A Note on the Estimation of Job Amenities and Labor Productivity*

Arnaud Dupuy[†] Alfred Galichon[‡]

First circulated version: October, 2015. This version: July, 2021

Abstract

This paper introduces a maximum likelihood estimator of the value of job amenities and labor productivity in a single matching market based on the observation of equilibrium matches and wages. The estimation procedure simultaneously fits both the matching patterns and the wage curve. While our estimator is suited for a wide range of assignment problems, we provide an application to the estimation of the Value of a Statistical Life using compensating wage differentials for the risk of fatal injury on the job. Using US data for 2017, we estimate the Value of Statistical Life at \$6.3 million (\$2017).

Keywords: Matching, Observed transfers, Structural estimation, Value of Statistical Life.

JEL Classification: C35, C78, J31.

*We thank two anonymous referees and the editor Chris Taber for insightful comments.

[†]CREA, University of Luxembourg and IZA, arnaud.dupuy@uni.lu. Dupuy gratefully acknowledges the support of a FNR grant C14/SC/8337045.

[‡]Departments of Economics and of Mathematics, New York University, and Department of Economics, Sciences-Po, ag133@nyu.edu. Galichon gratefully acknowledges the support of NSF grant DMS-1716489 and European Research Council Grant no. 866274.

1 Introduction

Identification and estimation of both agents' value of a match in one-to-one matching models with transferable utility have been the subject of increasing interest in the last decade. Two important applications are in the fields of family economics with the marriage market (where the econometrician observes matching patterns, but not the transfers) and labor economics with the labor market, or more generally the literature on hedonic models (where the econometrician observes both the matching patterns and the transfers), although the former has thus far received most of the attention.

In the case when transfers are not observed, thus in the case of the marriage literature, Choo and Siow (2006) is a seminal reference which allowed to bring theoretical models to the data. Subsequent references such as Chiappori et al. (2015), Galichon and Salanié (2015) and Dupuy and Galichon (2014) have extended the structure of the model in various dimensions. In particular, Dupuy and Galichon (2014) have provided a framework for estimation of a matching model where agents match on continuous characteristics, which they have applied to marriage market data.

In the case when transfers are observed, however, transfers may potentially provide useful supplementary information about the partners' values of a match. In the analysis of the labor market, for example, wages may be observed. The literature referred to above is not very explicit on how this information may be used. Many authors, such as Ekeland et al. (2004), Heckman et al. (2010) and Galichon and Salanié (2015), among others, suggest techniques that implicitly or explicitly require to perform nonparametric estimation ("hedonic regression") of the wage curve prior to the analysis. While this works well in the case when the relevant characteristics is single-dimensional, as in Ekeland et al. (2004) and Heckman et al. (2010), or discrete, as in Galichon and Salanié (2015), this is more involved when the characteristics are continuous and multivariate. In this framework, Salanié (2015) shows that this structure implies quite strong testable restrictions.

In this paper, we build a flexible and tractable model of equilibrium matching and wages on the labor market, and show how to estimate the model using a maximum likelihood approach. This work therefore extends our previous work, Dupuy and Galichon (2014), to the case when transfers are observed.

We illustrate our method by revisiting the literature on compensating wage differentials (CDW) initiated as an application of Rosen's (1974) hedonic model to the labor market by Lucas (1977) and Thaler and Rosen (1976), and soon followed by many others; see Rosen (1986) for an elegant presentation of the theory and a review of the early empirical literature, and Viscusi and Aldy (2003) for a more recent review of the empirical literature. The approach in this vein consists in performing the reduced form estimation of the risk-wage gradient to uncover workers' marginal willingness to accept certain levels of fatal injury risk at their job and herewith derive an estimate of the Value of Statistical Life. A crucial assumption of this approach is that the data contains rich enough information about a worker's skills to control for wage differentials due to productivity differentials across workers. Departure from this assumption implies an inherent bias in estimates of the compensating wage differentials, and Hwang et al. (1992) have shown this bias can be large in magnitude. Attempts to avoid this bias have consisted in either using panel data to estimate workers' fixed effects and control for unobserved heterogeneity (see for instance Brown, 1980) or an instrumental variable approach (see for instance Garen, 1988). More recently, Kniesner et al. (2007) have argued in favor of adopting a structural hedonic model to identify the "underlying fundamentals (preferences), [...] that would further generalize estimates of Value of Statistical Life".

To the extent of our knowledge, however, the structural hedonic model approach has been largely ignored in the applied literature. Our method contributes to this discussion by proposing a structural estimation of preferences for risky jobs that explicitly takes into account the matching of workers to jobs while estimating the equilibrium hedonic wage equation. This method comes at no cost on the data since information about who matches with whom is, by definition, already available in the data needed to perform the hedonic wage regression in the first place. Accounting for the matching of workers to jobs results in the likelihood of observing the data given parameters being expressed as a weighted sum of two terms: the contribution of the first term is to equate the predicted moments of the matching distributions to their sample counterparts whereas the contribution of the second is to equate the predicted wages with their sample counterparts.

Following Viscusi (2003, 2007 and 2013), we use US data and merge the Census of Fatal Occupational Injuries (CFOI) by occupation and industry to the 2017 Current Population Survey (CPS). This allows us to have access to data on hourly wages, workers' characteristics and the rate of fatal injuries in their job. Our main results quantify the extent to which US workers dislike risky jobs, their utility dropping by 0.023 log-points per hour of work as the probability of fatal injury on the job increases by 13.05 per $100,000$. This amounts to a Value of Statistical Life (VSL hereafter) of \$6.3 million (\$2017). This estimate is about \$3 million lower, though not statistically so, than the estimate obtained when applying a hedonic (log)wage regression that does not account explicitly for the sorting of workers into jobs, i.e. \$9.7 million (\$2017), which itself lies in the range of previous estimates using similar data (e.g. Viscusi, 2013).

Our model is also related to a growing empirical literature applying the celebrated estimation technique proposed in Abowd et al. (1999) to decompose workers' wage differentials into differentials due to observed workers' characteristics, unobserved workers' heterogeneity and firms' heterogeneity using matched employer-employee panel data, see among others Abowd et al. (2002), Andrews et al. (2008;2012), Gruetter and Lalive (2009), Woodcock (2010) and Torres et al. (2013). Workers and firms fixed effects capture reduced form notions of workers and firms types that are fixed over time and are identified using the mobility of workers across firms over time.

The outline for the rest of the paper is as follows. Section 2 introduces the model and characterizes equilibrium. Section 3 presents our parametric specification of the model and a maximum likelihood estimator on data about matches and wages. Section 4 presents the empirical application and section 5 summarizes and concludes.

2 The model

The purpose of this section is to succinctly present our model, which is a bipartite continuous matching model with transferable utility and logit unobserved heterogeneity. In our context, equilibrium transfers (wages) are observed. This is relevant for instance in the labor market, as opposed to the marriage market where transfers are typically unobserved. We limit ourselves to the introduction of the notation needed for the construction of our estimator, emphasize on the additional identification and estimation results obtained when transfers are observed and refer the interested reader to the original paper, i.e. Dupuy and Galichon (2014), for more details about its other main properties. To fix ideas, we use in the remainder of the paper the example of the labor market where transfers (wages) are observed, in line with the application of Section 4.

Populations and matching. We shall assume that workers' characteristics are contained in a vector of attributes $x \in \mathcal{X} = \mathbb{R}^{d_x}$, while firms' characteristics are captured by a vector of attributes $y \in \mathcal{Y} = \mathbb{R}^{d_y}$.[1]

[1] In our one-to-one matching model, the terms "job" and "firm" are interchangeable. Our model would continue to work in a one-to-many context where firms offer multiple jobs as long as there is perfect substitutability between jobs (workers) within firms, i.e. as long as the surplus of a firm is the sum of the surplus at each job (of each worker) in the firm. Our model can therefore be seen as a

Our first main assumption is about the distribution of workers' and firms' types in the economy.

Assumption 1. *There is a continuum of workers, with a density of type distribution f on $\mathbb{R}^{d_x}$, and a continuum of firms, with a density of type distribution g on $\mathbb{R}^{d_y}$. There is the same total mass of workers and firms, and this mass is normalized to one, hence*

$$\int_{\mathbb{R}^{d_x}} f(x)\,dx = \int_{\mathbb{R}^{d_y}} g(y)\,dy = 1.$$

Since workers of type x have a density of probability $f(x)$, and firms of type y have a density of probability $g(y)$ and workers and firms are in equal number, a *feasible matching* between workers and firms will consist in the probability density $\pi(x, y)$ of occurrence of a (x, y) pair, which should have marginal densities f and g. More formally, we define the set of feasible matchings as

$$\mathcal{M}(f, g) = \left\{ \pi : \pi(x, y) \geq 0, \ \int_{\mathcal{Y}} \pi(x, y)\,dy = f(x) \text{ and } \int_{\mathcal{X}} \pi(x, y)\,dx = g(y) \right\}.$$

Demand and supply. Let $w(x, y)$ denote the wage of a worker of type x when working for a firm of type y. It is assumed that a worker of type x not only values her wage but also the amenities of her job. The value of amenities is further assumed to be decomposed into a systematic value $\alpha(x, y)$, which is the same for all workers of type x, and a random value $\varepsilon(y)$ that is specific to a particular worker, holds for all firms of a given type y and is known by the worker at the time the matching occurs. This specification therefore contrasts with the literature on search and matching that typically introduces a match-specific shock that is revealed after the matching occurred.

matching model of jobs to workers within firms under perfect substitutability. A recent application of this idea is found in a model of polygamy in the marriage market context by André and Dupraz (2017).

In particular, we assume the value for a worker of type x of working for a firm of type y at wage $w(x,y)$ is given by $\alpha(x,y)+w(x,y)+\sigma_1\varepsilon(y)$ where $\alpha(x,y)+w(x,y)$ is deterministic, $\varepsilon(y)$ is a worker-specific random process and σ_1 is a scaling factor. As in Dupuy and Galichon (2014), we choose to model the random process $\varepsilon(y)$ as a *Gumbel random process*, introduced by Cosslett (1988) and Dagsvik (1988), which is constructed as follows.[2] Assume that, in a first step, workers form their demand by drawing a random pool of observable types of firms, along with the corresponding utility shocks. We model the random pool by a Poisson point process, so that its cardinality does not have to be fixed and finite. More specifically, we assume that this Poisson process is valued in $\mathcal{Y}\times\mathbb{R}$ with intensity $dye^{-\varepsilon}d\varepsilon$. The random pool sampled by a worker is $\{(y_k,\varepsilon_k), k\in\mathbb{N}\}$, where y_k is the type and ε_k the corresponding utility shock. Define $\varepsilon(y)=\max_k\{\varepsilon_k : y_k=y\}$, with the convention that $\max\emptyset=-\infty$.

By construction, the problem of a utility-maximizing worker of type x reads as

$$\max_{y\in\mathcal{Y}}\{\alpha(x,y)+w(x,y)+\sigma_1\varepsilon(y)\}.$$

Note that workers' preferences only depend on their potential partner's type. Once the desired type has been determined, workers are indifferent between firms of that type.

By symmetry, we assume the value for a firm of type y of hiring a worker of type x at wage $w(x,y)$ is given by $\gamma(x,y)-w(x,y)+\sigma_2\eta(x)$ where $\gamma(x,y)-w(x,y)$ is deterministic, and $\eta(x)$ is a firm-specific Gumbel random process.

Assuming that $\varepsilon(y)$ and $\eta(x)$ follow Gumbel random processes allows us to get a continuous logit framework. Indeed, Proposition A.1 in the appendix, which was obtained by Cosslett (1988) and Dagsvik (1988), shows that the density of demand for firms of type y originating from workers of type x, is proportional to $\exp(\alpha(x,y)+$

[2] See appendix A for more details about Gumbel random processes.

$w(x,y)$). This leads us assuming continuous logit demands from workers and firms, which we formalize in the following assumption.

Assumption 2 (Continuous logit demands)**.** *All agents are price-takers and utility maximizers and given the wage schedule $w(x,y)$, the conditional density demand for firms of type y by workers of type x is*

$$\pi(y|x) = \frac{\exp\left(\frac{\alpha(x,y)+w(x,y)}{\sigma_1}\right)}{\int_{\mathcal{Y}} \exp\left(\frac{\alpha(x,y')+w(x,y')}{\sigma_1}\right) dy'}. \tag{2.1}$$

Symmetrically, the conditional density demand for workers of type x by firms of type y is

$$\pi(x|y) = \frac{\exp\left(\frac{\gamma(x,y)-w(x,y)}{\sigma_2}\right)}{\int_{\mathcal{X}} \exp\left(\frac{\gamma(x',y)-w(x',y)}{\sigma_2}\right) dx'}. \tag{2.2}$$

Note that the conditional density demands rewrite as

$$\pi(y|x) = \exp\left(\frac{\alpha(x,y)+w(x,y)-u(x)}{\sigma_1}\right) \tag{2.3}$$

where $u(x)$ given by

$$u(x) = \sigma_1 \log \int_{\mathcal{Y}} \exp\left(\frac{\alpha(x,y')+w(x,y')}{\sigma_1}\right) dy' \tag{2.4}$$

interprets in the Gumbel framework as the expected indirect utility of a worker of type x and symmetrically,

$$\pi(x|y) = \exp\left(\frac{\gamma(x,y)-w(x,y)-v(y)}{\sigma_2}\right) \tag{2.5}$$

where $v(y)$ given by

$$v(y) = \sigma_2 \log \int_{\mathcal{X}} \exp\left(\frac{\gamma(x',y)-w(x',y)}{\sigma_2}\right) dx' \tag{2.6}$$

interprets as the expected indirect profits of a firm of type y.

In a second step, agents determine equilibrium by tatonnement over $w(x, y)$ using the demand functions defined in step 1. Note that since workers (resp. firms) are indifferent between firms (workers) of the same type, once the desired type has been determined, workers (resp. firms) can match with any of the firms (workers) of that type.

Note also that agents need not to observe other agents' idiosyncratic shocks ε and η to form their demand. Even if they had access to that information, they would not use it.

Equilibrium. At equilibrium, the wage curve $w(x, y)$ is such that the density $\pi(x, y)$ of pairs (x, y) emanating from the workers' problem coincides with the density of pairs (x, y) emanating from the firms' problem, and hence must satisfy

$$\exp\left(\frac{\alpha(x, y) + w(x, y) - a(x)}{\sigma_1}\right) = \pi(x, y) = \exp\left(\frac{\gamma(x, y) - w(x, y) - b(y)}{\sigma_2}\right), \tag{2.7}$$

where

$$\begin{cases} a(x) = u(x) - \sigma_1 \log f(x) \\ b(y) = v(y) - \sigma_2 \log g(y) \end{cases}. \tag{2.8}$$

Substituting out $w(x, y)$ in system (2.7) yields

$$\pi(x, y) = \exp\left(\frac{\phi(x, y) - a(x) - b(y)}{\sigma}\right), \tag{2.9}$$

where $\sigma := \sigma_1 + \sigma_2$, while substituting out $\pi(x, y)$ yields

$$w(x, y) = \frac{\sigma_1}{\sigma}\left(\gamma(x, y) - b(y)\right) + \frac{\sigma_2}{\sigma}\left(a(x) - \alpha(x, y)\right). \tag{2.10}$$

Note that σ is the amount of heterogeneity in the model and when the scaling factors of the random values σ_1 and σ_2 tend to zero, i.e. there is no heterogeneity in the model $\sigma \to 0$, the firm's problem and the worker's problem converge to the deterministic maximization problems

$$u(x) = \max_{y \in \mathcal{Y}} \{\alpha(x,y) + w(x,y)\} \text{ and } v(y) = \max_{x \in \mathcal{X}} \{\gamma(x,y) - w(x,y)\}$$

and the equilibrium problem consists in finding $w(x,y)$ and $\pi(x,y)$ which are compatible with optimality in these problems. See section 5.2 below.

Note also that in our model, equilibrium wages do not vary systematically with the idiosyncratic shocks of workers and firms. This is because the non-wage valuation of a job at a firm of type y by a worker of type x is given as $\alpha(x,y) + \sigma_1 \varepsilon(y)$ and only depends on the observable characteristics of the firm, not the firm's idiosyncratic shock (i.e. the process η). Symmetrically, the productivity of a job at a firm of type y when performed by a worker of type x is given as $\gamma(x,y) + \sigma_2 \eta(x)$, only depends on the observable characteristics of the worker, not her idiosyncratic shock (i.e. the process ε). Hence, by the law of one price, this implies that the wage of any worker of type x must be the same at every firm of the same type y.

We can now formally define an equilibrium outcome on this market.

Definition 1 (Equilibrium outcome)**.** *An equilibrium outcome* (π, w) *consists of an equilibrium matching* $\pi(x,y)$*, and an equilibrium wage* $w(x,y)$ *where there exist functions* $a(x)$ *and* $b(y)$ *such that:*

(i) matching π *is feasible: defined by (2.9) and* $\pi \in \mathcal{M}(f,g)$*, and*

(ii) wage w *is defined by (2.10).*

As a result, the equilibrium outcome problem consists of looking for functions $a(x)$ and $b(y)$ that are solution to the system

$$\begin{cases} \int_{\mathcal{Y}} \exp\left(\frac{\phi(x,y) - a(x) - b(y)}{\sigma}\right) dy = f(x) \\ \int_{\mathcal{X}} \exp\left(\frac{\phi(x,y) - a(x) - b(y)}{\sigma}\right) dx = g(y). \end{cases} \tag{2.11}$$

Two important remarks are in order.

Remark 2.1 (Location normalization)**.** If $a(x)$ and $b(y)$ are solutions of system (2.11), so are $a(x)+t$ and $b(y)-t$. Using equation (2.10), the equilibrium wages are $w\left(x,y\right)$ for the former solution and $w\left(x,y\right)+t$ for the latter. The nonuniqueness of the solution for system (2.11) requires a normalization which is reflected by the arbitrary choice $a\left(x_0\right)=0$ and a constant term t in the equilibrium wages equation (2.10). Uniqueness of such (a,b) upon normalization $a\left(x_0\right)=0$ is proved in Rüschendorf and Thomsen (1993), theorem 3.

Remark 2.2 (Continuous Mixed Logit demand)**.** It follows from formula (2.1) that the density of market demand for firms of type y is given by

$$\int_{\mathcal{X}} \frac{\exp\left(\frac{\alpha(x,y)+w(x,y)}{\sigma_1}\right)}{\int_{\mathcal{Y}} \exp\left(\frac{\alpha(x,y')+w(x,y')}{\sigma_1}\right) dy'} dx$$

which is a continuous Mixed Logit model. Likewise, the density of market demand for workers of type x has a similar expression. The equilibrium wage $w\left(x,y\right)$ equates these quantities to the respective densities of supply, $g(y)$ and $f(x)$ respectively.

3 Parametric estimation

3.1 Observations

Assume that one has access to a random sample of the population of matches of firms and workers. For each match, this sample contains information about the worker's characteristics, her wage and the firm's characteristics. The observations consist of $\left\{\left(X_i,Y_i,W_i\right),i=1,...,n\right\}$, where n is the number of observed matches, i indexes an employer-employee match, X_i and Y_i are respectively the vectors of employee's and employer's observable characteristics, which are assumed to be sampled from a continuous distribution, and W_i is a noisy measure of the true unobserved transfer

$w(X_i, Y_i)$ assumed to be such that

$$W_i = w(X_i, Y_i) + \epsilon_i \tag{3.1}$$

where measurement error ϵ_i follows a $\mathcal{N}(0, s^2)$ distribution and is independent of (X_i, Y_i).

Note that depending on the nature of preferences, observed transfers W_i can be a monotonic transformation of observed wages. This flexibility allows one to consider equation (3.1) as a hedonic wage regression using any known monotonic transformation of wages, i.e. identity (to estimate the model in levels), logarithm, power transformation etc.

Finally, note that, while we assume the analyst has access to data containing all variables in X and Y, in practice datasets only contain a subset of these variables. In such a situation, the analyst faces issues of unobserved heterogeneity that our current method does not account for. However, while this is a current limitation of our approach, one can expect that existing methods dealing with unobserved heterogeneity would adapt to be incorporated into the framework. This is left for future research.

3.2 Identification

In this section, we briefly discuss identification of the deterministic value of amenities α and productivity γ. Not that α and γ do not appear individually in the expression of the equilibrium matches in equation (2.9), only the joint value of a match ϕ appears in this equation. However, α and γ do appear separately and with opposite signs in the formula of the equilibrium wages in equation (2.10).

This clearly indicates that when only matches are observed, one cannot identify and hence estimate the deterministic value of amenities α separately from the deterministic value of productivity γ. In contrast, if transfers are observed, one actually can identify and estimate these objects separately.

It should be noted, however, that taking the values of σ_1 and σ_2 as known[3] and $\sigma_1 + \sigma_2 = 1$ for notational simplicity, equations (2.1) and (2.2) clearly indicate that $\alpha(x, y) + w(x, y)$ is identified up to a function $c(x)$ by $\sigma_1 \ln \pi(x, y) + c(x)$, and $\gamma(x, y) - w(x, y)$ is identified up to a function $d(y)$ by $\sigma_2 \ln \pi(x, y) + d(y)$. It follows that α is identified up to fixed effects $c(x)$ by

$$\alpha(x, y) = \sigma_1 \ln \pi(x, y) - w(x, y) + c(x),$$

while γ is identified up to fixed effects $d(y)$ by

$$\gamma(x, y) = \sigma_2 \ln \pi(x, y) + w(x, y) + d(y).$$

This result has been used in a nonparametric setting by Galichon and Salanié (2015) and Salanié (2015). In this paper, we exploit it in a parametric setting using basis functions of x and y (see Section 3.4). Indeed, since α is identified up to fixed effects $c(x)$, the parametrization of α can only include basis functions depending on both x and y or on y only but it cannot include basis functions depending on x only. By a similar reasoning, the parametrization of γ cannot include basis functions depending on y only.

3.3 Notation

For the sake of readability and to avoid additional notational burden, we propose the following change of notation. Replace α by $\sigma\alpha$, γ by $\sigma\gamma$, ϕ by $\sigma\phi$, a by σa, and b by σb, so that the equations of the model become

$$\pi(x, y) = \exp(\phi(x, y) - a(x) - b(y)), \tag{3.2}$$

[3]The σ parameters are not non-parametrically identified but they can be estimated using observed transfers once $\alpha(.,.)$ and $\gamma(.,.)$ have been parametrically specified.

where (a, b) is the unique solution to the system of equations

$$\begin{cases} \int_{\mathcal{Y}} \exp\left(\phi\left(x,y\right) - a\left(x\right) - b\left(y\right)\right) dy = f\left(x\right) \\ \int_{\mathcal{X}} \exp\left(\phi\left(x,y\right) - a\left(x\right) - b\left(y\right)\right) dx = g\left(y\right), \end{cases} \tag{3.3}$$

still normalized by $a\left(x_0\right) = 0$, and the terms a and b are related to u and v by

$$u\left(x\right) = \sigma a\left(x\right) + \sigma_1 \log f\left(x\right) + t, \text{ and } v\left(y\right) = \sigma b\left(y\right) + \sigma_2 \log g\left(y\right) - t \tag{3.4}$$

and the equilibrium transfer w is given by

$$w\left(x,y\right) = \sigma_1\left(\gamma\left(x,y\right) - b\left(y\right)\right) + \sigma_2\left(a\left(x\right) - \alpha\left(x,y\right)\right) + t. \tag{3.5}$$

This change of notation is without loss of generality since from equation (3.5) one can estimate parameters σ_1 and σ_2 and hence σ and therefore recover the initial values of α and γ. In the remainder of the paper, equations (3.2)–(3.5) will characterize the model to estimate.[4]

3.4 Parametrization

Let A and Γ be two vectors of $\mathbb{R}^K$ parameterizing the function of workers' systematic value of job amenities α and the function of firms' systematic value of productivity γ, in a linear way, so that

$$\alpha(x, y; A) = \sum_{k=1}^{K} A_k \varphi_k(x, y), \text{ and } \gamma(x, y; \Gamma) = \sum_{k=1}^{K} \Gamma_k \varphi_k(x, y),$$

where the basis functions φ_k are linearly independent, and may include functions that depend on x (respectively y) only. Note that by definition, the function of the joint

[4]The value of σ is therefore not imposed but estimated. Note however that the non-negativity of σ_1 and σ_2 should be imposed as a constraint, as in the application below.

value of a match reads as

$$\phi(x, y; \Phi) = \sum_{k=1}^{K} \Phi_k \varphi_k(x, y), \tag{3.6}$$

where $\Phi_k = A_k + \Gamma_k$. Inspection of equation (3.5) reveals that, given the parametric choice above, equilibrium matching and transfers are parameterized by $(A, \Gamma, \sigma_1, \sigma_2, t)$. The model is hence fully parameterized by $\theta = (A, \Gamma, \sigma_1, \sigma_2, t, s^2)$, which we make explicit by writing the predicted equilibrium transfer as $w(x, y; \theta)$.

3.5 Estimation

The main purpose of this exercise is to estimate the vector of parameters θ. To this aim we adopt a maximum likelihood approach. It follows from section (3.2) that the likelihood of observing a pair (x, y) only depends on $\Phi = A + \Gamma$, and is given by

$$\pi(x, y; \Phi) = \exp\left(\phi\left(x, y; \Phi\right) - a(x; \Phi) - b(y; \Phi)\right),$$

where $a(x; \Phi)$ and $b(y; \Phi)$ are uniquely determined by system of equations (3.3). Since, by assumption, measurement errors in transfers are independent of (X, Y), the log-likelihood of an observation (x, y, w) at parameter θ is therefore

$$\log L\left(x, y, w; \theta\right) = \log \pi\left(x, y; \Phi\right) - \frac{\left(w - w\left(x, y; \theta\right)\right)^2}{2s^2} - \frac{1}{2} \log s^2,$$

and hence, the log-likelihood of the sample reads as:

$$\log L\left(\theta\right) = n\mathbb{E}_{\hat{\pi}}\left[\phi\left(X, Y; \Phi\right) - a\left(X; \Phi\right) - b\left(Y; \Phi\right) - \frac{\left(W - w\left(X, Y; \theta\right)\right)^2}{2s^2}\right] - \frac{n}{2} \log s^2 \tag{3.7}$$

where $\hat{\pi}\left(x, y\right)$ is the observed density of matches in the data.

However, note that a, b and w that appear in (3.7) are computed in the population; here, we only have access to a sample. So, denoting a_i and b_j the sample analog of

$a(x)$ and $b(x)$, we compute the sample analog of system (3.3)

$$\begin{cases} \sum_{j=1}^{n} \exp\left(\phi_{ij}(\Phi) - a_i - b_j\right) = 1/n, \forall i = 1, ..., n \\ \sum_{i=1}^{n} \exp\left(\phi_{ij}(\Phi) - a_i - b_j\right) = 1/n, \forall j = 1, ..., n \end{cases} \tag{3.8}$$

with the added normalization $a_1 = 0$, which ensures uniqueness of the solution.[5] (Note that since we have assumed that the population distribution is continuous, each sampled observation occurs uniquely, hence the right-hand side here is $1/n$; however, this could easily be extended to a more general setting). We denote $(a_i(\Phi), b_i(\Phi))$ this solution at Φ. This allows us to compute a sample estimate of the equilibrium transfer $w_i(\theta)$ as

$$w_i(\theta) := \sigma_1 (\gamma_{ii}(\Gamma) - b_i(\Phi)) + \sigma_2 (a_i(\Phi) - \alpha_{ii}(A)) + t, \tag{3.9}$$

where the notation $\alpha_{ij}(A)$ substitutes for $\alpha(X_i, Y_j; A)$, and similarly for $\gamma_{ij}(\Gamma)$.

We are thus able to give the expression of the log-likelihood of the sample in our next result. Recall that $\theta = (A, \Gamma, \sigma_1, \sigma_2, t, s^2)$ and $\Phi = A + \Gamma$.

Theorem 1. *The log-likelihood of the sample is given by*

$$\log \hat{L}(\theta) = \log \hat{L}_1(\theta) + \log \hat{L}_2(\theta), \tag{3.10}$$

where

$$\log \hat{L}_1(\theta) = \sum_{i=1}^{n} (\phi_{ii}(\Phi) - a_i(\Phi) - b_i(\Phi)) \tag{3.11}$$

and,

$$\log \hat{L}_2(\theta) = -\sum_{i=1}^{n} \frac{(W_i - w_i(\theta))^2}{2s^2} - \frac{n}{2} \log s^2, \tag{3.12}$$

where $\phi_{ij}(\Phi) := \phi(X_i, Y_j; \Phi)$ is as in (3.6), $a_i(\Phi)$ and $b_i(\Phi)$ are obtained as the solution of (3.8), and where $w_i(\theta)$ is given by (3.9).

[5] An argument similar to theorem A.2 in Chernozhukov et al. (2017) would show that the solution to the system (3.8) converges uniformly to a and b as computed in the population, i.e. solution of system (3.3).

Proof. Immediate given the discussion before the theorem. ∎

Theorem 1 motivates the following remark.[6]

Remark 3.1 (Interpretation of the objective function)**.** Expression (3.10) has a straightforward interpretation. The term $\log \hat{L}_1(\theta)$, whose expression is given in equation (3.11) comes from the observed matching patterns. It only depends on θ through $\Phi = A + \Gamma$, and one has

$$\frac{1}{n}\frac{\partial \log \hat{L}_1}{\partial \Phi_k} = \mathbb{E}_{\hat{\pi}}\left[\varphi_k(X,Y)\right] - \mathbb{E}_{\pi^{\Phi}}\left[\varphi_k(X,Y)\right]$$

where $\mathbb{E}_{\hat{\pi}}$ is the sample average and $\mathbb{E}_{\pi^{\Phi}}$ the expectation with respect to

$$\pi_{ij}^{\Phi} := \exp\left(\phi_{ij}(\Phi) - a_i(\Phi) - b_j(\Phi)\right).$$

Hence, the contribution of the first term is to equate the predicted moments of the matching distributions to their sample counterparts. The term $\log \hat{L}_2(\theta)$, whose expression appears in equation (3.12) tends to match the predicted transfers $w_i(\theta)$ with the observed transfers W_i in order to minimize the sum of the square deviations $(W_i - w_i(\theta))^2$. Hence, the contribution of the second term is to equate the predicted transfers with their sample counterparts. Of course, s^2 will determine the relative weighting of those two terms in the joint optimization problem. If s^2 is high, which means transfers are observed with a large amount of noise, then the first term becomes predominant in the maximization problem. In the limit $s^2 \to +\infty$, the problem will boil down to a two-stage problem, where the parameter Φ is estimated in the first stage, and the rest of the parameters are estimated in the second stage by Non-Linear Least Squares conditional on $A + \Gamma = \Phi$. In the MLE procedure, s^2 is a parameter,

[6]Note that in most applications, the parameters of primary interest are those governing workers' deterministic values of amenities and firms' deterministic values of productivity, i.e. A and Γ respectively. The remaining parameters $\left(\sigma_1, \sigma_2, t, s^2\right)$ are auxiliary. Our MLE estimator can be concentrated on the parameters of interest. These results are available in an online Appendix.

and its value is determined by the optimization procedure.

4 Application

4.1 Data

We illustrate the usefulness of our method using an application to the estimation of the value of job amenities related to risks of fatal injury. This application requires access to a single cross-section of data containing a representative sample of worker-job matches with information about workers' characteristics (education, experience, gender etc.), their (hourly) wage and a measure of fatality rates associated to their job. Many surveys such as the CPS contain all required information but the fatal injury data. As a result, following Thaler and Rosen (1978), a large strand of the literature has compiled the required data by combining survey data with data about fatality per type of jobs from alternative sources.

In this paper, we follow the recent work by Viscusi (2003, 2007 and 2013) and construct measures of fatality rates by occupation-industry cells for the period 2012-2016. Unfortunately, data on fatal injuries by occupation within industries are not readily available. Instead, we rely on fatal injury data by occupation (4-digits SOC) and by industry (4-digits NAICS) provided by the U.S. Bureau of Labor Statistics (BLS) CFOI.[7] For each year in the period 2012-2016, we create a matrix of fatal injuries by occupation$\times$industry by simply multiplying the marginal distribution by occupation and industry hence assuming independence. To reduce measurement errors, the 4-digits occupational codes are aggregated into 25 major occupations and the

[7]In the CFOI data, a fatal injury is an injury leading to death within one year of the day of the accident. See Viscusi (2003) for more details about the CFOI data and its use in the present context.

4-digits industry codes into 80 major industries.[8] We then use the CPS March files[9] for 2012-2016 and compute matrices of hours-adjusted employment level by occupation and industry, combining person-weights, computed to this effect by the census and the BLS, and hours worked per week. The two sets of matrices are then merged allowing us to compute, for each year, fatality rates for a given occupation-industry cell as the ratio of the number of fatalities to total hours-weighted employment in that cell (see e.g. Viscusi, 2013). To attenuate further measurement errors, for each occupation-industry cell, we take the average fatality rate over time as our measure of risk.

We obtained our working dataset by merging the 2017 March CPS data with our measure of fatality rate by occupation-industry cells. This dataset therefore contains information about our main variables of interest: hourly earnings, hours of work, gender, years of schooling, age, ethnic group, marital status, whether one's job is in the public sector or not, and occupation-industry fatality rates. We follow the literature (e.g. Viscusi, 2013) and keep only full-time, non-agricultural, non-armed force workers[10] between 16 and 64 years old for the remainder of the analysis.[11]

Table (1) provides descriptive statistics of our working dataset. The average fatality rate in our sample is about 3.44 per $100,000$ which is close to the figure obtained in Viscusi (2013) for the year 2008, i.e. 3.29. To further compare our dataset with the literature, we run a hedonic (log)wage regression including the traditional controls (gender, years of schooling, age, age squared, ethnic group, marital status, union

[8]We use crosswalks provided by the BLS to perform the aggregation to major occupations and industry.

[9]The BLS advises to use March files of the CPS for computations of total employment.

[10]Assumption 2 requires all agents to be price-takers. This assumption is likely not to be met in the armed force industry whose sole employer is the US government. For this reason, we exclude armed force workers from the analysis. Note, however, that this exclusion is common in the hedonic wage regression literature, e.g. Viscusi (2013).

[11]As is standard when using March CPS wage data, see Katz and Murphy (1992) for instance, the sample excludes individuals with hourly earnings below one half of minimum wage and top coded earnings are imputed 1.45 times the top code value.

membership, public sector dummy, regional dummies, Metropolitan dummy) and our measure of hours-weighted fatality rates by occupation-industry. Using the estimate of the compensating wage differential for risk, we obtain an estimate of the VSL of \$9.7 million (\$2017). This figure falls in the range of estimates in the literature using similar data, i.e. Viscusi (2013)'s estimate of \$8.4 million (\$2017) using the 2008 CPS data.[12]

4.2 Estimates

We estimate the model using the maximum likelihood estimator presented in this paper. Observed transfers are assumed to be the logarithm of observed wages to be consistent with the hedonic regression literature that typically uses log wage regressions. We standardize other continuous variables to facilitate the comparison and interpretation (in terms of standard deviation) of the respective coefficients.

Estimation requires to specify the basis functions used to parameterize the values of job amenities $\alpha(x, y; A)$ and productivity $\gamma(x, y; \Gamma)$. We adopt a linear (in parameters) specification of the basis functions and present estimates for the following specification:[13]

$$\alpha(x, y; A) = \sum_{l=1}^{2} A_{0,l} x^{(0)} y^{(l)} + A_{1,2} x^{(1)} y^{(2)},$$

[12] The risk coefficient in the log wage hedonic regression reported in Viscusi (2013), i.e. 0.0024, is very close to the estimate obtained with our data, i.e. 0.0027. Using 0.0024 instead of 0.0027 to calculate the VSL with our data one would obtain \$8.6 million (\$2017).

[13] The fit of this specification can be compared with that of alternative (nesting/nested) specifications using likelihood ratio tests. For instance, the test statistic obtained when comparing the chosen specification with a richer specification, where both job amenities and productivity include interactions between workers' years of schooling, experience and gender with jobs' risk and sector, is equal to 1.358. This statistic is not significantly different from 0 at conventional levels. One concludes that the specification presented in the paper should be preferred. Note however, that the estimates of the VSL are similar across specifications.

and

$$\gamma(x, y; \Gamma) = \sum_{k=1}^{8} \Gamma_{k,0} x^{(k)} y^{(0)} + \sum_{k=1}^{4} \sum_{l=1}^{2} \Gamma_{k,l} x^{(k)} y^{(l)}$$

where x includes a constant (k=0), years of schooling (k=1), (potential[14]) experience (k=2), experience squared (k=8), a dummy variable for female (k=3), a dummy variable indicating whether one is married or not (k=4), and 3 ethnic dummy variables (white, black and asian, using others, incl. hispanic, as the reference group, k=5,6,7), whereas y includes a constant (l=0), our measure of fatality rates (l=1) and a dummy variable indicating the public sector (l=2).

Hence, our specification of job amenities includes the main effects of fatality rates and public sector as well as an interaction between a workers' years of schooling and jobs' sector. Our specification of perceived productivity includes the main effects of years of schooling, experience (squared), marital status and ethnic groups as well as interactions between workers' year of schooling, experience and gender with jobs' fatality rates and sector.

Estimates are presented in table (2). Note first that the model fits quite well the wage data with an R^2 of 0.235 which compares to that obtained for the standard hedonic wage regression, i.e. $R^2 = 0.255$.

Second, estimates of the value of perceived productivity show expected results.[15] The value of productivity increases with years of schooling (0.057), although the estimate is not significant,[16] and the experience-productivity gradient is positif (0.084) but decreasing, as indicated by the negative coefficient for experience squared (-0.051). These human capital effects, however, vary significantly across jobs: the years-of-schooling-productivity gradient is absent in risky jobs ($-0.002 = 0.057 - 0.059$)

[14] Age less years of schooling less 6.

[15] Unless stated otherwise, the significance level is 1%.

[16] We have also estimated the model including years of schooling squared. However, comparing the two specifications, the log-likelihood ratio test statistic is 3.360 and not significantly different from 0 at conventional levels. Our chosen specification should be preferred.

but greater for public sector jobs ($0.895 = 0.057 + 0.838$), whereas the experience-productivity gradient is higher in risky jobs ($0.158 = 0.084 + 0.074$).

Third, our estimates for perceived productivity show negative coefficients for female and black workers (-0.404 and -0.108 respectively) and a positive coefficient for white workers (0.046). These coefficients should be interpreted with care as they indeed reflect employers' perceived productivity of the underlying types of workers, revealing discrimination effects.[17] Our results are in line with the large literature showing discriminating wage differentials across gender and race. Interestingly, the gender perceived productivity gap varies significantly across jobs unlike the racial one: a one standard deviation increase in the probability of fatal injury more than triples the gender perceived productivity gap.

Fourth, regarding the value of job amenities, results show that the value of public sector jobs increases significantly with years of schooling: a one standard deviation in years of schooling generates a 0.081 log-points increase in the value of jobs in the public sector.

Finally, our main result shows that US workers' utility drops by 0.023 log-points per hour of work as the probability of fatal injury on the job increases by one standard deviation (i.e. 13.05 per $100,000$). We can use this coefficient to compute the VSL from the formula

$$VSL(x,y) = -\frac{\partial \alpha(x,y)}{\partial y^{(1)}}\overline{z},$$

where $\overline{z}$ are the average earnings in the sample.[18]

[17] We refer herewith to Becker (1971). The parameter γ reflecting both the true productivity of workers and employers' taste discrimination parameter.

[18] To derive this formula, remember that the systematic utility of a worker of type x working in job of type y and receiving a transfer $w(x,y)$ is given as $U(x,y) = \alpha(x,y) + w(x,y)$. Since transfers are specified in log wages, i.e. $w(x,y) = \ln z(x,y)$ where $z(x,y)$ are the equilibrium wages, we can then compute a worker of type x's trade-off between earnings and risk as

$$\frac{\partial U(x,y)}{\partial y^{(1)}} := \frac{\partial \alpha(x,y)}{\partial y^{(1)}} + \frac{\frac{\partial z(x,y)}{\partial y^{(1)}}}{z(x,y)} = 0.$$

Using $\frac{\partial\alpha(x,y)}{\partial y^{(1)}} = A_{1,1} = 0.023$ and the appropriate units, one obtains a VSL of \$6.3 million (\$2017).[19] This value lies in the range of estimates found in the literature using similar data, i.e. Viscusi, (2013). Nevertheless, it is about \$3 million lower than the estimate obtained using the classical hedonic wage regression. Though the difference is not statistically significant, it suggests that not accounting for the sorting of workers into jobs, as in the hedonic regression, may lead to an overestimation of the true VSL in our data.

To see this, note that, in our model, the differential value of job amenities with respect to fatality risk is identified as

$$\frac{\partial\alpha\left(x,y\right)}{\partial y^{(1)}} = \sigma_1 \frac{\partial \ln \pi\left(y|x\right)}{\partial y^{(1)}} - \frac{\partial w\left(x,y\right)}{\partial y^{(1)}}.$$

In contrast, the hedonic wage regression literature identifies this differential value using the coefficients of an (log) earnings regression as

$$\frac{\partial\alpha^h\left(x,y\right)}{\partial y^{(1)}} = -\frac{\partial w\left(x,y\right)}{\partial y^{(1)}} = \frac{\partial\alpha\left(x,y\right)}{\partial y^{(1)}} - \sigma_1 \frac{\partial \ln \pi\left(y|x\right)}{\partial y^{(1)}}.$$

As a result, the VSL as measured in the hedonic regression literature reads as

$$VSL^h\left(x,y\right) := -\frac{\partial\alpha^h\left(x,y\right)}{\partial y^{(1)}}\overline{z} = VSL\left(x,y\right) - \sigma_1 \frac{\partial \ln \pi\left(y|x\right)}{\partial y^{(1)}}\overline{z},$$

once substituting $\frac{\partial\alpha^h(x,y)}{\partial y^{(1)}}$ by its expression in terms of α and π.

Since average wages are positive, $\overline{z} > 0$, and $\sigma_1 > 0$, it follows that, compared to our method, estimates of VSL from hedonic wage regressions tend to be larger (lower) when, in equilibrium, conditional on workers' type, workers sort into safe (resp. risky)

Rearranging this equation, one can express the differential wage increase required to compensate the differential drop in job amenity due to an one unit increase in the risk of fatal injury as in the text using $\overline{z}$ to replace $z\left(x,y\right)$.

[19] In our preferred specification, the VSL does not vary with the type of workers nor with the fatality risk. Note that this method excludes productivity effects of fatality risk. It only reflects the valuation of life from the perspective of workers.

jobs, i.e. when $\frac{\partial \ln \pi(y|x)}{\partial y^{(1)}} < 0$ (resp. > 0).

Not only our structural approach allows one to explicitly take into account the sorting of workers to jobs when estimating the value of job amenities, it also allows to compute counterfactual equilibria. In particular, one could use the estimates of job amenities and productivity obtained above to compute the impact of a government intervention aiming at reducing fatality risk at work. For instance, consider the Site-Specific Targeting (SST) inspection plan proposed by the Occupational Safety and Health Administration (OSHA) in the US. If effective, this program would decrease fatality rates in the most risky jobs. For the sake of an example, suppose that the program ends up decreasing the fatality risk of all jobs whose fatality risk is at least 1 standard deviation above the mean (i.e. $\geq 16.5 = 3.4 + 1 \times 13.1$) down to 16.5 fatalities per 100,000 workers per year. As a result of this intervention, the distribution of types of jobs would change causing the equilibrium matching and wages to change too. We use our model to compute the equilibrium before (observed) and after the intervention and then compare the matching and distribution of wages. We find that, as a result of this intervention, about 3.1% of the workers would change jobs, the mean wage would drop by 3.9% and wage inequality, as measured by the Gini coefficient, would drop by 3.6%.

5 Discussion and conclusion

We conclude by addressing a few methodological remarks before summarizing our main results.

5.1 Job seekers and vacancies

A natural extension of the model is to allow for workers to be unemployed and jobs to be vacant. To do so, one needs first to allow for the total masses of workers and firms to be different, herewith relaxing that part of Assumption 1. Second, one needs to extend the definition of utilities for matched workers and firms to unemployed workers and vacant jobs, introducing reservation utilities. The reservation utility of a worker (firm) of type x (y) may be decomposed into a systematic part $\alpha(x, \emptyset)$ ($\gamma(\emptyset, y)$) and a random value $\varepsilon(\emptyset)$ ($\eta(\emptyset)$) following a Gumbel type I distribution. Assumption 2 should then simply be modified by replacing the choice sets of workers $\mathcal{Y}$ and firms $\mathcal{X}$ by $\mathcal{Y} \cup \{\emptyset\}$ and $\mathcal{X} \cup \{\emptyset\}$ respectively and adopting the convention that $w(x, \emptyset) = w(\emptyset, y) = 0$ for all types of workers and firms.

However, note that the conditional probabilities in (2.1) and (2.2) have a logit structure and hence satisfy the Independence of Irrelevant Alternative property. As a consequence, $\pi(y|x)$ in (2.1) is also the density of probability of choosing a firm of type y for a worker of type x conditional on participation. As shown in appendix D in Dupuy and Galichon (2014), this implies that in a market with unemployed workers and vacant jobs and reservation utilities $\alpha(x, \emptyset) + \varepsilon(\emptyset)$ and $\gamma(\emptyset, y) + \eta(\emptyset)$, at equilibrium, the probability density $\pi(x, y)$ of occurrence of a (x, y) pair among matched pairs (i.e. considering only active workers and filled jobs) is the same as the probability density $\pi(x, y)$ of occurrence of a (x, y) pair in a market with no outside options and where the masses of workers and firms are the same as the masses of active workers and filled jobs in the former market.

5.2 Related assignment models

When $\sigma \to 0$, the model converges to the classical model of Monge-Kantorovich, which is a continuous extension of the Becker-Shapley-Shubik model. Indeed, when

σ_1 and σ_2 tend to zero, the scaling coefficients of the random value of job amenities and productivity ε and η, tend to zero, then the model becomes nonstochastic. Intuitively, when $\sigma_1 \to 0$, the worker's expected indirect utility $u(x)$ tends to the deterministic indirect utility $\max_y \{\alpha(x,y) + w(x,y)\}$, and it follows from (2.1) that the conditional choice distribution $\pi(y|x)$ becomes concentrated around the optimal firm's type y such that $u(x) = \alpha(x,y) + w(x,y)$. Similarly, when $\sigma_2 \to 0$, a firm of type y expected indirect profits $v(y)$ tends to the deterministic indirect profits $\max_x \{\gamma(x,y) - w(x,y)\}$, and $\pi(x|y)$ becomes concentrated around the optimal worker's type x such that $v(y) = \gamma(x,y) - w(x,y)$. Combining these two results, $\pi(x,y)$ becomes concentrated around the set of pair (x,y) such that $u(x) + v(y) = \phi(x,y)$, hence, in the limit when σ_1 and σ_2 tend to zero, we have

$$\begin{cases} \pi & \in & \mathcal{M}(f,g) \\ u(x) + v(y) & \geq & \phi(x,y) & \forall x \in \mathcal{X}, y \in \mathcal{Y} \\ u(x) + v(y) & = & \phi(x,y) & \pi - a.s. \end{cases}$$

These are the classical stability conditions in the Monge-Kantorovich problem (see Villani, 2003 and 2009), whose variants have been applied in economics by Becker (1973), Shapley and Shubik (1963), Gretsky, Ostroy, Zame (1992). A particular example is Sattinger's workhorse model extensively used in the labor economics literature (see Sattinger, 1979 and 1993). This model indeed corresponds to a matching market with no unobserved heterogeneity ($\sigma \to 0$), unidimensional observed types ($d_x = d_y = 1$), in which workers only care about their compensation ($\alpha = 0$) and where the firm's value of productivity is smooth and supermodular (i.e. $\partial^2 \gamma(x,y) / \partial x \partial y$ exists and is positive). Under these restrictions, both the worker's and firm's problems become deterministic, and the conditional distribution $\pi(y|x)$ in this case is concentrated at one point $y = T(x)$, where $T(x)$ is the only assignment of workers to firms which is nondecreasing. The equilibrium wage w only depends on x and satisfies the differential wage equation $w'(x) = \frac{\partial \gamma(x, T(x))}{\partial x}$ and an explicit formula for the equilibrium wage is obtained by integration.

5.3 Conclusion

Over the last decade, a great deal of efforts has been made to bring matching models to data. In the transferable utility class of models, following Choo and Siow's seminal contribution, various extensions have been proposed to enrich the empirical methodology. These extensions were so far limited to the case when transfers are not observed. However, the observation of transfers allows to widen the scope of identified objects in this class of models, and in particular allows the analyst to separately identify the (pre-transfer) values of a match for each partner. Our paper proposes an intuitive and tractable maximum likelihood approach to structurally estimate these values of a match for each partner using data about matches and transfers from a single market.

We illustrate the usefulness of our methodology to the estimation of compensating wage differentials for the risk of fatal injury on the job. Using the 2017 March CPS data together with CFOI data on fatal injury per occupation and industry, our estimate of the value of job amenities related to risk translates into a Value of Statistical Life of \$6.3 million (\$2017). This estimate is \$3 million lower (though not significantly) than the one obtained by applying a classical hedonic regression technique on our data. Since the hedonic approach can be seen as an extreme version of our method where all the weight in the likelihood function is put on fitting transfers (wages) and none on fitting matching patterns, this suggests that not accounting explicitly for the sorting of workers to jobs can lead to biases in the estimation of the value of statistical life.

Tables

Table 1: Descriptive statistics of workers' and firms' attributes and hourly wages (in 2017 dollars).

	Mean	Std	Min	Max
Workers				
Years of Schooling (in years)	13.35	2.24	1.00	21.00
Experience (in years)	20.67	12.97	0.00	51.00
Female	0.52	0.50	0.00	1.00
Married	0.50	0.50	0.00	1.00
White	0.63	0.48	0.00	1.00
Black	0.12	0.32	0.00	1.00
Asian	0.06	0.24	0.00	1.00
Wage (hourly)	17.95	9.02	3.75	70.00
Firms				
Risk (per 100,000)	3.44	13.05	0.00	345.70
Public	0.12	0.33	0.00	1.00
N	3454			

Note: For measurement purposes, a fatal injury is an injury leading to death within one year of the day of the accident.

Table 2: Effect of firms' and workers' attributes on job amenities and (perceived) productivity (in 2017 dollars), specification 3.

Job Amenities (Alpha)	Main effects	Risk (in 100,000)	Public
Main effects		-0.023	-0.062
		(0.009)	(0.027)
YoS (in years)			0.081
			(0.031)
Productivity (Gamma)	Main effects	Risk (in 100,000)	Public
YoS (in years)	0.057	-0.059	0.838
	(0.035)	(0.020)	(0.099)
Experience (in years)	0.084	0.074	0.096
	(0.025)	(0.029)	(0.104)
Female	-0.404	-2.388	0.548
	(0.061)	(0.238)	(0.212)
Married	0.050		
	(0.020)		
White	0.046		
	(0.021)		
Black	-0.108		
	(0.039)		
Asian	0.069		
	(0.035)		
Experience squared (in years)	-0.051		
	(0.016)		
Salary constant	2.981		
	(0.373)		
Sigma 1	0.046		
Sigma 2	2.233		
R-square	0.235		

Notes: This table reports the estimates of the main effects of workers' characteristics on (perceived) productivity and firms' characteristics on job amenities as well as the interaction of workers' characteristics and firms' characteristics on (perceived) productivity and job amenities. All effects are measured in dollars (per hour of work). All non dummy covariates are standardized to have a standard deviation of 1. Standard errors, calculated from the Hessian of the likelihood, are in parentheses.

Appendix

A The continuous logit framework

Recall that the value for a worker x of the job amenities at firm y is given by $U(x,y) + \sigma_1 \varepsilon(y)$ where $U(x,y) = \alpha(x,y) + w(x,y)$ is deterministic, and $\varepsilon(y)$ is a worker-specific random process. As in Dupuy and Galichon (2014), we choose to model the random process $\varepsilon(y)$ as a *Gumbel random process*, introduced by Cosslett (1988) and Dagsvik (1988).

Assume that workers form their demand by drawing a random pool of observable types of firms, along with the corresponding utility shocks. We call this pool a worker's "random pool of prospects."

Let $k \in \mathbb{N}$ index firms in a worker's pool of prospects and $\{(y_k, \varepsilon_k), k \in \mathbb{N}\}$ be the points of a Poisson process on $\mathcal{Y} \times \mathbb{R}$ with intensity $dye^{-\varepsilon}d\varepsilon$. A worker of type x therefore chooses a firm's of type y by looking at her pool of prospects and solving the utility maximization program

$$\tilde{U} = \max_{y \in \mathcal{Y}} \{U(x,y) + \sigma_1 \varepsilon(y)\} = \max_{k \in \mathbb{N}} \{U(x,y_k) + \sigma_1 \varepsilon_k\},$$

where $\tilde{U}$ denotes the worker's (random) indirect utility. The worker's program induces conditional density of choice probability of firm's type given worker's type, which is expressed as follows:

Proposition A.1. *The conditional density of probability of choosing a firm of type y for a worker of type x is given by*

$$\pi(y|x) = \frac{\exp\left(\frac{U(x,y)}{\sigma_1}\right)}{\int_{\mathcal{Y}} \exp\left(\frac{U(x,y')}{\sigma_1}\right) dy'}$$

while the expected indirect utility of a worker of type x, denoted $u(x) = \mathbb{E}\left[\tilde{U}|x\right]$, is expressed as

$$u(x) = \sigma_1 \log \int_{\mathcal{Y}} \exp\left(\frac{U(x, y')}{\sigma_1}\right) dy'.$$

This result was obtained by Cosslett (1988) and Dagsvik (1988). The intuition of the result is that the c.d.f. of the random utility $\tilde{U}$ conditional on $X = x$ is given by $F_{\tilde{U}|X=x}(z|x) = \Pr\left(\tilde{U} \le z|X = x\right)$, which is the probability that the process (y_k, ε_k) does not intersect the set $\{(y, e) : U(x, y) + \sigma_1 e > z\}$. Hence, the log probability of the event $\tilde{U} \le z$ is minus the integral of the intensity of the Poisson process over this set, that is

$$\begin{aligned}
\log \Pr\left(\tilde{U} \le z|X = x\right) &= -\int_{\mathcal{Y}} \int_{\mathbb{R}} 1\{U(x, y) + \sigma_1 e > z\} e^{-\varepsilon} de dy \\
&= -\exp\left(-z + \log \int_{\mathcal{Y}} \exp\left(\frac{U(x, y)}{\sigma_1}\right) dy\right),
\end{aligned}$$

which is the c.d.f. of a Gumbel distribution with location parameter $\log \int_{\mathcal{Y}} \exp(U(x, y))\, dy$, and scale parameter σ_1.

B Proofs and additional results

Let Da and Db be the two $n \times K$ matrices of respective terms $\partial a_i(\Phi) / \partial \Phi_k$ and $\partial b_j(\Phi) / \partial \Phi_k$ respectively. Let Π be the matrix of terms $\pi_{ij}^{\Phi} = \exp\left(\phi_{ij}(\Phi) - a_i(\Phi) - b_j(\Phi)\right)$, and let $\tilde{\Pi}$ be the same matrix where the entries on the first row have been replaced by zeroes. Let E be the $n \times K$ matrix whose terms E_{ik} are such that $E_{1k} = 0$ for all k, and $E_{ik} = \sum_{j=1}^{n} \pi_{ij}^{\Phi} \varphi_k(x_i, y_j)$ for $i \ge 2$ and all k. Let F be the $n \times K$ matrix of terms such that $F_{jk} = \sum_{i=1}^{n} \pi_{ij}^{\Phi} \varphi_k(x_i, y_j)$.

Lemma 1. *The derivatives of the a_i's and the b_i's with respect to the Φ_k's are given* *by matrices Da and Db such that*

$$\begin{pmatrix} Da \\ Db \end{pmatrix} = \begin{pmatrix} I & \tilde{\Pi} \\ \Pi^\top & I \end{pmatrix}^{-1} \begin{pmatrix} E \\ F \end{pmatrix}. \tag{B.1}$$

Proof of lemma 1. Recall that

$$(Da)_{ik} := \frac{\partial a_i(\Phi)}{\partial \Phi_k} \text{ and } (Db)_{jk} := \frac{\partial b_j(\Phi)}{\partial \Phi_k}$$

for $1 \leq i \leq n$ and $1 \leq k \leq K$. Note that the system in equation (3.3) is normalized such that $a_1(\Phi) = 0$, one has that $\partial a_1(\Phi)/\partial \Phi_k = 0$ for all k. Differentiation yields

$$\begin{aligned} Da_{1k} &= 0 \\ Da_{1k} + \sum_{j=1}^{n} \pi_{ij}^{\Phi} Db_{jk} &= E_{ik},\ i \in \{2, ..., n\} \\ \sum_{i=1}^{n} \pi_{ij}^{\Phi} Da_{ik} + Db_{jk} &= F_{jk},\ j \in \{1, ..., n\}, \end{aligned}$$

where $\pi_{ij}^{\Phi} = \exp\left(\phi_{ij}(\Phi) - a_i(\Phi) - b_j(\Phi)\right)$. Recall that under the linear parameterization we have adopted in section 3.4, $\partial \phi_{ij}(\Phi)/\partial \Phi_k = \varphi_k(x_i, y_j)$ and let

$$\begin{aligned} E_{1k} &= 0,\ E_{ik} = \sum_{j=1}^{n} \pi_{ij}^{\Phi} \varphi_k(x_i, y_j) \text{ for } i \geq 2, \text{ and} \\ F_{jk} &= \sum_{i=1}^{n} \pi_{ij}^{\Phi} \varphi_k(x_i, y_j) \text{ for all } j, \end{aligned}$$

this system rewrites

$$\begin{pmatrix} I & \widetilde{\Pi} \\ \Pi^\top & I \end{pmatrix} \begin{pmatrix} Da \\ Db \end{pmatrix} = \begin{pmatrix} E \\ F \end{pmatrix} \tag{B.2}$$

where block $\tilde{\Pi}$ is the $n \times n$ matrix of term $\widetilde{\pi}_{ij}^{\Phi}$ so that $\widetilde{\pi}_{1j}^{\Phi} = 0$ for all $j \in \{1, ..., n\}$ and $\widetilde{\pi}_{ij}^{\Phi} = \pi_{ij}^{\Phi}$ for $i \geq 2$ and all $j \in \{1, ..., n\}$, and block Π is the $n \times n$ matrix of term π_{ij}^{Φ}. It is easily checked that the matrix on the left hand-side of (B.2) is invertible.

One therefore obtains Da and Db as

$$\begin{pmatrix} Da \\ Db \end{pmatrix} = \begin{pmatrix} I & \widetilde{\Pi} \\ \Pi^{\top} & I \end{pmatrix}^{-1} \begin{pmatrix} E \\ F \end{pmatrix}.$$

∎

Recall $\theta = (A, \Gamma, \sigma_1, \sigma_2, t, s^2)$ and $\Phi = A + \Gamma$.

Theorem 2. *(i) The partial derivatives of* $\log \hat{L}_1(\theta)$ *with respect to* A_k *and* Γ_k *are given by*

$$\frac{\partial \log \hat{L}_1(\theta)}{\partial A_k} = \frac{\partial \log \hat{L}_1(\theta)}{\partial \Gamma_k} = \sum_{i=1}^{n} \varphi_k(x_i, y_i) - n \sum_{i,j=1}^{n} \pi_{ij}^{\Phi} \varphi_k(x_i, y_j)$$

and the partial derivatives of $\log \hat{L}_1(\theta)$ *with respect to all the other parameters is zero.*

(ii) The partial derivatives of $\log \hat{L}_2(\theta)$ *with respect to any parameter entry* θ_k *other than* s *is given by*

$$\frac{\partial \log \hat{L}_2(\theta)}{\partial \theta_k} = s^{-2} \sum_{i=1}^{n} (W_i - w_i(\theta)) \frac{\partial w_i(\theta)}{\partial \theta_k}$$

(iii) The partial derivative of $\log \hat{L}_2(\theta)$ *with respect to* s^2 *is given by*

$$\frac{\partial \log \hat{L}_2(\theta)}{\partial s^2} = \sum_{i=1}^{n} \frac{(W_i - w_i(\theta))^2}{2s^4} - \frac{n}{2s^2}$$

(iv) The partial derivative of $w_i(\theta)$ *with respect to* t *is one, its derivative with respect to* σ_1 *is* $\gamma_{ii}(\Gamma) - b_i(\Phi)$*, its derivative with respect to* σ_2 *is* $a_i(\Phi) - \alpha_{ii}(\Gamma)$*. The partial derivative of* $w_i(\theta)$ *with respect to* A_k *and* Γ_k *are given by*

$$\begin{aligned} \frac{\partial w_i(\theta)}{\partial A_k} &= \sigma_2 \left(\frac{\partial a_i(\Phi)}{\partial \Phi_k} - \varphi_k(x_i, y_i) \right) - \sigma_1 \frac{\partial b_i(\Phi)}{\partial \Phi_k} \\ \frac{\partial w_i(\theta)}{\partial \Gamma_k} &= \sigma_1 \left(\varphi_k(x_i, y_i) - \frac{\partial b_i(\Phi)}{\partial \Phi_k} \right) + \sigma_2 \frac{\partial a_i(\Phi)}{\partial \Phi_k}. \end{aligned}$$

(v) The partial derivatives $\partial a_i(\Phi)/\partial\Phi_k$ *and* $\partial b_i(\Phi)/\partial\Phi_k$ *are given by expression (B.1) in lemma (1).*

Proof of theorem 2. The log-likelihood given in equation (3.10) is made of two terms, the first of which, $\log\hat{L}_1(\theta)$ only depends on θ through Φ, while the second one, $\log\hat{L}_2(\theta)$ depends on all the parameters of the model. The differentiations yielding points (i)-(v) are straightforward. ∎

References

[1] Abowd, J., R.H. Creecy, and F. Kramarz (2002). "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data,"Technical Paper 2002-06, Longitudinal Employer-Household Dynamics, Centre for Economic Studies, U.S. Census Bureau.

[2] Abowd, J., F. Kramarz, and D. Margolis (1999). "High Wage Workers and High Wage Firms,"*Econometrica*, 67(2), pp. 251–333.

[3] André, Pierre and Yannick, Dupraz (2019). "Education and polygamy : evidence from Cameroon," Working Paper. Coventry: University of Warwick. Department of Economics. Warwick economics research papers series (WERPS).

[4] Andrews, M.J., L. Gill, T. Schank, and R. Upward (2008). "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?," *Journal of the Royal Statistical Society, Series A*, 171 (3), pp. 673-697.

[5] Andrews, M.J., L. Gill, T. Schank, and R. Upward (2012). "High Wage Workers Matched with High Wage Firms: Clear Evidence of the Effects of Limited Mobility Bias," *Economics Letters,* 117 (3), pp. 824-827.

[6] Becker, G. S. (1971). *The Economics of Discrimination*. 2nd Edition, University of Chicago Press, Chicago.

[7] Becker, G. S. (1973). "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–46.

[8] Brown, C. (1980). "Equalizing Differences in the Labor Market," *Quarterly Economics Journal,*" 94 (1), pp. 113-134.

[9] Chiappori, P.-A., B. Salanié, and Y. Weiss (2015). "Assortative Matching on the Marriage Market: A Structural Investigation," working paper.

[10] Chiappori, P.-A., B. Salanié, and Y. Weiss (2017). "Partner Choice, Investment in Children, and the Marital College Premium," *American Economic Review*, 107(8), pp. 2109-67.

[11] Choo, E., and A. Siow (2006). "Who Marries Whom and Why," *Journal of Political Economy*, 114(1), pp. 175–201.

[12] Cosslett, S. (1988). "Extreme-value stochastic processes: A model of random utility maximization for a continuous choice set," Technical report, Ohio State University.

[13] Dagsvik, J. (1994). "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes," *Econometrica*, 62, 1179–1205.

[14] Dupuy, A., and A. Galichon (2014). "Personality Traits and the Marriage Market," the *Journal of Political Economy,* 122 (6), pp. 1271-1319.

[15] Ekeland, I., Heckman, J., and Nesheim, L. (2004). "Identification and estimation of Hedonic Models,"*Journal of Political Economy*, 112 (S1), pp. S60–S109.

[16] Galichon, A., and B. Salanié (2015). "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," working paper.

[17] Garen, J.E., (1988). "Compensating Wage Differentials and the Endogeneity of Job Riskiness," the *Review of Economics and Statistics,* 70(1), pp. 9-16.

[18] Gretsky, N., J. Ostroy, and W. Zame (1992). "The nonatomic assignment model," *Economic Theory*, 2(1), 103–127.

[19] Gruetter, M. and R. Lalive (2009). "The Importance of Firms in Wage Determination," *Labour Economics*, 16 (2), pp. 149-160.

[20] Katz, L. and K. Murphy (1992). "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107 (1), pp. 35-78.

[21] Heckman, J. J., Matzkin, R.L. and Nesheim, L. (2010). "Nonparametric Identi.cation and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78 No. 5, pp. 1569-1591.

[22] Hwang, H.S., Reed, W.R. and Hubbard, C. (1992). "Compensating Wage Differentials and Unobserved Productivity," *Journal of Political Economy*, 100 (4), pp. 835.58.

[23] Kniesner, Thomas, J., Viscusi, W. Kip, Woock, Christopher and Ziliak, James P. (2007). "Pinning Down the Value of Statistical Life,"IZA Discussion Paper No. 3107.

[24] Kniesner, Thomas, J., Viscusi, W. Kip, Woock, Christopher and Ziliak, James P. (2012). "The value of statistical life: Evidence from panel data," *Review of Economics and Statistics*, 94(1), pp. 74-87.

[25] Lucas, Robert E B, (1977). "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling," *American Economic Review*, 67(4), pp. 549-58.

[26] Rosen, Sherwin, (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82(1), pp. 34-55.

[27] Rosen, Sherwin, (1986). "The theory of equalizing differences," Handbook of Labor Economics, in: O. Ashenfelter & R. Layard (ed.), Handbook of Labor Economics, edition 1, volume 1, chapter 12, pp. 641-692 Elsevier.

[28] Rüschendorf, L., and Thomsen, W. (1993). "Note on the Schrodinger Equation and I-projections," *Statistics and Probability Letters*, 17, p. 369–375.

[29] Salanié, B. (2015). "Identification in Separable Matching with Observed Transfers," Mimeo.

[30] Sattinger, M. (1979). "Differential Rents and the Distribution of Earnings," *Oxford Economic Papers*, 31(1), pp. 60–71.

[31] Sattinger, M. (1993). "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31(2), pp. 831–80.

[32] Shapley, L., and M. Shubik (1972). "The Assignment Game I: The Core," *International Journal of Game Theory*, 1, 111–130.

[33] Thaler, R., and Rosen, S. (1976). "The Value of Saving a Life: Evidence from the Labor Market." In *Household Production and Consumption*, Nestor E. Terleckyj (ed.). New York: NBER pp. 265–298.

[34] Torres, S. P. Portugal, J.T. Addison and P. Guimarães, (2013). "The Sources of Wage Variation: A Three-Way High-Dimensional Fixed Effects Regression Model,"IZA working paper No. 7276.

[35] Villani, C. (2003). *Topics in Optimal Transportation,* Lecture Notes in Mathematics, AMS.

[36] Villani, C. (2009). *Optimal transport, old and new*. Grundlehren der Mathematischen Wissenschaften, 338, Springer-Verlag, Berlin.

[37] Viscusi, W.K. and J. E. Aldy, (2003). "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World," *Journal of Risk and Uncertainty*, 27(1), pp. 5-76.

[38] Viscusi, W.K. (2003). "Racial Differences in Labor Market Values of a Statistical Life,", *The Journal of Risk and Uncertainty*, 27(3), pp. 239-256.

[39] Viscusi, W.K. (2007). "The Value of Life: Estimates with Risks by Occupation and Industry,"*Economic Inquiry*, 42(1), pp. 29-48.

[40] Viscusi, W.K. (2013). "Using Data from the Census of Fatal Occupational Injuries to Estimate the 'Value of a Statistical Life,'"*Monthly Labor Review*, pp. 1-17.

[41] Woodcock, S.D. (2010). "Heterogeneity and Learning in Labor Markets," The B.E. Journal of Economic Analysis and Policy, Berkeley Electronic Press, 10 (1), pages 85.