Review of Economic Dynamics 16 (2013) 668–690

Contents lists available at SciVerse ScienceDirect

Review of Economic Dynamics

www.elsevier.com/locate/red

Learning about match quality and the use of referrals

Manolis Galenianos *,1

Pennsylvania State University, United States

ARTICLE INFO

Article history:
Received 8 May 2012
Received in revised form 7 November 2012
Available online 29 November 2012

JEL classification:
J0

Keywords:
Search
Matching
Referrals
Learning

ABSTRACT

The firm's decision to use referrals as a hiring method is studied in a theoretical model of the labor market. The labor market is characterized by search frictions and uncertain quality of the match between a worker and a job. Using referrals increases the arrival rate of applicants and provides more accurate signals regarding a worker's suitability for the job. Consistent with the data, referred workers are predicted to have higher wage, higher productivity and lower separation rates and these differentials decline with tenure. The model is extended by introducing heterogeneity in firm productivity and allowing the endogenous determination of signal accuracy. High productivity firms are predicted to invest more in increasing signal accuracy and use referrals to a lesser extent.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

It is well known that approximately half of all American workers find their jobs through referrals by friends, acquaintances, relatives, etc.[2] Most work on this topic focuses on the effect of referrals on workers' labor market outcomes. The focus of the present study is on the other side of the market, namely on the firms' decision to use referrals as a method of hiring workers.

This paper proposes a theoretical model where firms choose the intensity of referral use and yields three sets of results. First, the model's predictions regarding the differentials in wages, productivity and separation rates between referred and non-referred workers are consistent with the findings of many empirical studies. Second, high productivity firms are predicted to use referrals to a lesser extent which is consistent with the evidence and can reconcile seemingly contradictory findings regarding the wage premium of a referral. Third, the model points towards specific moments in the data that can be explored to better understand differences in the prevalence of referrals across sectors of economy.

A frictional model of the labor market is developed where a firm and a worker meet through the market or through a referral and the firm exerts effort in searching through each channel. The productivity of an employment relationship is match-specific and it is uncertain at the time of meeting. The firm and worker observe a public signal before deciding whether to form a match and the accuracy of the signal depends on the channel through which they met. Match quality is revealed over the course of the employment relationship. In the baseline model, firms are homogeneous.

Under the assumption that a referral leads to a more accurate signal regarding match quality, the model's predictions are consistent with a wealth of empirical evidence regarding worker outcomes. First, a referred applicant is more likely to be

* Fax: +1 (814) 863 4775.
E-mail address: manolis@psu.edu.

[1] I would like to thank an editor, three anonymous referees, Boyan Jovanovic, Giuseppe Moscarini, Uta Schoenberg, Giorgio Topa and seminar and conference participants for useful comments and the National Science Foundation for financial support (grant SES-0922215).

[2] For instance Granovetter (1995). See Ioannides and Datcher Loury (2004) or Topa (2011) for recent surveys.

1094-2025/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.red.2012.11.005

hired.[3] Conditional on being hired, the probability of a good match is greater if a worker is hired through a referral rather than the market which leads to the second prediction: ceteris paribus, a referred worker receives a higher wage and has higher productivity.[4] The third prediction is that a referred worker has lower separation rates as he is more likely to survive learning about match quality.[5] The fourth prediction is that the differentials between referred and non-referred workers in wages, productivity and separation rates decline with job tenure, because the bad matches are terminated when learning occurs.[6]

The model is subsequently extended along two dimensions: firms are heterogeneous in productivity and choose the accuracy of their signals subject to a cost. High productivity firms are predicted to choose a higher level of signal accuracy because they face a greater opportunity cost of employing a worker who turns out to be badly matched. This choice diminishes the informational advantage of referrals and leads to the prediction that high productivity firms use referrals to a lesser extent.[7]

A referred worker is still predicted to receive a higher wage conditional on his employer's productivity but referred workers are more likely to work in low productivity firms which pay lower wages. This leads to the prediction that the wage premium of a referral is reduced when the firm type is omitted and it is even possible to generate a negative correlation between a referral and the wage, which is consistent with the data.[8] Therefore, the seemingly contradictory evidence on whether a referral is associated with a higher or a lower wage can be reconciled by taking into consideration that different type of firms make different choices regarding the intensity of referral use.

Finally, the model offers a possible interpretation for the large differentials in referral use across sectors of the economy (such as occupations or industries).[9] According to the model, heavy use of referrals in some sector suggests that the referral's informational advantage is greater in that sector. This informational advantage should also lead to greater wage, productivity and separation differentials between referred and non-referred workers. There is some supportive evidence for the predicted positive correlation between the prevalence of referrals and these differentials but a systematic study has not been performed so far.

The closest paper to the current work is Dustmann et al. (2011). In addition to providing what is to data the most detailed empirical analysis of the effect of referrals on labor market outcomes, they also develop a theoretical model where match quality is uncertain and referrals lead to more accurate signals regarding match quality.[10] The present paper departs from their theoretical analysis in the following ways: first, firms choose the intensity with which they search through the market and referrals; second, firm productivity is heterogeneous; third, firms choose the accuracy of the signal regarding match quality. This paper complements their work in providing an interpretation for why high wage firms use referral to a lesser extent.

2. The labor market with homogeneous firms

This section presents the baseline model, characterizes the equilibrium and derives its main testable predictions.

2.1. The model

Time is continuous, the horizon is infinite and the labor market is in steady state. There is measure n of workers who are ex ante homogeneous, risk-neutral, maximize expected discounted utility and discount the future at rate r. A worker is either employed or unemployed. The flow utility of unemployment is $z > 0$ and the flow utility of employment is equal to the wage.

There is an endogenous measure of firms which is determined through free entry. Each firm hires one worker, is risk-neutral, maximizes expected discounted profits and discounts the future at rate r. A firm is either filled and producing or vacant and searching. The flow profit when vacant is $-K$ and the flow profit when producing is equal to output minus the wage.

[3] This is consistent with the evidence in the firm-level studies of Fernandez and Weinberg (1997), Castilla (2005), and Brown et al. (2012).

[4] For evidence regarding wages see Simon and Warner (1992), Bayer et al. (2008), Dustmann et al. (2011) and Brown et al. (2012); regarding productivity see Castilla (2005) and Pinkston (2012) who have data on direct measures of worker productivity.

[5] See Datcher (1983), Simon and Warner (1992), Dustmann et al. (2011) and Brown et al. (2012).

[6] See Simon and Warner (1992), Castilla (2005), Dustmann et al. (2011) and Brown et al. (2012).

[7] Barron et al. (1987) report that larger firms incur greater screening expenditures per hire and Marsden (1994a) reports that large firms use more formal methods to screen workers. Holzer (1987) and Marsden (1994b) find that larger firms use more formal methods to hire workers and Pellizzari (2010) finds that workers who report finding their jobs through a referral work at smaller firms. This evidence is supportive of both predictions, to the extent that firm productivity and size are positively correlated.

[8] Pistaferri (1999), Pellizzari (2010), Bentolila et al. (2010) and Dustmann et al. (2011) find a negative correlation between wages and a referral when controlling for worker observables and fixed effects but without firm fixed effects. Dustmann et al. (2011) find that the coefficient of a referral turns positive when they add controls for firm fixed effects. The firm-level studies mentioned earlier confirm this positive correlation. See Section 3.3 for a discussion.

[9] For instance, 35–40% of workers in construction and manufacturing report finding their jobs through a referral as opposed to 22–24% in education and health (Galenianos, 2012).

[10] Simon and Warner (1992) develop a similar model in a partial equilibrium setting.

The expected output of a match is given by $y\bar{x}$ where y is the firm's productivity and $\bar{x}$ is the expected match quality. Match quality can be good or bad: $x \in \{x_B, x_G\}$ with $x_G > x_B$. Denoting the probability that a match is good by p, the expected match quality is given by $\bar{x} = px_G + (1-p)x_B$. It is assumed that $x_G = 1$, $x_B = 0$ and that $y > z$.[11]

Match quality is determined when a worker and a firm first meet and it is good with probability $\gamma \in (0,1)$. At the time of the meeting, the worker and firm observe an informative signal regarding the realization of match quality and decide whether to form a match or to continue searching (the signal and matching decision are described in more detail below; the variable p in the previous paragraph corresponds to the posterior probability that the match quality is good).

The worker and the firm are symmetrically informed about the probability that their match is good. During production, match quality is perfectly revealed to the worker and firm at rate λ, at which point they decide whether to terminate or to continue their match. The assumptions on match quality imply that it is always optimal to terminate a match whose quality is revealed to be bad.[12]

Matches are destroyed exogenously at rate δ, in addition to the endogenous destruction that might occur after match quality is revealed. There is no on the job search. The surplus is split through Nash bargaining where the workers' bargaining power is denoted by β. The following assumption will be maintained:

$$\gamma \geqslant \overline{\gamma} = \frac{z}{y + (y-z)\frac{\lambda}{r+\delta}}.$$

This assumption guarantees that the initial uncertainty about match quality does not cause the market to shut down. Indeed, the condition is always satisfied when $\lambda \to \infty$ and learning becomes instantaneous.

A firm and a worker meet through two different channels: the market, M, or a referral, R. The two channels differ in the information that is transmitted about match quality. When a worker and a firm meet through channel $i \in \{M, R\}$, a binary signal $s_i \in \{g_i, b_i\}$ is generated regarding match quality. The signal s_i is correct with probability q_{Gi} if the match is good and q_{Bi} if it is bad. These probabilities will be referred to as the signal's accuracy.

Denote the unconditional probability that the worker generates a good signal by π_i and the posterior probability that the match is good conditional on a good signal by p_i. Therefore

$$\pi_i = \gamma q_{Gi} + (1-\gamma)(1-q_{Bi}),$$
$$p_i = \frac{\gamma q_{Gi}}{\pi_i}.$$

The following assumptions are made regarding the accuracy of the signals: $q_{Gi} \in [\frac{1}{2}, 1]$, $q_{Bi} \in [\frac{1}{2}, 1]$, $q_{ki} \in (\frac{1}{2}, 1)$ for some $k \in \{G, B\}$ and

$$\frac{q_{GR}}{1-q_{BR}} > \frac{q_{GM}}{1-q_{BM}}. \tag{1}$$

The assumptions on signal accuracy mean that $p_i \in (\frac{1}{2}, 1)$ and $p_R > p_M$. The overall level of signal accuracy is higher through the referral than the market channel and therefore the posterior probability of a good match after a good signal is higher after a referral.

Assumption (1) is motivated by the empirical evidence that referred workers receive higher wages, produce more output, have lower separation rates than non-referred workers and these differentials decline over time.[13] The fact that the advantage of referred workers is relatively short-lived suggests that a referral is mostly a mechanism for speeding up the transmission of information that will eventually be revealed in any case. The present paper takes this observation as the starting point and evaluates its implications with respect to firms' other decisions such as search effort and, in Section 3, the choice of signal accuracy in an equilibrium setting. Studying the strategic interaction between the referrer, the referred and the potential employer is beyond the scope of this paper.[14]

[11] Modeling worker heterogeneity as being purely match-specific is very convenient because every worker's value of unemployment is the same regardless of the level of output at the current match. See Galenianos (2012) for a model of referrals with permanent heterogeneity across workers.

[12] This model's qualitative predictions regarding labor market outcomes are very similar to those of Jovanovic (1979) and Moscarini (2005) despite the assumption that learning is lumpy: separation rates fall and wages increase with tenure, conditional on the match surviving. Of course, the dynamics leading up to learning are much simpler in the present model. See Pries (2004) for micro-foundations of the present model's assumptions on the structure of learning and production. Nagypal (2007) provides evidence that declining separation rates are primarily due to learning about match quality rather than learning by doing.

[13] See the evidence cited in the introduction and, especially, Dustmann et al. (2011). Note that the advantage of referred workers is most obvious when the firm type is controlled for, either in the firm-level studies or in the studies that control for firm fixed effects. This is explored further in Section 3 where firm heterogeneity is introduced.

[14] More detailed data regarding the source of the referral would be needed to inform such a study. As far as I know, Pinkston (2012) is the only paper which uses data about the source of a referral and he reports that workers referred by friends of the employer look statistically different from workers referred by other sources, such as current employees or labor unions. This suggests the presence of nepotism in the former case. However, only 5.4% of hires have been referred by the employer's friends (out of 55% of hires who were referred in total) suggesting that nepotism is not a first order issue, at least in that data. Additionally, in a firm-level study Fernandez et al. (2000) find that a referred worker becomes more likely to separate from the firm when his referrer separates himself, suggesting the presence of non-pecuniary benefits of referrals.

Example. A physical structure is presented where the difference between the signals received through the market and referral channels is made explicit.

When a worker and a firm meet, the worker is interviewed which generates an *interview signal* s_1. In the event the meeting occurred through a referral, an additional *personal signal* s_2 is generated. The two signals are conditionally independent. The interview signal is correct with probability q_{G1} if match quality is good and q_{B1} if it is bad. Similarly, the personal signal is correct with probability q_{G2} and q_{B2} if match quality is good or bad, respectively. Assume that $q_{G2} = 1$.

The connection between the model and this example is the following. The market signal in the model is the same as the interview signal from a meeting through the market:

$$q_{GM} = q_{G1},$$
$$q_{BM} = q_{B1}.$$

Turning to referrals, a good referral signal in the model is the same as a good personal and a good interview signal in the example. A bad referral signal in the model is the same as a bad personal signal (which means that match quality is bad for sure) or a good personal signal and a bad interview signal. The resulting probabilities are:

$$q_{GR} = q_{G1}q_{G2} = q_{G1} = q_{GM},$$
$$q_{BR} = 1 - (1 - q_{B1})(1 - q_{B2}) = q_{B1} + q_{B2}(1 - q_{B1}) > q_{B1} = q_{BM}.$$

Therefore, it is clear that $p_R > p_M$ in this example, consistent with assumption (1).

When the firm and worker meet they observe the signal and decide whether to match or to continue searching. Let $d_{si} \in [0, 1]$ denote the probability with which the match is formed when they meet through channel i and observe signal s_i.

Attention is restricted to equilibria where a bad signal leads to separation and a good signal leads to match formation with positive probability through both channels of search: $d_{bM} = d_{bR} = 0$, $d_{gM} > 0$, and $d_{gR} > 0$.[15] To simplify notation, let $d_R = d_{gR}$ and $d_M = d_{gM}$ from now on. Observe that the firm and worker can use lotteries to decide whether to form a match (however, randomization will turn out not to occur in equilibrium).

The aggregate flow of meetings between unemployed workers and vacancies through the two channels is given by a standard Cobb–Douglas function:

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

where u denotes the number of unemployed workers, v denotes the number of vacancies, $\mu > 0$ and $\eta \in (0, 1)$.

A vacancy chooses how much effort to exert in searching through the market (e_M) and through referrals (e_R). Let E_M and E_R denote the aggregate effort exerted in searching through the market and referrals, respectively. The flows of meetings through the market and referrals are given by:

$$m_M(v, u, E_M, E_R) = \frac{E_M}{E_M + E_R} m(v, u),$$
$$m_R(v, u, E_M, E_R) = \frac{E_R}{E_M + E_R} m(v, u).$$

It is assumed that aggregate search effort only affects the proportion of meetings that occur through each channel. A more general specification would also allow the total flow of meetings to be affected by aggregate effort, making μ a function of E_M and E_R. This is not pursued in this paper because the focus is on the *relative* use of referrals vis-a-vis the market as a search channel. Also, the referral process will not be modeled in a detailed way. See Galenianos (2012) for a model where the process of generating referrals is more explicit and the use of referrals has an effect on the aggregate matching efficiency.

When an individual firm exerts additional effort in searching through referrals it increases the arrival rate of workers. The rate at which vacancy j meets a worker through channel i when choosing effort e_M and e_R is:

$$\alpha^j_{Fi} = \frac{e_i}{E_i} \frac{E_i}{E_M + E_R} \frac{m(v, u)}{v}.$$

The cost of exerting effort e_M and e_R is given by

$$C_M(e_M) + C_R(e_R) = \frac{c_M e_M^2}{2} + \frac{c_R e_R^2}{2}.$$

[15] This formulation can be thought of as arising from the following specification. A signal s is drawn from a continuous distribution $F(s|x, i)$ which depends on match quality $x \in \{G, B\}$ and whose density satisfies the monotone likelihood ratio property. The match will be formed only if the signal is above some threshold $\bar{s}_i$ and $q_{Gi} = 1 - F(\bar{s}_i|G, i)$ and $q_{Bi} = F(\bar{s}_i|B, i)$. This specification differs from the model in that the threshold, and hence the probabilities, is endogenous but this does not materially affect any of the decisions that are considered here.

Attention is restricted to symmetric equilibria where every firm chooses $e_i = E_i$ for $i \in \{M, R\}$. Allowing the marginal cost of exerting additional effort to differ across the two channels breaks the connection between signal accuracy and the choice of effort.

A worker meets a vacancy through channel i at rate:

$$\alpha_{Wi} = \frac{E_i}{E_M + E_R} \frac{m(v,u)}{u}.$$

When matched, a worker–firm pair is in one of three states: a match where learning has not yet occurred and was created through the market or through a referral (call this an uncertain match); or a match where learning has occurred and match quality is good (recall that bad matches are immediately terminated), in which case the channel through which they originally met no longer matters. An unmatched worker is unemployed and an unmatched firm is vacant.

The value functions for each state are now determined. An unemployed worker meets firms through channel i at rate α_{Wi} and, if a good signal is emitted (which occurs with probability π_i), a match is formed with probability d_i. Denoting the value of unemployment by U yields:

$$rU = z + \alpha_{WM}\pi_M d_M(W_M - U) + \alpha_{WR}\pi_R d_R(W_R - U),$$

where W_M and W_R denote the worker's value of being in an uncertain match which was created through the market or a referral, respectively.

A vacancy j chooses effort e_M and e_R, meets workers through channel i at rate α^j_{Fi} and, if a good signal is emitted, forms a match with probability d_i. Its value is given by:

$$r\tilde{V}^j(e_M, e_R) = -K + \alpha^j_{FM}\pi_M d_M(J_M - V) + \alpha^j_{FR}\pi_R d_R(J_R - V) - \frac{c_M e_M^2 + c_R e_R^2}{2}, \tag{2}$$

where J_M and J_R denote the firm's value of being in an uncertain match which was created through the market or a referral, respectively.

The value of a vacancy which chooses effort optimally is:

$$V = \max_{e_M, e_R} \tilde{V}^j(e_M, e_R).$$

A worker and a firm in an uncertain match that was created through channel $i \in \{M, R\}$ produce flow output yp_i, determine the wage w_i through Nash bargaining and separate at exogenous rate δ. Furthermore, they learn their match quality at rate λ and transit to a good match with probability p_i or separate if the match quality is bad (with probability $1 - p_i$). The worker's and firm's values are given by:

$$rW_i = w_i + \lambda p_i(W_G - W_i) + \big(\delta + \lambda(1 - p_i)\big)(U - W_i), \tag{3}$$

$$rJ_i = yp_i - w_i + \lambda p_i(J_G - J_i) + \big(\delta + \lambda(1 - p_i)\big)(V - J_i), \tag{4}$$

where W_G and J_G denote the worker's and firm's value of being in a good match, respectively.

A worker and a firm in a good match produce flow output y, determine the wage w_G through Nash bargaining and separate at exogenous rate δ. The worker's and firm's values when in a good match are given by:

$$rW_G = w_G + \delta(U - W_G), \tag{5}$$

$$rJ_G = y - w_G + \delta(V - J_G). \tag{6}$$

Wages are determined by Nash bargaining which solves:

$$w_k = \arg\max_w (W_k - U)^\beta (J_k - V)^{1-\beta}, \tag{7}$$

where $k \in \{G, M, R\}$.

Turning to labor market flows, a worker can be in one of four states: unemployed, employed at an uncertain match which was created through the market or through a referral and employed at a good match. Denote the measure of workers at each state by u, n_M, n_R, n_G and note that

$$n = u + n_M + n_R + n_G. \tag{8}$$

The labor market is in steady state. The steady state is described by the following conditions which equate the flows in and out of the employment states:

$$u\alpha_{WM}d_M\pi_M = (\delta + \lambda)n_M, \tag{9}$$

$$u\alpha_{WR}d_R\pi_R = (\delta + \lambda)n_R, \tag{10}$$

$$\lambda(n_M p_M + n_R p_R) = \delta n_G. \tag{11}$$

Given Eqs. (9), (10) and (11) it is superfluous to equate the flows in and out of unemployment.

The equilibrium can now be defined.

Definition 2.1. An equilibrium is the steady state measure of unemployed u and vacancies v, the decision rules for forming a match $\{d_M, d_R\}$ and the effort levels E_M and E_R such that:

1. The labor market is in steady state as described in Eqs. (8)–(11).
2. The surplus is split according to (7).
3. The choice of effort maximizes (2).
4. A meeting through either channel leads to a match after a good signal and does not after a bad signal.
5. There is free entry of firms: $V = 0$.

2.2. Equilibrium characterization

This section proves the following result.

Proposition 2.1. *An equilibrium exists if $K \in (\underline{K}, \overline{K})$ where $\underline{K} < \overline{K}$.*

The surplus of a match between a firm and a worker who are at state $k \in \{M, R, G\}$ is:

$$S_k = W_k - U + J_k - V.$$

Note that it is possible for the surplus of an uncertain match to be negative, for instance if the signal accuracy of one channel is very low (it is shown below that the surplus of a good match is always positive).

For the match formation rules to be optimal, a match is created only if the surplus is non-negative. Therefore for $i \in \{M, R\}$:

$$S_i > 0 \quad \Rightarrow \quad d_i = 1,$$
$$S_i < 0 \quad \Rightarrow \quad d_i = 0,$$
$$S_i = 0 \quad \Rightarrow \quad d_i \in [0, 1].$$

Define the expected surplus of a meeting between a worker and a firm through channel i as

$$\bar{S}_i = \pi_i \max[S_i, 0].$$

Within a match (i.e. for $S_i \geqslant 0$) the solution to the Nash bargaining problem implies:

$$W_i - U = \beta S_i,$$
$$J_i - V = (1 - \beta) S_i.$$

It will prove useful to simplify notation by defining:

$$\alpha_W \equiv \frac{m(v,u)}{u} = \mu\left(\frac{v}{u}\right)^{\eta} \quad \Rightarrow \quad \alpha_{Wi} = \frac{\alpha_W E_i}{E_M + E_R},$$
$$\alpha_F \equiv \frac{m(v,u)}{v} = \mu\left(\frac{u}{v}\right)^{1-\eta} \quad \Rightarrow \quad \alpha^j_{Fi} = \frac{\alpha_F e_i}{E_M + E_R}.$$

The value functions of the unemployed worker and vacancy j can be rewritten as:

$$rU = z + \alpha_W \beta \left[\frac{E_M \bar{S}_M}{E_M + E_R} + \frac{E_R \bar{S}_R}{E_M + E_R}\right], \tag{12}$$
$$r\tilde{V}^j(e_M, e_R) = -K + \alpha_F(1-\beta)\left[\frac{e_M \bar{S}_M}{E_M + E_R} + \frac{e_R \bar{S}_R}{E_M + E_R}\right] - \frac{c_M e_M^2 + c_R e_R^2}{2}. \tag{13}$$

Consider the firms' effort choice. To calculate the optimal choice of effort, set the derivative of Eq. (13) with respect to e_i to zero and rearrange to arrive at:

$$e_i = \frac{\alpha_F(1-\beta)\frac{\bar{S}_i}{c_i}}{E_M + E_R}. \tag{14}$$

Notice that $S_i = 0$ implies that $e_i = 0$ and therefore an interior d_i never occurs in equilibrium.

Evaluating Eq. (14) at $e_i = E_i$ for both $i = M$ and $i = R$ yields:

$$\frac{E_i}{E_M + E_R} = \frac{\frac{\bar{S}_i}{c_i}}{\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R}}, \tag{15}$$

if $\bar{S}_M > 0$ and/or $\bar{S}_R > 0$. If $\bar{S}_M = \bar{S}_R = 0$, then $E_M = E_R = 0$. Introducing the firms' optimal effort choice into Eq. (12) leads to the following lemma:

Lemma 2.1. *In equilibrium, the value of unemployment is uniquely determined as a function of the expected match surplus, the measure of unemployed and the measure of vacancies by*:

$$rU = z + \frac{\mu(\frac{v}{u})^{\eta}\beta(\frac{\bar{S}_M^2}{c_M} + \frac{\bar{S}_R^2}{c_R})}{\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R}}, \tag{16}$$

if $\bar{S}_M > 0$ *and/or* $\bar{S}_R > 0$ *and* $rU = z$ *if* $\bar{S}_M = \bar{S}_R = 0$.

The next lemma uses the steady state equations to solve for the measure of unemployment and express the value of unemployment only as a function of the expected match surplus and measure of vacancies.

Lemma 2.2. *In equilibrium, the value of unemployment is uniquely determined as a function of the expected match surplus and the measure of vacancies by the solution to*:

$$\frac{n}{v} = \left(\frac{\mu\beta(\frac{\bar{S}_M^2}{c_M} + \frac{\bar{S}_R^2}{c_R})}{(rU - z)(\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R})}\right)^{\frac{1}{\eta}} + \left(\frac{\mu\beta(\frac{\bar{S}_M^2}{c_M} + \frac{\bar{S}_R^2}{c_R})}{(rU - z)(\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R})}\right)^{\frac{1-\eta}{\eta}} \frac{\mu(\frac{\Gamma_{1M}\bar{S}_M}{c_M} + \frac{\Gamma_{1R}\bar{S}_R}{c_R})}{\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R}},$$

where:

$$\Gamma_{1i} = \frac{\gamma q_{Gi}}{\delta} + \frac{(1-\gamma)(1-q_{Bi})}{\delta + \lambda}.$$

Proof. See Appendix A. □

Turning to the match surplus, the surplus from a good match is calculated by combining Eqs. (5) and (6):

$$(r + \delta)S_G = y - rU - rV. \tag{17}$$

Recall that in equilibrium $rU \geqslant z$ and further observe that $rU \leqslant y$ because y is the level of output. Since $V = 0$ in equilibrium and $y > z$ by assumption, $S_G > 0$ in equilibrium.

To get the surplus of a match of type $i \in \{M, R\}$, combine Eqs. (3) and (4) with the Nash bargaining solution, the free entry condition and Eq. (17):

$$S_i = \frac{yp_i}{r+\delta} - \frac{rU}{r+\delta}\frac{r+\delta+\lambda p_i}{r+\delta+\lambda} \quad \Rightarrow \quad \pi_i S_i = y\Gamma_{2i} - rU(\Gamma_{2i} + \Gamma_{3i}),$$

where

$$\Gamma_{2i} = \frac{\gamma q_{Gi}}{r+\delta},$$
$$\Gamma_{3i} = \frac{(1-\gamma)(1-q_{Bi})}{r+\delta+\lambda}.$$

To determine whether a meeting through channel i leads to a match or not ($d_i = 1$ or $d_i = 0$), the relevant endogenous variable is the value of unemployment. In particular,

$$S_i > 0 \quad \Leftrightarrow \quad U < \bar{U}_i \quad \Rightarrow \quad d_i = 1$$

where

$$\bar{U}_i = \frac{y\Gamma_{2i}}{r(\Gamma_{2i} + \Gamma_{3i})}.$$

Therefore, $\bar{U}_i$ is the maximum level of the value of unemployment such that a meeting with a good signal through channel i delivers positive surplus. Intuitively, if the posterior probability that a match is good (p_i) is not too high but the worker's value of unemployment is very high (say because channel k is more promising), then a meeting through channel i does not lead to a match regardless of the signal. Notice that the firm's value does not appear because $V = 0$ in equilibrium.

The assumptions on signal accuracy mean that $\overline{U}_M < \overline{U}_R$ and the assumption that $\gamma \geqslant \overline{\gamma}$ implies that $\overline{U}_M > z/r$. The expected surplus of a meeting can now be characterized as a function of the value of unemployment:

$$\begin{aligned}
U \in \left[\frac{z}{r}, \overline{U}_M\right) \quad &\Rightarrow \quad \bar{S}_i(U) = y\Gamma_{2i} - rU(\Gamma_{2i} + \Gamma_{3i}) \quad &&\Rightarrow \quad d_i = 1, \quad i \in \{M, R\},\\
U \in [\overline{U}_M, \overline{U}_R) \quad &\Rightarrow \quad \bar{S}_R(U) = y\Gamma_{2R} - rU(\Gamma_{2R} + \Gamma_{3R}) \quad &&\Rightarrow \quad d_R = 1,\\
&\Rightarrow \quad \bar{S}_M(U) = 0 &&\Rightarrow \quad d_M = 0,\\
U \in \left[\overline{U}_R, \frac{y}{r}\right) \quad &\Rightarrow \quad \bar{S}_M(U) = \bar{S}_R(U) = 0 &&\Rightarrow \quad d_M = d_R = 0.
\end{aligned}$$

In words, if the value of unemployment is low, then meetings through both channels lead to a match. If the value of unemployment takes an intermediate value, then meetings through the less informative channel (the market) do not lead to a match but meetings through a referral do. If the value of unemployment is high, then the surplus of a match is always negative and no match is created. Of course, the last case cannot occur in equilibrium because the value of unemployment is greater than z only if workers can occasionally find jobs. Furthermore, notice that $U < \overline{U}_M$ is necessary for hires to occur through both search channels, which is the case that this paper focuses on.

The following lemma combines the characterization of expected match surplus derived above with the result of Lemma 2.2 to determine the value of unemployment as a function of the measure of vacancies alone. Furthermore, it derives conditions such that a meeting through either channel leads to a match after a good signal and does not after a bad signal.

Lemma 2.3. *In equilibrium*:

1. *The value of unemployment is determined as a function of the measure of vacancies, $U(v)$.*
2. $\lim_{v\to 0} U(v) = \frac{z}{r}$ *and* $\lim_{v\to\infty} U(v) = \overline{U}_R$.
3. *There exists $\bar{v}$ such that $v < \bar{v} \Rightarrow U(v) < \overline{U}_M$.*
4. *There exists $\underline{v}$ such that if $v > \underline{v}$ then a bad signal does not lead to a match.*

Proof. See Appendix A. □

Finally, consider the firm's problem. Combine Eqs. (13), (14) and (15), to determine the value of a vacancy when search effort is chosen optimally:

$$rV = -K + \frac{1}{2}(1-\beta)\alpha_F \frac{\frac{\bar{S}_M^2}{c_M} + \frac{\bar{S}_R^2}{c_R}}{\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R}}.$$

Lemma 2.4. *If $K \in (\underline{K}, \overline{K})$ there exists v such that $rV = 0$, matches occur through both channels after a good signal and a bad signal does not lead to a match.*

Proof. See Appendix A. □

This completes the proof of Proposition 2.1.

2.3. Predictions

This section presents the first set of the model's predictions.

Let $\mathcal{P}$ denote the proportion of hires that occurs through the referral channel:

$$\mathcal{P} = \frac{\alpha_{FR}\pi_R}{\alpha_{FM}\pi_M + \alpha_{FR}\pi_R}.$$

$\mathcal{P}$ will be referred to as the prevalence of referrals.

Proposition 2.2. *In equilibrium*:

1. *A worker who is hired through a referral receives a higher wage than a worker hired through the market.*
2. *A worker who is hired through a referral produces more output than a worker hired through the market.*

3. *A worker who is hired through a referral has a lower separation rate than a worker hired through the market.*
4. *The differentials in wages, productivity and separation rates across hiring channels decline with the workers' tenure on the job.*

Proof. The assumptions on signal accuracy imply that $p_R > p_M$.

A higher posterior probability of a good match leads to higher output ($yp_R > yp_M$), higher wage ($w_R > w_M$) and lower separation rate ($\lambda(1-p_R)+\delta < \lambda(1-p_M)+\delta$). These differentials disappear after learning has occurred and the probability of being in a good match among surviving workers is equal to 1 regardless of the channel through which they were hired. □

There is a lot of evidence to support these predictions. Regarding point 1, Bayer et al. (2008) and Dustmann et al. (2011) find that referred workers earn higher wages than non-referred workers controlling for workers' observable characteristics and, in the case of Dustmann et al. (2011), also for firm fixed effects. Simon and Warner (1992) find a wage premium for being hired through a referral in the Survey of Natural and Social Scientists and Engineers and Brown et al. (2012) reach the same conclusion in a firm-level study. Regarding point 2, Castilla (2005) and Pinkston (2012) have direct measures of worker productivity and report that productivity is higher for workers who were referred to the firm controlling for worker observables and firm fixed effects.

Regarding point 3, Datcher (1983), Simon and Warner (1992), Dustmann et al. (2011) and Brown et al. (2012) all find that the separation rates are lower for referred workers, controlling for worker observables and, in the case of Dustmann et al. (2011), for firm fixed effects. Regarding point 4, the firm-level study of Castilla (2005) reports that productivity differentials decline with tenure and that of Brown et al. (2012) reports that wage differentials decline with tenure, conditional on worker observables. Furthermore, Simon and Warner (1992), Dustmann et al. (2011) report that differentials in wages and separation rates decline with tenure conditional on worker observables and firm fixed effects.

It is worth remarking that a model where referrals alleviate problems of adverse selection (by bringing in higher productivity workers) or moral hazard (by facilitating monitoring) will provide predictions that are consistent with points 1 and 2 but not with points 3 and, especially, 4.[16] The fact that the performance of referred and non-referred workers converges over time is strong evidence in favor of productivity uncertainty after a hire, which implies that learning is at the heart of what makes referrals useful.

Proposition 2.3. *If $\gamma(q_{GR}-q_{GM}) \geqslant (1-\gamma)(q_{BR}-q_{BM})$ then a meeting through a referral is more likely to lead to a match than a meeting through the market.*

Proof. Notice that

$$\gamma(q_{GR}-q_{GM}) \geqslant (1-\gamma)(q_{BR}-q_{BM}) \quad \Leftrightarrow \quad \pi_R \geqslant \pi_M.$$

If the condition holds, the unconditional probability that a good signal is generated is higher when a firm and a worker meet through a referral than through the market. □

The firm-level studies of Fernandez and Weinberg (1997), Castilla (2005), and Brown et al. (2012) have information about the universe of applicants, including whether the applicant was referred, for their respective firms during the time period under study. All three studies find that referred workers are more likely to be hired conditional on their observable characteristics, consistent with the prediction.

Proposition 2.4. *If*

$$\frac{c_R}{c_M} > \frac{(q_{GR}-(1-q_{BR}))(\gamma q_{GR}+(1-\gamma)(1-q_{BR}))}{(q_{GM}-(1-q_{BM}))(\gamma q_{GM}+(1-\gamma)(1-q_{BM}))},$$

and $K \geqslant \hat{K}$ for some $\hat{K}$ where $\hat{K} < \overline{K}$ then more workers are hired through the market than referrals ($\mathcal{P} < \frac{1}{2}$).

Proof. See Appendix A. □

This proposition makes the point that the informational advantage of referrals does not necessarily lead to a majority of hires occurring through referrals. In particular, if the cost of searching through referrals is sufficiently higher than the market, then firms exert more effort in searching through the market and more workers are hired through the market. This happens despite the fact that a worker is more likely to be hired conditional on meeting a firm through a referral. This observation is relevant because, though referrals are used across the board, they do not necessarily constitute a majority of hires.

---

[16] Of course, point 3 can be had if the adverse selection problem is with respect to workers' separation rates.

The model provides a framework to examine the differential in the prevalence of referrals across different sectors of the economy, such as industries or occupations. While the evidence in this regard is somewhat scattered, the model points to some moments that could be examined. This is important because these differentials have been shown to be significant and to correlate with the efficiency of matching in the case of industries (Galenianos, 2012).

Suppose that the economy consists of two sectors (a and b), that a signal through a referral is more accurate than a signal through the market in both sectors and that it is *relatively* more accurate in sector a. To make this distinction in a sharper way, three simplifying assumptions will be made. First, a signal is correct with symmetric probability regardless of the underlying match quality:

$$q^l_{Gi} = q^l_{Bi} = q^l_i, \quad l \in \{a, b\},\ i \in \{M, R\}. \tag{18}$$

Second, sector a has a more accurate signal through referrals but a less accurate signal through the market:

$$q^a_R > q^b_R > q^b_M > q^a_M. \tag{19}$$

Third, the value of unemployment is the same in both sectors:

$$U^a = U^b. \tag{20}$$

This could be the outcome of worker mobility, as will be the case in Section 3.

Proposition 2.5. *Suppose that Eqs.* (18), (19) *and* (20) *hold. Then in equilibrium*:

1. *A worker who is hired through a referral produces more output, receives a higher wage and has lower separation rates than a worker who is hired through the market in both sectors.*
2. *The differentials in productivity, wages and separation rates between workers who were hired through a referral or the market are larger in sector a.*
3. *If $\gamma > \frac{1}{2}$ then sector a exhibits greater prevalence of referrals.*

Proof. Eq. (19) implies that

$$p^a_R > p^b_R > p^b_M > p^a_M,$$

which proves points 1 and 2 and also means that $\bar{S}^a_R > \bar{S}^b_R > \bar{S}^b_M > \bar{S}^a_M$.

Note that:

$$\mathcal{P}^a > \mathcal{P}^b \quad \Leftrightarrow \quad \frac{\bar{S}^a_R \pi^a_R}{\bar{S}^a_M \pi^a_M} > \frac{\bar{S}^b_R \pi^b_R}{\bar{S}^b_M \pi^b_M}.$$

If $\gamma > \frac{1}{2}$ then $\pi^a_R > \pi^b_R > \pi^b_M > \pi^a_M$ which proves point 3. □

The predictions that the prevalence of referrals across sectors is positively correlated with differentials across referred and non-referred workers have not been tested in the data. However some supportive evidence exists in the literature.

Dustmann et al. (2011) find that the use of referrals are more common for workers who are young or low-skilled. Furthermore, they find that wages of referred workers are higher than non-referred workers only for young and low-skilled workers.

Brown et al. (2012) examine a large firm which hires workers at a variety of skill levels allowing the authors to interact whether a referral took place with the level of the job (support staff, junior staff, medium staff, senior staff or executive) and the educational attainment that is required (high school diploma, associate degree, bachelor's degree, or graduate degree). They find that a referral increases the probability of a job offer and that this effect is strongest for support stuff and for positions that require lower educational attainment. Furthermore, the wage premium of a referral is largest for support staff, declines with the seniority of the job and is actually negative for executives, although the number of executives is probably too small to draw firm conclusions. Finally, the separation differential is again largest for support staff (i.e. it is lowest for referred support staff workers) and is smaller or insignificant for more senior positions.

These observations are consistent with the well-known fact that referrals are more prevalent at the lower end of the labor market (see Bayer et al., 2008, among others). They are also consistent with the model's prediction that there is a positive correlation between the prevalence of referrals and the differentials between referred and non-referred workers. Of course, the evidence is far from conclusive at this point but it seems to be a promising avenue for future work. See the Conclusions for additional discussion.

3. Firm heterogeneity and endogenous signal accuracy

This section introduces two features to the baseline model: firm heterogeneity and endogenous choice of signal accuracy.

3.1. The extended model

There are two types of firm which differ in productivity. A firm of type H has productivity y^H and a firm of type L has productivity y^L. It is assumed that $y^H > y^L$ and that

$$\gamma \geqslant \overline{\gamma}^L = \frac{z}{y^L + (y^L - z)\frac{\lambda}{r+\delta}}.$$

When a firm of type $t \in \{H, L\}$ is searching for a worker, it is subject to a flow cost $K(v^t)$ where v^t is the measure of type-t vacancies and $K(0) = K'(0) = 0$, $K'(v) > 0$. This assumption will guarantee the coexistence of different types of firms in equilibrium.

There are two islands and each island is populated by firms that belong to one type. From now on the islands are identified by the type of firm that populate them. Every worker chooses one island to enter and the measure of workers in island t is denoted by n^t, where $n^H + n^L = n$.

Each island operates as in the baseline model with two simplifications and one major change. The simplifications are that, first, the cost of exerting search effort is assumed to be the same for both channels and normalized to unity: $c_M = c_R = 1$; second, the probability that a signal is correct is assumed to be symmetric across match qualities: $q_{Gi} = q_{Bi} = q_i$ for $i \in \{M, R\}$.

The major change is that the probability that a signal is correct is a choice variable for the firm. Specifically, each firm chooses h which determines the accuracy of the signals, $q_M(h)$ and $q_R(h)$. A firm that chooses h incurs cost $s(h) = \sigma h$ every time the firm meets with a worker. In other words, the cost is proportional to the frequency of generating signals. The choice of h represents the firm's investment on better screening technology, such as a better human resources department.[17]

It is assumed that $q_i(h)$ is strictly increasing and strictly concave in h, satisfies the Inada condition $\lim_{h\to 0} q_i'(h) = +\infty$ and, as in Section 2, $q_R(h) > q_M(h) > \frac{1}{2}$ for all h. The key assumption is that $q_M'(h) \geqslant q_R'(h) \geqslant 0$. The important component of this assumption is that increasing human resources reduces the gap between referred and market applicants for the job.

In terms of the example of Section 2, one can interpret an increase in h as improving the accuracy of the interviewing signal while leaving the accuracy of the personal signal unaffected. This would have the desired effect of diminishing the accuracy differential between the market and referral signals.[18]

Firms exert effort in searching through the market and referrals and learn about match quality as in Section 2. The value of vacancy j of type t is:

$$r\tilde{V}^{jt}(e_M, e_R, h) = -K(v^t) + \alpha_{FM}^{jt}\big(\pi_M(h) d_M^t\big(J_M^t(h) - V^t\big) - s(h)\big) + \alpha_{FR}^{jt}\big(\pi_R(h) d_R^t\big(J_R^t(h) - V^t\big) - s(h)\big) - \frac{e_M^2 + e_R^2}{2},$$

where

$$V^t = \max_{e_M, e_R, h} \tilde{V}^{jt}(e_M, e_R, h).$$

We focus on equilibria where all firms of a given type make the same decision regarding e_M, e_R and h.

Note that the aggregate choice of human resources affects the other value functions only through its effect on signal accuracy. Therefore, given q_M^t and q_R^t all value functions of workers and firms on island t are identical to the ones in Section 2.

The equilibrium is defined as follows:

Definition 3.1. An equilibrium is the steady state measure of workers n^t, unemployed u^t and vacancies v^t, the decision rules for forming a match $\{d_M^t, d_R^t\}$, the effort levels E_M^t and E_R^t, the choice of human resources h^t and the value of unemployment U^t for $t \in \{H, L\}$ such that in each island:

1. The labor market is in steady state.
2. The surplus is split according to the Nash bargaining solution.
3. The choice of effort and human resources maximizes the value of a vacancy.
4. A meeting through either channel leads to a match after a good signal and does not after a bad signal.
5. There is free entry of firms: $V^t = K(v^t)$;

[17] Note that there is a hold-up problem under this specification: the firm chooses h prior to meeting with the worker and only receives share $1 - \beta$ of the resulting surplus. As a result the joint surplus of a meeting is not maximized. This is consistent with the spirit of the model regarding the importance of frictions in labor markets. The prediction that high productivity firms choose higher levels of h does not qualitatively depend on the presence of the hold-up problem.

[18] The parallel is imperfect because the example is phrased in terms of signals whose accuracy is asymmetric across match qualities but the intuition is clear.

and

6. The value of unemployment is the same across islands ($U^H = U^L$) and $n^H > 0$ and $n^L > 0$ where $n^H + n^L = n$.

3.2. Equilibrium

This section proves that:

Proposition 3.1. *An equilibrium exists if $n \geqslant \underline{n}$ and $\sigma \leqslant \hat{\sigma}$.*

The analysis will be kept as close as possible to Section 2. The equilibrium will be characterized in each island for a given measure of workers n^t. The allocation of workers across islands will be subsequently considered.

Denote the surplus of a match between a worker and a firm that chose h by:

$$S_i^t(h) = J_i^t(h) + W_i^t(h) - V^t - U^t.$$

The value function of a type-t vacancy can be written as follows, taking into account the Nash bargaining solution:

$$\begin{aligned} r\tilde{V}^{jt}(e_M, e_R, h) &= -K(v^t) + \alpha_F^t\left(\frac{e_M}{E_M^t + E_R^t}\big((1-\beta)\pi_M(h)d_M^t S_M^t(h) - \sigma h\big)\right. \\ &\quad \left. + \frac{e_R}{E_M^t + E_R^t}\big((1-\beta)\pi_R(h)d_R^t S_R^t(h) - \sigma h\big)\right) - \frac{e_M^2 + e_R^2}{2} \\ &= -K(v^t) + \frac{\alpha_F^t(1-\beta)}{E_M^t + E_R^t}\big(e_M \hat{S}_M^t(h) + e_R \hat{S}_R^t(h)\big) - \frac{e_M^2 + e_R^2}{2}, \end{aligned} \tag{21}$$

where

$$\hat{S}_i^t(h) = \max\big[\pi_i(h) S_i^t(h) - \tilde{\sigma} h, 0\big], \tag{22}$$

and $\tilde{\sigma} = \frac{\sigma}{1-\beta}$. As in Section 2, $d_i^t = 0$ if $S_i^t < 0$ which is implicit in Eq. (22).

Setting the derivative of Eq. (21) with respect to e_M and e_R to zero yields:

$$e_i = \frac{\alpha_F^t(1-\beta)\hat{S}_i^t(h)}{E_M^t + E_R^t}. \tag{23}$$

Combining Eqs. (21) and (23) evaluated at $e_i = E_i^t$ leads to:

$$r\tilde{V}^{jt}(h) = \frac{1}{2}\left(\frac{(1-\beta)\alpha_F^t}{E_M + E_R}\right)^2 \big(\big(\hat{S}_M^t(h)\big)^2 + \big(\hat{S}_R^t(h)\big)^2\big). \tag{24}$$

The following lemma characterizes the optimal choice of h and the resulting value of unemployment.

Lemma 3.1. *In equilibrium*:

1. *The optimal choice of h of a type-t firm is determined as a function of the value of unemployment*: $h^t(U^t)$.
2. *The optimal choice of h is decreasing in σ.*
3. *The expected surplus of a meeting in island t is determined as a function of the value of unemployment*: $\hat{S}_i^t(U^t)$.

Proof. See Appendix A. □

The steps that lead to Lemma 2.2 can be replicated: impose the symmetric effort condition on Eq. (23) and introduce it inside the steady state conditions. The outcome of repeating these steps (which are omitted because they are identical to Section 2) is to derive an expression that determines island t's value of unemployment U^t as a function of the expected match surplus ($\hat{S}_M^t, \hat{S}_R^t$), the measure of vacancies (v^t) and the measure of island t's workers (n^t).

It is straightforward to show that:

Lemma 3.2. *There exists a value of unemployment $U^t(n^t)$ for island t if $n^t \in (\underline{n}^t, \bar{n}^t)$ for some $\underline{n}^t < \bar{n}^t$ such that the market is in steady state, the surplus is split through Nash bargaining, the choices of effort and human resources are optimal, a good signal leads to a match, a bad signal does not and the measure of vacancies is determined through free entry.*

The firm's flow cost depends on the measure of vacancies on the island and therefore the bounds on the measure of workers in Lemma 3.2 correspond to the bounds on the flow cost of vacancies in Proposition 2.1. Otherwise, the proof is very close to that of Proposition 2.1 and is therefore omitted.

The final step is to show that the value of unemployment across the two islands is equalized for an interior measure of workers which satisfies the conditions above.

Lemma 3.3. *If $n \geqslant \underline{n}$ and $\tilde{\sigma} \leqslant \hat{\sigma}$ then there exist $n^{H*} > 0$ and $n^{L*} > 0$ where $n^{H*} + n^{L*} = n$ and $U^H(n^{H*}) = U^L(n^{L*}) = U^*$.*

Proof. See Appendix A. □

This completes the proof of Proposition 3.1.

3.3. Predictions

This section presents the predictions of the extended model.

Proposition 3.2. *In equilibrium, high productivity firms choose a higher level of human resources: $h^H > h^L$.*

Proof. See Appendix A. □

The intuition for this result is as follows. The production function exhibits complementarities between the firm's productivity and match quality which means that high productivity firms face a greater opportunity cost of being badly matched. Therefore, high productivity firms have a greater incentive to limit the probability of hiring a worker who turns out to be a bad match and choose a higher level of human resources than low productivity firms.

Barron et al. (1987) report that larger firms spend more for screening per hire and Marsden (1994a) reports that large firms use more formal methods to screen workers. To the extent that firm size is positively correlated with firm productivity, this finding is consistent with the prediction. Section 4.2 describes an explicit model where firm productivity and size are positively correlated.

The previous result leads naturally to the following propositions.

Proposition 3.3. *In equilibrium, conditional on the firm's type:*

1. *A worker who is hired through a referral receives a higher wage than a worker hired through the market.*
2. *A worker who is hired through a referral produces more output than a worker hired through the market.*
3. *A worker who is hired through a referral has a lower separation rate than a worker hired through the market.*
4. *The differences in wages and separation rates across hiring channels decline with the workers' tenure on the job.*

Proposition 3.3 follows from $p_R^t > p_M^t$.

Proposition 3.4. *In equilibrium, if $\gamma \geqslant \frac{1}{2}$:*

1. *Low productivity firms exhibit greater prevalence of referrals than high productivity firms: $\mathcal{P}^L > \mathcal{P}^H$.*
2. *The wage premium of a referred worker is lower if the firm type is not controlled for than if it is.*
3. *If $y^L < \epsilon y^H$ for some $\epsilon \in (0, 1)$ then the average wage of a worker hired through a referral is lower than the average wage of a worker hired through the market.*

Proof. See Appendix A. □

Proposition 3.3 is a restatement of Proposition 2.4 for the case of firm heterogeneity. The prediction of Proposition 3.4 that low productivity firms use referrals to a larger extent is supported by Holzer (1987) and Marsden (1994b) who find that larger firms use more formal methods to hire workers and Pellizzari (2010) who finds that workers who report finding their jobs through a referral work at smaller firms.

Propositions 3.3 and 3.4 are consistent with the empirical finding that a referral is associated with a higher wage in studies that control for firm fixed effects and with a lower wage in studies that do not control for firm fixed effects. This finding is most clearly demonstrated in Dustmann et al. (2011) who perform wage regressions with and without firm fixed effects in a large matched employer–employee data set. They find that their estimate for the effect of a referral on the wage is positive when firm fixed effects are included but negative when they are not. The change in sign is not affected by whether worker fixed effects are included.

The findings of Dustmann et al. (2011) are supported by the contrast between firm-level studies, where the firm effect is controlled by design and where a positive correlations between a referral and the wage is reported (Brown et al., 2012) or productivity (Castilla, 2005; Pinkston, 2012), and studies where firm fixed effects are not included and a negative correlation is reported (Pistaferri, 1999; Pellizzari, 2010; Bentolila et al., 2010). It should be remarked, that including firm size, rather than firm fixed effects, in the wage regression still generates a negative correlation of wage and a referral, although the magnitude is smaller than if firm size is not included (Pellizzari, 2010).

What is more, Propositions 3.2, 3.3 and 3.4 provide an interpretation for why these studies find seemingly opposing conclusions: the informational friction in the market can be alleviated in an informal way (using referrals) or a formal way (investing in better interviewing) and firms sort according to productivity in their preferred option: low productivity firms find it easier to rely on referrals while high productivity firms improve their interviewing. In other words, there is no contradiction in the finding that the firms with greater value for “better” workers use referrals to a lesser extent, even though referrals are associated with “better” workers.

Finally, the model provides an additional prediction.

Proposition 3.5. *In equilibrium, the differentials in productivity, wages and separation rates between workers who were hired through a referral or the market are larger in low productivity firms.*

This is a direct consequence of the result that high productivity firms choose higher levels of human resources leading to smaller differences between newly-hired referred and non-referred workers. I am not aware of any publicly available work that has examined this prediction but evaluating it would be a useful test of this paper's theory.

4. Extensions

In this section a number of extensions are sketched and discussed: alternative modeling of the hiring process, non-trivial firm size and life-cycle dynamics of the firm. These extensions are intended to be suggestive rather than fully-developed equilibrium analyses.

4.1. An alternative model of the hiring process

This section provides an explicit micro-structure for the hiring process of Section 2 in an overall simplified model. The principal difference is that many workers simultaneously apply for the same job through both the market and the referral channel.[19]

Consider a discrete time model with the following three stages: first, workers apply for jobs; second, firms decide which worker to hire; third, production takes place and payments are made. The three stages are described in turn.

There are u unemployed workers and v firms. Every worker applies for one job. The application is sent through the market channel with probability ϕ_M and through the referral channel with probability $\phi_R = 1 - \phi_M$. The probabilities ϕ_M and ϕ_R depend on the firms' actions, as described below.

Conditional on the channel through which the application is sent, the recipient firm is chosen at random according to an urn-ball process. Firm j receives a_M applicants through the market channel and a_R applicants through the referral channel where a_M and a_R are random variables which are distributed according to Poisson distributions with parameters ψ_M and ψ_R, respectively. The parameters depend on the firm's search effort and the aggregate effort of all other firms according to the following expression:

$$\psi_i = \frac{e_i}{E_i}\frac{E_i}{E_M + E_R}\frac{u}{v},$$

where e_M and e_R denote the effort that firm j exerts in searching through the market and referrals, respectively, and E_M and E_R denote the aggregate effort levels. The cost of the search effort for firm j is given by $\frac{e_M^2 + e_R^2}{2}$. Attention is focused on equilibria where $e_i = E_i$ for all firms and $i \in \{M, R\}$.

The probability that an application is sent through channel i depends on the firms' aggregate effort levels according to:

$$\phi_i = \frac{E_i}{E_M + E_R}.$$

The information structure is the same as in Section 2. Denote the probability that a match is good conditional on signal g_i or b_i (i is the channel of meeting) by:

$$p_{gi} = \frac{\gamma q_{Gi}}{\pi_{gi}},$$

$$p_{bi} = \frac{\gamma(1 - q_{Gi})}{\pi_{bi}},$$

[19] I thank an anonymous referee for suggesting that I explore this route.

recall that the assumptions on signal accuracy imply that:

$$p_{gR} > p_{gM} > p_{bM} > p_{bR}.$$

A worker hired through channel i after signal s_i produces y with probability p_{si} and zero with the complementary probability. His outside option is the value of being unemployed which is equal to z. His wage is determined through Nash bargaining prior to the realization of output and is denoted by w_{si}. Denoting the worker's bargaining power by β we have: $w_{si} = \beta(p_{si} y - z)$.

The firm's expected profits from hiring a worker of type si are given by:

$$J_{si} = p_{si} y - w_{si} = (1-\beta) p_{si} y - \beta z,$$

which implies that:

$$J_{gR} > J_{gM} > J_{bM} > J_{bR},$$

and determines firms' ranking of applicants.

Furthermore, if

$$z \geqslant \frac{1-\beta}{\beta} \frac{y\gamma(1-q_{GM})}{\gamma(1-q_{GM}) + (1-\gamma) q_{BM}},$$

then a match is never formed when the worker emits a bad signal.

The optimal policy for the firm is to hire one of the referred applicants who emitted a good signal and, if there are none, to hire one of the market applicants who emitted a good signal. If no applicant has emitted a good signal, then the firm remains idle.

Note that, since each worker's type and signal is independent from other workers, the number of applicants through channel i who emit a good signal is distributed according to a Poisson distribution with parameter $\psi_i \pi_i$. The value of firm j which exerts search effort e_M and e_R is therefore given by:

$$\tilde{V}(e_M, e_R) = \left(1 - e^{-\psi_R^j \pi_R}\right) J_{gR} + e^{-\psi_R^j \pi_R}\left(1 - e^{-\psi_M^j \pi_M}\right) J_{gM} - \frac{e_M^2 + e_R^2}{2}.$$

The optimal choice of search effort delivers the value of a vacant firm:

$$V = \max_{e_M, e_R} \tilde{V}(e_M, e_R).$$

The free entry condition determines the number of firms in the market:

$$V = K.$$

The probability that a particular worker finds employment is now calculated. A worker who applies through a referral is hired if he emits a good signal and he is chosen among the referred applicants who emit a good signal. A worker who applies through the market is hired if there are no referred applicants who emit a good signal, he emits a good signal and he is chosen among the applicants through the market who emit a good signal.

The probability that an applicant through channel i emits a good signal and is chosen among the workers of the same type (channel i, good signal) is equal to:

$$\pi_{gi} \sum_{n=1}^{\infty} \frac{(\psi_i \pi_{gi})^n e^{-\psi_i \pi_{gi}}}{n!} = \frac{\pi_{gi}(1 - e^{-\psi_i \pi_{gi}})}{\psi_i \pi_{gi}}.$$

Therefore, the probability that an applicant is hired through a referral (H_R) or the market (H_M) is given by:

$$H_R = \frac{\pi_{gR}(1 - e^{-\psi_R \pi_{gR}})}{\psi_R \pi_{gR}},$$

$$H_M = \frac{\pi_{gM} e^{-\psi_R \pi_{gR}}(1 - e^{-\psi_M \pi_{gM}})}{\psi_M \pi_{gM}}.$$

The specification presented here leads to very similar results as Section 2 but provides a more explicit description of the hiring process. One difference is that the condition for referred applicants to be hired more often is weaker than in Section 2, where $\pi_{gR} > \pi_{gM}$ was necessary. Noting that $\frac{1-e^{-\psi_{gR}\pi_{gR}}}{\psi_{gR}\pi_{gR}} > e^{-\psi_{gR}\pi_{gR}}$ and $\frac{1-e^{-\psi_M \pi_{gM}}}{\psi_M \pi_{gM}} < 1$, it is clear that the condition is weaker in this specification. This is due to the fact that applicants are applying simultaneously and therefore the referred applicants' advantage is necessarily putting the market applicants at a disadvantage.

This extension provides a more explicit description of the hiring process. The main qualitative features are similar to the main model except that the condition for more frequent hiring of referred applicants has been relaxed.

4.2. Firm size

The model of Section 3 can be extended in a straightforward way to examine the predictions regarding the use of referrals by firm size.

A firm of type $t \in \{H, L\}$ operates a constant returns to scale production technology which has (potential) productivity y^t per worker and where actual output depends on match quality as before. There is free entry of firms, each firm incurs flow operation cost $K_f(f^t)$, where f^t is the number of firms in island t, and a firm can post multiple vacancies at cost $K_v(v)$ for v vacancies. $K_f(\cdot)$ and $K_v(\cdot)$ are strictly increasing and convex with $K_l(0) = K_l'(0) = 0$, $l \in \{f, v\}$.

A firm's vacancies operate independently from each other: each vacancy chooses the search effort levels (e_M, e_R) and signal accuracy (h) separately, is contacted independently by workers and, when filled, it closes down at an independent rate, for endogenous or exogenous reasons (learning or jobs destruction, respectively). The marginal cost of posting an additional vacancy depends on the total number of vacancies posted by the firm, which will determine firm size.

The value of a type-t firm that posts v vacancies is:

$$rV^t(v) = -K_f\big(f^t\big) - K_v(v) + v\bar{S}^t, \tag{25}$$

where $\bar{S}^t$ denotes the flow value of a posted vacancy and is defined similarly to Section 3:

$$\bar{S}^t = \max_{e_M, e_R, h} \alpha_{FM}^t(e_M)\big(\pi_M(h)\big(J_M^t(h) - \bar{S}^t\big) - s(h)\big) + \alpha_{FR}^t(e_R)\big(\pi_R(h)\big(J_R^t(h) - \bar{S}^t\big) - s(h)\big) - \frac{e_M^2 + e_R^2}{2}.$$

The optimal number of vacancies for a type-t firm is given by:

$$K_v'(v) = \bar{S}^t, \tag{26}$$

and attention is focused on symmetric equilibria where every vacancy makes the same choices and each firm posts the same number of vacancies conditional on type.

A vacancy of a high productivity firm is clearly more valuable than that of a low productivity firm ($\bar{S}^H > \bar{S}^L$). This implies that high productivity firms post more vacancies than low productivity firms ($v^H > v^L$) and, therefore, have larger steady state size.

The predictions of Section 3.3 regarding the use of referrals and firm productivity extend to firm size: larger firms use higher levels of human resources and hire through the market to larger extent. These predictions are consistent with the evidence presented in that section.

The model of the firm is quite stark: a firm is simply a collection of vacancies. Alternatively one could model a more elaborate production structure where there is more interaction among workers, such as production in teams. This might be particularly interesting in the case where referred workers fit better (on average) to the existing workers. Such a model would generate an additional incentive for firms to hire through referrals – or perhaps a micro-foundation for the better match quality of referred workers.

4.3. Life-cycle of the firm

This section introduces firm exit and entry into the framework of Section 4.2 in order to examine the use of referrals by firms at different points in their life-cycle.

Assume that a type-t firm exits at rate ϵ^t, where $\epsilon^L > \epsilon^H$, and that new firms enter with no employees. Therefore, a job at a type-t firm is exogenously destroyed at rate $\epsilon^t + \delta$ rather than δ.

In this formulation, a firm's stock of employees does not affect its decision of how many vacancies to post. Therefore, the optimal choice of how many vacancies to post is given by Eq. (26), with the value of a filled job appropriately adjusted for the different destruction rate, regardless of the firm's age or stock of employees. As a result, a firm's hiring rate only depends on its type.

The total attrition of a continuing firm is equal to the rate at which an individual worker separates from the firm for exogenous (δ) or endogenous $(\lambda(1 - p_i^t))$ reasons times the firm's stock of employees. Since a new firm has no workers, it experiences more hires than separations and grows in size. Eventually, it reaches its stationary size where the new hires exactly replace the separations (as in Section 4.2), unless it exits in the meantime.

Furthermore, low productivity firms are assumed to have higher exit rates which means that they have higher entry rates in steady state. This extension's steady state predictions are consistent with the broad evidence on the firm's life-cycle: new firms start small, grow faster than large firms and are more likely to exit (due to the fact that low productivity firms are over-represented in new entrants).

In this extension, older firms are larger on average, both conditional on type and because of selection of high types: older firms of both types have had more time to reach their "optimal size" where new hires simply replace workers who separated; furthermore, the proportion of high type firms increases with age due to their lower exit rates and high-type firms are on average larger. Therefore, older firms use referrals to a lesser extent than younger firms. Notice that in this simple model a firm's usage of referrals does not change with its own age, but rather the firm composition by age changes.

The predictions of Section 3.3 can, again, be extended, this time with respect to firms' age: older firms use higher levels of human resources and hire through the market to larger extent. I am not aware of any evidence that relates a firm's age to the use of referrals, over and above the correlation between size and age.

5. Conclusions

This paper presents a model where firms choose how intensely to search through referrals vs the market and the principal benefit of using referrals is that they provide more accurate signals regarding a worker's suitability for the job. The baseline model captures a large number of stylized facts regarding the correlation of referrals with wages, productivity, separation rates and the interaction of these variables with tenure.

The extension to include firm heterogeneity demonstrates that this framework can be used to examine the observed regularities in referral use across different firm types. When the signal's accuracy is an endogenous choice, high productivity firms will choose to improve the signal to a larger extent, thereby reducing the benefit of referrals. As a result, high productivity firms use referrals to a lesser extent which is consistent with the evidence. Furthermore, it provides an interpretation for why the correlation between wages and referrals is positive when firm types are controlled for but negative when they are ignored.

This framework provides a first step in studying the large differentials in referral use across industries and occupations. According to the model, greater use of referrals is driven by a larger informational advantage for referrals which also implies larger differentials in wages, productivity and separation rates between referred and non-referred workers. Some supportive evidence for this prediction is present in Dustmann et al. (2011) and Brown et al. (2012) but this is clearly an issue that needs to be studied further. Specifically, it is unclear why jobs differ in the amount of information that can be transmitted through a referral. One possibility is that the difference in signal accuracy might be due to the type of skills that are required for the job: perhaps some jobs require skills that are easy to observe and convey for the referrer (e.g. reliability) while the skills required for more complex jobs are harder to observe or convey (e.g. creativity), reducing the usefulness of referrals.

This framework can also be used to examine the use of referrals over the business cycle. Galenianos (2012) shows that the use of referrals might be helpful in interpreting the apparent pro-cyclical movements in matching efficiency and the present model can further analyze possible variations in the intensity of use. Depending on the specific assumptions, referrals might be used more or less intensely during a boom. Supposing that signal accuracy is fixed over the business cycle, then an increase in aggregate productivity (y in the model) would increase the intensity of referral search because of the informational advantage of that channel. If, however, one assumes that human resources can be adjusted at the business cycle frequency, then a boom leads to higher choice of h and, consequently, heavier use of the market. Which set of assumptions is better suited requires more data than is currently available regarding the cyclical behavior of the prevalence of referrals.

Appendix A

Lemma 2.2. *In equilibrium, the value of unemployment is uniquely determined as a function of the expected match surplus and the measure of vacancies by the solution to*:

$$\frac{n}{v} = \left(\frac{\mu\beta(\frac{\bar{S}_M^2}{c_M} + \frac{\bar{S}_R^2}{c_R})}{(rU - z)(\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R})}\right)^{\frac{1}{\eta}} + \left(\frac{\mu\beta(\frac{\bar{S}_M^2}{c_M} + \frac{\bar{S}_R^2}{c_R})}{(rU - z)(\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R})}\right)^{\frac{1-\eta}{\eta}} \frac{\mu(\frac{\Gamma_{1M}\bar{S}_M}{c_M} + \frac{\Gamma_{1R}\bar{S}_R}{c_R})}{\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R}}, \tag{27}$$

where:

$$\Gamma_{1i} = \frac{\gamma q_{Gi}}{\delta} + \frac{(1-\gamma)(1-q_{Bi})}{\delta + \lambda}.$$

Proof. The steady state conditions (9), (10) and (11) can be rearranged as follows:

$$n_i = \frac{u\alpha_W E_i \pi_i}{(\delta+\lambda)(E_M + E_R)}, \quad i \in \{M, R\},$$

$$n_G = \frac{u\alpha_W \lambda}{\delta(\delta+\lambda)}\left(\frac{E_M \gamma q_{GM}}{E_M + E_R} + \frac{E_R \gamma q_{GR}}{E_M + E_R}\right) \quad \Rightarrow \quad n_M + n_R + n_G = u\alpha_W \frac{E_M \Gamma_{1M} + E_R \Gamma_{1R}}{E_M + E_R}. \tag{28}$$

Combining Eqs. (8), (15) and (28) yields:

$$n = u + \mu v^\eta u^{1-\eta} \frac{\frac{\Gamma_{1M}\bar{S}_M}{c_M} + \frac{\Gamma_{1R}\bar{S}_R}{c_R}}{\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R}}. \tag{29}$$

The right-hand side of Eq. (29) is equal to zero when $u = 0$, is strictly increasing in u and is greater than n when $u = n$. Therefore, Eq. (29) uniquely determines the measure of unemployed given the measure of vacancies and expected match surplus.

Rearrange Eq. (16) as follows:

$$u = \left(\frac{\mu v^\eta \beta(\frac{\bar{S}_M^2}{c_M} + \frac{\bar{S}_R^2}{c_R})}{(rU - z)(\frac{\bar{S}_M}{c_M} + \frac{\bar{S}_R}{c_R})} \right)^{\frac{1}{\eta}},$$

and introduce this expression inside Eq. (29) to get Eq. (27). The right-hand side of Eq. (27) approaches infinity as $U \to z/r$, is strictly decreasing in U and it approaches zero as $U \to +\infty$. Therefore, there is a unique U that satisfies Eq. (27). □

Lemma 2.3. *In equilibrium*:

1. *The value of unemployment is determined as a function of the measure of vacancies,* $U(v)$.
2. $\lim_{v\to 0} U(v) = \frac{z}{r}$ *and* $\lim_{v\to\infty} U(v) = \bar{U}_R$.
3. *There exists* $\bar{v}$ *such that* $v < \bar{v} \Rightarrow U(v) < \bar{U}_M$.
4. *There exists* $\underline{v}$ *such that if* $v > \underline{v}$ *then a bad signal does not lead to a match.*

Proof. 1. Define

$$\Psi(U, v) = (\mu\beta)^{\frac{1}{\eta}} \left(\frac{Q_1(U)}{rU - z} \right)^{\frac{1}{\eta}} + \mu^{\frac{1}{\eta}} \beta^{\frac{1-\eta}{\eta}} \left(\frac{Q_1(U)}{rU - z} \right)^{\frac{1-\eta}{\eta}} Q_2(U) - \frac{n}{v}, \tag{30}$$

where

$$Q_1(U) = \frac{\frac{\bar{S}_M(U)^2}{c_M} + \frac{\bar{S}_R(U)^2}{c_R}}{\frac{\bar{S}_M(U)}{c_M} + \frac{\bar{S}_R(U)}{c_R}},$$

$$Q_2(U) = \frac{\frac{\Gamma_{1M}\bar{S}_M(U)}{c_M} + \frac{\Gamma_{1R}\bar{S}_R(U)}{c_R}}{\frac{\bar{S}_M(U)}{c_M} + \frac{\bar{S}_R(U)}{c_R}}.$$

Notice that $\Psi(U, v) = 0$ if and only if Eq. (27) holds. Furthermore:

$$\lim_{U\to z/r} \Psi(U, v) = +\infty. \tag{31}$$

When $U \in [\bar{U}_M, \bar{U}_R)$, we have $Q_1(U) = \bar{S}_R(U)$, $Q_2(U) = \Gamma_{1R}$ and Eq. (30) is simplified as follows:

$$\Psi(U, v) = (\mu\beta)^{\frac{1}{\eta}} \left(\frac{\bar{S}_R(U)}{rU - z} \right)^{\frac{1}{\eta}} + \mu^{\frac{1}{\eta}} \beta^{\frac{1-\eta}{\eta}} \left(\frac{\bar{S}_R(U)}{rU - z} \right)^{\frac{1-\eta}{\eta}} \Gamma_{1R} - \frac{n}{v}. \tag{32}$$

If $U \geqslant \bar{U}_R$ then $Q_1(U) = 0$ and:

$$\Psi(U, v) = -\frac{n}{v} < 0. \tag{33}$$

These observations together with the continuity of $\Psi(U, v)$ prove that there exists $U(v)$ such that $\Psi(U(v), v) = 0$ which therefore defines the value of unemployment as a function of the measure of vacancies. Note that it is the ratio of workers to vacancies that matters.

2. Eq. (31) together with $\lim_{v\to 0} \Psi(U, v) = -\infty$ implies that $\lim_{v\to 0} U(v) = z/r$. Eq. (33) together with

$$\lim_{v\to\infty} \Psi(U, v) = (\mu\beta)^{\frac{1}{\eta}} \left(\frac{Q_1(U)}{rU - z} \right)^{\frac{1}{\eta}} + \mu^{\frac{1}{\eta}} \beta^{\frac{1-\eta}{\eta}} \left(\frac{Q_1(U)}{rU - z} \right)^{\frac{1-\eta}{\eta}} Q_2(U) \geqslant 0,$$

implies that $\lim_{v\to\infty} U(v) = \bar{U}_R$.

3. From Eq. (32), $\Psi(U, v)$ is strictly decreasing in U so long as $U \in (\bar{U}_M, \bar{U}_R)$. If $U(v) < \bar{U}_M$ then a meeting where a good signal is generated leads to a match regardless of the channel of search. A sufficient condition is $\Psi(\bar{U}_M, v) < 0$ which corresponds to $v < \bar{v}$.

4. We now show that matches are not formed after a bad signal if $U \geqslant \underline{U}_M$ and that $\bar{U}_M > \underline{U}_M$ so that an equilibrium is possible. Denote the probability the match quality is good after a bad signal through channel i by:

$$p_{bi} = \frac{\gamma(1 - q_{Gi})}{\gamma(1 - q_{Gi}) + (1 - \gamma)q_{Bi}}.$$

Assumption (1) implies that $q_{GR} - q_{BR} > q_{GM} - q_{BM}$ which in turn means that $p_{bM} > p_{bR}$.

The surplus of hiring a worker through channel i after a bad signal is:

$$S_{bi} = \frac{yp_{bi}}{r+\delta} - \frac{rU}{r+\delta}\frac{r+\delta+\lambda p_{bi}}{r+\delta+\lambda}.$$

Noting that $S_{bM} > S_{bR}$, it is suboptimal to form any match after a bad signal if

$$rU \geqslant r\underline{U}_M = \frac{y\gamma(1-q_{GM})(r+\delta+\lambda)}{\gamma(1-q_{GM})(r+\delta+\lambda)+(1-\gamma)q_{BM}(r+\delta)}$$

and it can be readily verified that $\overline{U}_M > \underline{U}_M$.

If $\Psi(\underline{U}_M, v) > 0 > \Psi(\overline{U}_M, v)$, then $U(v) \in (\underline{U}_M, \overline{U}_M)$ and the desired result holds. This inequality can be straightforwardly (if tediously) be rewritten in terms of $v \in (\underline{v}, \bar{v})$. □

Lemma 2.4. *If $K \in (\underline{K}, \overline{K})$ there exists v such that $rV = 0$, matches occur through both channels after a good signal and a bad signal does not lead to a match.*

Proof. Define:

$$\Phi(v) = -K + \frac{1}{2}(1-\beta)\alpha_F \frac{\frac{[\bar{S}_M(U(v))]^2}{c_M} + \frac{[\bar{S}_R(U(v))]^2}{c_R}}{\frac{\bar{S}_M(U(v))}{c_M} + \frac{\bar{S}_R(U(v))}{c_R}},$$

so that $\Phi(v) = 0 \Leftrightarrow rV = 0$.

When $v \to 0$ a firm meets with a worker instantaneously and pays no vacancy costs. Furthermore, $\lim_{v\to 0} U(v) = \frac{z}{r}$ and in that case the surplus of a match is positive because $z < r\overline{U}_M < r\overline{U}_R$. Therefore:

$$\lim_{v\to 0} \Phi(v) > 0.$$

As $v \to \infty$, the surplus of a meeting through either channel goes to zero because $\lim_{v\to\infty} U(v) = \overline{U}_R$. Therefore:

$$\lim_{v\to\infty} \Phi(v) = -K < 0.$$

By the continuity of $\Phi(v)$, there exists v such that $\Phi(v) = 0$. Furthermore, the requirement of $v \in (\underline{v}, \bar{v})$ corresponds to $K \in (\underline{K}, \overline{K})$. □

Proposition 2.4. *If*

$$\frac{c_R}{c_M} > \frac{(q_{GR}-(1-q_{BR}))(\gamma q_{GR}+(1-\gamma)(1-q_{BR}))}{(q_{GM}-(1-q_{BM}))(\gamma q_{GM}+(1-\gamma)(1-q_{BM}))}, \tag{34}$$

and $K \geqslant \hat{K}$ for some $\hat{K}$ where $\hat{K} < \overline{K}$ then more workers are hired through the market than referrals ($\mathcal{P} < \frac{1}{2}$).

Proof. Notice that $U = \overline{U}_M \Rightarrow E_M = 0 \Rightarrow \alpha_{FM} = 0$. Therefore, the value of unemployment needs to be low enough for $\mathcal{P} > \frac{1}{2}$. A condition is derived such that $\mathcal{P} < \frac{1}{2}$ when $rU = z$. Then the upper bound on U is determined such that $\mathcal{P} < \frac{1}{2}$.

Observe that

$$\frac{\alpha_{FM}\pi_M}{\alpha_{FR}\pi_R} = \frac{\alpha_F \frac{E_M}{E_M+E_R}\pi_M}{\alpha_F \frac{E_R}{E_M+E_R}\pi_R} = \frac{\frac{\bar{S}_M}{c_M}\pi_M}{\frac{\bar{S}_R}{c_R}\pi_R}.$$

Furthermore:

$$\frac{\bar{S}_i}{c_i} = \frac{(y-rU)\gamma}{r+\delta}\frac{q_{Gi}}{c_i} - \frac{rU(1-\gamma)}{r+\delta+\lambda}\frac{1-q_{Bi}}{c_i},$$

and therefore when $rU = z$ the following holds:

$$\alpha_{FM}\pi_M > \alpha_{FR}\pi_R \quad\Leftrightarrow\quad \frac{(y-z)\gamma}{r+\delta}\left[\frac{q_{GM}\pi_M}{c_M} - \frac{q_{GR}\pi_R}{c_R}\right] > \frac{z(1-\gamma)}{r+\delta+\lambda}\left[\frac{(1-q_{BM})\pi_M}{c_M} - \frac{(1-q_{BR})\pi_R}{c_R}\right].$$

The assumption that $\gamma \geqslant \bar{\gamma}$ means that $\frac{(y-z)\gamma}{r+\delta} \geqslant \frac{z(1-\gamma)}{r+\delta+\lambda}$ and the terms in the square brackets above are equivalent to Eq. (34).

Finally, the highest level of U such that $\mathcal{P} \leqslant \frac{1}{2}$ is given by the solution to:

$$\frac{\pi_M}{c_M}\big(y\Gamma_{2M} - r\hat{U}(\Gamma_{2M}+\Gamma_{3M})\big) = \frac{\pi_R}{c_R}\big(y\Gamma_{2R} - r\hat{U}(\Gamma_{2R}+\Gamma_{3R})\big)$$
$$\Rightarrow \quad r\hat{U} = \frac{y(\frac{\pi_M}{c_M}\Gamma_{2M} - \frac{\pi_R}{c_R}\Gamma_{2R})}{\frac{\pi_M}{c_M}(\Gamma_{2M}+\Gamma_{3M}) - \frac{\pi_R}{c_R}(\Gamma_{2R}+\Gamma_{3R})}.$$

Therefore, if $U < \hat{U}$ then there are more hires through the market channel. For $U < \hat{U}$ the number of vacancies need to be below a threshold which corresponds to $K \geqslant \hat{K}$.

The last step is to check that $U < \hat{U}$ can occur in equilibrium. Specifically, this means checking that $\hat{U} > \underline{U}_M$ since $U > \underline{U}_M$ is a necessary condition for equilibrium:

$$\frac{y(\frac{\pi_M}{c_M}\Gamma_{2M} - \frac{\pi_R}{c_R}\Gamma_{2R})}{\frac{\pi_M}{c_M}(\Gamma_{2M}+\Gamma_{3M}) - \frac{\pi_R}{c_R}(\Gamma_{2R}+\Gamma_{3R})} > \frac{y\gamma(1-q_{GM})(r+\delta+\lambda)}{\gamma(1-q_{GM})(r+\delta+\lambda)+(1-\gamma)q_{BM}(r+\delta)}.$$

It is a matter of tedious algebra to show that if Eq. (34) holds, then $\hat{U} > \underline{U}_M$. □

Lemma 3.1. *In equilibrium*:

1. *The optimal choice of h of a type-t firm is determined as a function of the value of unemployment*: $h^t(U^t)$.
2. *The optimal choice of h is decreasing in* σ.
3. *The expected surplus of a meeting in island t is determined as a function of the value of unemployment*: $\hat{S}_i^t(U^t)$.

Proof. 1. Eq. (22) can be rewritten as follows (see Section 2):

$$\hat{S}_i^t(h) = \max\left[\frac{y^t\gamma q_i(h)}{r+\delta} - rU^t\left(\frac{\gamma q_i(h)}{r+\delta} + \frac{(1-\gamma)(1-q_i(h))}{r+\delta+\lambda}\right) - \tilde{\sigma}h, 0\right]. \tag{35}$$

Note that:

$$\hat{S}_i^{t\prime}(h) = q_i'(h)\left(\frac{y^t\gamma}{r+\delta} - rU^t\left(\frac{\gamma}{r+\delta} - \frac{1-\gamma}{r+\delta+\lambda}\right)\right) - \tilde{\sigma},$$
$$\hat{S}_i^{t\prime\prime}(h) = q_i''(h)\left(\frac{y^t\gamma}{r+\delta} - rU^t\left(\frac{\gamma}{r+\delta} - \frac{1-\gamma}{r+\delta+\lambda}\right)\right) < 0.$$

The assumptions about $q_i^t(h)$ mean that $\hat{S}_i^{t\prime}(0) > 0 > \lim_{h\to\infty} \hat{S}_i^{t\prime}(h)$. Define $h_i^t(U^t)$ by $\hat{S}_i^{t\prime}(h_i^t(U^t)) = 0$ and note that $h_i^t(U^t)$ also maximizes $[\hat{S}_i^t(h)]^2$.

Define $\overline{U}_i^t$ as follows:

$$r\overline{U}_i^t = \frac{\frac{y\gamma q_i^t(h_i^t(\overline{U}_i^t))}{r+\delta} - \tilde{\sigma}h_i^t(\overline{U}_i^t)}{\frac{\gamma q_i^t(h_i^t(\overline{U}_i^t))}{r+\delta} + \frac{(1-\gamma)(1-q_i^t(h_i^t(\overline{U}_i^t)))}{r+\delta+\lambda}},$$

and notice that $\overline{U}_i^t$ exists and it is unique. If $U^t \geqslant \overline{U}_i^t$ then $\hat{S}_i^t(h) \leqslant 0$ for all h. If $U^t < \overline{U}_i^t$ then $\hat{S}_i^t(h) > 0$ for $h \in (\underline{h}_i^t(U^t), \bar{h}_i^t(U^t))$ and $\hat{S}_i^t(h) = 0$ otherwise. The assumptions on $q_M(h)$ and $q_R(h)$ imply that $\overline{U}_M^t < \overline{U}_R^t$, $h_R^t(U^t) < h_M^t(U^t)$ and, when applicable, $\bar{h}_M^t(U^t) < \bar{h}_R^t(U^t)$ and $\underline{h}_M^t(U^t) \geqslant \underline{h}_R^t(U^t)$.

Define

$$\Omega^t\big(U^t, h\big) = \big[\hat{S}_M^t(h)\big]^2 + \big[\hat{S}_R^t(h)\big]^2. \tag{36}$$

If $U^t \geqslant \overline{U}_R^t$ then $\Omega^t(U^t, h) = 0$ and $h^t(U^t) = 0$. When $U^t \in [\overline{U}_M^t, \overline{U}_R^t)$ then $\Omega^t(U^t, h) = [\hat{S}_R^t(h)]^2$ and $h^t(U^t) = h_R^t(U^t)$. If $U^t < \overline{U}_M^t$, $\Omega^t(h)$ is strictly positive on $(\underline{h}_R^t(U^t), \bar{h}_R^t(U^t))$ and it is strictly increasing on $(\underline{h}_R^t(U^t), h_R^t(U^t))$, strictly decreasing on $(h_M^t(U^t), \bar{h}_M^t(U^t))$ and $\Omega^{t\prime}(h_R^t(U^t)) > 0 > \Omega^{t\prime}(h_M^t(U^t))$. In this case, the optimal $h^t(U^t)$ is characterized by the root of

$$\Omega^{t\prime}\big(U^t, h\big) = 2\big(\hat{S}_M^t(h)\hat{S}_M^{t\prime}(h) + \hat{S}_R^t(h)\hat{S}_R^{t\prime}(h)\big), \tag{37}$$

which is maximized on $(h_R^t(U^t), h_M^t(U^t))$. Further, notice that if $\underline{h}_M^t(U^t) \leqslant h_R^t(U^t)$, then $\hat{S}_M^t(h^t(U^t)) > 0$. This occurs if $U < \hat{U}_M^t$ for some $\hat{U}_M^t$ which will be the relevant condition for matches to occur through both channels.

2. It is now shown that Eq. (37) is decreasing in $\tilde{\sigma}$ when it is equal to zero. This suffices to prove that $h^t(U^t)$ is decreasing in $\tilde{\sigma}$ for the stable solutions,

$$\frac{\partial \Omega^{t\prime}(U^t,h)}{\partial \tilde{\sigma}} = 2\big(-h\hat{S}_M^{t\prime}(h) - \hat{S}_M^t(h) - h\hat{S}_R^{t\prime}(h) - \hat{S}_R^t(h)\big)$$
$$= 2\left(-\hat{S}_M^t - \hat{S}_R^t - h\hat{S}_M^{t\prime}\left(1 - \frac{\hat{S}_M^t}{\hat{S}_R^t}\right)\right) < 0,$$

where the second equality results from using the root of Eq. (37).

3. The expected surplus of a meeting through channel i is given by:

$$\hat{S}_i^t(U^t) = \max\left[\frac{y^t\gamma q_i(h^t(U^t))}{r+\delta} - rU^t\left(\frac{\gamma q_i(h^t(U^t))}{r+\delta} + \frac{(1-\gamma)(1-q_i(h^t(U^t)))}{r+\delta+\lambda}\right) - \sigma h^t(U^t), 0\right]. \quad \square$$

Lemma 3.3. *If $n \geqslant \underline{n}$ and $\sigma \leqslant \hat{\sigma}$ then there exist $n^{H*} > 0$ and $n^{L*} > 0$ where $n^{H*} + n^{L*} = n$ and $U^H(n^{H*}) = U^L(n^{L*}) = U^*$.*

Proof. It is clearly suboptimal for all workers to go to the low firm-type island. To rule out that every worker goes to the high-type island first recall that the vacancy cost at an island is an increasing function of the number of vacancies in that island. Therefore, if n is high enough ($n \geqslant \underline{n}$, for some $\underline{n}$) then the resulting measure of firms (and labor market tightness) is low enough so that $U^H(n) < \lim_{n^L \to 0} U^L(n^L)$ and therefore $n^{L*} > 0$.

The surplus of a match through channel i in island t after a bad signal is emitted is:

$$S_{bi}^t = \frac{y^t\gamma p_{bi}^t}{r+\delta} - \frac{rU^*}{r+\delta}\frac{r+\delta+\lambda p_{bi}^t}{r+\delta+\lambda},$$

where

$$p_{bi}^t = \frac{\gamma(1-q_i(h^t))}{\gamma(1-q_i(h^t)) + (1-\gamma)q_i(h^t)}.$$

If $q_i(h^t)$ is high enough then $S_{bi}^t \leqslant 0$ for all i and t. Additionally, if $q_M(h^t)$ is high enough, then $\hat{S}_M^t > 0$ and a hire occurs through both channels. Both conditions are satisfied if σ is low enough: $\sigma \leqslant \hat{\sigma}$ for some $\hat{\sigma}$. $\square$

Proposition 3.2. *In equilibrium, high productivity firms choose a higher level of human resources: $h^H > h^L$.*

Proof. The choice of h is characterized by the root of the following equation:

$$\frac{1}{2}\Omega' = \hat{S}_M\hat{S}_M' + \hat{S}_R\hat{S}_R', \tag{38}$$

where the type superscripts and arguments are omitted for notational convenience. It will be shown that Eq. (38) is increasing in y at its root. This shows that the optimal choice of h is increasing in y.

The derivative with respect to y is given by

$$\frac{1}{2}\frac{\partial\Omega'}{\partial y} = \frac{\partial\hat{S}_M}{\partial y}\hat{S}_M' + \frac{\partial\hat{S}_M'}{\partial y}\hat{S}_M + \frac{\partial\hat{S}_R}{\partial y}\hat{S}_R' + \frac{\partial\hat{S}_R'}{\partial y}\hat{S}_R$$
$$= \frac{\gamma}{r+\delta}\big(q_M\hat{S}_M' + q_M'\hat{S}_M + q_R\hat{S}_R' + q_R'\hat{S}_R\big).$$

At the root of Eq. (38):

$$\hat{S}_R' = -\frac{\hat{S}_M\hat{S}_M'}{\hat{S}_R},$$

where $\hat{S}_M' > 0 > \bar{S}_R'$ because $q_R > q_M$. Therefore:

$$\Omega' = 0 \quad\Rightarrow\quad \frac{1}{2}\frac{\partial\Omega'}{\partial y} = \frac{\gamma}{r+\delta}\left(\frac{\hat{S}_M'}{\hat{S}_R}(q_M\hat{S}_R - q_R\hat{S}_M) + q_M'\hat{S}_M + q_R'\hat{S}_R\right).$$

Finally:

$$q_M\hat{S}_R - q_R\hat{S}_M = q_M\left(\frac{yq_R}{r+\delta} - rU\left(\frac{q_R\gamma}{r+\delta} + \frac{(1-q_R)(1-\gamma)}{r+\delta+\lambda}\right) - h\right)$$
$$- q_R\left(\frac{yq_M}{r+\delta} - rU\left(\frac{q_M\gamma}{r+\delta} + \frac{(1-q_M)(1-\gamma)}{r+\delta+\lambda}\right) - h\right)$$
$$= \frac{rU(1-\gamma)}{r+\delta+\lambda}\big(q_R(1-q_M) - q_M(1-q_R)\big) + h(q_R - q_M) > 0,$$

which completes the proof. $\square$

Proposition 3.4. *In equilibrium, if* $\gamma \geqslant \frac{1}{2}$:

1. *Low productivity firms exhibit greater prevalence of referrals than high productivity firms:* $\mathcal{P}^L > \mathcal{P}^H$.
2. *The wage premium of a referred worker is lower if the firm type is not controlled for than if it is.*
3. *If* $y^L < \epsilon y^H$ *for some* $\epsilon \in (0,1)$ *then the average wage of a worker hired through a referral is lower than the average wage of a worker hired through the market.*

Proof. 1. Note that

$$\mathcal{P}^L > \mathcal{P}^H \quad \Leftrightarrow \quad \frac{\hat{S}_M^L \pi_M^L}{\hat{S}_R^L \pi_R^L} < \frac{\hat{S}_M^H \pi_M^H}{\hat{S}_R^H \pi_R^H}.$$

The following will be proven:

$$\frac{d}{dy}\left(\frac{\hat{S}_M \pi_M}{\hat{S}_R \pi_R}\right) > 0,$$

where the type superscripts and arguments are omitted to simplify notation.

First, note that:

$$\frac{d(\hat{S}_M/\hat{S}_R)}{dy} = \frac{\partial(\hat{S}_M/\hat{S}_R)}{\partial y} + \frac{\partial(\hat{S}_M/\hat{S}_R)}{\partial h}\frac{dh}{dy}.$$

As shown in Proposition 3.2:

$$\frac{\partial(\hat{S}_M/\hat{S}_R)}{\partial y} = \frac{1}{(\hat{S}_R)^2}\frac{\gamma}{r+\delta}(q_M \hat{S}_R - q_R \hat{S}_R) > 0.$$

Proposition 3.2 also proved that $\frac{dh}{dy} > 0$. Furthermore:

$$\frac{\partial(\hat{S}_M/\hat{S}_R)}{\partial h} = \frac{1}{(\hat{S}_R)^2}\Bigg[\left(q_M'(h)\left(\frac{y\gamma}{r+\delta} - rU\left(\frac{\gamma}{r+\delta} - \frac{1-\gamma}{r+\delta+\lambda}\right)\right) - 1\right)\hat{S}_R(h) - \left(q_R'(h)\left(\frac{y\gamma}{r+\delta} - rU\left(\frac{\gamma}{r+\delta} - \frac{1-\gamma}{r+\delta+\lambda}\right)\right) - 1\right)\hat{S}_M(h)\Bigg],$$

which is strictly positive because

$$\hat{S}_R > \hat{S}_M,$$
$$q_M'(h) > q_R'(h).$$

Finally:

$$\begin{aligned}\frac{d(\pi_M/\pi_R)}{dy} &= \frac{\partial(\pi_M/\pi_R)}{\partial h}\frac{dh}{dy} \\ &= \frac{dh}{dy}\frac{1}{(\pi_R)^2}(\pi_M'\pi_R - \pi_R'\pi_M) \\ &= \frac{dh}{dy}\frac{2\gamma-1}{(\pi_R^t)^2}\big((q_M' q_R - q_R' q_M)(2\gamma-1) + (q_M' - q_R')(1-\gamma)\big) > 0.\end{aligned}$$

2. When firm type is controlled for, the average wage premium of a referral at the time of hire is:

$$\Delta w_T = \frac{(w_R^H - w_M^H)(\alpha_M^H \pi_M^H + \alpha_R^H \pi_R^H) + (w_R^L - w_M^L)(\alpha_M^L \pi_M^L + \alpha_R^L \pi_R^L)}{\alpha_M^H \pi_M^H + \alpha_R^H \pi_R^H + \alpha_M^L \pi_M^L + \alpha_R^L \pi_R^L}.$$

When the firm type is not controlled for, the average wage premium of a referral is:

$$\Delta w_N = \frac{w_R^H \alpha_R^H \pi_R^H + w_R^L \alpha_R^L \pi_R^L}{\alpha_R^H \pi_R^H + \alpha_R^L \pi_R^L} - \frac{w_M^H \alpha_M^H \pi_M^H + w_M^L \alpha_M^L \pi_M^L}{\alpha_M^H \pi_M^H + \alpha_M^L \pi_M^L}.$$

It is a matter of algebra to show that:

$$\Delta w_T > \Delta w_N \quad \Leftrightarrow \quad (w_M^H + w_R^H - w_M^L - w_R^L)(\hat{S}_M^H \pi_M^H \hat{S}_R^L \pi_R^L - \hat{S}_R^H \pi_R^H \hat{S}_M^L \pi_M^L) > 0.$$

The term in the first parenthesis is positive because $w_i^H > w_i^L$. The term in the second parenthesis was shown to be positive in part 1 of this proposition. Therefore, when the firm type is not controlled for, the wage differential conflates the difference in firm productivity levels with the wage premium of referred workers, thereby reducing the wage premium of a referral.

3. For $\triangle w_N < 0$, the proportion of jobs that are found through a referral needs to be sufficiently higher for low productivity firms than high productivity firms. This corresponds to h^H being sufficiently higher that h^L. This occurs if y^H is sufficiently higher than y^L. $\square$

References

Barron, John M., Black, Dan A., Loewenstein, Mark A., 1987. Employer size: the implications for search, training, capital investment, starting wages and wage growth. Journal of Labor Economics 5 (1), 76–89.

Bayer, Patrick, Ross, Stephen L., Topa, Giorgio, 2008. Place of work and place of residence: informal hiring networks and labor market outcomes. Journal of Political Economy 116 (6), 1150–1196.

Bentolila, Samuel, Michelacci, Claudio, Suarez, Javier, 2010. Social contacts and occupational choice. Economica 77, 20–45.

Brown, Meta, Setren, Elizabeth, Topa, Giorgio, 2012. Do referrals lead to better matches? Evidence from a firm's employee referral system. Mimeo.

Castilla, Emilio J., 2005. Social networks and employee performance in a call center. American Journal of Sociology 110 (5), 1243–1283.

Datcher, Linda, 1983. The impact of informal networks on quit behavior. Review of Economics and Statistics 65 (3), 491–495.

Dustmann, Christian, Glitz, Albrecht, Schoenberg, Uta, 2011. Referral-based job search networks. Mimeo.

Fernandez, Roberto M., Weinberg, Nancy, 1997. Sifting and sorting: personal contacts and hiring in a retail bank. American Sociological Review 62, 883–902.

Fernandez, Roberto M., Castilla, Emilio J., Moore, Paul, 2000. Social capital at work: networks and employment at a phone center. American Journal of Sociology 105 (5), 1288–1356.

Galenianos, Manolis, 2012. Hiring through referrals. Mimeo.

Granovetter, Mark, 1995. Getting a Job: A Study of Contacts and Careers, 2nd edition. Harvard University Press, Cambridge, MA.

Holzer, Harry, 1987. Hiring procedures in the firm: their economic determinants and outcomes. NBER WP 2185.

Ioannides, Yannis, Datcher Loury, Linda, 2004. Job information networks, neighborhood effects, and inequality. Journal of Economic Literature XLII, 1056–1093.

Jovanovic, Boyan, 1979. Job matching and the theory of turnover. Journal of Political Economy 87 (5), 972–990.

Marsden, Peter V., 1994a. Selection methods in US establishments. Acta Sociologica 37, 287–301.

Marsden, Peter V., 1994b. The hiring process. American Behavioral Scientist 37 (7), 979–991.

Moscarini, Giuseppe, 2005. Job matching and the wage distribution. Econometrica 73 (2), 481–516.

Nagypal, Eva, 2007. Learning by doing vs. learning about match quality: can we tell them apart? Review of Economic Studies 74, 537–566.

Pellizzari, Michele, 2010. Do friends and relatives really help in getting a good job? Industrial and Labor Relations Review 63 (3), 494–510.

Pinkston, Joshua C., 2012. How much do employers learn from referrals? Industrial Relations 51, 317–341.

Pistaferri, Luigi, 1999. Informal networks in the Italian labor market. Giornale degli Economisti 58 (3–4), 355–375.

Pries, Michael, 2004. Persistence of employment fluctuations: a model of recurring job loss. Review of Economic Studies 71 (1), 193–215.

Simon, Curtis, Warner, John, 1992. Matchmaker, matchmaker: the effect of old boy networks on job match quality, earning and tenure. Journal of Labor Economics 10 (3), 306–330.

Topa, Giorgio, 2011. Labor markets and referrals. In: The Handbook of Social Economics. Elsevier/North-Holland, Amsterdam.