FACULTY OF LAW, ECONOMICS AND FINANCE

CREA
Discussion
Paper
2017-14
Economics

Center for Research in Economics and Management
University of Luxembourg

A Note on the Estimation of Job Amenities and Labor Productivity

Arnaud Dupuy, CREA, Université du Luxembourg
and IZA, Germany
Alfred Galichon, New York University, USA
and Toulouse School of Economics, France

May, 2017

For editorial correspondence, please contact: crea@uni.lu
University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faïencerie
L-1511 Luxembourg

A NOTE ON THE ESTIMATION OF JOB AMENITIES AND LABOR PRODUCTIVITY

ARNAUD DUPUY[§] AND ALFRED GALICHON[†]

Abstract. This note 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. Our estimator is suited for applications to a wide range of assignment problems.

Keywords: Matching, Observed transfers, Structural estimation, Value of job amenities, Value of productivity.

JEL Classification: C35, C78, J31.

Date: First circulated version: October, 2015. This version: July 2017. Dupuy's research has received funding from the FNR CORE grant C14/SC/8337045. Galichon gratefully acknowledges the support of NSF grant DMS-1716489, and ERC grants FP7-295298, FP7-312503, and FP7-337665.

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).

In the case when transfers are not observed, thus in the case of the marriage literature, Choo and Siow (2006) is a key 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 problematic when the characteristics are continuous and multivariate. In this framework, Salanié (2015) shows that this structure implies very strong testable restrictions.

In this note, we build a flexible and tractable model of equilibrium matching and wages on the labor market, and we 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.

Our paper is closely related to the literature on compensating wage differentials initiated as an application of Rosen's (1974) hedonic model to the labor market by Lucas (1977) and Thaler and Rosen (1976), 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 wage-amenities gradient to uncover the marginal willingness of workers to pay for or accept certain levels of (dis-)amenities at their job. 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. For instance, Brown (1980) proposed using panel data to estimate workers' fixed effects and control hence for unobserved heterogeneity whereas Garen (1988) suggested an instrumental variable approach. Departure from this assumption implies an inherent bias in the estimates of the compensating wage differentials, and Hwang et al. (1992) have shown that this bias can be large. Our method contributes to this literature by proposing a structural estimation technique to derive unbiased estimates of the value of job amenities and productivity in a matching market where agents match on multivariate and continuous characteristics using data about a single market.

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 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 for more details about its otherwise main properties. In the remainder, to fix ideas, we use the example of the labor market where typically transfers, in the form of wages, are observed.

2.1. **Set up.** Consider a one-to-one, bipartite matching model with transferable utility. Assume that a worker's characteristics are contained in a vector of attributes $x \in \mathcal{X} = \mathbb{R}^{d_x}$, while a firm's characteristics are captured by a vector of attributes $y \in \mathcal{Y} = \mathbb{R}^{d_y}$.

Let $w(x, y)$ be the wage of worker x when working for firm y. It is assumed that worker x not only values her wage at firm y but also the amenities of her job at firm y and let $\tilde{\alpha}(x, y)$ denote the value of these amenities at firm y for a worker x. As in Dupuy and Galichon (2014), the value of amenities is further assumed to be decomposed into a systematic value $\alpha(x, y)$ and a random value $\varepsilon(x, y)$ modelled as

Gumbel random process, whose definition is recalled in Appendix A. One therefore has $\tilde{\alpha}(x,y) = \alpha(x,y) + \sigma_1 \varepsilon(x,y)$ where σ_1 is a scaling factor. By Proposition A.1 in that appendix, this implies that the density of probability that choosing firm y is optimal for a worker of type x is given by

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

while the expected indirect utility of worker x is

$$u(x) = \sigma_1 \log \int_{\mathcal{Y}} \exp\left(\frac{\alpha(x,y') + w(x,y')}{\sigma_1}\right) dy'. \tag{2.2}$$

Similarly, it is assumed that $\widetilde{\gamma}(x,y)$ measures the value of productivity when a firm y matches with a worker x. This value consists of a systematic part $\gamma(x,y)$ and a random part $\eta(x,y)$ modelled as a Gumbel random process, i.e. $\widetilde{\gamma}(x,y) = \gamma(x,y) + \sigma_2 \eta(x,y)$ where σ_2 is a scaling factor. A firm y employing worker x generates therefore profits $\widetilde{\gamma}(x,y) - w(x,y)$. Under this specification, one can show that the density of probability that choosing worker x is optimal for a firm of type y is given by

$$\pi(x|y) = \frac{\exp\left(\frac{\gamma(x,y)-w(x,y)}{\sigma_2}\right)}{\int_{\mathcal{X}} \exp\left(\frac{\gamma(x',y)-w(x',y)}{\sigma_2}\right) dx'} \tag{2.3}$$

while the expected indirect profits of firm y is

$$v(y) = \sigma_2 \log \int_{\mathcal{X}} \exp\left(\frac{\gamma(x',y) - w(x',y)}{\sigma_2}\right) dx' \tag{2.4}$$

It is assumed that there is the same mass of workers and firms, and that the workers' types x have a density of probability $f(x)$, and the firms' types y have a density of probability $g(y)$. Therefore a *feasible matching* between workers and firms, which consists in the probability density $\pi(x,y)$ of occurrence of a (x,y) pair, should have marginal densities f and g. More formally,

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

We now define an equilibrium on this market.

Definition 1. *An equilibrium outcome* (π,w) *consists of an equilibrium matching* $\pi(x,y)$*, and an equilibrium wage* $w(x,y)$ *such that:*

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

(ii) Matching π *is optimal for workers and firms given wage curve* w*, i.e. (2.1) and (2.3) hold.*

Note that when the scaling factors of the random values σ_1 and σ_2 tend to zero, 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 Remark 2.4 below.

2.2. **Equilibrium Characterization.** Let

$$\phi(x,y):=\alpha(x,y)+\gamma(x,y)$$

be the deterministic joint value of a match between a worker of type x and a firm of type y.

Theorem 1. *Under the above assumptions:*

(i) The equilibrium matching $\pi\in\mathcal{M}(f,g)$ *is given by*

$$\pi(x,y)=\exp\left(\frac{\phi(x,y)-a(x)-b(y)}{\sigma}\right),\tag{2.5}$$

where (a, b) normalized so that $a(x_0) = 0$ for some $x_0 \in \mathcal{X}$ is the unique solution to the system of equations

$$\begin{aligned} \int_{\mathcal{Y}} \exp\left(\frac{\phi(x, y) - a(x) - b(y)}{\sigma}\right) dy &= f(x) \\ \int_{\mathcal{X}} \exp\left(\frac{\phi(x, y) - a(x) - b(y)}{\sigma}\right) dx &= g(y). \end{aligned} \tag{2.6}$$

(ii) The terms a and b that appear in (2.5) are related to the expected indirect utilities u and profits v that appear in (2.2) and (2.4) by

$$\begin{aligned} u(x) &= a(x) + \sigma_1 \log f(x) + t \\ v(y) &= b(y) + \sigma_2 \log g(y) - t \end{aligned}$$

for some $t \in \mathbb{R}$.

(iii) The equilibrium wage w is given by

$$w(x, y) = \frac{\sigma_1}{\sigma}(\gamma(x, y) - b(y)) + \frac{\sigma_2}{\sigma}(a(x) - \alpha(x, y)) + t. \tag{2.7}$$

Proof. See the Appendix. ∎

Theorem 1 suggests a few important remarks.

Remark 2.1 (Location normalization)**.** If $a(x)$ and $b(y)$ are solutions of system (2.6), so are $a(x)+t$ and $b(y)-t$. Using Equation (2.7) of Theorem 1, the equilibrium wages are $w(x, y)$ for the former solution and $w(x, y) + t$ for the later. The nonuniqueness of the solution for system (2.6) requires a normalization which is reflected in both the constant term t appearing in the equilibrium wages equation (2.7) and the arbitrary choice $a(x_0) = 0$.

Remark 2.2 (Scale normalization)**.** Equation (2.5) clearly indicates that the matching probabilities are scale invariant with respect to σ, i.e. two markets otherwise similar but one with (α, γ, σ) and one with $(\frac{\alpha}{\sigma}, \frac{\gamma}{\sigma}, 1)$ generate the same equilibrium matches. However, inspection of Equation (2.7) reveals that wages are not scale invariant with respect to σ. While the two markets discussed above generate the same equilibrium matches, they generate different equilibrium wages. As a results, with observed wages, one can identify the scale of the systematic value of a match ϕ together with the amount of heterogeneity necessary to rationalize the data. The normalization of ϕ required when only matches are observed (see Dupuy and Galichon, 2014), is not necessary when in addition wages are also observed. Interestingly enough, this implies that both the scale of the deterministic value of a match ϕ and the amount of unobserved heterogeneity necessary to rationalize the data can be compared across markets when both matches and wages are observed in those markets.

Remark 2.3 (Identification of α and γ)**.** The deterministic value of amenities α and productivity γ do not appear individually in the expression of the equilibrium matches in Equation (2.5), only the joint value of a match ϕ appears in this equation. However, α and γ do appear separately in the formula of the equilibrium wages in Equation (2.7). 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 whereas, in contrast, if wages are observed, one actually can identify and estimate these objects separately.

Remark 2.4 (The Becker-Shapley-Shubik model)**.** When $\sigma \to 0$, the model converges to the classical model of Monge-Kantorovich, which is a continuous extension of the model by Becker-Shapley-Shubik. 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 y such that $u(x) = \alpha(x,y) + w(x,y)$. Similarly, when $\sigma_2 \to 0$, firm y's 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 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

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

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).

Remark 2.5 (The Sattinger model)**.** Our model embeds 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 ($\sigma \to 0$; see Remark 2.4), 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, namely

$$w'(x) = \frac{\partial \gamma(x, T(x))}{\partial x}, \tag{2.8}$$

and an explicit formula for the equilibrium wage is obtained by integration.

Remark 2.6 (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.

Remark 2.7 (Identification)**.** Suppose the values of σ_1 and σ_2 are known, and $\sigma_1 + \sigma_2 = 1$ for notational simplicity. Equations (2.1) and (2.3) clearly indicate that $\alpha\left(x,y\right) + w\left(x,y\right)$ is identified up to a function $c\left(x\right)$ by $\sigma_1 \ln \pi\left(x,y\right) + c\left(x\right)$, and $\gamma\left(x,y\right) - w\left(x,y\right)$ is identified up to a function $d\left(y\right)$ by $\sigma_2 \ln \pi\left(x,y\right) + d\left(y\right)$. It follows that α is identified up to fixed effects $c\left(x\right)$ by

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

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

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

This result has been exploited in a nonparametric setting by Galichon and Salanie (2015) and Salanie (2015). In this paper, it is exploited in a parametric setting based on basis functions of x and y (see Section 3.3). Indeed, since α is identified up to fixed effects $c\left(x\right)$, 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. 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 in $\{(X_i, Y_i, W_i), i = 1, ..., n\}$, 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 wage W_i^* assumed to be such that

$$W_i = W_i^* + \epsilon_i$$

where measurement error ϵ_i follows a $\mathcal{N}(0, s^2)$ distribution and is independent of (X_i, Y_i). The latent wage W_i^* is assumed to be an equilibrium wage, and is thus related to (X_i, Y_i) by

$$W_i^* = w(X_i, Y_i)$$

where $w(.,.)$ is as given by Equation (2.7).

3.2. **Scaling.** For the sake of readability and to avoid additional notational burden, we propose the following rescaling of the model. 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

$$\pi(x, y) = \exp(\phi(x, y) - a(x) - b(y)), \tag{3.1}$$

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

$$\begin{cases} \int_{\mathcal{Y}} \exp(\phi(x, y) - a(x) - b(y))\, dy = f(x) \\ \int_{\mathcal{X}} \exp(\phi(x, y) - a(x) - b(y))\, dx = g(y), \end{cases} \tag{3.2}$$

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

$$u(x) = \sigma a(x) + \sigma_1 \log f(x) + t, \text{ and } v(y) = \sigma b(y) + \sigma_2 \log g(y) - t \tag{3.3}$$

and the equilibrium wage w is given by

$$w(x,y) = \sigma_1 (\gamma(x,y) - b(y)) + \sigma_2 (a(x) - \alpha(x,y)) + t. \tag{3.4}$$

This scaling is without loss of generality since from Equation (3.4) one can estimate parameters σ_1 and σ_2 and hence $\sigma = \sigma_1 + \sigma_2$ and therefore recover the pre-scaling values of α and γ. In the remainder of the paper, Equations (3.1)–(3.4) will characterize the model to estimate.

3.3. **Parametrization.** Let A and Γ be two vectors of $\mathbb{R}^k$ parameterizing the function of workers' systematic net 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 value of a match reads as

$$\phi(x,y;\Phi) = \sum_{k=1}^{K} \Phi_k \varphi_k(x,y), \tag{3.5}$$

where $\Phi_k = A_k + \Gamma_k$. Inspection of Equation (3.4) reveals that, given the parametric choice above, equilibrium matching and wages 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 wage as $w(x,y;\theta)$.

3.4. **Estimation by maximum likelihood.** The main purpose is to estimate the vector of parameters θ. To this aim we adopt a maximum likelihood approach. It follows from Remark (2.3) that the likelihood of observing a pair (x,y) only depends on $\Phi = A + \Gamma$, and is given by

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

where $a(x;\Phi)$ and $b(y;\Phi)$ are uniquely determined by system of equations (3.2). Since, by assumption, measurement errors in wages 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:

$$\log L\left(\theta\right)=n\mathbb{E}_{\hat{\pi}}\left[\phi\left(X,Y;\Phi\right)-a\left(X;\Phi\right)-b\left(Y;\Phi\right)-\frac{\left(W-w\left(X,Y;\theta\right)\right)^{2}}{2s^{2}}\right]-\frac{n}{2}\log s^{2} \tag{3.6}$$

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 (3.6) are computed in the population; here, we only have access to a sample, so we compute the sample analog of system (3.2), that is

$$\begin{cases}\sum_{j=1}^{n}\exp\left(\phi_{ij}\left(\Phi\right)-a_{i}-b_{j}\right)=1/n\\ \sum_{i=1}^{n}\exp\left(\phi_{ij}\left(\Phi\right)-a_{i}-b_{j}\right)=1/n\end{cases} \tag{3.7}$$

with the added normalization $a_1 = 0$, which ensures uniqueness of the solution. (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 $\left(a_i\left(\Phi\right),b_i\left(\Phi\right)\right)$ this solution at Φ. This allows us to compute a sample estimate of the equilibrium wage $w_i\left(\theta\right)$ as

$$w_{i}\left(\theta\right):=\sigma_{1}\left(\gamma_{ii}\left(\Gamma\right)-b_{i}\left(\Phi\right)\right)+\sigma_{2}\left(a_{i}\left(\Phi\right)-\alpha_{ii}\left(A\right)\right)+t, \tag{3.8}$$

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

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

Theorem 2. *The log likelihood of the sample is given by*

$$\log \hat{L}(\theta) = \log \hat{L}_1(\theta) + \log \hat{L}_2(\theta), \tag{3.9}$$

where

$$\log \hat{L}_1(\theta) = \sum_{i=1}^{n} \left(\phi_{ii}(\Phi) - a_i(\Phi) - b_i(\Phi)\right) \tag{3.10}$$

and,

$$\log \hat{L}_2(\theta) = -\sum_{i=1}^{n} \frac{\left(W_i - w_i(\theta)\right)^2}{2s^2} - \frac{n}{2}\log s^2, \tag{3.11}$$

where $\phi_{ij}(\Phi) := \phi(X_i, Y_j; \Phi)$ *is as in (3.5),* $a_i(\Phi)$ *and* $b_i(\Phi)$ *are obtained as the solution of (3.7), and where* $w_i(\theta)$ *is given by (3.8).*

Proof. See Appendix B. ∎

Theorem 2 motivates the following remarks.

Remark 3.1 (Interpretation of the objective function)**.** Expression (3.9) has a straightforward interpretation. The term $\log \hat{L}_1(\theta)$, whose expression is given in Equation (3.10) 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

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

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 (3.11) tends to match the predicted wages $w_i(\theta)$ with the observed wages W_i in order to minimize the sum of the square deviations $\left(W_i - w_i(\theta)\right)^2$. Hence, the contribution of the second term is to equate the predicted wages 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 wages 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, and its value is determined by the optimization procedure.

Remark 3.2 (Concentrated Likelihood)**.** 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 $(\sigma_1, \sigma_2, t, s^2)$ are auxiliary, and the focus of attention is the *concentrated log-likelihood*, which is given by

$$\log l\,(A, \Gamma) := \max_{\sigma_1,\sigma_2,t,s^2} \log \hat{L}\,(\theta) = \log \hat{L}_1\,(\Phi) + \max_{\sigma_1,\sigma_2,t,s^2} \log \hat{L}_2\left(A, \Gamma, \sigma_1, \sigma_2, t, s^2\right).$$

where as usual, $\Phi = A + \Gamma$. Denoting $\sigma_1^*, \sigma_2^*, t^*$ and s^{*2} the optimal value of the corresponding parameters given A and Γ, one gets

$$(\sigma_1^*, \sigma_2^*, t^*) = \arg\min_{\sigma_1,\sigma_2,t} \sum_{i=1}^{n} \left(W_i - w_i\,(\theta)\right)^2, \tag{3.12}$$

which is the solution to a Nonlinear Least Squares problem which is readily implemented in standard statistical packages, and $s^{*2} = n^{-1} \sum_{i=1}^{n} \left(W_i - w_i\,(\theta^*)\right)^2$. The partial derivative of the concentrated log likelihood with respect to A_k is given by

$$\frac{\partial \log l\,(A, \Gamma)}{\partial A_k} = \frac{\partial \log \hat{L}_1\,(\Phi)}{\partial \Phi_k} + \frac{\partial \log \hat{L}_2\,(A, \Gamma, \sigma_1^*, \sigma_2^*, t^*, s^{*2})}{\partial A_k}$$

and a similar expression holds for $\partial \log l / \partial \Gamma_k$. These formulas are proved in the Appendix B.

Remark 3.3 (Extension to missing wages)**.** In many applications, data will come from surveys where typically non response to questions about earnings are frequently encountered. Our proposed estimation strategy extends to the case where, for some matches, wages are randomly missing. The log likelihood expression presented in Theorem 2 offers a very intuitive way of understanding how missing wages for some observations will impact the estimation. To formalize ideas, let p be the probability that for any arbitrary match the wage is missing. The sample is still representative of the population of matches, but part of the sample consists of matches with observed wages , i.e. $(X_i, Y_i, W_i)_{i=1}^{n^o}$, and the other part of matches with missing wages, i.e. $(X_i, Y_i, .)_{i=n^o+1}^{n}$ where n^o is the number of matches with observed wages and n is as before the size of our sample of matches (we have re-ordered the observations such that those matches with observed wages are indexed first). The log likelihood in this situation is therefore

$$\log \hat{L}(\theta) = \log \hat{L}_1(\theta) + \log \hat{L}_2(\theta) + n^o \log p + (n - n^o)\log(1-p)$$

where $\log \hat{L}_1(\theta)$ is given as in Equation (3.10) and $\log \hat{L}_2(\theta)$ reads now as

$$\log \hat{L}_2(\theta) = -\sum_{i=1}^{n^o} \frac{(W_i - w_i(\theta))^2}{2s^2} - \frac{n^o}{2}\log s^2 \qquad (3.13)$$

thus $p = n^0/n$. As n^o tends to 0, and hence p tends to 0, the log likelihood function tends to $\log \hat{L}_1(\theta)$. In contrast, when n^o tends to n, and hence p tends to 1, the expression of $\log \hat{L}_2(\theta)$ in Equation (3.13) tends to that of $\log \hat{L}_2(\theta)$ in Equation (3.11) such that the log likelihood function tends to Equation (3.9).

3.5. **Gradient of the log-likelihood.** 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 $\pi_{ij}^{\Phi} = \exp\left(\phi_{ij}(\Phi) - a_i(\Phi) - a_j(\Phi)\right)$, 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 $i \geq 2$ and all k. Let F be the $n \times K$ matrix of terms such that $F_{jk} = \sum_{i=1}^{n} \pi_{ji}^{\Phi}\varphi_k(x_i, y_j)$.

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

$$\begin{pmatrix} Da \\ Db \end{pmatrix} = \begin{pmatrix} I & \tilde{\Pi} \\ \Pi' & I \end{pmatrix}^{-1} \begin{pmatrix} E \\ F \end{pmatrix}. \tag{3.14}$$

Proof. See Appendix B. ∎

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

Theorem 3. *(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 (3.14) in Lemma (1).*

Proof. See Appendix B. ∎

4. 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 a very 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.

[§] *CREA, University of Luxembourg, and IZA. Address: University of Luxembourg, 162a, avenue de la Faiencerie L-1511 Luxembourg. Email: arnaud.dupuy@uni.lu.*

[†] *Departments of Economics and of Mathematics, New York University, and Fondation Jean-Jacques Laffont, Toulouse School of Economics. Email: ag133@nyu.edu.*

Appendix A. The continuous logit framework

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

Assume that workers can only possibly have access to a countable random subset of firms. We call this subset of known firms a worker's "prospects". Hence, a worker can only choose to work for a firm within her prospects. Let $k \in \mathbb{N}$ index firms in a worker's 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 to work at firm l of type $y_l = y$ in her prospects if and only if l is a solution of the worker's utility maximization program

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

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 worker 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 $F_{\tilde{U}|X=x}(z|x) = \Pr\left(\tilde{U} \leq z|X = x\right)$, 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 $\tilde{U} \leq z$ is minus the integral of the intensity of the Poisson process over this set, that is

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

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

Appendix B. Proofs

Proof of Theorem 1. Part (i) is proved in Dupuy and Galichon (2014), Theorem 1, and recalled below for completeness. It follows from Equations (2.1) and (2.2) that

$$\pi\left(y|x\right) = \exp\left(\frac{\alpha\left(x,y\right) - u\left(x\right) + w\left(x,y\right)}{\sigma_1}\right), \tag{B.1}$$

thus

$$\sigma_1 \log \pi\left(x,y\right) = \alpha\left(x,y\right) - u\left(x\right) + \sigma_1 \log f\left(x\right) + w\left(x,y\right) \tag{B.2}$$

and similarly on the other side of the market,

$$\sigma_2 \log \pi\left(x,y\right) = \gamma\left(x,y\right) - v\left(y\right) + \sigma_2 \log g\left(y\right) - w\left(x,y\right). \tag{B.3}$$

Adding Equation (B.2) to (B.3) yields

$$\sigma \log\left(\pi\left(x,y\right)\right) = \phi\left(x,y\right) - a\left(x\right) - b\left(y\right) \tag{B.4}$$

where $a\left(x\right) = u\left(x\right) - \sigma_1 \log f\left(x\right) - t$ and $b\left(y\right) = v\left(y\right) - \sigma_2 \log g\left(y\right) + t$, where t is chosen so that $a\left(x_0\right) = 0$, thus $t = u\left(x_0\right) - \sigma_1 \log f\left(x_0\right)$. The potentials a and b that appear in Equation (B.4) are such that $\pi \in \mathcal{M}\left(f,g\right)$, i.e.

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

Uniqueness of such (a,b) such that $a\left(x_0\right) = 0$ is proved in Rüschendorf and Thomsen (1993), Theorem 3.

Let us now show (ii). Taking the logarithm of Equation (B.1) and rearranging yields

$$u(x) = \alpha(x, y) + w(x, y) + \sigma_1 \log f(x) - \sigma_1 \log \pi(x, y) \tag{B.5}$$

A similar procedure using the expression for $\pi(x|y)$ obtains,

$$v(y) = \gamma(x, y) - w(x, y) + \sigma_2 \log g(y) - \sigma_2 \log \pi(x, y). \tag{B.6}$$

Adding Equation (B.5) to (B.6) and using Equation (B.4) yields

$$u(x) + v(y) = \sigma_1 \log f(x) + \sigma_2 \log g(y) + a(x) + b(y). \tag{B.7}$$

By rearranging Equation (B.7) to obtain

$$u(x) - \sigma_1 \log f(x) - a(x) = \sigma_2 \log g(y) + b(y) - v(y)$$

one notes that since the right hand-side only depends on x and the left hand-side only depends on y, both these terms must be equal to a constant, denoted t. Hence,

$$u(x) = a(x) + \sigma_1 \log f(x) + t, \text{ and} \tag{B.8}$$

$$v(y) = b(y) + \sigma_2 \log g(y) - t. \tag{B.9}$$

Let us finally show (iii). Plugging Equation (B.8) into Equation (B.5) and rearranging yields

$$w(x, y) = a(x) - \alpha(x, y) + t + \sigma_1 \log \pi(x, y),$$

which, using Equation (B.4) to substitute for $\log \pi(x, y)$, provides the following expression of the equilibrium wages as a function of α, γ, a, b and t,

$$w(x, y) = \frac{\sigma_1}{\sigma}(\gamma(x, y) - b(y)) + \frac{\sigma_2}{\sigma}(a(x) - \alpha(x, y)) + t. \tag{B.10}$$

∎

Proof of Theorem 2. Immediate given the discussion before the Theorem. ∎

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 (3.2) 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 $\pi_{ij}^{\Phi} = \exp\left(\phi_{ij}(\Phi) - a_i(\Phi) - a_j(\Phi)\right)$. Recall that under the linear parameterization we have adopted in Section 3.3, $\partial \phi_{ij}(\Phi)/\partial \Phi_k = \varphi_k(x_i, y_j)$ 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

$$\begin{pmatrix} I & \widetilde{\Pi} \\ \Pi' & I \end{pmatrix} \begin{pmatrix} Da \\ Db \end{pmatrix} = \begin{pmatrix} E \\ F \end{pmatrix} \tag{B.11}$$

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.11) is invertible. One therefore obtains Da and Db as

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

∎

Proof of Theorem 3. The log-likelihood given in Equation (3.9) 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. ∎

Proof of the statement in Remark 3.2. Recall $\theta = (A, \Gamma, \sigma_1, \sigma_2, t, s^2)$. The maximum likelihood problem can be written as

$$\max_{\theta} \log \hat{L}(\theta) = \max_{A,\Gamma} \log l(A, \Gamma)$$

where $\log l(A, \Gamma) = \max_{\sigma_1,\sigma_2,t,s^2} \log \hat{L}(\theta)$ is the concentrated log-likelihood which can be rewritten as

$$\log l(A, \Gamma) = \log \hat{L}_1(\theta) + \max_{\sigma_1,\sigma_2,t,s^2} \log \hat{L}_2(\theta). \tag{B.12}$$

where

$$\max_{\sigma_1,\sigma_2,t,s^2} \log \hat{L}_2(\theta) = -\min_{s^2} \left(\frac{n}{2} \log s^2 + \frac{1}{2s^2} \min_{\sigma_1,\sigma_2,t} \sum_{i=1}^{n} (W_i - w_i(\theta))^2 \right) \tag{B.13}$$

The second minimization in Equation (B.13) is an Ordinary Least Squares problem whose solution given A, Γ, denoted $(\sigma_1^*, \sigma_2^*, t^*)$, is the vector of coefficients of the OLS regression of W on $(\gamma - b, a - \alpha, 1)$. The value of s^2, denoted s^{*2}, is given by

$$s^{*2} = \frac{\sum_{i=1}^{n} (W_i - w_i(\theta^*))^2}{n}.$$

The envelope theorem yields an expression for the gradient of the concentrated log-likelihood with respect to the concentrated parameters A and Γ, that is

$$\nabla_{A,\Gamma} \log l(A, \Gamma) = \nabla_{A,\Gamma} \log \hat{L}_1(\theta^*) + \nabla_{A,\Gamma} \log \hat{L}_2(\theta^*).$$

The elements of the first part of the gradient are given in Theorem 3 part (i) whereas parts (ii), (iv) and (v) of Theorem 3 provide the building blocks for the elements of the second part of the gradient. ∎

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] 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.

[4] 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.

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

[6] Becker, G. S. (1991). *A Treatise on the Family*. Harvard University Press.

[7] Berry, S, J. Levinsohn, and A. Pakes, (1995). "Automobile Prices in Market Equilibrium," *Econometrica*, vol. 63(4), pp. 841-90.

[8] Bojilov, R., and A. Galichon (2013). "Closed-form formulas for multivariate matching". Mimeo.

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

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

[11] Choo, E., and A. Siow (2006). "Who Marries Whom and Why," *Journal of Political Economy*, 114(1), 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,* vol. 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] Fox, J. (2010). "Identification in Matching Games," *Quantitative Economics* 1, 203–254.

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

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

[19] Graham, B. (2011). "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. Jackson. Elsevier.

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

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

[22] 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.

[23] Hwang, H.S., Reed, W.R. and Hubbard, C. (1992). "Compensating Wage Differentials and Unobserved Productivity". *Journal of Political Economy*, 100 No. 4, pp. 835.58.

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

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

[26] 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.

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

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

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

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

[31] Shapley, L., and M. Shubik (1972). "The Assignment Game I: The Core," *International Journal of Game Theory*, 1, 111–130.Sattinger, M. (1978). "Comparative Advantage in Individuals". *Review of Economics and Statistics*, 60 No. 2, pp. 259–67.

[32] 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.

[33] 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.

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

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

[36] 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*, vol. 27(1), pp. 5-76.

[37] 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.