School Characteristics and Teacher Turnover: Assessing the Role of Preferences and Opportunities*†

Stéphane Bonhomme Grégory Jolivet Edwin Leuven

Abstract

Job characteristics can affect worker turnover through their effect on utility and through their effect on outside job opportunities. The aim of this paper is to separately identify and estimate the roles of these two channels. Our method exploits information on job changes, and relies on an augmented sample selection correction. To illustrate our approach, we use an exhaustive register of Dutch primary school teachers and show a detailed picture of preferences for school characteristics. We also find that the dependence between current and outside job attributes can affect turnover and thus the allocation of teachers across schools.

JEL codes: C34, C36, J32, J40, J62, J63
Keywords: labour turnover, compensating differentials, teacher labour markets, sample selection.

The study of labour turnover plays a central role in labour market analysis.[1] A large literature has studied how the determinants of job quit decisions relate to wages and wage dynamics (Topel and Ward, 1992), or more generally to job satisfaction (Freeman, 1978; Akerlof *et al.*, 1988). A standard approach, grounded in job search theory, consists in modelling the job quit probability as a function of the characteristics of the current job. These characteristics can affect a worker's decision to leave her job through two channels: preferences (their effect on the worker's utility) and job offers (their effects on the worker's outside job opportunities). In this paper, we aim to separately identify and estimate the

*Corresponding author: Edwin Leuven. University of Oslo, Department of Economics, Postboks 1095 Blindern, 0317 Oslo, Norway. E-mail: edwin.leuven@econ.uio.no

†We thank Holger Sieg, Hélène Turon, Marty West, Frank Windmeijer, and numerous seminar audiences for excellent feedback and discussion. Stéphane Bonhomme gratefully acknowledges support from the European Research Council/ ERC grant agreement n^{0}263107.

[1]This topic has also received considerable attention in microeconomic theory, see Burdett (1978) or Jovanovic (1979), and in macroeconomics, see Hall (1972), to cite only a few early references. In this paper we focus on the determinants of labour turnover.

role of these two channels, and show the relevance of this decomposition for the analysis of preferences and turnover on the teacher labour market.

Individual preferences, the first of the two channels we consider, have been studied extensively in the literature, in particular to recover workers' Marginal Willingness to Pay (MWP hereafter) for amenities. If job characteristics only affect workers' utility, Gronberg and Reed (1994) show that the job hazard rate reveals how workers value different job amenities. The intuition of this approach is that if workers are more likely to stay in jobs with certain characteristics, then this reveals their preferences for these amenities. Unlike hedonic wage regressions, this approach is robust to the presence of search frictions on the labour market (Hwang *et al.*, 1998). Several studies have used job quit probabilities to estimate teachers' MWP for school characteristics and make policy recommendations on the level of compensation needed to have teachers stay in specific schools (see e.g. Hanushek *et al.*, 2004).

We argue that job characteristics may affect turnover not only through workers' preferences, but also through a second channel driven by outside job opportunities. If this is the case, the job hazard rate is no longer guaranteed to reveal workers' MWP. The intuitive argument is that a worker may be staying in her job not only because she likes it, but also because this job reduces her access to attractive job opportunities. In the context of a job search model this happens if, unlike in Gronberg and Reed (1994), the arrival rate or the distribution of job offers depend on the characteristics of the current job.[2]

In the context of the teacher labour market we study in this paper, there are reasons to expect that current job characteristics affect school turnover through job opportunities. Some school characteristics such as average student performance may directly affect student intakes, and thus schools' demand for teachers. Teachers' current job attributes may also constrain outside opportunities: teachers working in schools in poorer neighbourhoods may find it more difficult to get a job offer from a school in a more affluent neighbourhood, or private schools may prefer to hire teachers who already worked in

[2]Several recent contributions to the job search literature allow for the search environment to vary across jobs. See, for example, Meghir *et al.* (2012), and Bradley *et al.* (2013).

private schools before. This has potential implications for policy. While policy makers cannot directly affect preferences (embedded in workers' utility function), they may set mechanisms that facilitate or hinder teachers' access to specific schools. There are thus two policy instruments that can be used to improve teachers' allocation across schools: compensation policies (based on preferences and thus on the first channel), and "mobility" policies (based on the second channel).

We propose a method that separately identifies the roles of preferences and job opportunities in job turnover, and recovers workers' preferences as well as the correlations between current and outside job characteristics. Our estimation approach exploits information on job-to-job transitions. In the simple theoretical framework that we use as a motivation, we let the worker's decision to move from one job to another depend on current and outside job characteristics. However, unlike Gronberg and Reed (1994), we allow outside job offers and current job characteristics to be correlated. Recovering the model's parameters is then formally equivalent to the standard sample selection problem in econometrics (Heckman, 1976). This is because job characteristics posterior to a job change are selected within the set of available job opportunities.

Identification of this type of models can be achieved if one or several determinants of the mobility decision (cost shifters) can be excluded from the job offer equations. We show that one such exclusion restriction is sufficient to identify the distribution of outside jobs' characteristics, and that workers' preferences may then be recovered in a second step by "differencing out" the effect of job characteristics on job opportunities. Since we allow for a large set of job attributes (ten, in our application), our benchmark results rely on linear index structures for workers' utility and job characteristics equations, as well as on parametric (normal) assumptions for the shocks. We show however that our identification approach can be extended to a non-parametric setting and we conduct robustness checks to allow for non-normal shocks or for unobserved individual heterogeneity.

We apply our approach to an exhaustive administrative data set of primary school teachers in the Netherlands, where wages are rigid and other characteristics are therefore likely to influence teacher mobility. We estimate teachers' preferences for a large number of attributes, including the percentage of disadvantaged students and average school performance in a national exam.

The validity of our approach relies on the presence of convincing exclusion restrictions. We use two excluded covariates in our empirical work. The first one is based on an interaction between demographic shocks and funding rules which lead to shocks to schools' budget. The second excluded covariate is based on the fertility of teachers' colleagues, which affects the school's demand for teachers. In both cases we argue that teachers' outside job opportunities are unlikely to be influenced by these variables, conditional on a set of controls. Importantly, having two excluded covariates yields over-identification conditions that we use to provide evidence on the joint validity of our exclusion restrictions.

Our estimates of teachers' preferences show that the main characteristics driving teacher mobility between schools are the proportion of disadvantaged pupils, the pupil-teacher ratio, the support-teaching staff ratio, and teaching hours. According to our estimates, Dutch teachers also value the average student performance, based on centrally set and graded exit tests. In terms of sign, our estimates yields similar conclusions on individual preferences to those produced by the standard approach that ignores correlations between current and outside jobs. However, in terms of size the estimates differ. These differences are driven by significant correlations between the characteristics of current and outside jobs.

Since our approach delivers estimates of the correlations between current and outside job characteristics, it also provides a new set of results relevant for the analysis of worker turnover which, as far as we know, has not yet been reported. For example, we find that teachers working in a school with a larger proportion of students with low-educated parents have fewer opportunities to move to a school where this proportion is small.

To illustrate the benefits of identifying the effects of preferences and job opportunities, we conduct a counterfactual analysis of teacher turnover in a market where job offers no longer depend on current job attributes. This exercise allows us to assess the effect of a policy that aims at improving the access of teachers to a different set of schools. The results show that, if turnover was only driven by teachers' preferences, the relationship between job turnover and average student performance or the proportion of disadvantaged minority students in the school would be substantially stronger. We would also observe more mobility across the distribution of school characteristics. In the case of disadvantaged minority pupils, most of the increase is driven by downward mobility, as, under the counterfactual scenario, teachers at the top of the distribution (that is, in schools with a larger share of disadvantaged pupils) would have a better access to schools at the bottom. In contrast, for other job characteristics the increase in mobility is more evenly spread between upward and downward mobility.

We are not the first to argue that job turnover-based methods may provide biased estimates of workers' preferences. For example, Boyd *et al.* (2005) note that job transition probabilities reflect not only a teacher's choice to transfer, but also her opportunities to do so. Boyd *et al.* (2011) analyse teacher and school preferences separately thanks to a rich data set on the centralised transfer request system in New York City. An important advantage of the approach we propose in this paper is that it is widely applicable and can be used to analyse labour markets where there is no centralised application system, such as teacher labour markets in many European countries or – more generally – non-teacher labour markets. Our approach can be implemented on a standard labour force survey with information on amenities and a reliable exclusion restriction.[3]

The paper is organised as follows. In Section 1 we present the model and describe our identification and estimation strategies. Section 2 describes the Dutch teacher market and our data. In Section 3 we present estimates of teachers' preferences based on our benchmark specification. We present the results of alternative specifications in Section 4, conduct a counterfactual exercise in Section 5, and conclude in Section 6. Additional results may be found in an online appendix (attached at the end of this manuscript).

[3]Plausible exclusion restrictions may also be available in other labour markets. For example, Gibbons and Katz (1992) and Dustmann and Meghir (2005) use plant closure as an exogenous shock on workers' mobility.

1 The Framework

This section starts with a general description of the problem of interest. We then present the selection model and describe our identification and estimation strategies.

1.1 Statement of the Problem

Consider an economy where jobs are described by a vector of J attributes, denoted as $A = (a_1, ..., a_J)$. The value that a worker with individual characteristics X attaches to a job A is given by the value function $V(A, X)$. We are interested in the marginal rate of substitution between amenity a_j and amenity a_k, defined as:[4]

$$\text{MWP}_{jk}(A, X) = \frac{\partial V(A, X)}{\partial a_j} \bigg/ \frac{\partial V(A, X)}{\partial a_k}, \quad \text{for } j \neq k. \tag{1}$$

MWP_{jk} is the change in a_k needed to keep the value of the job constant when a_j increases marginally. It thus measures the worker's relative preferences for two job characteristics. When a_k is the wage, MWP_{jk} is the marginal willingness to pay for a_j. We use the notation MWP throughout the paper, although in the empirical analysis the "numeraire" a_k is not the wage.

Note that we define MWP_{jk} as the ratio of marginal derivatives of the value function $V(A, X)$, not of the instantaneous utility function $u(A, X)$. This distinction matters in our context. The objects we are interested in – the MWP's derived from V – reflect workers' preferences given the distribution of jobs in the economy.

Our approach relies on job change decisions as a source of identification for individual preferences. Suppose that, at a given point in time, an alternative job with characteristics A^*, and value $V(A^*, X)$, is available to the worker. In the following we refer to alternative jobs as "outside jobs" or "job offers", indistinctly. Suppose also that the worker decides to move if:

$$V(A^*, X) > V(A, X) + C, \tag{2}$$

[4]Job attributes are assumed continuous in this discussion. In the empirical analysis, only one out of the ten job characteristics is discretely distributed (the public school dummy).

where C is a stochastic mobility cost, which could for example reflect the current school's demand for teachers, or monetary/psychic costs associated with changing job.[5] Note that this representation is quite general. If workers receive multiple job offers in a period, $V(A^*, X)$ may be interpreted as the value of the best alternative.

Workers weigh the various attributes of their job in proportion to their preferences when deciding whether to change job or to stay in their current job. However, equation (2) cannot directly be exploited for identifying preferences, as the characteristics A^* of an alternative job are not observed in the data if the worker chooses to remain in her job. The literature on the estimation of worker preferences from labour turnover is based on the probability to change job conditional on current job attributes and individual characteristics (Gronberg and Reed, 1994). In our notation, this standard approach considers a job quit decision where A^* is integrated out in equation (2), so that variation in A can be used to identify worker preferences. Let Q denote the indicator that an individual decides to change job. Formally, we can compute the probability of changing job, conditionally on A and X, as follows:

$$\Pr(Q = 1|A, X) = \mathbb{E}_{A^*|A,X}\Big\{ \Pr\left[C < V(A^*, X) - V(A, X)|A^*, A, X\right] \Big\}.$$

We can then see that, in general, A affects the job change probability through three channels: preferences (the value function V), the distribution of job opportunities (A^*), and the distribution of mobility costs (C).

In this paper we propose a general approach to estimate workers' preferences when current and outside job characteristics are not independent. Our approach has two main features. First, we use data on job-to-job transitions, as opposed to data on job turnover only as in the standard approach. The availability of job characteristics posterior to job change provides relevant, though indirect, information on job opportunities. Second, our approach relies on the availability of "cost shifters" Z, i.e. determinants of mobility costs C that are unrelated to the attributes of potentially available job offers. This second

[5] Also, drawing a very large positive mobility cost C can be interpreted as not receiving an outside offer. Similarly, drawing a large and negative cost C can be seen as receiving an adverse shock which may lead the worker to lose her current job.

feature allows us to separate the effect of preferences from that of job opportunities in the job change decision.

1.2 The Sample Selection Model

In order to take the model to the data, we assume that value functions, mobility costs, and amenity offers are linear in their determinants. We specify the value function and the mobility cost as follows:

$$V(A, X) = \theta A + \xi_X X, \quad \text{and} \quad C(X, Z) = -\left(\theta_X X + \theta_Z Z + \nu\right), \tag{3}$$

where θ, ξ_X, θ_X and θ_Z are parameter vectors, and where ν is independent of X, Z, and A. The assumption that the unobserved mobility shock ν is uncorrelated with current job attributes is one of the two main requirements of our approach. Note that, if this assumption failed to hold then the standard approach based on job turnover would yield inconsistent estimates, even if outside and current job characteristics were independent. To strengthen the plausibility of this assumption, we will control for a number of time-varying covariates. In addition we also control for worker-specific unobserved heterogeneity, using a simple extension of the basic approach that we outline in the next section.

Using (2) and (3), we have:

$$Q = \mathbf{1}\left\{\theta(A^* - A) + \theta_X X + \theta_Z Z + \nu > 0\right\}. \tag{4}$$

The marginal willingness to trade for the various job attributes can directly be recovered from the vector $\theta = (\theta_1, ..., \theta_J)$, as $\text{MWP}_{jk} = \theta_j/\theta_k$. Similarly, we also impose a linear index structure on the distribution of amenity offers:

$$A^* = \alpha A + \alpha_X X + \varepsilon, \tag{5}$$

where ε can be correlated with ν in (4). Note that (5) is a system of J equations, where

J is the number of job attributes. In particular, α is a $J \times J$ matrix of coefficients which plays an important role here, as it measures to which extent amenity offers depend on current amenities. With some abuse of terminology we refer to α as a matrix of correlation coefficients.[6]

We assume that (ε, ν) is jointly independent of A, X, and Z. Independence between the unobserved determinants of amenity offers ε and cost shifters Z is the second main requirement of our approach. Under independence, (4) and (5) satisfy an exclusion restriction whereby an exogenous cost shifter, Z, affects mobility decisions but is not related to outside job opportunities. We shall provide an extensive discussion of our choice of excluded covariates in the empirical section. Moreover, we will use two excluded regressors, thus obtaining over-identifying restrictions implied by the exclusion.

The linear index restrictions in (4) and (5) are not necessary for identification. In Appendix A, we provide a nonparametric identification result that only relies on conditional independence assumptions. Nevertheless, index specifications are useful to deal with a relatively large number of job attributes– ten, in our application– while a fully nonparametric approach would face a severe curse of dimensionality in this case. Moreover, under index restrictions the model takes the form of a standard sample selection model, making identification and estimation simple and transparent.

Combining (4) and (5) we obtain the following reduced-form equation:

$$Q = \mathbf{1}\left\{\psi A + \psi_X X + \theta_Z Z + \eta > 0\right\}, \tag{6}$$

where $\psi_X = \theta_X + \theta\alpha_X$, and where $\eta = \nu + \theta\varepsilon$ is independent of A, X, and Z. The reduced-form parameter ψ is then linked to the preference parameter θ by the mapping:

$$\psi = \theta\left(\alpha - I_J\right), \tag{7}$$

where I_J is the $J \times J$ identity matrix. This mapping comes from combining (4) and (5) into (6). Equation (7) shows that ψ is a composite of workers' preferences (θ) and

[6]Though convenient for implementation, specification (5) is not directly motivated by an economic model. This specific linear form is not needed for identification of teachers' preferences, as we show below.

characteristics of the job offer process (α). This provides a clear separation of the effect of job characteristics on job turnover into a preference effect and a job opportunities effect.

Taking stock, we have a sample selection model where the selection equation is a reduced-form mobility decision, (6), and the outcome equation is given by (5). The parameters of this model are linked to the preference parameters by the mapping (7).

Combining the two equations of our selection model, (5)-(6), we can see that job-to-job transitions provide information on the mean amenity values among job changers:

$$\mathbb{E}\left(A^{*}|A,\, X,\, Z,\, Q=1\right) = \alpha A + \alpha_X X + \mathbb{E}\left(\varepsilon|\psi A + \psi_X X + \theta_Z Z + \eta > 0\right). \tag{8}$$

In general, ε and $\eta = \nu + \theta\varepsilon$ are correlated. As a result, an ordinary regression of job amenities for job changers on the attributes of their previous job does not provide a consistent estimate of the correlation coefficients α. However, the availability of one continuously distributed cost shifter Z is sufficient for both the correlation coefficients and the MWP for job amenities to be semi-parametrically identified. Formally, we have the following result.

PROPOSITION 1. *Let (6)-(7)-(8) hold. Suppose that* (ε, η) *is independent of* A, X, *and* Z. *Suppose in addition that* A, Z *and* (ε, η) *admit absolutely continuous densities, and that* $\theta_Z \neq 0$ *and* $\alpha \neq I_J$. *Then* α *is identified, and* ψ *and* θ *are identified up to scale.*

Proposition 1 follows directly from existing semi-parametric identification results for sample selection models (e.g., Das *et al.*, 2003). Its proof is given in the online appendix.

1.3 A Three-Step Estimation Method

Suppose that we have panel data on job attributes A_{it}, individual characteristics X_{it}, cost shifters Z_{it}, as well as data on job change decisions $Q_{it} \in \{0, 1\}$, where i and t denote individuals and time periods, respectively. Observations are assumed i.i.d. across individuals. Following the discussion in the previous subsection, the estimation procedure consists of three simple steps. Here we present the method assuming that unobservables are normally distributed. In Section 4 we will report the results of a non-normal specification.

Step 1. We estimate the reduced-form parameters ψ in (6) by Probit, assuming that η_{it} is normally distributed with variance equal to one. This means that we recover the vector ψ up to scale. The output of the first step consists of the parameter estimates $\widehat{\psi}$, and of the predicted probabilities $\widehat{\Pr}\left(Q_{it}=1|A_{it},\, X_{it},\, Z_{it}\right)$.

Step 2. To estimate α we start by noting that, by (8), α can be consistently estimated by regressing the job attributes A^*_{it} of teachers who have just moved (that is, for $Q_{it}=1$) on A_{it}, X_{it}, and a flexible function of the estimated job change probability $\widehat{\Pr}\left(Q_{it}=1|A_{it},\, X_{it},\, Z_{it}\right)$.

Under normality (8) becomes, for $j=1,...,J$:

$$\mathbb{E}\left(A^*_{ji,t}|A_{it},\, X_{it},\, Z_{it},\, Q_{it}=1\right)=\alpha_j A_{it}+\alpha_{Xj}X_{it}+\rho_j\sigma_j\lambda\left(\psi A_{it}+\psi_X X_{it}+\theta_Z Z_{it}\right),$$

where α_j and α_{Xj} are the jth rows of matrices α and α_X, respectively, where σ_j is the standard deviation of the jth element of ε_{it}, and where ρ_j is the correlation between the jth element of ε_{it} and η_{it}. The function $\lambda(\cdot)$ is the inverse Mills' ratio.[7]

Step 3. Finally, given $\widehat{\psi}$ and $\widehat{\alpha}$ we estimate the $1\times J$ vector θ as:

$$\widehat{\theta}=\widehat{\psi}\left(\widehat{\alpha}-I_J\right)^{-1}. \tag{9}$$

Note that $\widehat{\theta}$ depends on a scale normalisation which does not affect the MWP estimates because $\widehat{\mathrm{MWP}}_{jk}=\widehat{\theta}_j\big/\widehat{\theta}_k$.

Our three-step estimation method thus consists of a simple selection correction estimator, augmented with a final step where teachers' preferences are recovered. Under normality, the first two estimation steps follow the standard Heckman (1979) procedure except that we have a multidimensional outcome. For inference, we use the nonparametric bootstrap, since this conveniently takes into account the multi-step nature of the estimation algorithm and the clustering of the standard errors at the school level.

[7] That is, $\lambda(\cdot)=\frac{\phi(\cdot)}{\Phi(\cdot)}$, where $\phi(\cdot)$ (respectively. $\Phi(\cdot)$) denotes the probability (resp. cumulative) distribution function of the standard normal.

1.4 *Discussion*

1.4.1 *Exit from the labour market.*

A common drawback of using administrative labour data is that individuals may disappear from the sample when they leave the state of interest. For example, matched employer-employee data sets may lose track of individuals who become unemployed, go to work for the public sector or become self-employed. In our case, whilst we have a rich and exhaustive data set on teachers, we do not observe labour market outcomes for individuals who leave the teacher labour market. This empirical issue is pervasive in the literature on teacher turnover.[8]

Our framework is robust to exits from the labour market of interest, under certain assumptions. Specifically, for the identification result of Proposition 1 to remain valid, the assumed independence between (ε, η) and (A, X, Z) needs to hold conditionally on the individual not exiting the labour market of interest. In our case, this means that we assume away possible dependence between teaching and non-teaching job opportunities, conditionally on the current job's characteristics, teacher characteristics, and cost shifters. This assumption may not be too strong once the conditioning is taken into account. Indeed, we allow teaching and non-teaching job offers to be correlated through the presence of teacher characteristics, attributes of the current school (which may act as a signal), or demand shocks at the school or local labour market level. All these features are accounted for in the empirical analysis. In particular, the local labour demand shocks will be captured by a vector of local labour market conditions. Importantly, differences in unobserved teacher characteristics, for example teacher quality, will also be accounted for in the specifications where we control for an unobserved teacher fixed effect.

1.4.2 *Unobserved heterogeneity.*

In the benchmark specification we have assumed that the unobserved determinants in the job mobility equation – that is, ν_{it} in equation (4) – are independent of current job characteristics. There may however be unobserved factors that affect job mobility

[8] See Dolton and Van Der Klaauw (1995, 1999) for an analysis of teacher turnover in the UK.

decisions and are correlated with job attributes.[9] We use panel data techniques to allow for unobserved teacher-specific effects in our estimation approach as follows.

We introduce an heterogeneous intercept, denoted as β_i (J-vector), in the offer equation (5) and another one, denoted as μ_i (scalar), in the mobility equation (6). Note that the latter allows value functions and mobility costs in (3) to depend on unobserved individual-specific determinants. We assume that these teacher-specific effects remain constant during the period of observation (3 years in our data set).

We build on the approach suggested by Wooldridge (1995) and treat the unobserved intercepts as correlated random effects.[10] Specifically, we model each individual effect as a linear function of the first observed individual and job characteristics (X_{i1}, A_{i1}), plus a residual.[11] The individual effects are thus modelled as:

$$\beta_i = \beta_A A_{i1} + \beta_X X_{i1} + \widetilde{\beta}_i, \quad \text{and} \quad \mu_i = \mu_A A_{i1} + \mu_X X_{i1} + \widetilde{\mu}_i, \tag{10}$$

where $\widetilde{\beta}_i$ and $\widetilde{\mu}_i$ are independent of A_{it}, X_{it}, Z_{it} for $t \geq 1$. In addition we assume that ε_{it} and η_{it} are independent of (X_{i1}, A_{i1}).

We thus obtain a selection model with unobserved individual heterogeneity which is an extension of the benchmark model (5)-(6). Under normality, we can use the three-step estimation technique from subsection 1.3, period by period, to recover the preference parameters θ (up to scale), as well as the correlation coefficients α. The computational simplicity of our estimation approach is then preserved when allowing for unobserved heterogeneity.

Controlling for unobserved individual effects allows to account for teacher-specific sources of endogenous selection into jobs. The analysis may still be affected by the presence of time-varying unobserved confounders such as unobserved job attributes, correlated

[9] In particular, a drawback of our administrative data is that one has little access to information on a worker's family.

[10] See also Semykina and Wooldridge (2007). Another approach, suggested by Kyriazidou (1997), consists in treating the individual fixed effects as parameters. A comparison of these two methods is conducted in Dustmann and Rochina-Barrachina (2007).

[11] Wooldridge (1995) suggests conditioning individual effects on the whole sequence of regressors, X_{it}, A_{it} for all t. Because the A_{it}'s are not strictly exogenous in our case, we only condition individual effects on the initial values.

with the observed job characteristics. Because of this concern, we think it is important to allow for a large number of job attributes in the estimation. Compared to previous studies, we control for an unusually large number of job characteristics (ten different attributes). The ability of our approach to handle multivariate amenities – i.e., vectors of A and A^* – is thus essential in our view.[12]

1.4.3 Structural interpretation.

Our method recovers consistent estimates of the determinants of the value function V when job attributes affect outside opportunities. The MWP's that we identify, given by (1), reflect workers' preferences given the observed conditional distribution of job offers (that is, A^* given A). If the social planner were to change this distribution, the MWP's could change even though primitive preference parameters remain constant.

Formally, if $u(A, X)$ denotes the instantaneous utility function, the primitive preference parameters are given by the ratios $\frac{\partial u(A,X)}{\partial a_j} \Big/ \frac{\partial u(A,X)}{\partial a_k}$ and are not sensitive to changes in the offer distribution. These ratios are equal to our MWP's when the distribution of offers A^* and mobility costs C are independent of A. If offers A^* depend on A, however, then in general $\frac{\partial V(A,X)}{\partial a_j} \Big/ \frac{\partial V(A,X)}{\partial a_k}$ is different from $\frac{\partial u(A,X)}{\partial a_j} \Big/ \frac{\partial u(A,X)}{\partial a_k}$.[13] A fully structural approach to recover the primitive preference parameters $\frac{\partial u(A,X)}{\partial a_j} \Big/ \frac{\partial u(A,X)}{\partial a_k}$ would be to solve a challenging dynamic programming problem with multi-dimensional state variables and a search environment that depends on the current job. As far as we know, no study has yet tackled identification and estimation of worker preferences when jobs are characterised by a large number of attributes and the search environment varies across jobs.[14] Identifying and estimating determinants of the value function, as we do in this paper, can thus be

[12]One possible strategy to deal with the presence of unobserved job attributes would be to use lagged amenity values (e.g., characteristics of the first job) as instruments for current job characteristics in equation (6). This would require assuming a specific dynamic structure on the error terms. Given the short length of the panel we were not able to pursue this strategy in the empirical analysis, but we view this extension as an interesting avenue for future work.

[13]This follows from the Bellman equation (where δ is the discount rate):
$V(A, X) = u(A, X) + \delta \mathbb{E}_{A^*,C|A,X} \{\max [V(A, X), V(A^*, X) - C]\}$.

[14]Bonhomme and Jolivet (2009) estimate an on-the-job search model with five binary amenities and no dependence of job offers on the current job. Bradley *et al.* (2013) consider an on-the-job search model where the search environment depends on a binary amenity (private/public sector), but they only allow for that single amenity.

seen as a first step towards this goal.

2 Primary School Teachers in the Netherlands

2.1 The Dutch Market for Primary School Teachers

We use our approach to estimate the preferences of primary school teachers in the Netherlands. In this section, we present some features of the Dutch education system that are relevant for our analysis.[15]

First, there is financial and statutory equality between public and private schools. The latter, which are not governed by a public legal entity, are subject to private law, have discretion in their teaching content and practice (within rules and end goals set by the ministry of education) and can refuse admission to pupils. Otherwise, private schools do not differ from public schools. In particular, both types of schools are publicly funded and cannot charge student fees. Schools are governed by a school board which, for public schools, is the municipal authority. Some school boards administer more than one school.

Primary school teachers must have obtained a teaching certificate. They are qualified to teach all subjects with the exception of sports, arts, and foreign languages which are taught by special teachers. Teachers are employed by the school board which has full discretion in the management of its labour force (within rules set by the ministry of education). However, wage scales are set centrally by the government (in terms of full-time equivalents). Basically, teachers are on a wage ladder and move up one rung every year until they reach the top of the ladder and then move on to the next one (there are three wage scales overall). A teacher's wage is thus a deterministic function of her experience, rare exceptions being that some teachers skip the first rung when they move from one wage ladder to the next. There is no wage compensation for working in a given type of school. This is an important feature as teacher selection between schools is therefore only based on non-wage job characteristics.

The school year runs from August 1st to July 31st of the following calendar year. There

[15] For a detailed description of the Dutch education system, see Eurydice (2008).

is a 6-week holiday during the summer and other shorter holidays throughout the year. Primary schools receive government funding under three budget headings: running costs, accommodation and staff. The latter budget is a function of the number and types of pupils registered in the school on October 1st. Schools funding is driven by a compensatory policy aimed at giving more resources to schools with a larger number of disadvantaged pupils. The scheme is based on weights as follows: a weight of 1.9 is assigned to pupils with a non-Dutch cultural background and whose parents have a low level of education, 1.7 to pupils from traveler families, 1.4 to those living in a children's home or a foster family, 1.25 to children whose parents are Dutch and have a low level of education and 1 to everyone else.[16] Therefore, a school's demand for teachers depends both on the number and on the types of children who register. Below we use changes in this budget (which reflect changes in the pupil population) as an important source of variation in teachers' mobility.

Schools that have more disadvantaged students are allotted more funding for staff. However, they cannot offer a teacher a wage higher than what her experience grants her. Schools can thus spend this additional funding on support staff (increasing the support-teaching staff ratio), on teaching material (e.g. on computers), or on hiring more teachers. There is no class size rule in the Netherlands, so schools with large numbers of disadvantaged pupils can hire more teachers and make smaller classes. Schools can thus use their budget to compete for teachers on non-wage job attributes, which motivates our empirical analysis of teacher preferences for these characteristics.

2.2 Data

We use administrative data that contain every contract between a teacher and a primary school in the Netherlands. Merging this register with other data sets on schools, we construct a matched teacher-school panel with one observation per teacher i and year t.[17] We restrict our sample to female teachers since the overwhelming majority (over 80

[16] Our data span over the period 1999-2002. Since then, a new scheme has been introduced in August 2006.

[17] Since our data set covers the whole country, we do not have the attrition problem faced by studies based on state or district-level data (e.g. Hanushek *et al.*, 2004 or Boyd *et al.*, 2005).

percent) of primary school teachers in the Netherlands are women. We further consider only teachers whose age is between 20 and 60. There are essentially no teachers younger than 20, and to avoid potentially confounding effects of retirement we cut our sample at age 60.

We have access to data for three school years, from 1999 to 2002. For every teacher one observation per school year is kept, corresponding to her main employment. We assume that this is the observation for the school $s = s(i, t)$ where the teacher has the highest teaching load on December 31st of year t.[18] The choice of December 31st is motivated by an empirical regularity in teacher school-to-school transitions (the vast majority of school changes take place between July and November). The school change indicator Q_{it} equals 1 if $s(i, t) \neq s(i, t+1)$ and 0 otherwise. The total attrition rate is 14%, which includes retirements.

We have no information on teachers' outcomes once they stop working or take a non-teaching job. We thus abstract from individual decisions to leave the Dutch teacher labour market. In subsection 1.4, we discussed the assumption that allows us to conduct our analysis only for teachers who stay in the market. We assume that non-teaching job opportunities are not correlated with alternative teaching jobs conditionally on the current teaching job. To control for local labour market conditions and thus reinforce the validity of this assumption, we include four region dummies as well as the regional unemployment rate in levels and changes. We also control for the unemployment insurance rate and vacancy rate at the provincial level (12 provinces). In the online appendix we report several descriptive statistics on teacher exits.

Mobility rates are particularly high and nonlinear at the beginning of a teacher's career. For this reason, our controls X_{it} include age in a flexible manner with single year age dummies up to age 25, after which we have dummies for 26-30, 30-39, 40-49 and 50+. We also include the teacher's current wage as an individual characteristic. As we pointed out above, selection across jobs cannot operate through wages because these follow a rigid

[18] Most teachers work in one school but some arts, sports or foreign language teachers may be employed in several schools. We cannot identify these teachers but we expect them to have smaller teaching loads in each school they work at. Also, we drop observations posterior to an exit from and a re-entry in the teacher labour market.

Table 1: Descriptive statistics on teachers

Average age (years)	40.5
% <30 years old	19.3
% $30-39$ years old	22.6
% $40-49$ years old	37.8
% >50 years old	20.3
% Parental leave	3.4
% Movers	3.5
Number of observations	167,550
Number of individuals	70,159

Note: "Parental leave": at least one parental leave. "Movers": at least one school change.

scheme set by the government.[19]

In addition, our data allow us to compute a dummy that equals one if the teacher is on a parental leave during the second semester of year t (i.e. between July and December of year t), as we observe the starting and ending dates as well as the reason of all individual absence spells. We will document that the number of colleagues on parental leave affects a teacher's decision to change job on that year. Finally, we add relative seniority within the school and an extensive set of controls, which we discuss in the next subsection when motivating our exclusion restrictions. Table 1 shows a set of basic descriptive statistics for the teachers in our sample.

Most job transitions take place between school years. Hence in most cases, when teachers decide whether to leave a school, the information on this school's new student numbers and budget is known. It is also natural to assume that teachers care about the pupil population and school attributes of the school-year that is about to start and not of the school-year that just ended. We therefore assume that teachers base their job change decision on the upcoming school-year's attributes of their current school.[20] We also include the current teaching load (year t) in the vector of job attributes A_{it} since we do not observe i's counterfactual teaching load in her old school in case she moves.

Our data contain information on ten job attributes that may enter teachers' value

[19]The wage may thus be interpreted as an additional proxy for teaching experience.

[20]While it is rational for a teacher to care about the school's attributes of the year that is about to start, it is important to acknowledge that her information about some of these attributes might not be perfect. Allowing for uncertainty would be a significant extension of our framework.

Table 2: Descriptive statistics on schools in 2000

	Mean	Std.Dev.
Amenities		
Disadv. minority pupils (fraction)	0.163	0.250
Disadv. Dutch pupils (fraction)	0.138	0.116
Pupil-teacher ratio	20.1	3.8
Teacher hours (in full-time equivalents (FTE))	0.734	0.250
Population density - log(population/km^2)	6.7	1.2
Public school	0.322	0.467
Student achievement (percentile)	0.493	0.131
Age teachers (average)	41.4	3.8
Female teachers (fraction)	0.824	0.095
Support staff (in FTE as fraction of total staff)	0.098	0.113
Excluded covariates	Z^{bud}	Z^{pl}
Mean	-0.97	0.28
Quantile 25% / 50% / 75%	-11.6/-1.0/9.3	0/0/0.46
$\Pr(Z \leq 0)$	0.53	0.65
$\Pr(Z > 0)$	0.47	0.35
Number of schools	5,758	
Number of teachers per school (FTE)	10.4 (7.6)	
Number of pupils per school	223	

Note: Z^{bud} is the change in a school's budget. Z^{pl} is the total teaching loads of a teacher's colleagues who are on parental leave. FTE: Full-Time Equivalent.

function. These variables are presented in Table 2, where means and standard deviations are computed among the population of teachers. In the online appendix we report age-specific averages of job attributes.

We measure the socio-economic composition of the school through the proportions of disadvantaged children within the school. Disadvantaged minority pupils include all pupils with budget weights 1.9 or 1.7 (see subsection 2.1). Disadvantaged Dutch pupils are all children in categories with budget weights 1.25 or 1.4. Since there are very few children in categories with weights 1.7 or 1.4 we merge them with category 1.9 and 1.25 respectively. The proportion of children coming from a disadvantaged ethnic minority is around 16 percent on average. In comparison, the proportion of children coming from disadvantaged native Dutch families is 14 percent.

The pupil-teacher ratio – a proxy for class size – is 20 on average. Population density is defined as the logarithm of the number of inhabitants per square kilometre in the school's municipality. About one third of teachers work in public schools. Notice however that, as discussed above, private schools in the Netherlands are publicly financed. The teaching load is a variable taking values between zero and one, and giving the full-time equivalent number of teaching hours.

Student achievement is computed using a national exit test taken at the end of primary education (in February). We control for the average percentile score within the school. Some schools (14%) do not implement this exam. We drop these schools from our sample for our benchmark estimation results. We have run robustness checks in which we included these schools and dropped the test variable from the list of job attributes, and found qualitatively similar results.

Finally, since our data set contains the employment contracts of non-teaching staff, we compute a variable that gives the number of support staff per full-time teacher within the school. We also account for the average age and gender among teachers within the school.

2.3 Exclusion Restrictions

We rely on two covariates as determinants of job mobility that are excluded from the amenity offer equations. We now present these two variables in turn.

2.3.1 Shocks to the school's budget.

A school's staff budget B_t for a given year is computed as the weighted sum of the five groups of pupils registered at the school (the student numbers are taken on October 1st). We define our first excluded covariate as the change in the school budget, that is $Z_t^{bud} = B_{t+1} - B_t$. This variable exploits demographic shocks to the school's student body, both in terms of the number of pupils and of the distribution of types (such as the share of disadvantaged pupils).

The variable Z_t^{bud} captures how a school's demand for teachers changes from one year to the next. A set of descriptive statistics on schools, together with the distribution of

Z^{bud} across schools in 2000 is reported in Table 2. Ideally, we would like to know whether the school is closing (or opening) a class but this information is not available in our data. Schools probably smooth the impact of budget shocks to some extent. We expect however that a teacher is more likely to leave (resp. to stay in) a school if Z^{bud} is negative (resp. positive). This is confirmed by our estimation results which show that these shocks are a significant predictor of mobility. Indeed, the estimate of $\theta_{Z^{bud}}$ obtained from the first estimation step and shown in Appendix B, Table B1, is significantly negative.[21] To further illustrate the role of budget shocks on mobility, we have simulated the average quit probability whilst fixing Z^{bud} at different deciles of its distribution. The results are not fully reported here but are available upon request. The quit probability is above 4% at the first decile of Z^{bud} (large negative budget shocks) and falls to 3% at the 9$^{\text{th}}$ decile.

For our exclusion restriction to be valid, the shock Z^{bud} on individual mobility needs to be independent of the characteristics of alternative jobs available to a teacher. Alternatively, our assumption is that a school does not take the Z^{bud} of other schools into account in its hiring decisions. To strengthen the plausibility of this assumption, we control for a number of potential confounders, in addition to the controls presented above. A first potential concern is that if a given region is hit by an aggregate demographic shock, then a teacher who has to leave her school may have access to fewer outside jobs because the pupil population in other schools also decreases. We address this concern by controlling for two aggregate demand proxies: the sum of Z^{bud} among all the other schools that are in the same town as school s, and the sum of Z^{bud} among all the schools that are in the same district but not in the same town as school s.[22] A second concern is that, since there are no catchment areas in the Netherlands, a school's pupil population may decrease as a result of it being perceived as a "bad" school. In this case, other schools may be less inclined to hire its former teachers. We account for this possibility by controlling for the ranks of a given school in the distributions of school average test scores within the town, and within the district.

[21] Note that this type of exclusion restriction is not new in the education and labour economics literature. For example, Hoxby (2000) uses similar variation in student populations to study the effect of class size on test scores.

[22] Districts are administrative areas, larger than cities, defined by the Dutch ministry of education.

2.3.2 Fertility of colleagues.

We also construct a second variable based on fertility. At all dates we observe all the teaching loads in the school, and whether teachers are on parental leave.[23] For each teacher i we compute Z_i^{pl} as the sum of the teaching loads of her colleagues who are on a parental leave (between July and December). It is important to note that teacher i's own parental leave is not used to compute this variable.

For the exclusion restriction to be valid we need to assume that the parental leaves of a teacher's colleagues affect her probability to leave the school but not her outside job opportunities. This assumption seems likely to hold but one may think that colleagues' parental leave is too small a phenomenon to have an impact on teacher turnover. It turns out that Z^{pl} is positive for 35% of our observations (see Table 2). In 13% of our observations, Z^{pl} is greater than 1 which means that the cumulated teaching load of teachers on parental leave is larger than that of a full-time teacher. Moreover, our estimation results show that colleagues' fertility has a significant impact on a teacher's mobility decision. The estimate of $\theta_{Z^{pl}}$ shown in Appendix B, Table B1 is significantly negative. Similarly to what we did for budget shocks, we computed the average quit rate for different values of Z^{pl}. The quit probability is around 3.55% at the first decile of Z^{pl} and goes below 3.2% at the 9$^{\text{th}}$ decile.

Finally, while these arguments suggest that Z^{bud} and Z^{pl} may plausibly be excluded from amenity offer equations, the validity of the exclusion restrictions might be compromised if unobserved school factors, correlated with either of the two covariates, are taken into account by outside schools in their recruitment strategies. In this non-experimental setting, it is thus particularly useful to have two exclusion restrictions that rely on different sources of variation, and provide evidence on the joint validity of Z^{bud} and Z^{pl}.

[23] The parental leave policy is defined in the collective bargaining agreement that binds all primary schools. Women have 16 weeks of fully paid leave in connection to a birth. There is also a right to 13 weeks of unpaid parental leave, to be taken during a period of no longer than 6 months and for at most 50% of the contracted working time during a given week. In practice many women move from a full time to a part time contract after the birth of a child.

3 Main Estimation Results

3.1 Teacher Preferences

We start by reporting our estimates of the weights of each job characteristic in the value of a job. The first two columns in Table 3 present the estimates of the preference parameters θ. From now on, all reported standard errors (in parenthesis) are bootstrapped using 499 replications and take clustering at the school level into account. The last two columns in the table show the MWP parameter estimates together with their standard errors, using as reference characteristic the pupil-teacher ratio.

The sign and significance of the parameters θ convey information on teacher preferences. The results show the following general picture: teachers are less willing to work in schools with a high proportion of disadvantaged pupils, large classes or a large support-teaching staff ratio. They prefer to work in schools with higher average test scores, a more experienced staff (i.e. a higher average age) and a higher proportion of female teachers. They would also rather work more hours and in less densely populated areas.

The proportion of disadvantaged minority pupils is perceived as a disamenity by teachers as its θ coefficient is significantly negative. This is consistent with previous findings in the literature (e.g., Hanushek *et al.* 2004; Scafidi *et al.* 2007). Depending on the institutional context and/or data availability these previous studies typically use the proportion of minority pupils and of pupils eligible for subsidised lunch to control for students' socio-economic background. In our data, in contrast, we observe the proportions of pupils with low-educated parents from a Dutch or a non-Dutch background.

Not surprisingly, teachers prefer schools with a smaller pupil-teacher ratio. As we mentioned above, schools with a larger budget cannot post higher wages since wages are set at the national level and are tied to experience. However, schools can hire more teachers and reduce class size in order to attract teachers. Our results in the second column of Table 3 show the changes in pupil-teacher ratio required to compensate for a one unit change in each amenity (MWP). For example, to compensate for a 10 percentage point increase in the proportion of minority students one would need to reduce the pupil-

Table 3: Estimates of preference parameters (θ) and MWP (θ/θ_{PT})

	θ_j		θ_j/θ_{PT}	
Disadv. minority pupils	-0.410***	(0.079)	-43.2***	(15.6)
Disadv. Dutch pupils	-0.146	(0.092)	-15.4	(10.7)
Pupil-teacher ratio (PT)	-0.009**	(0.004)	ref.	
Teacher hours	0.254***	(0.070)	26.8**	(13.0)
Population density	-0.056***	(0.020)	-5.9*	(3.3)
Public school	-0.137**	(0.063)	-14.5	(8.9)
Student achievement	0.958**	(0.378)	101.0*	(55.8)
Age teachers	0.010***	(0.003)	1.0*	(0.6)
Female teachers	0.253**	(0.110)	26.7*	(16.2)
Support staff	-0.513***	(0.100)	-54.1**	(26.9)

Note: */**/*** statistically significant at the 10/5/1 percent level.

teacher ratio by more than 4 (0.1×43.2).

To put these results in perspective, remember that the weighting in the Dutch budget scheme is such that a school's budget almost doubles when the proportion of disadvantaged minority pupils goes from 0 to 100% (disadvantaged minority pupils have a weight of 1.9 in the funding scheme). In practice we observe that schools where this proportion is 0 have a pupil-teacher ratio of 23 on average whereas schools where this proportion is 100% have a pupil-teacher ratio of 12. It seems that these latter schools use most of their extra budget to reduce class size. This is consistent with our results in the sense that schools try to provide what teachers value. Yet this not enough to fully compensate teachers. A decrease of $23 - 12 = 11$ in the pupil-teacher ratio only compensates for a $100 * 11/43.2 \approx 25$ percentage point increase in the proportion of disadvantaged minority pupils. This simple calculation may explain why schools in disadvantaged areas can have problems retaining their teachers.

The average age of teachers within the school plays a positive and significant role in teachers' utility. This effect is almost equivalent to the effect of reducing the pupil-teacher ratio by one unit. It is difficult to interpret this effect without more detailed data. Since a teacher's age is a good indicator of her experience, one interpretation would be that teachers prefer more experienced colleagues. Another interpretation could be that teaching positions in schools with a more experienced staff are more secure than in other schools. We will present preference estimates for different age groups that shed more light on this. Also, teachers– who in our sample are all women– seem to prefer working in schools with larger proportions of female teachers.

Teachers prefer schools with a lower support-to-teaching staff ratio. Support staff can be seen as one of the many indicators of working conditions and we may expect teachers to prefer schools where the support staff is large. Indeed, the survey by Guarino *et al.* (2006) shows that schools with more administrative support for teachers tend to have a lower teacher attrition rate. Note that Table 3 shows that teachers in the Netherlands prefer the relative size of the support staff to be low. In other words, they would rather work in schools that spend their budget on hiring more teachers than on hiring support staff. This result is therefore not inconsistent with previous findings.

We find significant preferences for more teaching hours. This is intuitive given that wages are set in terms of full-time equivalents. We suspect that there may be heterogeneity by age in the preferences for this variable at the extensive margin (two-thirds of the teachers in our sample do not have a full-time contract), an issue that we come back to below. Population density seems to have a negative effect on teachers' utility. Since wages are set by a fixed national scheme, teachers may prefer less densely populated areas where they would enjoy a higher real wage.

The preference parameter estimate for public schools is negative and borderline significant at the 10% level (p-value = 0.102). As we mentioned in subsection 2.1, public and private schools in the Netherlands mainly differ with respect to religion and to discretion in the way teaching is organised. Funding, wages and curriculum are the same. It thus seems that the limited differences between the two types of schools still affect the teachers' utility. However, like teaching hours, preferences for public schools may differ across individuals.

Lastly, we find that the school's student achievement plays a major role in teachers' preferences, especially when compared with the proportion of disadvantaged minority students (who score on average 1 standard deviation lower than non-minority students). Hanushek *et al.* (2004) also find that student achievement is one of the drivers of teacher turnover. Scafidi *et al.* (2007) show that the effect of test scores on turnover may be due to the correlation between this variable and other school characteristics, especially ethnic composition. Our results show that in the Netherlands, even after controlling for the education and nationality of students' parents, test scores still play an important role in teachers' preferences for schools.

3.2 Job Opportunities: Dependence Between Current and Outside Job Characteristics

While teacher preferences for school characteristics are the main targets of the estimation, our analysis of teacher turnover between schools also accounts for the heterogeneity of the search environment that teachers face when making their mobility decisions. Our approach thus produces new results on the dependence between the current job characteristics of a teacher and her outside job opportunities. The results reported in Table 4 show that many correlation estimates are significantly positive. However, the extent to which current job characteristics affect outside opportunities shows substantial variation across amenities.

If we look at the elements on the diagonal of Table 4, we note that a_j^* significantly and positively depends on a_j for all job attributes. For example, a teacher working in a school with a larger proportion of disadvantaged minority pupils is more likely to have access to an alternative school with a large proportion of similar students. We saw in Table 3 that working in a school with a large proportion of disadvantaged students has a negative effect on teachers' utility. Here we see that this also decreases her chances of moving to a school where this proportion is low.

Table 4 also shows strong dependence between the current teaching load of a teacher and her job opportunities, between the status (public or private) of the current and the outside schools, and between the population density of the area of the current and the outside schools. The same goes for the average test score in the current school and that in the outside schools. In contrast, for the last three amenities in the table, the dependence between the current job and job opportunities is weaker.

Our reduced-form modelling of the dependence between current and outside job characteristics precludes a structural interpretation of the correlations in Table 4. Teachers

Table 4: Estimates of job offer parameters (α), equation (5)

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
a_1^*	0.204^{***}	0.024^{**}	-1.809^{***}	0.022	0.194^{**}	0.116^{***}	-0.050^{***}	-0.089	-0.013	0.036^{***}
a_2^*	-0.057^{*}	0.161^{***}	1.617^{***}	-0.008	0.089	0.033	-0.011	-0.489	-0.010	-0.015
a_3^*	0.001	0.001^{***}	0.052^{***}	0.000	0.008^{*}	-0.003	-0.001^{*}	-0.070^{***}	0.001*	0.001
a_4^*	0.059^{***}	-0.005	-0.728^{***}	0.519^{***}	0.211^{***}	0.011	-0.014^{*}	-0.047	-0.002	0.004
a_5^*	0.027^{***}	-0.007^{***}	-0.245^{***}	0.008^{**}	0.477^{***}	0.002	-0.004^{*}	-0.109^{*}	-0.002	0.012^{***}
a_6^*	0.064^{***}	0.002	-1.408^{***}	0.031^{***}	0.083^{***}	0.629^{***}	-0.023^{***}	0.554^{***}	-0.011^{***}	0.003
a_7^*	-0.279^{***}	-0.172^{***}	0.838	0.097	0.000	-0.173	0.317^{***}	5.589^{***}	-0.040	0.008
a_8^*	-0.000	-0.001^{***}	0.008	0.002^{***}	-0.001	0.002	0.001^{**}	0.089^{***}	0.000	0.000
a_9^*	-0.115^{***}	-0.037^{**}	1.433^{**}	0.038	-0.129	0.024	0.046^{**}	0.244	0.050^{***}	-0.021
a_{10}^*	0.112^{***}	-0.010	-1.831^{***}	0.031	0.352^{***}	0.159^{***}	-0.025	-0.033	-0.011	0.172^{***}

Note: */**/*** statistically significant at the 10/5/1 percent level (standard errors available on request). Amenities are abbreviated as follows, a_1: Disadv. minority pupils, a_2: Disadv. Dutch pupils, a_3: Pupil-teacher ratio, a_4: Teacher hours,a_5: Pop. density, a_6: Pub. school, a_7: Student achievement, a_8: Age teachers, a_9: Female teachers, a_{10}: Support staff.

Table 5: Estimates of preference parameters and MWP: θ vs. ψ

	θ/θ_{PT}		ψ/ψ_{PT}	
Disadv. minority pupils	-43.2***	(15.6)	-37.4**	(14.9)
Disadv. Dutch pupils	-15.4	(10.7)	-16.5	(12.5)
Pupil-teacher ratio (PT)	ref.		ref.	
Teacher hours	26.8**	(13.0)	25.5*	(13.3)
Population density	-5.9*	(3.3)	-1.9	(1.6)
Public school	-14.5	(8.9)	-3.1	(2.7)
Student achievement	101.0*	(55.8)	65.6	(46.6)
Age teachers	1.0*	(0.6)	1.1	(0.7)
Female teachers	26.7*	(16.2)	19.7	(16.6)
Support staff	-54.1**	(26.9)	-50.8*	(28.6)

Note: */**/*** statistically significant at the 10/5/1 percent level.

working in schools with high proportions of disadvantaged pupils may develop specific teaching skills that are less relevant for teaching at schools with larger classes and in more affluent neighbourhoods. Location may play a role as well. For example, teachers working in a school with more disadvantaged pupils may live further from other types of schools and thus have limited access to job opportunities arising in these schools.

The fact that the degree of dependence varies across amenities is relevant for measuring teachers' preferences. Table 5 compares our estimates of MWP with the estimates from a simple turnover regression, i.e. based on the reduced-form equation (6). The MWP are computed taking the pupil teacher ratio as the reference amenity. The MWP estimates are qualitatively similar, so it is fair to say that the Gronberg and Reed (1994) approach paints a relatively accurate picture of what teachers value in our data. Still, we note differences in the size of the MWP estimates between the two methods which, in the case of the average test score, can be large (101 vs. 66). We find that this difference is significant at the 5% level for population density, and at the 10% level for public school and the school's student achievement.

The magnitude of the coefficient estimates $\widehat{\theta}$ and $\widehat{\psi}$ in Table 5 are substantially different. While, as we have just seen, this does not necessarily imply large differences in preference estimates (which are defined up to scale), this difference in magnitude does have implications for teacher turnover. The counterfactual exercise in Section 5 will illustrate this point.

4 Alternative Specifications

4.1 Unobserved teacher heterogeneity.

The results of the previous section are based on a model specification that does not allow for unobserved teacher heterogeneity. As we discussed in subsection 1.4, it is easy to augment the selection and outcome equations of model (5)-(6) with an individual-specific effect modelled, following Wooldridge (1995), as a linear function of A_{i1} and X_{i1} plus a normally distributed term. Table 6 reports the estimates of preference parameters θ when the model allows for different specifications of individual unobserved heterogeneity: when the individual intercepts are assumed to depend on X_{i1}, on A_{i1}, or on both X_{i1} and A_{i1}, respectively. For comparison, we also show in the first column the benchmark estimation results, without unobserved heterogeneity.

The first two columns show that there are essentially no differences between the preference parameters estimated in a model without unobserved heterogeneity and those estimated in a model that allows for individual effects correlated only with individual characteristics' initial values X_{i1}. If we look at the next column, we see that differences do arise when the individual effects are allowed to be correlated with the first observed values of job characteristics A_{i1}. Teachers now show more significant and more negative preferences for the proportion of disadvantaged Dutch pupils in the school. We also see a sign reversal for population density. Meanwhile, preferences are now less precisely estimated. Once we allow for the individual intercepts to be correlated with both A_{i1} and X_{i1}, the parameter estimates for teaching hours, population density, public schools and the school's student achievement are no longer significant. We note that the point estimates remain quite large and the loss of significance seems to come from a loss of precision due to the large number of parameters we need to introduce in order to account for unobserved heterogeneity in that specification.

Overall, the general qualitative picture of teacher preferences remains similar when allowing for unobserved heterogeneity. The main difference arises from the parameter estimates associated with the proportion of disadvantaged Dutch pupils and pupil-teacher

Table 6: Estimates of preference parameters (θ) - Individual heterogeneity

	No unobserved heterogeneity		Individual effect correlated with					
			X_1		A_1		X_1, A_1	
Disadv. minority	-0.410***	(0.079)	-0.389***	(0.073)	-0.458**	(0.180)	-0.513***	(0.158)
Disadv. Dutch	-0.146	(0.092)	-0.125	(0.088)	-0.696**	(0.304)	-0.736***	(0.283)
Pup.-teach. ratio	-0.009**	(0.004)	-0.009**	(0.004)	-0.031***	(0.007)	-0.032***	(0.006)
Teacher hours	0.254***	(0.070)	0.261***	(0.068)	0.253*	(0.131)	-0.007	(0.129)
Pop. density	-0.056***	(0.020)	-0.054***	(0.020)	0.103*	(0.059)	0.053	(0.055)
Public school	-0.137**	(0.063)	-0.134**	(0.061)	-0.218	(0.198)	-0.198	(0.178)
Student achievement	0.958**	(0.378)	0.893**	(0.370)	0.949**	(0.393)	0.698	(0.440)
Age teachers	0.010***	(0.003)	0.009***	(0.003)	0.036***	(0.006)	0.036***	(0.006)
Female teachers	0.253**	(0.110)	0.243**	(0.098)	1.336***	(0.228)	1.374***	(0.234)
Support staff	-0.513***	(0.100)	-0.505***	(0.094)	-0.538***	(0.186)	-0.510***	(0.187)

Note: */**/*** statistically significant at the 10/5/1 percent level.

Table 7: Estimates of preference parameters (θ) - Exclusion restriction

	Benchmark specification		Excluded covariates: Budget		Excluded covariates: Mat. Leave	
Disadv. minority	-0.410***	(0.079)	-0.413***	(0.077)	-0.365***	(0.088)
Disadv. Dutch	-0.146	(0.092)	-0.146	(0.095)	-0.129	(0.084)
Pup.-teach. ratio	-0.009**	(0.004)	-0.010**	(0.004)	-0.008*	(0.004)
Teacher hours	0.254***	(0.070)	0.256***	(0.070)	0.227***	(0.068)
Pop. density	-0.056***	(0.020)	-0.057***	(0.020)	-0.051**	(0.021)
Public school	-0.137**	(0.063)	-0.139**	(0.065)	-0.123**	(0.058)
Student achievement	0.958**	(0.378)	0.967**	(0.377)	0.855**	(0.358)
Age teachers	0.010***	(0.003)	0.010***	(0.003)	0.009***	(0.003)
Female teachers	0.253**	(0.110)	0.255**	(0.107)	0.225**	(0.102)
Support staff	-0.513***	(0.100)	-0.518***	(0.104)	-0.459***	(0.115)

Note: */**/*** statistically significant at the 10/5/1 percent level.

ratio, which increase in magnitude as the specification of heterogeneity gets more flexible, and with the parameter estimates associated with teaching hours, population density and the school's student achievement, which are less precise when unobserved teacher-specific effects are accounted for.

4.2 Exclusion restriction.

The benchmark estimation results assume that the school budget shock, Z^{bud}, and colleagues' parental leaves, Z^{pl}, enter the job change decision (4) but are excluded from the job offer equations (5). Since only one exclusion restriction is needed to identify the model, we can assess the sensitivity of our results to changes in the set of excluded covariates.

In Table 7, we show estimation results for the benchmark specification, as well as for specifications where either the school budget shock, or colleagues' parental leaves, is excluded from the outcome equation. Looking at the estimates, we note that taking either Z^{bud} or Z^{pl} out of the instrument set yields estimates of preference parameters that are still close to those found in our benchmark specification.

4.3 Non-normal specification.

All results presented so far rely on normality assumptions. In order to check that our findings do not hinge on normality, we next report results based on a more flexible speci-

Table 8: Estimates of preference parameters (θ) - Non-normal specification

	Excluded covariates			
	Z^{bud} and Z^{pl}		None	
Disadv. minority pupils	-0.470***	(0.101)	-0.639	(11.81)
Disadv. Dutch pupils	-0.173	(0.114)	-0.072	(2.28)
Pup.-teach. ratio	-0.011**	(0.005)	-0.018	(0.34)
Teacher hours	0.327**	(0.138)	-0.357	(2.94)
Population density	-0.066***	(0.025)	-0.226	(3.21)
Public school	-0.158**	(0.076)	-0.267	(7.40)
Student achievement	1.109**	(0.465)	1.984	(57.15)
Age teachers	0.011***	(0.004)	0.011	(0.26)
Female teachers	0.295**	(0.121)	0.179	(1.36)
Support staff	-0.599***	(0.132)	-0.647	(15.80)

Note: */**/*** statistically significant at the 10/5/1 percent level.

fication. We build on Newey (2009) to adapt the three-step estimation strategy presented in Section 1.3. In the first step, we flexibly estimate the parameters of the reduced-form mobility equation following a series logit approach:

$$Q_{it} = \mathbf{1}\left\{\Lambda\left(\psi A_{it} + \psi_X X_{it} + \theta_Z Z_{it}\right) + \nu_{it} > 0\right\}, \tag{11}$$

where $\Lambda(\cdot)$ is a polynomial function, and where ν_{it} follows a logistic distribution. In the second step, we estimate equation (8) where the expectation on the right-hand side is now specified as a polynomial function of the index, $\psi A_{it} + \psi_X X_{it} + \theta_Z Z_{it}$, estimated in first step. In practice we use second-order polynomials. The third estimation step is unchanged.

The first column in Table 8 shows that the results that rely on the flexible specification and our two excluded covariates are similar to the benchmark estimates from Table 3. This provides evidence that the estimates shown in Section 3 are not driven by normality. In addition, the second column in the table shows the results of the same specification, but now removing the two excluded covariates. We can see that the point estimates are rather different. Moreover, standard errors become very large. This suggests that the estimates in the first column of the table are mostly driven by the power of the exclusion restrictions, as opposed to functional forms assumptions.

4.4 *Additional exercises.*

In the online appendix we report the results of specifications that include an indicator for full-time contract and school size (that is, number of pupils) as additional amenities. We also show a specification that allows the effect of the proportion of disadvantaged minority pupils, an amenity which according to our results is very relevant to teachers' decisions, to be nonlinear. We find that the parameters associated with the other amenities are very similar to the ones reported in Section 3. Lastly, we show the results of a specification where teachers' preferences vary with age, and document some preference heterogeneity, in particular for working hours.

5 Counterfactual Analysis

In this section, we show that disentangling the effects of preferences (θ) and opportunities (α) on teacher turnover can be relevant from a policy perspective. More specifically, although a social planner may not be able to affect individual preferences, it may be possible to manipulate the distribution of job offers by facilitating or blocking the access to specific job offers for some groups. Given that teacher labour markets are more regulated than most labour markets, such policies are realistic and are actually already implemented in some countries. As an example, in France teacher turnover is ruled by an experience rating system whereby teachers who have accumulated more points, for instance by working in a disadvantaged school, have access to a wider set of schools.[24]

We illustrate the effect of such policy interventions by changing the dependence between current and outside job offers, and document teacher turnover and post-mobility distributions of job characteristics for different values of the α parameters. We consider two cases: the benchmark case where the model parameters are set to their estimated values, and a counterfactual scenario where job offers are independent of the current school characteristics. With this scenario we attempt to capture, albeit in an artificial environment, the effect of policies that aim at improving the access of teachers to a different set

[24]For information on the French system, see for example
http://www.education.gouv.fr/cid53746/mutation-des-personnels-enseignants-du-premier-degre.html

of schools.

It is important to note that this exercise only captures short-term, partial-equilibrium effects. This is because, as we mentioned in Section 1, our approach recovers the MWP associated with individual value functions, which, unlike utility functions, depend on the distribution of outside job characteristics. This distribution may change in the long term because of demand-side effects. For instance, if the policy increases the quit rate for disadvantaged schools and does not affect the quit rate of other schools, the proportion of disadvantaged pupils among offers will increase and, because of congestion effects in richer schools, the correlation between this proportion and teaching hours could decrease. In the short term, however, it is realistic to assume that teachers have not yet factored in the change in the offer distribution when taking their mobility decision. Also, we are not modelling how a school characteristic (such as student achievement) may respond to changes in teacher turnover. Again, in the short term, we assume that these characteristics are not affected by changes in α. These issues must be kept in mind when interpreting the results below.

Since the α matrix drives the dependence between A^* and A, see equation (5), we set it to its estimated value (the benchmark, shown in Table 4) or to 0 (counterfactual). All the other model parameters are kept at their estimated value. The ex-ante distribution of teacher and school characteristics, (X, A), is taken from the data.

In this counterfactual exercise we choose to focus on two specific outcomes. First, we predict the probability that each teacher leaves her current job using equation (6), where the residual follows a normal distribution and where the ψ parameters are composites of preference parameters θ and of α (see equation (7)). Secondly, we compute the average job characteristics conditionally on changing job, $\mathbb{E}\left(A^*|A, X, Z, Q=1\right)$, using equation (8) and assuming normality.[25]

We start with the job quit probability. Table 9 reports, for each school characteristic

[25] To facilitate the comparison between job characteristics before and after a job change, we ensure that the marginal distribution of counterfactual offers is the same as the distribution of offers produced by our benchmark estimation. To do this, we set α to its counterfactual value and apply an affine transformation so that the counterfactual job offers have the same mean and variance as the estimated one (and thus the same distribution as we assume normality).

Table 9: Benchmark and counterfactual probabilities (in %) to leave a school, conditional on current job characteristic (quintile)

	Benchmark: $\alpha = \widehat{\alpha}$					Counterfactual: $\alpha = 0$				
	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
Disadv. minority	3.2	4.0	3.1	3.4	4.9	2.5	2.5	2.9	3.6	7.3
Disadv. Dutch	3.6	3.7	3.4	3.4	3.6	3.6	4.2	3.6	3.6	3.9
Pupil-teacher ratio	4.6	3.4	3.1	3.1	3.3	6.7	3.6	2.9	2.8	2.9
Teacher hours	4.0	3.1	3.1	3.6	4.9	4.1	3.3	3.4	3.9	5.7
Population density	2.9	3.1	3.1	3.9	4.8	2.6	2.8	3.0	4.6	6.1
Student achievement	4.8	3.7	3.3	3.0	2.9	7.3	4.1	3.1	2.5	2.0
Age teachers	4.3	3.7	3.4	3.2	2.9	4.7	3.9	3.6	3.5	3.2
Female teachers	3.3	3.6	3.6	3.6	3.6	3.8	4.0	3.9	3.7	3.4
Support staff	3.1	3.2	3.2	3.6	4.4	2.9	3.2	3.1	4.0	5.7

Note: All probabilities are in %. Q_k denotes the k-th quintile of a given amenity a_j.

a_j, the probability to leave one's school conditionally on the quintile of a_j. On the left (first five columns) we have the benchmark case, and on the right (next five columns) the counterfactual case with $\alpha = 0$. Looking at the first row, we see that in the benchmark case, teachers working in schools with a low (resp. high) proportion of disadvantaged minority students are less (resp. more) likely to leave their job. In the counterfactual case, we see that the job quit probability decreases for teachers in the lowest quintiles and increases for teachers in the highest quintiles. This illustrates the effect of job opportunities. When α goes from its estimated value to 0, teachers in schools with few disadvantaged pupils become more likely to receive offers from schools with many disadvantaged students. Since their preferences have not changed, they thus tend to stay more in their current schools (the quit probability goes from 3.2% to 2.5%). For teachers in schools with high proportions of disadvantaged pupils, the effect is the opposite as these teachers now have improved access to schools with fewer disadvantaged pupils and are thus more likely to leave their school (the probability goes from 4.9% to 7.3%). In contrast, severing the link between current job characteristics and job opportunities has little impact on turnover rates across schools with different fractions of disadvantaged Dutch students.

If we now consider a characteristic that teachers value positively, for example student achievement (on the sixth row of Table 9), we see that shutting down the dependence between A^* and A tends to increase turnover for teachers working in schools with low

Table 10: Job characteristics: Quintile transitions after a job change.

	Benchmark: $\alpha = \widehat{\alpha}$				Counterfactual: $\alpha = 0$			
	$\Delta \neq 0$	$\Delta > 0$	$\Delta < 0$	$\frac{\Delta>0}{\Delta\neq 0}$	$\Delta \neq 0$	$\Delta > 0$	$\Delta < 0$	$\frac{\Delta>0}{\Delta\neq 0}$
Disadv. minority	64.3	52.9	11.4	82.3	83.2	55.8	27.4	67.0
Disadv. Dutch	73.7	41.2	32.5	55.9	78.9	44.0	34.9	55.8
Pupil-teacher ratio	71.2	28.4	42.8	39.9	80.9	32.6	48.3	40.3
Teacher hours	53.9	21.1	32.8	39.1	74.9	33.7	41.2	45.0
Population density	46.3	19.1	27.2	41.3	73.1	32.7	40.4	44.7
Student achievement	77.9	37.4	40.5	48.0	84.7	40.3	44.4	47.6
Age teachers	74.4	33.3	41.1	44.8	78.2	35.7	42.5	45.7
Female teachers	77.2	41.6	35.6	53.9	79.0	41.8	37.2	52.9
Support staff	69.1	47.3	21.8	68.5	79.6	49.7	29.9	62.5

Note: All figures are in %. $\Delta \neq 0$ (resp. $\Delta > 0$, $\Delta < 0$) gives the proportion of teachers whose average amenity after a job change is in a different (resp. higher, lower) quintile than their original amenity.

average achievement, as these teachers get more access to better-performing schools. In contrast, teachers in schools with high student achievement tend to stay more in their current school, because the average student achievement among their outside opportunities has decreased with respect to the benchmark case.

Our approach allows us not only to study quit probabilities, but also to predict the distribution of job characteristics posterior to job change. In particular, it is well-suited to analyse whether changing job allows teachers to move up or down the distribution of job characteristics, and how this mobility along the distribution is affected by the dependence between current and outside jobs.

We address this issue as follows: for a teacher in a given amenity quintile we compute her average amenity after a job change and its corresponding quintile.[26] Table 10 shows summary statistics on transitions between quintiles. Its main message is that removing the dependence between current and outside job characteristics results in more mobility between school types. In the case of disadvantaged minority pupils, most of the increase is driven by downward mobility, as teachers in high quintiles (that is, those working in schools with many disadvantaged pupils) have more access to schools in lower quintiles. Teachers in low quintiles may have more offers from schools with a high proportion of

[26] All quintiles are computed with respect to the ex-ante distribution of job characteristics A.

disadvantaged pupils, but they can reject these offers (unless they are hit by a large shock and are forced to move). For other job characteristics we see that the increase in mobility is more evenly spread between upward and downward changes.

This counterfactual exercise illustrates that affecting the distribution of job opportunities can be used as a policy tool to affect the reallocation of teachers across schools through turnover. For a more thorough welfare analysis, one would need more structure. In particular, one would need to take a stand on the objective function (such as student achievement, teacher lifetime utility, or inequality between schools). This section provides a first illustration of the potential of affecting teacher turnover not only through compensation for school characteristics, and thus preferences, but also through the outside job opportunities of teachers working in specific schools.

6 Conclusion

In this paper we argue that job characteristics can affect worker turnover not only through their preferences, but also through their effect on job opportunities. We propose a simple three-step method to estimate these two effects. Taking our model to an administrative data set of primary school teachers in the Netherlands, we obtain estimates of teacher preferences for schools that complement earlier results in the literature (e.g. Hanushek *et al.*, 2004, Scafidi *et al.*, 2007). We also show that the dependence between current and outside job attributes has an impact on labour turnover. This suggests that affecting the availability of job opportunities may provide an effective policy instrument. As an illustration, we perform a counterfactual analysis where we remove the dependence between current and outside jobs.

We see two natural extensions to our work, both in a structural direction. First, as we mention earlier, our estimates of individual preferences are based on value functions, not on instantaneous utilities. They are thus sensitive to the offer distribution, and our counterfactual analysis is only valid in the short term. To recover the primitive preference parameters, one would need to solve a dynamic problem with potentially high-dimensional state variables (as there can be many amenities). The other extension would consist in putting more economic structure in order to study the full equilibrium effects of policies aiming at re-allocating teachers across schools. Indeed, these policies may lead schools to change the contracts they offer, and student achievement will probably be affected by the departures and arrivals of teachers induced by these new incentives.

Appendix

A Nonparametric Identification

The semi-parametric identification result of Proposition 1 relies on several linear index restrictions imposed on value functions, mobility costs, and outside job characteristics, respectively; see equations (4) and (5). The following result, proved in the online appendix, shows that it is possible to relax these assumptions and achieve fully nonparametric identification of the MWP for job amenities.

PROPOSITION 2. *Consider the general setup of equation (2). Suppose that the characteristics A^* of outside jobs are statistically independent of the cost shifters Z, conditionally on the current job's amenities A and worker characteristics X. In addition, suppose that mobility costs C are independent of current job's attributes A given $(A^*, V(A, X), X, Z)$. Lastly, suppose that the technical assumption 1 in the online appendix is satisfied. Then, the marginal willingness to trade $MWP_{jk}(A, X)$ given by equation (1) is non-parametrically identified for all $j \neq k$.*

As in the semi-parametric case, two key conditional independence assumptions are needed: between cost shifters Z and outside jobs' characteristics A^* on the one hand, and between mobility costs C and current jobs' characteristics A on the other hand. Nevertheless, the nonparametric setup of Proposition 2 is substantially more general than the setup of Proposition 1. In particular, it allows for a flexible formulation of preference parameters, as given by equation (1). Also, Proposition 2 shows that identification can be achieved without imposing that outside amenities A^* are linear in A and X.

Proposition 2 provides a basis to conduct a fully nonparametric analysis of workers' preferences. In the context of our empirical application, however, such a nonparametric approach raises practical problems. Since we allow teachers to base their mobility decisions on ten different school attributes, nonparametric estimation would face a severe curse of dimensionality in our data set.

B Additional Results

Table B1: Estimated reduced form turnover equation (6)

Amenities:		
Disadv. minority pupils	0.247***	(0.054)
Disadv. Dutch pupils	0.109	(0.066)
Pupil-teacher ratio	0.007**	(0.003)
Teacher hours	-0.169***	(0.028)
Population density	0.012	(0.009)
Public school	0.021	(0.016)
Student achievement	-0.434**	(0.213)
Age teachers	-0.007***	(0.002)
Female teachers	-0.130	(0.091)
Support staff	0.337***	(0.061)
Individual characteristics:		
Age 21	0.111*	(0.060)
Age 22	0.052	(0.046)
Age 23	-0.013	(0.039)
Age 24	-0.043	(0.038)
Age 25	0.005	(0.036)
Age 20-29	0.620***	(0.038)
Age 30-39	0.497***	(0.027)
Age 40-49	0.270***	(0.023)
ln(wage)	0.026	(0.109)
On maternity leave	-0.674***	(0.054)
Tenure (rank in school)	0.060**	(0.025)
Temporary contract	0.897***	(0.023)
School rank at municipality level	0.010	(0.051)
School rank at district level	0.057	(0.100)
Local labour market controls:		
Sum Z^{bud} at municipality level	0.063*	(0.036)
Sum Z^{bud} at district level	0.022	(0.020)
Region = North	-0.045	(0.071)
Region = South	-0.065*	(0.038)
Region = East	-0.075***	(0.024)
UI rate (Province)	-0.060*	(0.032)
Vacancy rate (Province)	3.453	(2.216)
Unemp. rate (Region)	0.019	(0.037)
Δ Unemp. rate (Region)	-0.104	(0.064)
Exclusion restrictions:		
Z^{bud}	-0.003***	(0.000)
Z^{pl}	-0.043***	(0.014)
Intercept	-2.068**	(0.880)

Note: */**/*** statistically significant at the 10/5/1 percent level.

University of Chicago
University of Bristol
University of Oslo, CEPR, CESifo, ESOP, IZA and Statistics Norway

References

Akerlof, R., Rose, A. and Yellen, J. (1988). 'Job switching and job satisfaction in the us labor market', *Brooking Papers on Economic Activity*, vol. 1988(2), pp. 495–594. (document)

Bonhomme, S. and Jolivet, G. (2009). 'The pervasive absence of compensating differentials', *Journal of Applied Econometrics*, vol. 24(5), pp. 763–795. 14

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M. and Wyckoff, J. (2011). 'The role of teacher quality in retention and hiring: Using applications-to-transfer to uncover preferences of teachers and schools', *Journal of Policy Analysis and Management*, vol. 30(1), pp. 88–110. (document)

Boyd, D., Lankford, H., Loeb, S. and Wyckoff, J. (2005). 'Explaining the Short Careers of High-Achieving Teachers in Schools with Low-Performing Students', *American Economic Review*, vol. 95(2), pp. 166–171. (document), 17

Bradley, J., Postel-Vinay, F. and Turon, H. (2013). 'Public sector wage policy and labor market equilibrium: a structural model', . 2, 14

Burdett, K. (1978). 'A theory of employee job search and quit rates', *American Economic Review*, vol. 68(1), pp. 212–220. 1

Das, M., Newey, W. and Vella, F. (2003). 'Nonparametric Estimation of Sample Selection Models', *Review of Economic Studies*, vol. 70(1), pp. 33–58. 1.2

Dolton, P. and Van Der Klaauw, W. (1995). 'Leaving teaching in the uk: A duration analysis', *Economic Journal*, vol. 105, pp. 431–444. 8

Dolton, P. and Van Der Klaauw, W. (1999). 'The turnover of teachers: A competing risks explanation', *Review of Economics and Statistics*, vol. 81(3), pp. 543–550. 8

Dustmann, C. and Meghir, C. (2005). 'Wages, Experience and Seniority', *Review of Economic Studies*, vol. 72(1), pp. 77–108. 3

Dustmann, C. and Rochina-Barrachina, M. (2007). 'Selection Correction in Panel Data Models: An Application to the Estimation of Females' Wage Equations', *Econometrics Journal*, vol. 10(2), pp. 263–293. 10

Eurydice (2008). 'Organisation of the Education System in the Netherlands', . 15

Freeman, R. (1978). 'Job satisfaction as an economic variable', *American Economic Review*, vol. 68(2), pp. 135–141. (document)

Gibbons, R. and Katz, L. (1992). 'Does Unmeasured Ability Explain Inter-Industry Wage Differentials', *Review of Economic Studies*, vol. 59(3), pp. 515–535. 3

Gronberg, T. and Reed, W. (1994). 'Estimating workers' marginal willingness to pay for job attributes using duration data', *Journal of Human Resources*, vol. 29(3), pp. 911–931. (document), 1.1

Guarino, C., Santibanez, L. and Daley, G. (2006). 'A Review of the Research Literature on Teacher Recruitment and Retention', *Review of Educational Research*, vol. 76, pp. 173–208. 3.1

Hall, R. (1972). 'Turnover in the labor force', *Brooking Papers on Economic Activity*, vol. 1972(3), pp. 709–764. 1

Hanushek, E., Kain, J. and Rivkin, S. (2004). 'Why public schools lose teachers', *Journal of Human Resources*, vol. 39(2), pp. 623–654. (document), 17, 3.1, 3.1, 6

Heckman, J. (1976). 'The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models', *Annals of Economic and Social Measurement*, vol. 5(4), pp. 475–492. (document)

Heckman, J. (1979). 'Sample Selection Bias as a Specification Error', *Econometrica*, vol. 47, pp. 153–161. 1.3

Hoxby, C. (2000). 'The Effects of Class Size on Student Achievement: New Evidence from Population Variation', *Quarterly Journal of economics*, vol. 115(4), pp. 1239–1285. 21

Hwang, H., Mortensen, D. and Reed, W. (1998). 'Hedonic Wages and Labor Market Search', *Journal of Labor Economics*, vol. 16(4), pp. 815–847. (document)

Jovanovic, B. (1979). 'Job matching and the theory of turnover', *Journal of Political Economy*, vol. 87(5), pp. 972–990. 1

Kyriazidou, E. (1997). 'Estimation of a Panel Data Sample Selection Model', *Econometrica*, vol. 65, pp. 1335–1364. 10

Meghir, M., Narita, R. and Robin, J. (2012). 'Wages and informality in developing countries', NBER. 2

Newey, W. (2009). 'Two-step series estimation of sample selection models', *Econometrics Journal*, vol. 12, pp. S217–S229. 4.3

Scafidi, B., Sjoquist, D. and Stinebrickner, T. (2007). 'Race, Poverty, and Teacher Mobility', *Economics of Education Review*, vol. 26(2), pp. 145–159. 3.1, 3.1, 6

Semykina, A. and Wooldridge, J. (2007). 'Estimation of Dynamic Panel Data Models with Sample Selection', Mimeo. 10

Topel, R. and Ward, M. (1992). 'Job mobility and the careers of young men', *Quarterly Journal of Economics*, vol. 107(2), pp. 439–479. (document)

Wooldridge, J. (1995). 'Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions', *Journal of Econometrics*, vol. 68(1), pp. 115–132. 1.4.2, 11, 4.1

Online Appendix — School Characteristics and Teacher Turnover: Assessing the Role of Preferences and Opportunities

This online appendix provides: the proofs of the propositions in the paper (Section C), estimates that allow for preference heterogeneity by age (Section D), descriptive statistics on amenities (Section E) and teacher exits (Section F), and estimates of additional specifications (Section G). In all tables, bootstrapped standard errors are shown in parenthesis.

C Proofs of Propositions

C.1 Proof of Proposition 1

Let F denote the cumulative distribution function of $-\eta$, and let f denote its pdf. We assume that $f > 0$. Then, by (6) we have:

$$\Pr(Q = 1|A, X, Z) = F\left(\psi A + \psi_X X + \theta_Z Z\right).$$

Hence $\psi_j/\psi_k = \frac{\partial \Pr(Q=1|A,X,Z)}{\partial a_j} / \frac{\partial \Pr(Q=1|A,X,Z)}{\partial a_k}$ is identified.

Moreover, by (8):

$$\mathbb{E}\left(A^*|A,\, X,\, Z,\, Q = 1\right) = \alpha A + \alpha_X X + G\left[F\left(\psi A + \psi_X X + \theta_Z Z\right)\right],$$

where $G(u) = \mathbb{E}\left[\varepsilon|F^{-1}(u) + \eta > 0\right]$. We will denote as $g(\cdot)$ the derivative of $G(\cdot)$ (a $J \times 1$ vector).

We have:

$$\begin{aligned}
\frac{\partial \mathbb{E}\left(A^*|A,\, X,\, Z,\, Q = 1\right)}{\partial A'} &= \alpha + f\left(\psi A + \psi_X X + \theta_Z Z\right) g\left[F\left(\psi A + \psi_X X + \theta_Z Z\right)\right]\psi, \\
\frac{\partial \mathbb{E}\left(A^*|A,\, X,\, Z,\, Q = 1\right)}{\partial Z'} &= f\left(\psi A + \psi_X X + \theta_Z Z\right) g\left[F\left(\psi A + \psi_X X + \theta_Z Z\right)\right]\theta_Z.
\end{aligned}$$

Hence, if the row vector θ_Z is not identically zero:

$$f\left(\psi A + \psi_X X + \theta_Z Z\right) g\left[F\left(\psi A + \psi_X X + \theta_Z Z\right)\right] = \frac{\partial \mathbb{E}\left(A^*|A,\, X,\, Z,\, Q = 1\right)}{\partial Z'} \frac{\theta_Z'}{\theta_Z \theta_Z'},$$

so that:

$$\alpha = \frac{\partial \mathbb{E}\left(A^*|A,\, X,\, Z,\, Q = 1\right)}{\partial A'} - \frac{\partial \mathbb{E}\left(A^*|A,\, X,\, Z,\, Q = 1\right)}{\partial Z'} \frac{\theta_Z' \psi}{\theta_Z \theta_Z'}$$

is identified.

Lastly, by (7) identification of α and identification of ψ_j/ψ_k imply that θ_j/θ_k is identified, provided $\alpha \neq I_J$.

C.2 Proof of Proposition 2

Throughout the proof, let f be a generic notation for a distribution function. We need the following technical assumption.

ASSUMPTION 1. *i)* $f(A^*|A, X) > 0$ almost surely. *ii)* $\Pr(Q = 1|A^*, A, X, Z) > 0$ almost surely. *iii)* C has an absolutely continuous distribution given (A^*, A, X, Z). *iv)* For some Z_1 and Z_2 in the support of Z:

$$\frac{\partial}{\partial a_j}\left[\frac{\Pr(Q = 1|A^*, a_j, A_{-j}, X, Z_2)}{\Pr(Q = 1|A^*, a_j, A_{-j}, X, Z_1)}\right] \neq 0 \quad \text{a.s.,} \qquad \text{for all } j.$$

Assumption 1*i)* rules out situations where the support of amenity offers depends on current amenities and worker characteristics. Assumption 1*ii)* states that for any values of the vector (A^*, A, X, Z) workers have a strictly positive probability of changing job. This assumption has a technical purpose, as our identification proof relies on ratios of job change probabilities to exploit the variation in the excluded variable. Assumption 1*iii)* requires mobility costs to be continuous. Lastly, Assumption 1*iv)* is a technical assumption that states that the ratio of job change probabilities for two values of the cost shifter is a non-trivial function of amenities of the current job. As the proof below shows, this ratio can be recovered from the data, so Assumption 1*iv)* is testable.

To prove Proposition 2, first define $B(A^*, A, X, Z)$, available in the data, as:

$$B(A^*, A, X, Z) = \Pr(Q = 1|A, X, Z) f(A^*|A, X, Z, Q = 1).$$

Using Bayes' rule we have and the conditional independence between A^* and Z we have:

$$B(A^*, A, X, Z) \;=\; \Pr(Q = 1|A^*, A, X, Z) f(A^*|A, X).$$

Using Assumptions 1*i)* and 1*ii)* we obtain, for Z_1 and Z_2 in the support of Z:

$$\frac{B(A^*, A, X, Z_2)}{B(A^*, A, X, Z_1)} \;=\; \frac{\Pr(Q = 1|A^*, A, X, Z_2)}{\Pr(Q = 1|A^*, A, X, Z_1)},$$

where the offer distribution has been "differenced out".

Lastly, using (2) we also have:

$$\begin{aligned}\Pr(Q = 1|A^*, A, X, Z) &= \Pr[C < V(A^*, X) - V(A, X) \mid A^*, A, X, Z] \\ &= \Pr[C < V(A^*, X) - V(A, X) \mid A^*, V(A, X), X, Z]\end{aligned}$$

by Assumption 1*iii)* and the fact that C and A are conditionally independent. This quantity depends on A only through $V(A, X)$. Taking (Z_1, Z_2) as in Assumption 1*iv)* thus implies that:

$$\text{MWP}_{jk}(A, X) = \frac{\partial V(A, X)}{\partial a_j}\Big/\frac{\partial V(A, X)}{\partial a_k} = \frac{\frac{\partial}{\partial a_j}\left[\frac{\Pr(Q=1|A^*, A, X, Z_2)}{\Pr(Q=1|A^*, A, X, Z_1)}\right]}{\frac{\partial}{\partial a_k}\left[\frac{\Pr(Q=1|A^*, A, X, Z_2)}{\Pr(Q=1|A^*, A, X, Z_1)}\right]} = \frac{\frac{\partial}{\partial a_j}\left[\frac{B(A^*, A, X, Z_2)}{B(A^*, A, X, Z_1)}\right]}{\frac{\partial}{\partial a_k}\left[\frac{B(A^*, A, X, Z_2)}{B(A^*, A, X, Z_1)}\right]},$$

so $\text{MWP}_{jk}(A, X)$ is non-parametrically identified for all $j \neq k$.

D Preference Heterogeneity by Age

Our approach may easily be extended to allow for observed teacher heterogeneity in preferences, simply by allowing preference parameters θ to depend on X. Teachers may

Table D1: Preference estimates (θ) by age group

	20-26		27-39		40-60	
Disadv. minority pupils	-0.308**	(0.155)	-0.329***	(0.111)	-0.527***	(0.166)
Disadv. Dutch pupils	-0.373*	(0.196)	0.005	(0.139)	-0.158	(0.185)
Pupil-teacher ratio	-0.009	(0.009)	-0.011*	(0.006)	-0.010	(0.007)
Teacher hours	0.884***	(0.181)	-0.146	(0.136)	-0.040	(0.200)
Population density	-0.023	(0.039)	-0.032	(0.026)	-0.120**	(0.051)
Public school	0.159	(0.107)	-0.150*	(0.087)	-0.412**	(0.204)
Student achievement	1.209	(0.783)	1.217*	(0.717)	1.464*	(0.781)
Age teachers	-0.021**	(0.008)	-0.005	(0.005)	0.042***	(0.008)
Female teachers	0.113	(0.238)	0.360**	(0.182)	0.312	(0.195)
Support staff	-0.069	(0.212)	-0.672***	(0.174)	-0.631***	(0.225)

Note: */**/*** statistically significant at the 10/5/1 percent level.

value different amenities at different points in their life. Table D1 presents preference estimates for three age groups that roughly correspond to young school teachers who are starting their career (aged 20 to 26), mid-career teachers who often need to combine work with raising young children (aged 27 to 39), and teachers aged 40 to 60.

As can be seen from Table D1, these estimates stratified by age group are less precise than the ones based on the whole population and shown in Table 3. We nevertheless see some interesting patterns arise. First, teachers in all age groups prefer schools with fewer disadvantaged minority pupils. Young teachers also prefer schools with fewer disadvantaged Dutch pupils. Though not always precisely estimated, teachers' taste for smaller classes and also for higher student achievement is remarkably similar across age groups.

We also observe some differences. With age teachers value working hours differently. In particular younger teachers seem to value more hours, whereas the point estimate suggests that mid-career teachers want to decrease the amount of time they work. We also see that older teachers prefer to work in less densely populated areas and appear to prefer working in private schools. As to the composition of the school staff, colleagues' age seems to be valued differently by younger and older teachers: the younger teachers prefer younger colleagues whereas the 40 to 60-year-old teachers prefer to work with older colleagues. The overall picture that emerges is that, although there is some heterogeneity across age groups, teachers' preferences for amenities such as student composition, class size, and average student achievement seem to vary little with age.

E Additional Descriptive Statistics

In Figure E1 we show the evolution of job characteristics over the life cycle of teachers. The graphs show the average level of each job attribute (on the vertical axis) by age (on the horizontal axis). Interestingly, we see no clear pattern for the "student achievement" variable. Looking at population density, we see that teachers work in more populated areas at the beginning and at the end of their working life, while going to less dense areas in their thirties and forties. We also observe a fall in teaching hours at around the same time, followed by an increase after forty. This may be related to the birth of children. The pupil-teacher ratio decreases with age after forty and the proportion of disadvantaged minority pupils increases. Finally, there is an increasing pattern for average teacher age in the school (which could be partly mechanical), and a declining pattern for the proportion of female teachers. Note that this evidence might partly reflect cohort effects in addition to life-cycle effects.

F Descriptives on Teacher Exits

We compute an indicator for exits from the teacher labour market by looking at whether a given teacher leaves our data (i.e. no longer has a contract with a school) between two consecutive years. Figure F2 shows the density of age for leavers (red, solid) and non-leavers (blue, dashed).

To get a more detailed view of exits, we regress our exit indicator on job (A) and individual (X) characteristics by probit. The results are in Table F2 (standard errors in parentheses). We see in particular that teachers in schools where average student achievement is high tend to stay in the market. We also note that exits from the teacher labour market are more likely when the local labour market tightness is high, as the probability of exit increases with vacancies and decreases with unemployment.

G Additional Specifications

Here we report the results of three specifications, which account for the effect of part-time versus full–time contracts, school size, and nonlinearities in the effect of an amenity.

G.1 Full- and part-time contracts

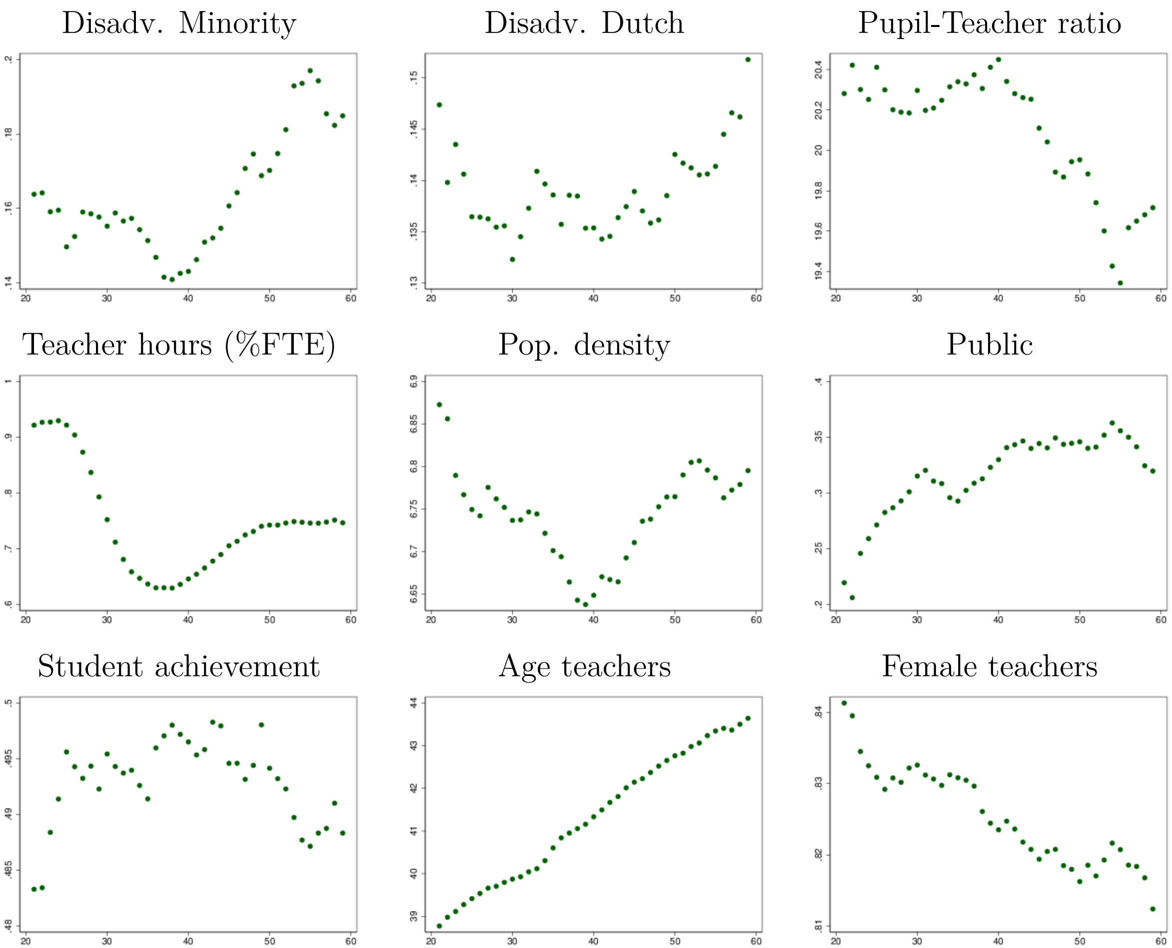
We observe the teaching load as a fraction of full-time. If one sets the threshold at 50% (resp. 60%, 70%, or 80%) of a full teaching load, the proportion of part-time teachers equals 19% (resp. 34%, 52%, or 62%). If we replace the "teacher hours" job characteristic with a full-time dummy, where the threshold was set at 60%, we get the preference parameter estimates θ in Table G3. Comparing these estimates with the benchmark ones (Table 3 in the main text), we see that the estimates for the other amenities are not affected and that the preference parameter for full time is positive, which is consistent with our baseline results (with teacher hours included as an amenity).

G.2 School Size as a Job Characteristic

We next include school size (number of pupils) in the job characteristics vector and run our estimation procedure. The estimates of the preference parameters θ are in Table G4. Comparing these results with those in Table 3 we note that the estimates for the other job attributes are not qualitatively affected by the inclusion of school size. We also note that the θ-parameter estimate for school size is positive and significant, indicating that teachers prefer to work in larger schools. It is difficult to interpret this result further, however, given that the pupil population also enters other amenities such as the proportion of disadvantaged pupils or the pupil-teacher ratio.

G.3 Nonlinear Preferences

A simple approach to allow for some nonlinearities is to create several amenities out of one, and to assume that the linear specification for job offers, equation (5), still holds



Note: age on the x-axis, average amenity on the y-axis.

Figure E1: Average amenities by teacher age

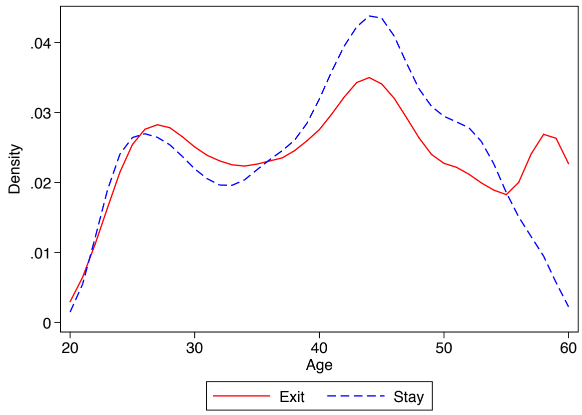


Figure F2: Density of age for teachers who leave or stay in the teacher labour market

Table F2: Probit regression of teacher exit on characteristics

Disadv. minority.	0.006	(0.036)
Disadv dutch	-0.097	(0.049)
Pupil-teacher ratio	-0.0169	(0.002)
Teacher hours	-0.340	(0.022)
Population density	-0.043	(0.006)
Public school	0.071	(0.012)
Student performance	-.454	(0.052)
Age teachers	-0.006	(0.002)
Female teachers	0.211	(0.061)
Support staff	0.553	(0.048)
Age 21	-0.203	(0.096)
Age 22	-0.152	(0.058)
Age 23	-0.024	(0.042)
Age 24	-0.030	(0.038)
Age 25	0.008	(0.036)
Age 20-29	-0.065	(0.029)
Age 30-39	-0.180	(0.021)
Age 40-49	-0.284	(0.016)
ln(wage)	0.079	(0.084)
On maternity leave	0.149	(0.029)
Tenure (rank in school)	0.037	(0.020)
Temporary contract	0.103	(0.023)
Region = North	0.604	(0.059)
Region = South	-0.303	(0.027)
Region = East	0.231	(0.017)
UI rate (Province)	0.530	(0.026)
Vacancy rate (Province)	78.079	(1.802)
Unemp. rate (Region)	-0.282	(0.029)
Δ Unemp. rate	0.555	(0.049)
Intercept	-3.401	(0.679)

Table G3: Preference estimates (θ) with a full-time dummy

	θ_j	
Disadv. minority pupils	-0.400	(0.079)
Disadv. Dutch pupils	-0.148	(0.093)
Pupil-teacher ratio (PT)	-0.009	(0.004)
Full time	0.065	(0.036)
Population density	-0.056	(0.019)
Public school	-0.141	(0.062)
Student achievement	0.926	(0.393)
Age teachers	0.009	(0.003)
Female teachers	0.239	(0.108)
Support staff	-0.515	(0.102)

Table G4: Preference estimates (θ) with school size as an amenity

	θ_j	
Disadv Minority pupils	-0.370	(0.087)
Disadv. Dutch pupils	-0.062	(0.094)
Pupil-teacher ratio (PT)	-0.009	(0.004)
Teacher hours	0.239	(0.077)
Population density	-0.084	(0.023)
Public school	-0.159	(0.064)
Student achievement	1.009	(0.420)
Age teachers	0.012	(0.003)
Female teachers	0.374	(0.118)
Support staff	-0.465	(0.104)
School size	0.001	(0.0001)

Table G5: Preference estimates (θ) allowing for nonlinearities

	θ_j	
Disadv Minority pupils (if $<$ 10%)	-0.004	(0.379)
Disadv Minority pupils (if $\geq$ 10%)	-0.401	(0.079)
Disadv. Dutch pupils	-0.150	(0.094)
Pupil-teacher ratio (PT)	-0.010	(0.004)
Teacher hours	0.251	(0.074)
Population density	-0.056	(0.019)
Public school	-0.140	(0.062)
Student achievement	0.964	(0.394)
Age teachers	0.010	(0.003)
Female teachers	0.256	(0.110)
Support staff	-0.512	(0.102)

for each of these new amenities. We illustrate this by allowing the effect of the job attribute "proportion of disadvantaged minority students" to vary depending on whether this proportion is less or more than 10%. Two-thirds of schools have less than 10% of disadvantaged minority pupils in our data. We run our estimation procedure with a job attribute vector now composed of 11 variables, and get the the preference parameter estimates in Table G5. There is a nonlinearity in the strong negative preference for proportion of disadvantaged minority students that we found in the paper, as we see that this job characteristic does not enter teachers' value function when it is lower than 10%. The preference estimates for the other amenities are not affected.