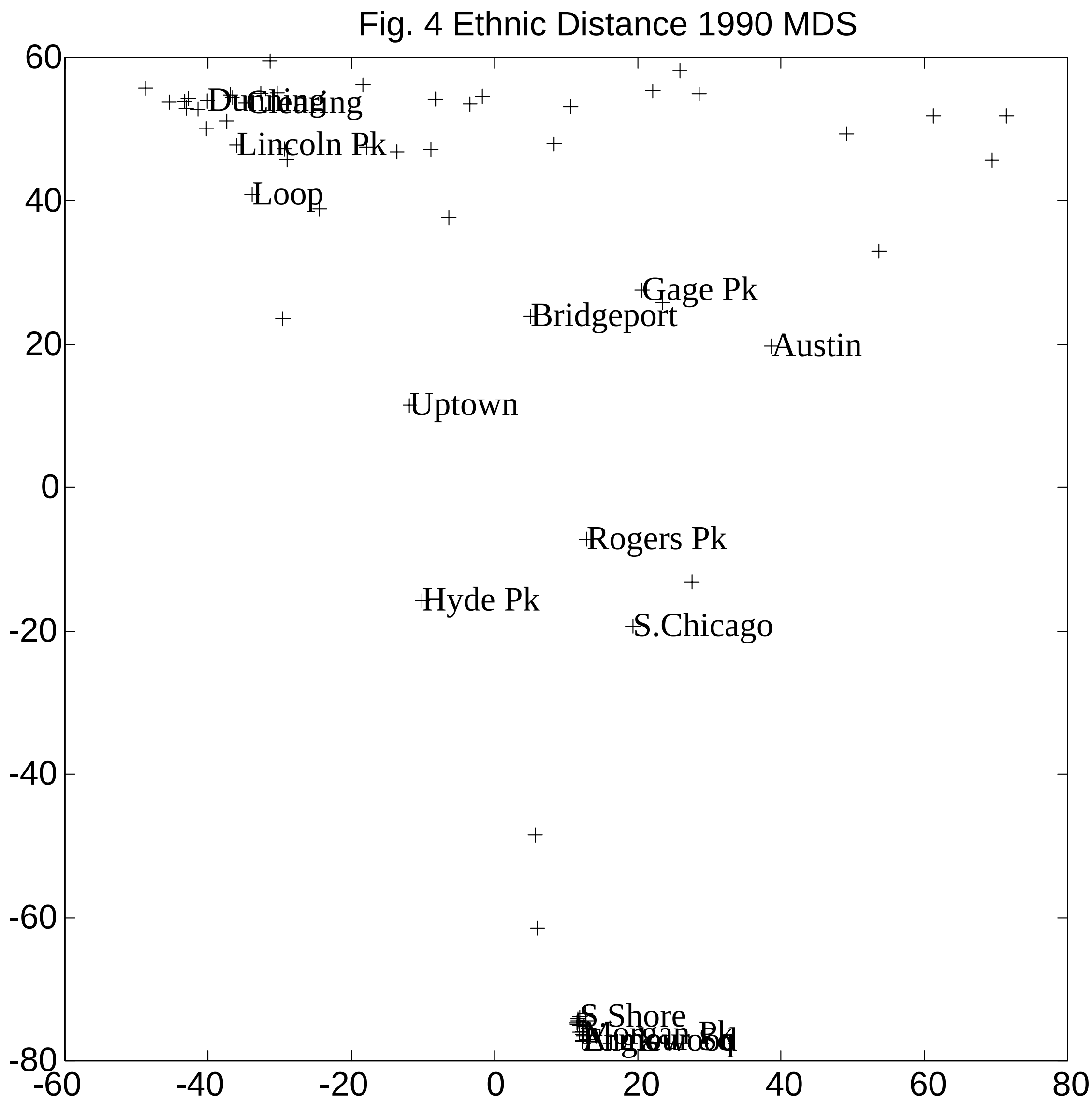


Figure 4: MDS configuration using ethnic distance in 1990.

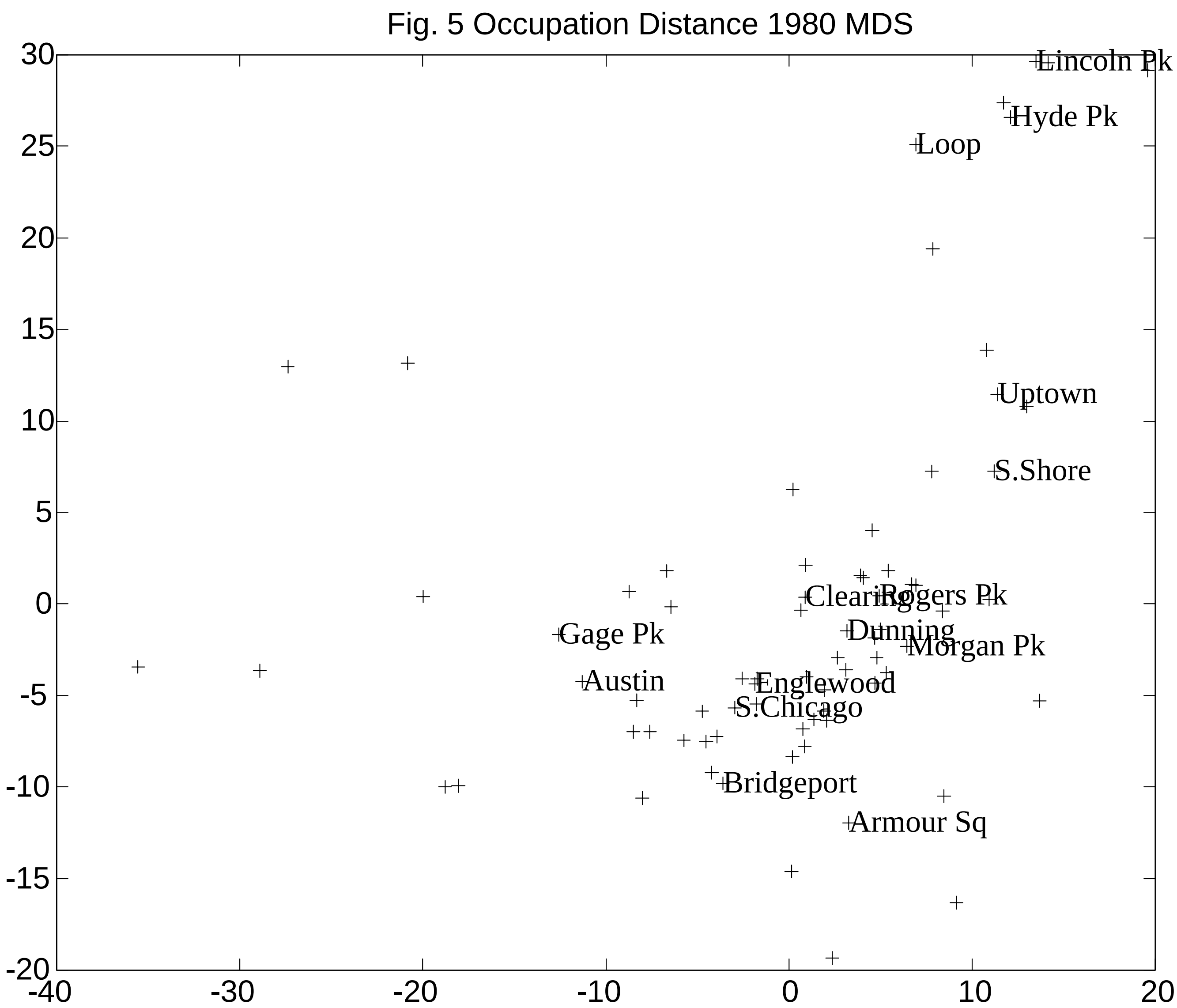


Figure 5: MDS configuration using occupation distance in 1980.

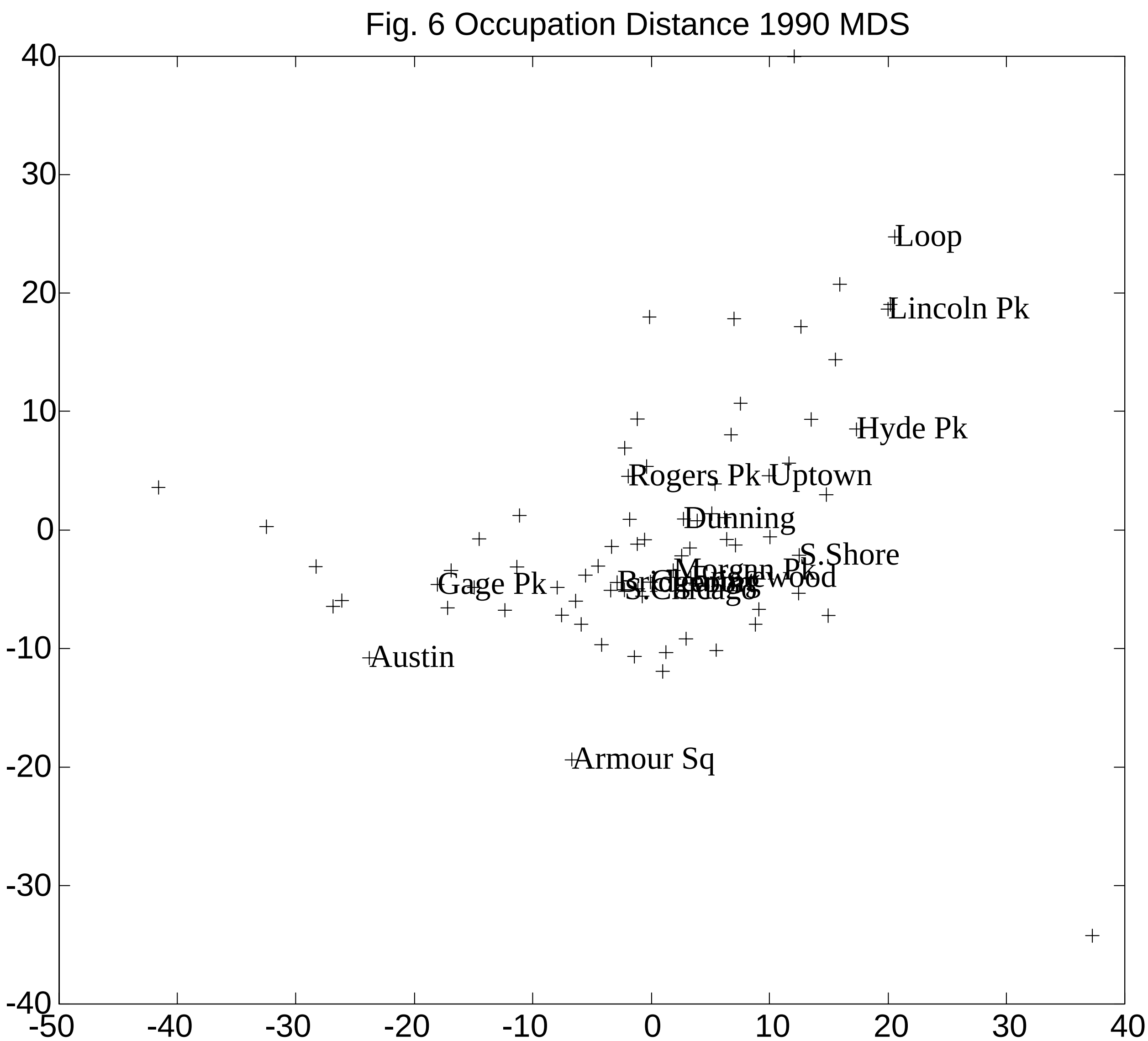


Figure 6: MDS configuration using occupation distance in 1990.

FIGURE 7 − ACF for Employment rate, 1980

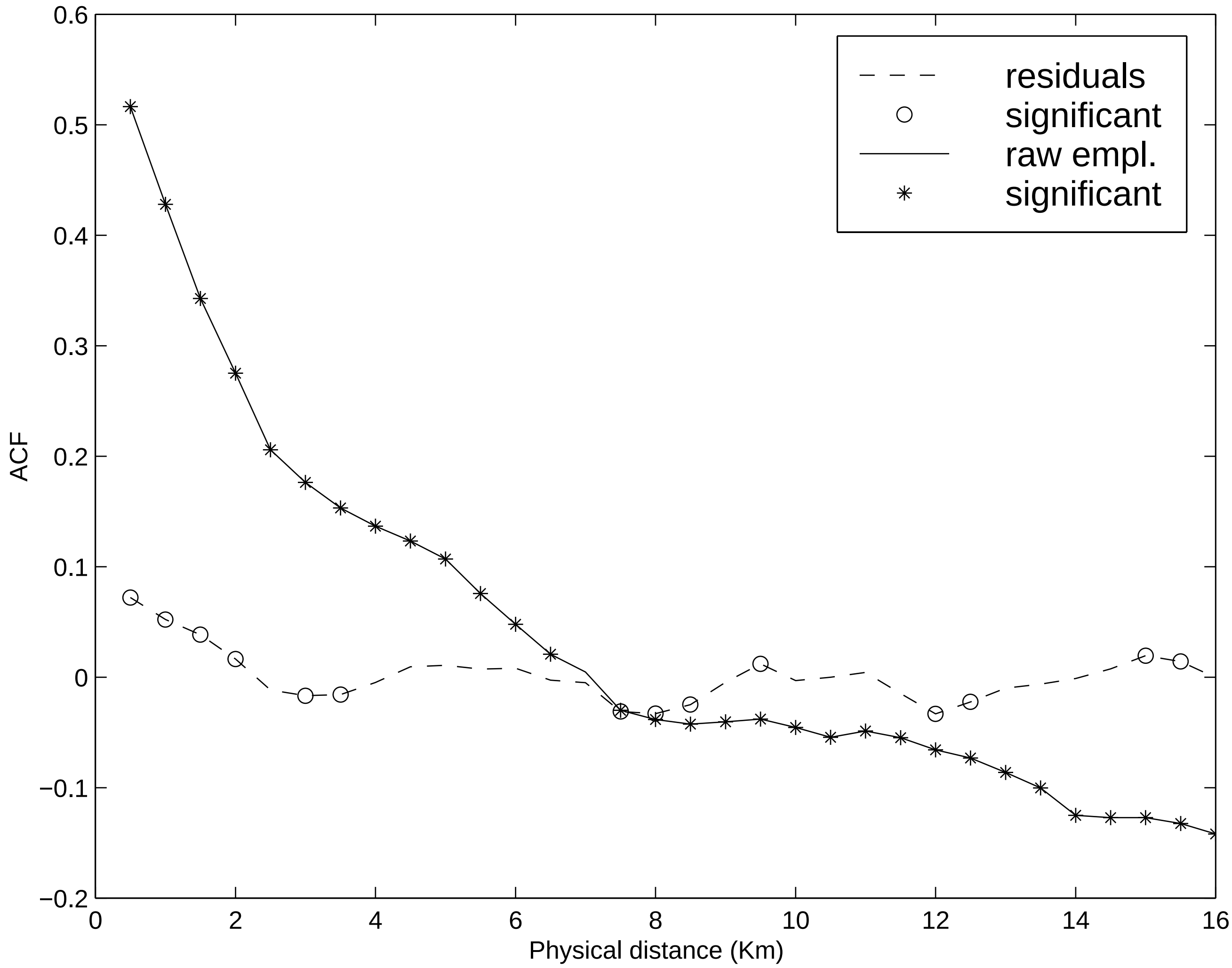
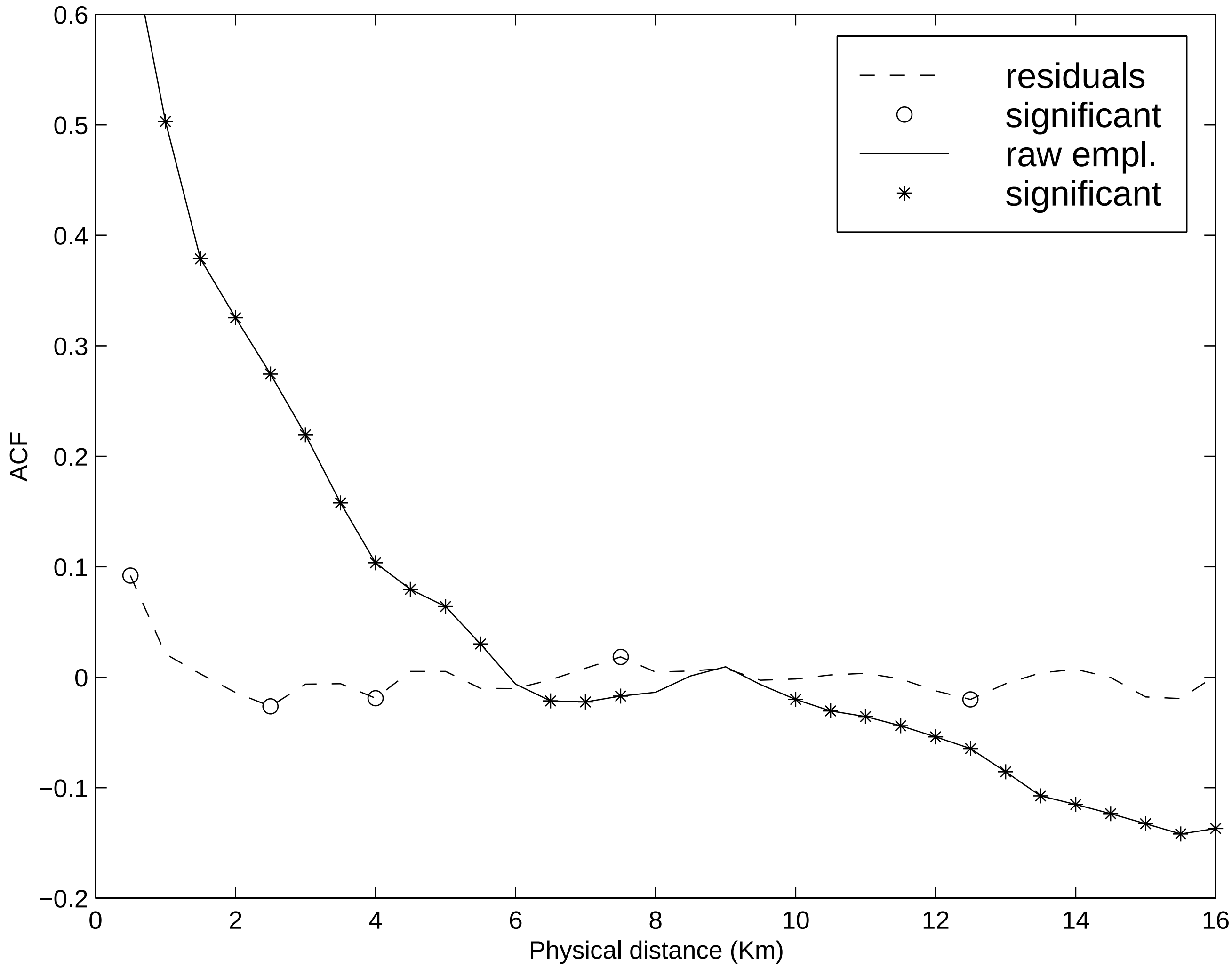


FIGURE 8 − ACF for Employment rate, 1990



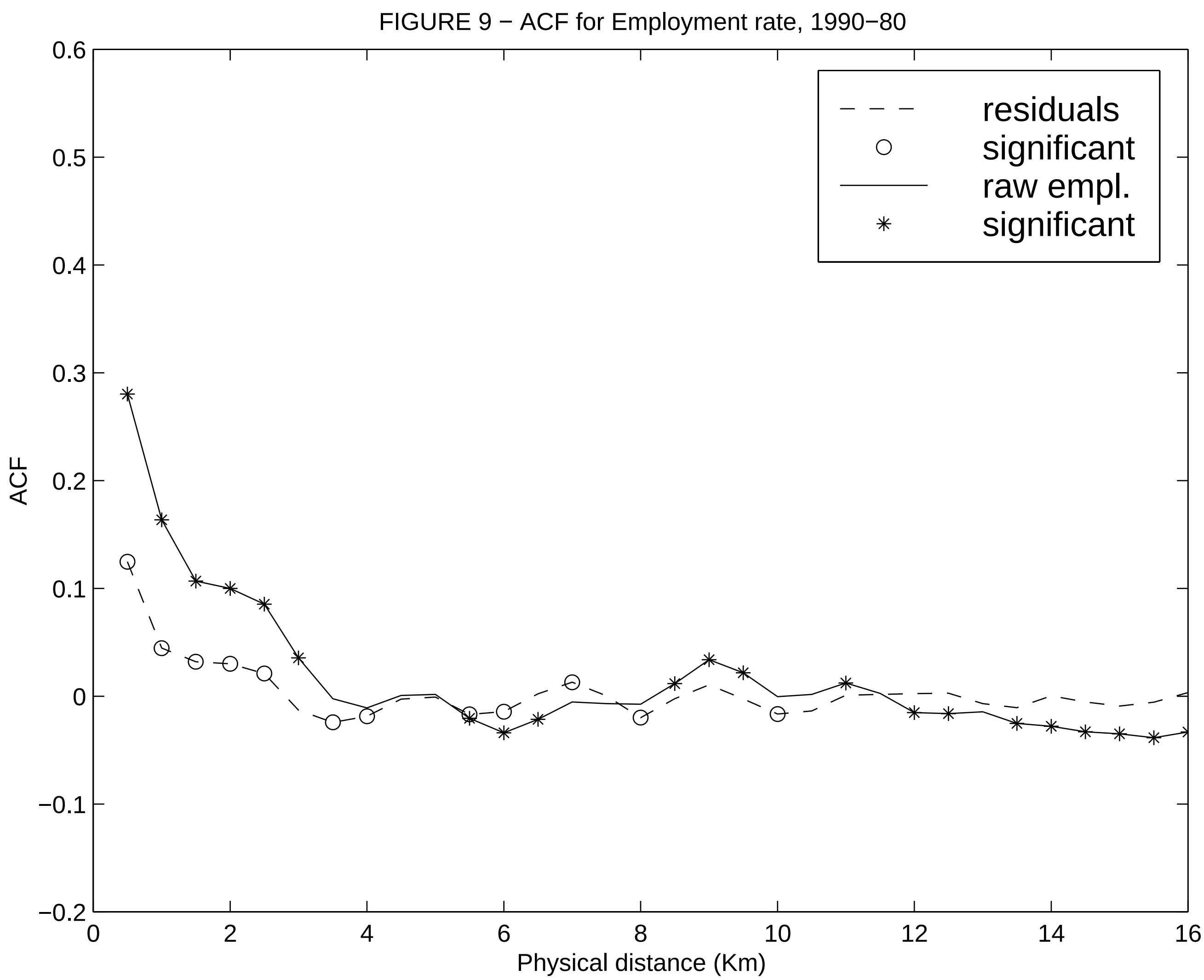


FIGURE 9 − ACF for Employment rate, 1990−80
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
ACF
0
2
4
6
8
10
12
14
16
Physical distance (Km)
residuals
significant
raw empl.
significant

FIGURE 10 − ACF for Employment rate, 1980

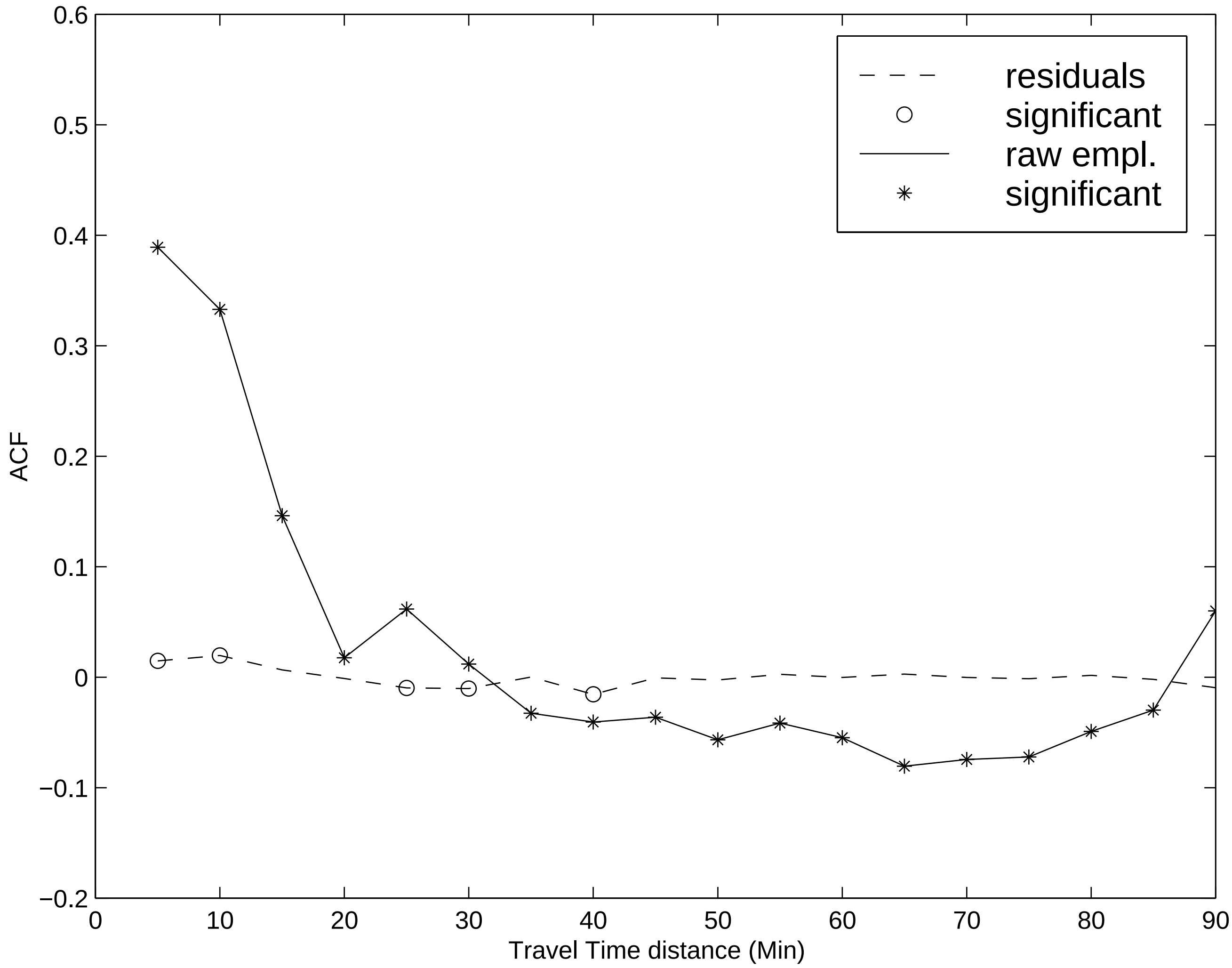


FIGURE 11 − ACF for Employment rate, 1990

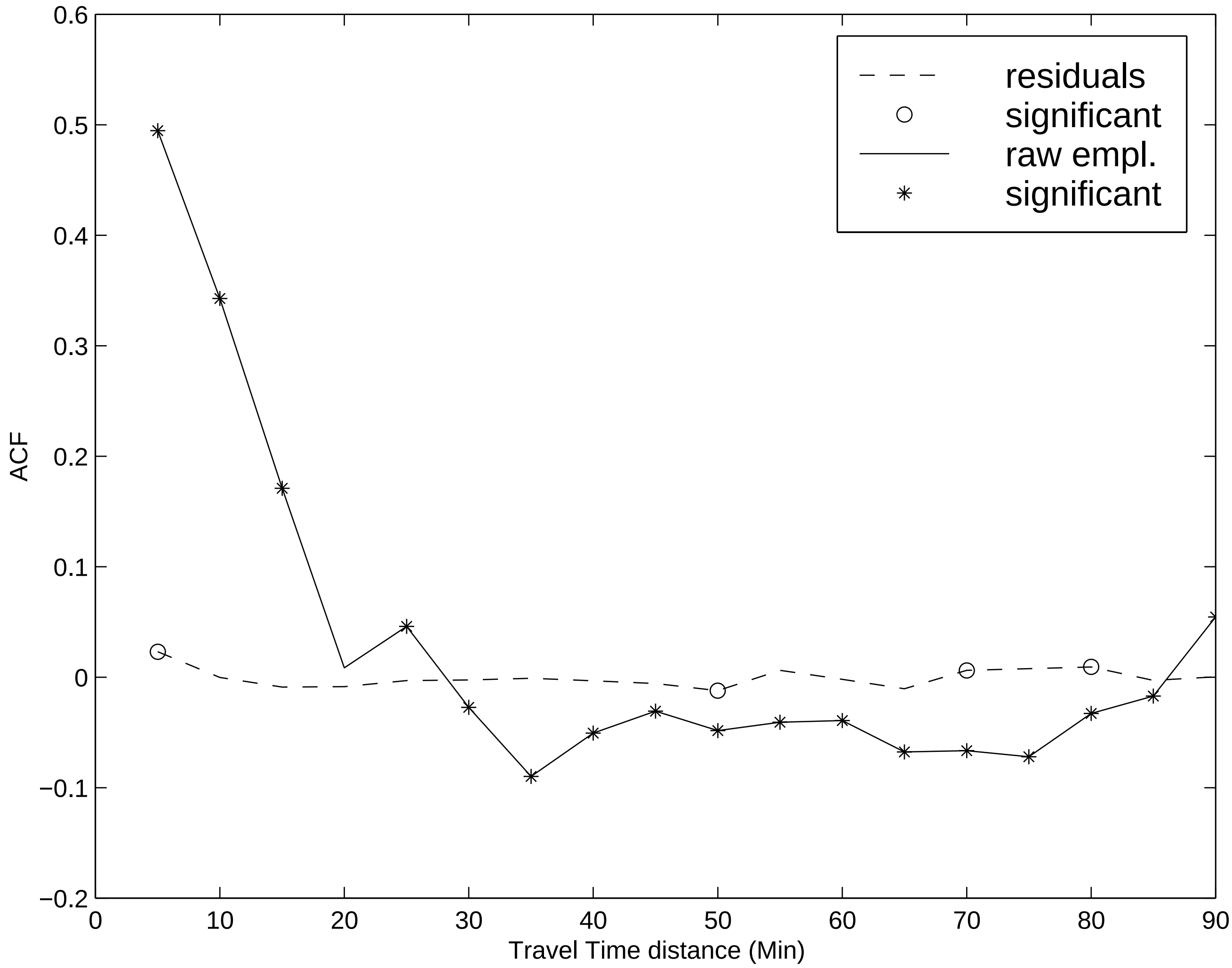


FIGURE 12 − ACF for Employment rate, 1990−80

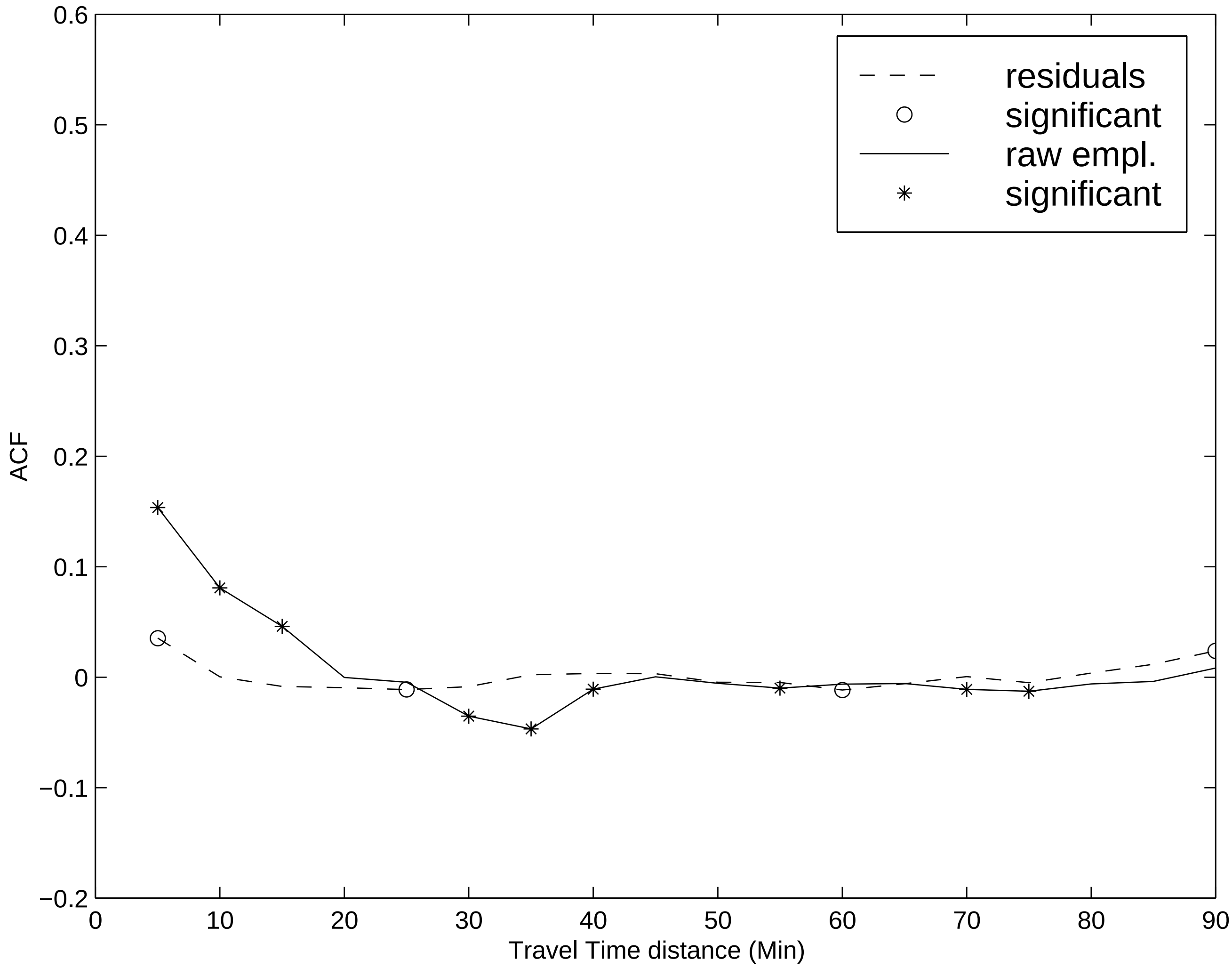


FIGURE 13 − ACF for Employment rate, 1980

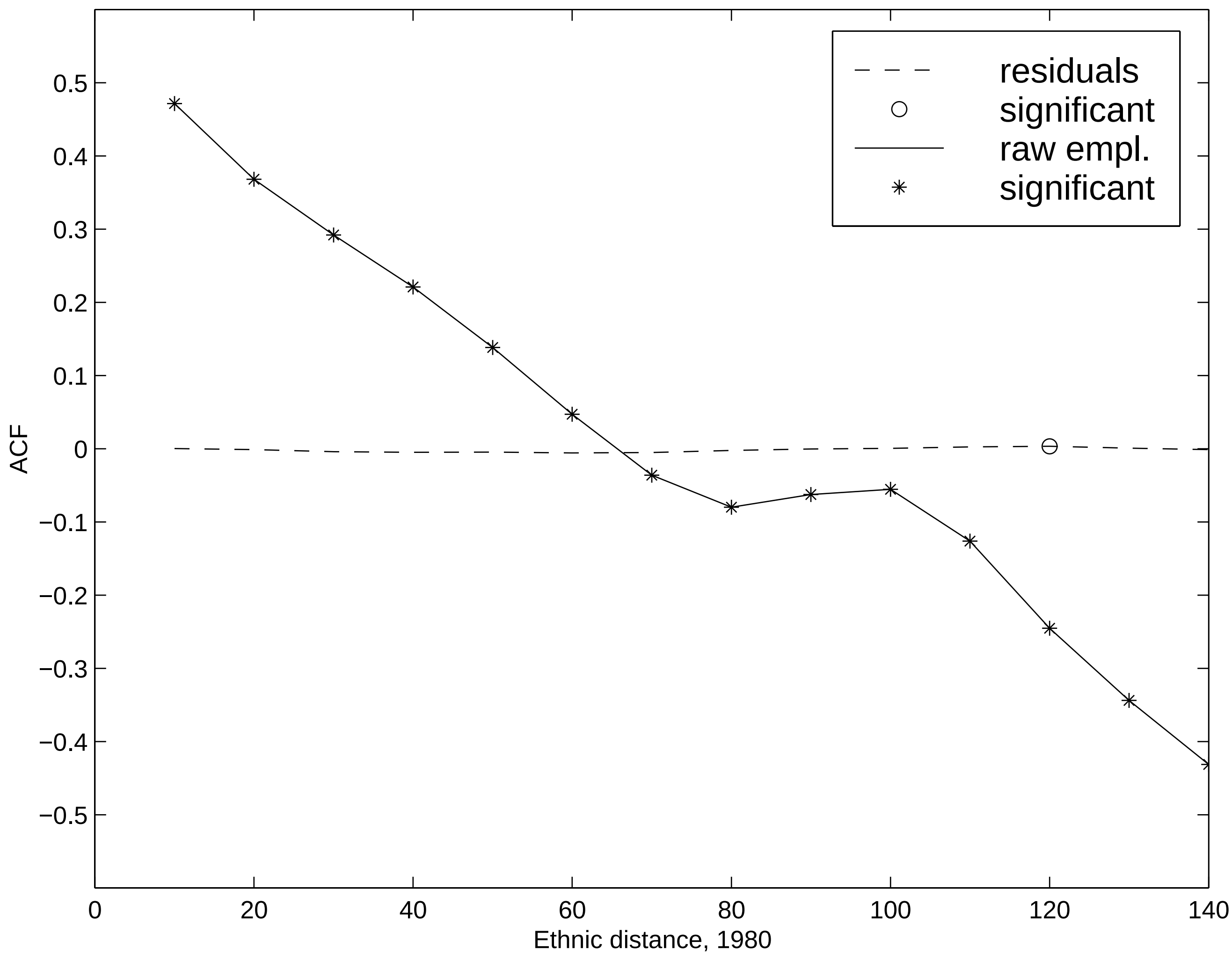


FIGURE 14 − ACF for Employment rate, 1990

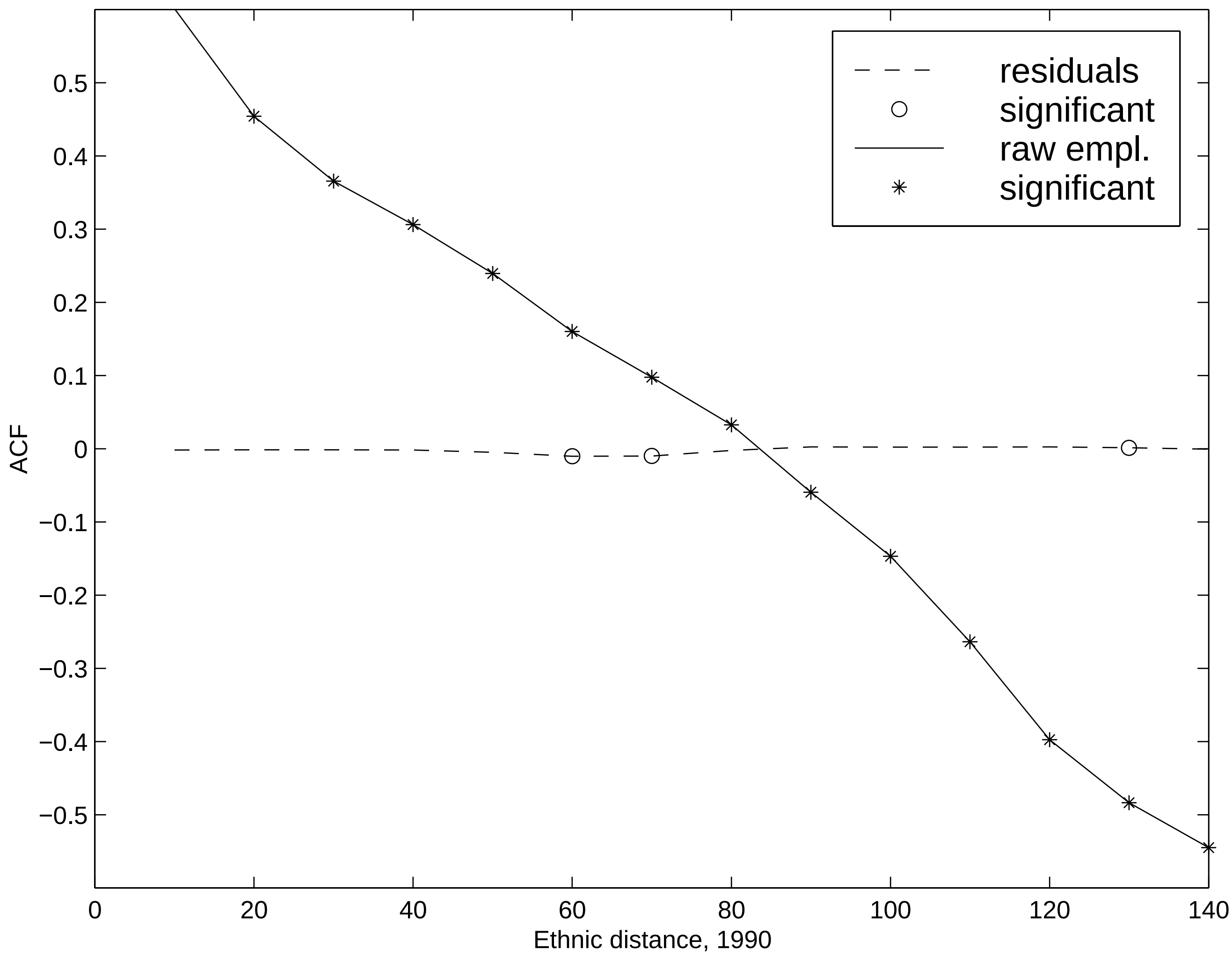
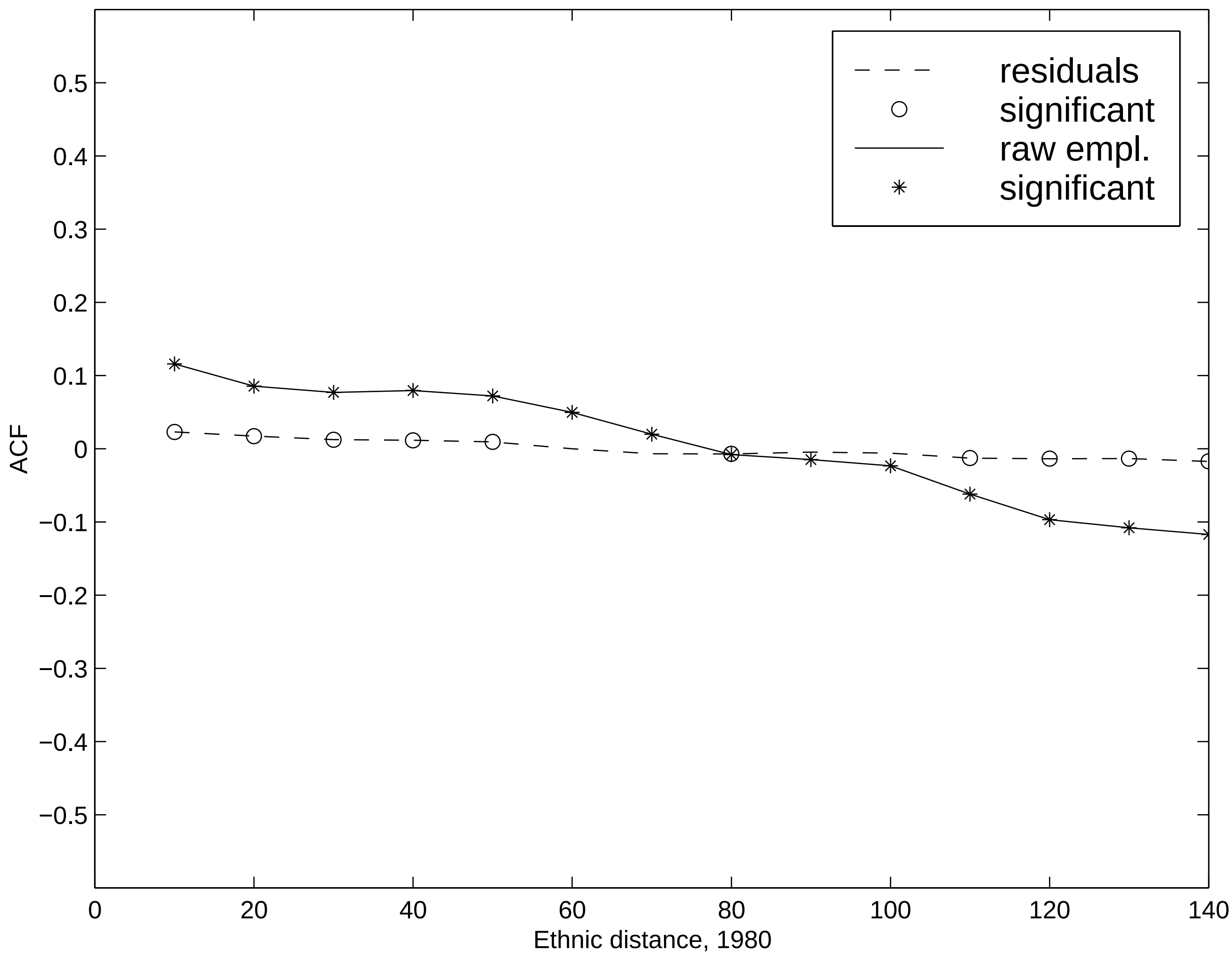


FIGURE 15 − ACF for Employment rate, 1990−80



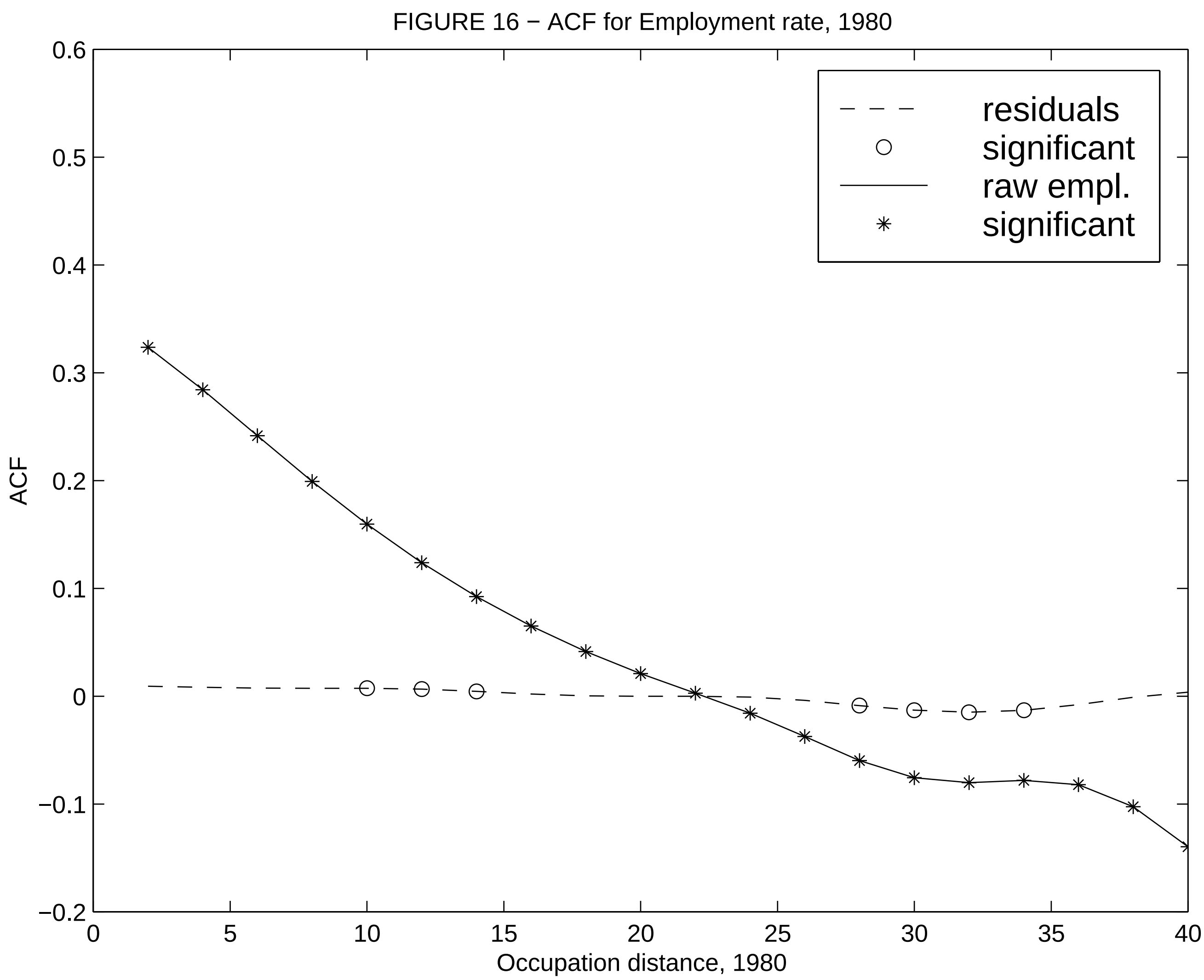


FIGURE 16 − ACF for Employment rate, 1980
residuals
significant
raw empl.
significant
ACF
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
0
5
10
15
20
25
30
35
40
Occupation distance, 1980

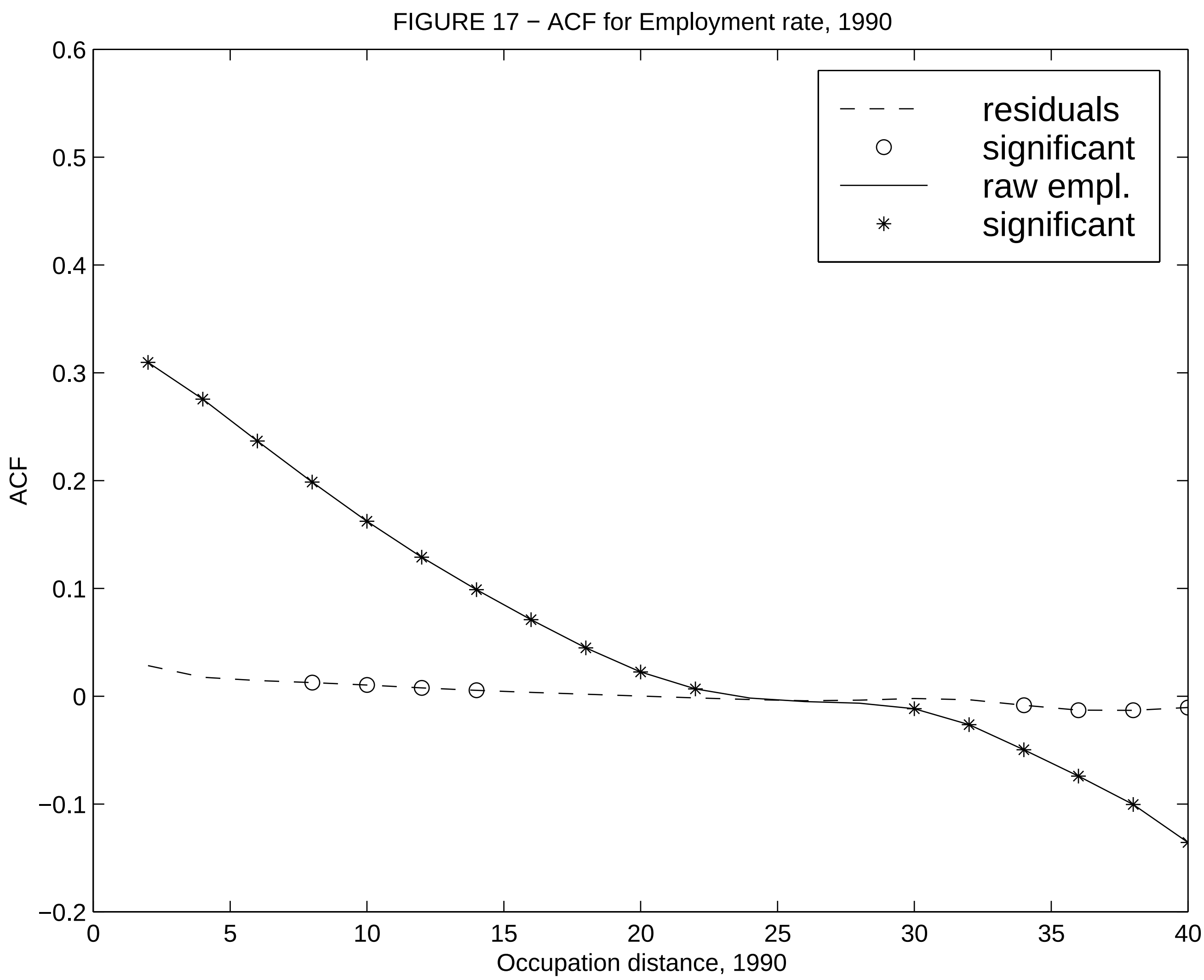


FIGURE 17 − ACF for Employment rate, 1990
ACF
Occupation distance, 1990
residuals
significant
raw empl.
significant
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
0
5
10
15
20
25
30
35
40

FIGURE 18 − ACF for Employment rate, 1990−80

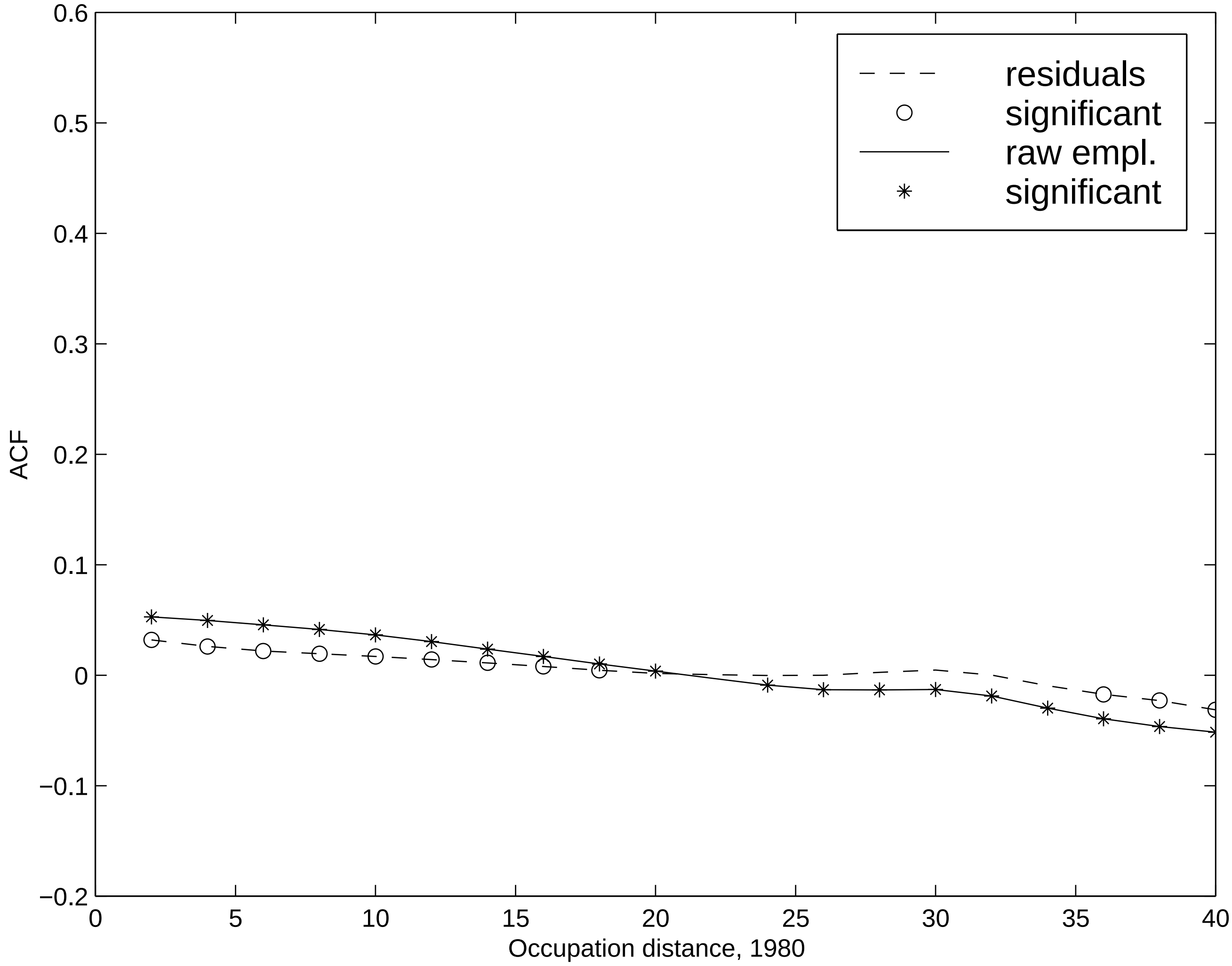


FIGURE 19 - ACF of Employment, 1980

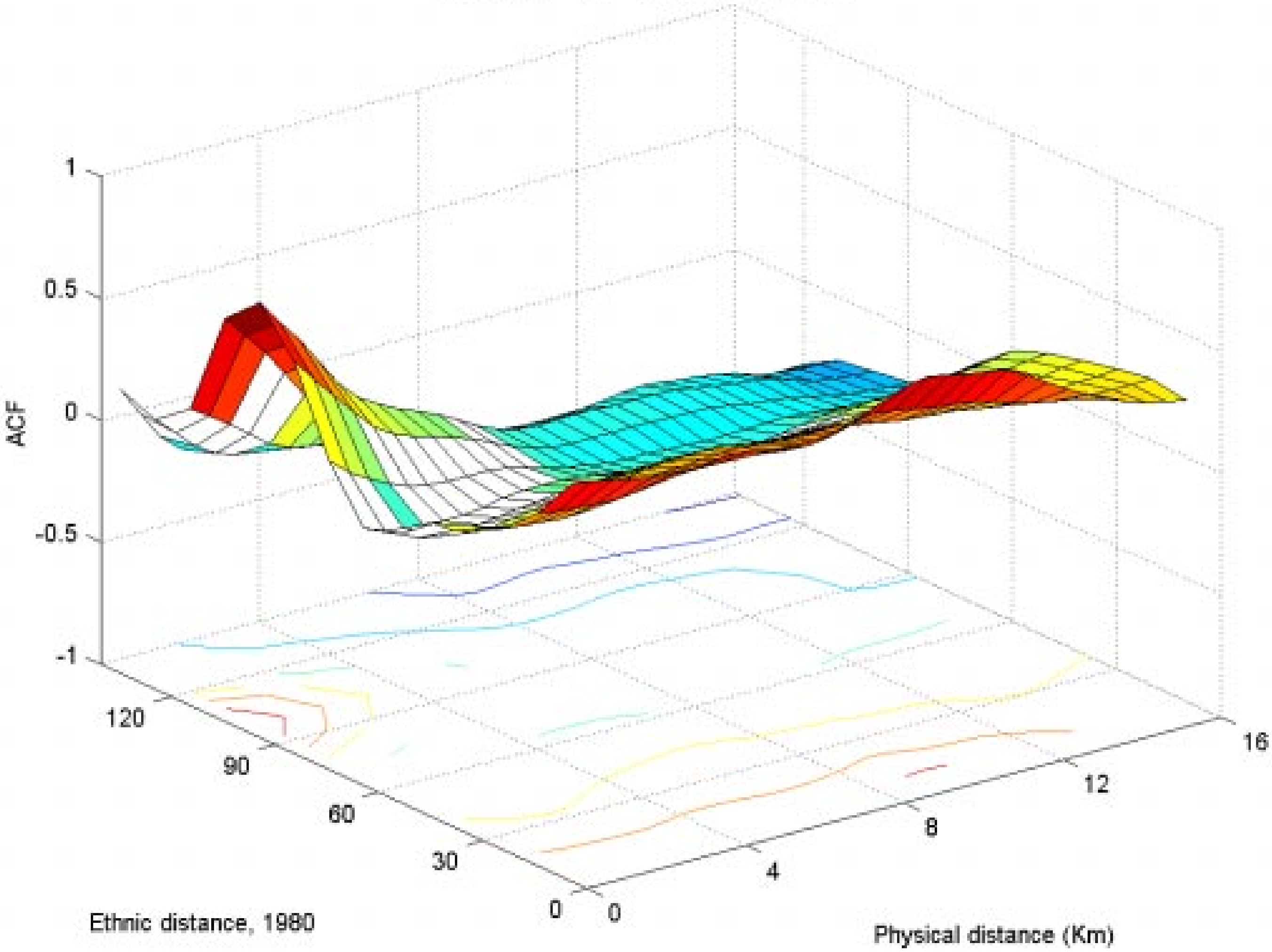


FIGURE 20 - ACF of Residuals of Empl(80) on X(80)

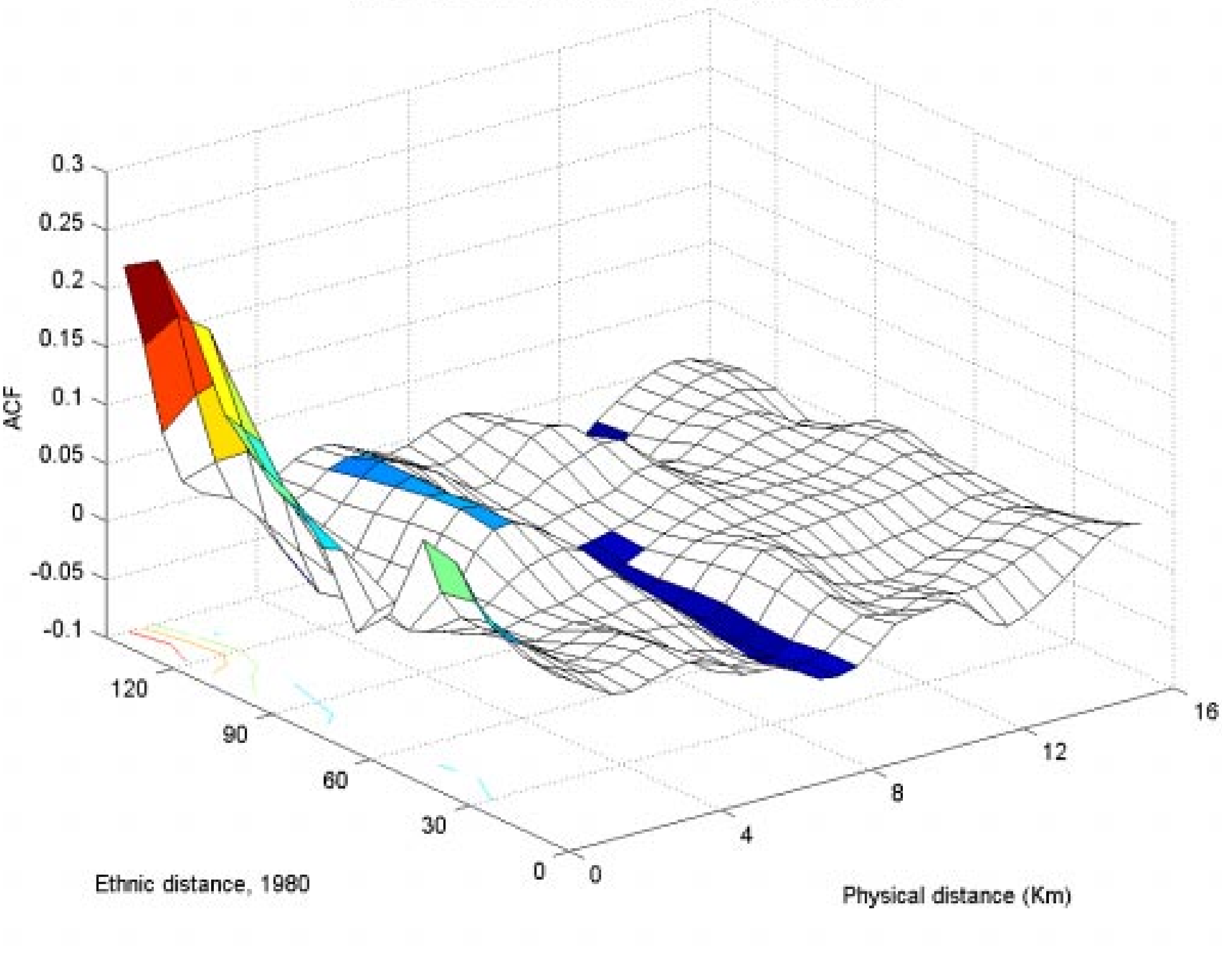


FIGURE 21 - ACF of Employment, 1990

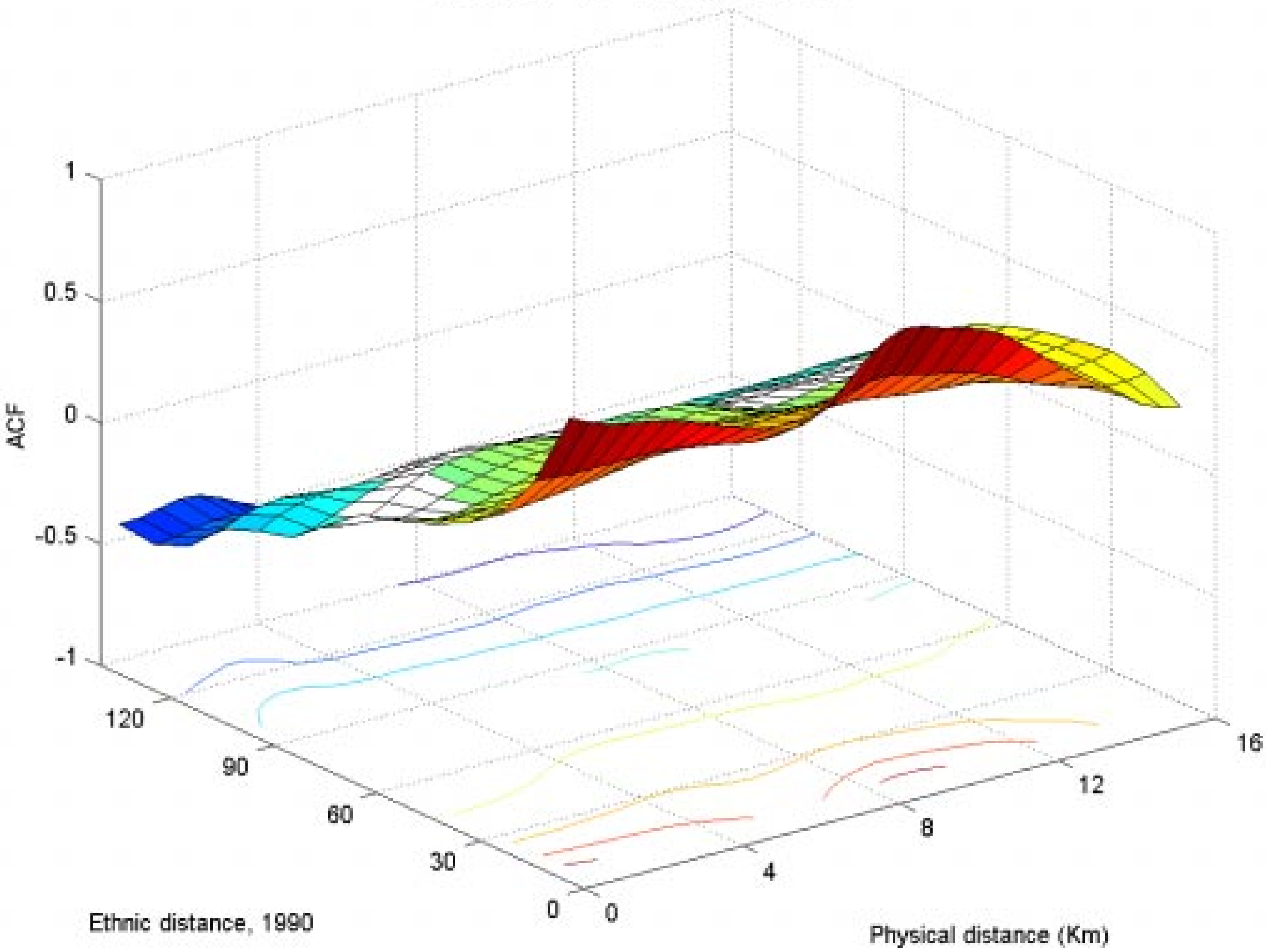


FIGURE 22 - ACF of Residuals of Empl(90) on X(90)

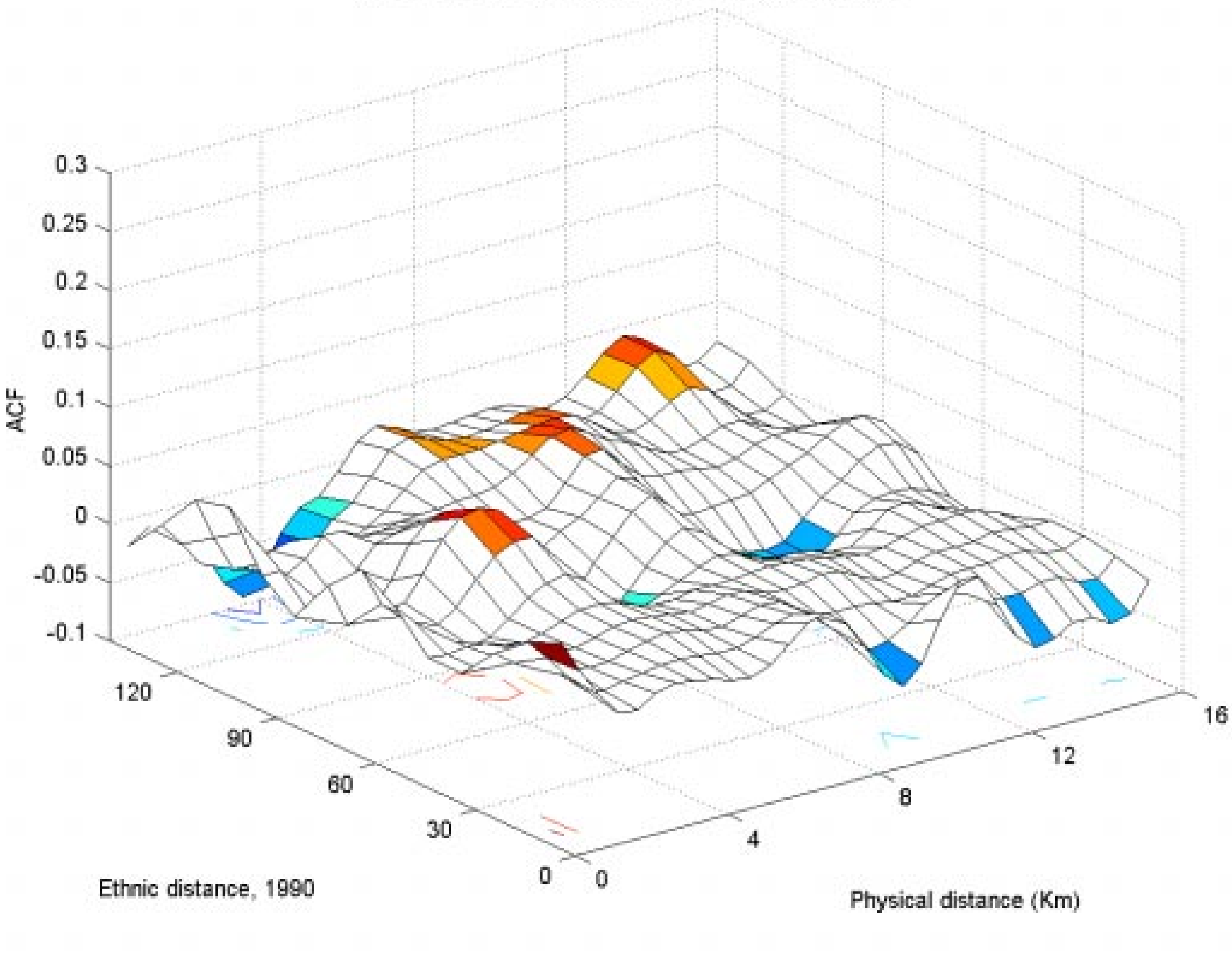


FIGURE 23 - ACF of Employment, 1990-80

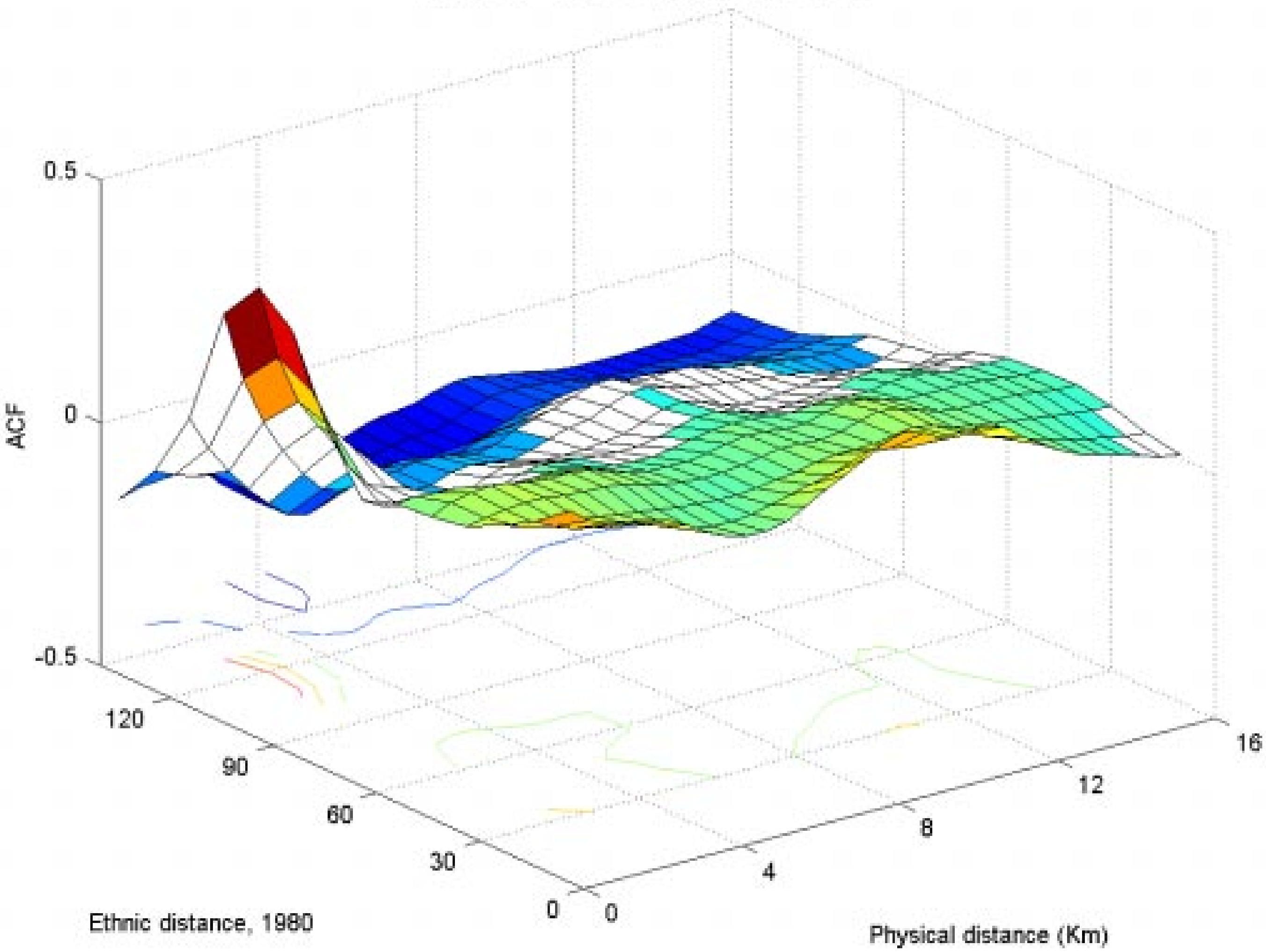


FIGURE 24 - ACF of Residuals of Empl(90-80) on X(90-80)

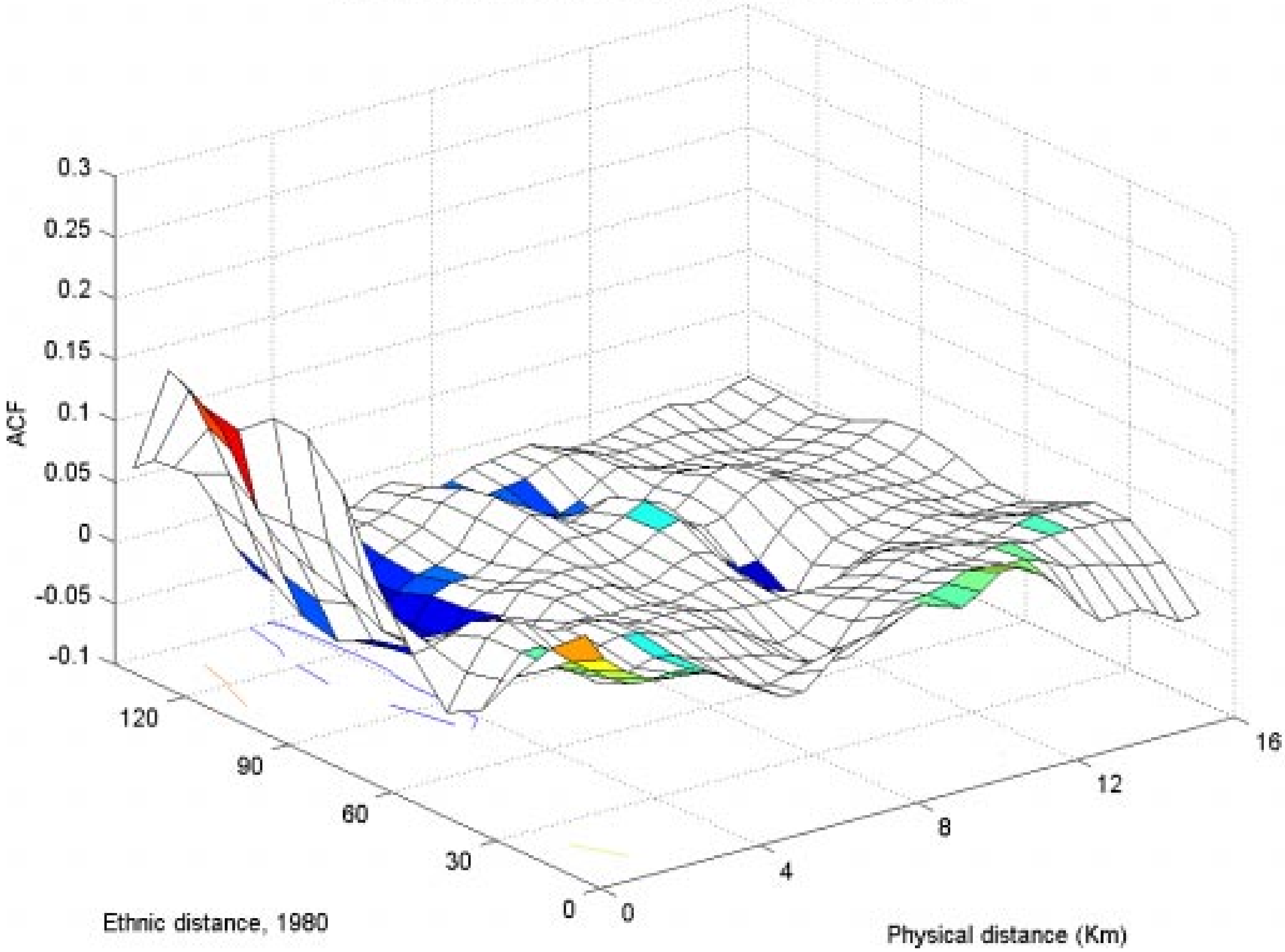


FIGURE 25 - ACF of Employment, 1980

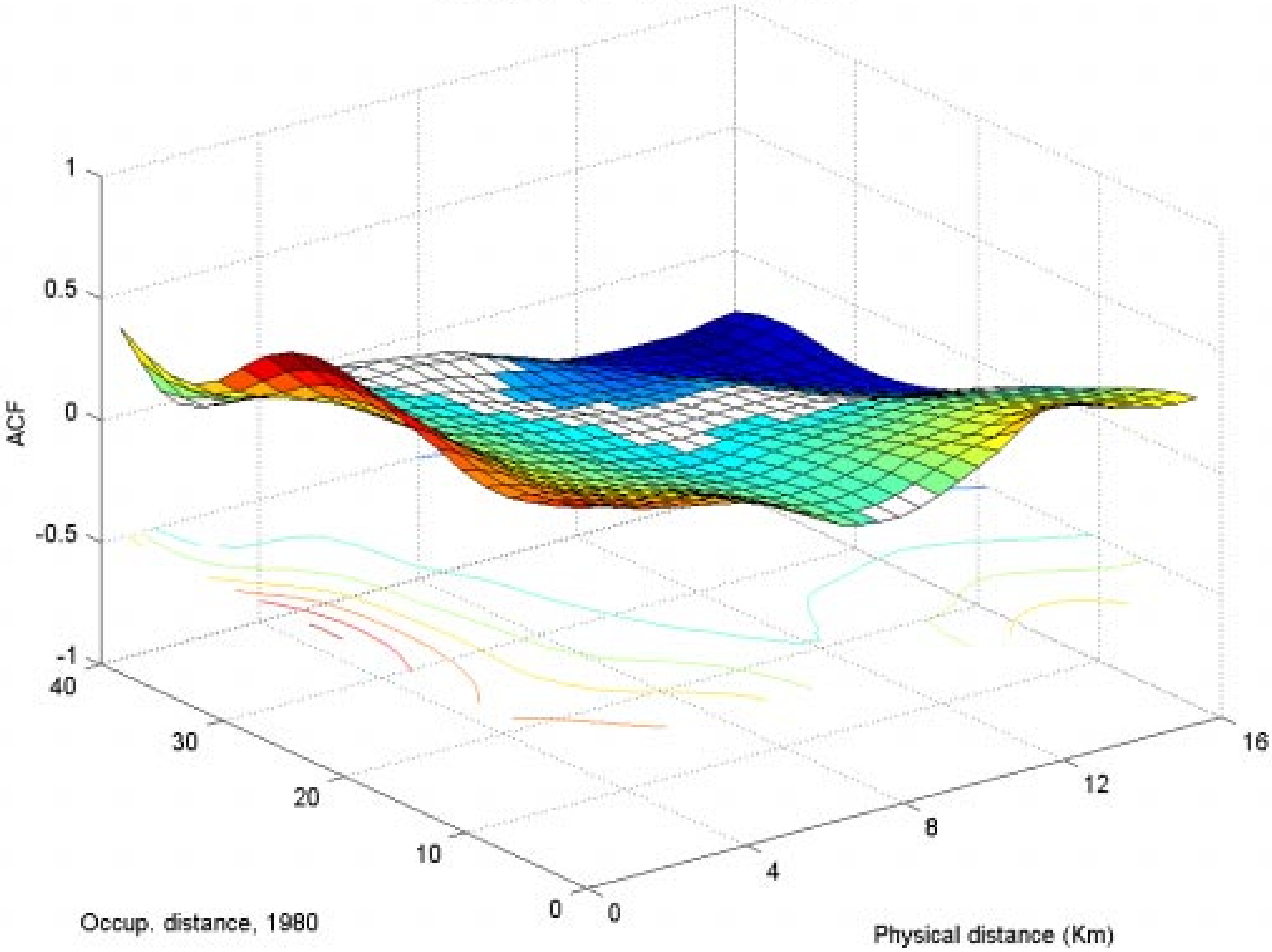


FIGURE 26 - ACF of Residuals of Empl(80) on X(80)

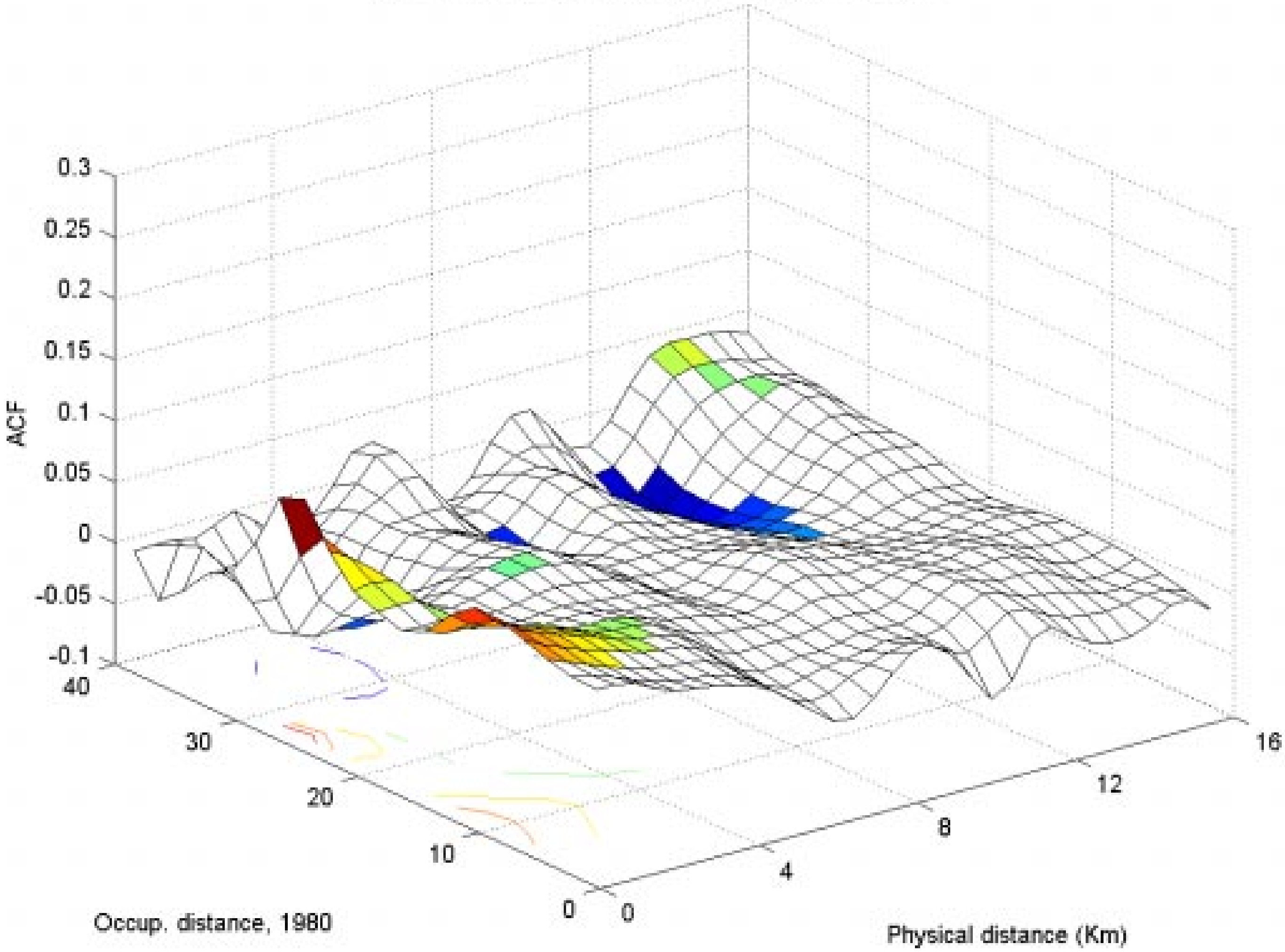


FIGURE 27 - ACF of Employment, 1990

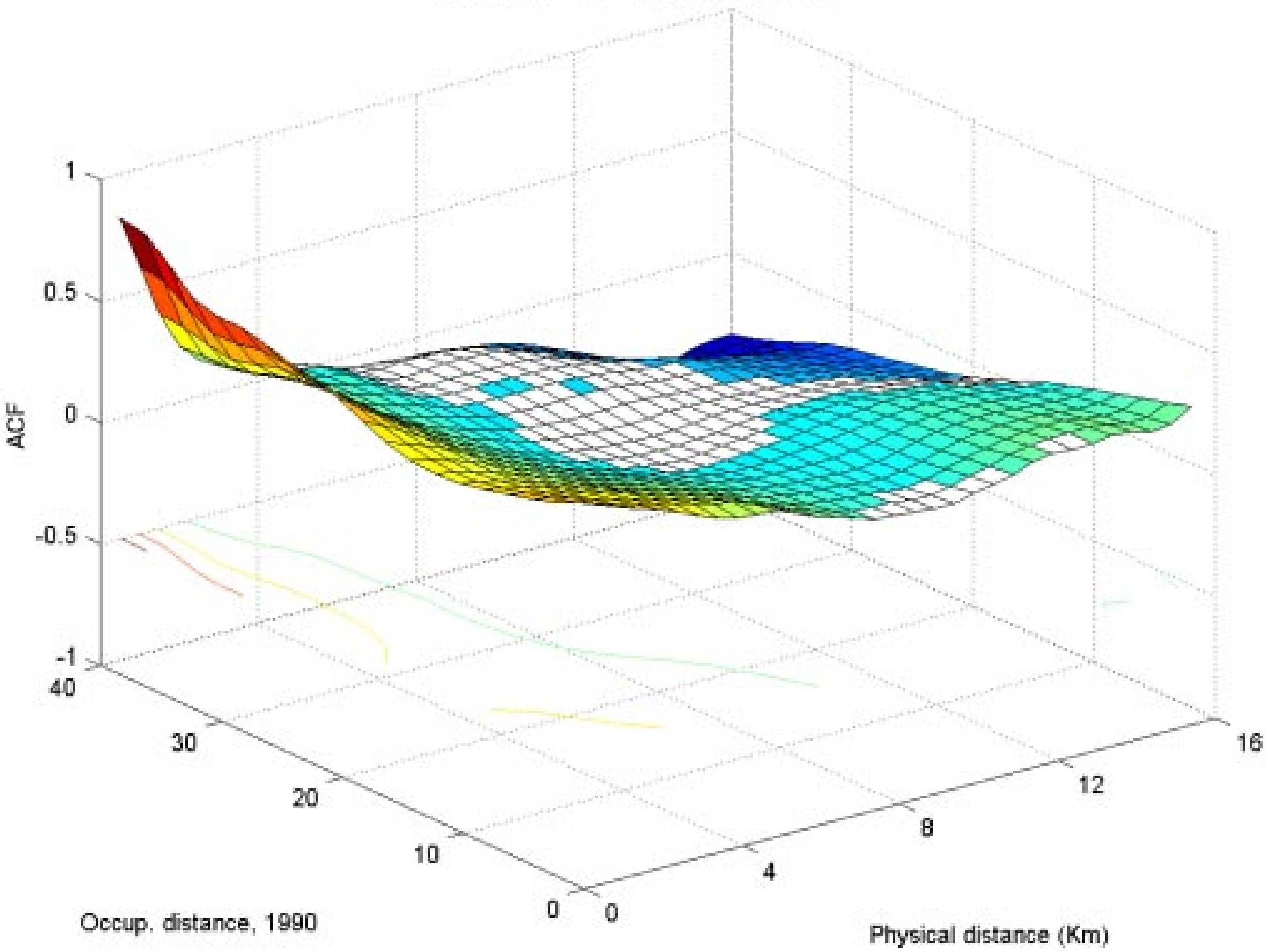


FIGURE 28 - ACF of Residuals of Empl(90) on X(90)

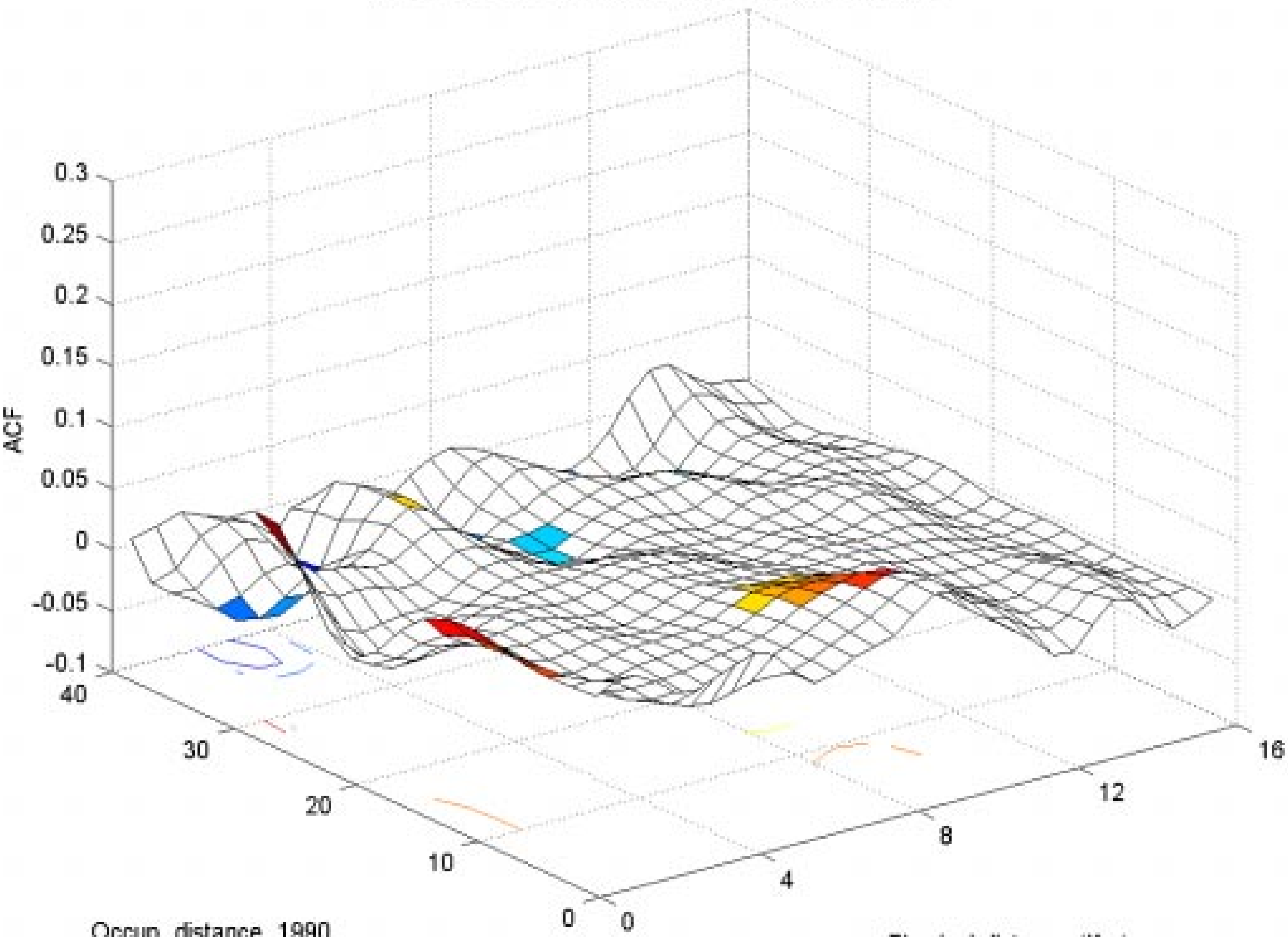


FIGURE 29 - ACF of Employment, 1990-80

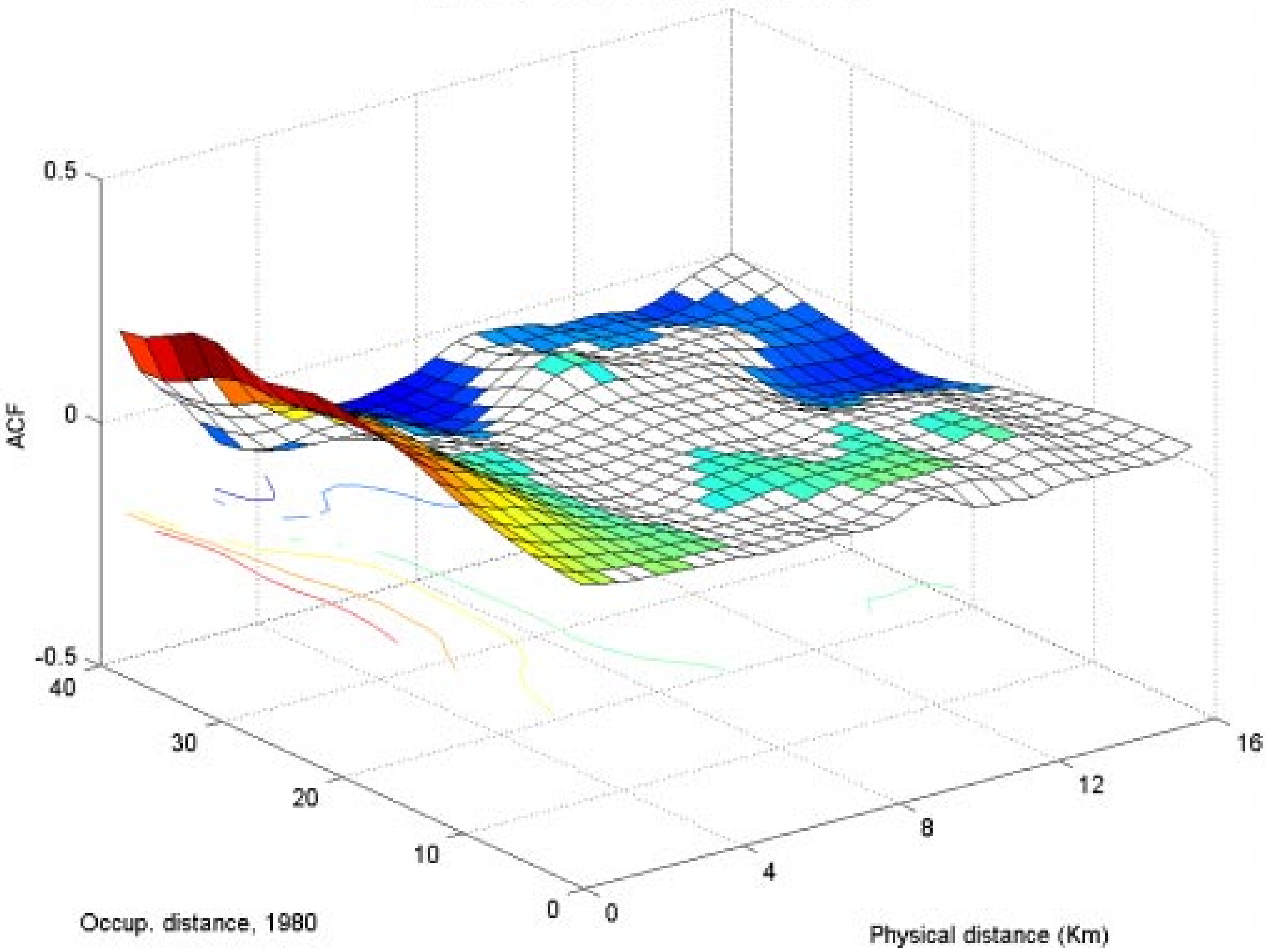


FIGURE 30 - ACF of Residuals of Empl(90-80) on X(90-80)

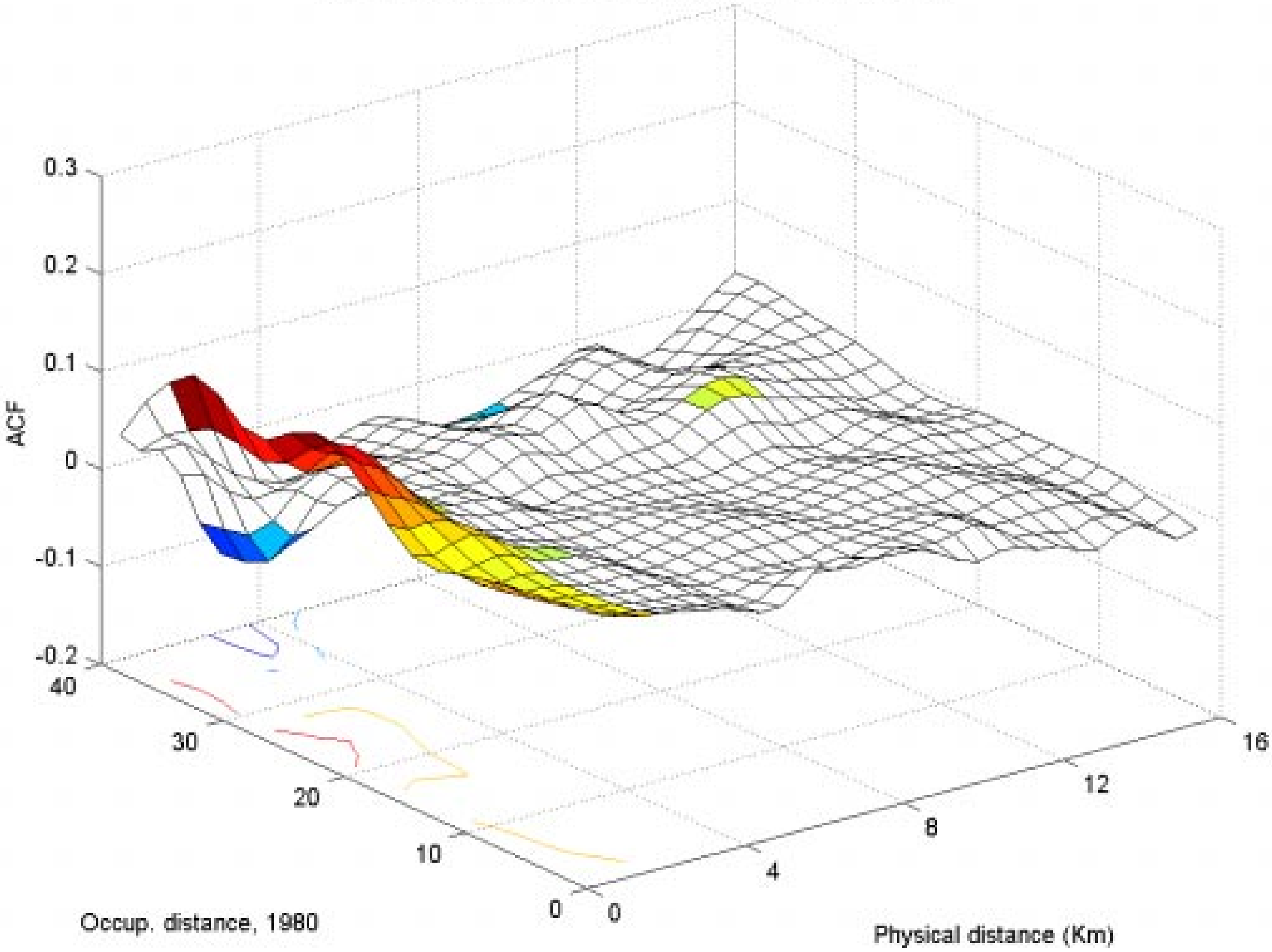


FIGURE 31 - ACF of Employment, 1980

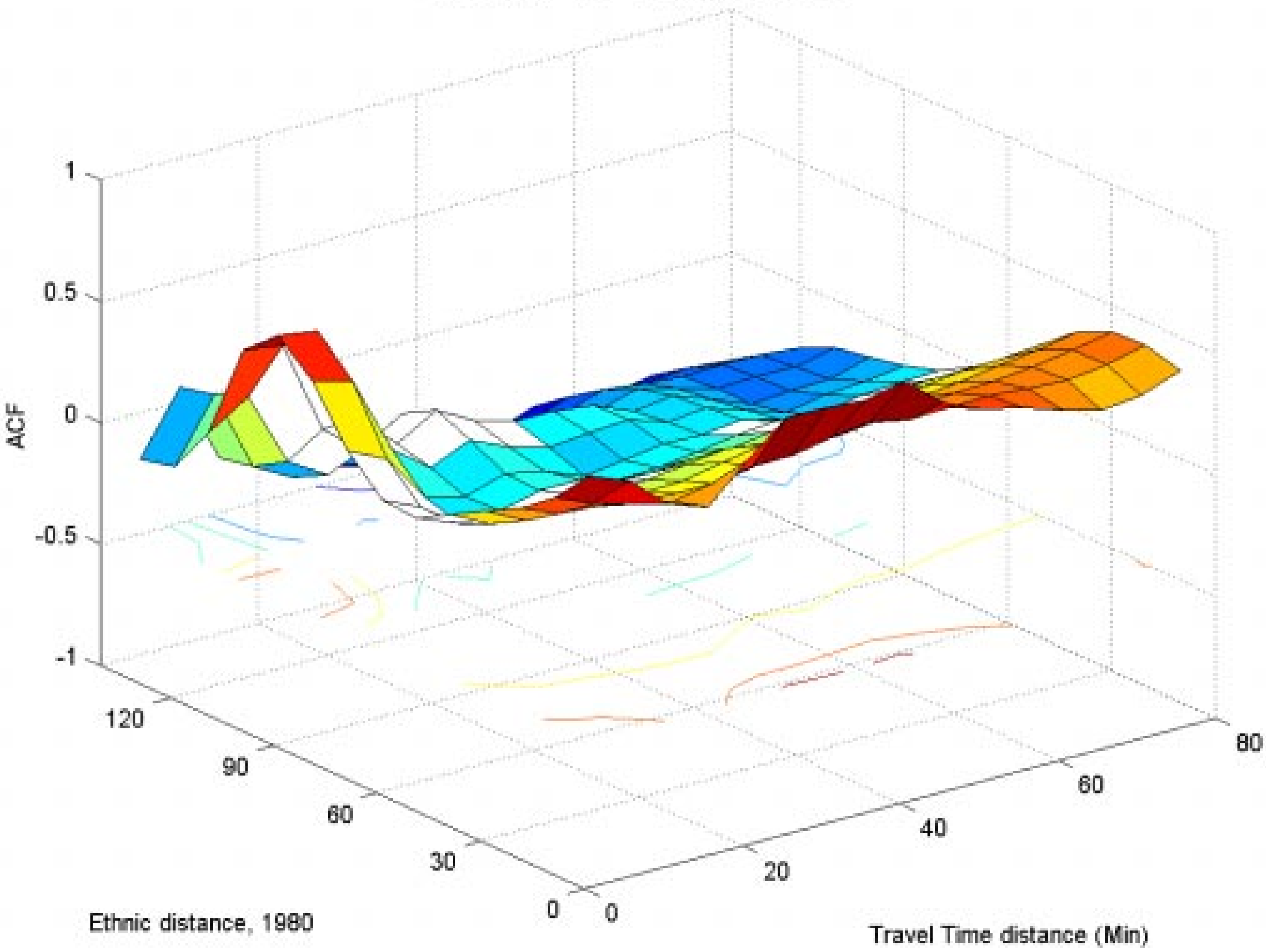


FIGURE 32 - ACF of Residuals of Empl(80) on X(80)

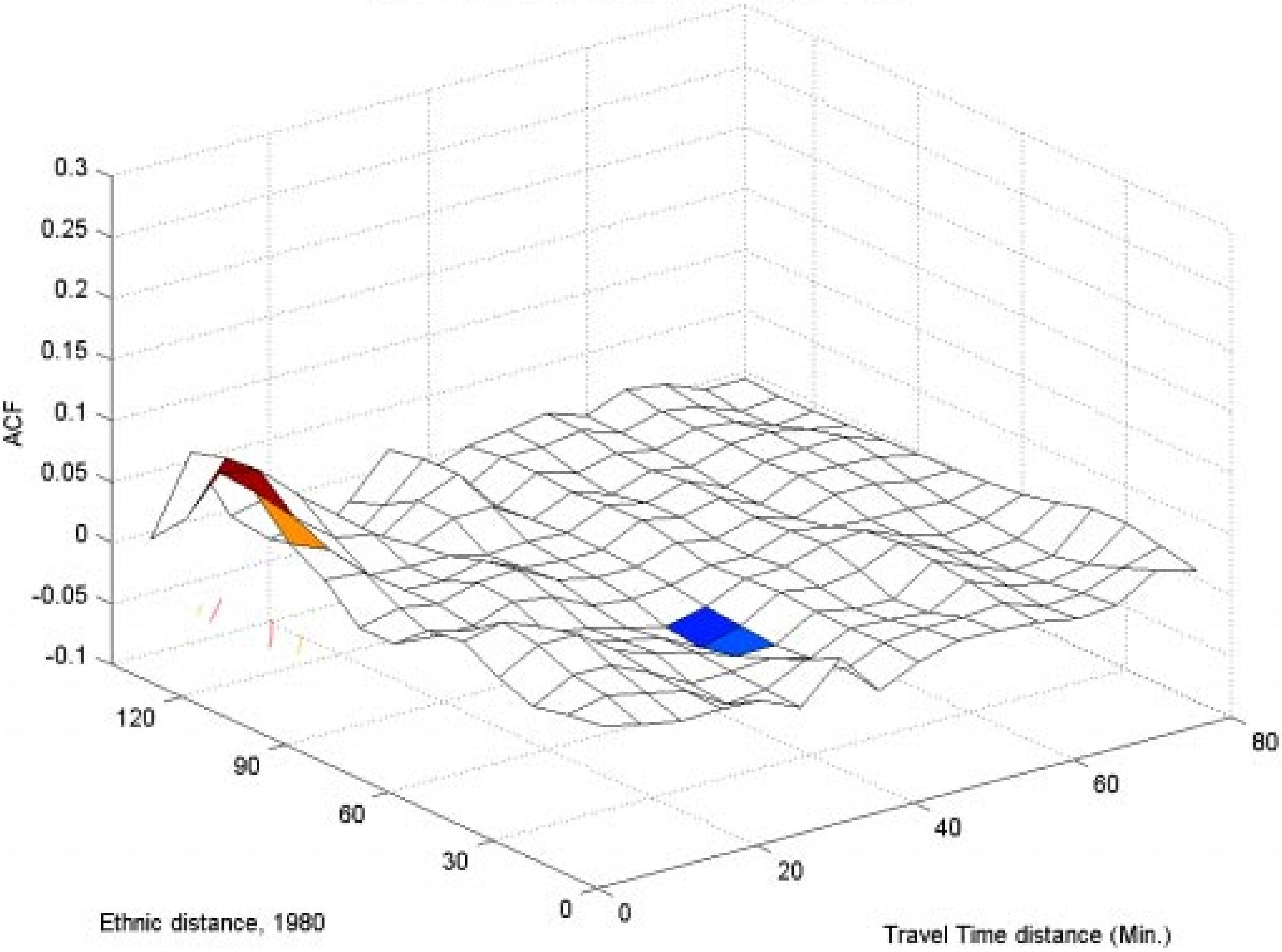


FIGURE 33 - ACF of Employment, 1990

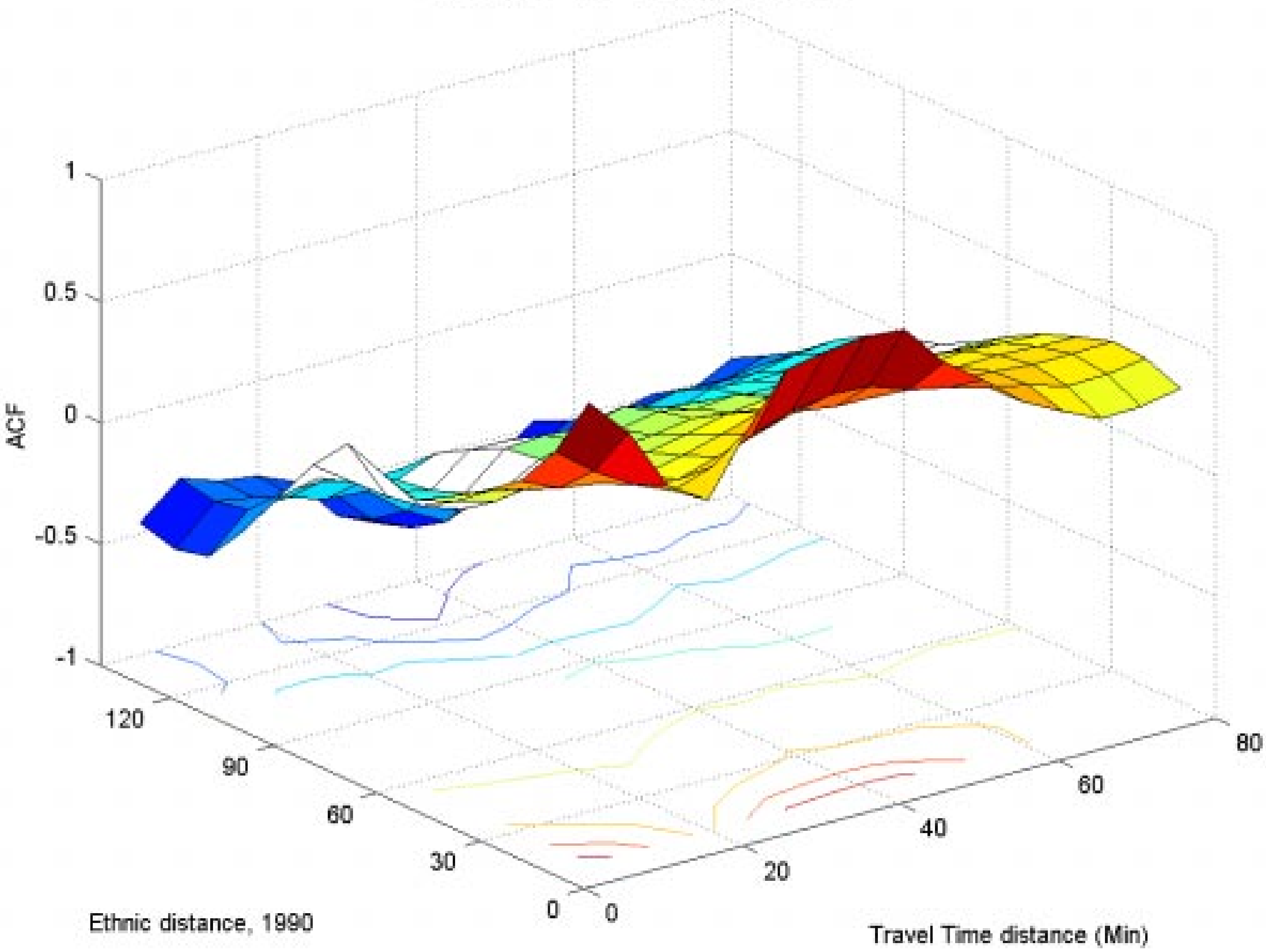


FIGURE 34 - ACF of Residuals of Empl(90) on X(90)

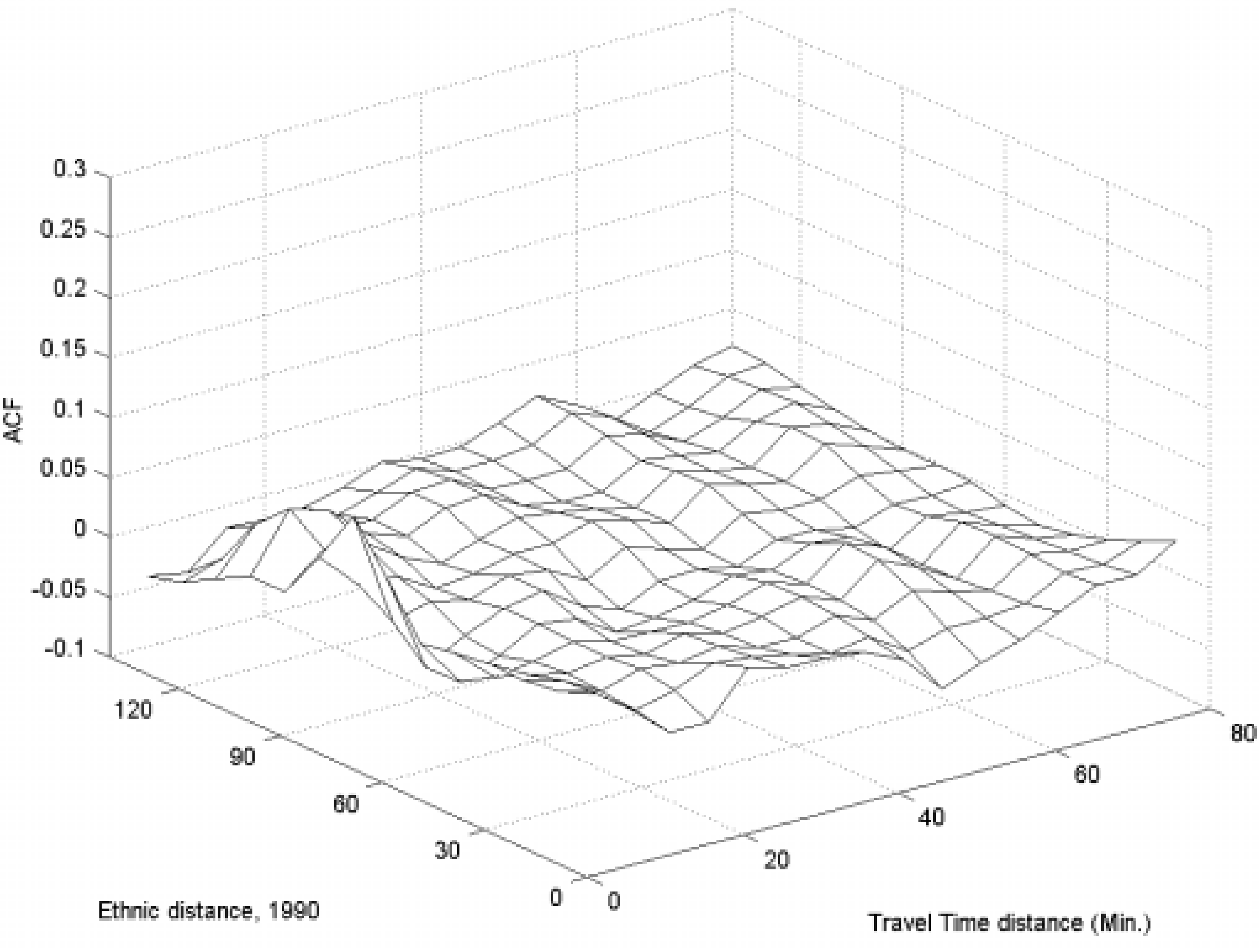
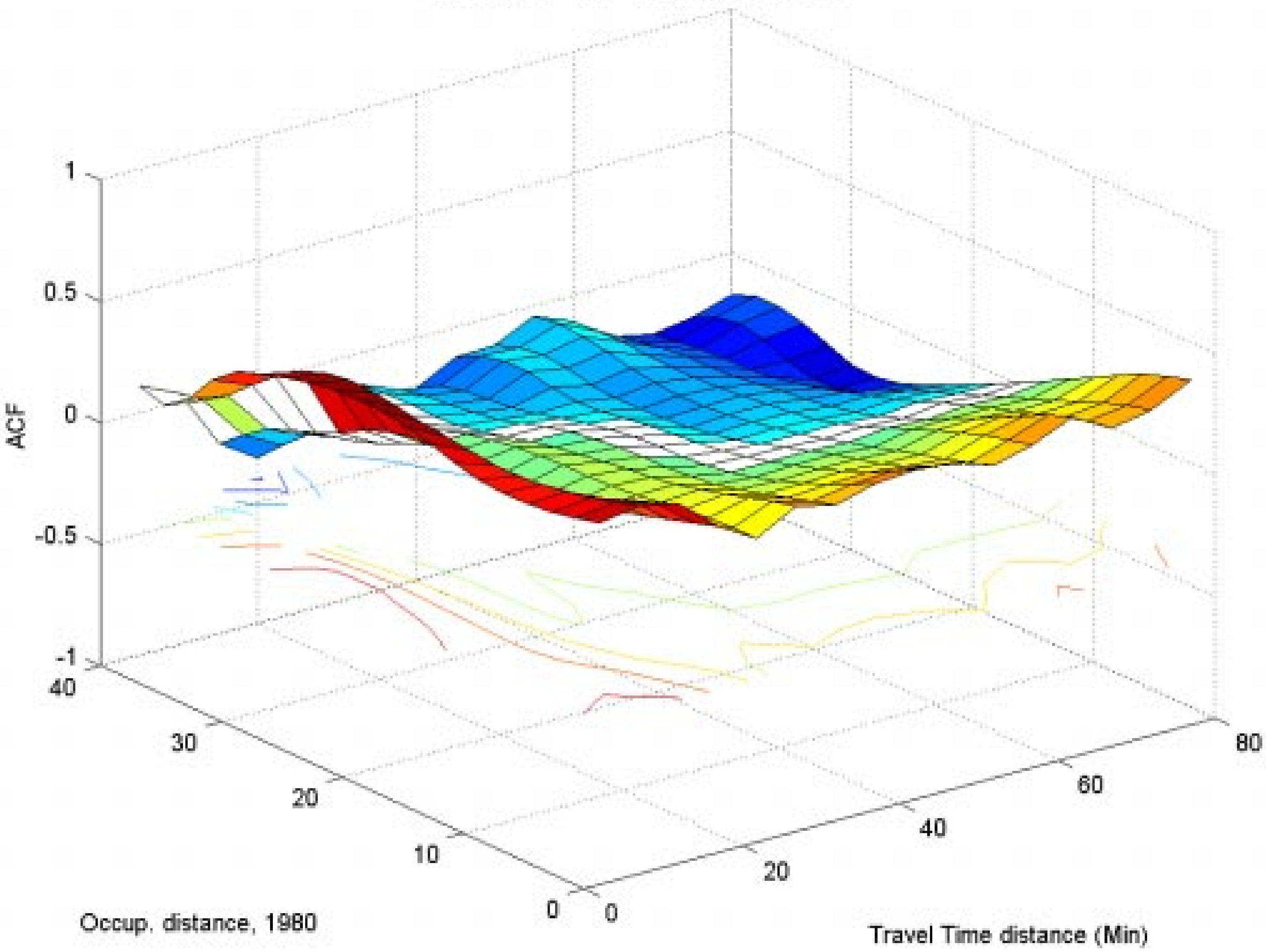


FIGURE 35 - ACF of Employment, 1980



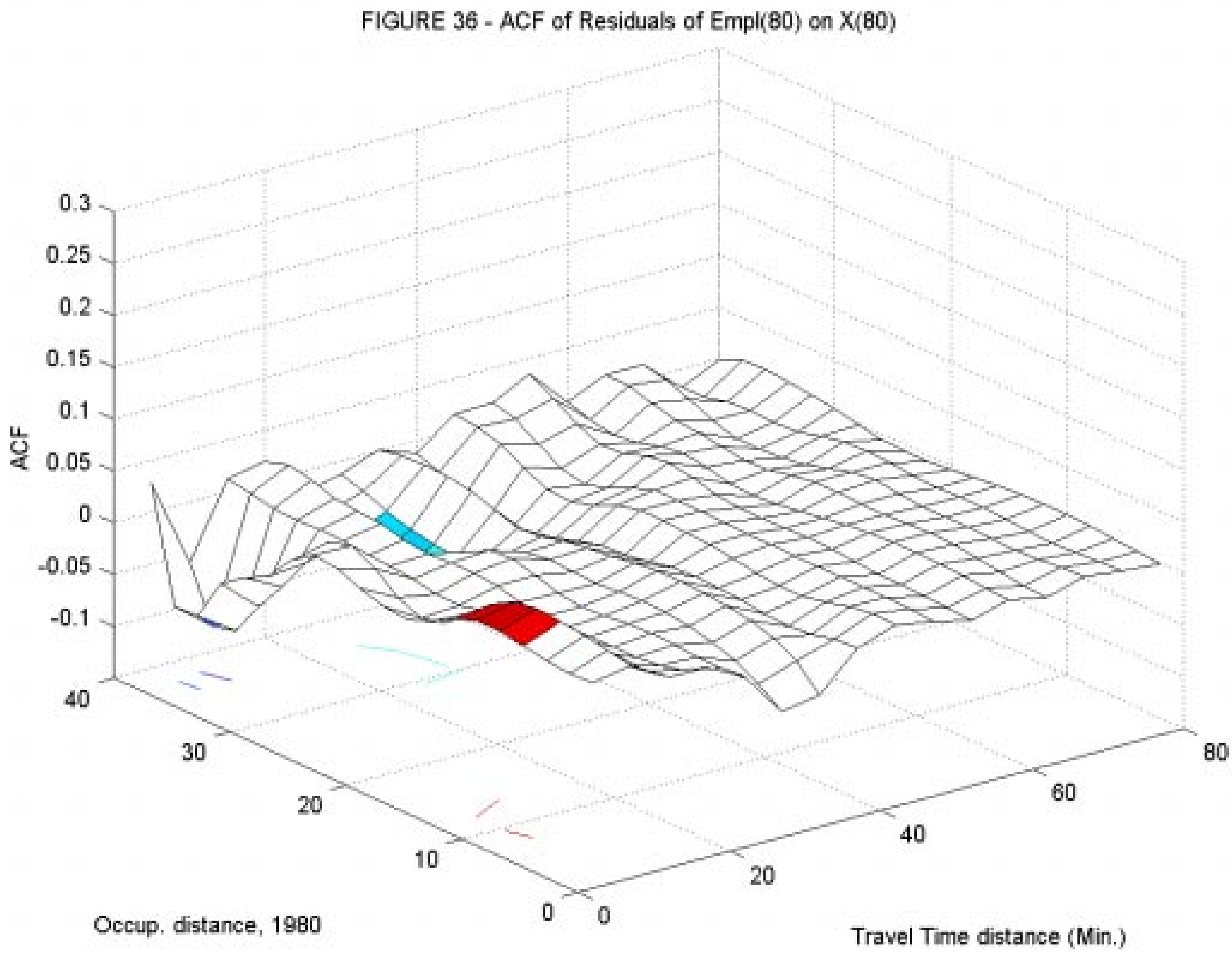


FIGURE 36 - ACF of Residuals of Empl(80) on X(80)
ACF
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
40
30
20
10
0
0
20
40
60
80
Occup. distance, 1980
Travel Time distance (Min.)

FIGURE 37 - ACF of Employment, 1990

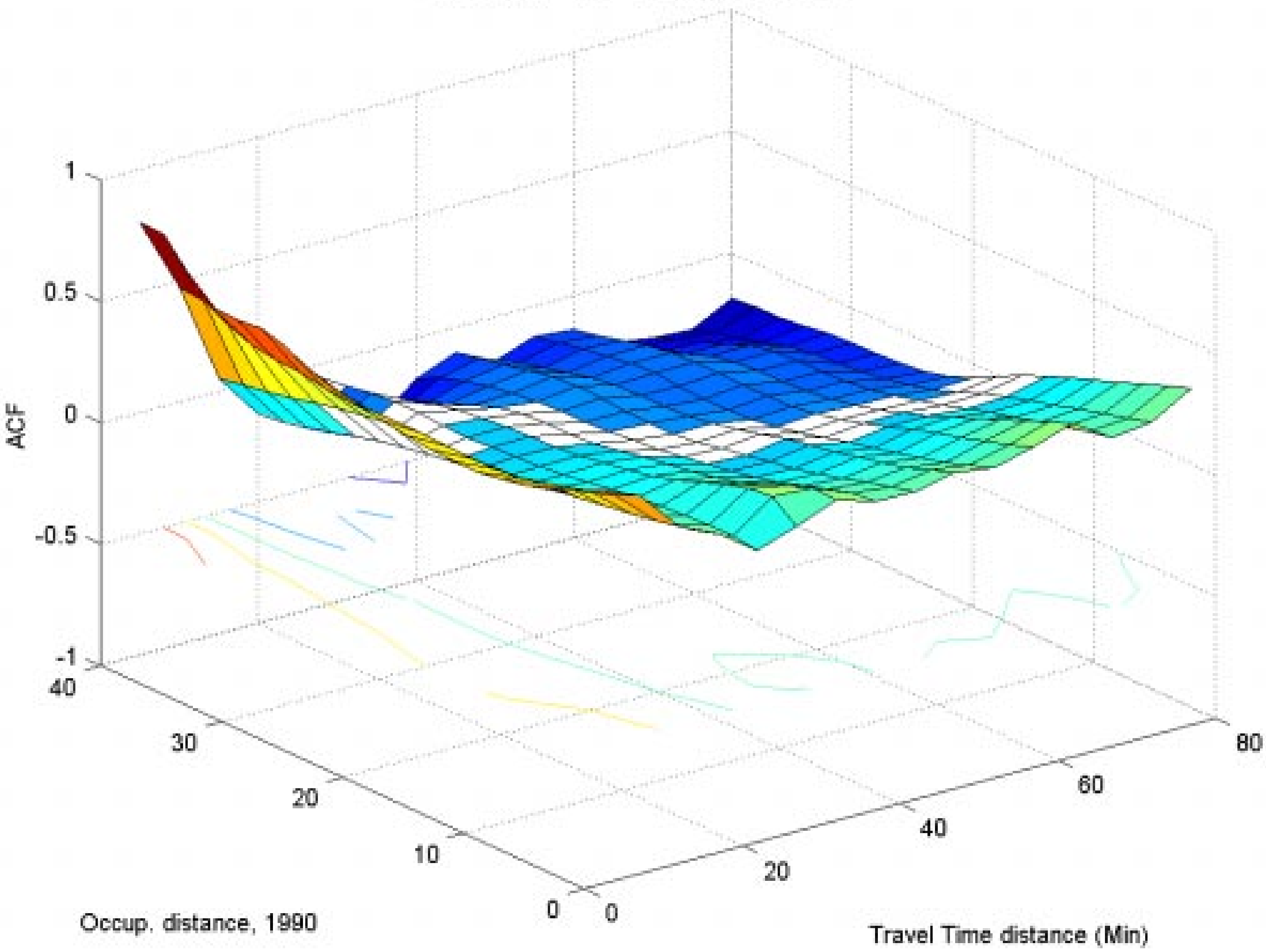
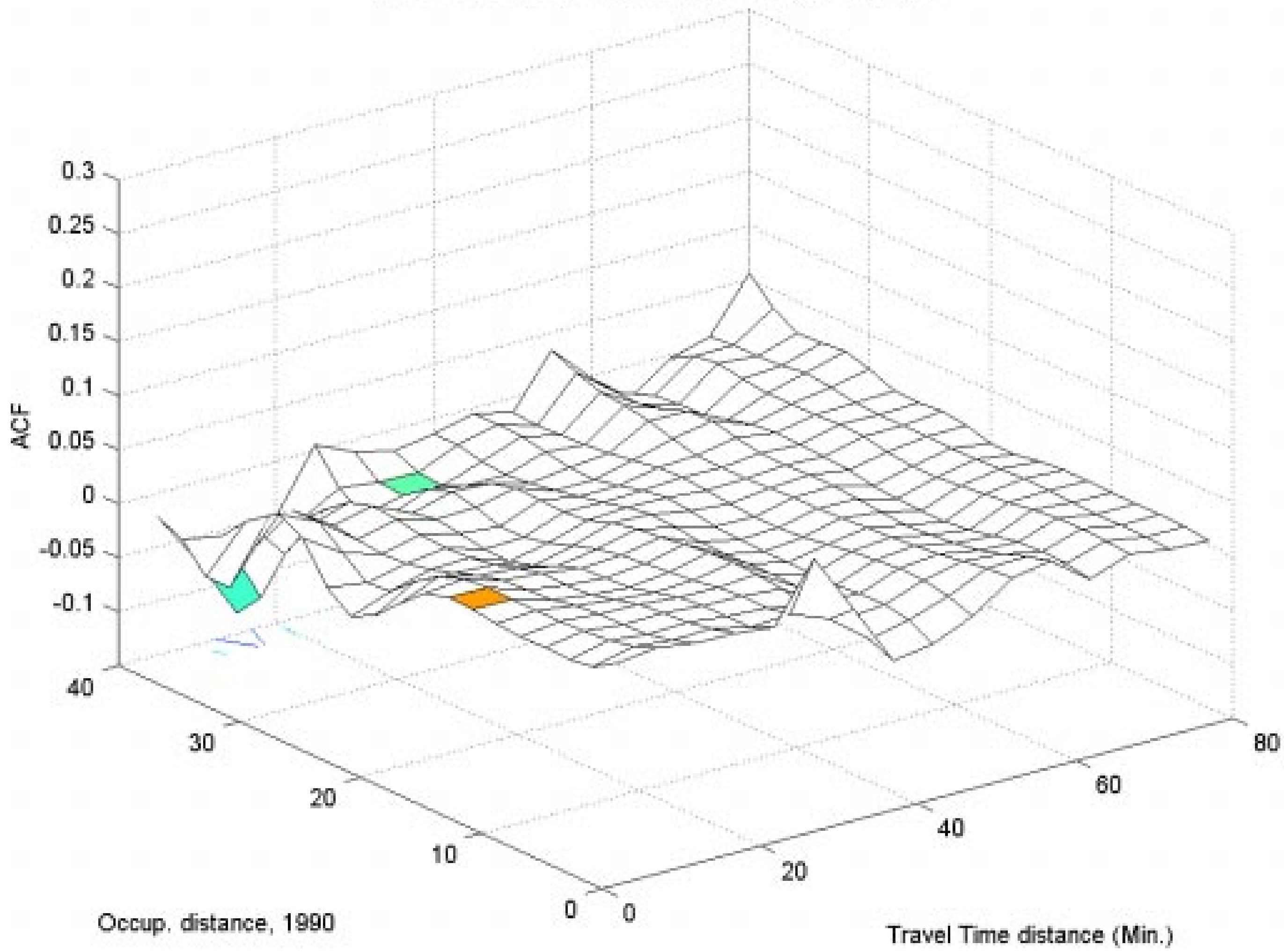


FIGURE 38 - ACF of Residuals of Empl(90) on X(90)



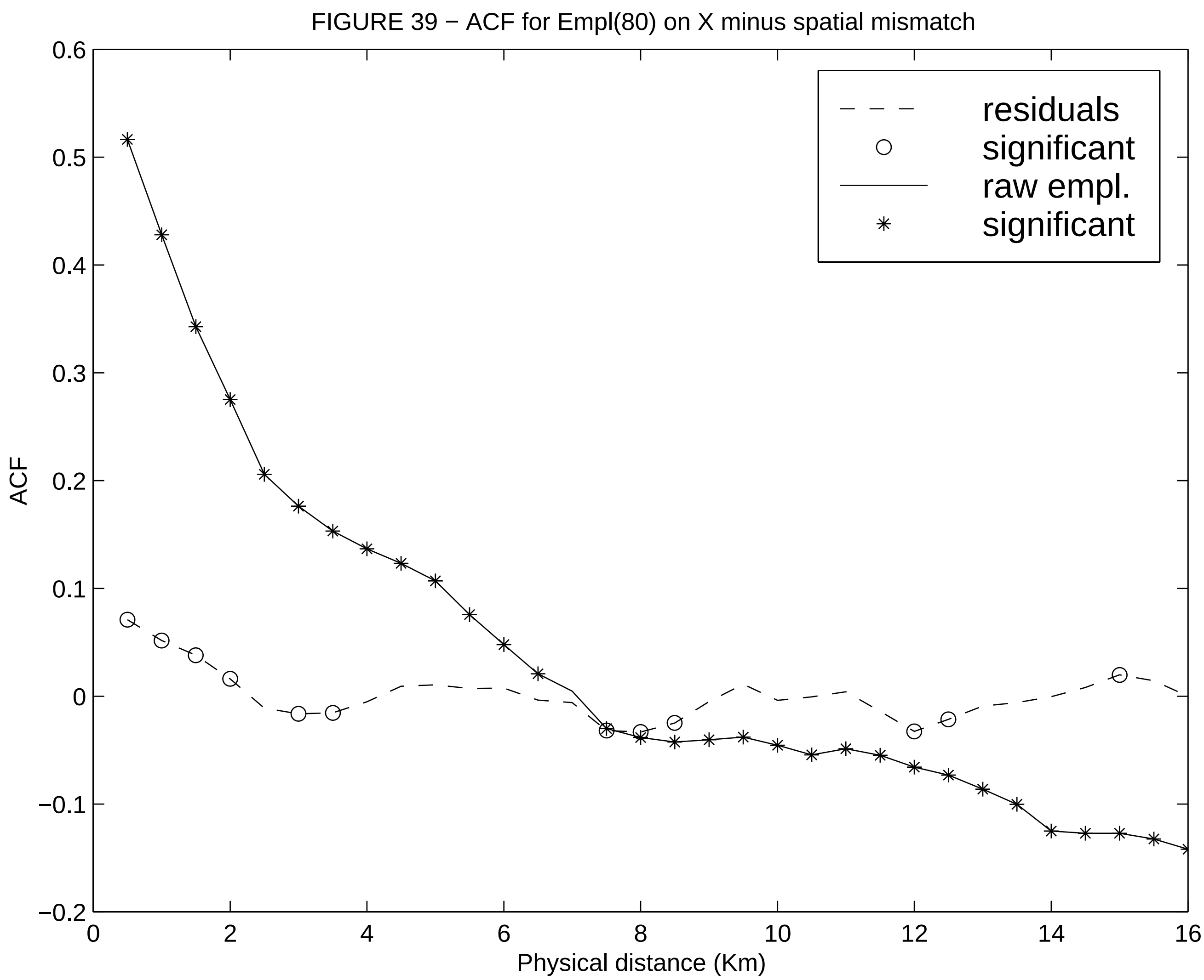


FIGURE 39 − ACF for Empl(80) on X minus spatial mismatch
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
ACF
0
2
4
6
8
10
12
14
16
Physical distance (Km)
residuals
significant
raw empl.
significant

FIGURE 40 − ACF for Empl(80) on X minus spatial mismatch

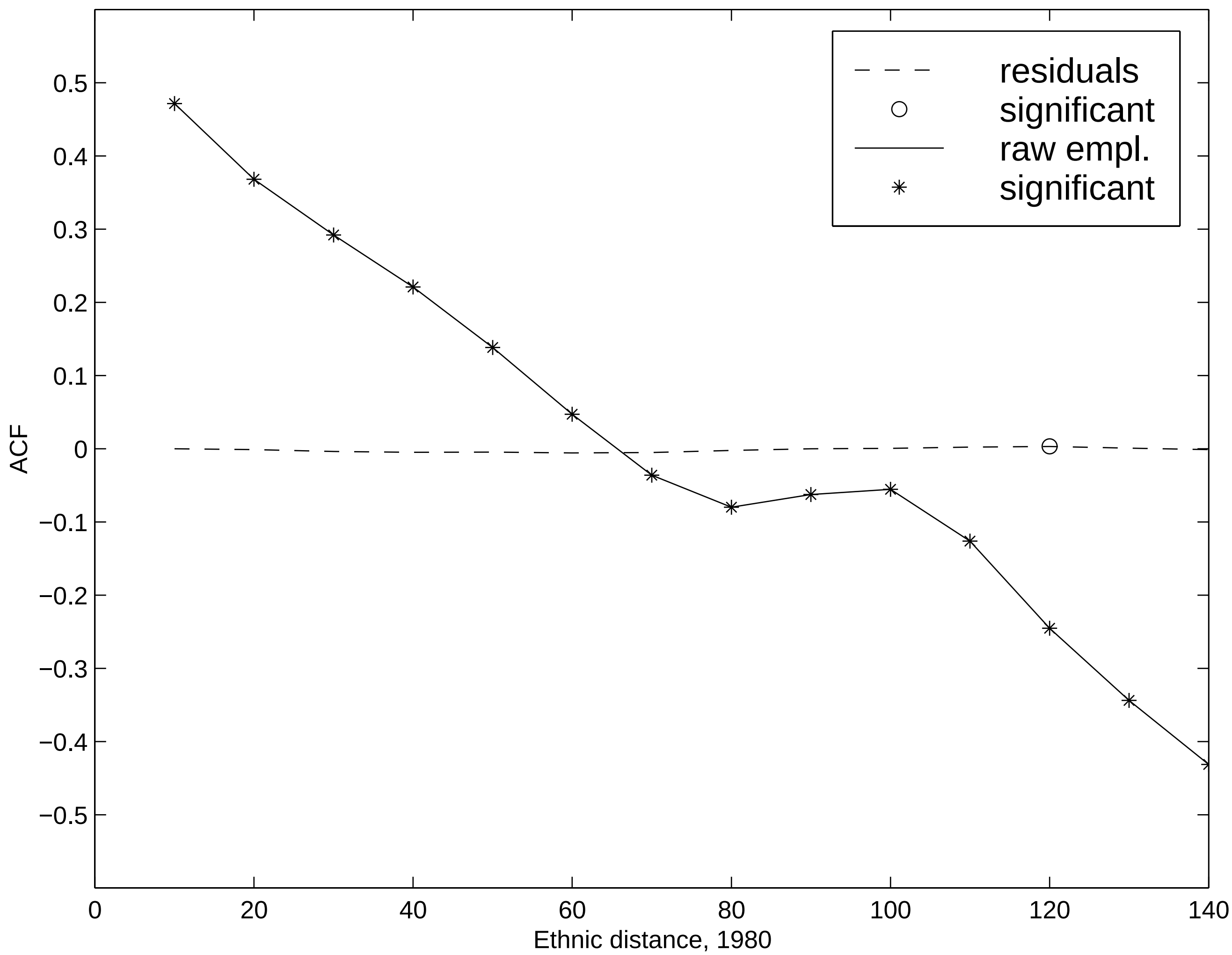
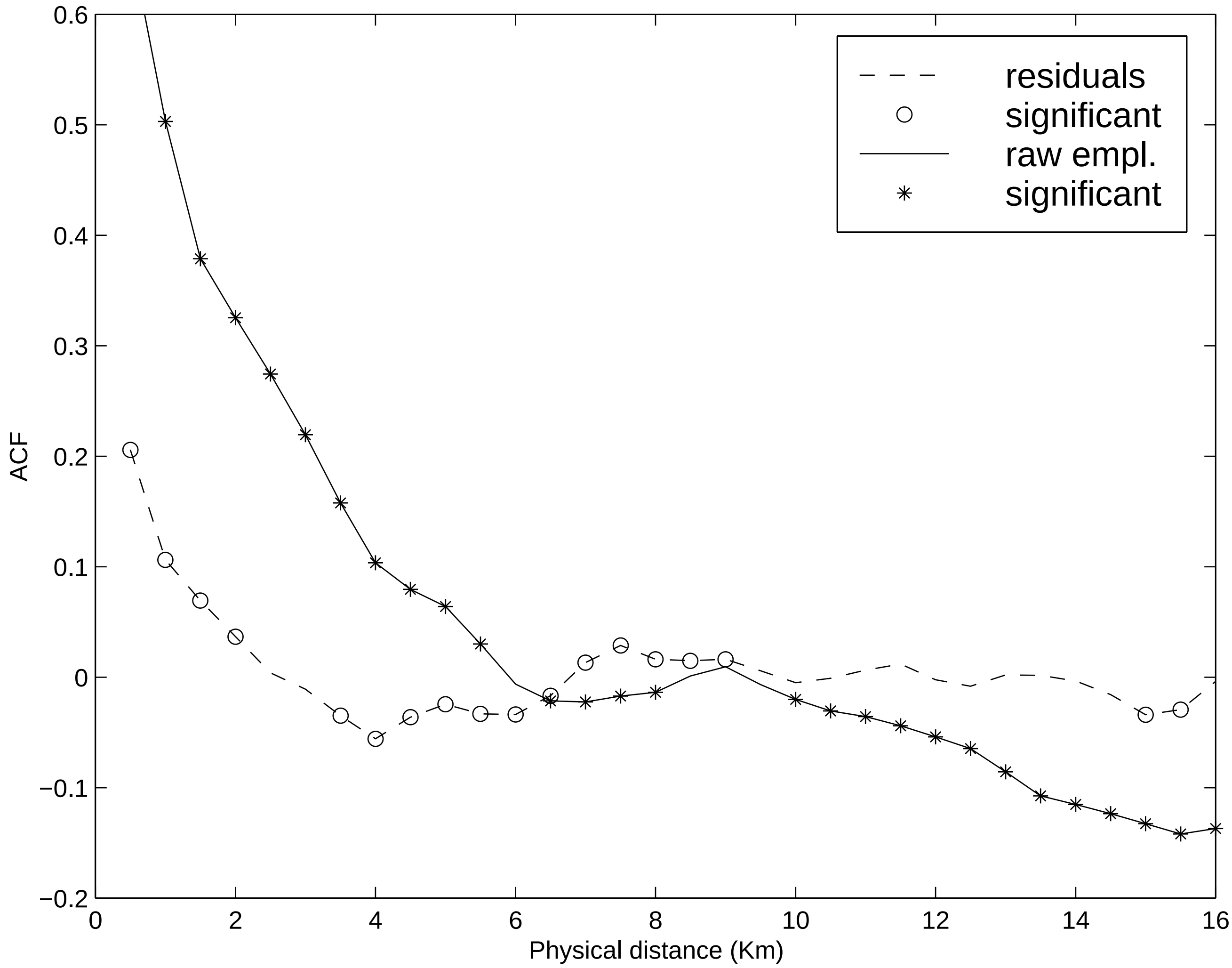


FIGURE 41 − ACF for Empl(90) on X minus racial/ethnic variables



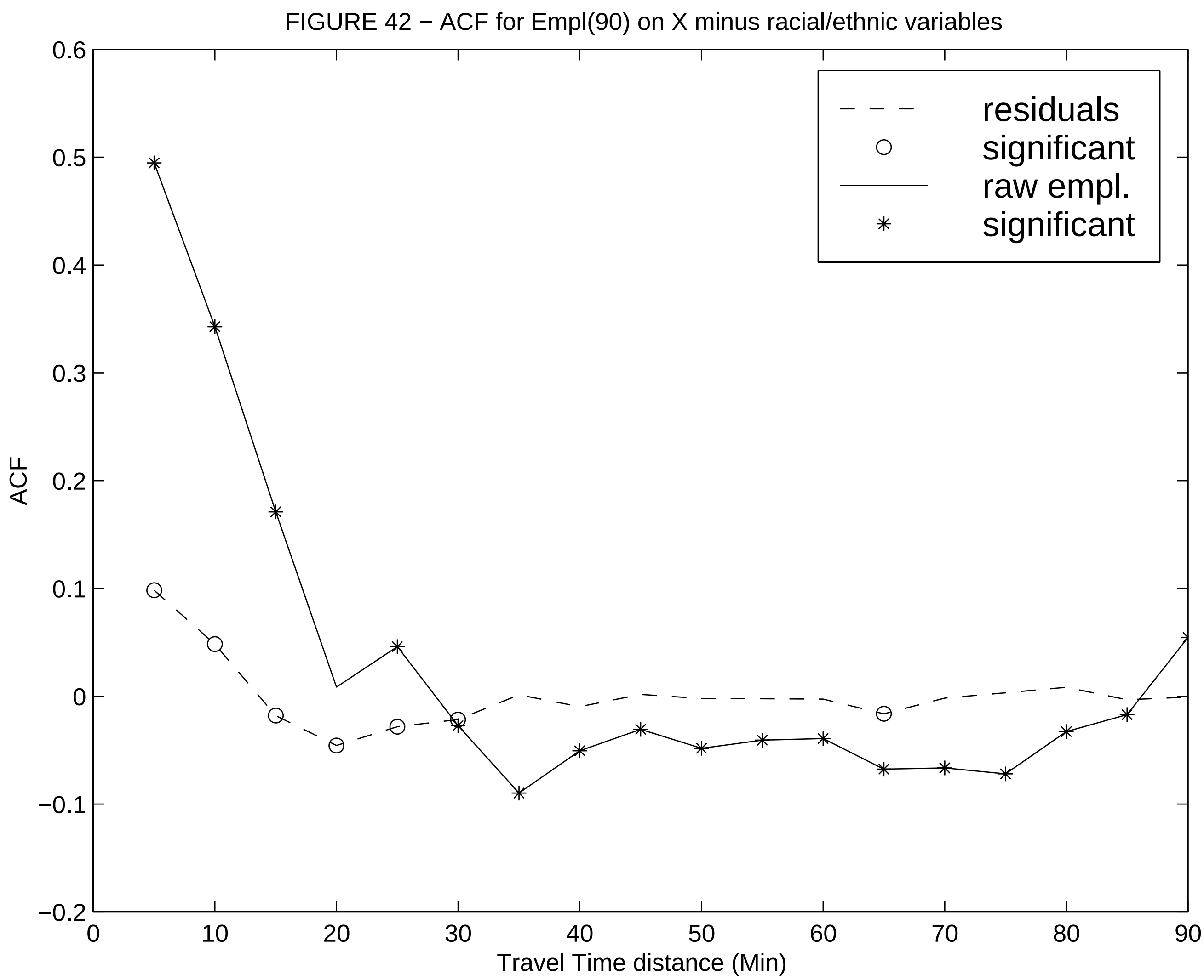


FIGURE 42 − ACF for Empl(90) on X minus racial/ethnic variables
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
ACF
0
10
20
30
40
50
60
70
80
90
Travel Time distance (Min)
residuals
significant
raw empl.
significant

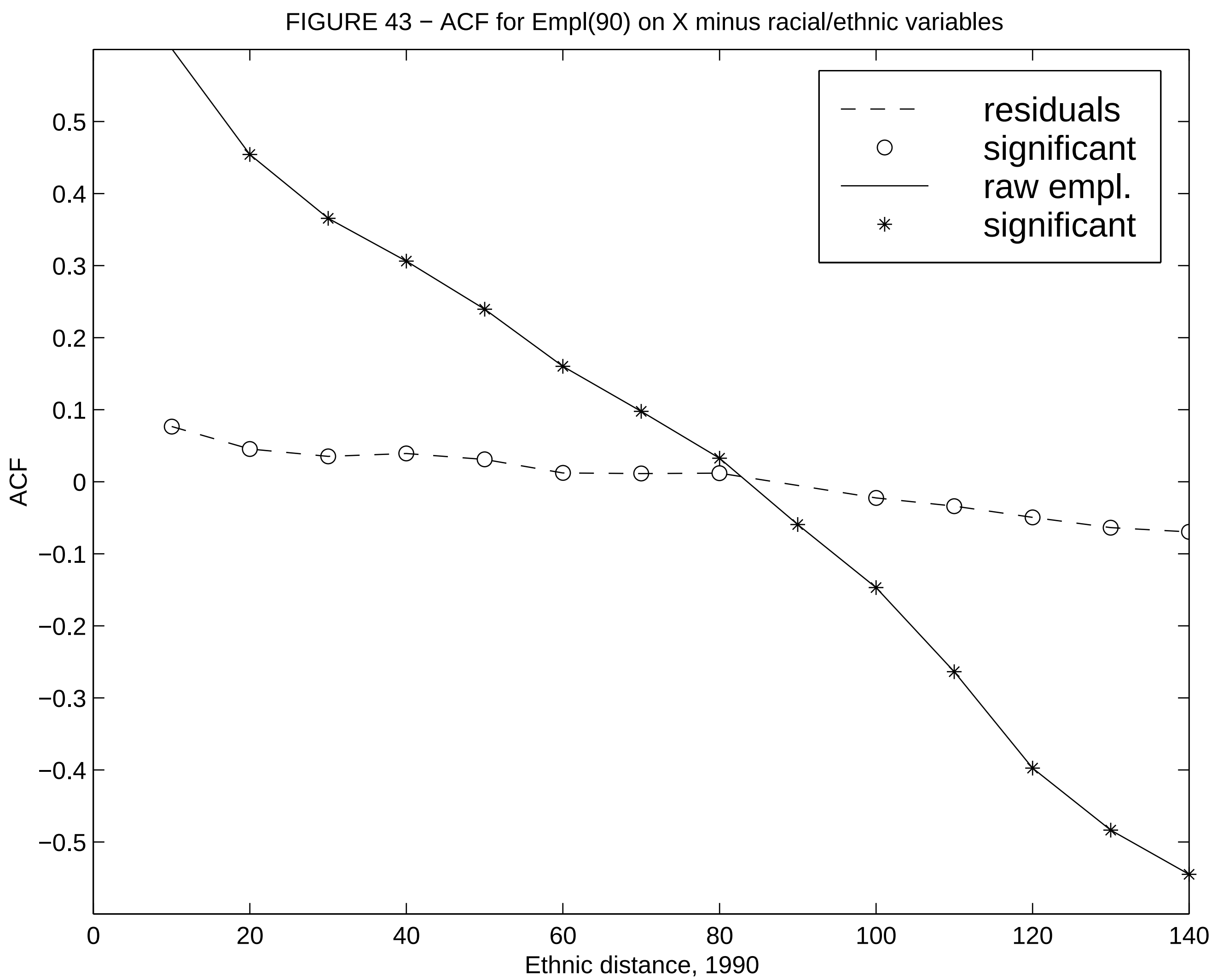


FIGURE 43 − ACF for Empl(90) on X minus racial/ethnic variables
residuals
significant
raw empl.
significant
ACF
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
−0.5
0
20
40
60
80
100
120
140
Ethnic distance, 1990

FIGURE 44 − ACF for Empl(90) on X minus racial/ethnic variables

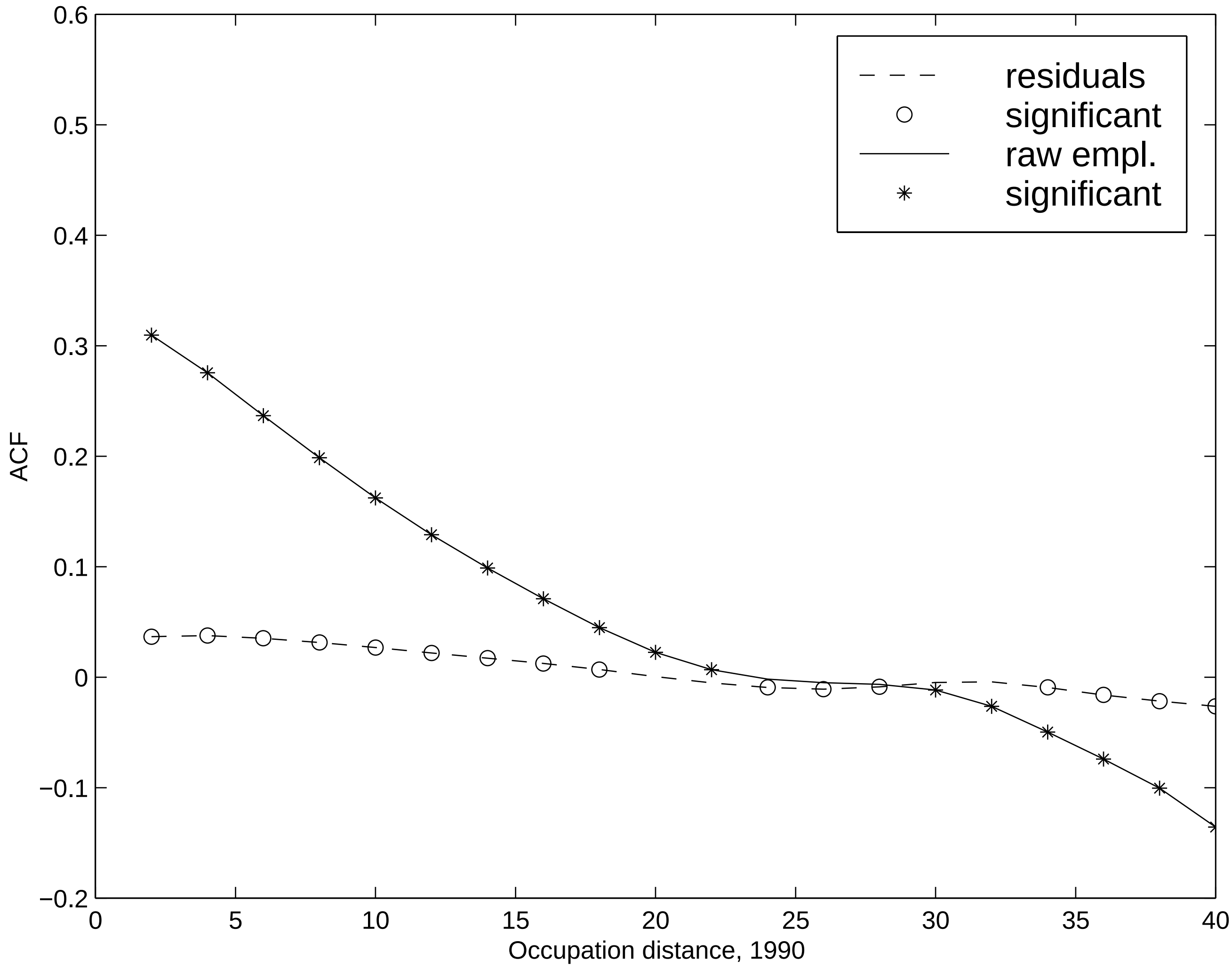


FIGURE 45 − ACF for Empl(90) on X minus education variables

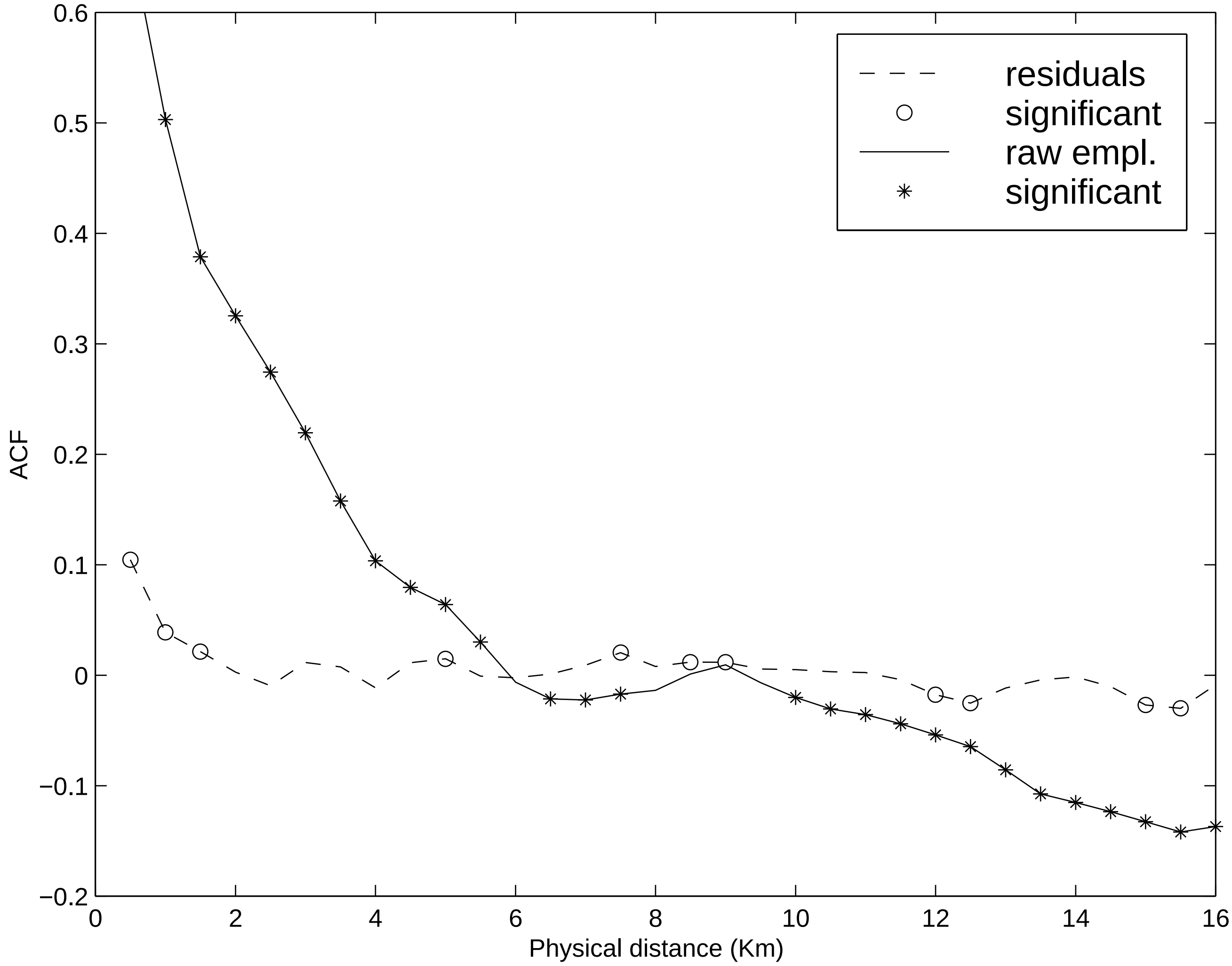


FIGURE 46 − ACF for Empl(90) on X minus education variables

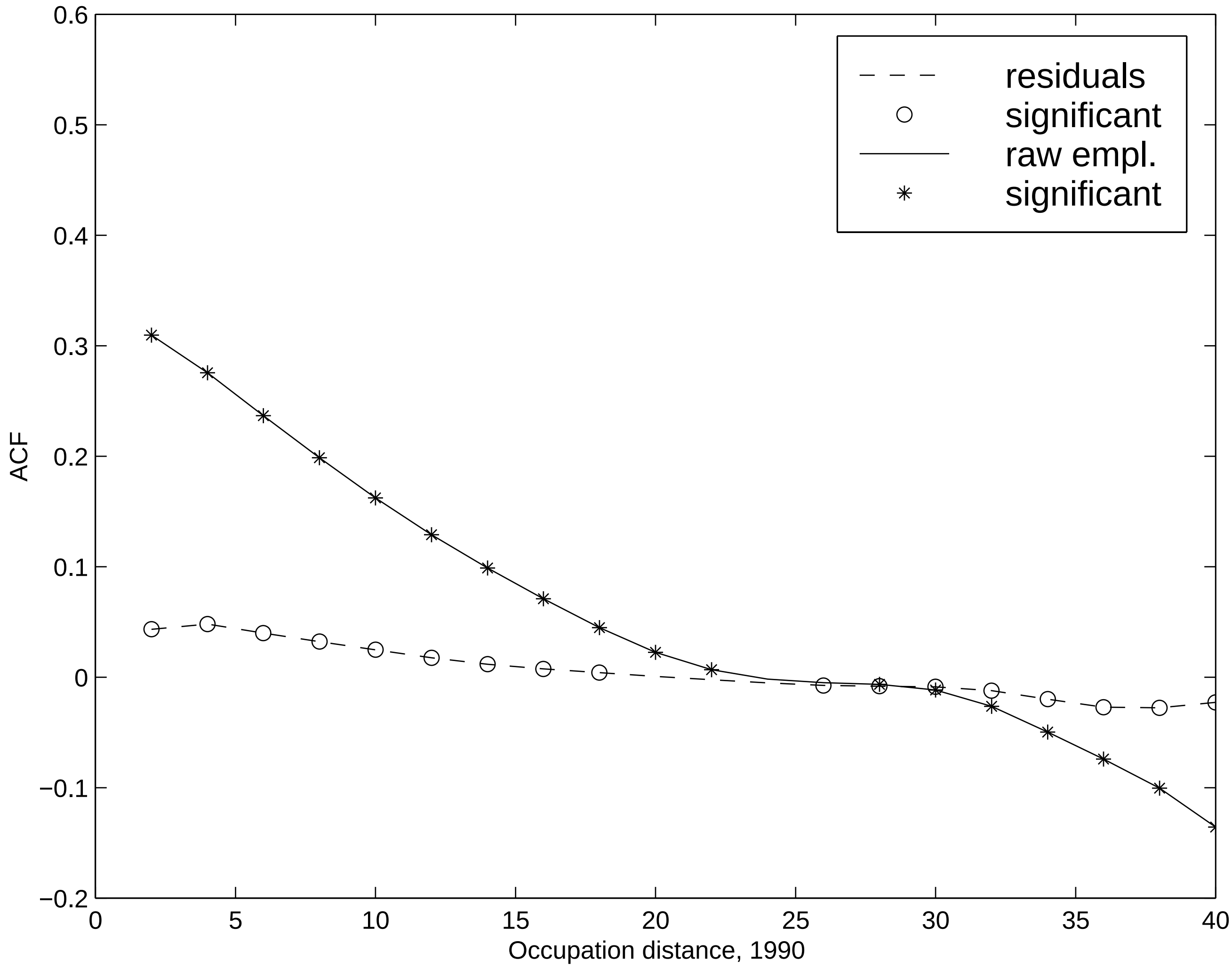


FIGURE 47 − ACF for fraction of non−white persons in 1980

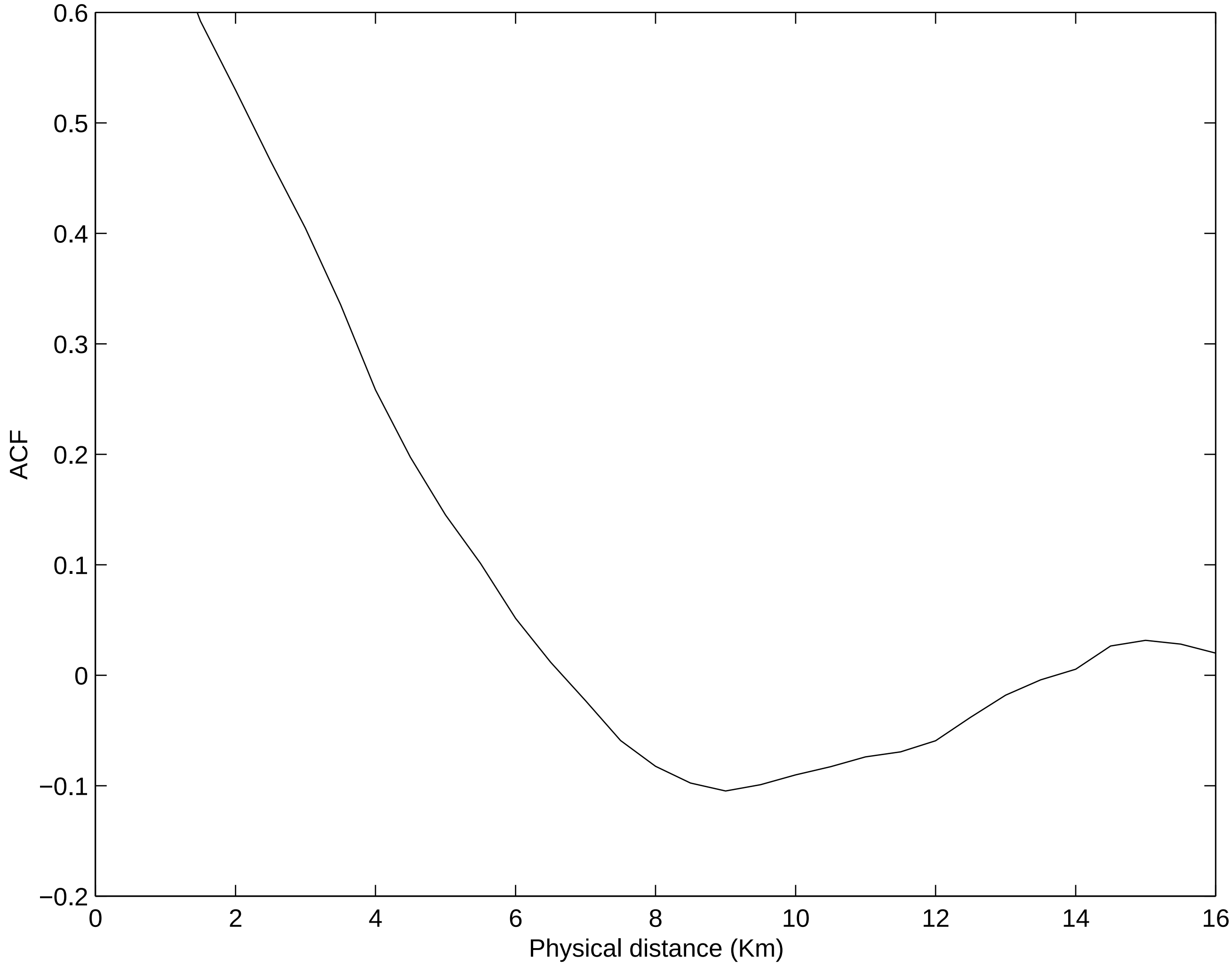


FIGURE 48 − ACF for fraction of Hispanic persons in 1980

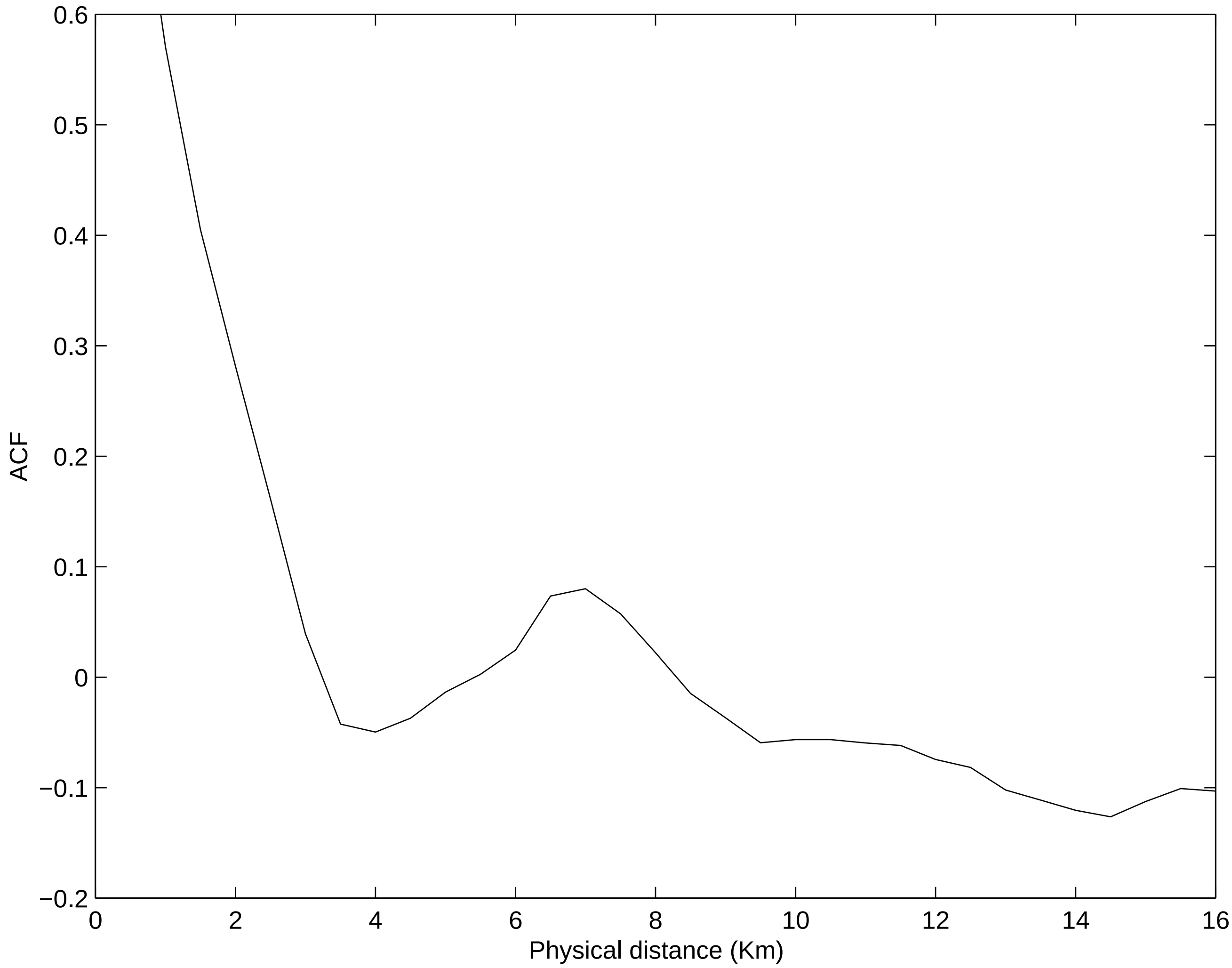


TABLE 1 - Dependent variable: unemployment rate among civilians 16 years and older.

	1980	1990	1990-80
Baseline adjusted R^2	0.6096	0.7322	0.3150
	change in adjusted R^2	**change in adjusted R^2**	**change in adjusted R^2**
segregation index	0.0046	0.0042	0.0050
% non whites	0.0470	0.0064	0.0053
% Hispanics	0.0081	0.0580	0.0487
% females	0.0039	0.0001	0.0103
% 0-18 year old	0.0019	0.0002	0.0155
% 0-24 year old	-0.0005	0.0007	0.0293
% 18-24 year old	0.0000	-0.0002	0.0351
% high school graduates	0.0300	0.0198	0.0243
% college graduates	0.0028	0.0026	0.0076
% managerial/profess. workers	0.0000	-0.0003	0.0060
% males out of labor force	-0.0005	0.0002	0.0170
% females out of labor force	0.0000	-0.0003	0.0217
persons per household	0.0078	0.0009	-0.0005
% vacant housing units	-0.0004	0.0219	0.0284
median gross rent	-0.0005	0.0051	-0.0003
average housing value	-0.0002	-0.0002	0.0072
median commuting time to work	-0.0001	0.0001	-0.0005

The baseline adjusted R^2 refers to the OLS regression of unemployment rate on a constant term and the full set of regressors listed in the first column.
Each cell reports the change in adjusted R^2 caused by eliminating the corresponding variable from the baseline regression. A positive sign indicates a decrease in adjusted R^2.