Hiring through referrals

Manolis Galenianos

Royal Holloway, University of London, United Kingdom

Received 8 October 2012; final version received 19 December 2013; accepted 27 February 2014

Available online 27 March 2014

Abstract

An equilibrium search model of the labor market is combined with a social network. The key features are that the workers’ network transmits information about jobs and that wages and firm entry are determined endogenously. Empirically, the inter-industry variation in aggregate matching efficiency is attributed to variation in referral use. The model predicts that the efficiency of the aggregate matching function is pro-cyclical which is consistent with empirical evidence.

© 2014 Elsevier Inc. All rights reserved.

JEL classification: J0

Keywords: Search; Referrals; Social networks

1. Introduction

Social networks are an important feature of labor markets (Granovetter [17]). Approximately half of all American workers report learning about their job through their social network (friends, acquaintances, relatives, etc.) and a similar proportion of employers report using the social

http://dx.doi.org/10.1016/j.jet.2014.03.009

0022-0531/© 2014 Elsevier Inc. All rights reserved.
networks of their current employees when hiring (the evidence is summarized in Section 2 and is surveyed in Ioannides and Loury [21], and Topa [32]).

Surprisingly, however, social networks are typically not included in the equilibrium models that are used to study labor markets. For instance, in their extensive survey of search-theoretic models of the labor market, Rogerson, Shimer and Wright [29] do not cite any papers that include social networks or referrals. Social networks have been extensively studied using graph theory (Jackson [22]). When applied to labor markets, however, these models usually restrict attention to partial equilibrium analyses where, for instance, wages or labor demand are exogenous (e.g. Calvo-Armengol and Jackson [7]).

The present paper proposes to bridge this gap by combining an equilibrium search model with a network structure that is simple enough to preserve tractability but also rich enough to deliver several predictions that can be confronted with the data.

In the model, workers are homogeneous in terms of their productivity and network. Each worker is linked with a measure of other workers and the network is exogenous. Vacancies are created both through the free entry of new firms and through the expansion of producing firms. A firm and a worker meet either through search in the frictional market, which is described by a standard matching function, or through a referral, which occurs when a producing firm expands and asks its current employee to refer a link. The flow surplus of a worker–firm match is equal to output plus the value of the referrals and the wage is determined through Nash bargaining.

The structure of the model is used to examine the large inter-industry differentials in aggregate matching efficiency documented in Davis, Faberman and Haltiwanger [12] and Sahin, Song, Topa and Violante [30]. This variation is decomposed into the two channels of matching, referrals and the market, whose relative importance is determined by the rate of referral generation and the efficiency of the market matching function, respectively.

A property of the model is that a higher rate of generating referrals increases both aggregate matching efficiency and the proportion of jobs found through a referral. However, a higher efficiency of the market matching function increases aggregate matching efficiency but reduces the prevalence of referrals. Therefore, the source of inter-industry variation in the speed of matching can be determined by examining the empirical correlation between referral prevalence and aggregate matching efficiency. To evaluate the above, a standard aggregate matching function (i.e. one where referrals are not explicitly modeled) is estimated and the rates of referral prevalence are calculated for every major industry. It turns out that the estimates for the aggregate matching efficiency are positively and significantly correlated with the prevalence of referrals across industries. According to the model, therefore, the variation in referral prevalence is the proximate cause for the observed inter-industry variation in aggregate matching efficiency. As a result, examining why referral use varies across industries is informative about the nature of labor market frictions.

The next property of the model is that an increase in the unemployment rate reduces the flow of referrals, in addition to increasing congestion among the unemployed. This leads to the prediction that the job finding rate is a decreasing function of the unemployment rate conditional on labor market tightness; in other words, aggregate matching efficiency is pro-cyclical. This prediction

---

2 Calvo-Armengol and Zenou [8], Fontaine [13] and Galeotti and Merlino [15] are exceptions and are discussed below.

3 However, the distribution of firm sizes is degenerate: each firm hires one worker and vacancies created through an expansion are immediately sold off.

4 Data from the National Longitudinal Survey of Youth, the Job Openings and Turnover Survey and the Current Population Survey is used. See Section 4 for the full data description.
is consistent with the findings of several recent papers that examine the cyclical properties of matching efficiency. Barnichon and Figura [1] find that the residual of the matching function moves pro-cyclically and Cheremukhin and Restrepo [10] find that matching efficiency falls in the aftermath of recessions. On the firm side, Davis, Faberman and Haltiwanger [12] find that a vacancy’s hire yield is lower in weak labor markets.

Calvo-Armengol and Zenou [8] and Fontaine [13] incorporate a finite network in an equilibrium search model. In both models workers search regardless of employment status and, if employed, forward job information to an unemployed member of their network. Neither paper focuses on the efficiency of the aggregate matching function which is an important feature of the present study. In both papers, however, networks induce persistence in unemployment similar to this paper’s cyclical property. In a similar framework, Galeotti and Merlino [15] show that the intensity of referral use is non-monotonic in the separation rate, which is supported in the data, in a model with endogenous network formation and exogenous wages. Mortensen and Vishwanath [24] consider a search market where workers learn about jobs through their contacts as well as the market and find that the former yield higher wages. However, they assume that the arrival rates of job opportunities are exogenous which distinguishes their work from the present paper.

2. Evidence about social networks and labor markets

Numerous studies in economics and sociology have documented the following salient facts about the interaction between social networks and labor markets.  

First, both workers and firms use referrals extensively when searching for a job or trying to fill a vacancy, respectively. More than 85% of workers use informal contacts when searching for a job according to the National Longitudinal Survey of Youth (NLSY) (Holzer [19]). In terms of outcomes, more than 50% of all workers found their job through their social network according to data from the Panel Study of Income Dynamics (Corcoran, Datcher and Duncan [11]) while the 24 studies surveyed by Bewley [4] put that figure between 30% and 60%. In most European countries 25–45% of workers report finding their jobs through referrals according to data from the European Community Household Panel (Pellizzari [26]).

On the firm side, between 37% and 53% of employers use the social networks of their current employees to advertise job vacancies according to data from the National Organizations Survey (Mardsen [23]) and the Employment Opportunity Pilot Project (EOPP) (Holzer [18]). According to the EOPP 36% of firms filled their last opening through a referral (Holzer [18]). Furthermore, using EOPP data Holzer [18] and Blau and Robins [5] find that referrals have a greater “hire yield” for firms than searching in the market.

Second, increasing access to referrals increases a worker’s job finding rate. Using census data Bayer, Ross and Topa [2] find that increasing a male individual’s access to social networks by one standard deviation raises his labor force participation by 3.3 percentage points and his hours worked by 1.8 hours after controlling for selection. A higher employment rate of the network’s members also increases access to referrals and employment rates. Topa [31] finds strong evidence of local spillovers in employment rates across different census tracks in the Chicago area. Weinberg, Reagan and Yankow [33] find that improving a neighborhood’s social characteristics by one standard deviation increases annual hours by 6.1% using NLSY data. Using data from the

---

5 See Section 4 for the data that is used to estimate the matching function parameters and the prevalence of referrals across industries.

6 Denmark, Finland and the Netherlands are the only countries, out of a sample of 14, where that ratio falls below 25%.
British Household Panel Survey Cappellari and Tatsiramos [9] show that an additional employed friend is associated with an increase in the probability of finding a job of 3.7 percentage points and a 5% increase in wages.

Third, the social ties that are most useful for transmitting information about job opportunities are a person’s more numerous “weak” ties, e.g. acquaintances, as opposed to his “strong” ties, such as close friends (Granovetter [16,17]).

3. The labor market

This section adds referrals to a standard equilibrium search model of the labor market.

3.1. The model

Time runs continuously, the horizon is infinite and the labor market is in steady state. There is free entry of firms and each firm hires one worker, is risk-neutral and maximizes expected discounted profits using discount rate $r > 0$. A firm is either filled and producing or vacant and searching and the flow profit when searching is $-k$.

There is a unit measure of workers who are homogeneous, risk-neutral, maximize expected discounted utility and discount the future at the same rate $r$. A worker is either employed or unemployed and the flow utility of unemployment is $b$. Every worker is linked with a measure $\nu$ of other workers, where $\nu \leq 1$.

Modeling a worker’s network as a continuum of links is consistent with the (spirit of the) sociology literature’s finding that it is a person’s more numerous weak ties that help most with finding a job (Granovetter [16,17]) and is crucial for the model’s tractability. A worker’s employment opportunities will in general depend on how many of his links are employed which necessitates keeping track of each link’s time-varying employment status. Having a continuum of links means that the aggregate (un)employment rate of a worker’s social contacts does not change over time due to the law of large numbers, thereby greatly simplifying the analysis.\footnote{In finite models additional assumptions are needed to preserve tractability. In Calvo-Armengol and Zenou [8] each worker is assumed to draw a new network every period so as to avoid keeping track of transitions in the network’s employment rate. In Fontaine [13] each worker belongs to one of a large number of disjoint finite networks and the focus is on the steady state distribution of employment rates across the population of networks.}

It is certainly true that some of the richness in the predictions that graph-theoretic models of networks can generate, especially with respect to network architecture, is lost by the assumption of a continuum of links.\footnote{Note, however, that network heterogeneity is not one of them: see Igarashi [20].} However, many of these additional predictions would be difficult to empirically verify or refute given the data that is currently available.

Vacancy creation occurs in two ways: a new firm enters the market or an existing firm expands which occurs at exogenous rate $\rho$. The position that is created by the expansion is immediately sold off which keeps firms’ employment at one. A firm and a worker meet either through search in the market or through a referral, which occurs when a firm expands and asks its current employee to refer a link. The rate of meeting through the market is determined by a matching function. The rate of meeting through referrals is determined by the rate at which firms expand.\footnote{Extending this model to non-degenerate firm size distribution would provide large firms with more opportunities to elicit a referral. However, large firms use referrals to a lesser extent than small firms (Holzer [18], Pellizzari [26]). Therefore a simple extension does not suffice to study the interaction between firm size and referral use. The interaction...}
An expansion can be interpreted in (at least) a couple of different ways. At rate $\rho$, the firm meets an entrepreneur who wants to enter the market at which point the firm expands and sells him the new position (that entrepreneur would otherwise create a new firm through free entry). Alternatively, at rate $\rho$ the firm identifies a business opportunity and expands to take advantage of it. However, it is subject to decreasing returns and finds it profitable to sell the new position to some new entrepreneur. Either way, the new position has zero value for the expanding firm, leading to a sale of that position to a new firm. For this paper’s purposes it makes little difference which interpretation is adopted.

When an expansion occurs, one of the links of the incumbent worker is contacted at random. If the link is employed then the referral opportunity is lost and search in the market begins; if the link is unemployed then he is hired by the firm. In other words, creating a vacancy through an expansion might lead to an immediate hire while a new firm’s entry is necessarily followed by time-consuming search in the market. This is consistent with the findings in Holzer [18] and Blau and Robins [5] that using referrals as a method of searching for workers exhibits a greater “hire yield” than the alternatives.\(^{10}\)

The assumption that the referrer contacts one of his links at random regardless of that link’s employment status captures the frictions which are present when the referral channel is used. One interpretation is that, consistent with the weak ties view of the network, the referring employee does not know which of his links is currently looking for a job and starts contacting them at random to find out if they are interested in the job. Because this is costly, he will only try a finite number of times and with positive probability will fail to find someone interested in the job. In this paper it is assumed for simplicity that the referring worker stops after a single try but allowing for further tries can be easily accommodated by appropriately modifying the referral function below.

The flow value of a match is given by the value of the worker’s output, $y$, and the value of the referrals that he generates. The worker and the firm split the surplus according to the Nash bargaining solution and all payoff-relevant information, including the worker’s network, are assumed to be common knowledge within the match. The worker’s bargaining power is denoted by $\beta \in (0, 1)$. Matches are exogenously destroyed at rate $\delta$, where $\delta > \rho$.\(^{11}\) There is no on the job search.\(^{12}\) Finally, to avoid trivial outcomes, it is assumed that $y > b$.

Denote the expected surplus generated during an expansion by $E$. When a firm expands, it creates a vacancy whose value is denoted by $V$. The incumbent worker contacts one of his links and a match is created if that worker is unemployed, the probability of which is denoted by $u$. Denoting the firm’s value of a match by $J$:  

$$E = V + u(J - V).$$  \hspace{1cm} (1)
The new position is immediately sold off and the incumbent firm receives share $\gamma \in [0, 1]$ of that surplus (the remaining $(1 - \gamma)E$ is captured by the buyer). Therefore a match’s flow value is given by $y + \rho \gamma E$.\footnote{The implicit assumption is that referrals will be used whenever the opportunity arises: a producing firm will expand at rate $\rho$ and it will ask for a referral from its current employee. Alternatively, one can model the decision of whether to expand and/or ask the current employee for a referral as a decision to be jointly taken by the firm and the worker. Since the use of referrals increases flow surplus by $\rho \gamma E$ and this gain is shared by the worker and the firm they will endogenously choose to do so.}

Consider worker $j$ who is linked with $v^j$ workers, each of whom is in turn linked with $v$ workers. The number of employed links of worker $j$ is equal to $(1 - u)v^j$. The employer of each link expands at rate $\rho$ in which case one of the incumbent employee’s $v$ links receives the referral at random. Therefore, the rate at which worker $j$ is referred to a job is $\alpha_R^j = \rho (1 - u)/v$. The network’s homogeneity ($v^j = v$, $\forall j$) implies

$$\alpha_R = \rho (1 - u).$$

Note that the network’s size does not affect the equilibrium.\footnote{The size of one’s network does play a role in the case of heterogeneous network sizes under this specification for the arrival rate of referrals: imagine that half the workers are “insiders” with $v_I = 1$ (i.e. they are linked to everyone) while the other half are “outsiders” with $v_O = 1/2$ (and, for consistency, they are only linked to insiders). It is clear that $\alpha_{RI} > \alpha_{RO}$ and insiders have better job prospects.} \footnote{In the finite model of Calvo-Armengol [6] the arrival rate of referrals is non-monotonic in the size of the symmetric network due to small number properties. In the finite-network model of Galeotti and Merlino [15] a worker’s network size depends endogenously on the separation rate which leads to a non-monotonic relationship between the arrival rate of referrals and the unemployment rate.}

Consider the rate of meeting in the market and let $v$ denote the number of vacancies. The flow of meetings in the market between a vacancy and a worker is determined by a Cobb–Douglas function

$$m(v, u) = \mu v^n u^{1-n},$$

where $\mu > 0$ and $\eta \in (0, 1)$.

The rate at which a firm meets with a worker through the market is

$$\alpha_F = \frac{m(v, u)}{v} = \mu \left( \frac{u}{v} \right)^{1-\eta},$$

and the rate at which a worker meets a firm through the market is

$$\alpha_M = \frac{m(v, u)}{u} = \mu \left( \frac{v}{u} \right)^{\eta}.$$ 

The aggregate matching function, which includes both meetings through referrals and meetings through the market, is given by

$$M(v, u) = \mu v^n u^{1-n} + \rho u (1 - u).$$

(2)

The second term is derived by noting that when the number of producing firms is $1 - u$, the rate of vacancy creation through expansion is equal to $\rho (1 - u)$ and each referral leads to a new match with probability $u$. Notice that the number of matches through referrals is a non-monotonic function of the unemployment rate.
The steady state condition is that the flows in and out of unemployment are equal:

$$u(\alpha_M + \alpha_R) = (1 - u)\delta.$$  

(3)

The steady state condition is that the flows in and out of unemployment are equal:

$$u(\alpha_M + \alpha_R) = (1 - u)\delta.$$  

(3)

The agents’ value functions are now described. When vacant, a firm incurs flow cost $k$, searches in the market and meets with a worker at rate $\alpha_F$. When producing, the firm’s flow payoffs are $y + \rho \gamma E - w$ where $w$ denotes the wage. The match is destroyed at rate $\delta$. The firm’s value of a vacancy ($V$) and production ($J$) are given by:

$$rV = -k + \alpha_F (J - V),$$

$$rJ = y + \rho \gamma E - w + \delta(V - J).$$

When unemployed, a worker’s flow utility is $b$ and job opportunities appear at rate $\alpha_M + \alpha_R$. When employed, the worker’s flow utility is equal to the wage and the match is destroyed at rate $\delta$. The worker’s value of unemployment ($U$) and employment ($W$) are given by:

$$rU = b + (\alpha_M + \alpha_R)(W - U),$$

$$rW = w + \delta(U - W).$$

The wage solves the Nash bargaining problem

$$w = \arg \max_w (W - U)^\beta (J - V)^{1-\beta}.$$  

(4)

The equilibrium is now defined.

**Definition 3.1.** An equilibrium is the steady state level of unemployment $u$ and the number of vacancies $v$ such that:

- The labor market is in steady state as described in (3).
- The surplus is split according to (4).
- There is free entry of firms: $V = 0$.

### 3.2. Existence and uniqueness of equilibrium

The characterization of equilibrium is fairly standard.

The condition that describes the steady state can be rewritten as follows:

$$u \left[ \mu \left( \frac{v}{u} \right)^{\eta} + \rho (1 - u) \right] = (1 - u)\delta \quad \Rightarrow \quad v = \left[ \frac{1 - u}{\mu} \left( \frac{\delta}{u^{1-\eta} - \rho u^\eta} \right) \right]^{1/\eta}.$$  

(5)

Eq. (5) shows that the steady state rate of unemployment is uniquely determined given $v$ and it is strictly decreasing in $v$.\footnote{It is more convenient mathematically to write $v$ as a function of $u$, although conceptually the measure of unemployed workers is the dependent variable (determined through the steady state condition for a given $v$) and the measure of vacancies is the independent variable (to be eventually determined through free entry). The assumption $\delta > \rho$ is sufficient for the steady state to be unique.} As a result, in steady state $\alpha_M$ and $\alpha_R$ are strictly increasing in $v$ while $\alpha_F$ is strictly decreasing in $v$.

The surplus of a match is given by $S = W + J - U - V$. Nash bargaining implies...
\( W - U = \beta S, \)
\( J - V = (1 - \beta)S. \)

The value functions can be rearranged to yield
\[
(r + \delta)S = y + \rho \gamma E - b - (\alpha_M + \alpha_R)\beta S - r V. 
\]

Combining Eq. (6) with the definition of \( E \) from Eq. (1) and the free entry condition \( (V = 0) \) and going through some algebra yields an expression for match surplus that only depends on the number of vacancies in the market (recall that \( u \) is a function of \( v \)):
\[
S = \frac{y - b}{r + \delta + (\alpha_M + \alpha_R)\beta - \rho \gamma u(1 - \beta)}. 
\]

The denominator of the right-hand side is strictly increasing in \( v \) which means that \( dS/dv < 0 \) when the steady state and free entry conditions hold. The value of a vacancy is
\[
rV = -k + \alpha_F (1 - \beta)S. \tag{8} \]

The right-hand side of (8) is strictly decreasing in the measure of vacancies since \( \alpha_F \) and \( S \) are both strictly decreasing in \( v \). To prove the existence of equilibrium, the only remaining step is to show that the right-hand side is positive when \( v \) is small and negative when it is large.

The proposition summarizes the previous statements and completes the proof:

**Proposition 3.1.** An equilibrium exists and it is unique.

**Proof.** It will prove convenient to use the steady state equation and substitute out the number of vacancies. Rearranging Eq. (5) yields:
\[
\frac{v}{u} = \left(1 - u \frac{\delta}{u - \rho}\right)^{\frac{1}{\eta}}. 
\]

Introduce this expression and Eq. (7) into Eq. (8), set \( V = 0 \) and rearrange to arrive at:
\[
\frac{(y - b)(1 - \beta)\mu}{k} = \left(1 - u \left(\frac{\delta}{u - \rho}\right)^{\frac{1}{\eta}}\right)^{\frac{1 - \eta}{\eta}} \left( r + (1 - \beta)\delta + \frac{\delta \beta}{u} - \rho \gamma (1 - \beta)u \right). \tag{9} 
\]

Eq. (9) determines the level of unemployment that satisfies both the steady state and free entry conditions as a function of the model’s parameters.

Denote the right-hand side of Eq. (9) by \( Q(u) \) and note that \( \lim_{u \to 0} Q(u) = +\infty, Q'(u) < 0 \) and \( Q(1) = 0 \). As a result there is a unique level of unemployment such that Eq. (9) holds and it satisfies the steady state and free entry conditions. Introducing that level of unemployment in Eq. (5) delivers the measure of vacancies that are consistent with equilibrium.

3.3. Welfare properties of equilibrium

This section sets up the planner’s problem and shows that constrained efficiency is achieved only for a non-generic set of parameters.

The constrained efficient allocation is given by the solution to the planner’s problem. The planner chooses how many vacancies to create and is subject to the constraint that, conditional on
the number of searchers, the flow from the unemployment to the employment state is identical to that of the decentralized market: the speed of matching in the market is determined by $\mu v^{\eta} u^{1-\eta}$ while producing firms expand at exogenous rate $\rho$ and hire immediately with probability $u$.

The former restriction is standard in search and matching models (e.g. Pissarides [28]); the latter restriction extends the notion of constrained efficiency to include the referral channel. Notice that if the planner could speed up expansions by increasing $\rho$ then he would be able to eliminate frictions which runs counter to the spirit of constrained efficiency.

The planner chooses $u$ and $v$ to maximize:

$$W(u, v) = \int_0^\infty e^{-rt} \left( y(1-u) + bu - kv \right) dt$$

subject to the constraint:

$$\dot{u} = \delta (1-u) - \mu v^{\eta} u^{1-\eta} - (1-u)u\rho$$

Attention is focused to the steady state where $\dot{u} = 0$.

**Proposition 3.2.** We have:

1. The equilibrium is generically inefficient.
2. Constrained efficiency is attained for an appropriate choice of the two bargaining parameters $\beta$ and $\gamma$.
3. The value of $\beta$ that leads to constrained efficiency does not generically equal the elasticity of the market matching function.

**Proof.** See Appendix A. \qed

4. **Theoretical properties and empirical evidence**

The model’s properties are developed and compared with the data.

4.1. **Cross-sectional matching efficiency**

The flow of matches between workers and firms varies significantly across industries. More importantly, when seen through the lens of a standard matching function it varies more than the number of searchers across industries, as documented in Davis, Faberman and Haltiwanger [12] and Sahin, Song, Topa and Violante [30]. This observation is typically rationalized by allowing the efficiency parameter of the aggregate matching function to vary across industries.\(^{17}\)

In the present model matches are formed through two distinct channels: referrals and the market. The model’s structure is used to determine whether the variation in aggregate matching efficiency across industries is due to variation in the rate of referral generation or variation in other factors, which are grouped in the market matching function. This decomposition is useful in analyzing the frictions that are captured by the “black box” of the aggregate matching function.

Specifically, if variation in aggregate matching efficiency is related to variation in referral generation, then examining why referral use differs across industries is informative about the nature of labor market frictions. If, on the other hand, variation in aggregate matching efficiency is due to other factors, then referrals are probably not a leading explanatory factor of cross-industry variation, although they may still play a role in individual workers’ labor market outcomes.

Two polar cases are considered in the theoretical model: In the first case, the rate of referral generation ($\rho$) is equal across industries and the inter-industry heterogeneity is due to variation in the efficiency of the market matching function ($\mu$). In the second case, the rate of referral generation varies across industries and the efficiency of the market matching function is constant. In each case the model is used to derive the predicted correlation between two quantities that can be observed in the data: the proportion of jobs found through referrals and the aggregate matching efficiency, to be estimated using a standard matching function (i.e. one where referrals are not explicitly modeled). The predictions are then confronted with the data. This is, of course, a very stark exercise so the results should be viewed with caution. A discussion of the conclusions to be drawn is below.

The following proposition describes the interaction according to the model between the two observable quantities and variation in the parameters $\mu$ and $\rho$.

**Proposition 4.1.** We have:

1. Keeping the rate of generating referrals ($\rho$) constant, a higher matching efficiency through the market ($\mu$):
   - reduces the proportion of jobs that are found through a referral;
   - increases aggregate matching efficiency.
2. Keeping the matching efficiency through the market ($\mu$) constant, a higher rate of generating referrals ($\rho$):
   - increases the proportion of jobs that are found through a referral;
   - increases aggregate matching efficiency.

**Proof.** See Appendix A. $\square$

Therefore, an increase in either $\mu$ or $\rho$ increases aggregate matching efficiency but the two variables have opposite effects on the proportion of jobs found through a referral. This distinguishes whether the variation in aggregate matching efficiency is driven by variation in $\mu$ or in $\rho$, according to the model. It is worth noting that, though intuitive, these results are not immediate because a change in either variable affects the entry decision of firms.

These observations are summarized in the first property:

**Property 1.** According to the model, if the correlation between the proportion of jobs found through a referral and the aggregate matching efficiency

- is positive, then variation in $\rho$ is the source of variation of aggregate matching efficiency across industries;
- is negative, then variation in $\mu$ is the source of variation of aggregate matching efficiency across industries.
The proportion of jobs found through a referral for each industry is calculated from the 1994 wave of the NLSY. The interviewees were asked which method of search led to being offered their current job and the response “contacted friends and relatives” is interpreted as evidence that a referral took place. The proportion of interviewees who report finding their job through referral is calculated for 15 major industries and they are presented in Table 1.9

A Cobb–Douglas matching function is estimated using monthly data on industry-specific vacancies and hires from JOLTS, and monthly data on industry-specific unemployed from the Current Population Survey (CPS). The data covers all major industries except for agriculture from January 2001 to June 2011.

The following matching function is estimated using OLS:

\[
\ln \left( \frac{m_{it}}{u_{it}} \right) = \ln(\hat{\kappa}_i) + \hat{\phi} \ln \left( \frac{v_{it}}{u_{it}} \right) + \hat{\zeta}_1 t + \hat{\zeta}_2 t^2,
\]

where \( m_{it} \), \( v_{it} \) and \( u_{it} \) are the number of hires, vacancies and unemployed workers, respectively, in industry \( i \) at time \( t \), \( \hat{\phi} \) is the elasticity of the matching function, \( \hat{\kappa}_i \) is the industry-specific efficiency of the aggregate matching function and there is a quadratic time trend (note that the data series are not de-trended).

---

9 The definition of major industries changed in 2000. The numbers reported here are consistent with the most recent definition.

10 See Davis, Faberman and Haltiwanger [12] for a detailed description of JOLTS. Note that an establishment is counted by JOLTS as having a vacancy when it is making ‘word of mouth’ announcements in order to hire. Therefore, in principle there is no systematic undermeasurement of vacancies that end up being filled through a referral.

11 The CPS assigns unemployed workers to the industry where they were last employed. Sahin, Song, Topa and Violante [30] estimate the transition matrix across industries and use it to adjust for the industry where an unemployed worker searches. They find that their estimates of matching function parameters do not change appreciably.
The following tables summarize the results. Table 1 presents the estimates of industry-specific matching efficiency (with standard errors in parentheses) and the proportion of referral hires (sample size in parentheses).\(^{22}\)

For completeness, Table 2 presents the estimates of the other parameters.\(^{23}\)

The estimated efficiency parameters of the matching function are positively and significantly correlated with the proportion of referrals as can be seen in Fig. 1 and from the outcome of the following regression (standard errors in parentheses)\(^{24}\):

\[
\hat{\kappa}_i = \tau_0 + \tau_1 PR_i
\]

\[
0.071 \ (0.093) \quad 3.713 \ (0.333)
\]

where \(\hat{\kappa}_i\) is the estimate of matching efficiency for industry \(i\) and \(PR_i\) is the proportion of jobs in industry \(i\) that are found through a referral.

According to the model, these results mean that the variation in the inter-industry matching efficiency is due to the variation in the rate of referral generation. Taking the quantitative estimates at face value, a ten percentage point increase in the prevalence of referrals is associated with an increase in the efficiency parameter of 0.37, which corresponds to a 30% increase in the matching efficiency of the average industry.

Although variation in the use of referrals is the dominant factor, the efficiency of the market matching function (\(\mu\)) might also differ across industries. The extent of that variation is not assessed because the “correct” matching function according to the model, which is given in Eq. (2), has not been estimated. The reason is that some of the assumptions that have been made for theoretical convenience should be modified to give a good empirical fit (e.g. the assumption of instantaneous matching when a referral is used) which introduces complications that are beyond the present paper’s scope.

Of course, the rate of referral generation is not an exogenous variable: it is the outcome of decisions made by firms and workers and therefore studying the determinants of these decisions has the potential of being informative about the nature of frictions in the labor market. Furthermore, the factors affecting the agents’ referral choices are likely to also affect the ease of matching through the market (which, incidentally, means that the magnitudes of the estimates reported above are likely to be biased). A first step in studying these effects is in a companion

\(^{22}\) The estimates for matching efficiency are practically identical (correlation of 0.96) with those of Sahin, Song, Topa and Violante [30]. They are also highly correlated (0.83) with the estimates of the vacancy yield in Davis, Faberman and Haltiwanger [12].

\(^{23}\) The estimate of the elasticity of the matching function is higher than other estimates in the literature. This is a feature of the JOLTS data and has been observed before: Sahin, Song, Topa and Violante [30] find a similar estimate for the elasticity to that of Table 2; when they use CPS data their estimate drops by half (see their Table B4).

\(^{24}\) Each industry is equally weighted in the regression and the standard errors are calculated using bootstrap. Weighting each industry by its employment share changes the estimate of \(\tau_1\) to 3.67.
paper (Galenianos [14]) where the informativeness of referrals about match quality affect how often they are used and additional predictions regarding wages and separation rates are supported by the data.

4.2. Job destruction and referral prevalence in the cross-section

An alternative source of heterogeneity across industries is the job destruction rate. A higher job destruction rate reduces the incentive to create a vacancy which leads, under some conditions, to a greater proportion of jobs being found through a referral. This intuition holds for the case where the elasticity of the matching function with respect to vacancies is relatively high and the unemployment rate is “low enough.” The main theoretical result is the following:

**Proposition 4.2.** The prevalence of referrals is an increasing function of the job destruction rate if $\eta > \beta$ and $k < \bar{k}$ for some finite $\bar{k}$.

**Proof.** See Appendix A.  

The second property follows:

**Property 2.** The job destruction rate and the prevalence of referrals are positively correlated, if the conditions of Proposition 4.2 hold.
There is some evidence to support this mechanism. Regressing referral prevalence on JOLTS data for the annual separation rate of different industries leads to:

\[ PR_i = \sigma_0 + \sigma_1 \delta_i \]

0.242 (0.037) 0.178 (0.086)

where \( \delta_i \) is the separation rate of industry \( i \) and \( PR_i \) is the proportion of jobs in industry \( i \) that are found through a referral. The coefficient \( \sigma_1 \) is significant at the 10% level but is (just) insignificant at the 5% level.

A more complete model of this interaction would include the joint determination the sources of differential separation rates (e.g. differential demand volatility) and the firms’ response in terms of hiring methods. Galenianos [14] introduces endogenous search intensity across the two different channels and could be extended to endogenize job separations.

### 4.3. Cyclical properties

The model’s cyclical properties are now examined.

**Proposition 4.3.** Conditional on labor market tightness, the job finding rate is decreasing in the unemployment rate. The aggregate matching function exhibits decreasing returns to scale.

**Proof.** See Appendix A. □

To see the intuition for the first part of Proposition 4.3, consider how an unemployed worker’s job finding rate is affected by a proportional increase in the number of unemployed workers and vacancies. Such a change does not affect his meeting rate through the market, since labor market tightness is unchanged; however, it reduces his meeting rate through referrals, since the number of employed links that act as the source of a referral is lower. In other words, an increase in the unemployment rate reduces the effectiveness of the referral channel which corresponds to a decrease in the efficiency of the aggregate matching function.

Identifying the (inverse of the) unemployment rate with the business cycle, leads to the third property.

**Property 3.** The efficiency of the aggregate matching function is pro-cyclical.

This is consistent with the evidence presented in several recent papers that have empirically examined the cyclical properties of matching efficiency. Barnichon and Figura [1] estimate a matching function and calculate its residual in a way similar to the calculation of the Solow residual and find that the residual exhibits cyclical regularities, increasing in the late parts of expansions and falling in the aftermath of recessions. Furthermore, they find that this variation cannot be explained by differential cyclicity across industries.

Cheremukhin and Restrepo [10] perform a business cycle accounting exercise of the standard search and matching model to decompose the labor wedge into several components one of which is variations in matching efficiency. They find that matching efficiency drops significantly in the aftermath of recessions. On the firm side, Davis, Faberman and Haltiwanger [12] find that the vacancy filling rate is lower in weak labor markets than would be predicted by a matching function with constant efficiency.
These findings are consistent with Property 3 which provides a natural interpretation for why they occur. This interpretation is complementary to the explanations that have already been proposed and assessing the relative importance of referrals and the alternatives is a quantitative issue which is left for future work.25

4.4. Further properties

The final two properties relate to the effects of network heterogeneity on wages and employment rates.

Property 4. Ceteris paribus, increasing the size of a worker’s network leads to a drop in the probability that he is unemployed and an increase in his wage.

Property 5. Ceteris paribus, increasing the employment rate of a worker’s network leads to a drop in the probability that he is unemployed and an increase in his wage.

The logic is straightforward: increasing the size of a worker’s network or the employment rate of his links raise his job finding rate which reduces his unemployment probability and raises his wage by increasing his value of unemployment. These prediction are consistent with the finding of Bayer, Ross and Topa [2] about network size and Topa [31], Weinberg, Reagan and Yankow [33] and Cappellari and Tatsiramos [9] about the network’s employment rate.

Properties 4 and 5 are essentially partial equilibrium predictions due to the ceteris paribus statement, however the results also hold in a fully specified model with network heterogeneity of Igarashi [20]. He analyzes a similar model where workers have homogeneous productivity but heterogeneous network sizes and finds that workers who have greater access to networks enjoy higher wages and lower unemployment.

5. Conclusions

The aim of this paper is to combine social networks, which have long been recognized as an important feature of labor markets, with the equilibrium models that are used to study labor markets. This is achieved in a tractable theoretical framework which, despite its simplicity, is consistent with empirical findings and yields further empirical insights.

Of particular interest is the result that variation in the prevalence of referrals is a source of variation in the speed of matching. This is relevant in light of the finding in Davis, Faberman and Haltiwanger [12] and Sahin, Song, Topa and Violante [30] that there is significant variation in the speed with which vacancies are filled across different industries and, especially, that this variation cannot be explained by differences in the number of searchers across industries. An empirical exploration of the source of cross-industry variation in the prevalence of referrals is left for future work.

A theoretical study of the determinants of intensity of search through referrals is performed in a companion paper, Galenianos [14]. In that paper, forming a match through a referral alleviates a learning friction by facilitating the hiring of workers who are better-matched to the job. In the

25 Davis, Faberman and Haltiwanger [12] propose that the variation is due to lower recruiting intensity by employers. Barnichon and Figura [1] attribute a large part of the variation to the varying composition of the unemployment pool.
types of jobs where this learning advantage is greater, one should expect high use of referrals and larger differentials in wages and separation rates between referred and non-referred workers. There is some empirical evidence to support that prediction which suggests that approaching this question through the lens of learning is a fruitful way to proceed.

A further avenue for future work is to introduce social networks in the study of individuals’ migration decisions. There is ample empirical evidence to suggest that social networks affect these decisions. For instance, Munshi [25] finds that Mexican migrants are more likely to move to locations with more people from their village of origin and this helps them with finding employment while Belot and Ermisch [3] show that an individual is less likely to move if he has more friends at his current location. Therefore, it seems natural to combine the decision to migrate with an explicit model of how the social network helps a worker to find a job.

Finally, this paper’s focus is on the positive implications of combining social networks and labor market models. Having provided a theoretical framework, one can move towards asking normative questions. A first step is taken in Igarashi [20] who studies the effect of banning referrals in a market where some workers have no access to networks. Surprisingly, he finds that non-networked workers might become worse off even though they have no direct access to referrals.

Appendix A

Proposition 3.2. We have:

1. The equilibrium is generically inefficient.
2. Constrained efficiency is attained for an appropriate choice of the two bargaining parameters \( \beta \) and \( \gamma \).
3. The value of \( \beta \) that leads to constrained efficiency does not generically equal the elasticity of the market matching function.

Proof. 1. The Hamiltonian of the planner’s problem can be written as:

\[
H(u, v, \lambda) = e^{-rt}(y(1 - u) + bu - kv) + \lambda \left( \delta(1 - u) - \mu v^\eta u^{1-\eta} - (1 - u)u\rho \right)
\]

where \( \lambda \) is the co-state variable. The partial derivatives of the Hamiltonian with respect to the co-state, state and control variables are:

\[
\frac{\partial H}{\partial \lambda} = \delta(1 - u) - \mu v^\eta u^{1-\eta} - (1 - u)u\rho = \dot{u}
\]

\[
\frac{\partial H}{\partial u} = -e^{-rt}(y - b) - \lambda \left( \delta + (1 - \eta)\mu \left( \frac{v}{u} \right)^\eta + (1 - 2u)\rho \right) = -\dot{\lambda}
\]

\[
\frac{\partial H}{\partial v} = -e^{-rt}k - \lambda \eta \mu \left( \frac{u}{v} \right)^{1-\eta} = 0
\]

Rearranging the above and introducing the steady state condition \( \dot{u} = 0 \) lead to:
\[
\frac{y - b}{k} = r + \delta + (1 - \eta)(1 - u)(\frac{\delta}{u} - \rho) + (1 - 2u)\rho \\
\eta\mu\left(\frac{\mu}{(1-u)(\frac{\delta}{u} - \rho)}\right)^{1-\eta}\]

The right-hand side is strictly decreasing in \(u\), approaches zero as \(u \to 1\) and approaches infinity as \(u \to 0\). Therefore this equation uniquely defines the planner’s optimal level of unemployment.

The equilibrium level of unemployment is given by Eq. (9) which is rearranged as:

\[
\frac{y - b}{k} = r + \delta + \beta(1 - u)(\frac{\delta}{u} - \rho) + \rho((1 - u)\beta - \gamma u(1 - \beta)) \\
(1 - \beta)\mu\left(\frac{\mu}{(1-u)(\frac{\delta}{u} - \rho)}\right)^{1-\eta}\]

It is clear that the right-hand side of Eqs. (11) and (12) are generically not equal to each other and therefore the equilibrium is generically inefficient.

2. Equating the right-hand side of Eqs. (11) and (12), the planner and equilibrium levels of unemployment are equal to each other if and only if

\[T(\beta, \gamma) = 0\]

where:

\[
T(\beta, \gamma) = (1 - \beta - \eta)(r + \frac{\delta}{u}) + \rho((1 - \beta)(\eta(1 - u) - u(1 - \gamma)))\]

It is now shown that \(\beta\) and \(\gamma\) can be chosen so that \(T(\beta, \gamma) = 0\) for an arbitrary \(u\). Notice:

\[
\frac{\partial T(\beta, \gamma)}{\partial \gamma} = \rho(1 - \beta)\eta u > 0
\]

\[
\frac{\partial T(\beta, \gamma)}{\partial \beta} = -r - \rho\eta(1 - u) - \left(\frac{\delta}{u} + \rho u(1 - \eta\gamma)\right) < 0
\]

\[T(1, 0) = -\eta\left(r + \frac{\delta}{u}\right) < 0\]

\[T(0, 1) = (1 - \eta)\left(r + \frac{\delta}{u}\right) + \rho(\eta - u)< 0\]

Finally, \(T(0, 1) > 0\) for all \(u\) because, defining \(\hat{T}(\eta) \equiv T(0, 1)\), we have:

\[
\hat{T}(0) = r + \frac{\delta}{u} - \rho u > 0
\]

\[
\hat{T}(1) = \rho(1 - u) > 0
\]

\[
\hat{T}'(\eta) = -(r + \frac{\delta}{u}) + \rho < 0
\]

3. Immediate from the above.  

**Proposition 4.1.** We have:

1. **Keeping the rate of generating referrals (\(\rho\)) constant, a higher matching efficiency through the market (\(\mu\))**:
   - reduces the proportion of jobs that are found through a referral;
   - increases the aggregate matching efficiency.

2. **Keeping the matching efficiency through the market (\(\mu\)) constant, a higher rate of generating referrals (\(\rho\))**:
   - increases the proportion of jobs that are found through a referral;
   - increases the aggregate matching efficiency.
Proof. The prevalence of referrals is studied first. The prevalence of referrals in steady state is equal to:

\[ PR = \frac{\alpha_R}{\alpha_M + \alpha_R} = \frac{(1 - u)\rho}{\frac{(1-u)\delta}{u}} = \frac{u\rho}{\delta} \]

where \( u \) is given by the solution of Eq. (9). Tedious algebra leads to:

\[
\frac{du}{d\mu} = -\frac{1}{\eta\mu\left(\frac{1}{\eta} - \frac{1}{\beta}\right)}\left(\frac{r + (1 - \beta)\delta + \frac{\beta\delta}{u} - \rho\gamma(1 - \beta)u}{\eta\mu\left(\frac{1}{\eta} - \frac{1}{\beta}\right)}(r + (1 - \beta)\delta + \frac{\beta\delta}{u} - \rho\gamma(1 - \beta)u + (\frac{\beta\delta}{u} + \rho\gamma(1 - \beta)))\right) < 0
\]

and therefore:

\[
\frac{dPR}{d\mu} = \frac{\rho}{\delta} \frac{du}{d\mu} < 0
\]

Furthermore:

\[
\frac{du}{d\rho} = -\frac{1}{\eta\mu\left(\frac{1}{\eta} - \frac{1}{\beta}\right)}\left(\frac{r + (1 - \beta)\delta + \frac{\beta\delta}{u} - \rho\gamma(1 - \beta)u}{\eta\mu\left(\frac{1}{\eta} - \frac{1}{\beta}\right)}(r + (1 - \beta)\delta + \frac{\beta\delta}{u} - \rho\gamma(1 - \beta)u + (\frac{\beta\delta}{u} + \rho\gamma(1 - \beta)))\right)
\]

Using Eq. (14) it can be shown that

\[
\frac{dPR}{d\rho} = \frac{1}{\delta}\left(u + \rho \frac{du}{d\rho}\right) > 0
\]

Aggregate matching efficiency is defined as:

\[ ME = \frac{\alpha_M + \alpha_R}{\theta \eta} = \frac{(1-u)\delta}{u \left(\frac{\delta}{u} - \rho\right)} = \frac{\delta\mu}{\delta - \rho u} \]

which implies that:

\[
\frac{dME}{d\rho} = \frac{\delta\mu}{(\delta - \rho u)^2}\left(u + \rho \frac{du}{d\rho}\right)
\]

\[
\frac{dME}{d\mu} = \frac{\delta}{(\delta - \rho u)^2}\left(\delta - \rho u + \mu \rho \frac{du}{d\mu}\right)
\]

The earlier calculations prove that aggregate matching efficiency is increasing in \( \rho \).

Aggregate matching efficiency is increasing in \( \mu \) when:

\[
u\left(\frac{\delta}{u} - \rho\right) > -\mu \rho \frac{du}{d\mu}
\]

\[\Leftrightarrow \left(\frac{r + (1 - \beta)\delta + \frac{\beta\delta}{u} - \rho\gamma(1 - \beta)u}{1 - u}\right)\left(\delta(1 - \eta) - \rho u(1 - \eta u)\right)
\]

\[+ \eta\left(\frac{\delta}{u} - \rho\right)\left(\frac{\beta\delta}{u} + u\rho\gamma(1 - \beta)\right) > 0\]

where Eq. (13) was used. The inequality holds if:
\[ \delta (1 - \eta) - \rho u(1 - \eta u) > 0 \]

This inequality holds because the left-hand side is a quadratic in \( u \) which takes a strictly positive value when \( u = 0 \) and whose determinant is negative. This completes the proof. \( \square \)

**Proposition 4.2.** The prevalence of referrals is an increasing function of the job destruction rate if \( \eta > \beta \) and \( k < \bar{k} \) for some finite \( \bar{k} \).

**Proof.** Taking the derivative of referral prevalence with respect to \( \delta \) yields:

\[ \frac{d PR}{d \delta} = \frac{d(u \rho)}{d \delta} = \frac{\rho}{\delta^2} \left( \frac{du}{d \delta} \delta - u \right) \Rightarrow \frac{d PR}{d \delta} > 0 \iff \frac{du}{d \delta} > \frac{u}{\delta} \]

where \( u \) is given by the solution of Eq. (9).

Some algebra leads to:

\[ \frac{du}{d \delta} = \frac{1 - \eta}{\eta} \frac{1 - \eta + \rho \gamma (1 - \beta)u}{1 - \beta + \rho \gamma (1 - \beta)u} \]

Using this expression, it can be shown that \( \frac{d PR}{d \delta} > 0 \) if

\[ \delta (\eta - \beta) > u(1 - \beta)(\delta - u \rho \gamma) + ((1 - \beta) \rho \gamma \eta + r(1 - \eta)) \]

This inequality holds if \( \eta > \beta \) and if \( u \) is small enough which, using Eq. (9), corresponds to \( k \) being low enough. \( \square \)

**Proposition 4.3.** Conditional on labor market tightness, the job finding rate is decreasing in the unemployment rate. The aggregate matching function exhibits decreasing returns to scale.

**Proof.** Consider the effect of an increase in the number of unemployed workers and vacancies by a factor \( \xi > 1 \) on the aggregate matching function (Eq. (2)):

\[ M(\xi v, \xi u) = \mu(\xi v)^{\eta} (\xi u)^{1-\eta} + \rho(\xi u)(1 - (\xi u)) 
= \xi M(v, u) - \xi(\xi - 1)\rho u^2 \]

which proves the decreasing returns to scale. Increasing \( u \) while keeping labor market tightness constant requires that \( \xi > 1 \). An unemployed worker’s job finding rate is given by:

\[ \frac{M(\xi v, \xi u)}{\xi u} = \frac{M(v, u)}{u} - (\xi - 1)\rho u \]

which is decreasing in \( u \) when labor market tightness is constant. \( \square \)

**References**

