WHY IT MATTERS WHO YOU KNOW: EVIDENCE ON THE MECHANISMS UNDERLYING JOB REFERRAL

Ian M. Schmutte
Department of Economics
Terry College of Business
University of Georgia
schmutte@uga.edu

October 31, 2012

Abstract

The ubiquity of job referrals suggest that they fulfill an important intermediation function in the labor market, but the nature of that intermediation function remains poorly understood. Part of the problem, which this paper addresses, is that the use of referrals depends on three parties: the referring worker, the referred worker, and the referred employer. Different models of intermediation imply different forms of correlation between referral and unobservable characteristics of the three relevant parties. Using geographically detailed longitudinal matched employer-employee data, I show that referrals are more likely among high ability referrers, high ability referees, and to involve moves to high-paying firms. These findings are consistent with models in which referral networks are used to screen workers on the basis of ability, and where workers use referrals to find better paying jobs. These results are not consistent with local referrals being either a search method of last resort for workers with bad outside opportunities nor with referrals as a substitute for raising wages among firms facing monopsonistic conditions.

JEL Codes: J31, J64, R23.
Keywords: Social Interactions; Informal Hiring Networks; Wage Variation; Neighborhood Effects.

This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau, its program sponsors or data providers. All results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

It is now well-established that referral networks are an important and persistent feature of labor markets. The functions that referrals perform in mediating labor market relationships are less understood. Existing empirical work indicate a complex incidence of referral use by workers and by firms. The relationship between referral use and labor market outcomes is also hard to pin down. Referrals are sometimes associated with increased earnings, and sometimes with decreased earnings(Ioannides 2004; Loury 2006). Sometimes referrals lead to apparently higher productivity, and sometimes lower. Since referrals are just one channel of information in labor markets, it is not surprising that they play different roles in different circumstances. Given the potential for referrals to shape both the efficiency and equity of labor market outcomes, we need a better understanding of why referrals are associated with particular outcomes in particular contexts.

One gap in the existing literature, which this study helps fill, is that it is rare to observe the characteristics of all parties to a referral relationship. At its core, a referral requires three parties: the person making the referral (the *referrer*), the person receiving the referral (the *referee*), and the firm accepting the referral (the *employer*). Using geographically-detailed matched employer-employee data, I document the characteristics of referee, referrer, and employer that increase the probability of a referral. More specifically, I show how the probability of a referral relationship between neighbors is associated with unobservable characteristics that affect earnings.

Referrals presumably relieve some information problem important to the parties involved. Otherwise, the presence or absence of a referral is completely redundant to employment outcomes. Any observed relationship between referrals and earnings depends on the type of information problem that the referral resolves. One theory that has received a great deal of attention in the economics literature is that referrals help firms screen workers on the basis of unobservable skill Simon and Warner (1992); Dustmann et al. (2011). Referred workers should then earn more because there is less uncertainty about their productivity. Another possibility is referrals are simply an additional channel of search for firms trying to hire in a frictional labor market. In that case, referrals are complementary to firm size, which is correlated with firm-specific variation in pay. Referred workers earn more because referral is a proxy for an unobserved firm-specific component of pay.[1] Referrals may also be used by workers with limited outside options (Elliott 2001). If workers use referral when the outcomes of formal job search are weak, then referrals

[1]Mortensen and Vishwanath (1994) develop such a model.

are negatively selected on the characteristics of the referrer and the referee. Furthermore, if referrals are substitutable for wage increases in attracting employees, then in a basic wage-posting model, referrals can be negatively correlated with employer pay.

My methodology is an extension of that used by Bayer et al. (2008). I observe pairs of workers who live in the same neighborhood. When one of those workers changes jobs, I see whether they become their neighbors coworker. I measure the contrast between the probability of becoming a neighbor's coworker when the pair live in the same block and when the pair live in different blocks. This contrast identifies the presence of a social interaction in job search under assumptions that I outline below. Unlike Bayer et al. (2008), I observe which worker was at the job first, so can infer the direction of the implied referral relationship. I use heterogeneity components from an earnings decomposition to measure unobservable characteristics of workers and firms that are correlated with earnings. I then go on to assess whether the contrast that identifies the social interaction effect is stronger when the referred worker is a high-wage worker, or when the firm is a high-wage firm.

I find that referrals are associated with high paying firms, highly paid referrers, and highly paid referees. These findings are consistent with neighborhood-level referrals screening both on the basis of productivity (Montgomery 1991; Simon and Warner 1992; Oyer and Schaefer 2011), but also with referrals directing workers into firms that pay higher wages(Beaman and Magruder 2011; Schmutte 2012). The data thus reject a model in which referrals are a search method of last resort for workers, or the preferred search method for low-paying firms.

At a more basic level, I extend and verify the findings of Bayer et al. (2008), who document the presence of local interactions in job search using cross-sectional data that describe where people work, and who their neighbors are. Bayer et al. (2008) do not directly observe the identity of the employer, can not distinguish the referrer from the referee, and have difficulty addressing the problem of reverse causality. My empirical design is closely related to their design, and as such, this paper will make continued reference to their model and results. Where appropriate, I will indicate where the two analyses are related.[2]

[2]Hellerstein et al. (2011) also use cross-sectional employer-employee matched data very similar in origin to those used by Bayer et al. (2008) to measure the influence of local social interactions in job finding. Their methodology is different, so my findings can not be understood as an extension of their results.

2 Data

The analysis in this paper requires data in which it is possible to identify when two workers are neighbors, when they are coworkers, and changes in these relationships over time. To allow for the possibility that neighbors become coworkers due to correlated job search processes requires observing multiple workers and job changers in the same neighborhood. In other words, we need population-level data that include information on both place of residence and place of work. The LEHD program of the U.S. Census Bureau provides data with this level of detail. The data used in this paper are nearly identical to Schmutte (2012), which includes a comprehensive appendix describing these data for the interested reader.

2.1 Data Sources

The primary source of data for this paper is the Longitudinal Employer-Household Dynamics (LEHD) program of the U.S. Census Bureau. The essential feature of the LEHD data is that they record for each job the identity of both the worker and employer involved. The LEHD Infrastructure files are based on state Unemployment Insurance (UI) records, and are, with some important exceptions that I discuss below, nearly universal in their coverage of private-sector employment in the U.S. between 1990 and 2004. I augment the LEHD Infrastructure with data on place of residence for 2002–2003. From the merged dataset, I select all workers with positive UI earnings in at least one quarter of 2002–2003 who lived in one of thirty large Metropolitan Statistical Areas (MSAs).

The purpose of this paper is to evaluate the extent to which referrals are associated with observable and unobservable determinants of labor market outcomes. To measure unobservable characteristics, I use estimates of worker- and employer-specific earnings components from the Abowd-Kramarz-Margolis (AKM) log earnings decomposition:

$$\ln Y = X\beta + D\theta + F\psi + \varepsilon. \tag{1}$$

This model is estimated on the set of all LEHD work histories for workers aged 18-70. These data cover 30 states between 1990-2003, and include 660 million wage records for 190 million workers and 10 million employers. Y is a vector of annualized earnings on the dominant job, and ε is a statistical residual. D and F are design matrices of the worker and employer effects. X is a matrix of time-varying controls consisting of a quartic in experience, year effects, and the exact within-year pattern of positive earnings. All of these measures are interacted with sex.[3]

The AKM decomposition yields a measure, θ_i, for each worker and ψ_j for each employer, of the average deviation of earnings from their conditional mean. A worker with a high relative value of θ_i is a high-wage worker. An employer with a high relative value of ψ is a high-wage employer. A structural interpretation of these parameter estimates requires the assumption that job mobility is exogenous. Here, since I am using these as control variables in a downstream model, I require a structural interpretation. The exogenous mobility assumption is likely false, but recent research by Abowd and Schmutte (2012), and Card et al. (2012) indicate that the estimated worker- and employer-heterogeneity are not very sensitive to its violation.

2.2 Analysis Sample

To conduct the main analysis, I restrict attention to workers who are employed full-time for the full year, who lived in one of 30 large MSAs, and for whom I can measure the Census block of residence in 2002 and 2003. I further restrict attention to workers who do not change place of residence over the two years. This yields a sample of about 24 million workers.

Of these 24 million workers, I observe 2,206,421 who change employers. My goal is to measure how the probability that these workers become coworkers of their neighbors increases with spatial proximity, and how it increases with unobservable correlates of earnings for the referral seeker, referral provider, and referral taker. The event of beginning to work with someone in the neighborhood is actually rather common. Out of my sample of 2,206,421 job changers, 360,289 become the coworker of someone residing in the same block group. There are 60,280 block groups in the sample.

Analysis proceeds on a sample of matched pairs of workers. A pair of workers (ℓ, m) appears when ℓ changes jobs between 2002 and 2003, and m resides in the same block-group, is employed, and does not change jobs. Following Bayer et al. (2008), I define variables

- $R_{\ell,m}$: an indicator equal to 1 if ℓ and m live on the same Census block and zero otherwise.

[3]This decomposition as applied to matched employer-employee data was first introduced by Abowd et al. (1999) as a means of correcting biases in the estimation of industry and other more aggregated types of wage premia. The estimates used in this paper were conducted as part of the Human Capital Estimates Project within LEHD according to the estimation procedure described in Abowd et al. (2002) and Abowd et al. (2003).

- $W_{\ell,m}$: an indicator equal to 1 if ℓ and m share the same employer in 2003 and zero otherwise.

The sample of matched pairs contains 1,558,436,893 observations.

3 Empirical Method

The analysis takes advantage of time-sequencing in the LEHD data to infer referral events – events where a worker starts a job with the employer of one of his neighbors. I adapt the Bayer et al. (2008) research design to this setting, controlling for block-group level heterogeneity to identify the effect of residential proximity on the likelihood of moving to the same employer as one of your neighbors. I control for block-group specific variation in the propensity for a mover to end up employed in the same firm as someone in her block group. The baseline empirical specification is the linear probability model

$$W_{\ell,m} = \rho_{G(\ell)} + \alpha_0 R_{\ell,m} + \varepsilon_{\ell,m}, \tag{2}$$

which is denoted explicitly to correspond to Equation 1 in Bayer et al. (2008).

The effect, α_0 is the estimated effect of living on the same block on the propensity for a mover to become the coworker of a stayer. This is evidence of some kind of social interaction under the identifying assumption that all spatial variation in job search outcomes occurs at the neighborhood level, so is swept out by the group effect, $\rho_{G(\ell)}$. The essence of this assumption is that workers do not sort within neighborhoods on the basis of employment-relevant characteristics, and that the constraints on job search are the same for workers in the same neighborhood. These assumptions are plausible given the high degree of geographic detail in the data. I also show below that there is very little evidence of sorting within neighborhoods on the basis of observable demographic characteristics.

Next, I extend the specification in Equation (2) to account for the characteristics of the pair. If workers are more or less likely to be involved in referral when they have particular characteristics, for instance if high-skilled workers are more likely to be chosen for referral, then the influence of spatial proximity on the probability of becoming a neighbors coworker will be increasing in skill. This observation motivates the following specification.

$$W_{\ell,m} = \rho_{G(\ell)} + \beta' X_{\ell,m} + (\alpha_0 + \alpha_1' X_{\ell,m}) R_{\ell,m} + \varepsilon_{\ell,m}, \tag{3}$$

I estimate this model including in $X_{\ell,m}$ the race, gender, ethnicity, age and person effect from the AKM decomposition (θ) for both the mover and stayer. I also include the employer wage premium, ψ, of the stayer. The main novelty of this model with respect to the analysis in Bayer et al. (2008) is that the directed nature of the job move allows me to infer who is doing the referring (the stayer) and who is being referred the mover) in these matched pair.

3.1 Identification

The identification of α as a referral effect relies on the correlation in job search outcomes being exogenous to spatial proximity other than through referral. Intuitively, the assumption means that when two workers live on the same block in the same neighborhood, they have the same ex ante probability of becoming coworkers as if they live in different blocks in the same neighborhood. Importantly, this research design allows for workers to sort into neighborhoods (here, block groups). However, it requires that workers not sort within neighborhoods on the basis of characteristics that predict where they are employed.

One economic rationale for this identifying assumption is that individuals are more likely to choose a neighborhood to live in than they are a particular block within that neighborhood. Individuals are attracted to parts of the city by schools, parks, transportation links, and other amenities, and of course by prices. However, within a neighborhood, individuals are restricted to choose among housing units available at the time.

The assumption that workers do not sort within neighborhoods on the basis of characteristics relevant to employment outcomes is untestable. However, Bayer et al. (2008) show that there is very little evidence of sorting within neighborhoods on the basis of observable demographic characteristics for Boston in the Decennial Census. In the companion paper to this one, Schmutte (2012), I show that there is little sorting on the basis of observable characteristics within neighborhoods, and even less sorting on the basis of unobservable characteristics correlated with earnings. I replicate this evidence below in support of my identification strategy.

3.1.1 Sorting across Blocks

Table II reproduces evidence on the extent of sorting within block groups that appears in Schmutte (2012). I measure within-neighborhood sorting as follows:

1. Construct subsample of prime-age male full-time full-year workers in 2002.

2. From each block, draw a random individual, i.

3. For characteristic y, compute block-level mean net of i, $\bar{y}^{-i}$.

4. Estimate

$$y_i = \alpha + \beta \bar{y}^{-i} + \mu_g + \eta_i$$

where μ_g is a block group effect.

5. Table II reports R^2 from this model with and without μ_g.

The column in Table II labeled 'Raw' reports the model without block group controls. Intuitively, if there is no sorting within neighborhoods, the raw estimate should be eliminated by the block group controls. Indeed, sorting is heavily attenuated after introducing block group controls. This is consistent with the assumption that identifies the local interaction effect . Like Bayer et al. (2008), these within-neighborhood sorting measures are not identically zero. However, I also present a rough measure of the amount of sorting on unobservables. The amount of within neighborhood sorting on the AKM residual is zero. This indicates that sorting on unobservable characteristics that influence earnings is less strong than sorting on observable characteristics. Sorting on unobservables will need to be much stronger than sorting on observables than seems likely, given the data evidence, to explain my estimated effects.

4 Results

Directed referral relationships are more likely to occur when the 'referring' worker is employed on a job with a higher wage premium, which I interpret as evidence in support of the mechanism proposed in the theoretical model. I also find that the 'person effect', θ, of the stable and moving worker also both increase the influence of spatial proximity on the likelihood that the mover starts working with the stayer, consistent with other 'screening' models of referral.

The baseline probability that $W_{\ell,m} = 1$ among pairs of workers that do not reside on the same block is 0.0013, or 0.13 percent of such pairs. This probability increases to 0.0016 among pairs of workers that live on the same block. Thus, among workers who change jobs, the probability of being employed with someone from your neighborhood is 23 percent higher when the neighbor lives on the same block. This estimate does not account for any variation in the baseline probability of taking a job in a firm that employs your neighbors, however.[4]

The results from including block-group controls to identify the referral effect appear in Table III. Specification (1) estimates to the basic model in Equation (2). The probability that a job changer moves into a firm employing one of his neighbors increases by a very precisely-estimated 0.0002, reported in the table as .02 percentage points, when they live on the same block. This corresponds to an 18 percent increase over the baseline probability of 0.0013.[5] Under the identifying assumptions, this is evidence of direct social interactions in job search.

I turn next to the main results of the paper, which take advantage of the fact that I know which of the two workers involved was in a job first. Specification (2) presents estimates of Equation 3, which measure heterogeneity in the influence of spatial proximity associated with characteristics of the job changer, characteristics of the stayer, and characteristics of the employer. My presentation focuses on variables associated with unobservable characteristics that are correlated with earnings. Specifically, the estimates of the person effect and firm effect from the AKM decomposition. If referral are more actively sought by workers with few outside options, then the interaction between R and θ should be negative. If referrals are more commonly accepted or used by low-paying firms, then the interaction between R and ψ_{ref} should be negative.

What we observe is the opposite. The referral effect is stronger when the worker changing jobs is a 'high-wage worker', when the worker staying is a 'high-wage worker', and when the employer involved is a 'high-wage firm'. Evaluated at the sample means, the estimated referral effect is 0.032. The probability of becoming a coworker of your block-level neighbor increases by 0.012 pp with a one standard deviation increase in θ_{ref}. This is a 60 percent increase over the raw social interaction effect. A one standard deviation increase in ψ_{ref} is associated with a 0.008 pp increase in the referral effect, or a 40 percent over the raw effect. Finally, a one standard deviation increase in the person effect of the changer, θ, is associated with a 0.009 pp increase in the referral effect; roughly a 50

[4]Note that the baseline estimate is much smaller than the baseline estimate in Bayer et al. (2008). Among pairs of individuals who live in the same block group, they find 0.36 percent work in the same block. This increases to 0.94 percent if the pair reside on the same block. They find that most of this increase is due to a mechanical correlation between block size and the probability that any two people live on the same block. Because their analysis counts all pairs of people living in the same neighborhood, doubling block size will quadruple the number of pairs. The larger block has a larger baseline probability that any two people work together, so the total effect is a large upward bias.

[5]Bayer et al. (2008) find a roughly 33 percent increase over the baseline in their most conservative specification. While different in magnitude, these results should be interpreted as supportive of the basic Bayer et al. (2008) findings.

percent increase over the raw effect.

5 Conclusion

The results in this paper suggest a range of further avenues for investigating referral relationships in longitudinal matched employer-employee data. While the presence and importance of referrals in job search are well-established at this point, the reasons referrals are used are still very poorly understood. I have documented that referral-use among workers in urban labor markets appears to be associated with selection of high-wage workers and the selection of high-wage employers. These findings are inconsistent with models in which referrals are used as a search method of last resort for workers with few outside options (Loury 2006). They are also not consistent with a model in which low-paying firms use referrals as a cost-minimizing strategy to attract workers.

These findings are consistent with many different models of referral use. The selection of high-wage workers is consistent with referrals as a form of screening on the part of firms for unobservable productive characteristics of workers. The selection of high-wage firms is consistent with referral use as a hiring strategy of firms paying efficiency wages (Kugler 2003) but also consistent with opportunism in job-search among workers trying to find a higher return to their characteristics (Schmutte 2012).

These findings may well be due to the particular nature of the sample. Having restricted the sample to workers who are already employed full-time for the full year, the data likely over-represent workers with good outside options, and therefore those for whom referrals are used by choice rather than by necessity. Other research using LEHD data indicate that part of the earnings gap between native and non-native workers is due to negative selection of non-native workers into lower-paying firms. We might expect then that when focusing on non-native workers who are marginally attached, the effect associated with referral use may be negative rather than positive. Evaluating the data for evidence consistent with such predictions is an objective of future work with these data.

References

Abowd, J. M., Creecy, R. H. and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data, *Technical Report TP-2002-06*, LEHD, U.S. Census Bureau.

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999). High wage workers and high wage firms, *Econometrica* **67**(2): 251–333.

Abowd, J. M., Lengermann, P. and McKinney, K. L. (2003). The measurement of human capital in the U.S. economy, *Technical Report TP-2002-09*, LEHD U.S. Census Bureau.

Bayer, P., Ross, S. L. and Topa, G. (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes, *Journal of Political Economy* **116**: 1150–1194.

Beaman, L. and Magruder, J. (2011). Who gets the job referral? Evidence from a social networks experiment, *American Economic Review* .

Dustmann, C., Glitz, A. and Schnberg, U. (2011). Referral-based job search networks, *IZA Discussion Papers 5777*, Institute for the Study of Labor (IZA).

Elliott, J. R. (2001). Referral hiring and ethnically homogeneous jobs: How prevalent is the connection and for whom?, *Social Science Research* **30**(3): 401 – 425.

Hellerstein, J. K., McInerney, M. and Neumark, D. (2011). Neighbors and coworkers: The importance of residential labor market networks, *Journal of Labor Economics* **29**(4): 659 – 695.

Ioannides, Y. M. (2004). Neighborhood income distribution, *Journal of Urban Economics* **56**: 435–457.

Kugler, A. D. (2003). Employee referrals and efficiency wages, *Labour Economics* **10**: 531–556. Secondary.

Loury, L. D. (2006). Some contacts are more equal than others: Informal networks, job tenure, and wages, *Journal of Labor Economics* **24**: 299–318.

Montgomery, J. D. (1991). Social networks and labor market outcomes: Toward an economic analysis, *American Economic Review* **81**: 1408.

Mortensen, D. and Vishwanath, T. (1994). Personal contacts and earnings: It is who you know!, *Labour Economics* **1**: 187–201.

Oyer, P. and Schaefer, S. (2011). Firm/Employee matching: An industry study of american lawyers. November 2010.

Schmutte, I. (2012). Job referral networks and the determination of earnings in local labor markets. University of Georgia Department of Economics.

Simon, C. J. and Warner, J. T. (1992). Matchmaker, matchmaker: The effect of old boy networks on job match quality, earnings, and tenure, *Journal of Labor Economics* **10**(3): 306–30.

Tables

Table I: Descriptive Statistics

Variable	Mean	Std. Dev
White	0.65	(.477)
Black	0.11	(.316)
Hispanic Origin	0.13	(.333)
Male	0.58	(.492)
Born in U.S.	0.81	(.388)
Age in 2002	37.10	(11.213)
Real Earnings in 2003	37422.62	(85819.14)
$\exp(\theta)$	1.44	(1.653)
$\exp(\psi_{2002})$	1.30	(6.581)
$\exp(\psi_{2003})$	1.35	(1.839)
N	$N = 2,206,421$	

Summary statistics for a sample of workers with reported UI earnings in one of 30 large MSAs between 2002 and 2003. The sample is restricted to workers who did not move MSAs during 2002-2003, were at least 14 years of age in 2002, and had valid data for block of residence in 2002 and 2003. The data are restricted to workers who change their dominant employer between 2002–2003, and who lived on blocks where at least 10 other workers contribute data to compute the block-level average ψ.

Table II: Sorting within neighborhoods, R^2 method.

Variable	Raw	Block Group Controls
White	.292	.013
Hispanic	.286	.013
Born U.S.	.225	.011
Age	.030	.007
θ	.160	.016
ψ	.055	.001
ε	.004	.000
$N = 394,305$		

Measures of sorting within Census block groups. The input dataset contains one individual-level observation per block and the fraction of people (not including the individual) in the block who share the listed characteristic, or its average. Each entry is the R-squared from a regression of the individuals characteristic on the block-level average. The second column controls for block group specific effects. The sample is restricted to blocks with more than six individuals.

Table III: Estimates of Direct Referral Effects: With and Without Covariates.

		No Covariates (1)		Pair Covariates (2)	
	Variable	Coeff.	t-Stat.	Coeff.	t-Stat.
Reside on same block	R	0.02	27.81	0.013	36.10
Wage premium of non-changer	$R \times \psi_{ref}$			0.014	63.78
Wage effect of non-changer	$R \times \theta_{ref}$			0.023	46.73
Wage effect of changer	$R \times \theta$			0.018	33.47
Block Group Effects		YES		YES	
Sample Size		$1,558,436,893$			

The table reports results for two regressions in which an observation is a pair of adults employed between 2002–2003 who live in the same Census block group. The pairs are drawn from 30 large CBSAs described in the text. The first worker in the pair (changer) is observed to change jobs between 2002–2003. The second worker in the pair (non-changer) did not. The sample includes all such pairs within each block group, restricted to block groups with at least ten non-changers. The coefficients have been multiplied by 100 to reflect percentage point changes. Block group heterogeneity is estimated by fixed-effects in both models. In addition to the variables listed, specification (2) also includes controls for race, ethnicity, age, AKM person effect (θ) and AKM firm effect (ψ) for both the changer and non-changer members of the pair.