



## Personality traits and the marriage market

Arnaud Dupuy, Alfred Galichon

► **To cite this version:**

| Arnaud Dupuy, Alfred Galichon. Personality traits and the marriage market. 2012. hal-01070393

**HAL Id: hal-01070393**

**<https://hal-sciencespo.archives-ouvertes.fr/hal-01070393>**

Preprint submitted on 1 Oct 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

IZA DP No. 6943

## Personality Traits and the Marriage Market

Arnaud Dupuy  
Alfred Galichon

October 2012

# Personality Traits and the Marriage Market

**Arnaud Dupuy**

*Reims Management School,  
Maastricht School of Management and IZA*

**Alfred Galichon**

*Sciences Po Paris*

Discussion Paper No. 6943  
October 2012

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### Personality Traits and the Marriage Market<sup>\*</sup>

Which and how many attributes are relevant for the sorting of agents in a matching market? This paper addresses these questions by constructing indices of mutual attractiveness that aggregate information about agents' attributes. The first  $k$  indices for agents on each side of the market provide the best approximation of the matching surplus by a  $k$ -dimensional model. The methodology is applied on a unique Dutch household survey containing information about education, height, BMI, health, attitude towards risk and personality traits of spouses. Three important empirical conclusions are drawn. First, sorting in the marriage market is not unidimensional: individuals face important trade-offs between the attributes of their spouses which are not amenable to a single-dimensional index. Second, although education explains a quarter of a couple's observable surplus, personality traits explain another 20%. Third, different personality traits matter differently for men and for women.

JEL Classification: D3, J21, J23, J31

Keywords: multidimensional sorting, saliency analysis, marriage market, personality traits, continuous logit

Corresponding author:

Arnaud Dupuy  
Reims Management School (RMS)  
59, rue Pierre Taittinger  
51100 Reims  
France  
E-mail: [arnaud.dupuy@reims-ms.fr](mailto:arnaud.dupuy@reims-ms.fr)

---

<sup>\*</sup> We thank Jim Heckman, Jean-Marc Robin, and seminar participants at Paris 1 Panthéon – Sorbonne, University of Alicante, Tilburg University and Sciences Po for their comments and Lex Borghans for useful discussions about the DNB data. Jinxin He provided excellent research assistance. Dupuy warmly thanks ROA at Maastricht University where part of this paper was written. Galichon acknowledges support from Chaire EDF-Calyon “Finance and Développement Durable,” and FiME, Laboratoire de Finance des Marchés de l’Energie ([www.fime-lab.org](http://www.fime-lab.org)).

## 1. INTRODUCTION

Marriage plays an important role in the distribution of welfare across individuals and in the intergenerational transmission of economic opportunities. In-depth understanding of marriage patterns is therefore of crucial importance for the study of a wide range of economic issues and in particular income inequality.

A growing body of the economic literature studies both empirically and theoretically the determinants of marriage as a competitive matching market. This literature draws insights from the seminal model of the marriage market developed by Becker (1973). At the heart of this theory lies a two-sided assignment model with transferable utilities where agents on both sides of the market (men and women) are characterized by a set of attributes which is only partly observed by the econometrician. Each agent aims at matching with a member of the opposite sex so as to maximize her own payoffs. This model is particularly interesting since under certain conditions (mainly restrictions on the shape of the surplus function), one can identify and estimate features of agents' preferences. A central question in this market is which and how many attributes are relevant for the sorting of agents?

Empirical studies in this literature (see among others, Becker, 1991, Wong, 2003, Anderberg, 2004, Choo and Siow, 2006, Browning et al., 2007, Chiappori and Oreffice, 2008, Hitsch et al., 2010, Chiappori et al., 2010, Chiappori et al., 2012, Oreffice and Quintana-Domeque, 2010, Charles et al., 2011, Jacquemet and Robin, 2012) have attempted to identify and estimate preferences in the marriage market, or in other matching markets (Fox, 2010, 2011, Terviö, 2008, Gabaix and Landier, 2008). So far, however, this literature has faced two important limitations that impede answers to the central question raised above.

The first limitation is related to the quantitative methods available to identify and estimate features of the surplus function. Up until recently<sup>1</sup>, these tools were only designed to study matching markets where sorting is unidimensional, i.e. occurs on a single index. Due

---

<sup>1</sup>Recently, two papers have studied markets where sorting occurs on more than one dimension. Chiappori et al. (2011) study sorting on a single continuous index of socio-economic variables and a binary variable capturing smoking behavior. Nesheim (2012) focuses on the identification of multivariate hedonic models without heterogeneity, and based on the observation of the price.

to this limitation, empirical studies to date have therefore either focused on one attribute at a time, hence ignoring the effect of other attributes on sorting (see e.g. Charles et al., 2011), or assumed that all observed attributes matter but only through a single index of mutual attractiveness (see e.g. Wong, 2003, Anderberg, 2004 and Chiappori, Oreffice and Quintana-Domeque, 2012).

The second limitation of the current empirical literature is related to the set of observable attributes available in the data. Most studies have only access to education and earnings<sup>2</sup>, and only a few observe a few more dimensions such as anthropometric measures captured by height and BMI or self-assessed measures of health (Chiappori, Oreffice and Quintana-Domeque, 2012 and Oreffice and Quintana-Domeque, 2010 are notable exceptions).

In the present paper, we contribute to the literature on three accounts. First, on the modeling front, we extend the Choo and Siow matching model to account for continuous types. In turns, this allows us to extend Galichon and Salanié’s (2010, 2012) inference of the Choo and Siow model to the continuous case. This extension allows us to consider at the same time attributes that are discrete and attributes that are continuous.

Second, on the data analysis front, we introduce a new technique to determine the most relevant dimensions on which sorting occurs in a matching market. We thus derive “indices of mutual attractiveness” by performing a Singular Value Decomposition of the estimated quadratic matching surplus. We call this new technique *Saliency Analysis*. The first  $k$  indices (for males and females) provide the best approximation of the matching surplus by a model where attributes are vectors of only  $k$  dimension. As a consequence, we can perform inference on the number of dimensions that are required in order to explain the sorting on the observed market by testing how many singular values differ from zero.

Third, on the empirical front, we make use of a unique dataset that allows us to observe a wide range of attributes for both spouses. The set of attributes we observe in the data includes socio-economic variables (education, age), anthropometric measures such as height and BMI, a measure of self assessed health, but also psychometric attributes such

---

<sup>2</sup>Bruze (2011) shows that sorting on education occurs even in segments (movie stars) where it does not relate to financial traits of spouses. He concludes that men and women have strong preferences for nonfinancial attributes of their spouse correlated with education (potentially personality traits).

as risk aversion and the “Big Five” personality traits, well-known in the Psychology literature: conscientiousness, extraversion, agreeableness, emotional stability and autonomy. This paper is, to the extent of our knowledge, the first attempt to evaluate the importance of personality traits in the sorting of men and women in the marriage market.

Saliency Analysis requires to estimate the surplus function, which will be chosen quadratic with respect to the male and female attributes. These attributes are assumed continuously distributed in the population. In the discrete case, one could apply a technique developed by Galichon and Salanié (2010, 2012) in the setting of Choo and Siow (2006) to parametrically estimate this surplus function. The key assumption is that there exist on both sides of the market unobservable attributes that matter for the choice of a partner but leave agents from the other side indifferent. If these attributes enter additively in the agents’ utility function, and follow independent and centered Gumbel distributions, then preferences are identified. Galichon and Salanié have proposed a parametric estimation technique for preferences, which allows the researcher to study the cross-differential effect of variation in the attributes of the two sides of the market. For instance, it allows to answer questions such as whether education matters more for conscientious men/women than for extravert ones. Since the attributes observed in our data are continuous, we extend the Galichon and Salanié technique to the continuous case. The resulting continuous version of the Choo and Siow model, based on extreme value stochastic processes, extends many of the enjoyable properties of that setting.

This paper relates to the literature showing the importance of personality traits in economic decisions (Borghans et al., 2008 for instance). Bowles et al. (2001) and Mueller and Plug (2006) among others have shown the importance of personality traits for earnings inequality. More related to our work, Lundberg (2010) studies the impact of personality traits on the odds in and out of a relationship (marriage and divorce). She finds empirical evidence that personality traits affect significantly the extensive margin in the marriage market. In particular, conscientiousness increases the probability of marriage at 35 for men and extraversion increases the odds of marriage at 35 for women. Our work is complementary to that of Lundberg (2010) in that we study the intensive margin, that is to whom conscientious men and extravert women are the most attractive. We show among

other things that conscientious men have preferences for conscientious and agreeable women whereas extravert women have preferences for extravert and less agreeable men.

The remaining structure of the paper is as follows. Section 2 presents an important extension of the model of Choo and Siow to the case with continuously distributed observables. Section 3 deals with parametric estimation of the surplus function in this setting. Section 4 presents a methodology to derive *indices of mutual attractiveness* that determine the principal dimensions on which sorting occurs. The problem of inferring the number of dimensions on which sorting occurs is dealt with in section 5. Section 6 presents the data used for our empirical estimation and Section 7 discusses the results. Section 8 concludes.

## 2. THE CONTINUOUS CHOO AND SIOW MODEL<sup>3</sup>

**2.1. The Becker-Shapley-Shubik model of marriage.** The setting is a one-to-one, bipartite matching model with transferable utility. Men and women are characterized by vectors of attributes, respectively denoted  $x \in \mathcal{X} = \mathbb{R}^{d_x}$  for men and  $y \in \mathcal{Y} = \mathbb{R}^{d_y}$  for women. Men and women are assumed to be in equal number; let  $P$  and  $Q$  be the respective probability distribution of their attributes.  $P$  and  $Q$  are assumed to have densities with respect to the Lebesgue measure denoted respectively  $f$  and  $g$ . Without loss of generality, it is assumed throughout that  $P$  and  $Q$  are centered distributions, that is  $\mathbb{E}_P[X] = \mathbb{E}_Q[Y] = 0$ .

A matching is the probability density  $\pi(x, y)$  of drawing a couple with characteristics  $(x, y)$  from the population. Quite obviously, this imposes that the marginals of  $\pi$  be  $P$  and  $Q$ , which we write  $\pi \in \mathcal{M}(P, Q)$ , where

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

Let  $\Phi(x, y)$  be the joint surplus generated when a man  $x$  and a woman  $y$  match, to be shared endogenously between them. It is assumed that the utility of unmatched individuals

---

<sup>3</sup>For the purposes of the subsequent analysis, we present a variant of the Choo and Siow model with no singles. However, all of our results apply to the model with singles.



is  $-\infty$  for all  $x$  and  $y$ , so that every market participant is matched. In this setting, Shapley and Shubik (1972) have shown that the optimal matching  $\pi$  maximizes the total surplus

$$\max_{\pi \in \mathcal{M}(P, Q)} \mathbb{E}_{\pi} [\Phi(X, Y)]. \quad (2.1)$$

Optimality condition (2.1) leads to very strong restrictions<sup>4</sup> on  $(X, Y)$ , which are rarely met in practice. We need to incorporate some amount of unobserved heterogeneity in the model.

**2.2. Adding heterogeneities.** Bringing the model to the data requires the additional step of acknowledging that sorting might also occur on attributes that are unobserved to the econometrician. In the case where men and women attributes are *discrete*, Choo and Siow (2006), introduced unobservable heterogeneities, say “sympathy shocks”, into the matching problem by considering that if a man  $m$  of attributes  $x_m = x$  and a woman  $w$  of attributes  $y_w = y$  match, they create a joint surplus  $\Phi(x, y) + \varepsilon_m(y) + \eta_w(x)$ , where  $(\varepsilon_m(y))_y$  and  $(\eta_w(x))_x$  are i.i.d. Gumbel (extreme value type I) distributed of scaling parameter  $\sigma/2$ . In this setting, Galichon and Salanié (2010; 2012) have shown that the equilibrium matching  $\pi \in \mathcal{M}(P, Q)$  is solution to

$$\max_{\pi \in \mathcal{M}(P, Q)} \mathbb{E}_{\pi} \left[ \frac{\Phi(X, Y)}{\sigma} \right] - \mathbb{E}_{\pi} [\ln \pi(X, Y)]. \quad (2.2)$$

However, in many applied settings, the attributes are continuous random vectors. In the present paper, we shall present an application where  $x$  and  $y$  measure height, BMI and various personality traits, which have a continuous multivariate distribution, and hence we need to model the random processes for  $\varepsilon_m(x)$  and  $\eta_w(y)$ .

A natural choice is the continuous logit setting. Although very natural and particularly tractable, this setting has been surprisingly little used in economic modelling, with some notable exceptions. McFadden (1976) initiated the literature of continuous logit models by extending the definition of IIA beyond finite sets. Ben-Akiva and Watanatada (1981)

---

<sup>4</sup>A basic result in the theory of optimal transportation (Brenier’s theorem) implies that in this case, the optimal matching is characterized by  $AY = \nabla V(X)$  where  $V$  is some convex function. Hence as soon as  $A$  is invertible, the matching is *pure*, in the sense that no two men of the same type may marry women of different types. This is obviously never observed in the data.

and Ben-Akiva, Litinas and Tsunekawa (1985) define continuous logit models by taking the limits of the discrete choice probabilities, with applications in particular to the context of spatial choice models. Cosslett (1988) and Dagsvik (1988) have independently suggested using max-stable processes to model continuous choice. This idea was further pursued by Resnick and Roy (1991). We follow their insights in what follows.

Assume that the utility surplus of a man  $m$  of observed attributes  $x$  (that is, such that  $x_m = x$ ) who marries a woman of observed attributes  $y$  can be written as

$$\alpha(x, y) + \tau + \varepsilon_m(y), \quad (2.3)$$

where the sympathy shock  $\varepsilon_m(\cdot)$  is a stochastic process on  $\mathcal{Y}_0$  modelled as an extreme value stochastic process

$$\varepsilon_m(y) = \max_k \left\{ -\lambda(y_k - y) + \frac{\sigma}{2} \varepsilon_k^m \right\}$$

where  $(y_k, \varepsilon_k^m)$  is the enumeration of a Poisson point process on  $\mathcal{Y} \times \mathbb{R}$  of intensity  $dy \times e^{-\varepsilon} d\varepsilon$ , and

$$\lambda(z) = 0 \text{ if } z = 0 \text{ and } \lambda(z) = +\infty \text{ otherwise.}$$

Let us provide an interpretation of this setup. The set of elements  $(y_k, \varepsilon_k^m)$  represents the network of “acquaintances” of  $m$ ; this is a set of partners  $y_k$  of man  $m$ , along with the corresponding sympathy shocks  $\varepsilon_k^m$ . These are drawn randomly according to a Poisson point process. The set of acquaintances of  $m$  is infinite, discrete, and dense in  $\mathcal{Y}$ . The specification of  $\lambda$  implies that the sympathy shock for any potential partners outside of his network of acquaintances is minus infinity:  $m$  only considers potential partners from his network of acquaintances.

Similarly, and with the same interpretation, the utility of a woman  $w$  of attributes  $y_w = y$  who marries a man of attributes  $x$  is

$$\gamma(x, y) - \tau + \eta_w(x),$$

with

$$\eta_w(x) = \max_l \left\{ -\lambda(x_l - x) + \frac{\sigma}{2} \eta_l^w \right\}$$

where  $(x_l, \eta_l^w)$  is the enumeration of a Poisson point process on  $\mathcal{X} \times \mathbb{R}$  of intensity  $dx \times e^{-\eta} d\eta$ .

We now give the main result of this section, which extends Galichon and Salanié (2010) to the continuous case.

**Theorem 1.** *The following holds:*

(i) *Under the assumptions stated above, the optimal matching  $\pi$  maximizes the social gain*

$$\iint_{\mathcal{X} \times \mathcal{Y}} \Phi(x, y) \pi(x, y) dx dy + \mathcal{E}(\pi) \quad (2.4)$$

over  $\pi \in \mathcal{M}(P, Q)$ , where  $\mathcal{E}$  is given by

$$\mathcal{E}(\pi) = -\sigma \iint_{\mathcal{X} \times \mathcal{Y}} \log \left( \frac{\pi(x, y)}{\sqrt{f(x)g(y)}} \right) \pi(x, y) dx dy.$$

(ii) *In equilibrium, for any  $x \in \mathcal{X}, y \in \mathcal{Y}$*

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

where the potentials  $a(x)$  and  $b(y)$  are determined such that  $\pi \in \mathcal{M}(P, Q)$ . They exist and are uniquely determined up to a constant.

(iii) *A man  $m$  of attributes  $x$  who marries a woman of attributes  $y$  obtains utility*

$$U(x, y) + \varepsilon_m(y) = \max_{z \in \mathcal{Y}} (U(x, z) + \varepsilon_m(z)) \quad (2.6)$$

where

$$U(x, y) = \frac{\Phi(x, y) + a(x) - b(y)}{2}. \quad (2.7)$$

Similarly, a woman  $w$  of attributes  $y$  who marries a man of attributes  $x$  obtains utility

$$V(x, y) + \eta_w(y) = \max_{z \in \mathcal{X}} (U(z, y) + \eta_w(z)) \quad (2.8)$$

where

$$V(x, y) = \frac{\Phi(x, y) - a(x) + b(y)}{2}. \quad (2.9)$$

As in Galichon and Salanié (2010), and independently, Decker et al. (2012), part (i) of this result expresses the fact that the optimal matching reflects a trade-off between sorting on the observed characteristics, which tends to maximize the term  $\int \Phi(x, y) \pi(x, y) dx dy$ , and sorting on the unobserved characteristics, which tends to maximize the entropic term  $\mathcal{E}(\pi)$ , i.e. the relative entropy of  $\pi(x, y)$  with respect to  $f(x)g(y)$ , which is the “random

matching”, or independent coupling of  $P$  and  $Q$ . The second term will therefore “pull” the solution towards the random matching; the parameter  $\sigma$ , which measures the intensity of the unobserved heterogeneity, measures the intensity of this trade-off. The smaller  $\sigma$ , the less unobserved heterogeneity in the model, and the closer the solution will be to the solution without heterogeneity. On the contrary, the higher  $\sigma$ , the larger the probabilistic independence between the characteristics of men and women.

Part (ii) of the result is an expression of the first order optimality conditions. The problem is an infinite dimensional linear programming problem;  $a(x)$  and  $b(y)$  are the Lagrange multipliers corresponding to the constraints  $\int \pi(x, y) dy = f(x)$  and  $\int \pi(x, y) dx = g(y)$  respectively. Taking the logarithm of Equation (2.5) yields

$$\log \pi(x, y) = \frac{\Phi(x, y) - a(x) - b(y)}{\sigma} \quad (2.10)$$

which will be the basis of our estimation strategy.

Along with the constraint  $\pi \in \mathcal{M}(P, Q)$ , equation (2.10) can be seen as a nonlinear equation in  $a(\cdot)$  and  $b(\cdot)$ . It is known in the applied mathematical literature as the Schrödinger-Bernstein equation. Existence and unicity (up to a constant) is well studied under very general conditions on  $P$  and  $Q$ , see for instance Ruschendorf and Thomsen (1998) and references therein.

From an identification perspective, equation (2.10) implies that  $\pi(x, y)$  allows to identify  $\Phi(x, y)/\sigma$  up to a separatively additive function: as we do not observe singles and assume their reservation utilities to be minus infinity,  $\Phi(x, y)/\sigma$  is observationally indistinguishable from  $\Phi(x, y)/\sigma + \mu(x) + \nu(y)$ .

Part (iii) of the result explains how the surplus is shared at equilibrium. Unsurprisingly, just as in Choo and Siow (2006) and Galichon and Salanié (2010; 2012), individuals do not transfer their sympathy shock at equilibrium, which is expressed by formulas (2.6) and (2.8). Formulas (2.7) and (2.9) provide the formulas for the systematic part of the surplus. As previously noted,  $a(x)$  and  $b(y)$  are the Lagrange multipliers of the scarcity constraint

of men's observable attributes  $x$  and women's attributes  $y$ . Hence a higher  $a(x)$  shall mean a higher relative scarcity for  $x$ , and hence a greater prospect for surplus extraction.

### 3. PARAMETRIC ESTIMATION

In this section we shall deal with parametric estimation of the surplus function  $\Phi$ . The technique we apply here was introduced by Galichon and Salanié (2010); we discuss here its extension to the continuous case, which does not raise particular conceptual challenges. In the remainder of the paper, we shall assume a quadratic parametrization of  $\Phi$ : for  $A$  a  $d_x \times d_y$  matrix, we take

$$\Phi_A(x, y) = x' Ay,$$

where we term the matrix  $A$  the *affinity matrix*. We need to impose some normalization on  $A$ , and hence we shall fix  $\|A\| = 1$ , where  $\|\cdot\|$  is the Froebenius norm:  $\|A\| = (\text{Tr}(A^*A))^{1/2}$ . One has

$$A_{ij} = \frac{\partial^2 \Phi(x, y)}{\partial x^i \partial y^j},$$

so  $A_{ij}$  measures the intensity of the complementarity/substitutability between attribute  $x^i$  of the man  $x$  and attribute  $y^j$  of the woman. If  $A_{ij} > 0$ ,  $x^i$  and  $y^j$  are complementary, and (all things else being equal) high  $x^i$  tend to match with high  $y^j$ . For instance, the education level of one of the partners may be complementary with the risk aversion of the other partner. On the contrary, if  $A_{ij} < 0$ , then  $x^i$  and  $y^j$  are substitutable.

Note that attributes should not be interpreted as a positive quality (where a greater value of  $x^i$ , the  $i$ -th dimension of  $x$ , would be more socially desirable than a smaller value of  $x^i$ ) as is sometimes done in the literature. The model above is indeed observationally indistinguishable from the same model but where  $x$  is replaced by  $-x$  and  $y$  by  $-y$ . Instead,  $x$  and  $y$  account for the strength of mutual attractiveness on various dimensions.

In order to estimate  $A$ , introduce the *cross-covariance matrix*

$$\Sigma_{XY} = (\mathbb{E}_\pi [X_i Y_j])_{ij} = \mathbb{E}_\pi [XY'] \tag{3.1}$$

which is computed at the optimal  $\pi$  solution of (2.2) and note that the optimal matching  $\pi$  maximizes the social gain

$$\mathcal{W}(A) = \max_{\pi \in \mathcal{M}(P, Q)} \mathbb{E}_\pi \left[ \frac{\Phi_A(X, Y)}{\sigma} \right] - \mathbb{E}_\pi [\ln \pi(X, Y)].$$

By the envelope theorem

$$(\Sigma_{XY})_{ij} = \frac{\partial \mathcal{W}}{\partial A_{ij}} \left( \frac{A}{\sigma} \right). \quad (3.2)$$

In a setting with discrete observables, Galichon and Salanié (2012) showed that  $B = A/\sigma$  is identified as a solution to the following optimization program

$$\min_{B \in \mathcal{M}_{d_x, d_y}(\mathbb{R})} \{ \mathcal{W}(B) - \text{Tr}(B' \Sigma_{XY}) \} \quad (3.3)$$

whose first-order conditions are precisely (3.2). In the present setting with continuous observables, things work in an exactly similar fashion. Using the normalization  $\|A\| = 1$ ,  $A$  and  $\sigma$  can then be found by

$$A = \frac{B}{\|B\|}, \quad \sigma = \frac{1}{\|B\|}.$$

Note that if needed, this way of identifying  $A$  extends by continuity to the case  $\sigma = 0$ . Let us denote  $A^{XY}$  the (unique) solution to this problem, and call it *affinity matrix*. Intuitively,  $A_{ij}^{XY}$  indicates the marginal change in a couple's surplus when increasing the man's  $i^{\text{th}}$  attribute and woman's  $j^{\text{th}}$  attribute by 1 unit. This parameter  $A^{XY}$  is “dual” to  $\Sigma_{XY}$  in the sense that Equation (3.2) is invertible, but unlike  $\Sigma_{XY}$ , it has a structural interpretation: it is the vector of weights of interactions of the various attributes.

Once the affinity matrix  $A^{XY}$  has been estimated, two questions arise. First, what is the rank of  $A^{XY}$ ? This question is of importance since one would like to know on how many dimensions of  $x$  and  $y$  the sorting of men and women occurs. Second, how can we construct “indices of mutual attractiveness” such that each pair of indices for men and women explains a mutually exclusive part of the surplus of matches?

Many papers in the literature resort to a technique called “Canonical Correlation,” which essentially relies on a singular value decomposition of  $\Sigma^{XY}$ . In Dupuy and Galichon (2012), we argue that this technique is not well suited for studying assortative matching, and that the resulting procedure is inconsistent. Instead we propose in Section 4 a method we

call ‘‘Saliency Analysis’’ in order to correctly answer these two questions. This method is essentially based on the singular value decomposition of the affinity matrix  $A^{XY}$  (instead of  $\Sigma^{XY}$  as in Canonical Correlation). Testing the rank of the affinity matrix is equivalent to testing the number of (potentially multiple) singular values different from 0. Performing this decomposition allows one to construct ‘‘indices of mutual attractiveness’’ that each explain a separate share of the surplus.

#### 4. SALIENCY ANALYSIS

In this section we set out to determine what is the rank of the affinity matrix  $A^{XY}$ , and what are the principal dimensions in which it operates. For this we introduce and describe a novel technique we call *Saliency Analysis*, which is close to Canonical Correlation in spirit but does not suffer the pitfalls of the latter. Instead of performing a singular value decomposition of the (renormalized) cross-covariance matrix  $\Sigma_{XY}$ , we shall perform a singular value decomposition of the renormalized affinity matrix  $A^{XY}$ .

Recall that we have defined the cross-covariance matrix  $\Sigma_{XY} = \mathbb{E}_\pi [XY']$ , and introduce

$$\Sigma_X = \mathbb{E}_\pi [XX'], \quad \Sigma_Y = \mathbb{E}_\pi [YY'] .$$

We shall in fact work with the rescaled attributes  $\Sigma_X^{-1/2}X$  and  $\Sigma_Y^{-1/2}Y$ . For this, we need a formula which expresses the affinity matrix of the rescaled attributes as a function of the affinity matrix between  $X$  and  $Y$ . Hence we need the following lemma:

**Lemma 1.** *For  $M$  and  $N$  two invertible matrices, one has:*

$$A^{MX,NY} = (M')^{-1} A^{XY} N^{-1}. \tag{4.1}$$

Applying Lemma 1 with  $M = \Sigma_X^{-1/2}$  and  $N = \Sigma_Y^{-1/2}$ , the affinity matrix of the rescaled attributes is

$$\Theta = \Sigma_X^{1/2} A^{XY} \Sigma_Y^{1/2},$$

whose singular value decomposition yields

$$\Theta = U' \Lambda V,$$

where  $\Lambda$  is a diagonal matrix with nonincreasing elements  $(\lambda_1, \dots, \lambda_d)$ ,  $d = \min(d_x, d_y)$  and  $U$  and  $V$  are orthogonal matrices. Define the vectors of *indices of mutual attractiveness*

$$\tilde{X} = U\Sigma_X^{-1/2}X \text{ and } \tilde{Y} = V\Sigma_Y^{-1/2}Y$$

and let  $A^{\tilde{X}\tilde{Y}}$  be the affinity matrix on the rescaled vectors of characteristics  $\tilde{X}$  and  $\tilde{Y}$ . From Lemma 1, it follows that  $A^{\tilde{X}\tilde{Y}} = \Lambda$ . In other words, there are no cross-complementarities (or substitutabilities) between the indices  $\tilde{x}_i$  and  $\tilde{y}_j$  for  $i \neq j$ . This justifies the terminology chosen:  $\tilde{x}_i$  and  $\tilde{y}_i$  are “mutually attractive” because they are complementary with each other and only with each other. All things being equal, a man with a higher  $\tilde{x}_i$  tends to match with a woman with a higher  $\tilde{y}_i$ .

The weights of each index of mutual attractiveness constructed by Saliency Analysis are given by the associated row of  $U\Sigma_X^{-1/2}$  for men and  $V\Sigma_Y^{-1/2}$  for women. The value  $\lambda_i/(\sum_i \lambda_i)$  indicates the share of the observable surplus of couples explained by the  $i^{th}$  pair of indices since by construction we have

$$\Phi_A(x, y) = x' Ay = \tilde{x}' \Lambda \tilde{y}.$$

In particular, this implies that testing for multidimensional sorting versus unidimensional sorting is equivalent to testing whether at least two singular values  $\lambda_i$  are significantly larger than 0.

## 5. INFERRING THE NUMBER OF SORTING DIMENSIONS<sup>5</sup>

Assume that a finite sample of size  $n$  is observed. The vector of mutual attraction weights estimated on the sample is denoted  $\hat{\Lambda}$ <sup>6</sup>, while the vector of mutual attraction weights in the population is denoted  $\Lambda$ . Similarly  $\hat{A}$  is the estimator of  $A$ . Let  $\hat{\Sigma}_X$ ,  $\hat{\Sigma}_Y$  and  $\hat{\Sigma}_{XY}$  be the sample estimators of  $\Sigma_X$ ,  $\Sigma_Y$  and  $\Sigma_{XY}$ , respectively. For a given quantity  $M$ , we shall

---

<sup>5</sup>In a discussion with one of the authors, Jim Heckman suggested the intuition of the approach proposed in this paper to test for multidimensional sorting.

<sup>6</sup>Dependence in  $n$  is omitted in the notation.



denote  $\delta M$  the difference between  $\hat{M}$ , the estimator of  $M$  and  $M$ , that is

$$\delta M = \hat{M} - M.$$

In the sequel, we shall represent linear operators acting on matrices as tensors, for which we shall use the bold notation to distinguish them from matrices, and we shall treat matrices as vectors, an operation which is called *vectorization* in matrix algebra. If  $\mathbb{T}$  is a tensor  $\mathbb{T}_{kl}^{ij}$ , then  $\mathbb{T} \cdot M$  will denote the matrix  $N$  such that  $N^{ij} = \sum_{kl} \mathbb{T}_{kl}^{ij} M^{kl}$ . We recall the definition of the Kronecker product: for two matrices  $A$  and  $B$ ,  $A \otimes B$  is the tensor  $\mathbb{T}$  such that

$$\mathbb{T}_{kl}^{ij} = A_{ik} B_{jl}.$$

Define two important tensors associated to the large sample properties of the model. The Fisher Information matrix is defined by

$$\mathbb{F}_{kl}^{ij} = \mathbb{E}_\pi \left[ \frac{\partial \log \pi(X, Y)}{\partial A_{ij}} \frac{\partial \log \pi(X, Y)}{\partial A_{kl}} \right],$$

and the variance-covariance matrix of the quadratic moments of  $(X, Y)$  is defined as

$$\begin{aligned} (\mathbb{K}_{XX})_{ij}^{kl} &= \text{cov}_\pi(X^i X^j, X^k X^l), \quad (\mathbb{K}_{YY})_{ij}^{kl} = \text{cov}_\pi(Y^i Y^j, Y^k Y^l), \\ \text{and } (\mathbb{K}_{XY})_{ij}^{kl} &= \text{cov}_\pi(X^i Y^j, X^k Y^l), \end{aligned}$$

so that the following Theorem holds:

**Theorem 2.** *The following convergence holds in distribution for  $n \rightarrow +\infty$ :*

$$n^{1/2} \left( \hat{A} - A, \hat{\Sigma}_X - \Sigma_X, \hat{\Sigma}_Y - \Sigma_Y \right) \Longrightarrow \mathcal{N} \left( 0, \begin{pmatrix} \mathbb{F}^{-1} & 0 & 0 \\ 0 & \mathbb{K}_{XX} & \mathbb{K}_{XY} \\ 0 & \mathbb{K}'_{XY} & \mathbb{K}_{YY} \end{pmatrix} \right)$$

Next, denoting

$$\hat{\Theta} = \hat{\Sigma}_X^{1/2} \hat{A} \hat{\Sigma}_Y^{1/2},$$

whose singular value decomposition is

$$\hat{\Theta} = \hat{U}' \hat{\Lambda} \hat{V},$$

and using the fact that

$$\delta\Sigma_X^{1/2} = \left( I \otimes \Sigma_X^{1/2} + \Sigma_X^{1/2} \otimes I \right)^{-1} \delta\Sigma_X,$$

one has

$$\begin{aligned} \delta\Theta &= \left( \Sigma_Y^{1/2} \otimes \Sigma_X^{1/2} \right) \delta A + \left( \Sigma_Y^{1/2} A' \otimes I \right) \delta\Sigma_X^{1/2} + \left( I \otimes \Sigma_X^{1/2} A \right) \delta\Sigma_Y^{1/2} \\ &= \mathbb{T}_{XY} \delta A + \mathbb{T}_X \delta\Sigma_X + \mathbb{T}_Y \delta\Sigma_Y, \end{aligned}$$

where

$$\mathbb{T}_X = \left( \Sigma_Y^{1/2} A' \otimes I \right) \left( \Sigma_X^{1/2} \otimes I + I \otimes \Sigma_X^{1/2} \right)^{-1} \quad (5.1)$$

$$\mathbb{T}_{XY} = \Sigma_Y^{1/2} \otimes \Sigma_X^{1/2} \quad (5.2)$$

$$\mathbb{T}_Y = \left( I \otimes \Sigma_X^{1/2} A \right) \left( \Sigma_Y^{1/2} \otimes I + I \otimes \Sigma_Y^{1/2} \right)^{-1}, \quad (5.3)$$

and we get as a consequence:

**Corollary 1.** *As  $n$  tends to infinity,*

$$n^{1/2} \left( \hat{\Theta} - \Theta \right) \implies \mathcal{N}(0, \mathbb{V})$$

where

$$\mathbb{V} = \mathbb{T}_{XY} \mathbb{F}^{-1} \mathbb{T}'_{XY} + \mathbb{T}_X \mathbb{K}_{XX} \mathbb{T}'_X + \mathbb{T}_Y \mathbb{K}_{YY} \mathbb{T}'_Y + \mathbb{T}_X \mathbb{K}_{XY} \mathbb{T}'_Y + \mathbb{T}_Y \mathbb{K}'_{XY} \mathbb{T}_X.$$

We would like to use this asymptotic result to test the rank of the affinity matrix  $\Lambda$ . This is an important problem with a distinguished tradition in econometrics (see e.g. Robin and Smith, 2000 and references therein). Here, we use results from Kleibergen and Paap (2006). One would like to test the null hypothesis  $H_0$ : the rank of the affinity matrix is equal to  $p$ . Following Kleibergen and Paap, the singular value decomposition  $\hat{\Theta} = \hat{U}' \hat{\Lambda} \hat{V}$  is written blockwise

$$\hat{\Theta} = \begin{pmatrix} \hat{U}'_{11} & \hat{U}'_{21} \\ \hat{U}'_{12} & \hat{U}'_{22} \end{pmatrix} \begin{pmatrix} \hat{\Lambda}_1 & 0 \\ 0 & \hat{\Lambda}_2 \end{pmatrix} \begin{pmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{pmatrix}$$

where the blocks are dimensioned so that  $\hat{U}'_{11}$  and  $\hat{V}_{11}$  are two  $p \times p$  square matrices. Define

$$\begin{aligned}\hat{T}_p &= \left(\hat{U}'_{22}\hat{U}_{22}\right)^{-1/2} \hat{U}'_{22}\hat{\Lambda}_2\hat{V}_{22} \left(\hat{V}'_{22}\hat{V}_{22}\right)^{-1/2} \\ \hat{A}_{p\perp} &= \begin{pmatrix} \hat{U}'_{21} \\ \hat{U}'_{22} \end{pmatrix} \left(\hat{U}'_{22}\right)^{-1} \left(\hat{U}'_{22}\hat{U}_{22}\right)^{1/2} \\ \hat{B}_{p\perp} &= \left(\hat{V}'_{22}\hat{V}_{22}\right)^{1/2} \hat{V}_{22}^{-1} \begin{pmatrix} \hat{V}_{21} & \hat{V}_{22} \end{pmatrix}\end{aligned}$$

so that we get, as a consequence of Kleibergen and Paap, Theorem 1:

**Theorem 3.** *Under  $H_0$*

$$n^{1/2}\hat{T}_p \implies \mathcal{N}(0, \Omega_p)$$

where  $\Omega_p = \left(B_{p\perp} \otimes A'_{p\perp}\right) \mathbb{V} \left(B_{p\perp} \otimes A'_{p\perp}\right)'$ . As a result, the test-statistic

$$n\hat{T}'_p\hat{\Omega}_p^{-1}\hat{T}_p$$

converges under the null hypothesis  $H_0$ :  $\text{rank}(\Lambda) = p$  to a  $\chi^2((d_x - p)(d_y - p))$  random variable.

## 6. THE DATA

In this paper, we use data from the DNB Household Survey (DHS), in particular the waves 1993-2002, to estimate preferences in the marriage market<sup>7</sup>. This data is a representative panel of the Dutch population with respect to region, political preference, housing, income, degree of urbanization, and age of the head of the household among others. The DHS data was collected via on-line terminal sessions and each participating family was provided with a PC and a modem if necessary. The panel contains on average about 2,200 households in each wave.

This data has three main features that are particularly attractive for our purpose. First, within each household, all persons aged 16 or over were interviewed<sup>8</sup>. This implies that the data contains detailed information not only about the head of the household but about

<sup>7</sup>For a thorough description of the set up and quality of this data we refer the reader to Nyhus (1996)

<sup>8</sup>The section General Information on the Household includes all members of the household (also those under 16 years of age).

all individuals in the household. In particular, the data identifies “spouses” and “permanent partners” of the head of each household. This information reveals the nature of the relationship between the various individuals of each household and allows us to reconstruct “couples”.

Second, this data contains very detailed information about individuals. This rich set of information includes socio-demographic variables such as birth year and education, but also variables about the morphology of respondents (height and weight), a self-assessed measure of health and above all, information about personality traits and risk attitude.

Finally, as for most panel data, the DHS data suffers from attrition problems. The attrition of households is on average 25% each year<sup>9</sup>. To remedy attrition, refreshment samples were drawn each year, such that, over the period 1993-2002, about 7,700 distinct households were interviewed at least once. Since the methodology implemented in this paper relies essentially on the availability of a cross-section of households, attrition and its remedy is in fact an asset of this data as it allows us to have access to a rather large pool of potential couples.

**6.1. Variables.** Educational attainment is measured from the respondent’s reported highest level of education achieved. The respondents could choose among 13 categories<sup>10</sup>, ranging from primary to university education. We coded responses as follows:

- (1) Lower education<sup>11</sup>,
- (2) Intermediate education<sup>12</sup> and,

---

<sup>9</sup>See Das and van Soest (1999) among others.

<sup>10</sup>In the later waves, respondents could only choose among 7 categories: 1 (continued) special education, 2 kindergarten/primary education, 3 VMBO (pre-vocational education), 4 HAVO, VWO (pre-university education), 5 senior vocational training or training through apprentice system, 6 vocational colleges and 7 university education. Categories 1-2 are coded as low education, categories 3-6 are coded as intermediary education and category 7 as higher education.

<sup>11</sup>Which consists of: other education, special (low-level) education, vocational training through apprentice system, other sort of education/training, kindergarten/primary education, continued primary education [VGLO] or elementary secondary education [LAVO].

<sup>12</sup>Which consists of: continued special (low-level) education [MLK , VSO, LOM], secondary education [MAVO/MULO], pre-university education [HAVO, VWO, Atheneum, Gymnasium, HBS, MMS, Lyceum],

(3) Higher education<sup>13</sup>.

The respondents were also asked about their height and weight. From the responses to these questions we calculated the Body Mass Index of each respondent as the weight in Kg divided by the square of the height measured in meters. The respondents were also asked to report their general health. The phrasing of the question was: “How do you rate your general health condition on a scale from 1, excellent, to 5, poor?”<sup>14</sup>. We defined our measure of health by subtracting the answer to this question to 6.

The DHS panel contains three lists of items that would allow one to assess a respondent’s personality traits.

- (1) The first list contains 150 items and refers to the Five-Factor Personality Inventory measure, developed by Hendriks et al. (1999). This list was included in a supplement to the 1996 wave.
- (2) The second list refers to the 16 Personality Adjective (16PA) scale developed by Brandstätter (1988) and was included in the module “Economic and Psychological Concepts” from 1993 until 2002.
- (3) From 2003 on, the panel replaced the 16PA scale by the International Personality Item Pool (IPIP) developed by Golberger (1999). The 10-item list version of the IPIP scale is used except for the 2005 wave where the 50-item list was implemented.

Of the three scales, the 16PA scale covers the largest sample of individuals. For that reason, the 16PA scale was chosen to measure personality traits. This scale offers the respondents the opportunity to locate themselves on 16 personality dimensions. Each dimension is represented by two bipolar scales so that the full scale contains 32 items. Nyhus

---

junior vocational training [e.g. LTS, LEAO, Lagere Land- en Tuinbouwschool], senior vocational training [e.g. MTS, MEAO, Middelbare Land- en Tuinbouwschool], vocational colleges [e.g. HTS, HEAO, opleidingen MO-akten] and vocational colleges 2nd tier [e.g. accountant NIVRA, actuaris, opleidingen MO-B-akten].

<sup>13</sup>University education.

<sup>14</sup>We make use of the panel structure to deal partly with nonresponses on socio-economic and health variables. When missing values for height, weight, education, year of birth etc. were encountered, values reported in adjacent years were imputed.

and Webley (2001) showed that this scale distinguishes 5 factors<sup>15</sup>. They labelled these factors as:

- Emotional stability: a high score indicates that the person is less likely to interpret ordinary situations as threatening, and minor frustrations as hopelessly difficult,
- Extraversion (outgoing): a high score indicates that the person is more likely to need attention and social interaction,
- Conscientiousness (meticulous): a high score indicates that the person is more likely to be meticulous,
- Agreeableness (flexibility): a high score indicates that the person is more likely to be pleasant with others and go out of their way to help others and,
- Autonomy (tough-mindedness): a high score indicates that the person is more likely to be direct, rough and dominant.

Of the 32 items associated with the 16PA measure, the first half was asked in 1993, 1995 and each year between 1997 and 2002 while the other half was asked in 1994 and 1996 only. Constructing the full scale, therefore, requires losing all respondents but those who responded in two successive years between 1993 and 1996. To avoid throwing out too many observations, we constructed the five dimensions using only those 16 items included in waves 1993, 1995 and 1997-2002<sup>16</sup>. Since answers given to the same item by the same person in different waves are strongly correlated (see Nyhus and Webley, 2001), we simply collapse the data by individual using the person’s median answer to each item. We then construct our five factors by adding the items identified by Nyhus and Webley (2001) for the respective scales. In other words, “Emotional stability” is constructed using items:

- “Oriented toward reality”/“dreamer”,
- “Happy with myself”/“doubtful”,

---

<sup>15</sup>Using the 1996 wave that contains both the FFPI module and the 16PA module, Nyhus and Webley (2001) checked the correlation between the 5 factors identified by the 16PA scale and the (big) five factors identified by the FFPI. The correlation is generally high though not perfect. This suggests that both sets of factors assess slightly different aspects of the latent factors. We followed Nyhus and Webley and use a slightly less general wording for the various dimensions identified from the 16PA scale.

<sup>16</sup>See the Appendix for the list of items.

- “need to be supported”/“independent”,
- “well-balanced”/“quick-tempered”,
- “slow-thinker”/“quick thinker” and,
- “easily worried”/“not easily worried”.

“Agreeableness,” or flexibility, is constructed using items:

- “creature of habit”/“open to changes”,
- “slow thinker”/“quick thinker”,
- “quiet, calm”/“vivid and,
- vivacious”.

“Autonomy,” or tough-mindedness, is constructed based on:

- “direct, straightforward”/“diplomatic,
- “quiet, calm”/“vivid, vivacious” and,
- “shy”/“dominant”.

“Extraversion” is based on:

- “oriented towards things”/“towards people”,
- “flexible”/“stubborn” and,
- “trusting, credulous”/“suspicious”.

“Conscientiousness” is constructed using:

- “little self-control”/“disciplined”,
- “carefree”/“meticulous” and,
- “not easily hurt”/“easily hurt, sensitive”<sup>17</sup>.

---

<sup>17</sup>As a robustness check, we constructed the full scale using the 1993, 1994, 1995 and 1996 waves. We followed Nyhus and Webley (2001) and constructed the five factors using Principal Component Analysis and varimax rotation on the five main factors. The correlation between each of the factors we constructed using only 16 items and the corresponding factor using the full scale varies between 0.42 for Agreeableness and 0.76 for emotional stability.

The data also contains information about attitude toward risk<sup>18</sup>. The attitude toward risk can be assessed in the data using a list of 6 items of the type “I am prepared to take the risk to lose money, when there is also a chance to gain money..... from a scale from 1, totally disagree, to 7, totally agree”. As for the 16PA items, there is a great stability of answers by the same person over time (see Nyhus and Webley, 2001). We therefore simply collapse the data by individual using the person’s median answer to each item. We then construct an index of risk aversion by adding the three items formulated positively (aversion towards risk) and subtracting the three items formulated negatively. The Cronbach’s alpha is 0.86 which indicates the great reliability of this scale.

**6.2. Identification of “couples”.** Our definition of a couple is a man and woman living in the same household and reporting being either head of the household, spouse of the head or a permanent partner (not married) of the head. To construct our data set of couples, we first pool all the waves selected (1993-2002). We then keep only those respondents that report being head of the household, spouse of the head or a permanent (not married) partner of the head. This sample contains roughly 13,000 men and women and identifies about 7,700 unique households. We then create two data sets, one containing women and one containing men. Each data set identifies about 6,500 different men and women. We create our working dataset of couples by merging the men dataset to the women dataset using the household identifier. We identify 5,445 unique couples while roughly 1,250 men and 1,250 women remain unmatched.

Table 1 reports the number of identified couples and the number of couples for which we have complete information on the various dimensions. For nearly all couples we have information on both spouses’ educational attainment. However, out of 5,445 couples only 3,214 have complete information on education, height, health and BMI. We lose another 641 couples for which personality traits are not fully observed. Another 195 couples are lost when attitude towards risks is also taken into consideration.

---

<sup>18</sup>See the Appendix for the list of items corresponding to the attitude towards risk. The data also contains various lists of items from which measures of time preference can be constructed. However, the initial lists have been replaced in 1996 by significantly different lists. Using an homogenous measure of time preference in our analysis would imply losing about 40% of our couples.



Table 2 presents summary statistics for men and women. In our sample, on average, men are slightly more educated than women, taller by 12 centimeters, have a BMI of 1Kg/m<sup>2</sup> higher, are less conscientious (meticulous), less extravert but more emotionally stable and more risk averse. In our sample, on average, men and women have similar (good) health and a comparable degree of agreeableness and autonomy.

Oreffice and Quintana-Domeque (2010) estimate features of the (observed) matching function between men and women in the marriage market using the PSID data for the US. Their strategy consists in regressing each attribute of men on all attributes of women and vice versa. This procedure can easily be replicated with our data in an attempt to compare features of the matching function in both data sets (US versus Dutch marriage market). In doing so, interestingly enough, we find very similar results as those obtained by Oreffice and Quintana-Domeque (2010). For instance, Oreffice and Quintana-Domeque (2010) find that an additional unit in the husband's BMI is associated with a 0.4 additional unit in the wife's BMI. Using our sample, our estimate is also significant and of similar magnitude, i.e. 0.3. Also, they find that an additional inch in the husband's height is associated with an additional 0.12 inch in the wife's height. Here too, our estimate is significant and of similar magnitude, i.e. 0.18. Yet as another example, Oreffice and Quintana-Domeque (2010) find that richer men (higher educated men in our case) tend to be married with wives of lower BMI (an increase of 10% in the husband's earnings is associated with a decrease of 0.21 points in his wife's BMI). In our sample, we find that higher educated men (interpreting education as permanent income) are matched with women of lower BMI, i.e. a man with one additional level of education is matched with a woman whose BMI is 0.5 units lower.

## 7. EMPIRICAL RESULTS

We apply the Saliency Analysis outlined in the previous section on our sample of couples. The procedure requires first to estimate the affinity matrix. This is done by applying the technique presented in section 3. The estimation results are reported in Table 3. It is important to note that the estimates reported in the table are obtained using standardized variables rather than the original ones. The main advantage of using standardized variables

is that the magnitude of the coefficients is directly comparable across variables, allowing a direct interpretation in terms of comparative statics.

The estimates of the affinity matrix reveal four important and remarkable features:

- (1) First, education is the single most important factor in the marriage market. The largest coefficient of the affinity matrix is indeed observed on the diagonal for the variable education. This coefficient is twice as large as the second largest coefficient obtained on the diagonal for the variables height and BMI. In words, this means that increasing the education of both spouses by 1 standard deviation increases the couple's surplus by 0.45 units. To achieve a similar increase in surplus, either the height or the BMI of both spouses should be increased by 2 standard deviations each.
- (2) Second, the table clearly indicates the importance of interaction between the various attributes as many off-diagonal coefficients of the affinity matrix are significantly different from 0. This implies that important trade-offs take place between the various attributes. For instance, men's emotional stability interacts positively with women's conscientiousness, i.e. 0.19. Stated otherwise, this means that increasing the husband's emotional stability only increases the surplus of couples whose wives are relatively conscientious. Other examples are noticeable: men's autonomy interacts negatively with women's conscientiousness, i.e.  $-0.09$  but positively with women's extraversion, i.e. 0.09. Conversely, men's agreeableness interacts positively with women's conscientiousness, i.e. 0.10, but negatively with women's extraversion, i.e.  $-0.11$ . Men's risk aversion interacts positively with women's autonomy, i.e. 0.08.
- (3) Third, the affinity matrix is not symmetric indicating that preferences for attributes are not similar for men and women. For instance, increasing a wife's education by 1 standard deviation increases the surplus of couples with less meticulous men more (significant coefficient of magnitude  $-0.07$ ) while increasing the husband's education by 1 standard deviation has the same impact on a couple's surplus, indifferently of the wife's degree of conscientiousness, but increases the surplus of couples whose wives have lower BMI, i.e.  $-0.06$ .

- (4) Finally, personality traits matter for preferences, not only directly (terms on the diagonal are significant for conscientiousness, extraversion, autonomy and risk aversion and of respective magnitude, 0.16, 0.08, -0.10 and 0.14) but also indirectly through their interactions with other attributes. For instance, the single most important interaction between observable attributes of men and women is found between the emotional stability of husbands and the conscientiousness of women, i.e. 0.19, a magnitude that matches with the direct effect of height, BMI and health. Also, personality traits interact not only with other personality traits but also with anthropometry and education. For instance, the conscientiousness of men interacts negatively with women's education, i.e. -0.07, whereas, the agreeableness of women interacts negatively with men's BMI.

Using the estimated affinity matrix, we then proceed to the Saliency Analysis as introduced in the previous section. This enables us to i) test whether sorting is unidimensional, i.e. occurs on a single-index and ii) construct pairs of indices of mutual attractiveness for men and women.

We first test the dimensionality of the sorting in the marriage market. For  $p = 1$ , that is testing against the null hypothesis that sorting occurs on a single index, we find that  $n\hat{T}'_1\hat{\Omega}_1^{-1}\hat{T}_1 = 508.4$  which is significant at the 1% level. This implies that sorting in the marriage market does not occur on a single index as has been assumed in the previous literature. In fact, our test-statistic never becomes insignificant. Even for  $p = 9$  we have  $n\hat{T}'_9\hat{\Omega}_9^{-1}\hat{T}_9 = 7.4$  which is still significant at the 1% level. This suggests that the affinity matrix has full rank and that sorting occurs on at least 10 observed indices. Our results therefore clearly highlight that sorting in the marriage market is multidimensional and individuals face important trade-offs between the attributes of their spouses.

Each pair of indices derived from Saliency Analysis explains a mutually exclusive part of the total observable surplus of couples. The share explained by each of our 10 indices is reported in Table 4. The table shows that the share of the first 8 pairs of indices is significantly different from 0 at the 1% level.

As for the Principal Component Analysis, the labeling of each dimension is subjective and becomes increasingly difficult to interpret as one considers more dimensions. Table 5 therefore only contains the 3 pairs of indices explaining most of the surplus. Together these 3 pairs of indices explain about half of the total surplus. The first pair, indexed I1, explains about 26% of the surplus. These indices load heavily (in bold weight  $>0.5$ ) on education and the weights on education are of similar magnitude for men and women. This confirms that education plays the most important role in the sorting in the marriage market. However, the second pair of indices, which explains another 19% of the surplus, loads heavily on personality traits (i.e. conscientiousness and emotional stability for men and only conscientiousness for women). Personality traits play a very important role in the sorting in the marriage market. Interestingly enough, while conscientiousness is a factor of mutual attraction for both men and women, more so for women, emotional stability only matters on the men's side. The third pair explains another 12% of the surplus and loads on height and subjective health with similar magnitudes for men and women. This result corroborates Chiappori, Oreffice and Quintana-Domeque's (2012) finding that anthropometry and health are important for the sorting in the marriage market.

## 8. SUMMARY AND DISCUSSION

This paper has introduced a novel technique to test for the dimensionality of the sorting in the marriage market and derive indices of mutual attractiveness, namely Saliency Analysis. This technique is grounded in the structural equilibrium model of Choo and Siow (2006) which we extended to the continuous case in this paper. Indices of mutual attractiveness derived in Saliency Analysis, in contrast to Canonical Correlation for instance, have a structural interpretation and are therefore informative about agents' preferences.

Saliency Analysis has been performed on a unique dataset of Dutch households containing information about the education, height, BMI, health, attitude towards risk and five personality traits of both spouses. The empirical results of this paper show two important features of the marriage market. First, our results clearly show that sorting occurs on multiple indices rather than just a single one as assumed in the current literature. This implies that individuals face important trade-offs between the attributes of their potential

spouse. For instance, in the dataset we studied, women with lower BMI prefer men with higher education (-0.06) and lower BMI (0.21) whereas, although agreeable women do not care about a man's education (-0.01), they do prefer men with lower BMI (-0.05). This means that, although a man could increase his education in order to compensate a loss of attractiveness towards low BMI women, due to an increase in his own BMI, doing so, this man would become less attractive to more agreeable women.

Second, personality traits and attitude towards risk matter for the sorting of spouses in the marriage market. In fact, although education explains the largest share of the (observable) surplus of spouses, i.e. 26%, personality traits explain a relative large share too, i.e. 19%. Interestingly enough, different traits matter differently for men and women. For instance, women find conscientious and emotionally stable men more attractive. Yet, men prefer conscientious women but are indifferent about the emotional stability of women.

## APPENDIX A. PROOFS

In all the proofs we shall use positive homogeneity: without loss of generality one may assume  $\sigma = 1$ .

### A.1. Proof of Theorem 1.

*Proof.* (i) The first part of the argument is essentially an extension of Galichon and Salanié (2010) to the continuous case. By the results of Shapley and Shubik, extended to the continuous case by Gretsky, Ostroy and Zame, the optimal matching solves the dual transportation problem which expresses the social welfare

$$\mathcal{W} = \inf_{u_m + v_w \geq \Phi(x_m, y_m) + \varepsilon_m(y) + \eta_w(x)} \int u_m dm + \int v_w dw \quad (\text{A.1})$$

now, the constraint can be rewritten as

$$U(x, y) + V(x, y) \geq \Phi(x, y)$$

where  $U$  and  $V$  have been defined as

$$U(x, y) = \inf_m (u_m - \varepsilon_m(y)) \quad \text{and} \quad V(x, y) = \inf_w (v_w - \eta_w(x))$$

which implies that  $u_m$  and  $v_w$  can be expressed in  $U(x, y)$  and  $V(x, y)$  by

$$u_m = \sup_{y \in \mathcal{Y}} (U(x, y) + \varepsilon_m(y)) \quad \text{and} \quad v_w = \sup_{x \in \mathcal{X}} (V(x, y) + \eta_w(x)). \quad (\text{A.2})$$

Therefore, replacing  $u_m$  and  $v_w$  by their expression in  $U$  and  $V$ , (A.1) rewrites as

$$\mathcal{W} = \inf_{U(x, y) + V(x, y) \geq \Phi(x, y)} \int G_x(U(x, \cdot)) dP(x) + \int H_y(V(\cdot, y)) dQ(y) \quad (\text{A.3})$$

where

$$\begin{aligned} G_x(U(x, \cdot)) &= \mathbb{E} \left[ \sup_{y \in \mathcal{Y}} (U(x, y) + \varepsilon_m(y)) \mid x_m = x \right] \\ H_y(V(\cdot, y)) &= \mathbb{E} \left[ \sup_{x \in \mathcal{X}} (V(x, y) + \eta_w(x)) \mid y_w = y \right] \end{aligned}$$

where  $x_m$  and  $y_w$  denote respectively the vectors of attributes of man  $m$  and woman  $w$ .

Write (A.3) as a saddlepoint problem

$$\mathcal{W} = \inf_{U, V} \sup_{\pi} \left( \begin{array}{c} \int \Phi d\pi \\ + \int G_x(U(x, \cdot)) dP(x) - \int U d\pi \\ + \int H_y(V(\cdot, y)) dQ(y) - \int V d\pi \end{array} \right)$$

or in other words

$$\mathcal{W} = \sup_{\pi} \int \Phi d\pi + \mathcal{E}(\pi)$$

where

$$\begin{aligned} \mathcal{E}(\pi) &= - \sup_U \left( \int U d\pi - \int G_x(U(x, \cdot)) dP(x) \right) \\ &\quad - \sup_V \left( \int V d\pi - \int H_y(V(\cdot, y)) dQ(y) \right). \end{aligned}$$

The results derived thus far do not depend on the particular choice of stochastic processes  $\varepsilon_m(y)$  and  $\eta_w(x)$ . Now, it remains to compute  $\mathcal{E}$ , and thus  $G_x$  and  $H_y$  with the distributional assumptions made on these stochastic processes. Introduce

$$Z = \max_{y \in \mathcal{Y}} (U(x, y) + \varepsilon_m(y)),$$

so that  $G_x(U(x, \cdot)) = \mathbb{E}[Z]$ . One has for  $c \geq 0$ ,

$$\begin{aligned} \Pr(Z \leq c) &= \Pr\left(U(x, y) - \lambda(y_k - y) + \frac{\sigma}{2}\varepsilon_k^m \leq c \forall y \in \mathcal{Y}, \forall k\right) \\ &= \Pr\left(U(x, y_k) + \frac{\sigma}{2}\varepsilon_k^m \leq c \forall k\right). \end{aligned}$$

Note that this is exactly the probability that the Poisson point process  $(y_k, \varepsilon_k^m)$  has no point in  $\{(y, \varepsilon) : U(x, y) + \frac{\sigma}{2}\varepsilon > c\}$ , which is equal to

$$\exp\left(-\iint_{\mathcal{Y}_0 \times \mathbb{R}} 1\left(U(x, y) + \frac{\sigma}{2}\varepsilon > c\right) dy e^{-\varepsilon} d\varepsilon\right) = \exp\left(-\exp\left(-c - \log \int_{\mathcal{Y}} \exp \frac{U(x, y)}{\sigma/2} dy\right)\right)$$

thus conditional on  $x_m = x$ ,  $Z$  is a  $\left(\log \int_{\mathcal{Y}} \exp \frac{U(x, y)}{\sigma/2} dy, 1\right)$ -Gumbel, hence

$$G_x(U(x, \cdot)) = \log \int_{\mathcal{Y}} \exp \frac{U(x, y)}{\sigma/2} dy.$$

Now, in order to get an expression for  $\mathcal{E}(\pi)$  it remains to compute

$$\sup_{U(x, y)} \iint_{\mathcal{X} \times \mathcal{Y}} U(x, y) \pi(x, y) dx dy - \int G_x(U(x, \cdot)) f(x) dx.$$

By F.O.C.,

$$\pi(x, y) = \frac{f(x) \exp \frac{U(x, y)}{\sigma/2}}{\int_{\mathcal{Y}} \exp \frac{U(x, y)}{\sigma/2} dy}$$

which implies that the value of the problem is infinite unless  $\int \pi(x, y) dy = f(x)$ , in which case it is

$$(\sigma/2) \iint_{\mathcal{X} \times \mathcal{Y}} \pi(x, y) \log \frac{\pi(x, y)}{f(x)} dx dy.$$

A symmetric expression is obtained for the other side of the market, and finally

$$\mathcal{E}(\pi) = -\sigma \iint_{\mathcal{X} \times \mathcal{Y}} \log \frac{\pi(x, y)}{\sqrt{f(x)g(y)}} \pi(x, y) dx dy.$$

(ii) Letting

$$a(x) = \frac{-\sigma}{2} \log \frac{f(x)}{\int_{\mathcal{Y}} \exp U(x, y) dy} \text{ and } b(y) = \frac{-\sigma}{2} \log \frac{g(y)}{\int_{\mathcal{X}} \exp V(x, y) dx},$$

one has

$$\log \pi(x, y) = \frac{U(x, y) - a(x)}{\sigma/2} \text{ and } \log \pi(x, y) = \frac{V(x, y) - b(y)}{\sigma/2}$$

and by summation

$$\pi(x, y) = \exp\left(\frac{\Phi(x, y) - a(x) - b(y)}{\sigma}\right).$$

(iii) One has

$$U(x, y) = \frac{\sigma \log \pi(x, y)}{2} + a(x) = \frac{\Phi(x, y) + a(x) - b(y)}{2}$$

and similarly

$$V(x, y) = \frac{\Phi(x, y) - a(x) + b(y)}{2}.$$

By (A.2), one sees that if man  $m$  of type  $x$  marries a woman of type  $x$ , he gets surplus

$$u_m = \sup_{y' \in \mathcal{Y}} (U(x, y') + \varepsilon_m(y')) = U(x, y) + \varepsilon_m(y).$$

□

## A.2. Proof of Lemma 1.

*Proof.* Recall that every affinity matrix  $A^{XY}$  is characterized by the fact that:

$$\frac{\partial \mathcal{W}^{P,Q}}{\partial A_{ij}}(A^{XY}) = \Sigma_{XY}^{ij}. \quad (\text{A.4})$$

Let  $P_M$  (resp.  $Q_N$ ) be the distribution of  $MX$  (resp  $NY$ ). We therefore have that:

$$\frac{\partial \mathcal{W}^{P_M, Q_N}}{\partial A_{ij}}(A^{MX, NY}) = \Sigma_{MX, NY}^{ij} = M \Sigma_{XY}^{ij} N' = M \frac{\partial \mathcal{W}^{P,Q}}{\partial A_{ij}}(A^{XY}) N', \quad (\text{A.5})$$

where the second equality follows by definition and the third by using (A.4). A simple calculation shows that

$$\mathcal{W}^{P_M, Q_N}(A^{MX, NY}) = \mathcal{W}^{P,Q}(M' A^{MX, NY} N).$$

Taking the derivative with respect to  $A$ , yields

$$\frac{\partial \mathcal{W}^{P_M, Q_N}}{\partial A}(A^{MX, NY}) = M \frac{\partial \mathcal{W}^{P,Q}}{\partial A}(M' A^{MX, NY} N) N'. \quad (\text{A.6})$$

And, by comparing (A.5) and (A.6), one gets

$$\frac{\partial \mathcal{W}^{P,Q}}{\partial A}(M' A^{MX, NY} N) = \frac{\partial \mathcal{W}^{P,Q}}{\partial A}(A^{XY}).$$



From the strict convexity of  $\mathcal{W}^{P,Q}$ , we therefore have that  $M'A^{MX,NY}N = A^{XY}$ , and given that  $M$  and  $N$  are invertible, it follows that

$$A^{MX,NY} = (M')^{-1} A^{XY} N^{-1}.$$

QED. □

**A.3. Proof of Theorem 2.** The proof of the theorem will make use of the following useful facts, which directly follow from the properties of the log-likelihood and the constraints on the marginal distributions.

**Lemma 2.** *Let  $\pi_A \in \mathcal{M}(P, Q)$  be the optimal matching computed for surplus function  $\Phi_A$ . Then*

$$\mathbb{E}_\pi \left[ \frac{\partial \log \pi_A(X, Y)}{\partial A_{ij}} \frac{\partial \log \pi_A(X, Y)}{\partial A_{kl}} \right] = \mathbb{E}_\pi \left[ \frac{\partial \log \pi_A(X, Y)}{\partial A_{ij}} x^k y^l \right] = \mathbb{E}_\pi \left[ x^i y^j \frac{\partial \log \pi_A(X, Y)}{\partial A_{kl}} \right], \quad (\text{A.7})$$

and

$$\mathbb{E} \left[ \frac{\partial \log \pi_A}{\partial A_{ij}} | X = x \right] = \mathbb{E} \left[ \frac{\partial \log \pi_A}{\partial A_{ij}} | Y = y \right] = 0. \quad (\text{A.8})$$

*Proof of Theorem 2.* In the sequel, we let

$$\hat{\pi}(x, y) = \frac{1}{n} \sum \delta(x - X_k) \delta(y - Y_k)$$

be the distribution of the empirical sample under observation, and  $\pi_A$  is the optimal matching computed for surplus function  $\Phi_A$  (we shall drop the subscript  $A$  when there is no ambiguity). Recall that the (population) affinity matrix  $A$  is characterized by

$$\frac{\partial \mathcal{W}(A)}{\partial A_{ij}} = \Sigma_{XY}^{ij}$$

and its sample estimator  $\hat{A}$

$$\frac{\partial \mathcal{W}(\hat{A})}{\partial A_{ij}} = \hat{\Sigma}_{XY}^{ij}.$$

By the Delta method, we get

$$(\mathbb{F} \cdot \delta A)^{ij} = \int \frac{\partial \log \pi_A}{\partial A_{ij}} (\hat{\pi} - \pi) dx dy + o_D(n^{-1/2}) \quad (\text{A.9})$$

where  $\mathbb{F}$  is the Hessian of  $\mathcal{W}$  at  $A$ , whose expression is

$$\mathbb{F}_{kl}^{ij} = \mathbb{E}_\pi \left[ \frac{\partial \log \pi_A(X, Y)}{\partial A_{ij}} \frac{\partial \log \pi_A(X, Y)}{\partial A_{kl}} \right]$$

where  $\pi \in \mathcal{M}(P, Q)$  is the optimal matching computed for the surplus function  $\Phi_A$ . Further,

$$\begin{aligned} (\delta \Sigma_X)^{ij} &= \int x^i x^j (\hat{\pi} - \pi) dx dy + o_D(n^{-1/2}) \\ (\delta \Sigma_Y)^{ij} &= \int y^i y^j d\pi (\hat{\pi} - \pi) dx dy + o_D(n^{-1/2}) \end{aligned}$$

hence

$$\begin{aligned} \mathbb{E} \left[ (\mathbb{F} \cdot \delta A)_{ij} (\delta \Sigma_X)_{kl} \right] &= \text{cov} \left( \frac{\partial \log \pi}{\partial A_{ij}}, X^i X^j \right) \\ &= 0, \end{aligned}$$

where we have used (A.8), and similarly,  $\mathbb{E} \left[ (\delta A)_{ij} (\delta \Sigma_Y)_{kl} \right] = 0$ . This proves the asymptotic independence between  $\delta A$  and  $(\delta \Sigma_X, \delta \Sigma_Y)$ . The conclusion follows by noting that the asymptotic variance-covariance matrix of  $\delta A$  is  $\mathbb{F}^{-1}$ , and that of  $(\delta \Sigma_X, \delta \Sigma_Y)$  is

$$\begin{pmatrix} \mathbb{K}_{XX} & \mathbb{K}_{XY} \\ \mathbb{K}'_{XY} & \mathbb{K}_{YY} \end{pmatrix}.$$

□

## APPENDIX B. TABLES

TABLE 1. Number of identified couples and number of couples with complete information for various subset of variables.

	N
Identified couples	5,445
Couples with complete information on:	
Education	5,409
The above + Health, Height and BMI <sup>a</sup>	3,214
The above + Personality traits (Big 5)	2,573
The above + measure of risk aversion	2,378

Notes: The selected sample for our analysis is the one from the last row.

a: Excluding health produces exactly the same number of couples at this stage.

Source: DNB. Own calculation.

TABLE 2. Sample of couples with complete information: summary statistics by gender.

	Husbands			Wives		
	N	mean	S.E.	N	mean	S.E.
Educational level	2378	2.0	0.6	2378	1.8	0.6
Height	2378	180.8	7.2	2378	168.4	6.5
BMI	2378	24.8	2.9	2378	23.9	4.1
Health	2378	4.1	0.7	2378	4.0	0.7
Conscientiousness	2378	-0.1	0.7	2378	0.1	0.7
Extraversion	2378	-0.1	0.7	2378	0.2	0.6
Agreeableness	2378	-0.1	0.6	2378	-0.1	0.6
Emotional stability	2378	0.1	0.6	2378	-0.2	0.5
Autonomy	2378	-0.0	0.7	2378	-0.0	0.7
Risk aversion	2378	0.1	0.7	2378	-0.2	1.0

TABLE 3. Estimates of the Affinity matrix: quadratic specification (N = 2378).

	Wives	Education	Height.	BMI	Health	Consc.	Extra.	Agree.	Emotio.	Auto.	Risk
Husbands											
Education	<b>0.46</b>	0.00	<b>-0.06</b>	0.01	-0.02	0.03	-0.01	-0.03	0.04	0.01	0.01
Height	0.04	<b>0.21</b>	0.04	0.03	<b>-0.06</b>	0.03	0.02	0.00	-0.01	0.02	0.02
BMI	-0.03	0.03	<b>0.21</b>	0.01	0.03	0.00	<b>-0.05</b>	0.02	0.01	-0.02	-0.02
Health	-0.02	0.02	-0.04	<b>0.17</b>	-0.04	0.02	-0.01	0.01	-0.00	0.03	0.03
Conscientiousness	<b>-0.07</b>	-0.01	<b>0.07</b>	-0.00	<b>0.16</b>	0.05	0.04	0.06	0.01	0.01	0.01
Extraversion	0.00	-0.01	0.00	0.01	<b>-0.06</b>	<b>0.08</b>	-0.04	-0.01	0.02	<b>-0.06</b>	<b>-0.06</b>
Agreeableness	0.01	0.01	-0.06	0.02	<b>0.10</b>	<b>-0.11</b>	0.00	0.07	-0.07	-0.05	-0.05
Emotional	0.03	-0.01	0.04	<b>0.06</b>	<b>0.19</b>	0.04	0.01	-0.04	<b>0.08</b>	<b>0.05</b>	<b>0.05</b>
Autonomy	0.03	0.02	0.01	0.02	<b>-0.09</b>	<b>0.09</b>	-0.04	0.02	<b>-0.10</b>	0.03	0.03
Risk	0.03	-0.01	-0.03	-0.01	0.00	-0.02	-0.03	-0.03	<b>0.08</b>	<b>0.14</b>	<b>0.14</b>

Note: Bold coefficients are significant at the 5 percent level.

TABLE 4. Share of observed surplus explained.

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
Share of surplus explained	25.8***	18.5***	12.4***	11.0***	9.5***	7.6***	6.7***	4.8***	2.4	1.4
Standard deviation of Shares	(1.7)	(1.2)	(1.1)	(1.1)	(1.1)	(1.2)	(1.0)	(1.4)	(1.4)	(1.4)

\*\*\* significant at 1 percent

TABLE 5. Indices of attractiveness.

Attributes	I1		I2		I3	
	M	W	M	W	M	W
Education	<b>0.91</b>	<b>0.93</b>	0.15	0.13	-0.34	-0.32
Height	0.15	0.08	-0.13	-0.08	<b>0.58</b>	<b>0.60</b>
BMI	-0.24	-0.31	0.08	0.06	-0.15	-0.19
Health	0.12	0.13	-0.01	0.14	<b>0.64</b>	<b>0.64</b>
Conscientiousness	-0.23	-0.11	<b>0.58</b>	<b>0.90</b>	0.03	0.07
Extraversion	-0.00	0.02	-0.27	-0.06	-0.03	0.18
Agreeableness	0.08	-0.02	0.39	0.26	0.22	0.06
Emotional	0.05	-0.03	<b>0.63</b>	0.23	0.24	0.16
Autonomy	0.07	0.08	-0.17	0.09	0.21	-0.12
Risk	0.20	0.14	0.04	0.19	0.00	0.24
Cum. share	0.258		0.443		0.567	

\*\*\* significant at the 1 percent

Note: M means Men and W means women. Bold coefficients indicates coefficients larger than 0.5.

APPENDIX C. QUESTIONNAIRE ABOUT PERSONALITY AND ATTITUDES<sup>19</sup>*Personality traits, the 16PA scale.*

Now we would like to know how you would describe your personality. Below we have mentioned a number of personal qualities in pairs. The qualities are not always opposites. Please indicate for each pair of qualities which number would best describe your personality. If you think your personality is equally well characterized by the quality on the left as it is by the quality on the right, please choose number 4. If you really don't know, type 0 (zero).  
Scale: 1 2 3 4 5 6 7

TEG1: oriented towards things oriented towards people.

TEG2 slow thinker quick thinker.

TEG3: easily get worried not easily get worried.

TEG4: flexible, ready to adapt myself stubborn, persistent.

TEG5: quiet, calm vivid, vivacious.

TEG6: carefree meticulous.

TEG7 shy dominant.

TEG8: not easily hurt/offended sensitive, easily hurt/offended.

TEG9: trusting, credulous suspicious.

TEG10: oriented towards reality dreamer.

TEG11: direct, straightforward diplomatic, tactful.

TEG12: happy with myself doubts about myself.

TEG13: creature of habit open to changes.

TEG14: need to be supported independent, self-reliant.

TEG15: little self-control disciplined.

TEG16: well-balanced, stable irritable, quick-tempered.

---

<sup>19</sup>The following website: [http://www.centerdata.nl/en/TopMenu/Databank/DHS\\_data/Codeboeken/](http://www.centerdata.nl/en/TopMenu/Databank/DHS_data/Codeboeken/) provides a link to the complete description of the questionnaire.



*Attitude towards risk.*

The following statements concern saving and taking risks. Please indicate for each statement to what extent you agree or disagree, on the basis of your personal opinion or experience.

totally disagree                      totally agree: 1 2 3 4 5 6 7

SPAAR1: I think it is more important to have safe investments and guaranteed returns, than to take a risk to have a chance to get the highest possible returns.

SPAAR2: I would never consider investments in shares because I find this too risky.

SPAAR3: if I think an investment will be profitable, I am prepared to borrow money to make this investment.

SPAAR4: I want to be certain that my investments are safe.

SPAAR5: I get more and more convinced that I should take greater financial risks to improve my financial position.

SPAAR6: I am prepared to take the risk to lose money, when there is also a chance to gain money.

## REFERENCES

- [1] Anderberg, D. (2004): “Getting Hitched: The Equilibrium Marriage Market Behaviour of a British Cohort,” *Royal Holloway, University of London: Discussion Papers in Economics*.
- [2] Becker, G. S. (1973): “A Theory of Marriage: Part I,” *Journal of Political Economy*, 81(4), 813–46.
- [3] Becker, G. S. (1991): *A Treatise on the Family*. Harvard University Press.
- [4] Ben-Akiva, M., N. Litinas, and K. Tsunekawa (1985): “Continuous spatial choice: The continuous Logit model and distributions of trips and urban densities,” *Transportation Research Part A: General*, 19, 83–206.
- [5] Ben-Akiva, M., and T. Watanatada (1981): *Application of a continuous spatial choice logit model) in Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.
- [6] Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008): “The Economics and Psychology of Personality Traits,” *Journal of Human Resources*, 43(4).
- [7] Bowles, S., H. Gintis, and M. Osborne (2001): “The Determinants of Earnings: A Behavioral Approach,” *Journal of Economic Literature*, 39(4), 1137–1176.

- [8] Brandstatter, H. (1988): "Sixteen personality adjective scales as a substitute for the 16PF in experiments and field studies," *Zeitschrift fr Experimentelle und Angewandte Psychologie*, 35, 370–390.
- [9] Browning, M., P.-A. Chiappori, and Y. Weiss (2007): "Family Economics," Working paper.
- [10] Bruze, G. (2011): "Marriage Choices of Movie Stars: Does Spouses Education Matter?," *Journal of Human Capital*, 5(1), 1–28.
- [11] Charles, K. K., E. Hurst, and A. Killewald (2011): "Marital Sorting and Parental Wealth," *NBER Working Paper* 16748, National Bureau of Economic Research.
- [12] Chiappori, P.-A., and S. Oreffice (2008): "Birth Control and Female Empowerment: An Equilibrium Analysis," *Journal of Political Economy*, 116(1), 113–140.
- [13] Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2011): "Matching with a Handicap: The Case of Smoking in the Marriage Market," working paper.
- [14] Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012): "Fatter attraction: anthropometric and socioeconomic matching on the marriage market," *Journal of Political Economy*, forthcoming.
- [15] Chiappori, P.-A., B. Salanié, and Y. Weiss (2010): "Assortative Matching on the Marriage Market: A Structural Investigation," working paper.
- [16] Choo, E., and A. Siow (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114(1), 175–201.
- [17] Cosslett, S. (1988): "Extreme-value stochastic processes: A model of random utility maximization for a continuous choice set," Technical report, Ohio State University.
- [18] Dagsvik, J. (1994): "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes," *Econometrica*, 62, 1179–1205.
- [19] Das, M., and A. van Soest (1999): "A panel data model for subjective information on household income growth," *Journal of Economic Behavior & Organization*, 40(4), 409–426.
- [20] Decker, C., Lieb, E., McCann, R. and Stephens, B. (2012). "Unique equilibria and substitution effects in a stochastic model of the marriage market". To appear in the *Journal of Economic Theory*.
- [21] Dupuy, A., and A. Galichon (2012): "Canonical Correlation and Assortative Matching: A Remark," Working paper.
- [22] Echenique, F., Lee, S., Yenmez, B., and Shum, M. (2012): "The Revealed Preference Theory of Stable and Extremal Stable Matchings," *Econometrica*, forthcoming.
- [23] Fox, J. (2010): "Identification in Matching Games," *Quantitative Economics* 1, 203–254.
- [24] Fox, J. (2011): "Estimating Matching Games with Transfers," mimeo.
- [25] Gabaix, X., and Landier, A. (2008). "Why Has CEO Pay Increased So Much?," *Quarterly Journal of Economics*, 123, 49–100.
- [26] Galichon, A., and B. Salanié (2010): "Matching with Trade-offs: Revealed Preferences over Competing Characteristics," Technical report.

- [27] Galichon, A., and B. Salanié (2012): “Cupid’s Invisible Hand: Social Surplus and Identification in Matching Models,” Working paper.
- [28] 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.
- [29] Gretsky, N., J. Ostroy, and W. Zame (1992): “The nonatomic assignment model,” *Economic Theory*, 2(1), 103–127.
- [30] Hendriks, A. A. J., W. K. B. Hofstee, B. De Raad, and A. Angleitner (1999): “The Five-Factor Personality Inventory (FFPI),” *Personality and Individual Differences*, 27, 307–325.
- [31] Hitsch, G. J., A. Hortasu, and D. Ariely (2010): “Matching and Sorting in Online Dating,” *American Economic Review*, 100-1, 130–163.
- [32] Jacquemet, N., and J.-M. Robin (2012): “Marriage with Labor Supply,” working paper.
- [33] Kleibergen, F., and R. Paap (2006): “Generalized reduced rank tests using the singular value decomposition,” *Journal of Econometrics*, 133, 97–126.
- [34] Lundberg, S. (2010): “Personality and Marital Surplus,” *IZA Discussion Paper* 4945.
- [35] McFadden, D. (1976): “The Mathematical Theory of Demand Models,” in *Behavioral Travel-Demand Models*, ed. by P. Stopher, and A. Meyburg, pp. 305–314. Heath and Co.
- [36] Mueller, G., and E. Plug (2006): “Estimating the effect of personality on male and female earnings,” *Industrial and Labor Relations Review*, 60(1), 3–22.
- [37] Nesheim, L. (2012): “Identification in multidimensional hedonic models,” Working paper.
- [38] Nyhus, E. K. (1996): “The VSB-CentER Savings Project: Data Collection Methods, Questionnaires and Sampling Procedures,” VSB-CentER Savings Project. Progress Report.
- [39] Nyhus, E. K., and P. Webley (2001): “The role of personality in household saving and borrowing behaviour,” *European Journal of Personality*, 15, S85–S103.
- [40] Oreffice, S., and C. Quintana-Domeque (2010): “Anthropometry and socioeconomic among couples: Evidence in the United States,” *Economics & Human Biology*, 8(3), 373–384.
- [41] Resnick, S., and R. Roy (1991): “Random USC functions, Max-stable processes and continuous choice,” *Annals of Applied Probability*, 1(2), 267–292.
- [42] Robin, J.-M., and R. J. Smith (2000): “Tests of rank,” *Econometric Theory*, 16, 151–175.
- [43] Rüschemdorf, L., and W. Thomsen (1998): “Closedness of Sum Spaces and the Generalized Schrödinger Problem,” *Theory of Probability and its Applications*, 42, 483–494.
- [44] Shapley, L., and M. Shubik (1972): “The Assignment Game I: The Core,” *International Journal of Game Theory*, 1, 111–130.
- [45] Shimer, R., and L. Smith (2000): “Assortative matching and Search,” *Econometrica*, 68, 343–369.

- [46] Terviö, M. (2008): “The difference that CEO make: An Assignment Model Approach,” *American Economic Review*, 98, 642–668.
- [47] Wong, L. Y. (2003): “Structural Estimation of Marriage Models,” *Journal of Labor Economics*, 21(3), 699–728.