I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11329

Migration in China: To Work or to Wed?

Arnaud Dupuy

FEBRUARY 2018

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11329

Migration in China: To Work or to Wed?

Arnaud Dupuy
CREA, University of Luxembourg and IZA

FEBRUARY 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Migration in China: To Work or to Wed?*

This paper develops a model encompassing both Becker's matching model, and Tinbergen-Rosen's hedonic model. We study its properties and provide identification and estimation strategies. Using data on internal migration in China, we estimate the model and compute equilibrium under counter-factual alternatives to decompose the migration surplus. Our findings reveal that about 1/5 of the migration surplus of migrant women is generated in the marriage market and 3/5 in the labor market. We also find that the welfare of urban men married with a migrant wife would have been 10% lower had their migrant wives not entered the urban marriage market.

JEL Classification: D3, J21, J23, J31

Keywords: sorting in many local markets, marriage market, hedonic and matching models

Corresponding author:
Arnaud Dupuy
University of Luxembourg
162a, avenue de la Faïencerie
L-1511 Luxembourg
Luxembourg
E-mail: arnaud.dupuy@uni.lu

* We thank participants at the EEA 2014, the 2014 Conference on Optimization, Transportation, and Equilibrium in Economics at the Fields Institute, and seminars at Collegio Carlo Alberto, ECARES, Aix-Marseille University, KU Leuven, UC Louvain, CREA University of Luxembourg and THEMA-Cergy-Pontoise for valuable comments. We have benefited from numerous insightful discussions with Alfred Galichon and Lipping Zhao provided great data assistance. This project has received funding from the FNR under the CORE grant no C14/SC/8337045.

1. Introduction

Migration is an important aspect of most economies. While there are many motives for migration, work-related motives have received most of the attention in the literature. Yet, a recent emerging literature points towards marital motives as an important source of migration. For instance, Weiss, Yi and Zhang (2013) explains that distributional imbalances across regions, such as the economic gap between Hong Kong and mainland China, can induce significant welfare effects through marriage patterns. A different argument is put forward in Gautier et al. (2010), who argue that cities can act as more efficient marriage markets because of their density and hence attract migration from rural areas.

Many traits matter in the sorting of men and women on the marriage market (see Dupuy and Galichon, 2014 for instance). These traits are not only important for determining one's outcome on the marriage market but also one's outcome on the labor market. For instance, education affects outcome on both these markets by generating marital surplus and income. Rural men and women with higher education may have better labor market prospects in cities, contributing to higher incentives to migrate. However, the marital prospects in the city of men with higher education may be less attractive than that of women with higher education due to the sex-ratios imbalances (cf. Weiss, Yi and Zhang, 2013 and Porter, 2016).

This paper aims at quantifying the relative importance of labor and marital motives of migration. This question is of interest in countries like China, where internal rural-to-urban migration is significant. We develop a structural model encompassing both the matching model by Becker (1973) and Shapley and Shubik (1972), along with the hedonic model of Tinbergen (1956) and Rosen (1974). In our model, men and women are initially distributed over various locations indicating their birthplace. To each location corresponds a local marriage market and labor market. Individuals can choose which marriage market to enter with the option to remain single, which labor market to enter and where to live but face migration costs whenever they enter a labor market and live at a location other than their birthplace.

Migration not only induces costs, it also generates benefits in the form of better opportunities on the destination's local markets and hedonic enjoyment at destination. These costs and benefits are individual-specific, depending on the attributes of individuals and their marital status. For instance, consider individuals that migrate and remain single at destination. The net migration benefits of these single migrants must be related to entering the labor market at destination and/or higher hedonic enjoyment derived from amenities (health care, public goods etc.) at destination. We label this migration effect the "work" effect as we expect the main component to be related to labor market opportunities. Consider now migrants who enter the marriage market at their birthplace and marry with a spouse of same birthplace. The net benefits of migration for these individuals could come from two sources: from the "work" effect by entering the labor market at destination, and/or from a higher hedonic enjoyment for couples at destination.[1] We label this latter effect the "marriage hedonic enjoyment" effect. Finally, consider migrants that enter the marriage market at destination and marry a native spouse, i.e. migrants in mixed-couples. In addition to the two aforementioned effects, these migrants generate net benefits upon entering the destination's marriage market when their type is in short supply on that market. We label this third effect of migration the "marrying-up" effect. Note that the "marrying-up" effect does not only concern migrants who entered the marriage market at destination but all individuals at destination and birthplace. Indeed, the decision of a migrant man (woman) to enter the marriage market at destination rather than at birthplace affects the supply of men (women) and hence equilibrium on both marriage markets. As a result, quantifying the "marrying-up" effect requires to compare equilibrium when all individuals enter the marriage market at birthplace to equilibrium when they may enter the marriage market elsewhere.

[1]For instance in the form of better quality of schools for their children.

Our contribution. The starting point of our model is the observation that, for couples, the migration decision is a *joint* decision. In our model, couples will choose the destination that maximizes the joint satisfaction of both spouses, i.e. the sum of the man's and the woman's. Our model therefore has the flavor of a matching model, where individuals care about their partner's attributes; it boils down to a matching model when there is no joint decision to take, i.e. if migration is not an option. But our model also has the flavor of a hedonic model, as the two sides of the market jointly choose a quality characteristic; it boils down to a classical hedonic model when individuals do not care about their partners' characteristics, but only about the joint decision.

As we shall see, our setting is in fact a natural extension of both the matching and the hedonic model. This class of models has received little attention in the literature,[2] and bridging the gap between hedonic models and matching models in a unified model is the main methodological contribution of this paper. This paper characterizes stability and investigates a closely related notion of equilibrium.

We provide identification results for the marriage surplus at each location as well as the migration surplus of singles and couples. Under the assumption that the migration surplus of single migrants is merely due to better labor market opportunities at destination, these identification results allow us to separately identify the migration surplus generated in the labor market and the migration surplus due to higher hedonic enjoyment of couples at destination. We parameterize the marriage surplus of couples and the migration surplus of singles and couples and estimate the parameters using a matching moment estimator and data from the Rural Urban Migration in China (RUMiC) longitudinal dataset. We then use these estimates and simulate migration and marriage market equilibrium under alternative situations to quantify the three effects of migration: the "marrying-up" effect, the "work" effect and the "marriage

[2]See Dupuy (2010) and Quintana-Domeque (2011).

hedonic enjoyment" effect. The "marrying-up" effect is quantified by comparing the estimated indirect utility of each individual to the indirect utility obtained from a first simulation in which migrants in mixed-couples enter the local marriage market at their birthplace instead of the destination's marriage market. The "work" effect is quantified by comparing the indirect utility obtained in the first simulation with the indirect utility obtained from a second simulation in which, we further impose that the parameters of the surplus from work are set to zero, hence eliminating labor market opportunities differences across locations. Finally, the "marriage hedonic enjoyment" effect is measured by comparing the indirect utility obtained in the second simulation to the indirect utility from a third simulation where we further impose that the parameters of the "marriage hedonic enjoyment" of migrant couples are equal across locations.

This decomposition exercise therefore provides a quantification of the extent to which people migrate to work or to wed, hence answering the question raised in the title. First, for migrant women in mixed-couples, we find that roughly 1/5 of the migration surplus is due to the "marrying-up" effect, 3/5 to the "work" effect and the remaining 1/5 to the "marriage hedonic enjoyment" effect. Second, for migrant men in mixed-couples, the "marrying-up" effect is neglectable, and 4/5 of the migration surplus is due to the "work" effect, the remaining 1/5 due to the "marriage hedonic enjoyment". Third, by entering the urban marriage market rather than the rural marriage market, women in mixed-couples not only improve their welfare but actually improve the welfare of both urban men on the urban marriage market and rural women on the rural marriage market: 10% of the indirect utility of urban native men is due to the "marrying-up" effect whereas roughly 10% of the migration surplus of women in non mixed-couples can be attributed to the "marrying-up" effect.

Relation to the literature. As previously noted, the model we introduce here relates on the one hand to the classical Becker model (Becker 1973, Shapley and Shubik, 1972). Empirical analysis on the Becker model was greatly facilitated by the analysis of Choo and Siow (2006), who superimpose a discrete choice structure on the model of matching. We will show how to extend Choo and Siow's insight in our generalized setting. On the other hand, the paper connects to the literature on hedonic models, started by Tinbergen (1956), pursued by Rosen (1974), and revived by Ekeland et al. (2004).

The encompassing matching model relates to Mourifié and Siow (2014) and Jaffe and Weber (2018) who respectively study peer effects and differential meeting rates in matching markets. In these models, peer effects and meeting rates affect equilibrium matching by reducing individuals' choice set. In the encompassing model, migration plays a similar role, it affects equilibrium matching on the local marriage markets by changing individuals' choice set.

Our point of view differs from a number of papers, such as Chiappori et al. (2017), Lafortune (2013), McCann et al. (2015) and Zhang (2014), who also address the connection between the marriage market and the labor market but examine how agents invest in human capital prior to entering a matching market. We will take human capital as given and exogenous and will not investigate incentives to invest in human capital for marital or employment purposes.

There is a growing literature using natural experiments or instrumental variable approach to quantify the "marrying-up" effect of sex-ratios differences on the marriage market (Angrist, 2002, Porter, 2007, Abramitzky et al., 2011 for instance). A related literature also analyzes the effects of sex ratios differences on the labor market (Angrist, 2002, Chang and Zhang, 2012, Amuedo-Dorantes and Grossbard, 2008). We contribute to this literature by considering how variations in local marriage market conditions (sex ratios but also more generally the distribution of types of men and women) and labor market conditions affect migration and matching on the marriage market.

Our application to rural-urban migration in China also relates to work by Ying et al. (2014) about the recent increase in inter-province marriages among migrants, to Nie and Xing (2010) that document the impact of a change in the attribution of a Hukou (resident permit in China) on inter Hukou marriages and Banerjee et al. (2013) that study marriages inter casts in India.

The outline for the rest of the paper is as follows. Section 2 introduces the encompassing matching model and presents our identification, inference and computation strategy. Section 3 describes the data and section 4 discusses the parametric specification and empirical results. Section 5 summarizes and concludes.

2. Encompassing matching and hedonic models

2.1. **Equilibrium and identification.** We consider a population of men and women who differ by various characteristics including age, education (taken as exogenous), health and physical characteristics, and also birthplace. We shall model two decisions of these individuals: whom to marry and where to live. These decisions obviously are not independent.

Let $x_i \in \mathcal{X} = \mathbb{R}^{d_x}$ denote the vector of observable type of a man i, and let $y_j \in \mathcal{Y} = \mathbb{R}^{d_y}$ denote the observable type of woman j. The vectors x and y include socioeconomic and physical attributes, but also record the birthplace, as we assume that men and women were initially born and raised over various locations. Let $\mathcal{Z}$ be the (finite) set of locations. We let $Z(x)$ and $Z(y)$ denote the birthplace of a man of type x and a woman of type y, respectively.

To each location corresponds a local marriage market. Men and women can choose to enter any of the local marriage markets. A local marriage market is a place to meet, and it is assumed that agents can only meet each other if both are on the same local market and can only be present on one and only one local market at any time. Agents have the option to marry a member of the opposite sex or to remain single; the set of marital options available to men is therefore $\mathcal{Y}_0 = \mathcal{Y} \times \{0\}$, where 0 is the option of singlehood; and the set of marital options available to women is $\mathcal{X}_0 = \mathcal{X} \times \{0\}$, where 0 is singlehood again.

A household type is the specification of the observable types of the partners (or of the single member if it is a single household) and the location chosen by this household. Hence, the set of household types is $\mathcal{XYZ}_0 = \mathcal{XY}_0 \times \mathcal{Z}$ with $\mathcal{XY}_0 = (\mathcal{X} \times \mathcal{Y} \cup \mathcal{X} \times \{0\} \cup \{0\} \times \mathcal{Y})$.

We assume that if man x and woman y marry and live in location z, they enjoy respective utilities

$$\alpha(x, y, z) + t(x, y, z) + \varepsilon(x, y, z)$$

$$\gamma(x, y, z) - t(x, y, z) + \eta(x, y, z)$$

while if man x and woman y remain unmatched, they enjoy respectively payoffs

$$\alpha(x, 0, z) + \varepsilon(x, 0, z) \text{ and } \gamma(0, y, z) + \eta(0, y, z)$$

where:

- $t(x, y, z)$ is the marital utility transfer (positive or negative) in market z from a woman of type y to a man of type x. This is an endogenous quantity which is determined at equilibrium and depends deterministically on x, y and z.[3]

- $\alpha(x, y, z)$ (respectively, $\gamma(x, y, z)$) is an exogenous term that accounts for the combination of marital *and* labor market utility of the man (respectively of the woman). In a leading case, the utility of the man is simply decomposed as the sum of the marital utility $\alpha^{wed}(x, y, z)$ and the labor market utility $\alpha^{work}(x, z)$, i.e. $\alpha(x, y, z) = \alpha^{wed}(x, y, z) + \alpha^{work}(x, z)$, with a similar decomposition for the utility of the woman.

[3]As shown in Theorem 1 (i) below, it is a property of the equilibrium, given our assumptions, that transfers only depend on (z, x, y) and not on the identity of spouses (i, j).

- for every type $x \in \mathcal{X}$, $\varepsilon(x,.,.)$ (respectively, for every type $y \in \mathcal{Y}$, $\eta\left(.,y,.\right)$) is an idiosyncratic term modelled has a Gumbel random process on $\mathcal{Y}_0 \times \mathcal{Z}$ (respectively on $\mathcal{X}_0 \times \mathcal{Z}$) expressing the preference heterogeneity of men and women, respectively. Gumbel random processes are defined in Appendix A; they were introduced by Cosslett (1988) and Dagsvik (1994), and used in a matching context by Dupuy and Galichon (2014).

Without loss of generality, we normalize the systematic utility of non-migrant single individuals to zero, i.e.

$$\alpha(x,0,Z(x)) = \gamma(0,y,Z\left(y\right)) = 0.$$

This normalization allows us to refer to $\alpha(x,y,z)$ and $\gamma(x,y,z)$ in terms of surplus rather than utility.

Note that the transfer between spouses enters additively and with opposite sign in their respective surplus. As a result, transfers cancel each other out in the expression of the joint surplus to yield

$$\alpha(x,y,z) + \gamma(x,y,z) =: \Phi(x,y,z).$$

Let $f(x)$ be the mass density of men of type x and $g(y)$ the mass density of women of type y. For every $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $z \in \mathcal{Z}$, we define $\mu\left(x,y,z\right)$ as the mass density of a household of type (x,y,z). It follows that the mass density of matched men of type x reads as $\int_{\mathcal{Y}\times\mathcal{Z}} \mu(x,y,z)dydz$ which by definition has to be lower or equal than $f(x)$ the mass density of men of type x. A similar constraint holds for matched women of type y. This motivates the following definition of a feasible matching.

Definition 1. *A feasible matching in this economy is therefore*

$$\mathcal{M}(f,g) = \left\{\mu \geq 0 : \int_{\mathcal{Y}\times\mathcal{Z}} \mu(x,y,z)dydz \leq f(x) \text{ and } \int_{\mathcal{X}\times\mathcal{Z}} \mu(x,y,z)dxdz \leq g(y)\right\}.$$

Men and women choose whether to marry, with whom, and where to live, in a way which maximizes their individual surplus. A man of type x and a woman of type y

therefore choose both their partner's type and place to live, which yields respectively

$$\max_{yz\in\mathcal{Y}_0\times\mathcal{Z}}\left\{\alpha(x,y,z)+t(x,y,z)+\varepsilon(x,y,z)\right\} \tag{2.1}$$

$$\max_{xz\in\mathcal{X}_0\times\mathcal{Z}}\left\{\gamma(x,y,z)-t(x,y,z)+\eta(x,y,z)\right\}, \tag{2.2}$$

where $t(x,0,z)=t(0,y,z)=0$ for all $xz\in\mathcal{X}\times\mathcal{Z}$ and $yz\in\mathcal{Y}\times\mathcal{Z}$.

Note that given a transfer function t, the optimal choice of a man of type x is a random variable $(Y^x,Z^x)\in\mathcal{Y}_0\times\mathcal{Z}$, such that (Y^x,Z^x) is optimal for (2.1), that is

$$(Y^x,Z^x)\in\arg\max_{yz\in\mathcal{Y}_0\times\mathcal{Z}}\left\{\alpha(x,y,z)+t(x,y,z)+\varepsilon(x,y,z)\right\}$$

and similarly, the optimal choice of a woman of type y is a random variable $(X^y,Z^y)\in\mathcal{X}_0\times\mathcal{Z}$, which is optimal for (2.2), that is

$$(X^y,Z^y)\in\arg\max_{xz\in\mathcal{X}_0\times\mathcal{Z}}\left\{\gamma(x,y,z)-t(x,y,z)+\eta(x,y,z)\right\}.$$

It follows that at equilibrium, the matching μ and transfer t should be such that the distribution of (Y^x,Z^x) coincides with the conditional distribution of men of type x whose choice is y and z under μ, that is

$$(Y^x,Z^x)\sim\mu\left(y,z|x\right):=\mu\left(x,y,z\right)/f\left(x\right).$$

Similarly, at equilibrium, the distribution of (X^y,Z^y) should coincide with the conditional distribution of women of type y whose choice is x and z under μ, that is

$$(X^y,Z^y)\sim\mu\left(x,z|y\right):=\mu\left(x,y,z\right)/g\left(y\right).$$

Definition 2. *An equilibrium outcome consists of a feasible matching $\mu\in\mathcal{M}(f,g)$ and a transfer $t\in\mathbb{R}$ such that $\mu(x,y,z)$ is both the mass density of men of type $x\in\mathcal{X}$ such that $(y,z)\in\mathcal{Y}_0\times\mathcal{Z}$ is solution to*

$$\max_{yz}\left\{\alpha(x,y,z)+t(x,y,z)+\varepsilon(x,y,z)\right\}$$

and the mass density of women $y\in\mathcal{Y}$ such that $(x,z)\in\mathcal{X}_0\times\mathcal{Z}$ is solution to

$$\max_{xz}\left\{\gamma(x,y,z)-t(x,y,z)+\eta(x,y,z)\right\}.$$

In our continuous logit setting, the density of the conditional distributions $\mu(y, z|x)$ and $\mu(x, z|y)$ are proportional to the exponential of the systematic part of the surplus, cf. Appendix A. Hence, the conditional distribution of men of type x whose choice is y and z under μ reads as

$$\mu(y, z|x) = \frac{\exp(\alpha(x, y, z) + t(x, y, z))}{\int_{\mathcal{Z}} \left(\exp \alpha(x, 0, z') + \int_{\mathcal{Y}} \exp(\alpha(x, y', z') + t(x, y', z'))\, dy'\right) dz'}, \tag{2.3}$$

whereas the conditional distribution of women of type y whose choice is x and z under μ obtains as

$$\mu(x, z|y) = \frac{\exp(\gamma(x, y, z) - t(x, y, z))}{\int_{\mathcal{Z}} \left(\exp \gamma(0, y, z') + \int_{\mathcal{X}} \exp(\gamma(x', y, z') - t(x', y, z'))\, dx'\right) dz'}. \tag{2.4}$$

Note that because of the normalization to 0 of the systematic utility of native singles, one has the log-odds ratio formula

$$\log \frac{\mu(x, y, z)}{\mu(x, 0, Z(x))} = \alpha(x, y, z) + t(x, y, z), \tag{2.5}$$

and a similar expression holds for women of type y,

$$\log \frac{\mu(x, y, z)}{\mu(0, y, Z(y))} = \gamma(x, y, z) - t(x, y, z), \tag{2.6}$$

which, after combining both expressions, yields

$$\alpha(x, y, z) + \gamma(x, y, z) := \Phi(x, y, z) = \log \frac{\mu^2(x, y, z)}{\mu(x, 0, Z(x))\, \mu(0, y, Z(y))}. \tag{2.7}$$

Appendix A further shows that the assumption that ε and η follow Gumbel random processes carries through to the distributions of the indirect utility of men of type x and women of type y, such that the expected indirect utility of a man of type $x \in \mathcal{X}$ and woman of type $y \in \mathcal{Y}$ obtain respectively as

$$\begin{aligned} G_x(t) &= \log \int_{\mathcal{Z}} \left(\exp \alpha(x, 0, z) + \int_{\mathcal{Y}} \exp(\alpha(x, y, z) + t(x, y, z))\, dy\right) dz, \\ H_y(t) &= \log \int_{\mathcal{Z}} \left(\exp \gamma(0, y, z) + \int_{\mathcal{X}} \exp(\gamma(x, y, z) - t(x, y, z))\, dx\right) dz. \end{aligned}$$

We are now ready to present the following theorem characterizing an equilibrium outcome.

Theorem 1. *(i) Outcome (μ, t) is an equilibrium outcome if and only if:*

- μ is a solution to the social planner's primal problem

$$\mathcal{W}(\Phi) := \sup_{\mu \geq 0} \int_{\mathcal{Z}} \begin{pmatrix} \int_{\mathcal{XY}} \Phi(x,y,z)\mu(x,y,z)dxdy \\ + \int_{\mathcal{X}} \alpha(x,0,z)\mu(x,0,z)dx + \int_{\mathcal{Y}} \gamma(0,y,z)\mu(0,y,z)dy \end{pmatrix} dz - \mathcal{E}(\mu) \tag{2.8}$$

where

$$\begin{aligned} \mathcal{E}(\mu) \;:&= \int_{\mathcal{Z}} \left(2 \int_{\mathcal{XY}} h(\mu(x,y,z))\, dxdy + \int_{\mathcal{X}} h(\mu(x,0,z))\, dx + \int_{\mathcal{Y}} h(\mu(0,y,z))\, dy \right) dz \\ & - \left(\int_{\mathcal{X}} h(f(x))\, dx + \int_{\mathcal{Y}} h(g(y))\, dy \right) \end{aligned}$$

where $h(x) := x \log x$, and

- $t = (t(x,y,z))$ is a minimizer of the dual problem

$$\inf_t \int_{\mathcal{X}} G_x(t) f(x) dx + \int_{\mathcal{Y}} H_y(t) g(y) dy, \tag{2.9}$$

where the infimum extends to all functions $(x,y,z) \to t(x,y,z)$ where the objective is defined.

(ii) The first order conditions yield

$$\log \frac{\mu(x,0,z)}{\mu(x,0,Z(x))} = \alpha(x,0,z) \text{ for } xz \in \mathcal{XZ}, \tag{2.10}$$

$$\log \frac{\mu(0,y,z)}{\mu(0,y,Z(y))} = \gamma(0,y,z) \text{ for } yz \in \mathcal{YZ}, \tag{2.11}$$

$$\log \frac{\mu(x,y,z)}{\mu(x,0,Z(x))} = \alpha(x,y,z) + t(x,y,z) \text{ for } xyz \in \mathcal{XYZ}, \tag{2.12}$$

$$\log \frac{\mu(x,y,z)}{\mu(0,y,Z(y))} = \gamma(x,y,z) - t(x,y,z) \text{ for } xyz \in \mathcal{XYZ}, \tag{2.13}$$

and hence

$$\Phi(x,y,z) = \log \frac{\mu^2(x,y,z)}{\mu(x,0,Z(x))\,\mu(0,y,Z(y))} \text{ for } xyz \in \mathcal{XYZ}, \tag{2.14}$$

for the primal problem and

$$\mu\left(y,z|x\right)f(x)=\mu\left(x,z|y\right)g(y), \tag{2.15}$$

for the dual problem, where $\mu\left(.,.|x\right)$ *and* $\mu\left(.,.|y\right)$ *are given by (2.3) and (2.4).*

(iii) At equilibrium, each man of type x *is in a household* (x,y,z) *that maximizes his surplus* $\alpha(x,y,z)+t(x,y,z)+\varepsilon(x,y,z)$*. Similarly, each woman of type* y *is in a household* (x,y,z) *that maximizes her surplus* $\gamma(x,y,z)-t(x,y,z)+\eta(x,y,z)$*.*

Theorem (1) motivates the following remarks.

Remark 1 (Entropic penalization)**.** *The unobserved heterogeneity on both sides of the market manifests itself by the entropic penalization* $\mathcal{E}\left(\mu\right)$ *in the social planner's problem. In the absence of unobserved heterogeneity, the social planner's problem would consists in pairing men and women and allocating them to locations so as to maximize the total observable surplus*

$$\int_{\mathcal{Z}}\left(\int_{\mathcal{XY}}\Phi(x,y,z)\mu(x,y,z)dxdy+\int_{\mathcal{X}}\alpha(x,0,z)\mu(x,0,z)dx+\int_{\mathcal{Y}}\gamma(0,y,z)\mu(0,y,z)dy\right)dz$$

subject to feasibility constraints $\mu\in\mathcal{M}(f,g)$*. The presence of unobserved heterogeneity adds an entropic penalization to this objective function.*

Remark 2 (Marginal constraints in primal problem)**.** *One may wonder why the marginal constraints* $\mu\in\mathcal{M}(f,g)$ *do not appear in (2.8). In fact,* $\mathcal{E}\left(\mu\right)=+\infty$ *as soon as any of these constraints are not met, so the penalization by* $\mathcal{E}\left(\mu\right)$ *automatically implies the margin constraints. There is no need for budget constraints in (2.8) as they are already taken care of by* $\mathcal{E}\left(\mu\right)$*.*

Remark 3 (Expected utilities)**.** *Using Eqs. (2.10-2.13), the expected indirect utility of a man of type* $x\in\mathcal{X}$ *and woman of type* $y\in\mathcal{Y}$ *obtain respectively as*

$$G_x(t) = -\log\frac{\mu\left(x,0,Z(x)\right)}{f\left(x\right)}, \tag{2.16}$$

$$H_y(t) = -\log\frac{\mu\left(0,y,Z(y)\right)}{g\left(y\right)}. \tag{2.17}$$

2.2. **Identification.** We consider identification of migration surplus using a single cross-section of data. The data is assumed to contain information about single men and women as well as couples, both natives and migrants. Recall that $Z(x)$ (resp. $Z(y)$) indicates the birthplace of a (wo-)man of type x (resp. y), whereas z denotes his (her) residence. The main objects of interest are:

- $\alpha(x,0,z)$, the surplus of a migrant single man of type x living at location $z \neq Z(x)$,
- $\gamma(0,y,z)$, the surplus of a migrant single woman of type y living at location $z \neq Z(y)$ and,
- $\Phi(x,y,z)$, the joint surplus of a migrant couple (x,y) living at location z such that $z \neq Z(x)$ and $z \neq Z(y)$.

Using Equations (2.10), (2.11) and (2.14) in Theorem (1), one obtains immediate identification results for these objects:

(1) The mass of native single men of type x, $\mu(x,0,Z(x))$, and the mass of migrant single men of type x migrating from $Z(x)$ to location z, $\mu(x,0,z)$, identify the surplus of migration from $Z(x)$ to location z for single men of type x as

$$\alpha(x,0,z) = \log \frac{\mu(x,0,z)}{\mu(x,0,Z(x))}. \tag{2.18}$$

(2) The mass of native single women of type y, $\mu(0,y,Z(y))$, and the mass of migrant single women of type y migrating from $Z(y)$ to location z, $\mu(0,y,z)$, identify the surplus of migration from $Z(y)$ to location z for single women of type y as

$$\gamma(0,y,z) = \log \frac{\mu(0,y,z)}{\mu(0,y,Z(y))}. \tag{2.19}$$

(3) The mass of native single men of type x, $\mu(x,0,Z(x))$, the mass of native single women of type y, $\mu(0,y,Z(y))$, and the mass of couples (x,y), born at location $Z(x)$ and $Z(y)$ and residing at location z, identify the joint migration

surplus of couples (x, y) at z as

$$\Phi(x,y,z) = \log \frac{\mu^2(x,y,z)}{\mu(x,0,Z(x))\,\mu(0,y,Z(y))} \text{ for } xyz \in \mathcal{XYZ}. \tag{2.20}$$

Note that applying Equation (2.20) to native couples of location z, $z = Z(x) = Z(y)$, allows one to identify the marriage surplus for a couple (x, y). This particular case boils down to Choo and Siow's (2006) result obtained in the context of a closed matching market.

Using the three identification results above, we now proceed to the separate identification of the systematic migration surplus from work and the "marriage hedonic enjoyment" surplus of married men and women. To this aim, we further decompose men and women's utilities into additive terms capturing the surplus from work and the marriage hedonic enjoyment surplus as follows

$$\begin{aligned}\alpha(x,y,z) &= \alpha^{wed}(x,y,z) + \alpha^{work}(x,y,z),\\ \gamma(x,y,z) &= \gamma^{wed}(x,y,z) + \gamma^{work}(x,y,z),\end{aligned}$$

for all $(x, y) \in \mathcal{XY}_0$, and $(Z(x), Z(y), z) \in \mathcal{Z}^3$ and defining $\Phi^w(.,.,.) := \alpha^w(.,.,.) + \gamma^w(.,.,.)$ for $w = \{wed, work\}$.

By definition, the surplus of singles is merely related to work so that one has

$$\begin{aligned}\alpha^{work}(x,0,z) &: = \alpha(x,0,z) \text{ for all } x \in \mathcal{X}\\ \gamma^{work}(0,y,z) &: = \gamma(0,y,z) \text{ for all } y \in \mathcal{Y}.\end{aligned}$$

Moreover, in the absence of a marriage hedonic enjoyment surplus, the joint surplus of a man x and a woman y residing at z should be the same irrespective of their marital status. This suggests using the following definition for the surplus from work:

Definition 3. *The surplus from work of an individual is the part of her surplus that is independent of her marital status. In other words, the surplus from work of a* *(wo)man x (resp. y) living at z equates that of a single (wo)man of type x (resp. y). Formally, one therefore has*

$$\begin{aligned}\alpha^{work}(x,y,z) &= \alpha^{work}(x,z)=\alpha(x,0,z),\\ \gamma^{work}(x,y,z) &= \gamma^{work}(y,z)=\gamma(0,y,z),\end{aligned}$$

for all $(x,y)\in\mathcal{XY}$, and $(Z(x),Z(y),z)\in\mathcal{Z}^3$.

Definition 3 suggests two remarks.

Remark 4 (Men's marital wage-premium)**.** *An extensive literature (e.g. Korenman and Neumark, 1991 and for China, Wang, 2013) provides evidence that married (cohabiting) men earn higher wages than single men, due to, for instance, a decrease in housework time after marriage. Definition 3 indicates that these marital wage-premia will be attributed to the "marriage hedonic enjoyment" surplus of husbands.*

Remark 5 (Marriage hedonic enjoyment)**.** *Men and women may enjoy different life styles at different locations. This may occur when, for instance, different locations offer different amenities (i.e. public goods such as theaters, opera house etc.). For singles, by definition, the hedonic enjoyment at each location will be attributed to the surplus from work. Hence, the surplus from work should be understood as a residual surplus, i.e. the surplus not related to marital status. We label this surplus, the surplus from work as we expect its main component to be related to one's opportunities on the labor market. Couples may experience an additional hedonic enjoyment corresponding to, for instance, the additional pleasure to share common cultural activities or school quality for their children. Following Definition 3, this additional surplus is what we label "marriage hedonic enjoyment" of couples as this hedonic enjoyment clearly relates to the marital status.*

It is assumed that one can fix the non-geographical type of individuals, while varying their birthplace. This allows one to compare (wo)men born at different locations but otherwise observationally similar. We adapt the previous notation to make this dependence explicit: let $x = (z, \widetilde{x})$ where $z = Z(x)$ indicates the birthplace and $\widetilde{x}$ the non-geographical type of a man of type x. A similar notation is used for women. To crystallize ideas, the implications of Definition 3 are illustrated in Table 1 for two locations, say z is urban and z_0 is rural. The rows of the table correspond to the possible household combinations in terms of birthplace and residence,[4] holding the non geographical type of women $\widetilde{y}$ and men $\widetilde{x}$ constant and dropping these for notational simplicity.

The identification of the systematic surplus of migration from work for husbands and wives follows directly from Definition 3 using households of types "#1" and "#2" in Eqs. (2.18) and (2.19) respectively.

The remaining types of households listed in Table 1 can be used to identify the "marriage hedonic enjoyment" for the corresponding types of couples using Equation (2.20). First, using households of type "#3" and "#7", together with Definition 3, one identifies the "marriage hedonic enjoyment" of native couples at z_0 and z respectively as

$$\begin{aligned} \Phi^{wed}\left((z_0, \widetilde{x}), (z_0, \widetilde{y}), z_0\right) &= \Phi\left((z_0, \widetilde{x}), (z_0, \widetilde{y}), z_0\right), \\ \Phi^{wed}\left((z, \widetilde{x}), (z, \widetilde{y}), z\right) &= \Phi\left((z, \widetilde{x}), (z, \widetilde{y}), z\right). \end{aligned}$$

Using households of type "#4", one identifies the "marriage hedonic enjoyment" for migrant couples from z_0 to z as

$$\Phi^{wed}\left((z_0, \widetilde{x}), (z_0, \widetilde{y}), z\right) = \Phi\left((z_0, \widetilde{x}), (z_0, \widetilde{y}), z\right) - \alpha\left((z_0, \widetilde{x}), 0, z\right) - \gamma\left(0, (z_0, \widetilde{y}), z\right).$$

Similarly, using households of type "#5", one identifies the "marriage hedonic enjoyment" for mixed-couples with a migrant husband from z_0 and native wife from

[4]There is virtually no urban-to-rural migration in China such that, given our normalization of native single utilities, only 7 non trivial possibilities remain.

z as

$$\Phi^{wed}\left(\left(z_0,\widetilde{x}\right),(z,\widetilde{y}),z\right)=\Phi\left(\left(z_0,\widetilde{x}\right),\left(z,\widetilde{y}\right),z\right)-\alpha\left(\left(z_0,\widetilde{x}\right),0,z\right).$$

Finally, using households of type "#6", one identifies the "marriage hedonic enjoyment" for mixed-couples with a migrant wife from z_0 and native husband from z as

$$\Phi^{wed}\left(\left(z,\widetilde{x}\right),(z_0,\widetilde{y}),z\right)=\Phi\left(\left(z,\widetilde{x}\right),\left(z_0,\widetilde{y}\right),z\right)-\gamma\left(0,\left(z_0,\widetilde{y}\right),z\right).$$

As a result, for any non-geographical type of couple $(\widetilde{x},\widetilde{y})$, our identification results allow not only to quantify the systematic joint (migration) surplus between any two locations, but also to decompose this surplus into the part due to work and the part due to hedonic enjoyment from marriage.

2.3. **Inference.** Assume the surplus function is linear in its parameters $\lambda\in\mathbb{R}^K$ and reads as

$$\Phi\left(x,y,z;\lambda\right)=\sum_{k=1}^{K}\lambda_k\varphi^k(x,y,z)$$

where $\left\{\varphi^k(x,y,z)\right\}_{k=1}^{K}$ are K linearly independent basis functions.

Under this parametric specification, the social welfare function can be rewritten as

$$\mathcal{W}(\lambda):=\sup_{\mu\geq 0}\int_{\mathcal{Z}}\left(\begin{array}{c}\int_{\mathcal{X}\mathcal{Y}}\Phi\left(x,y,z;\lambda\right)\mu(x,y,z)dxdy\\+\int_{\mathcal{X}}\alpha\left(x,0,z;\lambda\right)\mu(x,0,z)dx+\int_{\mathcal{Y}}\gamma\left(0,y,z;\lambda\right)\mu(0,y,z)dy\end{array}\right)dz-\mathcal{E}\left(\mu\right). \tag{2.21}$$

where $\mathcal{E}\left(\mu\right)$ is as defined in Theorem 1.

Denoting $\mu^\lambda(x,y,z)$ the equilibrium matching given parameters λ, by the envelope theorem one has

$$\frac{\partial\mathcal{W}\left(\lambda\right)}{\partial\lambda_k}=\mathbb{E}_{\mu^\lambda}\left[\varphi^k(x,y,z)\right]. \tag{2.22}$$

This suggests using a matching moment estimator that is: find λ such that the moments generated by the model match with the moments observed in the data, i.e.

$$\mathbb{E}_{\mu^{\lambda}}\left[\varphi^{k}(x,y,z)\right]=\mathbb{E}_{\hat{\mu}}\left[\varphi^{k}(x,y,z)\right] \text{ for all } k=1,...,K \tag{2.23}$$

where $\hat{\mu}$ denotes the observed matching in the data.

This matching moment estimator, denoted λ^{MM}, is then obtained as the solution to the following problem

$$\min_{\lambda}\mathcal{W}(\lambda)-\mathbb{E}_{\hat{\mu}}\left[\Phi\left(x,y,z;\lambda\right)\right], \tag{2.24}$$

since indeed, the first order conditions of this problem yield exactly Equation (2.23).

Note that $\mathcal{W}(.)$ is strictly convex in Φ and $\Phi\left(.,.,.;\lambda\right)$ is linear in λ such that the objective function in Problem (2.24) is strictly convex and hence admits a unique solution.

In our leading case, the surplus function is specified such that the parametrization satisfies Definition 3 and in particular

$$\begin{aligned}
\Phi\left(x,y,z;\lambda\right) &: =\Phi^{wed}\left(x,y,z;A\right)+\alpha^{work}\left(x,z;B\right)+\gamma^{work}\left(y,z;C\right), \qquad (2.25)\\
\Phi^{wed}\left(x,y,z;A\right) &: =\sum_{k=1}^{K_A}A_k\varphi_A^k(x,y,z),\\
\alpha^{work}\left(x,z;B\right) &: =\sum_{k=1}^{K_B}B_k\varphi_B^k(x,z),\\
\gamma^{work}\left(y,z;C\right) &: =\sum_{k=1}^{K_C}C_k\varphi_C^k(y,z),
\end{aligned}$$

such that $\lambda=(A,B,C)$ where $A\in\mathbb{R}^{K_A}$, $B\in\mathbb{R}^{K_B}$ and $C\in\mathbb{R}^{K_C}$ and where each set $\left\{\varphi_A^k(x,y,z)\right\}_{k=1}^{K_A}$, $\left\{\varphi_B^k(x,z)\right\}_{k=1}^{K_B}$ and $\left\{\varphi_C^k(y,z)\right\}_{k=1}^{K_C}$ is composed respectively of K_A, K_B and K_C basis functions, which are jointly independent.

2.4. **Computation.** Our estimation strategy consists in finding the parameter λ such that the associated moments match the observed moments in the data. This strategy requires fast computation of the equilibrium matching. We propose to extend the procedure introduced in Galichon and Salanié (2015) to the context of the encompassing model in the following way. Consider that each individual in our sample defines its own type such that there is one and only one (wo-)man of each type. Given a sample of men and women, let $\mu_{ijz} = \mu(x_i, y_j, z)$ and $\Phi^{\lambda}_{ijz} = \Phi\left(x_i, y_j, z; \lambda\right)$ for $i \in \mathcal{I}$, $j \in \mathcal{J}$ and $z \in \mathcal{Z}$, $\mu_{i0z} = \mu(x_i, 0, z)$ and $\Phi^{\lambda}_{i0z} = \Phi\left(x_i, 0, z; \lambda\right)$ for $i \in \mathcal{I}$ and $z \in \mathcal{Z}$ and $\mu_{0jz} = \mu(0, y_j, z)$ and $\Phi^{\lambda}_{0jz} = \Phi\left(0, y_j, z; \lambda\right)$ for $j \in \mathcal{J}$ and $z \in \mathcal{Z}$, such that the (sample) feasibility constraints for any matching in our sample read as

$$\begin{aligned} \sum_{i\in\mathcal{I}}\sum_{z\in\mathcal{Z}}\mu_{ijz} + \sum_{z\in\mathcal{Z}}\mu_{0jz} &= 1, \\ \sum_{j\in\mathcal{J}}\sum_{z\in\mathcal{Z}}\mu_{ijz} + \sum_{z\in\mathcal{Z}}\mu_{i0z} &= 1. \end{aligned} \tag{2.26}$$

The equilibrium matching given values of λ is then computed by first noting that, from Theorem 1, in equilibrium one has

$$\mu_{ijz} = K_{ijz}\sqrt{\mu_{i0}\mu_{0j}},$$

$$\mu_{i0z} = L_{iz}\mu_{i0},$$

and

$$\mu_{0jz} = P_{jz}\mu_{0j},$$

where $\mu_{i0Z(x_i)} = \mu_{i0}$, and $\mu_{0jZ(y_j)} = \mu_{0j}$, and $K_{ijz} = \exp\left(\frac{\Phi^{\lambda}_{ijz}}{2}\right)$, $L_{iz} = \exp\left(\Phi^{\lambda}_{i0z}\right)$, and $P_{jz} = \exp\left(\Phi^{\lambda}_{0jz}\right)$.

The matching μ defined by these equilibrium relations should also satisfy the feasibility constraints in Equation (2.26), a condition that yields the following system of equations

$$\begin{cases} \sqrt{\mu_{i0}}\sum_{j\in\mathcal{J},z\in\mathcal{Z}} K_{ijz}\sqrt{\mu_{0j}} + \sum_{z\in\mathcal{Z}} L_{iz}\mu_{i0} = 1 \\ \sqrt{\mu_{0j}}\sum_{i\in\mathcal{I},z\in\mathcal{Z}} K_{ijz}\sqrt{\mu_{i0}} + \sum_{z\in\mathcal{Z}} P_{jz}\mu_{0j} = 1 \end{cases}.$$

A bit of algebra provides a solution for μ_{i0} as a function of the vector (μ_{0j}) and conversely, a solution for μ_{0j} as a function of the vector (μ_{i0}) that read as

$$\begin{cases} \mu_{i0} = \left(\sqrt{\frac{1}{\sum_{z \in \mathcal{Z}} L_{iz}} + \left(\frac{\sum_{j \in \mathcal{J}, z \in \mathcal{Z}} K_{ijz} \sqrt{\mu_{0j}}}{2 \sum_{z \in \mathcal{Z}} L_{iz}} \right)^2} - \frac{\sum_{j \in \mathcal{J}, z \in \mathcal{Z}} K_{ijz} \sqrt{\mu_{0j}}}{2 \sum_{z \in \mathcal{Z}} L_{iz}} \right)^2 \\ \mu_{0j} = \left(\sqrt{\frac{1}{\sum_{z \in \mathcal{Z}} P_{jz}} + \left(\frac{\sum_{i \in \mathcal{I}, z \in \mathcal{Z}} K_{ijz} \sqrt{\mu_{i0}}}{2 \sum_{z \in \mathcal{Z}} P_{jz}} \right)^2} - \frac{\sum_{i \in \mathcal{I}, z \in \mathcal{Z}} K_{ijz} \sqrt{\mu_{i0}}}{2 \sum_{z \in \mathcal{Z}} P_{jz}} \right)^2 \end{cases} . \tag{2.27}$$

The solution $(\mu_{i0}^{\lambda}, \mu_{0j}^{\lambda})$ for this problem can be attained by the following a simple algorithm:

(1) take an initial guess of μ_{0j},
(2) update the values of μ_{i0} using the current values of μ_{0j} in the first formula of System (2.27),
(3) update the values of μ_{0j} using the current values of μ_{i0} in the second formula of System (2.27),
(4) go back to step 2 until convergence.

3. Empirical analysis

3.1. **Context.** This section provides a brief historical account of China's rural-to-urban migration; for a thorough account and a survey of the related literature, we refer the reader to Zhao (2005). Migration used to be free in China up until the Great Famine in the late 1950s, when the Chinese government implemented a permanent geographic registration of Chinese citizens, i.e. the "Hukou system", in an effort to control for rural-to-urban migration. One's Hukou ties an individual to either a rural or urban location, is determined by birth and inherited from one generation to the next. Up until the end of the 1980s, migration was extremely difficult: Changing one's Hukou was merely impossible and strict restrictions were imposed that inhibited rural households from migrating.[5] In particular, migration through marriage was very limited by the rule that children's Hukou should be determined by the mother's Hukou.

In the last three decades, the booming economy led to a sharp increase in rural-to-urban migration. At the same time, China's policy on rural migration changed gradually allowing more people to change their Hukou status. As a result, the number of rural-to-urban migrants rose rapidly, doubling between 1989 and 1993 according to some estimates (Shi, 2008). Interestingly, the rule that children's Hukou be determined by the mother's Hukou was abolished in 1998 leading to a significant increase in inter-hukou marriages (see Nie and Xing, 2012).

The Hukou system creates two types of rural-to-urban migrants. Rural migrants that succeed in changing their Hukou, the so-called *permanent migrants*, and those that do not, the *floating migrants*.[6] Changing one's Hukou from rural to urban (still today) is both very difficult, as it requires very specific conditions to be met such as holding a higher education diploma, a high military grade, or owning a house in urban area, and very important for assimilation, since it opens access to the same amenities and job market as urban people. As a result, while permanent migrants assimilate quite rapidly, floating migrants usually hold jobs in the informal sector, the so-called four "Ds": Dirt, Drain, Danger, and Disgrace, and have only limited access to urban public goods such as health care(insurance) and education.

The analysis of rural-to-urban migration in China requires to take into account this dual migration situation. In particular, the current rules and regulations in most cities in China are such that a rural born individual must decide whether or not to

[5]For instance, the provision of food coupons in urban areas was strictly limited to urban residents (holding a urban Hukou) which made it very difficult for rural people to have access to food in urban areas.

[6]See Liang and Ma (2004) for a discussion of the differences between the two groups of migrants.

migrate before knowing whether he/she will be granted a change of Hukou. Indeed, as indicated by the US Congressional-Executive Commission on China (2017):

> Migrants must still meet locally-set criteria in order to transfer their Hukou registration to a given urban area. Generally, these reforms require that rural migrants have 1) a "stable job or source of income" and 2) lived in a "stable place of residence" for over two years as conditions for obtaining local Hukou in urban areas.

Rural born individuals need therefore to form expectations about their probability of being granted a urban Hukou when deciding whether to migrate or not.

3.2. **Data.** We use data from the Rural Urban Migration in China (RUMiC) longitudinal dataset.[7] This dataset consists of three surveys ran in China, i.e. the Urban Household Survey (UHS), the Rural Household Survey (RHS), and the Migrant Household Survey (MHS), collected since 2008. The RHS and UHS have been conducted in collaboration with the National Bureau of Statistics of China (NBS), while the MHS has been conducted in partnership between a professional survey company on the one hand and scholars around the world and in particular from the Australian National University and IZA on the other hand.

The RHS comprises $8,000$ households, while the UHS and MHS each cover $5,000$ households. Using this data, we define urban (resp. rural) natives those individuals interviewed in the UHS (resp. RHS) file who possess a urban (resp. rural) Hukou and report not having changed their Hukou. A rural-to-urban floating migrant is defined as an individual interviewed in the MHS or the UHS file who, by definition, still holds a rural Hukou although lives in a city at the time of the survey.[8] In contrast,

[7]We refer the interested reader to Akgüç et. al (2013) for a detailed description of the data.

[8]Note that less than 20% of young floating migrants (aged less then 32 years old for women and 34 years old for men in 2009) report having been back to their home village for longer than 3 months after migrating to the city. Of those reporting having returned for more than 3 months to their

a permanent rural-to-urban migrant is an individual interviewed in the UHS file who reports a change of Hukou from rural to urban.

The RHS was conducted in 9 provinces: Anhui, Chongqing, Guangdong, Hebei, Henan, Hubei, Jiangsu, Sichuan, and Zhejiang. The MHS was conducted in 15 cities, either provincial capitals or cities with the largest floating migrants inflow: Bengbu, Chengdu, Chongqing, Dongguan, Guangzhou, Hefei, Hangzhou, Luoyang, Nanjing, Ningbo, Shanghai, Shenzen, Wuhan, Wuxi, Zhengzhou. The UHS was conducted in 19 cities: the 15 cities listed above but also Anyang, Jiande, Leshan and Mianyang. UHS households living in one of the 4 cities not listed in the MHS are excluded from the analysis.

Our methodology must be applied on a representative sample of households for the provinces and cities covered in the three samples. Although the MHS, RHS and UHS surveys are representative of the respective targeted populations,[9] their relative size does not match with the actual relative proportions of these populations. Following Song and Yue (2011), we account for this issue by constructing weights for the respective populations using the 2005 mini census. In the 2005 1% mini census of households, there were about 982,000 rural households in the 9 provinces covered by the RHS survey and, about 832,000 urban households (including permanent migrants) and 121,000 floating migrant households in the 15 cities covered by the MHS (and restricted UHS). This implies that a representative sample of households for the

village, about 40% did so for unforseen reasons: lost their job in the city, got sick, had to take care of sick family member, found it hard to earn money or secure a suitable job. Therefore, the bulk of the floating migrants in our sample have long term migration plans.

[9]The RHS and UHS are subsamples drawn from the national household survey of the NBS, whereas the MHS survey was conducted separately drawing migrant households at their work place to account for the fact that many migrants live in dormitories adjacent to their workplace. For more details about the sampling procedure see the RUMiCI Project's homepage, http://rse.anu.edu.au/research-projects/rural-urban-migration-in-china-and-indonesia/, or Kong (2010).

provinces and cities of interest should contain 50.7% of rural households, 43.0% of urban and permanent migrant households and 6.3% of floating migrant households. Using the UHS sample size as the benchmark, i.e. about 4, 700 households, the size of our representative sample should therefore be about 11, 000 households; 5, 500 rural, 700 floating migrant and 4, 700 urban and permanent migrant households. We therefore randomly select 5, 500 rural households from the RHS files and 700 floating migrant households from the MHS files.

3.3. **Couples and singles.** In the RHS and UHS surveys all adult persons in the household were interviewed whereas in the MHS, only those adult persons currently living with the respondent at the urban location were interviewed. Each respondent was asked about his/her marital status and could choose out of 6 categories. We use the answer to that question and consider herewith as "single" all respondents that reported being the head of the household and being either divorced, widowed or never married/single. The remaining respondents, i.e. those who either reported being married, remarried or cohabiting, are considered as being "in couple".

Our dataset of couples is then constructed as follows. For each of the three surveys, we split the sample containing respondents being "in couple" and create two datasets, one containing women and one containing men. We then merge the men dataset to the women dataset using the household identifier. We use the respondents' reported relationship to the head of their household and keep only the "correct" matches: head of household with spouse of the head, parent of the head with parent-in-law of the head, and biological child of the head with child-in-law of the head. We discard all other matches.

Following the literature (see Chiappori et al. 2017, for instance), we further restrict our sample to relatively young adults, i.e. women (reps. men) aged between 20 and 32 (resp. 24 and 34) years old. This age selection reflects legal marriageable age in

China, the fact that migration really took up at the turn of the 21st century and that one's migration decision is frequently taken in one's twenties.

Table 2 reports the number of couples, single women and single men in our working data, distinguishing between rural and urban natives and, floating and permanent migrants, both for men and women. Note that our data contains 49 mixed-couples. There are 33 mixed-couples composed of a migrant wife (17 floating and 16 permanent) and an urban native husband and 16 composed of a migrant husband (2 floating and 14 permanent) and an urban native wife. We use these couples to quantify the "marrying-up" effect.

3.4. **Selected Variables.** Educational attainment is measured using the highest level of education completed. The UHS and RHS surveys use a 9 categorical scale to classify individuals whereas in the MHS survey uses a 28 categorical scale. Nevertheless, the two scales have a similar decomposition into 5 standard educational levels: (1) primary education, (2) junior secondary education, (3) senior and specialized secondary education, (4) polytechnic college and (5) university education. We use this scale as a measure of education.

The data also contains information about body weight and height which enables use to calculate each individual's Body Mass Index (BMI), i.e. the individual's body mass in kg divided by the square of its height in meters.

The respondents were also asked to report their general health. The phrasing of the question was: "What is your current health status (compared to people your age)" and were presented with 5 alternatives:

(1) Excellent,

(2) Good,

(3) Average,

(4) Poor,

(5) Very poor.

Using answers to this question we created a health variable by subtracting the answer to the question to 5.

Our working dataset consists of those households with complete information on age, height, BMI, health and education. Table 3 reports descriptive statistics by groups (rural, urban and types of migrants, floating or permanent) and gender. The anticipated results are noticeable:

- rural men (women) have lower education than urban men (resp. women),
- permanent migrants are more educated than floating migrants which merely reflects the fact that having a University degree is one condition to obtain an urban Hukou,
- permanent migrants are older than floating migrants which is due, at least partly, to the fact that on average permanent migrants studied longer.

4. Results

4.1. **Parametric specification.** The context of internal migration in China, highlighted in Section 3.1, and the data at our disposal impose several modelling restrictions. First, our working dataset allows us to distinguish between two locations, namely rural and urban. We therefore set $\mathcal{Z} = \{0, 1\}$, where $z = 1$ for urban locations by convention. Second, the bulk of internal migration in China is rural-to-urban migration; there is virtually no urban-to-rural migration in our data. In the logit model that we are using, the only utility that is compatible with the absence of urban-to-rural migration is

$$\begin{aligned} \alpha^{work}(x, z; B) &= -\infty \text{ if } Z(x) = 1 - z = 1, \\ \gamma^{work}(y, z; C) &= -\infty \text{ if } Z(y) = 1 - z = 1. \end{aligned}$$

Third, given parameter λ, the algorithm presented in Section 2.4 computes the associated equilibrium matching. The equilibrium matching indicates for instance whether an individual migrates or not. In China, individuals do not know whether the government will grant them a change of Hukou before they move to the city. Individuals must therefore decide whether to migrate or not based on their probability of obtaining an urban Hukou and becoming a permanent migrant. For each rural born man of type x (resp. woman of type y), denote $W(x) \in [0,1]$ (resp. $W(y) \in [0,1]$ the probability that he (she) would obtain an urban Hukou and hence become a permanent migrant. To compute these probabilities, we make use of the fact that gender and education are important determinants of acquiring of an urban Hukou. In particular, we set $W(x)$ and $W(y)$ equal to the observed conditional share of permanent migrants in the data, by gender and education and obtain:

(1) $W(x) = 0.214$ ($W(y) = 0$) if he (resp. she) has primary education,
(2) $W(x) = 0.071$ ($W(y) = 0.082$) if he (resp. she) has junior secondary education,
(3) $W(x) = 0.583$ ($W(y) = 0.521$) if he (resp. she) has senior and specialized secondary education,
(4) $W(x) = 0.842$ ($W(y) = 0.868$) if he (resp. she) has graduated from a polytechnic college,
(5) $W(x) = 0.941$ ($W(y) = 0.957$) if he (resp. she) has graduated from university.

Fourth, Section 3.1 also provides evidence suggesting that permanent migrants assimilate relatively quickly in cities unlike floating migrants, who face limited access to urban amenities and are often only eligible for the jobs in the informal sector. We therefore choose the following modelling strategy:

(1) different vectors of parameters to model the systematic surplus from work of the respective types of migrants: B^P and C^P for permanent migrant men and

women respectively and, B^F and C^F for floating migrant men and women respectively, and

(2) different vectors of parameters to model the "marriage hedonic enjoyment" of rural native couples, A^R, urban native couples, A^U, floating migrant couples, A^F, and permanent migrant couples, A^P.

In particular we choose the following parametric specification for both the marriage and the work surplus. For the systematic joint surplus from marriage, letting $A = (A_0, A_1, A_2, A_3)$ and $A_i = \left(A_i^U, A_i^R, A_i^P, A_i^F\right)$ for $i = 1,2,3$ and $A_0 = \left(A_0^R, A_0^U, A_0^{M_m}, A_0^{M_w}, A_0^P, A_0^F\right)$, we distinguish between 5 types of couples:

(1) Rural native couples $Z(x) = Z(y) = z = 0$ for which

$$\Phi^{wed}(x, y, z; A) = \sum_{k,l} A^R_{1(k,l)} \left|\tilde{x}^{(k)} - \tilde{y}^{(l)}\right| + A_2^R \tilde{x} + A_3^R \tilde{y} + A_0^R,$$

(2) Urban native couples $Z(x) = Z(y) = z = 1$ for which

$$\Phi^{wed}(x, y, z; A) = \sum_{k,l} A^U_{1(k,l)} \left|\tilde{x}^{(k)} - \tilde{y}^{(l)}\right| + A_2^U \tilde{x} + A_3^U \tilde{y} + A_0^U,$$

(3) Mixed-couples whose husband is a migrant $1 - Z(x) = Z(y) = z = 1$ for which

$$\begin{aligned}\Phi^{wed}(x, y, z; A) &= \sum_{k,l} A^U_{1(k,l)} \left|\tilde{x}^{(k)} - \tilde{y}^{(l)}\right| \\ &+ \left[W(x) A_2^P + (1 - W(x)) A_2^F\right] \tilde{x} + A_3^U \tilde{y} \\ &+ \left[W(x) A_0^P + (1 - W(x)) A_0^F\right] + A_0^{M_m},\end{aligned}$$

(4) Mixed-couples whose wife is a migrant $Z(x) = 1 - Z(y) = z = 1$ for which

$$\begin{aligned}\Phi^{wed}(x, y, z; A) &= \sum_{k,l} A^U_{1(k,l)} \left|\tilde{x}^{(k)} - \tilde{y}^{(l)}\right| \\ &+ A_2^U \tilde{x} + \left[W(y) A_3^P + (1 - W(y)) A_3^F\right] \tilde{y} \\ &+ \left[W(y) A_0^P + (1 - W(y)) A_0^F\right] + A_0^{M_w},\end{aligned}$$

(5) Migrant couples $1-Z(x)=1-Z(y)=z=1$ for which

$$\begin{aligned}\Phi^{wed}(x,y,z;A) &= (1-W(x))(1-W(y))\sum_{k,l}A^{F}_{1(k,l)}\left|\tilde{x}^{(k)}-\tilde{y}^{(l)}\right| \\ &+[1-(1-W(x))(1-W(y))]\sum_{k,l}A^{P}_{1(k,l)}\left|\tilde{x}^{(k)}-\tilde{y}^{(l)}\right| \\ &+\left[W(x)A^{P}_{2}+(1-W(x))A^{F}_{2}\right]\tilde{x}+\left[W(y)A^{P}_{3}+(1-W(y))A^{F}_{3}\right]\tilde{y} \\ &+\left[W(x)A^{P}_{0}+(1-W(x))A^{F}_{0}\right]+\left[W(y)A^{P}_{0}+(1-W(y))A^{F}_{0}\right].\end{aligned}$$

The matrices A^i_1 for $i=\{R,U,F,P\}$, indicate the "marriage hedonic enjoyment" resulting from the adequation of both spouses' attributes whereas the vectors A^i_2 and A^i_3 for $i=\{R,U,F,P\}$, indicate the "marriage hedonic enjoyment" associated with the attributes of husbands and wives respectively. The constant terms A^i_0 for $i=\{R,U,M_m,M_w,F,P\}$ indicate "marriage hedonic enjoyment" at mean values of spouses' attributes for each type of couples and the associated moment conditions pin down the mass of couples of each type. Note that the matrix A^U_1 for urban couples also applies to mixed-couples as we assume that in these couples, the migrant spouse assimilates instantaneously. Using a similar argument, the matrix A^P_1 for permanent couples applies to couples consisting of a floating spouse and a permanent spouse.

The systematic surplus from work for men and women are modelled respectively as

$$\begin{aligned}\alpha^{work}(x,z;B) &= \left[B^{P}W(x)+B^{F}(1-W(x))\right]\begin{pmatrix}\tilde{x}\\1\end{pmatrix}\text{ if }Z(x)=1-z=0 \\ &= 0\text{ else,}\end{aligned}$$

where $B=\left(B^F,B^P\right)$, and

$$\begin{aligned}\gamma^{work}(y,z;C) &= \left[C^{P}W(y)+C^{F}(1-W(y))\right]\begin{pmatrix}\tilde{y}\\1\end{pmatrix}\text{ if }Z(y)=1-z=0 \\ &= 0\text{ else,}\end{aligned}$$

where $C = \left(C^F, C^P\right)$.

The model is therefore fully parameterized by the vector

$$\lambda = (A, B, C) .$$

4.2. **Estimates.** We estimate the vector of parameters λ using the Matching Moment estimator introduced in Section 2.3. We use height, health, education, BMI and age as the observable attributes for men and women such that λ contains 170 parameters.[10] Note that, without loss of generality, these attributes are standardized. This facilitates the comparison of the magnitude of the parameters, which are now expressed in standard deviation units for the corresponding attribute. Note also that all coefficients discussed below are statistically significant at 1% unless stated otherwise.

Parameters of the marriage surplus of rural and urban natives and floating and permanent migrants are presented in three tables. Tables 4 and 5 presents the estimates of the matrices A_1^U and A_1^R and, A_1^P and A_1^F respectively. Each entry (k, l) of these matrices indicates the relative weight of the absolute difference between the k^{th} (row) attribute of the husband and the l^{th} (column) attribute of the wife in generating the marriage hedonic surplus of the associated type of couples. Table 6 presents the estimates (A_2, A_3) of the marriage surplus related to the direct effects of observable attributes of each spouse for rural, urban, permanent and floating migrants respectively.

Inspection of Tables 4 and 5 reveals several important results. First, all coefficients on the diagonal of the matrices A_1^i $i = \{R, U, F, P\}$ are negative and significant, clearly indicating that like attracts like on all five attributes. Second, not surprisingly, we find that the adequation of spouses' education is an important contributor to marriage

[10]Each of the 4 affinity matrices A_1^i contains 5×5 parameters, vectors A_2^i and A_3^i for $i = U, R, P, F$ contain 5 parameters each, whereas vectors B^F, B^P, C^F and C^P contain $5 + 1$ parameters each and A_0^i for $i = U, R, M_m, M_w, P, F$ contain 1 parameter each.

surplus for all four types of couples distinguished. However, our third result is that the adequation of spouses' health and age is at least as important as the adequation of spouses' education.

Inspection of Table 6 reveals two additional results. First, for all types of couples, the marriage surplus increases with the age of both spouses. Second, we note that the marriage surplus increases with the education of husbands, more so if the husband is a permanent migrant, but decreases with the education of wives although only significantly so for urban wives.

Table 7 presents the estimates $\left(B^F, C^F\right)$ and $\left(B^P, C^P\right)$ related to the systematic surplus from work by gender and types of migrants. Not surprisingly, the surplus from work of both permanent migrant men and women is mostly driven by their educational level. A one standard deviation in education increases the surplus from work for both gender by more that 8 units (8.72 for men and 8.12 for women). The surplus from work of floating migrant women is also positively and significantly related with education, though to a lower extent (4.24) whereas education does not significantly relate with the surplus from work of floating migrant men.

4.3. **Migration in China: to work or to wed?** The remaining exercise consists in quantifying the welfare contribution of the three migration effects: "marrying-up", "work" and "marriage hedonic enjoyment". To this aim, we compute the expected indirect utility of men and women under four alternative values of the parameter λ representing each a different state of the world. Using the parameter estimates of the previous section $\hat{\lambda}$, we compute the expected indirect utility of each man and woman in the current state of the world, i.e. current situation on both the marriage and labor markets. Using λ^u defined such that $\Phi^{\lambda}_{ij1} = -\infty$ for all (i, j) with $Z(x_i) = 1 - Z(y_j)$, we compute the expected indirect utility of each man and woman in the absence of the "marrying-up" effect by excluding the option of entering a destination's marriage market and forming a "mixed-couple". Using $\lambda^{u,l}$ defined such that we further impose that the work surplus of migrants is equal to 0, i.e. each entry of B and C is set to 0, we compute the expected indirect utility of each man and woman in the absence of the "marrying-up" effect and the "work" effect. Finally, using $\lambda^{u,l,h}$ defined such that we further impose that the "marriage hedonic enjoyment" of migrants is set equal to that of rural natives, i.e. setting $A_i^P = A_i^F = A_i^R$ for $i = 0, 1, 2, 3$, we compute the expected indirect utility of each man and woman in the absence of all three effects of migration.

For each value $\lambda \in \left\{\hat{\lambda}, \lambda^u, \lambda^{u,l}, \lambda^{u,l,h}\right\}$, the expected indirect utilities of each man and woman are computed as follows. First, we use the algorithm presented in Section 2.4 to compute μ^λ the equilibrium matching associated with parameter λ. Second, plugging μ^λ into the empirical counterparts of Eqs. (2.16) and (2.17), we obtain the expected indirect utility of each man and woman as

$$\begin{aligned} G_i(\lambda) &= -\log \mu_{i0}^\lambda, \\ H_j(\lambda) &= -\log \mu_{0j}^\lambda, \end{aligned}$$

where we recall that $\mu_{i0}^\lambda = \mu_{i0Z(x_i)}^\lambda$ and $\mu_{0j}^\lambda = \mu_{0jZ(y_j)}^\lambda$.

Note that each individual observed in our dataset is either a urban native individual married with a permanent migrant UP,[11] married with a floating migrant UF, married with a urban native spouse UU, or a urban native single $U0$ or, a permanent migrant married with a urban native spouse PU, married with a permanent migrant PP, married with a floating migrant PF or single $P0$, or, a floating migrant married with a urban native spouse FU, married with a permanent migrant FP, married with floating migrant FF, or single $F0$ or, a rural native individual married with a rural native spouse RR or single $R0$. Denote T_i (T_j) the type of household formed by a man i (resp. woman j), and $\mathcal{T} = \{UP, UF, UU, U0, PU, PP, PF, P0, FU, FP, FF, F0, RR, R0\}$ the

[11]By convention, the labelling of each type of household is such that the first character indicates the agent's own type, i.e. either urban native U, permanent migrant P, floating migrant F, or rural native R, and the second character indicates the spouse's type, including 0 to indicate singlehood.

set of possible types of households. One can compute the average indirect utility of men and women observed in each type of households as

$$\bar{G}^T(\lambda) = \frac{\sum_{i\in\mathcal{I}|T_i=T} G_i(\lambda)}{\sum_{i\in\mathcal{I}|T_i=T} 1}, \quad (4.1)$$

$$\bar{H}^T(\lambda) = \frac{\sum_{j\in\mathcal{J}|T_j=T} H_j(\lambda)}{\sum_{j\in\mathcal{J}|T_j=T} 1}, \quad (4.2)$$

for $T \in \mathcal{T}$.

It follows that for observed migrants in households of type $T \in \{PU, PP, PF, FU, FP, FF\}$

(1)

$$\bar{G}^T(\hat{\lambda}) - \bar{G}^T(\lambda^u),$$
$$\bar{H}^T(\hat{\lambda}) - \bar{H}^T(\lambda^u),$$

quantify the contribution of the "marrying-up" effect of migration in the welfare of men and women of household type T,

(2)

$$\bar{G}^T(\lambda^u) - \bar{G}^T(\lambda^{u,l}),$$
$$\bar{H}^T(\lambda^u) - \bar{H}^T(\lambda^{u,l}),$$

quantify the contribution of the "work" effect of migration in the welfare of men and women of household type T,

(3)

$$\bar{G}^T(\lambda^{u,l}) - \bar{G}^T(\lambda^{u,l,h}),$$
$$\bar{H}^T(\lambda^{u,l}) - \bar{H}^T(\lambda^{u,l,h}),$$

quantify the contribution of the "marriage hedonic enjoyment" effect of migration in the welfare of men and women of household type T,

(4) whereas

$$\bar{G}^T(\hat{\lambda}) - \bar{G}^T(\lambda^{u,l,h}),$$
$$\bar{H}^T(\hat{\lambda}) - \bar{H}^T(\lambda^{u,l,h}),$$

quantify the total welfare effect of migration for men and women of household type T.

Table 8 presents the decomposition of the migration welfare for observed married migrants by type of households $T \in \{PU, PP, PF, FU, FP, FF\}$. We first note that the "work" effect of permanent migrants is systematically larger than that of floating migrants, a finding that (partly) reflects the better opportunities of the former in the urban labor market. We also find that the "work" effect of migrant men is larger than that of women regardless of the types of migrants considered. This finding is in line with empirical studies documenting a large (and rising) gender pay gap in China (Chi and Li, 2013 for instance). Nonetheless, the share of the "work" effect in the total migration surplus roughly varies between 3/5 and 4/5 across types of migrants and gender.

Our counter-factual results also show that the "marrying-up" effect is small and negative for men while relatively large and positive for women, compressing up to 21% (17%) of the migration surplus of permanent (resp. floating) migrant women in mixed-couples. These findings are consistent with a situation where the types of the migrant women observed in mixed-couples are in relative short supply in the urban marriage market. As these women enter the urban marriage market to "marry-up", they also decrease the supply of women (of their types) in the rural marriage market. This worsens the position of men such that for some types, i.e. those of the migrant men observed in mixed-couples, it becomes relatively preferable to enter the urban marriage market. In this case, the model would indeed predict a higher indirect utility for women observed in mixed-couples and a lower indirect utility for men observed in mixed-couples when rural natives can enter the urban marriage market and form mixed-couples compared to when they cannot.

Interestingly, the "marrying-up" effect is also sizeable for migrant women that are not observed being in mixed-couples. These spillover effects are explained by the fact that, when mixed-couples cannot form, individuals observed in mixed-couples enter the rural marriage market instead. Since migrant women in mixed-couples are relatively more numerous than migrant men in mixed-couples, competition in the rural marriage market increases relatively more on the women side, leading to a loss of utility for all women. These losses are sizeable and represent 17% (8%) of the migration surplus of observed permanent women married with a floating (resp. permanent) husband and 18% (9%) of the migration surplus of observed floating women married with a permanent (resp. floating) husband. Similarly, the welfare of urban natives is affected by whether or not mixed-couples are allowed. As anticipated from the aforementioned results, the "marrying-up" effect for urban native women is negative but small, their welfare decreasing slightly when rural native women observed in mixed-couples enter the urban marriage market, and positive and large for urban native men, compressing about 10% of their total welfare.

Finally, we find that the "marriage hedonic enjoyment" is generally larger for men than for women which is consistent with the male marital wage-premium observed in China (see Wang, 2013). We also find that the "marriage hedonic enjoyment" is larger for permanent migrants than for floating migrants which reflects the fact that permanent migrants do enjoy urban amenities (incl. health care, school quality, theater etc.) whereas floating migrants generally do not.

5. Summary and Discussion

This paper contributes to the literature on two counts. First, this paper presents a marriage matching model encompassing the classical matching model a la Becker

(1973) and Shapley and Shubik (1972) with the hedonic model a la Rosen (1974). As in the classical matching model, the marriage market is viewed as a matching market where men and women sort according to their attributes. Unlike the classical matching model, in the unified model there are several locations each with its own marriage market and labor market. The hedonic attribute in this model is then the location where individuals decide to live at. Individuals can choose to stay at their current location and enter both the local marriage market and labor market. Alternatively, individuals can enter the local marriage market at their birthplace and then migrate as a couple to enter the destination's local labor market. Finally, individuals can enter both the destination's local marriage market and labor market. Migration induces additional costs but may also generates benefits in the form of better perspectives on the destination's local markets. We draw insights from Choo and Siow (2006) and Galichon and Salanié (2015) in order to introduce unobserved heterogeneity in the model. We show that a stable equilibrium is equivalent to a Walrasian equilibrium and is Pareto optimal. We also provide identification results for the migration surplus of singles and couples.

Our second contribution is empirical. We apply our methodology on China's marriage market using data from the Rural Urban Migration in China (RUMiC) longitudinal dataset. In particular we study the extent to which individuals migrate to work or to wed.

We estimate the structural parameters of the model using an extension to the unified model of the matching moment estimator applied in Dupuy and Galichon (2014). Our estimates indicate that like attracts like on all five attributes used in the analysis: education, height, BMI, subjective health and age; the adequation of spouses' health and age is at least as important as the adequation of spouses' education. We also find that the marital surplus increases with the age of both spouses for rural, urban and migrant couples, increases with the education of husbands, more so if the husband is a permanent migrant, but decreases with the education of urban wives.

Not surprisingly, the migration surplus from work is mostly driven by education for permanent migrants (more so for men) and only increases with education for floating migrant women, though at 50% of the rate estimated for permanent migrant women.

Counter-factual exercises allow us to decompose the migration surplus into the "marrying-up", the "work" and the "marriage hedonic enjoyment" effects. As expected, our results show that both the "work" and "marriage hedonic enjoyment" effects are larger for permanent migrants and for men, reflecting the more favorable position of permanent migrants and men in the urban labor market and the marital wage premium of husbands. Comparing the "work" and the "marrying-up" effect also allows us to answer the question raised in the title. While the "marrying-up" effect is neglectable for migrant men, it accounts for about 1/5 of the migration surplus of migrant women in mixed-couples and 10% of that of other migrant women. The "work" effect accounts for roughly 4/5 of the migration surplus of both permanent and floating migrant men, the remaining 10% being ascribed to the "marriage hedonic enjoyment" effect. For permanent migrant women, the share of the "work" effect in total migration surplus ranges between 3/5 for permanent migrant women in mixed-couples and 4/5 for other migrant women whereas the share of the "marriage hedonic enjoyment" varies between 1/5 for the former and 10% for the latter. Interestingly, we also find that urban men married with a migrant wife would suffer a 10% welfare loss were their spouses not to enter the urban marriage market.

The encompassing model also has potential applications well outside family economics and the marriage market. The model developed in this paper, indeed, could easily be applied to other differentiated market such as the labor market, goods and services markets etc. While the hedonic approach has been favored in applications related to product differentiation, the growing importance of "fair trade" indicates that buyers actually care not only about the attributes of the product but also about the attributes of the seller. In labor economics, the encompassing model is also needed when workers' and firms' attributes can reflect both productivity and tastes, and sorting is driven by two simultaneous forces: compensating wage differentials and productive complementarities.

Appendix A. Gumbel random processes

Dupuy and Galichon (2014) introduced the continuous logit formalism in the classical Becker matching model. This Appendix extends this formalism to the encompassing model presented in this paper. Assume that each man m of type $x_m = x$ is acquainted with an infinite but countable random subset of the population of women and locations from which he chooses his potential partner (with the option to remain single) and place of residence. Acquaintances are indexed by $k \in \mathbb{N}$. An acquaintance k is defined by the tuple $\{y_k^m, z_k^m, \varepsilon_k^m\}$, where y_k^m is the observable type of the associated woman, z_k^m is the associated place of residence and ε_k^m is the "idiosyncratic enjoyment" provided by acquaintance k to man m, beyond the systematic enjoyment $\alpha(x, y_k^m, z_k^m)$ all men of type x would enjoy when married to a woman of type y_k^m and living at place z_k^m.

Let $\{y_k^m, z_k^m, \varepsilon_k^m, k \in \mathbb{N}\}$ denote the set of acquaintances of a man m. Assume that each point in this set is drawn from a Poisson point process on $\mathcal{Y}_0\mathcal{Z} \times \mathbb{R}$ with intensity $d\lambda_0\left(y\right) dz e^{-\varepsilon} d\varepsilon$ defined such that, for $S \subseteq \mathcal{Y}_0$,

$$\lambda_0\left(S\right) = 1\left(0 \in S\right) + \lambda\left(S \backslash \left\{0\right\}\right)$$

where λ is the Lesbegue measure on $\mathcal{Y}$.

Given our definition, the problem (2.1) faced by a man m of type x reads as

$$\max_k \left\{\alpha(x, y_k^m, z_k^m) + t(x, y_k^m, z_k^m) + \varepsilon_k^m\right\}. \tag{A.1}$$

Let u^m be the indirect utility of man m. The probability that man m's indirect utility is lower than $u \in \mathbb{R}$, is given as

$$\Pr\left(u^m \leq u, \forall k \in \mathbb{N}\right),$$

and coincides with the probability that the Poisson point process $\{y_k^m, z_k^m, \varepsilon_k^m, k \in \mathbb{N}\}$ has no points in $\{y, z, \varepsilon : \alpha(x, y, z) + t(x, y, z) + \varepsilon > u\}$. Hence, by definition of the

Poisson distribution, one has

$$\Pr\left(u^m \leq u, \forall k \in \mathbb{N}\right) = \exp\left(-\Lambda\left(u\right)\right)$$

where

$$\Lambda\left(u\right) = \int\int\int_{\mathcal{Y}_0 \mathcal{Z} \mathbb{R}} 1\left(\alpha(x,y,z) + t(x,y,z) + \varepsilon > u\right) d\lambda_0\left(y\right) dz e^{-\varepsilon} d\varepsilon,$$

is the intensity measure above utility level u.

Solving the integral over ε and rearranging obtains

$$\Lambda\left(u\right) = \exp\left(-u + B\left(x\right)\right),$$

where

$$B\left(x\right) := \log \int_{\mathcal{Z}} \left(\exp\alpha(x,0,z) + \int_{\mathcal{Y}} \exp\left(\alpha(x,y,z) + t(x,y,z)\right) dy\right) dz.$$

One concludes that random variable u^m follows a $\left(B\left(x\right),1\right)$-Gumbel type 1 distribution. By a similar reasoning, one can easily show that $\{\varepsilon_k^m, k \in \mathbb{N}\}$ follows a $(0,1)$-Gumbel type 1 distribution.

The above results imply that the expected indirect utility of a man of type $x \in \mathcal{X}$ obtains as

$$\begin{aligned} G_x(t) &:= E\left[\max_k \{\alpha(x, y_k^m, z_k^m) + t(x, y_k^m, z_k^m) + \varepsilon_k^m\}\right] \\ &= B\left(x\right) \\ &:= \log \int_{\mathcal{Z}} \left(\exp\alpha(x,0,z) + \int_{\mathcal{Y}} \exp\left(\alpha(x,y,z) + t(x,y,z)\right) dy\right) dz. \end{aligned} \tag{A.2}$$

Also, the conditional distribution of men of type x whose choice is y and z reads as

$$\mu\left(y,z|x\right) = \frac{\exp\left(\alpha(x,y,z) + t(x,y,z)\right)}{\int_{\mathcal{Z}} \left(\exp\alpha(x,0,z') + \int_{\mathcal{Y}} \exp\left(\alpha(x,y',z') + t(x,y',z')\right) dy'\right) dz'}.$$

By a similar token, denoting $\{x_k^w, z_k^w, \eta_k^w, k \in \mathbb{N}\}$ the set of acquaintances of a woman w, and the problem (2.2) faced by a woman w of type y reads as

$$\max_k \{\gamma(x_k^w, y, z_k^w) - t(x_k^w, y, z_k^w) + \eta_k^w\}. \tag{A.3}$$

One therefore obtains the expected indirect utility of a woman of type $y \in \mathcal{Y}$ as

$$\begin{aligned} H_y(t) &:= E\left[\max_k \{\gamma(x_k^w, y, z_k^w) - t(x_k^w, y, z_k^w) + \eta_k^w\}\right] \\ &= \log \int_{\mathcal{Z}} \left(\exp \gamma(0, y, z) + \int_{\mathcal{X}} \exp\left(\gamma(x, y, z) - t(x, y, z)\right) dx\right) dz, \end{aligned} \tag{A.4}$$

and the conditional distribution of women of type y whose choice is x and z as

$$\mu\left(x, z|y\right) = \frac{\exp\left(\gamma(x, y, z) - t(x, y, z)\right)}{\int_{\mathcal{Z}} \left(\exp \gamma(0, y, z') + \int_{\mathcal{X}} \exp\left(\gamma(x', y, z') - t(x', y, z')\right) dx'\right) dz'}.$$

Appendix B. Proof of Theorem 1

Proof. Part (i) is proved using step 1-3 below. Part (ii) follows from step 4 whereas part (iii) is proved in step 5.

Step 1: The first step in the proof is to define the stochastic processes $\varepsilon(x,y,z)$ and $\eta(x,y,z)$ used in the theorem, using the continuous logit formalism exposed in Appendix A. For each man m of type $x^m = x$, the stochastic process $\varepsilon(x,y,z)$ is defined on $\mathcal{Y}_0 \times \mathcal{Z}$ as

$$\varepsilon(x,y,z) = \max_k \{\varepsilon_k^m : y_k^m = y, z_k^m = z\}$$

if the set $\{k : y_k^m = y, z_k^m = z\}$ is nonempty and $\varepsilon(x,y,z) = -\infty$ else. Similarly, for each woman w of type $y^w = y$, the stochastic process $\eta(x,y,z)$ is defined on $\mathcal{X}_0 \times \mathcal{Z}$ as

$$\eta(x,y,z) = \max_k \{\eta_k^w : x_k^w = x, z_k^w = z\}$$

if the set $\{k : x_k^w = x, z_k^w = z\}$ is nonempty and $\eta(x,y,z) = -\infty$ else.

We can now use this definition in Equations (A.1) and (A.3) of Appendix A, to rewrite the indirect utility of a man m of type x as

$$u^m = \max_{yz \in \mathcal{Y}_0 \times \mathcal{Z}} \{\alpha(x,y,z) + t(x,y,z) + \varepsilon(x,y,z)\} \tag{B.1}$$

and the indirect utility of a woman w of type y as

$$v^w = \max_{xz \in \mathcal{X}_0 \times \mathcal{Z}} \{\gamma(x,y,z) - t(x,y,z) + \eta(x,y,z)\}. \tag{B.2}$$

Suppose that transfers would depend on the identity of both spouses, which we denote with a slight abuse of notation $t(m,w,z)$, rather than, as assumed so far, on their types x and y only. Note that (u^m, v^w) would then be such that

$$\begin{aligned} u^m - \varepsilon(x^m,y,z) - \alpha(x^m,y,z) &\geq t(m,w,z)\, \forall m, z \in \mathcal{Z}, y \in \mathcal{Y}_0, \\ \gamma(x,y^w,z) + \eta(x,y^w,z) - v^w &\leq t(m,w,z)\, \forall w, z \in \mathcal{Z}, x \in \mathcal{X}_0, \end{aligned}$$

with equality whenever m and w are married to each other.

Combining these two inequalities yield

$$u^m - \varepsilon(x^m, y, z) - \alpha(x^m, y, z) \geq t\,(m, w, z) \geq \gamma(x, y^w, z) + \eta(x, y^w, z) - v^w$$

$\forall m, w, z \in \mathcal{Z}, y \in \mathcal{Y}, x \in \mathcal{X}$.

Clearly, the term on the left hand side does not depend on the identity w of the wife, whereas the term on the right hand side does not depend on the identity m of the husband. One concludes that transfers t do not vary with the identity of spouses, only with their observable types (x, y) and location z.

Finally, the expected indirect utility of men of type x and women of type y are respectively given as

$$G_x(t) = E\left[\max_{yz \in \mathcal{Y}_0 \times \mathcal{Z}} \{\alpha(x, y, z) + t(x, y, z) + \varepsilon(x, y, z)\}\right],$$

and

$$H_y(t) = E\left[\max_{xz \in \mathcal{X}_0 \times \mathcal{Z}} \{\gamma(x, y, z) - t(x, y, z) + \eta(x, y, z)\}\right].$$

Step 2: From Lemma 1 in Chiappori et al. (2010), we know that the indirect utilities (u^m, v^w) solve Kantorovich's dual problem which in our case reads as

$$\mathcal{W}(\Phi) := \inf_{\tilde{u}^m, v^w} \int \tilde{u}^m dm + \int \tilde{v}^w dw,$$

$$s.t.$$

$$\begin{aligned} \tilde{u}^m + \tilde{v}^w &\geq \alpha(x^m, y^w, z) + \gamma(x^m, y^w, z) + \varepsilon(x^m, y^w, z) + \eta(x^m, y^w, z) \\ \tilde{u}^m &\geq \alpha(x^m, 0, z) + \varepsilon(x^m, 0, z) \\ \tilde{v}^w &\geq \gamma(0, y^w, z) + \eta(0, y^w, z). \end{aligned}$$

$\forall m, w, z \in \mathcal{Z}, x \in \mathcal{X}, y \in \mathcal{Y}$.

The constraints of the dual problem rewrite as

$$\begin{aligned} U(x,y,z)+V(x,y,z) &\geq \alpha(x,y,z)+\gamma(x,y,z) \\ U(x,0,z) &\geq \alpha(x,0,z) \\ V(0,y,z) &\geq \gamma(0,y,z) \end{aligned}$$

where

$$\begin{aligned} U(x,y,z) &= \inf_{m|x^m=x} \left(\tilde{u}^m - \varepsilon(x,y,z)\right), \\ V(x,y,z) &= \inf_{w|y^w=y} \left(\tilde{v}^w - \eta(x,y,z)\right), \end{aligned}$$

such that one has

$$\begin{aligned} \tilde{u}^m &= \sup_{yz\in\mathcal{Y}_0\times\mathcal{Z}} \left(U(x,y,z)+\varepsilon(x,y,z)\right), \\ \tilde{v}^w &= \sup_{xz\in\mathcal{X}_0\times\mathcal{Z}} \left(V(x,y,z)+\eta(x,y,z)\right). \end{aligned}$$

Replacing $(\tilde{u}^m, \tilde{v}^w)$ by its expression in terms of U and V in the Kantorovich's dual problem one obtains

$$\mathcal{W}(\Phi) := \inf_{U,V} \int_{\mathcal{X}} G_x(U-\alpha)f(x)dx + \int_{\mathcal{Y}} H_y(\gamma - V)g(y)dy \tag{B.3}$$

where the function G_x and H_y are defined above and we have used the fact that $U(x,y,z) = \alpha(x,y,z) + t(x,y,z)$ when y and z solve Eq. (B.1) and $V(x,y,z) = \gamma(x,y,z) - t(x,y,z)$ when x and z solve Eq. (B.2).

One concludes that the dual problem rewrites as stated in Theorem 1 as

$$\mathcal{W}(\Phi) := \inf_{t} \int_{\mathcal{X}} G_x(t)f(x)dx + \int_{\mathcal{Y}} H_y(t)g(y)dy. \tag{B.4}$$

Step 3: Rewriting the dual program (B.3) as a saddle point problem, one obtains

$$\begin{aligned}
\mathcal{W}(\Phi) &= \inf_{U,V}\sup_{\mu}\int\int\int_{\mathcal{X}\mathcal{Y}\mathcal{Z}}\left(\Phi\left(x,y,z\right)-U\left(x,y,z\right)-V\left(x,y,z\right)\right)\mu\left(x,y,z\right)dxdydz \\
&+\int\int_{\mathcal{X}\mathcal{Z}}\left(\alpha\left(x,0,z\right)-U\left(x,0,z\right)\right)\mu\left(x,0,z\right)dxdz \\
&+\int\int_{\mathcal{Y}\mathcal{Z}}\left(\gamma\left(0,y,z\right)-V\left(0,y,z\right)\right)\mu\left(0,y,z\right)dydz \\
&+\int_{\mathcal{X}}G_x(U-\alpha)f(x)dx+\int_{\mathcal{Y}}H_y(\gamma-V)g(y)dy.
\end{aligned}$$

This can be written as

$$\begin{aligned}
\mathcal{W}(\Phi) &= \sup_{\mu}\int\int\int_{\mathcal{X}\mathcal{Y}\mathcal{Z}}\Phi\left(x,y,z\right)\mu\left(x,y,z\right)dxdydz \\
&\int\int_{\mathcal{X}\mathcal{Z}}\alpha\left(x,0,z\right)\mu\left(x,0,z\right)dxdz+\int\int_{\mathcal{Y}\mathcal{Z}}\gamma\left(0,y,z\right)\mu\left(0,y,z\right)dydz \\
&-\mathcal{E}\left(\mu\right)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{E}\left(\mu\right) &= -\inf_{U}\int_{\mathcal{X}}\left(G_x(U-\alpha)f(x)-\int_{\mathcal{Z}}\left(U\left(x,0,z\right)\mu\left(x,0,z\right)+\int_{\mathcal{Y}}U\left(x,y,z\right)\mu\left(x,y,z\right)dy\right)dz\right)dx \\
&-\inf_{V}\int_{\mathcal{Y}}\left(H_y(\gamma-V)g(y)-\int_{\mathcal{Z}}\left(V\left(0,y,z\right)\mu\left(0,y,z\right)+\int_{\mathcal{X}}V\left(x,y,z\right)\mu\left(x,y,z\right)dx\right)dz\right)dy.
\end{aligned}$$

Using the closed form expressions for G_x and H_y in Eqs. (A.2) and (A.4), the first order conditions of the underlying infimum read respectively as

$$\frac{\exp\left(U\left(x,y,z\right)\right)}{\int_{\mathcal{Z}}\left(\exp U(x,0,z')+\int_{\mathcal{Y}}\exp\left(U(x,y',z')\right)dy'\right)dz'}f(x) = \mu\left(x,y,z\right) \quad \text{(B.5)}$$

$$\frac{\exp\left(V(x,y,z)\right)}{\int_{\mathcal{Z}}\left(\exp V(0,y,z')+\int_{\mathcal{X}}\exp\left(V(x',y,z')\right)dx'\right)dz'}g\left(y\right) = \mu\left(x,y,z\right), \quad \text{(B.6)}$$

$\forall x\in\mathcal{X}, yz\in\mathcal{Y}_0\mathcal{Z}$ and $\forall y\in Y, xz\in\mathcal{X}_0\mathcal{Z}$ respectively.

Hence, the value of the respective infimum is $-\infty$ unless $\mu \in \mathcal{M}(f,g)$. In the latter situation, note that, taking logs on both sides, the first order conditions rewrite as

$$\begin{aligned} U(x,y,z) &= G_x(U-\alpha) - \log f(x) + \log \mu(x,y,z) \, \forall x \in \mathcal{X}, yz \in \mathcal{Y}_0\mathcal{Z} \\ V(x,y,z) &= H_y(\gamma - V) - \log g(y) + \log \mu(x,y,z) \, \forall y \in Y, xz \in \mathcal{X}_0\mathcal{Z}. \end{aligned}$$

Plugging these into the expression of $\mathcal{E}(\mu)$ obtains

$$\begin{aligned} \mathcal{E}(\mu) &= 2\int_{\mathcal{X}}\int_{\mathcal{Z}}\int_{\mathcal{Y}} \log \mu(x,y,z)\, \mu(x,y,z)\, dydzdx \\ &+ \int_{\mathcal{X}}\int_{\mathcal{Z}} \log \mu(x,0,z)\, \mu(x,0,z)\, dzdx \\ &+ \int_{\mathcal{Y}}\int_{\mathcal{Z}} \log \mu(x,0,z)\, \mu(0,y,z)\, dzdy \\ &- \int_{\mathcal{X}} f(x) \log f(x) dx \\ &- \int_{\mathcal{Y}} g(y) \log g(y)\, dy. \end{aligned}$$

Step 4: Using the FOCs (B.5) and (B.6) for $(x,0,z)$ and $(x,0,Z(x))$ and $(0,y,z)$ and $(0,y,Z(y))$ respectively, one directly sees that

$$\begin{aligned} \log(\alpha(x,0,z)) &= \log \frac{\mu(x,0,z)}{\mu(x,0,Z(x))} \forall x \in \mathcal{X}, z \in \mathcal{Z} \\ \log(\gamma(0,y,z)) &= \log \frac{\mu(0,y,z)}{\mu(0,y,Z(y))} \forall y \in Y, z \in \mathcal{Z} \end{aligned}$$

Then, using the FOCs (B.5) and (B.6) for (x,y,z) and $(x,0,Z(x))$ and (x,y,z) and $(0,y,Z(y))$ respectively, one obtains

$$\begin{aligned} U(x,y,z) &= \log \frac{\mu(x,y,z)}{\mu(x,0,Z(x))} \forall x \in \mathcal{X}, yz \in \mathcal{Y}_0\mathcal{Z} \\ V(x,y,z) &= \log \frac{\mu(x,y,z)}{\mu(0,y,Z(y))} \forall y \in Y, xz \in \mathcal{X}_0\mathcal{Z}. \end{aligned}$$

Adding the two expressions yields

$$\alpha(x,y,z) + \gamma(x,y,z) = 2\log \frac{\mu(x,y,z)}{\mu(x,0,Z(x))\, \mu(0,y,Z(y))} \forall xyz \in \mathcal{X}\mathcal{Y}\mathcal{Z}.$$

Step 5: Part (iii) follows directly from equations (B.1) and (B.2). ∎

References

[1] Abramitzky, A., Delavande, A. and L. Vasconcelos, (2011): "Marrying Up: The Role of Sex Ratio in Assortative Matching," *American Economic Journal: Applied Economics*, 3, 124-157.

[2] Akgüç, M., Giulietti, C. and Klaus F. Zimmermann, (2013): "The RUMiC Longitudinal Survey: Fostering Research on Labor Markets in China", IZA DP No. 7860.

[3] Amuedo-Dorantes, Catalina, and Shoshana, Grossbard, (2008): "Cohort-level sex ratio effects on women's labor force participation," *Review of Economics of the Household*, 6(3), 309-309.

[4] Angrist, Josh, (2002)."How Do Sex Ratios Affect Marriage And Labor Markets? Evidence From America's Second Generation," *The Quarterly Journal of Economics,* 117(3), 997-1038.

[5] Banerjee, A., Duflo, E., Ghatak, M. and J. Lafortune, (2013): "Marry for What? Caste and Mate Selection in Modern India", *American Economic Journal: Microeconomics*, 5(2), 33-72.

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

[7] Chang, S. and X. Zhang, (2012): "The Economic Consequences of excess Men Evidence from a Natural Experiment in Taiwan", IFPRI Discussion Paper No. 01203.

[8] Chiappori, P-A., Iyigun, M. and Y. Weiss, (2017): "Investment in schooling and the marriage market." *American Economic Review*, 99(5), 1689-713.

[9] Chiappori, P-A., McCann, R. and L. Nesheim, (2010): "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, 42(2), 317-354.

[10] Congressional-Executive Commission on China, (2017): "Recent Chinese Hukou Reforms", *https://www.cecc.gov/recent-chinese-hukou-reforms.*

[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., (2010): "Sorting on Skills and Preferences: Tinbergen Meets Sattinger," IZA working paper, No. 5143.

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

[16] Ekeland, I., and Heckman, J.J., and L. Nesheim, (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(S1), S60-S109.

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

[18] Gautier, Pieter A., Svarer, Michael, and Coen N. Teulings, (2010): "Marriage and the city: Search frictions and sorting of singles," *Journal of Urban Economics*, 67(2), 206-218.

[19] Jaffe, Sonia and Simon Weber, (2018): "The Effect of Meeting Rates on Matching Outcomes", working paper.

[20] Kong, Sherry Tao, (2010), "Rural-Urban Migration in China: Survey Design and Implementation," in X. Meng, C. Manning, S. Li, and T. N. Effendi, eds., The Great Migration: Rural-Urban Migration in China and Indonesia, 135-150, Northampton, MA: Edward Elgar.

[21] Korenman, S. and D. Neumark, (1991). "Does Marriage Really Make Men More Productive?" *The Journal of Human Resources*, 26, 282-307.

[22] Lafortune, J., (2013): "Making Yourself Attractive: Pre-Marital Investments and the Returns to Education on the Marriage Market," *American Economic Journal: Applied Economics,* 5(2), 151-78.

[23] Liang, Z. and Z. Ma, (2004): "China's floating population: new evidence from the 2000 census", *Population and Development Review*, 30(3), 467-488.

[24] McCann, R., Shi, X., Siow, A. and R. Wolthoff, (2015): "Becker meets Ricardo: Multisector matching with communication and cognitive skills", *Journal of Law, Economics and Organization*, 31(4), 690-720.

[25] Mourifié, I. and A. Siow, (2014): "Cohabitation versus marriage: Marriage matching with peer effects", working paper.

[26] Nie, H. and C. Xing, (2010): "When City Boy Falls in Love with Country Girl: Baby's Hukou, Hukou Reform, and Inter-Hukou Marriage" working paper.

[27] Porter, M., (2007): "The Effect of Sex-Ratio Imbalance in China on Marriage and Household Bargaining," *University of Chicago Working paper.*

[28] Porter, M., (2016): "How do sex ratios in China influence marriage decisions and intra-household resource allocation?", *Review of Economics of the Household*, 14(2), 337-371.

[29] Quintana-Domeque, C., (2011): "Preferences, Comparative Advantage, and Compensating Wage Differentials for Job Routinization", *Oxford Bulletin of Economics and Statistics*, 73(2), 207-229.

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

[31] Shi, L., (2008): "Rural Migrant Workers in China: Scenario, Challenges and Public Policy", Working Paper No. 89.

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

[33] Song Jin, Terry Sicular, and Yue Ximing, (2011): "Appendix II: The 2002 and 2007 CHIP Surveys: Sampling, Weights, and Combining the Urban, Rural, and Migrant Samples", working paper.

[34] Tinbergen, J., (1956): "On the Theory of Income Distribution," *Weltwirtschaftliches Archiv,* 77, 155-175.

[35] Wang, Shing-Yi, (2013): "Marriage Networks, Nepotism, and Labor Market Outcomes in China", *American Economic Journal: Applied Economics,* 5(3), 91-112.

[36] Weiss, Y., Yi, J., and J. Zhang, (2013): "Hypergamy, Cross-boundary Marriages, and Family Behavior", *IZA DP No. 7293.*

[37] Ying, H., Shuzhuo, L. and M.W. Feldman, (2014):"Trends and determinants of female marriage migration in contemporary China", working paper.

[38] Zhang, H., (2014): "Premarital Investments and Multidimensional Matching", working paper.

[39] Zhao, Z., (2005): "Migration, Labor Market Flexibility, and Wage Determination in China: A Review", *Developing Economies*, 43(2).

Appendix C. Tables

Table 1. Identification of the surplus of migration from work and marriage: Illustration of definition 3 with two locations, z =Urban and z_0 =Rural.

$Z(x)$	$Z(y)$	Live	#	Φ	=	α^{work}	+	γ^{work}	+	Φ^{wed}
z_0	0	z	1	$\Phi(z_0,0,z)$	=	$\alpha^{work}(z_0,0,z)$	+	0	+	0
0	z_0	z	2	$\Phi(0,z_0,z)$	=	0	+	$\gamma^{work}(0,z_0,z)$	+	0
z_0	z_0	z_0	3	$\Phi(z_0,z_0,z_0)$	=	0	+	0	+	$\Phi^{wed}(z_0,z_0,z_0)$
z_0	z_0	z	4	$\Phi(z_0,z_0,z)$	=	$\alpha^{work}(z_0,0,z)$	+	$\gamma^{work}(0,z_0,z)$	+	$\Phi^{wed}(z_0,z_0,z)$
z_0	z	z	5	$\Phi(z_0,z,z)$	=	$\alpha^{work}(z_0,0,z)$	+	$\gamma^{work}(0,z,z)$	+	$\Phi^{wed}(z_0,z,z)$
z	z_0	z	6	$\Phi(z,z_0,z)$	=	$\alpha^{work}(z,0,z)$	+	$\gamma^{work}(0,z_0,z)$	+	$\Phi^{wed}(z,z_0,z)$
z	z	z	7	$\Phi(z,z,z)$	=	$\alpha^{work}(z,0,z)$	+	$\gamma^{work}(0,z,z)$	+	$\Phi^{wed}(z,z,z)$

Table 2. Number of couples and singles by geographical types of households and gender.

	Urban		Rural		Float.	Migrants	Perm.	Migrants
	Men	Women	Men	Women	Men	Women	Men	Women
Singles	384	442	356	304	117	102	82	76
Couples	274	257	434	434	201	218	143	143
incl. Mixed-couples	33	16	0	0	2	17	14	16

Table 3. Summary statistics for the selected variables by geographical types of households and gender.

	Urban			Rural			Float.	Migrants		Perm.	Migrants	
	Obs.	Mean	std.	Obs.	Mean	std.	Obs.	Mean	std.	Obs.	Mean	std.
Men												
Height	658	173.89	5.07	790	168.58	5.54	318	171.23	5.24	225	171.88	4.88
Health	658	3.00	0.65	790	3.22	0.72	318	3.12	0.74	225	2.97	0.63
Education	658	3.96	0.94	790	2.75	0.62	318	2.22	0.64	225	3.88	1.03
BMI	658	22.68	2.82	790	22.22	2.20	318	22.50	2.87	225	22.75	2.87
Age	658	27.71	3.59	790	27.23	3.55	318	27.40	3.61	225	28.60	3.60
Women												
Height	699	162.03	5.06	738	160.45	5.46	320	160.69	4.59	219	160.13	4.40
Health	699	2.97	0.60	738	3.18	0.67	320	3.07	0.68	219	2.94	0.65
Education	699	4.02	0.87	738	2.70	0.74	320	2.19	0.67	219	3.84	0.96
BMI	699	20.55	2.45	738	21.16	2.23	320	20.65	2.62	219	20.43	2.11
Age	699	25.52	3.61	738	25.19	3.66	320	25.43	3.50	219	26.64	3.53

TABLE 4. Estimates of the marriage surplus: adequation of spouses' attributes.

Men/Women	Height	Health	Education	BMI	Age
Urban					
Height	-1.07	0.24	0.16	-0.15	0.06
	(0.151)	(0.178)	(0.156)	(0.140)	(0.172)
Health	0.04	-3.16	0.12	0.28	-0.28
	(0.162)	(0.174)	(0.182)	(0.165)	(0.205)
Education	-0.13	-0.32	-2.50	0.18	-0.62
	(0.169)	(0.198)	(0.152)	(0.153)	(0.167)
BMI	0.47	0.72	-0.33	-0.45	-0.55
	(0.145)	(0.169)	(0.146)	(0.124)	(0.157)
Age	-0.64	0.01	-0.21	-0.41	-2.80
	(0.174)	(0.202)	(0.164)	(0.156)	(0.184)
Rural					
Height	-0.80	0.10	0.29	-0.36	0.03
	(0.115)	(0.142)	(0.141)	(0.135)	(0.153)
Health	-0.10	-3.36	0.53	0.31	-0.46
	(0.132)	(0.152)	(0.178)	(0.138)	(0.156)
Education	0.43	1.03	-2.50	0.39	-0.37
	(0.161)	(0.195)	(0.158)	(0.206)	(0.240)
BMI	-0.04	0.09	0.53	-1.37	-0.18
	(0.123)	(0.140)	(0.169)	(0.123)	(0.141)
Age	0.24	-0.12	0.51	0.04	-3.34
	(0.129)	(0.146)	(0.195)	(0.127)	(0.155)

Note: Standard errors, calculated from the Hessian of the likelihood, are in parentheses.

Table 5. Estimates of the marriage surplus: adequation of spouses' attributes cont..

Men/Women	Height	Health	Education	BMI	Age
Permanent					
Height	-0.64	0.15	-0.12	0.13	0.01
	(0.231)	(0.264)	(0.233)	(0.220)	(0.277)
Health	0.08	-3.45	0.55	0.21	-0.27
	(0.236)	(0.271)	(0.259)	(0.236)	(0.309)
Education	0.07	-0.03	-2.29	-0.52	-1.47
	(0.265)	(0.291)	(0.225)	(0.247)	(0.290)
BMI	-0.54	1.05	0.10	-0.39	-0.21
	(0.223)	(0.269)	(0.222)	(0.204)	(0.256)
Age	-0.23	-0.17	-0.16	0.38	-2.98
	(0.270)	(0.297)	(0.250)	(0.246)	(0.305)
Floating					
Height	-0.41	0.73	-0.87	0.33	-0.10
	(0.207)	(0.237)	(0.346)	(0.203)	(0.234)
Health	-0.26	-2.82	1.02	0.11	-0.03
	(0.229)	(0.247)	(0.395)	(0.222)	(0.258)
Education	0.36	0.88	-2.25	0.70	0.32
	(0.339)	(0.442)	(0.539)	(0.333)	(0.443)
BMI	-0.05	0.46	0.47	-0.29	0.00
	(0.203)	(0.231)	(0.373)	(0.182)	(0.217)
Age	0.41	-0.33	-0.86	-0.40	-3.16
	(0.233)	(0.263)	(0.507)	(0.215)	(0.288)

Note: Standard errors, calculated from the Hessian of the likelihood, are in parentheses.

Table 6. Estimates of the marriage surplus: individual spouses' attributes.

	Height	Health	Education	BMI	Age	constant
Men						
Rural	-0.13	0.16	0.80	-0.41	0.96	-3.66
	(0.142)	(0.179)	(0.337)	(0.141)	(0.165)	(0.185)
Urban	-0.19	0.45	0.54	-0.13	2.26	-4.10
	(0.170)	(0.211)	(0.204)	(0.142)	(0.203)	(0.241)
Floating	0.84	-1.06	0.80	-0.10	2.12	-3.61
	(0.307)	(0.297)	(0.850)	(0.320)	(0.442)	(0.730)
Permanent	0.06	0.19	2.70	1.05	2.72	-3.89
	(0.372)	(0.381)	(0.476)	(0.316)	(0.459)	(0.679)
Women						
Rural	-0.09	-0.48	-0.01	0.56	2.73	-3.66
	(0.122)	(0.170)	(0.250)	(0.160)	(0.217)	(0.185)
Urban	-0.68	0.27	-1.57	0.59	2.34	-4.10
	(0.171)	(0.219)	(0.217)	(0.154)	(0.221)	(0.241)
Floating	0.09	-0.71	-0.60	-0.27	0.93	-3.61
	(0.272)	(0.304)	(0.678)	(0.278)	(0.383)	(0.730)
Permanent	-1.46	-0.19	-0.17	0.15	4.00	-3.89
	(0.382)	(0.367)	(0.530)	(0.359)	(0.555)	(0.679)

Notes: The constant for husbands and wives are equal to each other by convention as the moment estimator only identifies one constant per couple. Standard errors, calculated from the Hessian of the likelihood, are in parentheses.

TABLE 7. Estimates of the migration surplus from work.

	Height	Health	Education	BMI	Age	constant
Men						
Floating	1.06	-0.35	-0.94	-0.13	0.21	2.00
	(0.232)	(0.205)	(0.572)	(0.207)	(0.228)	(0.778)
Permanent	0.56	-0.24	8.72	0.04	0.14	-3.37
	(0.352)	(0.304)	(0.720)	(0.317)	(0.378)	(0.540)
Women						
Floating	0.21	0.27	4.24	0.02	0.73	7.65
	(0.196)	(0.195)	(0.591)	(0.185)	(0.267)	(0.953)
Permanent	0.01	-0.55	8.12	-1.06	-0.25	-7.68
	(0.273)	(0.244)	(0.763)	(0.307)	(0.364)	(0.734)

Note: Standard errors, calculated from the Hessian of the likelihood, are in parentheses.

TABLE 8. Migration in China: to Work or to Wed? Decomposition of the migration surplus.

Source	Wed	Work	Hedonic	Total	Share (%)
Men					
Permanent in mixed-couple: PU	-0.01	4.31	0.82	5.12	3
Permanent with floating spouse: PF	-0.03	1.96	0.51	2.44	15
Permanent couple: PP	-0.02	5.62	0.64	6.23	24
Floating in mixed-couple: FU	-0.04	0.39	0.61	0.96	1
Floating with permanent: FP	-0.04	1.62	0.43	2.01	11
Floating couple: FF	-0.03	1.18	0.33	1.48	46
Women					
Permanent in mixed-couple: PU	0.18	0.52	0.15	0.85	5
Permanent with floating spouse: PF	0.14	0.62	0.08	0.84	10
Permanent couple: PP	0.23	2.30	0.21	2.74	23
Floating in mixed-couple: FU	0.10	0.49	-0.02	0.58	4
Floating with permanent: FP	0.10	0.42	0.02	0.55	14
Floating couple: FF	0.06	0.64	-0.00	0.70	44

Note: The abbreviation for types of households is such that the first letter indicates the type of the individual considered (U for urban native, R for rural native, P for permanent migrant and F for floating migrant) and the second letter indicates the type of his/her spouse.