Measuring match quality using subjective data

Priscila Ferreira
Mark Taylor

December 2010
Measuring match quality using subjective data

Priscila Ferreira
NIMA and Department of Economics
University of Minho
Portugal
Email: priscila@eeg.uminho.pt

and

Mark Taylor
(corresponding author)
Institute for Social and Economic Research
University of Essex
Colchester CO4 3SQ
UK
Email: taylm@essex.ac.uk
Tel: +44(0)1206 873553
Fax: +44(0)1206 873151

23 December 2010

Abstract: We examine whether data routinely collected in household surveys and surveys of workers can be used to construct a measure of underlying match quality between worker and firm which helps test matching models and predict subsequent labour market outcomes of workers. We use subjective data from employees both on reported levels of job satisfaction with various aspects of the current job and on whether they would like a new job with a new employer to construct a measure of underlying match quality. We then use this to test several implications of matching models relating to wage-tenure profiles, wages, separations.

Keywords: panel data; wages; job mobility; match effects; BHPS.

JEL classification: C33; J28; J31; J62: J63

Acknowledgements: This research was partially funded by the UK Economic and Social Research Council (ESRC) through the Research Centre on Micro-Social Change (MiSoC) (award no. RES-518-28-001).
1. Introduction

Job matching models suggest that the quality of the match between a worker and a firm can explain a number of stylised facts in the labour market, such as the positive correlation between wages and job tenure and the inverse relationship between separation rates and job tenure. However the nature of most datasets does not allow us to directly identify or measure the value of match quality and tests of the predictions of matching model and of the impacts of matching on several labour market outcomes are still rare.

Our contribution in this paper is to examine whether data routinely collected in household surveys and surveys of workers can be used to construct a measure of underlying match quality which helps test matching models and predict subsequent labour market outcomes of workers. We use subjective data from employees both on reported levels of job satisfaction with various aspects of the current job and on whether they would like a new job with a new employer to construct a measure of the underlying match quality. We then use this to test several implications of matching models relating to wage-tenure profiles, wages, and separations.

The basis of matching models is that workers and firms are heterogeneous, which implies that workers differ in their suitability to different firms and consequently their productivity will vary across firms. The difficulty in workers identifying the characteristics of the firm with precision (and vice versa) generates issues related to selection, sorting, and turnover. Matching models therefore rely on imperfect information and job mobility results as information is revealed about the current match or about possible alternative matches.

The quality of the match between a worker and a firm is determined when the worker enters the firm. If the match is an experience good, then its quality is revealed to both worker and firm over time as the match continues. Good (or productive) matches are likely to continue and to receive higher wage offers within the firm, while poor (or unproductive) matches are likely to receive lower wage offers and to result in separations from the firm. Because the quality of the match is likely to be identified early, poor matches are dissolved earlier than later, which generates the negative correlation between tenure and rates of separation (Johnson 1978; Jovanovic 1979a;
Viscusi 1980). However it also generates the positive association between wages and tenure. This correlation appears not because there are payoffs to accumulated tenure but because, as time goes by, bad matches (with lower wages) are destroyed which raises the average wage within a cohort. This also generates, all else equal, lower wages for a worker who experiences many job separations because the worker rarely extracts the rents associated to the progressive identification of the quality of the match. If the match is a search good, and workers learn of possible alternative matches by searching on the job, then turnover is a process by which worker-firm matches are improved (Burdett 1978; Mortensen 1978; Jovanovic 1979b). In this case the inverse relationship between job tenure and probability of separation results from the growth in match-specific human capital.

Matching models also provide a framework for within-firm career progress as they allow employers to establish individual contracts with their workers. Examples of individual contracting include a system of promotions or delayed pay increases based on the quality of the match. Promotions can also arise from a system of counter-offers, and are a means of eliciting cooperative behaviour between worker and firm when alternative matches are found (Jovanovic 1979a,b; Mortensen 1978).

The empirical implications of matching theory are that estimates of the determinants of wages, job separations or promotions will be biased due to the presence of unobserved match heterogeneity. Previous research has focused on testing the empirical importance of match quality in determining wages and job mobility. Early studies attempted to do this using cross-sectional data, while the availability of panel data sets of individuals allowed worker unobserved heterogeneity to be explicitly modelled (Topel 1986, 1991; Flinn 1986; Altonji and Shakotko 1987; Abraham and Farber 1987; Mortensen 1988). Using a structural model of wage dynamics, Flinn (1986) reports that 38% of the variance in wages of new labour market entrants is attributable to worker-firm-specific heterogeneity. Altonji and Shakotko (1987), using an IV approach, find that match quality has a large negative effect on the quit and separation probability (and thus positively correlated with job tenure) and that the cross-sectional relationship between tenure and wages is a consequence of heterogeneity biases.
More recently the availability of longitudinal linked employer-employee data (LEED) allows for more direct estimation of match effects, together with unobserved worker and firm heterogeneity (Abowd et al 1999; Ferreira 2009; Woodcock 2008). Using LEED, some authors attempt to estimate the effect of matching and its relative importance in explaining wages of workers. Ferreira (2009) finds that unobserved match effects account for 6% of wage dispersion of workers that have entered a firm, while Woodcock (2008) estimates that they explain about 16% of observed variation in wages.

Some authors are attempting new approaches regarding the measurement of job-match quality. These have focussed on the use of subjective job-satisfaction measures collected at the individual level in longitudinal household surveys (Gielen, 2008; Gesthuizen 2008). However, in these papers match quality is being inferred from separate job satisfaction variables which are measured on a Likert scale. We find the partitioned and discrete nature of such measures insufficient and adopt a different approach to measuring match quality. We find the partitioned and discrete nature of such variables insufficient and unsatisfactory for identifying and quantifying the impact of match quality. Also using integer scores of subjective measures may be inappropriate as they assume even spacing between scale points, and the distribution of scores may be skewed (Bradley et al 1962; Snell 1964). Instead we use factor analysis to derive a single measure to identify a worker’s (unobserved) match quality. Our model specification follows closely that of Flinn (1986) who uses longitudinal data on individuals to identify match quality. However, instead of using structural econometric models, in our approach we use factor analysis to create an index of match quality using subjective job satisfaction data. We then test this index against some predictions of the matching model, relating to the evolution of match quality with seniority, individual contracting and separation rates. Our measure of match quality behaves as predicted by matching models, as does its effect on compensation and job mobility.

2. Data
We use nine years of the British Household Panel Survey (BHPS), covering 1998–2006. This is a nationally representative sample of 5,500 private households originally recruited in 1991 containing approximately 10,000 adults interviewed each successive
year. If anyone splits from their original household to form a new household, then all adult members of the new household are also interviewed. Children in original households are interviewed when they reach the age of 16. The core questionnaire elicits information on income and earnings, labour market status, housing tenure and conditions, household composition, education and health at each annual interview. For the purpose of our analysis we will only consider information on men that are in full-time employment. We do not consider women in our analysis because of issues relating to sample selection, non-random participation in the labour market, and secondary earners. We focus on working age men (21–59) in full-time employment who do not hold a second job or multiple jobs (to prevent possible contamination of responses across jobs).

There are several advantages to using the BHPS for these analyses. Firstly the data are longitudinal, which allow us to follow the same group of workers over time and identify job mobility. Secondly, the information collected is very rich, and include information on labour income, time with employer and a range of individual, household and job-related characteristics. Finally, respondents in employment are asked a range of questions relating to their subjective evaluation of the quality of the match with their current employer. These include questions about job satisfaction with various aspects of their current (main) job. In particular at each date of interview, employees are asked to rank on a scale of one (completely dissatisfied) to seven (completely satisfied) their current job satisfaction overall, their satisfaction with total pay, with job security, with the work itself and with the total hours worked in their current job. Furthermore, from 1998 onwards, people in employment are asked whether or not they would like a new job with a new employer. As we wish to include this variable in assessing match quality, we are restricted to using BHPS data from 1998 onwards. This yields an unbalanced sample which, when excluding cases with missing information on any variables of interest, has a sample size 5,192 men contributing 21,867 person-year observations.

3. Measuring match quality
We use subjective data from employees both on reported levels of job satisfaction in the current job and on whether they would like a new job with a new employer to construct a measure of the underlying match quality. We expect workers who match
better with their current employer will report higher levels of satisfaction with the various aspects of their job and be less likely to report wanting a new job with a new employer. The main assumption in our approach is that the underlying process generating the answers to the questions related to job satisfaction is the latent quality of the match between the worker and the firm. Therefore, our objective is to derive from a set of separate ordinal responses to a series of questions, one single continuous measure that reflects the variability observed in these responses. To do so we use two strategies: (i) Cronbach’s alpha and (ii) factor analysis.

Cronbach’s alpha is a coefficient of consistency and measures how well a set of variables or items measures a single, unidimensional latent construct, in our case the quality of the match. It creates a summative rating scale composed of a set of variables – the scale is the sum of the individual item scores (reversing the scoring for variables that have negative correlations with the factor being measured). It takes a value between 0 and 1, with one indicating perfect internal consistency. The literature suggests that a good summary indicator should have a value of alpha of at least 0.7 (Nunnally and Bernstein 1994). The internal consistency of the principle underlying factor reflected in responses to variables capturing job satisfaction and the desire for a new job with a new employer yields a Cronbach’s alpha of 0.79 and an average inter-item correlation of 0.38. This suggests that the individual variables all contribute to the underlying match quality component in the same way and empirically supports combining them into a single indicator.

We derive a single continuous measure of match quality that reflects the variability observed in responses to these questions using factor analysis, and regression scoring in particular. Factor analysis is a data reduction method used to describe variability among observed variables in terms of fewer unobserved (latent) variables called factors. Therefore, it is used to uncover the latent structure (dimensions) of a set of variables. It reduces attribute space from a larger number of variables to a smaller number of factors and as such is a “non-dependent” procedure. Following Zinbarg et al. (2005) and Zinbarg et al. (2006) we use higher order exploratory factor analysis.
using the principal factor method of extraction. The loadings on the highest order rotated factor are then used as the estimated match-quality loadings.\textsuperscript{1}

Table 1 summarises levels and changes in our match quality indicator between two consecutive BHPS dates of interview (\(t\) and \(t+1\)) by type of job mobility. This shows that on average match quality falls slightly from one year to the next while workers that remain with the same employer have higher than average match quality (0.02). Workers that subsequently separate from their employer have significantly lower match quality. Those that go on to quit a job have an average match quality of \(-0.30\) compared with \(-0.17\) for those that will be laid off. This is consistent with theory – separations occur where the quality of the worker-firm match is low. As predicted by matching theories, workers who separate and find new employment within the next year on average increase their match quality, which is largest (0.56) for those who quit.

These descriptive patterns indicate that the constructed measure of match quality behaves as predicted by economic theory. In Figures 1 and 2 we document other evidence that suggests the match quality variable is consistent with the predictions of matching models. In Figure 1 we plot average match quality by seniority, defined as years with the current employer. Matching models suggest that the wage-seniority profile arises because well-matched workers remain with the firm while poorly matched workers leave. If this is true, then the average match quality should be higher for workers with more seniority than for those with less seniority. Figure 1 indicates that this relationship emerges clearly in our data. Average match quality initially falls as workers learn about the qualities of the firm and their working environment. At this time workers in poor matches leave the firm, either voluntarily or involuntarily, and then average match quality improves as only well-matched workers remain with the firm. Figure 2 instead plots the average variance in match quality by seniority. If match quality is an experience good, then matching models suggest that the variance in match quality should fall with seniority, as the flow of new information to the worker on the quality of the match declines over time with the firm. Initially there will be a large flow of new information to the worker on the quality of the match, and so

\textsuperscript{1} Results presented are robust to constructing the continuous match quality indicator using a number of different methods such as principle component analysis and Cronbach’s alpha.
reported match quality will vary as the worker learns and absorbs this information and re-evaluates the match quality towards its ‘true’ value, resulting in a high match quality variance. Over time with the firm, this information flow slows as the worker becomes familiar with his working environment, and hence the true match quality is revealed. This pattern is exactly replicated in Figure 2, which shows that the variance in match quality falls with seniority.

Therefore descriptive statistics indicate that our match quality indicator behaves in ways that are consistent with the predictions of matching models. Workers in worse matches separate from their employer, while those in good matches are promoted. Similarly, workers who separate on average find employment in firms with whom they match better. In the remainder of the paper we test these descriptive findings more robustly using multivariate analysis.

4. Estimation

To test our measure of match quality against the predictions of matching theory, we estimate a series of three models. The first series of models examines the relationship between match quality and wages. If the quality of the match has a non-zero impact on wages, and if it is correlated with observable characteristics, then procedures that do not allow for match quality can result in biased estimates. In a matching model, the log of real wages of the $i$th worker at the $j$th firm at time $t$ ($y_{ijt}$) is given by:

$$y_{ijt} = a + x_{it} \beta + z_{jt} \sigma + \theta_i + \psi_j + \gamma_{ij} + \epsilon_{ijt}$$

[1]

where $a$ is the general mean, $x_{it}$ is a vector of worker characteristics, $z_{jt}$ is a vector of firm characteristics, $\theta_i$ captures worker-specific time-invariant unobserved effects (such as motivation, ability and ambition), and $\psi_j$ captures firm-specific time-invariant unobserved effects (such as firm-specific remuneration policies). $\gamma_{ij}$ is the unobserved match effect (which can be, for example, the worker-firm specific productivity, a production complementarities component or performance on the job) and measures the time-invariant heterogeneity associated to the match of a worker with a firm. $\epsilon_{ijt}$ is random error.
Empirical estimations of wage equations differ in their treatment of the (unobserved) worker ($\theta_i$), firm ($z_{jt}$) and match effects ($\gamma_j$). In particular, results based on cross-sectional individual level data and estimated using OLS assume that each is orthogonal to the covariates. However if the unobserved effects are correlated with covariates then OLS estimates are biased and inconsistent. The availability of individual-level panel data allows unobserved worker effects to be controlled for, and thus removing one source of potential bias. More recently, longitudinal linked employer-employee data (LEED) have become available in some countries which allow for more direct estimation of match effects, together with unobserved worker and firm heterogeneity (Abowd et al 1999; Ferreira 2009; Woodcock 2008). However, such data typically lack contextual information on workers and their background, potentially introducing omitted variable biases.

By attempting to measure match quality indirectly from data on workers, we fall somewhere in between. Our data are rich enough to include a large amount of contextual information on observed worker characteristics and the firms in which they are employed, as well as unobserved worker effects. In addition we include our measure of match quality. However as our sample is based on workers, we are unable to include firm-specific unobservables. Thus the model to be estimated becomes:

$$y_{ijt} = a + x_{it}\beta + z_{jt}\sigma + \theta_i + \gamma_j\eta + \epsilon_{ijt}$$  \[2\]

We estimate [2] using within-group fixed effects, which relies on the assumption of strict exogeneity but allow worker-specific unobserved effects to be arbitrarily correlated with the covariates. Matching models predict that $\eta$ is positive – workers in a better worker-firm match enjoy higher wages than those in a poor worker-firm match.

As well as receiving higher wages, matching models predict that well matched workers are unlikely to separate. We test our measure of match quality against this theoretical prediction by estimating a job separation model. Here the observed dependent variable, $s_{it}$, is binary, taking the value one if the worker separated from
his employer between \( t \) and \( t+1 \) and zero if he remained with the same employer. A separation is defined as a quit or a layoff, as matching models predict both are more likely when there is a poor worker-firm match. The model is specified as:

\[
\hat{s}_{it} = x_i \beta_s + \theta_i + \gamma_i \eta_s + \epsilon_{it}
\]  

[3]

where \( \hat{s}_{it} \) denotes the unobservable propensity for the worker to separate between \( t \) and \( t+1 \), \( x_i \) is a vector of observable characteristics that influence \( \hat{s}_{it} \), \( \theta_i \) denotes the individual-specific time-invariant unobservable effect and \( \epsilon_{it} \) is random error. By treating the \( \theta_i \) as random, this can be estimated using a random effects probit model under the common assumption that \( \epsilon_{it} \sim \text{IN}(0, \sigma^2) \) and are orthogonal to the other covariates.

This framework assumes that the time-invariant unobserved individual-specific effects are independent of the observable characteristics. This is quite realistic here as more able and highly motivated people are, for example, more likely to have higher education levels, be more aware of alternative employment opportunities, more likely to be promoted, and less likely to be laid off. In this case some of the estimated coefficients of interest (\( \beta \) and \( \eta \)) will pick up some of the effects of the unobservable \( \theta_i \). To avoid this problem we relax the assumption that \( \theta_i \) is independent of the observable time-varying covariates, following Chamberlain (1984) and Mundlak (1978). We model dependence between \( \theta_i \) and observables by assuming that the regression function of \( \theta_i \) is linear in the mean values of the time-varying covariates:

\[
\theta_i = a_i + \bar{x}_i b + \bar{\gamma}_i \eta + \mu_i
\]  

[4]

We assume that \( \mu \) are independent of \( x \) and \( \gamma \), \( \bar{x}_i \) refers to the vector of mean values of the time-varying covariates over time and \( \bar{\gamma}_i \) refers to the vector of mean values of match quality for individual \( i \) over time. The coefficients in \( b \) that correspond to the time-invariant variables are set equal to zero. Equation [3] therefore becomes:
This is equivalent to the random effects probit with additional regressors \( \xi \) and \( \gamma \).

Our models take into account a wide range of other individual and job-related characteristics that are likely to determine wages, separations and promotions. These include variables intended to capture labour market attachment, job search intensity, job offer arrival and job retention rates. All models include controls for age and its square, highest education qualification, marital status and number of children, and whether the worker experienced an employment interruption in the previous year. Job and employer characteristics include sector of employment, firm size, trade union coverage, pension scheme membership, place of work, seniority, occupation, whether receive increments or bonus payments, and whether there are opportunities for promotion in the current job. Region of residence and year dummies are also included.

5. Results
Table 2 presents estimated coefficients on the match quality variable from wage equations, where we normalise match quality in terms of standard deviations from the mean. Our estimates are consistent with matching models. In the OLS regression, having match quality of one standard deviation above the mean relates to earning 4.1% higher wages. However this is positively biased as more motivated and able workers are both more likely to earn higher wages and to search for suitable worker-firm matches. Within-group fixed effects estimates, which remove such bias, reveal a smaller coefficient that remains statistically significant – having match quality of one standard deviation above the sample mean relates to earning 3.2% higher wages. Therefore we find the measure of match quality is statistically significantly associated with wages, consistent with theory. Furthermore, robustness and specification tests reveal that the match quality measure yields a higher model $R^2$ than including each of the relevant component variables in separate models (results available from the authors on request).
Table 3 shows the importance of match quality relative to other observed characteristics in explaining wages. Row 1 presents the $R^2$ from the fully specified model and row 2 the error. Row 3 (5) presents the $R^2$ when just the covariates (match quality) are included, and row 4 (6) allocates the difference between the fully specified model to match quality (other covariates). According to OLS estimates match quality explains 1–2% of the total variance in wages, while other observed characteristics explain 45%. The fixed effects model attributes 1.1% of the total variance in wages to match quality. These estimates of the importance of match quality in explaining wages are in line with those from LEED data (Ferreira 2009).

Table 4 shows marginal effects from pooled and random effects probit models of job separation. A job separation is defined as either a quit for a better job, a layoff (redundancy) or a quit for other reasons (look after family or home, for example). It is intended to capture between-firm mobility. Consistent with matching models, the probability of separation is negatively related to match quality and the sizes of the effects are relatively large and statistically significant. In the pooled probit having match quality one standard deviation above the mean reduces the probability of separation by 3.2 percentage points. This is a large impact, given that the average separation probability was 0.2. In the random effects specification the size of the effect is larger, reducing the probability of separation by 4.4 percentage points. Therefore, consistent with matching models, we find that workers in a better worker-firm match have a lower probability of subsequent separation than workers in poorer matches. Again, specification tests reject models including each component variable separately in favour of the model including the match quality measure.

6. Conclusions
Our purpose is to show that combining several distinct indicators of employees’ subjective perceptions of match quality into a single composite measure yields a variable that behaves according to matching models, and which allows us to identify and quantify the impact of match quality on wages and turnover. We argue that using a single ordinal measure of job satisfaction as an indicator of match quality is too simplistic and unsatisfactory. Instead latent match quality is better captured using satisfaction across different dimensions of the job and other subjective information reported by the employee. Our approach of constructing a single ‘match quality’
measure from responses to a number of different questions related to job satisfaction and desire for a new job with a new employer allows us to identify the relationships between match quality and various outcomes in a simple and direct way. The estimated latent match-quality behaves as predicted by matching models, as does its effect on compensation and job mobility. Furthermore, specification and robustness tests indicate that this composite measure improves the fit of wage and job separation models relative to models in which each component variable is included separately. We therefore suggest its inclusion in analyses of these topics both in order to reduce omitted variable bias in the estimates obtained, and to directly identify the impact of match quality on outcomes of interest.
Table 1: Average match quality variable by job mobility

<table>
<thead>
<tr>
<th>Mobility between t and t+1</th>
<th>Mean</th>
<th>Change t to t+1</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>-0.004</td>
<td>-0.020</td>
<td>21867</td>
</tr>
<tr>
<td>None</td>
<td>0.022</td>
<td>-0.028</td>
<td>15307</td>
</tr>
<tr>
<td>Will quit</td>
<td>-0.301</td>
<td>0.558</td>
<td>1064</td>
</tr>
<tr>
<td>Will be laid off/dismissed</td>
<td>-0.174</td>
<td>0.101</td>
<td>221</td>
</tr>
<tr>
<td>Will separate for other reasons</td>
<td>-0.213</td>
<td>0.025</td>
<td>827</td>
</tr>
</tbody>
</table>


Table 2: Impact of match quality on wages

<table>
<thead>
<tr>
<th>Match quality</th>
<th>OLS</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.041</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>[11.18]</td>
<td>[14.58]</td>
</tr>
</tbody>
</table>

| R²            | 0.460| 0.163|
| N observations| 21867|
| N individuals | 5192 |

Notes: BHPS 1998–2006. Dependent variable is log(real hourly wage). See text for details of other control variables and for how match quality is defined.

Table 3: Contribution of match quality to total wage variation

<table>
<thead>
<tr>
<th>Source of wage variation</th>
<th>Share of TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
</tr>
<tr>
<td>1 Covariates &amp; match (A)</td>
<td>0.460</td>
</tr>
<tr>
<td>2 Error (1–A)</td>
<td>0.540</td>
</tr>
<tr>
<td>Covariates first</td>
<td></td>
</tr>
<tr>
<td>3 Covariates (B)</td>
<td>0.454</td>
</tr>
<tr>
<td>4 Match (A–B)</td>
<td>0.006</td>
</tr>
<tr>
<td>Match first</td>
<td></td>
</tr>
<tr>
<td>5 Match (C)</td>
<td>0.016</td>
</tr>
<tr>
<td>6 Covariates (A–C)</td>
<td>0.444</td>
</tr>
<tr>
<td>N observations</td>
<td>21867</td>
</tr>
<tr>
<td>N covariates</td>
<td>63</td>
</tr>
</tbody>
</table>

Table 4: Impact of match quality at t on probability of job separation t to t+1

<table>
<thead>
<tr>
<th>Match quality</th>
<th>OLS</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.032</td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td>[12.07]</td>
<td>[13.28]</td>
</tr>
</tbody>
</table>

| Rho           | 0.108|
| Log-likelihood| -9671.1| -9365.6|
| N observations| 21867|
| N individuals | 5192 |

Notes: BHPS 1998–2006. Marginal effects from probit models. Dependent variable is experiencing a job separation between t and t+1. See text for details of other control variables and for how match quality is defined.
Figure 1: Average match quality by seniority

Figure 2: Match quality variance by seniority
References

Ferreira, P. (2009), ‘Returns to job mobility: The role of observed and unobserved factors’, ISER WP No. 2009-12, Colchester.
Zinbarg, R., Revelle, W., Yovel, I., Li, W. (2005), “Cronbach’s alpha, Revelle’s beta, and Macdonald’s omega(h): their relations with each other and two alternative conceptualizations of reliability” *Psychometrika*, 70(1), 123-133