Networking Off Madison Avenue

Mohammad Arzaghi and J. Vernon Henderson[1]

Brown University

March 25, 2005

New York County, or Manhattan, accounts for 24% of ad agency receipts in the USA. Years ago the popular image had large agencies on Madison Avenue generating this business. But today there are over 1000 agencies in different clusters spread over different neighborhoods in the southern half of Manhattan. In clustering, advertising is known for the key role that networking plays in the operation of agencies, where interviews suggest that informal networking among close neighbors is critical to many agencies' success. Networking involves the exchange of ideas; and while there is no market place for these ideas, exchanges may be more deliberate, rather than just Marshall's (1890) "mysteries" of the "air" type of information spillovers.

This paper examines the effect on productivity of having more near advertising agency neighborhoods and hence better opportunities for meetings and exchange within Manhattan. We will show that there is extremely rapid spatial decay in the benefits of having more near neighbors even in the close quarters of southern Manhattan, a finding that is new to the empirical literature and indicates our understanding of scale externalities is still very limited. The finding indicates that having a high density of commercial establishments is important in enhancing local productivity, an issue in Lucas and Rossi-Hansberg (2002), where within business district spatial decay of spillovers plays a key role. We will argue also that in Manhattan advertising agencies trade-off the higher rent costs of being in bigger clusters nearer "centers of action", against the lower rent costs of operating on the "fringes" away from high concentrations of other agencies. Introducing the idea of trade-offs immediately suggests heterogeneity is involved. We will show that higher quality agencies are the ones willing to pay more rent to locate in greater size clusters, specifically because they benefit more from networking. While all

---

[1] The research in this paper was conducted while the authors were Census Bureau research associates at the Boston Research Data Center. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau. The paper has been screened to ensure that no confidential data are revealed. We acknowledge helpful comments from Andrew Foster on aspects of the paper and from participants in seminars at Brown University, LSE, and the University of Toronto.

this is an exploration of neighborhood and networking externalities, the findings relate to the economic anatomy of large metro areas like New York—the nature of their buzz.

This study differs in several ways from traditional work on urban scale externalities which is carried out at either the metropolitan or county level, as reviewed in Rosenthal and Strange (2004). First we look at one specific service industry, advertising, while most of the literature focuses on manufacturing. Given the high concentrations of business service firms in large metro areas compared to manufacturers, we might expect to find much greater degrees of scale effects in service industries such as advertising, compared to manufacturing. Second, with the exception of Rosenthal and Strange (2003), the literature assumes scale externalities operate at a county or metro area level, with no spatial decay within geographic units. We have data at the level of the census tract in Manhattan that allows us to distinguish locations at no more than 250 meter increments. With this fine level of geographic detail, we can infer the high benefits of close social interactions and track their quick spatial decay.

Third, while the empirical literature debates what the source of scale externalities is in any context, what we find here seems more clearly to be information spillovers formulated as external benefits of networking, as opposed to scale effects from labor market search and matching externalities or from diversity of local intermediate inputs in monopolistic competition models. Southern Manhattan is all one labor market, drawing workers from all over the New York PMSA and even beyond. There is evidence of very modest wage gradients (over longer distances than just within Manhattan) in large cities; but it is ascribed to the fact that those living further from the center earn lower wages because they work locally and have lower commuting costs, rather than there being different labor markets (McMillan and Singell, 1992). We will argue that Manhattan is also generally all one market for intermediate business service suppliers and buyers, with, for example, any graphic service providers or headquarters a short cab ride away.

Rather, within Manhattan, the spatial variation in scale effects that matters within these close quarters is the availability of networking opportunities, or the information "spillovers" involved in frequent, repeated interactions within the different clusters of advertising agencies as described below. We will argue these networking benefits are so localized, that they are exhausted within the confines of sub-areas of southern Manhattan:

firms on the far west side of southern Manhattan derive no direct benefits from firms on the east side. Thus this paper raises the question of what is to be inferred from studies based on county or MSA level data. To what extent do findings in such studies reflect labor market externalities, diversity externalities, the greater choice of different clusters to choose from when locating in a large metro area, or the benefits of greater density of establishments ?

Fourth, cross city and county studies which try to infer the relationship between productivity and covariates such as local own industry scale are plagued by the problem of unmeasured characteristics which vary across geographic units such as legal, regulatory and business climate, access to specific markets, and specific infrastructure including the transport system which affects commuting and work effort. These characteristics affect left and right hand side variables in such studies. In a panel context one can try to control for these by fixed effects (Henderson, 2003), but these effects are not necessarily time invariant, especially over longer panels. Southern Manhattan is all one legal-regulatory-business climate and infrastructure system, so the usual cross-section unobservables are not an issue in identification. There will be tract level unobservables; but we will argue they are easier to deal with econometrically, so that identification here is more straightforward than county or PMSA level analysis.

Finally there is the issue of heterogeneity, upon which the scale externality literature is rather silent. Suppose typical results suggest that being in the midst of a bigger cluster of ad agencies improves productivity. Can we be sure that this is a "pure" scale effect, or is there a problem that higher quality firms are the ones in bigger clusters? Then a scale coefficient may be also capturing other factors to do with quality, and scale effects may be heterogeneous benefiting different types of firms differentially. In this paper, we will try to disentangle these effects.

1. Advertising Agencies

In this section we give some background information on the advertising agency industry, in general and in New York City. Then we look at New York in more detail.

1.1 General Background

In 1997, advertising expenditures in the USA totaled $187b, about 2.2% of GDP. Of this, only $17b involves advertising agency receipts (NIPA, BEA).[2] However of the $17b, $10b in advertising agency income comes from commissions on $89b of media billings carried out by advertising agencies, implying that over half of advertising expenditures in the USA go through advertising agencies. This reflects the fact that many advertising agencies serve primarily as intermediaries, or matchmakers between advertisers and media outlets. However advertising agencies are better known for the design of advertising campaigns. From the industry point of view, top notch firms provide "great creative thinking campaigns and ideas" for advertisers to market their products and to develop a "unique selling proposition" (Schemetter, 2003). Advertising agencies operate with internal teams, where a team typically provides the full range of services for a client – market research, creation of an advertising campaign, production, media, client management, and the like, where most agency costs are labor costs. Apart from potential media commissions from operating ad campaigns, advertising agencies derive a significant portion of the rest of their income from fees and billings for creative services rendered in the design of campaigns.

Like R&D (Jaffe et al., 1993), with creativity playing a central role, information sharing and information diffusion are thought to be critical to an agency's success. That sharing occurs within and across agencies, where agencies belong to formal and informal networks.[3] Based on informal interviews with agencies in New York and Chicago, we ascertained that informal networking appears to occur at the design stage of advertising campaigns where an agency has received a "request for proposal" [RFP] for an

---

[2] Advertising is split about 55-45 between national versus local advertising (Berndt and Silk, 1994) and is focused on automobiles (18%), retail, department and discount stores (15%) and then movies, cosmetics and toiletries, medicines, food, financial services and restaurants with shares each of 4.3-5%. In terms of media, TV, radio and cable TV account for 36%, magazines and newspapers 32% and yellow pages and direct mail 30%. The last two sets of numbers are from the AdAge website, based on reports by respectively TNS Media Intelligence and Coen/McCann-Erikson. Both data sets are property of Crain Communication, Inc.

.

[3] Data exist on the formal network structure of advertising agencies, where members pay dues to networks which have full-time managers and they attend network meetings several times a year. Hameroff (1998) lists eleven huge national and international formal networks of larger agencies. These formal networks primarily exchange information about marketing conditions across cities and countries.

advertising campaign (Arzaghi, 2004a). The agency turns to members of its local network for information (who by the odds are unlikely to have received the same RFP) on how best to respond to the request for proposal and how to design the potential campaign. In New York, we believe clustering occurs because of localized networking to enhance creativity; agencies share information and ideas where repeated face-to-face contact is critical. Contacts may be "official" such as meetings among members of a neighborhood network, unofficial such as informal contacts in local lunch and coffee places, or both.

The notion that networking is so important to advertising agencies compared to most other activities may be related to the fact that advertising agencies have a low degree of industrial concentration, both in terms of establishment and firm size. Table 1 gives relevant numbers for advertising agencies relative to business services more generally, noting business services are much less concentrated than manufacturing. Establishments under five employees account for 58% of all establishments and 14% of sales, higher than for business services generally. The average establishment size is 11 employees compared to about 50 in manufacturing and to 21 in the rest of business services. Single unit firms [SU's] account for 90% of establishments and 55% of sales in the advertising agency industry, compared to 81% and 45% in other business services. At the upper tail, the 17 largest (in terms of number of establishments) multi-unit firms only account for 22% of industry sales.

While the advertising agency industry is a non-concentrated industry with many small firms, it is heavily spatially concentrated in New York County, or Manhattan. In 1997, New York has 7.3% of advertising agency establishments in the USA and 20% of their employment. The difference between the 7.3 and 20 percents means establishment sizes are much larger in New York than in the rest of the country. As such, New York has 24% of all advertising agency receipts and 31% of media billings. For comparison, we note that New York has about 1.9% of all private employment in the USA; so location quotients for New York compared to the nation for advertising agencies, whether based on employment, sales, or establishment counts, are all extremely high.

In the empirical work, in examining the advertising agency business, we separate out and focus on the activities of single unit [SU] firms, as opposed to multiple unit [MU] firms. Why don't we pool the groups? In the text and Appendix B, we will argue that, in

econometric work, firms in each group seem to interact within the own group (SU and MU), but not across groups. MU and SU's have very different establishment sizes and types of activities. In New York, SU's have an average size of 11 employees, while MU *establishments* (not firms) have an average size of 155. For the approximately 16 % of establishments for which we have data in 1997, MU's are much more reliant on income from media receipts indicating their greater role as intermediaries between advertisers and the media, getting 48% of receipts from media billings, while SU's get 28%. In contrast, in NY, SU's get 21% of their income on commissions and fees for services provided by *other* agencies, compared to just 6% for MU's, an indication of greater "sharing" and cross-firm indirect sales by SU's. And most of the rest of SU income is from fees for their own services. In general, SU's are the firms where localized cross-firm networking is widespread and close spatial clustering is important. In contrast, for their more media based activities, MU's may rely more on intra-firm networks across their own establishments. Thus we separate MU's from SU's. Given MU's have a limited sample size, it is harder to make inferences about their behavior, especially in analyzing where births and movers locate; we generally drop them from consideration.

1.2 New York City

Within New York City, advertising agencies are closely clustered. Because of disclosure difficulties, we are not allowed to map agency locations using Census data, although the empirical work to follow is based on Census data. For maps, we use Advertising Red Books online for 2003. The Red Books give the location of about 800 of the 1000 or so agencies in Manhattan.[4] Figure 1 shows advertising agency locations in the southern half of Manhattan, defined as the area below 90th Street. Almost all agencies in New York are located below 90th Street. Overall, clustering is heaviest within a block or two of Madison and Fifth Avenues, but there appear to be many "sub-clusters" as we move north-south along the Avenues, looking both east and west. The Red Books do not give a reliable breakdown between single and multi-unit firms, so as to map them separately. If we compare Figure 1 with the rent gradient map in Figure A-2, we can see there are large clusters of agencies in medium rent areas, with somewhat fewer in the highest rent areas. In low rent areas on the fringes, there are also agencies, but much

---

[4] This is a publication of National Register Publishing.

fewer in number. Overall, for Census data in 1992, the simple correlation coefficient between the number of SU's in a tract and the log of rent price is .23.

In order to further explore the spatial distribution of agencies within Manhattan, we return to Census data. We examine similarity of location patterns by zip code, asking whether any comparison pair (e.g., SU stocks in 1977 versus SU stocks in 1992, or SU vs. MU stocks in 1992) is drawn from the same distribution (across zip codes). We use zip code location information here, rather than the Census tracts we use later, since only the former are available in all Census comparison years (1977, 82, ...97). First, Pearson χ^2 tests[5] show MU and SU establishment stocks have different location patterns across zip codes from one another in all years (1977, 82, 87, 92, 97), consistent with the hypothesized lack of SU-MU interaction noted earlier.[6]

Next we examine key aspects of SU location patterns, which are relevant to the analysis in the sections to follow. Again we rely on χ^2 tests to compare location patterns across zip codes for all of Manhattan. These following items show that location patterns of SU's have changed overtime in Manhattan, which is useful information in discussing identification later.[7]

(1) Comparing 1987 (or 1982) with 1997, the SU stock location pattern in 1987 differs from the SU pattern in 1997. Patterns change from 1987 to 1992, reputedly in response to relative rental cost changes at that time, moving advertising agencies out of some of the highest rent areas (off Madison Avenue).[8]

(2) Corresponding to (1), births of SU's in 1982-87 vs. 1992-97 have different spatial distributions.

---

[5] The test statistic is $T = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(m_{ij} - e_{ij})^2}{e_{ij}}$ where $e_{ij} = N \frac{n_{j\cdot}}{N} \frac{n_{\cdot j}}{N}$ is the ML estimate of probability of cell ij where $n_{i\cdot}$ is sum across rows (say, SU and MU) and $n_{\cdot j}$ is sum across columns (say, locations), and m_{ij} is the count in any cell. T is distributed χ^2 with degree of freedom (r-1) (c-1).

[6] We cannot reject, for each of the periods 92-97, 87-92, and 82-87, that SU and MU births have the same location patterns as each other. However for both of 87-92 and 92-97, the p-values are .052, on the border of rejection of similarity.

[7] These results on the first three items also apply (separately) to MU's.

[8] There is also a reduction in the official media commission rate for ad agencies at that time, which arguably deprived them of income.

(3) In the analysis, while we start by focusing on location patterns of births, we will also look at the smaller sample of movers. In all years, SU births and movers have similar location patterns.

Finally we look at whether there are significant differences across Census tracts in 1992 in average establishment sizes of SU's as measured by payroll and what the relationship is between tract counts of agencies and average size. As explained later, we will use payroll size as a measure of establishment quality. While F-tests suggest there is no variation in average establishment size in southern Manhattan, there is a correlation of .38 between the number of SU's in a tract and log of the payroll of the median size establishment. This correlation arises because small firms are found disproportionately in tracts with few firms; once we are in tracts with more than 8 firms, there is no relationship between establishment size and number of establishments in a tract. These facts suggest that larger (high quality) firms group together in larger clusters in key neighborhoods, while small firms are more on the fringe areas of southern Manhattan. What keeps the latter from moving to the better neighborhoods are presumably higher rent costs, which given their lower quality they can't afford. So there is spatial separation in the Manhattan market with the lowest quality firms operating on the fringes. Not only are rents and tract counts of ad agencies positively correlated as noted earlier, but so are rents and establishment size.[9]

2. A Model of Expected Networking and Location Choice

In this section, we develop a model of location choice within New York City for single unit advertising agencies, to help us think about empirical issues. We start with the specification of networking technology and location choice. In section 3, we turn to the details of implementation, data, estimation, and results for a simple model with no firm heterogeneity. In section 4, we turn to issues of heterogeneity.

2.1 Formulating Local Network Contacts

We specify a simple model for a new firm that is choosing a location in anticipation of informally networking with other agencies nearby. The distance between

---

[9] The simple correlation coefficient between rents and median payroll size both in logs for tracts with firms is .27.

agencies plays a key role in the extent of networking. While the specification and jargon is informed by the large theoretical literature on networks (e.g., Jackson and Wolinsky, 1996, Bala and Goyal, 2000, and Goyal and Moraga-Gonzalez, 2002), we look at just one firm and how to specify the benefits, costs and extent of networking for it. We model networking as overlapping with direct linkages only, where each pair of establishments decides on its pair-wise extent of contact, independent of other pair-wise arrangements.

The idea is that a new firm anticipates networking with advertising agencies in close-by locations, where this networking can involve social "meetings" with employees of other agencies (going to the bar after work), scheduled business meetings where some team members in a firm get together with some team members in another firm, or contacts through intermediaries such as local graphics firms or contract designers serving several agencies. A new firm in networking with any other firm has a match parameter, revealed after the firm has picked its location and established contact with the other firm. Given the ex post chosen level of networking for each pair, networking is costly in terms of time, where that cost rises with distance between pairs. This could represent commuting time for the requisite face-to-face meetings or monitoring of one another's activities.

Consider a firm operating at location j. The value of its contact with a firm at location i is

$$A_j[V_{ij}^{\gamma} \, u_{ij} - \, V_{ij}c(\delta_{ij})] - F \, . \tag{1}$$

V_{ij} is the volume of communications j gets from a firm at i and $0 < \gamma < 1,$ so there are decreasing returns to volume in any pair of contacts. u_{ij} is a "match" parameter drawn on the interval [0,1], which is firm specific between the firm at j and a specific firm at $i,$ and gives variation in the extent of contacts across firms. $c(\delta_{ij})$ is the cost per unit of communication, where δ_{ij} is the distance from i to j. Total communication costs really comprise firm labor costs in this simple formulation. Differences in A_j could represent

9

differences across locations in amenities. F is the initial fixed cost of making a contact with a firm at another location, upon which u_{ij} is learned.[10]

Once the firm spends F and learns u_{ij}, it chooses any V_{ij} so that

$V_{ij}^* = (\gamma/c(\delta_{ij}))^{1/1-\gamma} \ (u_{ij})^{1/1-\gamma}$. If $c(\delta_{ij}) = c(\delta_{ji})$ and $u_{ij} = u_{ji}$, then in this simple case $V_{ij} = V_{ji}$ and communications are reciprocal and two sided. In the present specification, ex post profit from contact with any firm is $A_j(1-\gamma)\gamma^{\gamma/(1-\gamma)}c(\delta_{ij})^{-\gamma(1-\gamma)}(u_{ij})^{1/(1-\gamma)} - F$ and ex ante profit is

$$A_j(1-\gamma)\gamma^{\gamma/(1-\gamma)}c(\delta_{ij})^{-\gamma(1-\gamma)}E[u^{1/(1-\gamma)}] - F \qquad (2)$$

where we denote the expected value of $(u_{ij})^{1/(1-\gamma)}$ as $E[u^{1/(1-\gamma)}]$, which we assume is common across all i, for now. Note the expected volume of communications and profits are decreasing in distance and communication costs.

Equation (2) relates to the key set of coefficients we will attempt to estimate— how expected profit to a firm at j is increased if a firm is added at i and how that profit declines with distance, δ_{ij}. Equation (2) also can be used to define the maximum distance, δ_{max}, over which a firm makes contacts and scale externalities are relevant. Empirically we will see that there appears to be such a cut-off point. Given the positive term in (2) is declining in δ_{ij}, δ_{max} is defined by

$$\delta_{max} = \max\{\delta_{ij} \mid A_j(1-\gamma)\gamma^{\gamma/(1-\gamma)}c(\delta_{ij})^{-\gamma(1-\gamma)}E[u^{1/(1-\gamma)}] - F > 0\}. \qquad (3)$$

δ_{max} is increasing in firm productivity, A_j, and in expected match productivity $E[u^{1/(1-\gamma)}]$, and is declining in costs.

In this simple context, total firm profits are simply the value of all contacts the firm makes less costs. Recognizing that there may be more than one firm at distance δ_{ij},

---

[10] While u_{ij} can take the value zero to represent "no match" possibility, if we want to incorporate issues of conflict of interest across agencies, we would want to have a probability p of any contact resulting in "no match" and formulate an expected fixed cost per contact as F/p, with only actual retained contacts contributing to variable profits.

or that distance may be discretely measured so there are n_i firms at i, for a firm at

location j ex post profits are $\Pi_j = A_j \{\sum_i^{\delta_{max}} n_i (V_{ij}^\gamma u_{ij} - V_{ij} c(\delta_{ij}))\} - R_j - F \sum_i^{\delta_{max}} n_i$,

where R_j is the rental cost of the location at j. Substituting in the firm's choice of V_{ij} and examining ex ante, or expected profits yields

$$E[\Pi_j] = A_j \gamma^{\frac{\gamma}{1-\gamma}} (1-\gamma) \, E[u^{1-\gamma}] \sum_i^{\delta_{max}} n_i c(\delta_{ij})^{-\frac{\gamma}{1-\gamma}} - F \sum_i^{\delta_{max}} n_i - R_j. \qquad (4)$$

In this context, within δ_{max}, the more firms there are (i.e., the greater the density), the higher profits. Adding a specific firm within δ_{max} adds the positive expression in equation (2) to the sum in eq. (4). This presents two issues: why there is more than one cluster in New York and how might we start to think about heterogeneity of firms?

Multiple clusters. The fact that profits are increasing in density is a force which would lead all firms in New York City to cluster together in one neighborhood. There are several limits to this. First, in the empirics, the strategy is to have a firm-location specific matching parameter drawn in the usual logit fashion, so firms have preferences over locations per se, and thus spread out, some in distinct clusters and some more in isolation. Second, in Manhattan the number of available spaces within a neighborhood is limited and it isn't clear there could be a neighborhood of 1000 agencies. Third, one could respecify the model so that there are decreasing returns to scale to a firm in overall volume of all communications, which then makes δ_{max} dependent on density, where the overall spatial span of contacts increases as density declines. Finally, further motivation for having multiple clusters comes from heterogeneity

Heterogeneity. A key issue in estimation is heterogeneity across firms. Section 4 below discusses how we parameterize how better quality firms are willing to pay more to be in bigger clusters, such as allowing γ to vary by firm quality. A context where there is variation in what different quality firms are willing to pay for features of ad agency clusters has the potential to deliver a separating equilibrium as in Arzaghi (2005). There, high quality firms cluster in high cost neighborhoods and lower quality ones operate on the fringes because they can't afford the higher costs in larger clusters.

Here differences in quality could be reflected by differences across firms in γ or in the specification of matching parameters. For the last, we could assume each firm is endowed with its own "communication ability" parameter u_j, a measure of firm quality. For any pair, the match gets $u_{ij} = \min[u_j, u_i]$, where u_j is the communication parameter for our firm at j and u_i is the parameter for a specific firm at i. Given $u \in [0,1]$, one can show that

$$E[(u)^{1/(1-\gamma)}] = (u_j)^{1/(1-\gamma)} - \int_0^{u_j} \frac{1}{1-\gamma} u^{\gamma/(1-\gamma)} G(u) du \qquad (5)$$

where $G(u)$ is the distribution function. Equation (5) suggests several considerations. First higher quality firms have higher $E[(u)^{1/(1-\gamma)}]$; and they benefit more from neighbors in equations (2) or (4), so that their scale effects are enhanced. They also benefit more from being in a neighborhood with better quality neighbors, in the sense of a better distribution function, $G(u)$.[11]

Whichever of these ways differences in quality are specified, higher quality firms will have more contacts and be larger, motivating the idea in the empirical section that ex ante size can be a measure of inherent quality. In the empirical work we will develop both neighborhood quality measures and ex ante individual firm quality measures for a sub-set of firms to see if higher quality firms compared to lower quality ones get greater benefits from either greater neighborhood scale, better neighbor quality, or both.

2.2 Empirical Implementation and Data: Homogeneous Firms

The model we estimate is based on eq. (4). In the treatment of space, locations are census tracts, which in NY are at a finer level of spatial detail than zip codes. Moreover a census tract is intended to represent a true micro "neighborhood". Each census tract has a rental cost per square foot of class A office space, availability of slots measured by the stock of private establishments, and then distance measures to other ad agencies and other

---

[11] If a firm chooses a neighborhood with a better distribution, $F(u)$, of neighbors in the sense of first degree stochastic dominance, then $E[(u)^{1/(1-\gamma)}]$ rises by $\int_0^{u_j} \frac{1}{1-\gamma} u^{\gamma/(1-\gamma)} (G(u) - F(u)) du > 0$. This rise in $E[(u)^{1/(1-\gamma)}]$ is greater for higher quality firms by $(1-\gamma)^{-1}(u_j)^{\gamma/(1-\gamma)}(G(u_j) - F(u_j))$.

activities where access might be important, such as broadcasting establishments (where ads are placed), headquarters (clients), or graphic services (inputs).

In the empirical implementation of equation (4) we want to allow office rental space to be a variable, rather than fixed cost and to introduce an error component. To do so, we redefine production technology. If l_j is the amount of space consumed at a unit cost, R_j, suppose ex post profits are specified as

$$A_j l_j^\alpha \exp\left(\sum_i^{\delta_{max}} n_i \{[V_{ij}^\gamma u_{ij} - V_{ij} \mathrm{c}(\delta_{ij})] - F\} \right) - R_j l \quad \text{Optimizing with respect to choice of } V_j$$

gives $A_j l_j^\alpha \exp\left(\sum_i^{\delta_{max}} n_i \{(1-\gamma)\gamma^{\gamma/(1-\gamma)} c(\delta_{ij})^{-\gamma(1-\gamma)} u^{1/(1-\gamma)} - F\} \right) - R_j l_j$. For ex ante profits,

taking expected values, maximizing profits with respect to choice of l_j and taking logs gives, after some rearrangement,

$$\tilde{\pi}_j \equiv \ln\Pi_j = C_j - \frac{\alpha}{1-\alpha} \ln R_j + \frac{1}{1-\alpha} \sum_i^{\delta_{max}} n_i \ln \mathrm{E}[\exp\{(1-\gamma)\gamma^{\frac{\gamma}{1-\gamma}} c(\delta_{ij})^{\frac{-\gamma}{1-\gamma}} u^{\frac{1}{1-\gamma}} - F\}] \quad (6)$$

The empirics will focus on estimating the effect for a location i of increasing the number of firms at that location on firm j profits, or $\ln \mathrm{E}[\exp\{(1-\gamma)\gamma^{\gamma/1-\gamma} c(\delta_{ij})^{-\gamma/1-\gamma} u^{1/1-\gamma} - F]$, and on estimating how those effects decline with distance. Note in (6) the term A_j is now buried in the C_j term, which below we will specify as containing a separable logit type error term, known to the firm.

2.2.1 Formulation.

In estimation we look at the location choices of births, or new firms in Manhattan, where firms pick the census tract which maximizes profits. In viewing location choices across tracts, as is traditional, we assume that the firm looks at each tract as "one" choice with a tract-firm idiosyncratic match parameter, ε_j, drawn from an extreme value distribution. Tracts are defined to be neighborhoods in New York centered along the avenues. However, spaces differ within a tract and rather than assuming the same draw for a firm within a tract, we assume each location, z, in tract j is associated with its own idiosyncratic matching parameter to the firm, ε_{jz}, drawn from a generalized extreme

value distribution. Office spaces on the same avenue in New York may be somewhat similar (in the same building) and certainly face similar street conditions, so it is natural to assume that these error drawings are correlated within tracts. In this specification a critical feature is that tracts have differing availability of spaces, where land and building usage in Manhattan is tightly regulated and inflexible over time (Glaeser and Gyourko, 2002). If there are more available commercial spaces in a tract, then a firm enjoys more idiosyncratic draws and better matching possibilities (Holmes, 2004).

This treatment of tract error drawings leads to the nested logit formulation where the probability that a firm chooses tract j (at the upper level of nesting) is given by

$$P_j = \frac{\exp[\pi_j + (1-\sigma)\ln m_j]}{\sum_i \exp[\pi_i + (1-\sigma)\ln m_i]} . \qquad (7)$$

In (7), π_j is $\tilde{\pi}_j$ in (6) without the C_j term which contained the separable ε_{jz} error term.[12] As such, m_i is the number of vacancies in tract i, which we will represent by (as proportional to) the total number of private establishments in the neighborhood; we have no separate information on vacancy rates by tract. The coefficient of m_i is $1-\sigma$, for σ approximately the within tract correlation of error drawings in a nested logit framework.

With homogeneous births, all covariates are tract characteristics and are not firm specific. In that case, if the expected number of births in a neighborhood, λ_j, takes the usual form ($\exp[\pi_j + (1-\sigma)\ln m_j]$), the problem may be equivalently estimated as a standard Poisson model of birth counts per tract (Guimarães, Figueiredo, and Woodward, 2000). (Note estimation of (6) as an upper level nested logit is computationally equivalent to estimating a model where profits for a firm in tract j equal $\pi_j + (1-\sigma)\ln m_j + \eta_j$, where η_j is drawn from an extreme value distribution.) We start by presenting ordinary Poisson results. However the usual first order problem is perceived to be that error draws for neighborhoods may be correlated with neighborhood covariates. As such we want to define a set of instruments that influence these covariates but are exogenous to contemporaneous error draws. Assuming such instruments exist, we can estimate a (non-

---

[12] That is, $C_j = \ln(Ce^{\varepsilon_{jz}})$.

linear) moment condition for the number of births, which also no longer imposes the Poisson assumption. If $\upsilon_j \equiv B_j - \lambda_j = B_j - \exp[\pi_j + (1-\sigma)\ln m_j]$, where B_j is the actual number of births and λ_j the expected number, the moment condition is

$$E\left[\upsilon_j \mid Z_j\right] = 0 \qquad (8)$$

where Z_j are instruments (Windmeijer and Silva, 1997 and Mullahy, 1997). Before discussing identification, we discuss the birth context.

2.2.2 Data.

We start by examining births of SU's in NY from 1992 to 1997. We look at births in the 164 census tracts south of 90th Street, which cover 97.4% of all SU's; the reason for the exclusion of the other 132 tracts is that we couldn't infer rent data per tract for them with any degree of confidence. Rents are based on zip code asking rents for class A office space in Manhattan in 1992; we take zip codes for which we have data and use a GIS rent-contour fitting routine to infer rents for all our Census tracts (see Appendix A). Tract rental prices vary threefold within southern Manhattan. In terms of births, a birth is a new firm to Manhattan—a firm that did not exist in the 1992 SSEL but did in the 1997 one. The SSEL is the listing from social security records of all private establishments. We chose the SSEL in Census years because in those years additional information from the Census is available for firms; and more critically the records on existence and location are updated and most accurate for Census years.

The 1992 SU stock in our tracts is 949 and the number of births is 502. A birth is a new company identifier, and as such includes buy-outs, so actual births may be overstated.[13] We base 1992-1997 births on 1992 covariates, which implies that these births have naïve expectations concerning economic magnitudes, which generally fits the data best. But admittedly there is a lot of turnover in firms and their locations so firms born in, say, the last year (1997) may be basing their decisions on a more updated pattern than the 1992 one. Thus based on 1997 Census data which contain a year of birth, we

---

[13] There is more of a problem for MU's where there is a lot of corporate turnover. For example, for SU's about 10% of establishments have been around for 25 years or more. For MU's the number is much smaller, probably reflecting the general waves of buy-outs in the mid-1980's: the establishment may still be there but with a different owner, or firm identifier.

also look at a sample of SU births for the 148 establishments who give their birth year as 1993 or 1994.[14]

For each tract, we then construct measures of access of that tract to nearby advertising agencies, dividing the stock of agencies in 1992 relative to that tract into rings. For SU's for a tract in which they might locate, we define 5 rings moving out in increments of 250 meters. Ring 1 is the count of existing SU advertising agencies in census tracts whose centroid is within 250 meters of the own census tract centroid. Mostly ring 1 is just the own census tract (but it can include up to 3 tracts). Ring 2 is the count of SU's in census tracts whose centroids are between 250 and 500 meters of the own tract, and so on for ring's 3-5, up to 1250 meters. Each additional ring typically contains 3-4 more tracts. For other activities such as broadcasting or advertising agencies belonging to multi-unit firms, such a fine level of detail did not yield results and in fact other activities seem to play little role, although we will experiment with distance measures for these, usually in increments of 500 meters. Note southern Manhattan is about 2600 meters across and about 11,000 meters long.

2.2.3. Instruments

In terms of identification, as noted earlier, by focusing just on Manhattan, we have minimized the usual fixed effects problem in cross-county or PMSA work, where localities have different unobserved characteristics in terms of labor markets, transport infrastructure, and regulatory-business-political climates. Now, the problem reduces to one of unobserved neighborhood characteristics within Manhattan, such as what the trendy eating and socializing spots for networking are, where construction is underway making face-to-face meetings more time costly, or where garbage, pick-pocketing, street people and the like are currently a more troublesome problem for employees and perspective clients. For these types of unobservables, it is realistic to assume they are constantly changing, so that draws from, say, 1977 are uncorrelated with those 15 years later. We will report a variety of tests concerning this assumption below, but granting us that for now, the issue is what is endogenous and what are strong instruments.

---

[14] SSEL records for those years show almost no births, reflecting the fact that the records were delayed in being up-dated.

In the empirical work, the most critical variable will turn out to be rents. The idea is that the effects of current neighborhood unobservables on businesses in general are capitalized into rents. Better conditions are reflected in higher rents, biasing the rent coefficient upwards (making it less negative). Given Manhattan's highly regulated housing situation, rent has strong instruments: total housing units in 1970, and share of those units in 1970 in buildings with 5 or fewer housing units. The latter two items reflect the potential, despite regulation, to convert residences to office space. In first stage regressions for rent, these variables have significant negative coefficients. We also control for distance to Rockefeller center as the commercial center of the city (which itself has no direct effects on advertising agency profits), but its effect in first stage regressions is weak.

The stocks by ring of advertising agencies are also potentially endogenous variables. Whatever trendy things influence rents and births today may have enough persistence into the past to affect some of the accumulation of current stocks. We will present results with and without instrumenting for stocks. Specification tests below suggest instrumenting for stocks may be unnecessary; but, given our strong priors on their potential endogeneity, we generally instrument for them. For instruments we use 1977 values of ring stocks.[15] Why should these be strong instruments, given we are asserting there is no persistence in unobservables influencing stocks in 1977 and 1992 and advertising agencies have shifted locations in Manhattan overtime, as documented earlier by comparing 1987 and 1997 location patterns? This shift in location patterns appears to be influenced by shifting relative rent costs over time, which induced many advertising agencies to leave the highest rent tracts. This effect of changing rents is also enhanced by an "exogenous" change in revenue structure for agencies. Until the early 1980's, agencies received a 15% commission rate on media purchases they made for advertisers. In part following through on a Justice Department decree, this rate was abandoned in the earlier 80's and by 1997 had fallen to 11%. This supposedly induced cost cutting measures (e.g., moving to lower rent districts), as well as helped refocus activities of some agencies.

---

[15] Actually we can only assign about 70% of establishments in 1977 to a 1992 census tract; and it is the counts of these 70% that we use.

Given all this, why are 1977 ring stock values strong instruments for later year values? While there are distinct overall shifts, for many clusters there is also a reasonable degree of persistence, driven by the difficulty (co-ordination failure) in moving clusters en masse. So as rents change, while over time there is erosion of existing clusters, there is persistence based on holding clusters together, because of the loss in social capital for any firm of deviating from the cluster (such as is the reinvestment of all the fixed costs of initiating new contacts in eq. (4)). While 1977 ring stock variables are strong instruments, there is still the question of their validity in achieving identification. We note results on that issue, as well as the strength of instruments below.

2.2.4. Heterogeneity and Quality.

Our focus is on births, where it might be reasonable to assume firms have no ex ante information on their quality and thus their locations decisions are well represented by our homogeneous birth model. However, we would like to allow for the possibility that firms have ex ante information on their quality and then see if higher quality firms are willing to pay more to be in more highly clustered neighborhoods. We have data for 95 SU firms who move from one tract to another within Manhattan between 1992 and 1997. We call these agencies, movers. For these movers, their size in the prior location is a predetermined measure of their inherent quality and is uncorrelated with shocks to profits in the tract of choice in 1997. In section 4 below we will estimate an empirical model that allows for firm heterogeneity.

3. Results: Homogeneous Firm Model

In this section we start with results for the basic birth specification and check the robustness of results, by looking at different samples and specifications, as well as direct productivity results. Finally, we consider the problem of spatial correlation.

3.1 Basic SU Results.

Table 4 contains the basic results. The first two columns are for ordinary Poisson models (equivalent to ordinary conditional logits, when the covariates all relate to choice characteristics) and the last two are for IV estimation based on the moment condition in eq. (7). We start by discussing the specification in Table 4 and the instruments used in the last 2 columns. Then we turn to results for our main variables, rent and scale variables in

columns (1) and (3). At the end we consider other variables which may influence location decisions (see columns (2) and (4)).

Identification.

In Table 4, as listed at the bottom of the table, instruments are 1970 housing market variables and ad agency and total establishment count ring stock counts in 1977. All covariates are treated as endogenous, except for 1992 own ring *total* establishments. By all criteria total spaces for commercial establishments in a tract are exogenously given by New York zoning constraints and not sensitive to immediate market conditions. For the column (3) formulation, first stage regressions indicate the instruments are strong: partial R^2 and F's average .50 and 13.0 respectively for the covariates (for column (4) partial R^2 and F's average .55 and 16.3 respectively). There may be concern regarding the validity of using 1977 ring stock variables as instruments. The Sargan test statistics in the columns (3) and (4) in the table are very low, indicating that the specification and our assumption that current errors are orthogonal to past stocks are valid. But we also conducted other experiments.

For example in ordinary Poissons in column (1), inclusion of 1977 ring stock variables leaves the coefficients on 1992 ring stock variables unaffected (with coefficients on 1977 variables being small and insignificant). This suggests the 1977 variables are valid instruments, since the 1992 coefficients would decline if 1977 variables are correlated with the unobserved influences on profits that are hypothesized to affect 1992 stock variables. Similarly, we divided the 1992 ring stock variables into new (post 1982 birth date) versus old firms. In ordinary Poisons, the former strongly dominate the latter, since they more likely to be correlated with current shocks. In IV estimation such as in column (3), with so many variables the results are unstable, but the pattern of dominance by new stocks disappears as it should, since the IV estimation removes any correlation of the newer stocks with current shocks.[16]

Rents and local neighborhood size.

---

[16] In the first experiment, for the current ring variables, the coefficients are respectively .023, .017, .0052, .0034 and -.15, close to column 1 estimates; while for 1977 ring stock variables, the coefficients are .00046, .0035, -.0031, -.0037, and -.0025, all insignificant. For new stocks versus old, the ordinary Poisson coefficients for new are .28, .042, .022, -.0051, and -.00092; while for old stocks they are -.12, .087, .075, -.12, and -.052. For IV results, in the basic set, old stocks have coefficients of .020, .018, .0035, -.00067, and -.00053, while new have .032, .024, -.0015, -.012, and -.011.

Before proceeding, we note that we use two interpretations of the coefficients of the estimated profit function, in our discrete choice framework. First, if one adopts a count model interpretation, coefficients tell by what particular percent a change in a covariate raises or lowers the expected number of births in a tract. Or equivalently, in a logit framework, the effect is to change the probability of a birth by a particular percent (assuming the base probability of a birth in any tract is small). Second, we can compare relative magnitudes of effects on profits. So we can ask what percentage increase in rent per square foot of office space a firm would be willing to pay to have one more neighbor. Or, similarly, we can calculate how many standard deviations increase in unit rent a firm would pay for a one standard deviation increase in number of neighbors. Later, for a very small sample of firms, we will actually look at the effects of covariates on ex post absolute profits, which will give us a sense of how our coefficients, which are normalized parameters of the profit function, compare in absolute magnitude to actual parameters. Means and standard deviations of covariates are in Appendix B.

In Table 4, the rent variable has the appropriate negative sign. But, as expected, a comparison of columns (1) and (3) shows that rent is the one variable whose coefficient is affected by instrumenting: the absolute value of the column (3) coefficient is threefold that in the column (1). The coefficient in column (3) is a high elasticity with respect to rental costs, where a 1% increase in rents would lead to a 2.6% decrease in expected numbers of births, indicating births within Manhattan are very sensitive to rents. In particular, a one standard deviation change in the rent variable (.17), leads to a 44% drop in expected births in a tract. This magnitude helps explain the shifting pattern of advertising agency locations in Manhattan in response to relative rent changes across tracts over the last 15 years, as noted earlier.

The number of commercial establishments of all types in a tract represents the relative, potential number of slots available for rental for a new agency. A 1% increase in this raises expected births by 0.5%, again a high effect. The interpretation is that tracts with more slots offer better idiosyncratic matching of individual establishments to the neighborhood. Under a nested logit interpretation, the coefficient implies a significant positive correlation between error drawings within tracts. We will later look at spatial correlation across tracts. We note for the record the main effect of dropping the control

on total commercial establishments in column (3) is to raise the ring 1 coefficient on ad agency stocks from .0198 to .0224, a modest effect.

Scale Effects

The main results of interest are the effect of neighboring SU's on expected births and their importance in influencing profits relative to other variables. For these variables, while instrumenting has little effect on coefficients it does modestly improve Sargan values, so in column (3) we treat these ring counts as endogenous. Later in Table 5, we will consider results where only rents are endogenous.

In Table 4, column (3) shows that, as we move from the first ring out in 250 meter increments, the coefficients are .020, .023, .0042 and then zero beyond 750 meters. That is, any inferred networking effects end with ring 3. Based on numbers of ad agencies per ring in Appendix B and changes in ring areas as we move out, the number of ad agencies per unit spatial area does not decline as we move to outer rings, so there is no sense that rings 4 and 5 have anything less to potentially offer the typical firm in the sample; they are just too far away. Interactions occur within 750 meters, perhaps a 15 minute journey of elevator rides and walking during the day within Manhattan with its crowded conditions. This finding is robust to respecifications of the ring distance measures, such as making rings equal area or using finer or grosser definitions of rings. As noted earlier each additional ring encompasses 3-4 more tracts, which is a small increment. For example, on a lattice, ring 1 which is typically the own tract has 8 contiguous neighbors (tracts to the N, S, E, and W and then to the NE, NW, SE, and SW), but when we go to ring 2 we typically add in only 3-4 tracts (whose centroid falls between 250 and 500 meters away).

We also note that, while in some tables to follow, the ring 1 effect is larger than the ring 2 effect, the numbers are always close. However at 500 meters, the drop to the third ring is always large, with the coefficient being small and insignificant in IV estimation in this table, although it is more typically significant in the tables to follow. For the hypothesized networking effects, close spatial proximity is critical and thus so is having greater densities of establishments in commercial neighborhoods. But why does the ring 1 effect not dominate the ring 2 one? While, for a firm it is great to have a network "member" 5-10 minutes walk down the block, having an agency right next door

raises the problem of "poaching". A next door neighbor benefits from being able to "advertise" to clients who visit the own firm, and hence potentially to poach them, which would then reduce the net benefit of this neighbor.

In terms of magnitudes, as noted above there are several interpretations. First is the effect on expected births. So if the number of neighbors in ring 1 increases by one that increases the expected number of births by 2%. But it may be more helpful to place this in the context of the likely variation across tracts in numbers of neighbors. For that, we note the standard deviation of ring 1 neighbors is 10 and of ring 2 is 20 (and ring 3 is 49).[17] If we increase the number of immediate neighbors by one standard deviation that raises the expected number of births by 20% and if we increase the number in ring 2 by a standard deviation that raises the expected number of births in the own tract by 46%. Even for ring 3, with its point estimate of .004, the effect of a one standard deviation increase on expected number of births is 20%. These are large absolute effects.

Alternatively one can view the effects in terms of willingness to pay more rent in order to enhance networking possibilities. This gives a sense of the trade-off involved in choosing higher rent locations with more neighbors versus fringe locations with fewer neighbor but lower rents. The coefficients tell us that a firm would be willing to pay 0.8% (.0198/2.57x100) increase in rent per square foot to have one more neighbor in the first ring, with a similar magnitude for ring 2. Firms would be willing to pay just under a half a standard deviation increase in unit rent (or .078 of .17) for a one standard deviation increase in immediate neighbors. Or if a firm moves from the fringe with no neighbors to the height of the action in our sample with 50 immediate neighbors, it would be willing to pay over 2 standard deviations in terms of unit rental increases. Of course, that move is likely to bring more neighbors in rings 2 and 3, as well. A key issue examined in the next section is whether different types of agencies benefit differentially from more neighbors and hence are willing to pay differentially.

---

[17] Note the standard deviation and average of ring counts rise as we move out, since the ring area ($\pi(r_1 - r_2)(r_1 + r_2)$) rises as we move out where $(r_1 - r_2)$ is a constant (the 250 m. increment) but $(r_1 + r_2)$ rises with distance. Also these are standard deviations across tracts, unweighted by the number of births per tract. Standard deviations of tract characteristics weighted by births are generally higher as are the means.

These networking-scale results are based on a linear formulation of ring counts. We also experimented with a quadratic formulation and a log-linear one. We don't highlight these results because their IV versions are statistically weak.[18] But for ordinary Poissons both formulations "work". Under a log-linearity, the coefficients (and standard errors) for the first three rings are .391 (.066), .205 (.070), and .140 (.0603); these are very high scale elasticities. Quadratic scale effects are also large. For example for the first ring, the coefficients on the linear and squared terms are .0470 and -.000509 (both significant at 10% or better level). If we increase the number of neighbors by a standard deviation from 6 to 16, in the first ring expected births rise by 41%; and marginal effects persist until 45 neighbors within the first ring.

Other Activities.

In the rest of Table 4, we explore whether proximity to other types of activities matters for SU's. First trying these at a scale of 250 meters produced unstable and insignificant results. Thus we try these at a scale of 500 meter increments. We experimented with access to graphic services and headquarters, but neither of these had significant or noticeable effects. That leaves MU's and broadcasting. Ordinary Poisson results suggest that having MU's nearby helped in the first ring (significant coefficient of .027) and hurt in the second; but IV results give small and insignificant coefficients (first ring coefficient of .0045).

We focused on a specification with broadcasting (employment) in an immediate ring of 500 meters. In column (4), the number of broadcasting employees within the first ring positively and significantly affects profits. The IV results show that a one standard deviation increase in broadcasting employment (1785) raises expected births by 18%, a noticeable effect (although the ratio of standard deviation to mean is high (3.8) reflecting zeros in many rings, with very high concentrations in a few).

3.1.1 Robustness

In Table 5, we carry out a number of robustness checks as to specification and sample. First, in column (1) of Table 5, we re-estimate the column (3) Table 4 specification, treating just rents as endogenous, with ring stocks as exogenous. The

---

[18] No coefficients are significant in the quadratic and only the coefficient for the first ring is significant in the log-linear formulation.

Sargan test statistic is still excellent and results very similar; one impact is to make the ring 3 effect significant, due to lowering its standard error. In column (2) we turn to a major concern in the treatment literature: do samples of possible locations include ones that are not relevant or are infeasible, potentially giving overstated results? In column (2), we choose as the sample the 138 tracts of the 164 in our data that ever have any ad agencies (SU's or MU's). Again results are very similar to Table 4. The rent coefficient is about 25% lower and the ring 1 stock variable coefficient is about 11% lower, so that the net result is a slightly higher increased willingness-to-pay for more neighbors in terms of rent increases. Also, the ring 3 coefficient is now significant at a better than 10% level.

In column (3) and (4) we turn to another issue. So far we have been examining births over a five year interval. This could be problematic because there is so much turnover and change in conditions over 5 years. Therefore, we drew a sample of SU births which are firms that appear in the 1997 Census and list a birth year of 1993 or 1994. This includes a few firms we didn't count as births in the prior sample (for reasons that are unclear) and, of course, excludes the last 3 years of births in the prior sample. For these we report ordinary Poisson (column (3)) and IV estimation (column (4)). The IV results in column (4) of Table 5 are very similar to those in column (3) of Table 4. The only noticeable differences are that now the ring 1 coefficient exceeds that for the ring 2 (by 25%) and the ring 3 coefficient is noticeably larger (43%) and is now significant at a 10% level.

In the second experiment, for the 30 SU *existing* establishments in 1992 in Manhattan for which the Census data exist to calculate profits as sales minus all operating expenses including rent, we regress these log profits on the variables in Tables 5. Note that controlling for the number of establishments is no longer relevant per se for this non-birth specification. Given the small sample size, we focus on OLS results in column (5), but give 2SLS results for rents (only) endogenous in column (6). Point estimates of coefficients in Table 5 column (5) are larger in absolute terms; and the pattern of coefficients is similar to those in prior estimations, with a large negative rent coefficient and high benefits of near neighbors, which decline sharply with distance. The results hint at what absolute coefficients might be, compared to the discrete count estimation to date. In the OLS profit function, coefficients are at least twice ordinary Poisson coefficients for

rents and the first two ring variables. In the 2SLS results, coefficients for the first two ring variables are again about twice those in other estimation, or more. While the sample size is small, the numbers are suggestively large.

Finally in Appendix B, we present results for MU's where the stock is 132 establishments and the number of births is 67. There the table shows that MU's have similar patterns of signs and significance to covariates as SU's. The discussion in the Appendix indicates that, just as SU's benefit from nearby SU's and not MU's, MU's benefit from nearby MU's and not SU's. This separation implies that MU's may for example interact with other MU's to share information about media purchases, of which SU's do relatively less; while SU's may interact with each other at the creative design stage across firms. For that, MU's may rely on intra-firm interactions at the design stage.

3.1.2 Spatial Correlation.

When using data at this fine level of detail, and in addition when asserting within tract correlation of error terms, the issue of spatial correlation of error terms across tracts arises. Before addressing that directly, note that in Table 4, when we instrument for ring variables, we instrument for them all. So if the shock to the own ring affects ring stocks and shocks are correlated across tracts, then the shock to the own ring could be correlated with ring 2 or ring 3 covariates. Instrumenting with historical variables takes care of that problem; even if the instruments are historically spatially correlated, the concern is that the shocks today are independent of these historical variables.

The key issue with spatial correlation is the potential for biased estimates of standards errors. Thus we checked for spatial correlation, using a Moran I test on the residuals from both the ordinary Poisson and the GMM models in Table 4, columns (1) and (3). The Moran I test examines whether a regression of neighbors' average residuals on own tract residuals is significant. Near neighbors in the averaging are given a weight 1 and others a weight 0. In denoting what is a near neighbor, the usual procedure is to think of each tract as having 8 neighbors on a lattice. We use a radius from the own tract that gives an average of 8.3 neighbor tracts as neighbors. For this radius, the hypothesis of no correlation cannot be rejected at a p-value of .10 for an ordinary Poisson. The result for

GMM coefficients is similar—no rejection with a p-value of .11.[19] Thus within our data, while, given tracts are neighborhoods, there is within tract correlation, the problem of cross tract correlation does not seem compelling.

4. **Firm Heterogeneity**

Do different quality firms value scale externalities differently and are higher quality firms willing to pay more to be in better neighborhoods? These questions raise two sets of issues. What is a useful sample of firms that knows their quality and how might we measure quality? Second, how do we incorporate firm heterogeneity into the empirical estimation? In answer to the latter question, we follow the approach in Berry (1994) and Bayer, McMillan, and Rueben (2004) and postulate a heterogeneous coefficients model. In the profit function appearing in (6) we assume the coefficient vector ϕ for neighborhood covariates now takes the form

$$\phi^k = \phi_0 + \phi_1 z^k. \qquad (9)$$

ϕ^k are firm k's specific coefficients on the X_j covariates, where those coefficients vary with firm characteristics, z^k, which in this case will be a measure of firm quality. For scale effects in equation (6), we are allowing $\ln E[\cdot]$ to change as either γ or the expected u drawing from equation (5) vary with firm quality; for rental effects we are allowing α to vary.

In thinking about relevant samples and measures of quality, first we note that, given the earlier discussion in Section 2, controlling for neighborhood quality, better firms benefit from networking more, have more communications and hence more employees and larger payrolls. That is, expected volumes of communications rise as either γ or $E[u^{1/1-\gamma}]$ rise; and hence establishment sizes of higher quality firms are larger. For births we have no ex ante information on quality and might not expect them to have learned their inherent quality; and, for births, ex post size measures are likely to be influenced by contemporaneous error drawings. However we have a sample of 95 movers

---

[19] Only if we drop the radius so the average number of neighbors is 3.3, can the hypothesis of no correlation be rejected (p-value of .015). We also regressed residuals of pair-wise neighbors on each other, getting exactly the same pattern of results. For those we also looked at a sample requiring all pairs to both have counts greater than 1 (worrying about the discreteness problem with count data). Again the results are the same—no correlation for distances where the number of neighbors averages 7, or more.

who change location in the 92-97 time period within Manhattan. These more experienced firms presumably have more informed notions of their quality; and, for 82 of them, we have a size measure of quality-- payroll in their prior location in 1992, where that variable will be uncorrelated with the error drawing in the new location but will be a function of underlying firm quality. We note payroll and employment information from social security records are the only non-imputed census variables on firm economic magnitudes available for more than a handful of firms. We use payroll (where there is a uniform wage for uniform workers in Manhattan), to allow better quality firms to also hire better quality, or higher paid employees, but there is little difference in results.

Once we introduce firm heterogeneity and quality issues, we might also ask whether firms not only value counts of neighbors as assumed so far, but also their quality. In the "matching specification" discussed in equation (5), firms benefit more from being in neighborhoods with better distributions of quality and higher quality firms benefit the most. For an incoming firm, payroll levels of existing firms are a signal as to neighborhood quality. As such, we use the median firm payroll by ring for SU's in 1992 as a measure of neighborhood quality, recognizing this is an approximation.[20] We instrument for this with the 1977 measure, given both that actual quality may be influenced by unmeasured neighbor characteristics and given that 1992 average neighborhood payroll measures actual quality with error. In the empirics we focus on ring quality for just the first ring. Not only is ring quality of the 2nd and 3rd rings insignificant in estimation work, it seems less well defined. While it is easy to represent the scale effect in the second ring by a count of SU's from say 250-500 meters away, for quality, quality to the north may be rather different from quality to the south and picking the median in the ring seems at best a noisy measure of quality.

To estimate the model, one can proceed to estimate (7) amended by (9) directly by conditional logit, which we will do as a baseline case. However as discussed in Section 3, neighborhood characteristics are endogenous. Here in Section 4, we employ a different, two stage identification strategy, given we now have firm specific covariates as

---

[20] The issue is characterizing quality by a single measure of size. For example, this is a correct measure if the stock of firms is spread across neighborhoods so as to perfectly segment the line representing the quality types in the market. With perfect segmentation, the lowest quality firm in a neighborhood is at least the same quality as the highest quality firm in the next best quality neighborhood. Then greater size signals a first degree stochastic dominant improvement in the distribution.

well as tract ones. First, we estimate (7) with tract fixed effects, so the basic logit equation for firm k considering location j is

$$P_j^k = \frac{\exp[d_j + \phi_1 z^k X_j]}{\sum_i \exp[d_i + \phi_1 z^k X_i]} \qquad (10)$$

In (10), d_j is a neighborhood dummy variable, which controls for all neighborhood characteristics, observed or not. As long as the z^k are exogenous (see below), the ϕ_1 are then identified. But that leaves the base coefficients, ϕ_0, for the X_j covariates. These are given by the equation

$$d_j = b_j + \phi_0 X_j, \qquad (11)$$

where the LHS dummy variables are estimated in (10). To identify the ϕ_0, given the X_j are correlated with unobserved neighborhoods contemporaneous attributes, equation (11) is estimated by 2SLS, using as instruments the variables we used before in estimating the moment condition (9).

Table 6 contains the results. Because of the limited sample, to sharpen the reported results, we truncate ring effects at 3 rings, as is indicated by preliminary results on movers, as well as all prior tables. The first two columns give ordinary logit results for movers. Column (1) presents results without heterogeneity, where scale is for 3 rings and quality for 1. Results for movers, except for the rent variable, are similar to the results in column (1) of Table 4 for births (also another robustness check on Table 4). Column (2) presents the ordinary conditional logit results for the full model in (7), amended by (9) to allow for heterogeneous coefficients, so all variables in column (1) are now additionally interacted with the quality measure— payroll size in the prior location. If we evaluate these at median size of movers, results in column (2) are similar to those in column (1) on all variables. For that comparison,[21] we use a "synthetic" median, which is 4.75; this is the log of the mean size of the 10 firms around the median. However there are enormous differences by firm quality. In particular ordinary logit results suggest better quality firms

---

[21]At median size, the establishment, rent, and ring 1-3 scale and ring quality (heterogeneous) coefficients are respectively .21, -.36, .022, .015 and .0075, while those in column (1) are .30, -.093, .020, .018, and .0059, indicating here a formulation without heterogeneity tends to capture the coefficients of the "representative" firm .

are much less sensitive to rent differentials and much more sensitive to immediate ring 1 neighbors. We explore this next in columns (3) and (4), where issues of endogeneity are handled.

In column (3) we present results of the tract fixed effect logit model in equation (10). Notice the coefficients for these interactive variables are similar to those in column (2), but with stronger rent and ring 1 scale coefficients. In column (4), we recover the ϕ_0 coefficients by estimating equation (12) by 2SLS. For this, the sample size is only 40 tracts which raises a small sample problem for IV estimation, so in column (5) we also give the OLS results, which are similar to IV ones. To evaluate effects of any variable, we must combine the results in columns (3) and (4).

First we note that the results indicate that higher quality firms are less concerned about rent costs. The strongly negative rent coefficient for a small firm declines in absolute magnitude to -.36 for the median size mover. This could indicate that there is also a fixed cost portion to rent which effect diminishes with firm size.

Second, higher quality firms benefit more from scale externalities at least in ring 1. The ring 1 scale effect becomes positive at a very small firm and then rises sharply, so that at median size it is a high .034 (cf. Table 4 coefficient of .02) and at median size plus one standard deviation of size it is a huge .078. High quality firms are much more likely to choose neighborhoods with more SU's. Given the rent effect declines with firm quality, the willingness to pay in terms of rents for more neighbors is very high for higher quality firms. Point estimates suggest the median firm willing to pay a 9.4% increase in unit rent to have one more neighbor.

Finally, in the Table, we look at results on neighborhood quality for our movers, as measured by the median payroll size of existing establishments in 1992 in the first ring. There neighborhood quality doesn't seem to matter. Not only in column (3) is there no interactive effect whereby higher quality firms would pay more for higher quality neighbors, overall the effect is negative and insignificant. Scale matters a lot to high quality firms, but not the quality of neighbors, as currently measured. Of course we are looking within New York, which nationally has a high end cut of larger and presumably higher quality firms (Arzaghi, 2005).

5. **Conclusions**

There are two fundamental findings in the paper. First at a micro spatial level, scale externalities are very large but they dissipate very quickly with distance for advertisers and are gone by 750 meters. The effects seem to strongly suggest networking, or information "spillover" effects. That begs the questions of what one measures with more aggregate data, such as county scale effects. There may labor market externalities or it may be that counties with more advertising agencies and more clusters of agencies offer both more choice of neighborhoods and perhaps denser neighborhoods. The point is that what aggregate measures capture still remains ill-defined.

The second finding has to do with how different quality firms value neighborhood attributes. Estimates on movers suggest there may be segmentation across neighborhoods in New York, whereby higher quality firms value having near neighbors more than low quality firms and, as such, may separate themselves into large clusters away from low quality firms, by locating in high rent neighborhoods that low quality firms can't afford.

References

Arzaghi, Mohammad (2004), "Knowledge Sharing among Small Business Service Providers: The Case of Advertising Agencies," Working Paper, Brown University.

Arzaghi, Mohammad (2005), "Quality Sorting and Networking: Evidence from the Advertising Agency Indsutry," Working Paper, Brown University.

Bala, Venkatesh; Goyal, Sanjeev (2000), "A Noncooperative Model of Network Formation," *Econometrica*, 68, 1181-1229.

Bayer, Patrick; McMillan, Robert; Rueben, Kim (2004), "Residential Segregation in Equilibrium," Yale University mimeo.

Berndt, Ernest R; Alvin J. Silk (1994), "Costs, Institutional Mobility Barriers, and Market Structure: Advertising Firms as Multiproduct Firms," *Journal of Economics and Management Strategy*, 3, 437-450.

Berry, Steven (1994), "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25, 242-262.

Duranton, Gilles; Overman, Henry (2004), "Testing for Localization Using Micro Economic Data," LSE Mimeo.

Ellison, Gllen; Glaeser, Edward (1997), "The Geographic Concentration of Industry: A Dartboard Approach," *Journal of Political Economy*, 105, 889-927.

Glaeser, Edward; Gyourko, Joseph (2002), "The Impact of Zoning on Housing Affordability," Harvard Institute of Economic Research Paper # 1948.

Goyal, Sanjeev; Moraga-Gonzalez, Jose Luis (2001), "R&D networks," *RAND Journal of Economics*, 32(4), p 686-707.

Guimarães, Paulo; Figueirdo, Octávio; Woodward, Douglas (2000), "A Tractable Approach to the Firm Location Decision", *Review of Economics and Statistics*, 85, 201-204.

Hameroff, Eugene (1998), *The Advertising Agency Business*, New York: McGraw-Hill.

Henderson, J. Vernon (2003), "Marshall's Scale Externalities," *Journal of Urban Economics*, 53(1), 1-28.

Holmes, Thomas (2004), "The Location of Sales Offices and the Attraction of Cities," *Journal of Political Economy*, forthcoming

Jackson, Matthew; Wolinsky, Asher (1996), "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44-74.

Jaffe, Adam B.; Trajtenberg, Manuel; Henderson, Rebecca (1993), "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations", *Quarterly Journal of Economics,* 108, 577-598.

Lucas, Robert E. Jr. and Estaban Rossi-Hansberg (2002), "On the Internal Structure of Cities," *Econometrica*, 70, 1445-1476.

McMillen, Daniel P. and Larry D. Singell, Jr. (1992), "Work Location, Residence Location, and The Intraurban Wage Gradient," *Journal of Urban Economics*, 32, 195-213.

Mullahy, John (1997), "Instrumental Variables Estimation of Count Data Models, With an Application to Models of Cigarette Smoking Behavior," Review of Economics and Statistics, 79(4), 586-593.

Rosenthal, Stuart; Strange, William (2004), "Evidence on the Nature and Sources of Agglomeration Economies," in *Handbook of Regional and Urban Economics*, J.V. Henderson and J-F Thisse (eds.), North Holland.

Rosenthal, Stuart; Strange, William (2003), "Geography, Industrial Organization and Agglomeration," *Review of Economics and Statistics*, 85, 178-188.

Schemetter, Bob (2003), *Leap! A Revolution in Creative Business Strategy*, Hoboken, NJ: John Wiley & Sons, Inc.

Windmeijer, Frank; Santos Silva, Joya (1997), "Endogeneity in Court Data Models: An Application to Health Care," *Journal of Applied Econometrics*, 12, 281-294.

Table 1. 1992 Economic Census Numbers on Concentration in Advertising[*]

	advertising agencies		all other business services	
	share of total	no. of units	share	no. of units
establishments with under 5 employees: establishments	58%	6890	52%	133,380
sales by establishments with under 5 employees	14%	6890	10%	133,380
establishments with 100+ employees: establishments	1.3%	156	4.1%	10,590
establishment with 100+ employees: sales	33%	156	40%	10,590
single unit firms: establishments	90%	12,453	81%	236,212
single unit firms: sales	55%	12,453	45%	236,212
multi-unit firms: sales	45%	581 firms 1481 est.	55%	12,146 firms 56,460 est.
multi-unit firms with 10+ establishments: sales	22%	17 firms 408 est.	33%	771 firms 31,325 est.

_____

[*] The establishment numbers cover just "operating establishment". The firm numbers cover all establishments. Numbers are from the 1992 Census of Services.

Table 2. New York City's Role in the Ad Agency Industry 1992

Share of US establishments	7.3%
share of US employment	20%
share of US receipts	24%
share of media billing	31%

Table 3. New York City Single Unit versus Multi-Unit Establishments

	1997 Stock	**Born Before**		**'92-97 Births**
		'73	**'88**	
single unit est.	1088	104	357	545
multi-unit est.	149	3	77	67

Table 4. Birth Models for SU's.

	Ordinary Poisson	Ordinary Poisson	IV estimation	IV estimation
Ln(total no. of	.462**	.472*	.498**	.563**
establishments)	(.0756)	(.0767)	(.117)	(.120)
Ln(rent/sq ft)	-.819**	-1.06**	-2.57**	-2.60**
	(.289)	(.329)	(.773)	(.850)
Stock SU's	.0206**	.0208**	.0198**	.0221**
0-250 m.	(.00430)	(.00431)	(.00758)	(.00916)
Stock SU's	.0167**	.0167**	.0228**	.0190**
250-500 m.	(.00264)	(.00264)	(.00538)	(.00530)
Stock SU's	.00387**	.00368**	.00419	.00281
500-750 m.	(.00137)	(.00139)	(.00259)	(.00261)
Stock SU's	-.00149	-.00171	-.00149	-.00298
750-1000 m.	(.00158)	(.00160)	(.00274)	(.00292)
Stock SU's	-.00219	-.00184	-.00241	-.000945
1000-1250 m.	(.00147)	(.00149)	(.00251)	(.00290)
Broadcast emp.		.0000382*		.000103**
<500 m.		(.0000233)		(.000035)
Constant	.116	.816	5.46**	5.23*
	(.918)	(1.02)	(2.61)	(2.69)
N [births]	164	164	164	164
Pseudo R^2	.504	.507		
Sargan [p-value]			1.71 [.945]	1.57 [.954]

Instruments are distance to Rockefeller Center, 1970 log of total housing units, 1970 share of housing units in buildings with less than 5 units, stocks of SU ad agencies in each of the 5 rings in 1977, total count of all establishments in each of the first four rings in 1977, and 1992 total number of establishments in the own tract. For column 4 we add the 1977 broadcast employment within 500 meters.

Table 5. Robustness: Initial Births and Profits

|	IV estimation: only rents endogenous	IV estimation: "relevant" tracts only	Ordinary Poisson: 93-94 births	IV estimation: 93-94 births	OLS profits: '92 SU's	2SLS profits (rent endogenous)
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(total no. of establishments)	.563**	.458**	.411**	.498**	n.a.	n.a.
	(.0820)	(.110)	(.137)	(.201)		
Ln(rent/sq ft)	-2.30**	-1.96**	-1.72**	-2.71**	-3.38	-2.35
	(.623)	(.719)	(.572)	(1.24)	(2.62)	(2.95)
Stock SU's < 250 m.	.0166**	.0176**	.0334**	.0218**	.0796*	.0836**
	(.00489)	(.00747)	(.00817)	(.0104)	(0448)	(.0400)
Stock SU's 250-500 m.	.0176**	.0180**	.0147**	.0164**	.0357	.0354
	(.00290).	(.00524)	(.00494)	(.00755)	(.0255)	(.0224)
Stock SU's 500-750 m.	.00335**	.00399*	.00451*	.00735*	-.0327**	-.0328**
	(.00167)	(.00224)	(.00250)	(.00404)	(.0125)	(.0110)
Stock SU's 750-1000 m.	-.00203	-.00270	.000702	-.00158	-.00764	-.00812
	(.00200)	(.00254)	(.00281)	(.00501)	(.0133)	(.0118)
Stock SU's 1000-1250 m.	-.000594	-.00146	-.00577**	-.00512	.0160	.0139
	(.00177)	(.00224)	(.00272)	(.00467)	(.0132)	(.0122)
Constant	4.27**	3.98*	2.26	5.96	22.6**	19.3**
	(1.83)	(2.37)	(1.81)	(4.35)	(8.61)	(9.61)
N	164	138	164	164	30	30
[Pseudo/adj.] R^2			.343		.138	
Sargan [p-value]	2.65 [.851]	1.84 [.934]		3.86 [.570]		

For instruments, see Table 4.

Table 6. Heterogeneity: 92-97 Movers

	Logit	Logit	Logit, tract fixed effects	2SLS, tract dummies	OLS tract dummies
Ln(total no. of	.296	.617		.867**	.872**
establishments)	(.196)	(.496)		(.303)	(.160)
Ln(rent/sq ft)	-.0933	-3.91**		-5.35**	-5.31**
	(.633)	(1.82)		(1.04)	(.482)
Stock SU's <	.0198**	-.0449*		-.0490**	-.0677**
250 m.	(.00929)	(.0256)		(.0190)	(.00827)
Stock SU's	.0181**	.0166		-.00297	-.000387
250-500 m.	(.00575)	(.0161)		(.0142)	(.00443)
Stock SU's	.00592**	.0193**		.0139**	.0141**
500-750 m.	(.00253)	(.00677)		(.00495)	(.00191)
Ln(quality)	.259**	.336		-.320	.125
<250m	(.0970)	(.262)		(.271)	(.0841)
Ln(size)*ln(est)		-.0859	-.193		
		(.0895)	(.122)		
Ln(size)*ln(rent)		.748**	1.05**		
		(.313)	(.400)		
Ln(size)*ring 1		.0141**	.0175**		
stock		(.00475)	(.00578)		
Ln(size)*ring 2		-.000430	-.000481		
stock		(.00287)	(.00137)		
Ln(size)*ring 3		-.00247**	-.00220*		
stock		(.00123)	(.00131)		
Ln(size)*ring 1		-.0104	-.0176		
quality		(.0478)	(.0723)		
N [movers]	15580 [95]	13448 [82]	13448 [82]	40	40
*R*sq.		.209	.0774	.828	.826
"Sargan" (p-value)				1.10 (.89)	

For the second to last column, instruments are distance to Rockefeller Center, 1970 ln(total housing units), share of 1970 units in buildings with less than 5 units, 1977 ad agency stock for rings 1-3, 1977 total establishments for rings 1-2, ln(median ad agency payroll size in 1977) for first ring, and 1992 total own tract establishment count.

Figure 1: Locations of Advertising Agencies in Manhattan.
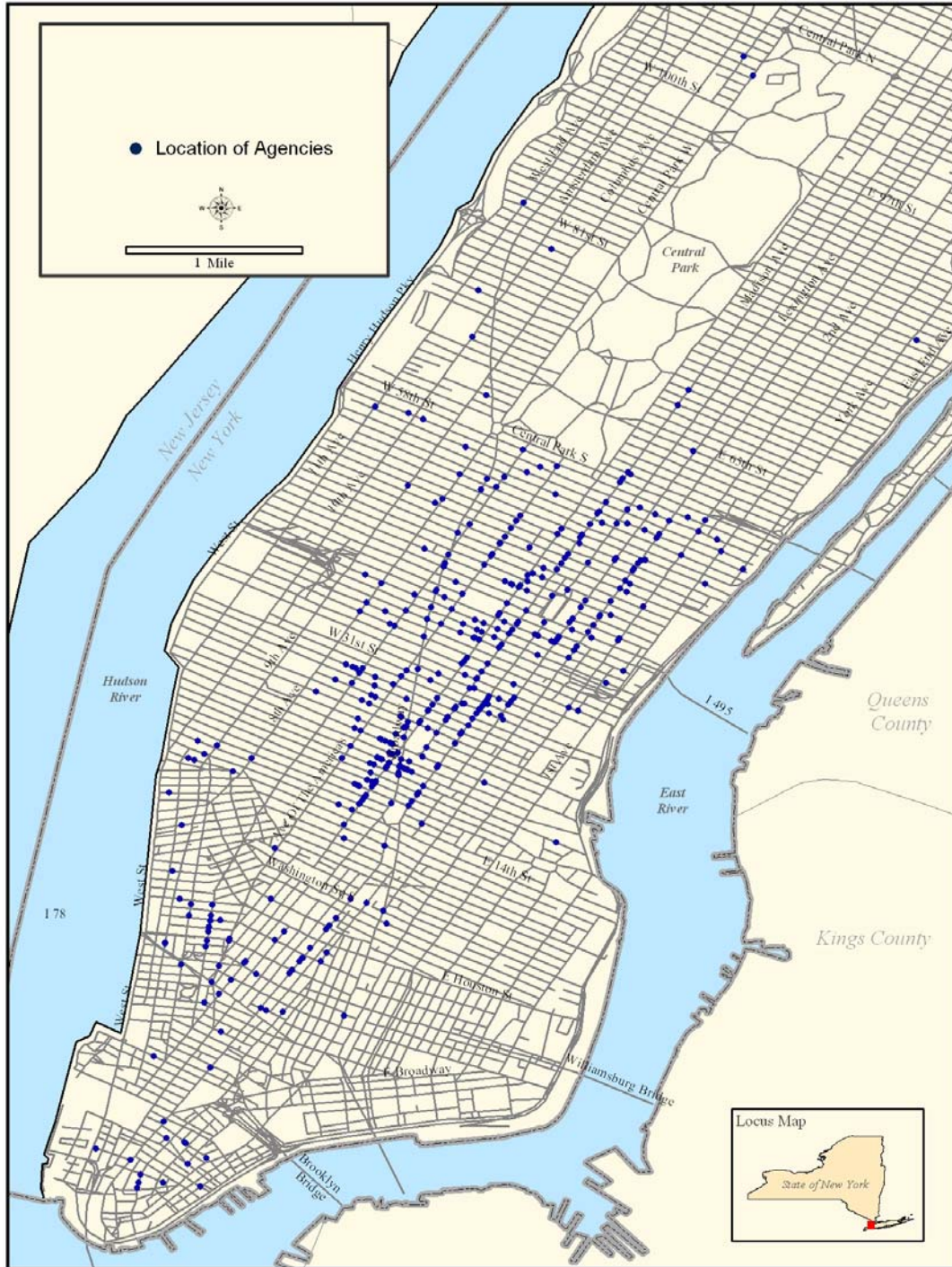
Appendix A. Office Rent Interpolation for NY City

We use the rent for class A office space as an appropriate rent measure for advertising agencies. The data is provided to us by Torto Wheaton Research. It covers 51 ZIP codes in Manhattan in 1992. The data includes the asked price for class A office space in dollar per squared feet. However the data do not provide us with the rent at Census tract level. Thus we interpolate the rent data for tracts in the Southern Manhattan. First, we assign the rent to the centroids of the ZIP codes. Second, we constructed a smoothed rent surface using the spline method (spline estimates values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points). Finally we interpolated the rent for a tract by averaging the value of the rent surface over the tract.

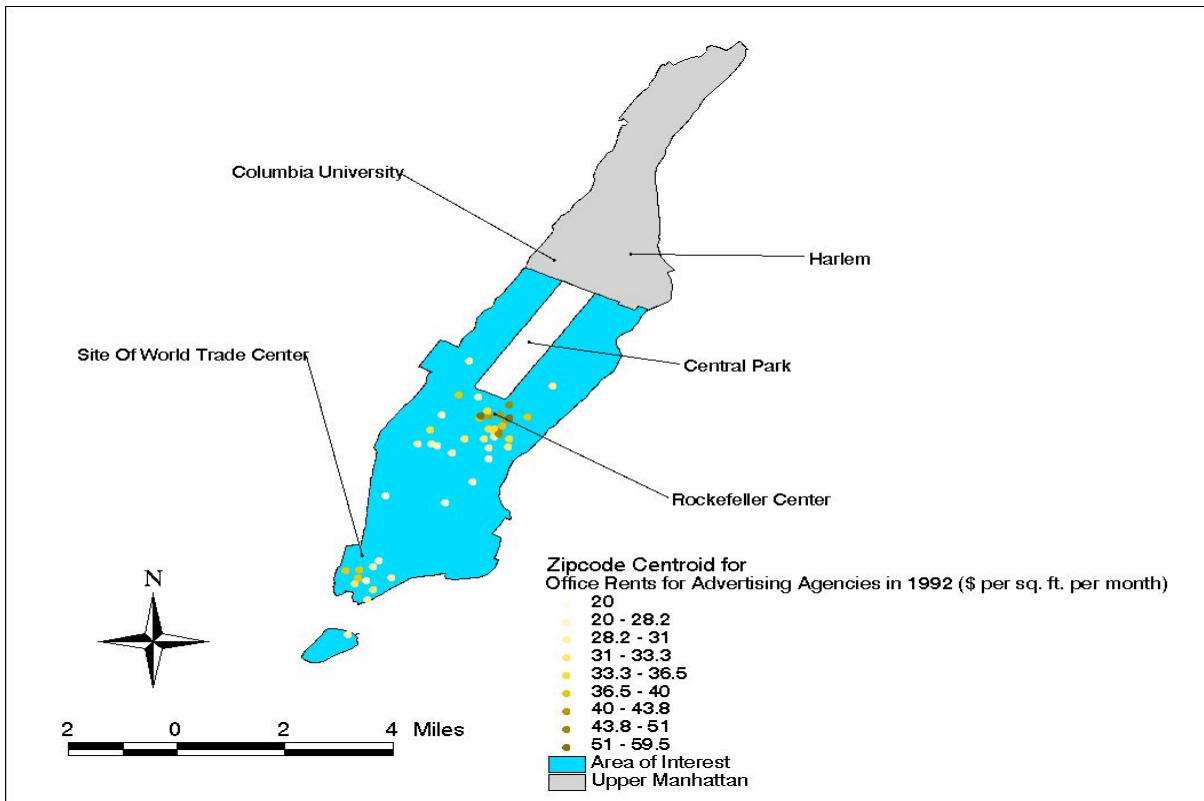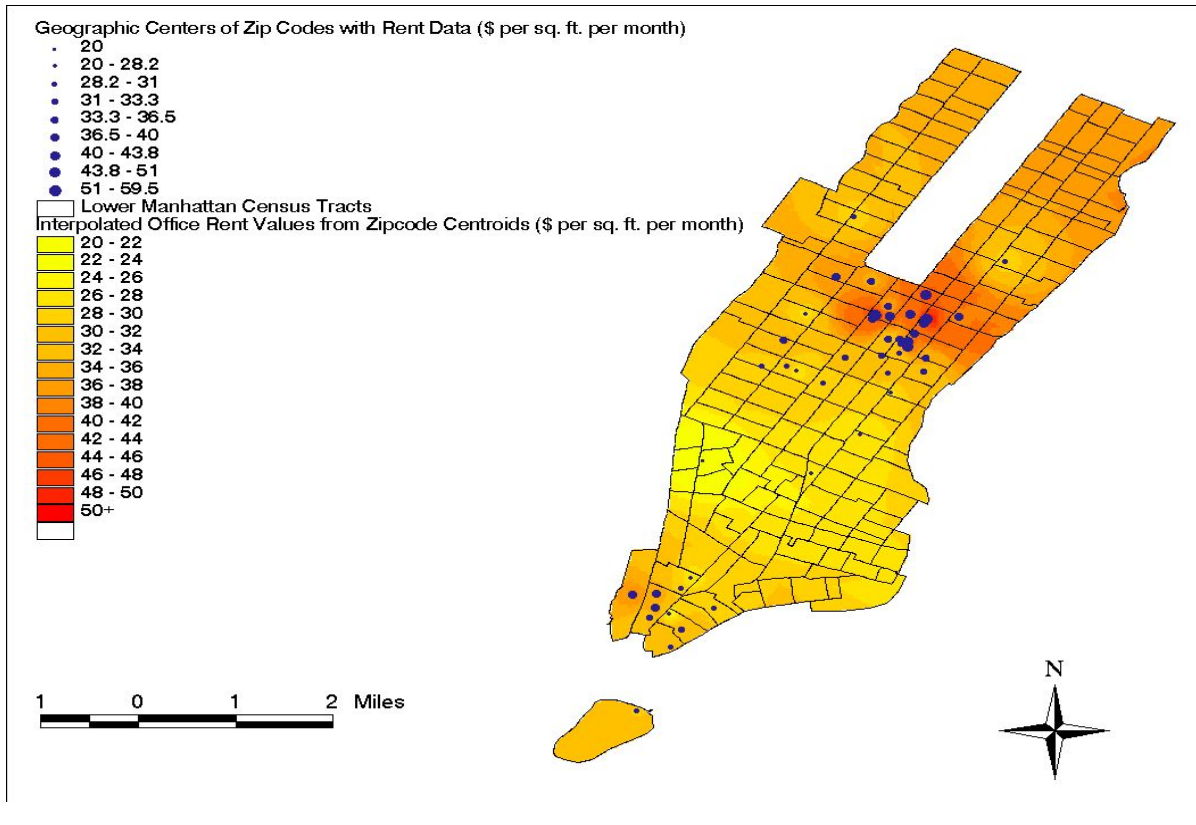**Figure A-1. Class A Office Rents for Lower Manhattan ZIP codes.**

Figure A-2. Class A Office Rent Surface for Lower Manhattan.

Appendix B. Basic Statistics for Covariates and Results for MU's

Means and Standard Deviations

Tract characteristic	Raw mean of tract characteristics	Raw standard deviation	Mean across tracts (weighted by number of births)	Standard deviation (weighted by births)
SU Stock <250m.	5.80	9.74	17.5	15.8
Stock 250-500m.	17.3	20.3	36.8	22.2
Stock 500-750m.	37.4	49.0	91.4	69.0
Stock 750-1000m.	34.3	34.6	52.6	33.4
Stock 1000-1250m	54.3	52.9	96.3	61.4
Ln(total no. establishments)	6.07	1.3	7.1	.88
Ln(rent/sq. ft.)	3.32	.17	3.36	.19
Ln(median payroll,stock of SU's)	3.30	2.73	5.24	1.98
N	164	164	502	502

Results for MU's

	Ordinary Poisson	Ordinary Poisson	IV estimation
Ln(total no. of establishments)	.528**	.576**	.679
	(.230)	(.0767)	(.574)
Ln(rent/sq ft)	-.1.61**	-2.25*	-2.68
	(.807)	(1.28)	(2.98)
Stock MU's < 250 m.	.262**	.339**	.343**
	(.0439)	(.0657)	(.111)
Stock MU's 250-500 m.	.118**	.166**	.0865
	(.0248)	(.0416)	(.0611)
Stock MU's 500-750 m.	.0216	.0307	-.00216
	(.0190)	(.0197)	(.0498)
Stock MU's 750-1000 m.	-.0461*	-.0465*	.00271
	(.0257)	(.0260)	(.0696)
Stock MU's 1000-1250 m.	-.0215	-.0198	-.00187
	(.0197)	(.0211)	(.0406)
Stock SU's <500 m.	-.00219	-.0148	
	(.00147)	(.00988)	
Broadcast Emp.<500 m.		-.0000547	
		(.0000567)	
Constant	.0228	1.97	2.00
	(3.05)	(4.08)	(12.0)
N [births]	164	164	164
Pseudo Rsq [Sargan p-val]	.460	.468	[.206]

Column 1 contains the ordinary Poisson and column 3 the IV Poisson for a baseline specification. There are very strong effects in rings 1 and 2 for ordinary Poissons and in ring 1 for IV results. While for SU's if we increase the number of relevant neighbors in ring 1 by 1 expected births rise by 2%, here for MU's, they rise by 34%. The high coefficients appear to arise from small numbers and an implied non-linearity in scale effects, where the average number of MU's in ring 1 is .8, versus 6 for SU's. In ordinary Poissons, a quadratic for MU's results in significant ring 1 coefficients of .822 and -.0478, so that in moving from 6 to 7 neighbors expected births rise by 2.0%, and positive marginal effects disappear by 9 neighbors. For SU's in a quadratic, if we increase neighbors from their mean number of 6 to 7, expected births rise by 2.8%. Column 2 shows MU's do not benefit from being near SU's. In fact in all formulations we tried such as having 3 or more SU ring variables in the ordinary Poisson or results under IV estimation, we never got positive, even modestly significant results for SU variables.