Matching models II: Empirics[1]

Alexander Wintzéus[2]

Department of Economics
University of Leuven

June 2, 2025

[1] Slides provided by Thimo De Schouwer – Based on slides by Aloysius Siow

[2] Email: alexander.wintzeus@kuleuven.be

Table of contents

Table of contents

Introduction – Empirical models of matching

- The matching model we studied last class makes extremely strong predictions
    - supermodularity of the surplus predicts **perfect** Positive Assortative Matching
    - which will obviously never hold in real world data
    - so we need to introduce uncertainty into the model
- We do so by allowing for the joint surplus to contain an unobserved component
- This yields a highly tractable model – under some limiting assumptions
    - this was the insight of a highly influential paper by Choo and Siow (2006)
    - subsequent work has relaxed these assumptions – but we'll stick to their model
- In the tutorial, we'll take this model to the data

Introduction – Choo and Siow model

- The Choo and Siow (2006) model is a frictionless transferable utility model
- The random surplus component is an additively separable random preference shock
  - as in McFadden (1974)'s Random Utility Model (RUM) you studied earlier
- The model is identified from observing matching patterns in a single (large) market
- We cannot recover male and female spousal preferences – only the joint surplus
  - in some extensions these can be recovered (not discussed today)
  - they are also recoverable if we observe marital transfers (which we typically never do)

Table of contents

Preferences

- Let the utility of man i of type I who is married to a woman of type J be

$$U^i_{IJ} = \tilde{\alpha}_{IJ} - \tau_{IJ} + \epsilon^i_{IJ} \tag{1}$$

  - with $\tilde{\alpha}_{IJ}$ the systematic value of type J women for type I men
  - with τ_{IJ} the equilibrium transfer made by type I men to type J women
  - and ϵ^i_{IJ} an idiosyncratic preference shock specific to man i

- This is the standard additive RUM you studied in previous classes:
  - there is a systematic component common to all men in (I, J) marriages
    - neither the systematic return nor the transfer depend on the specific woman chosen
  - and an idiosyncratic component specific to man i
    - ϵ^i_{IJ} is also independent of any particular woman j of type J

Preferences (cont.)

- The payoff to man i from remaining unmarried ($J = 0$) is:

$$U^i_{I0} = \tilde{\alpha}_{I0} + \epsilon^i_{I0} \tag{2}$$

- Man i will choose his partner according to:

$$U^i_I = \max_K \{U^i_{I0}, \ldots, U^i_{IK}, \ldots, U^i_{IJ}\} \tag{3}$$

Optimal choices

- The probability that man i will choose a type J woman is:

$$\Pr\left(U^i_{IJ} - U^i_{IK} \geq 0 \text{ for all } K = 0, \cdots, J\right) \tag{4}$$

$$\Pr\left(\tilde{\alpha}_{IJ} - \tau_{IJ} - (\tilde{\alpha}_{IK} - \tau_{IK}) \geq \epsilon^i_{IK} - \epsilon^i_{IJ} \text{ for all } K = 0, \cdots, J\right) \tag{5}$$

- Assuming McFadden (1974)'s EV Type I errors, this is known in closed form:

$$\frac{\mu^d_{IJ}}{m_I} = \frac{\exp(\tilde{\alpha}_{IJ} - \tau_{IJ})}{\sum_K \exp(\tilde{\alpha}_{IK} - \tau_{IK})} \tag{6}$$

  - μ^d_{IJ} the *demanded* number of matches between type I men and type J women by type I men
  - m_I the number of type I men

Optimal choices (cont.)

- We obtain the quasi-demand for type J women as:

$$\ln \mu_{IJ}^{d} - \ln \mu_{I0}^{d} = \alpha_{IJ} - \tau_{IJ}, \tag{7}$$

  - where $\alpha_{IJ} = \tilde{\alpha}_{IJ} - \tilde{\alpha}_{I0}$
  - i.e., the systematic payoff to type I man married to a type J women relative to singlehood

- Similar reasoning and calculations leads to a quasi demand for type I men:

$$\ln \mu_{IJ}^{s} - \ln \mu_{0K}^{s} = \gamma_{IJ} + \tau_{IJ}, \tag{8}$$

  - where $\gamma_{IJ} = \tilde{\gamma}_{IJ} - \tilde{\gamma}_{0J}$
  - i.e., systematic payoff to type J woman married to a type I men relative singlehood

Equilibrium

- Note that we will not directly observe these demanded quantities
  - what type of partner is man i looking for?
- But we can assume that our data resembles an equilibrium snapshot of a cleared market:

$$\mu^d_{IJ} = \mu_{IJ} = \mu^s_{IJ} \tag{9}$$

  - this is a market clearing condition
  - remaining singles are so voluntarily

Matching function

- The market clearing condition allows us to recover a matching function
- Some algebra after summing the two demand functions yields:

$$\ln \mu_{IJ} - \frac{\ln \mu_{I0} + \ln \mu_{0J}}{2} = \frac{\alpha_{IJ} + \gamma_{IJ}}{2} \tag{10}$$

$$\frac{\mu_{IJ}}{\sqrt{\mu_{I0}\mu_{0J}}} = \Pi_{IJ} \tag{11}$$

  where $\Pi_{IJ} = \exp(\frac{\alpha_{IJ}+\gamma_{IJ}}{2})$

- This matching function relates the number of singles and matches to the *surplus*
    - Recall that $\alpha_{IJ} + \gamma_{IJ}$ *is* the surplus
    - Choo and Siow (2006) label Π_{IJ} the match *surplus*

Matching function (cont.)

- The marriage matching function has constant returns to scale
  - doubling the number of singles and matches leaves surplus estimate unchanged
- The function can fit any observed matching distribution
- The surplus can be estimated within a single marriage market
  - note that there are $I \times J$ endogenous data points
  - and there are $I \times J$ parameters in the model
  - estimation inherently non-parametric

Estimates

Figure: Surplus Estimates from Choo and Siow (2006)

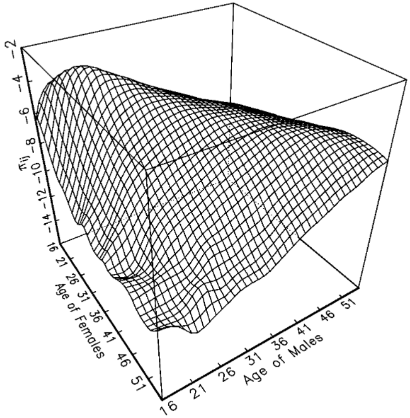

FIG. 4.—Smoothed π_{ij} for 1971/72

Sorting

- Suppose that I and J denote whether a man or a woman has a college degree, resp.
- λ_C and ω_C: denote the share of college educated men and women
- $\pi_{C,C}$ and $\pi_{N,N}$: share of couples where both have (resp. don't have) a college degree
- The stable matching is (positively) assortative, in the sense that

$$\pi_{C,C} + \pi_{N,N} \geq \lambda_C \omega_C + (1 - \lambda_C)(1 - \omega_C) \tag{12}$$

if and only if

$$\Pi_{C,C} + \Pi_{N,N} \geq \Pi_{C,N} + \Pi_{N,C} \tag{13}$$

Sorting (cont.)

- In others words, assortativeness is defined relative to *random matching*
  - logical arrow works in both directions
  - sorting only a function of the surplus

- More generally, we can show how matching patterns relate to the surplus:

$$\ln \frac{\mu_{I+1,J+1} \times \mu_{IJ}}{\mu_{I+1,J} \times \mu_{I,J+1}} = (\Pi_{I+1,J+1} + \Pi_{IJ}) - (\Pi_{I+1,J} + \Pi_{I,J+1}) \tag{14}$$

  under the implicit assumption that
  - the surplus Π_{IJ} is increasing in both I and J
  - there are complementarities (assortativeness)

Table of contents

Extending the model

- The model of Choo and Siow (2006) is appealing but simple
- We look at an important extension by Dupuy and Galichon (2014)
- Their model allows for *continuous* types
  - discretizing continuous attributes is not always appealing
  - assume that we want to look at assortative matching on *height*
  - do we round height to the nearest 20cm or 1cm?
  - important because of the iid assumption on the preference shocks
- They introduce a new technique to determine the *most relevant dimensions* of sorting
  - which they call saliency analysis

Preferences

- We still have preferences of the form:

$$U^i_{IJ} = \tilde{\alpha}_{IJ} - \tau_{IJ} + \epsilon^i_{IJ} \tag{15}$$

$$V^j_{IJ} = \tilde{\gamma}_{IJ} + \tau_{IJ} + \eta^j_{IJ} \tag{16}$$

- Compared to the discrete types model, two aspects change:
  - we now assume that types (I, J) are measured continuously
  - we now assume that idiosyncratic shocks follow *continuous* logit processes

Shocks

- The continuous logit shocks have an intuitive interpretation
- Choices can be represented as a two-stage process:
  - first, each man i draws a set of *acquaintances* $\mathcal{J}_i$, elements of which are indexed by k
  - each acquaintance is associated with a *sympathy shock* ϵ^i_k, distributed iid Gumbel
  - then each man chooses his preferred partner from these acquaintances by solving:

$$\max_{k \in \mathcal{J}_{ik}} \{\tilde{\alpha}_{IJ_k} - \tau_{IJ_k} + \epsilon^i_k\} \tag{17}$$

Continuous demands

- The probability that a type I man chooses a type J woman ($=$ demand) is:

$$\mu_{J|I} = \frac{\exp\left(\tilde{\alpha}_{IJ} - \tau_{IJ}\right)}{\int_J \exp(\tilde{\alpha}_{IJ} - \tau_{IJ})} \tag{18}$$

- Similarly, we have the probability that a type J women chooses a type I man:

$$\mu_{I|J} = \frac{\exp\left(\tilde{\gamma}_{IJ} + \tau_{ij}\right)}{\int_I \exp(\tilde{\gamma}_{IJ} + \tau_{IJ})} \tag{19}$$

Matching function

- We can again equate supply and demand to obtain a matching function:

$$\frac{\mu_{IJ}}{\mu_{I0}\mu_{0J}} = \exp\left(\frac{\alpha_{IJ} + \gamma_{IJ}}{2}\right) = \Pi_{IJ} \tag{20}$$

- Before, we would estimate the surplus non-parametrically
  - simply using data on matches within each combination of discrete types
- Clearly we need a different approach when types are continuous
  - cannot estimate Π as a matrix

Parameterizing the surplus

- Following Dupuy and Galichon (2014) we first parameterize the surplus quadratically as:

$$\Pi_{IJ} = I'AJ = \sum_{I_k, J_l} A_{kl} \times I_k \times J_l \tag{21}$$

- Entry A_{kl} of affinity matrix captures complementarities between characteristics I_k and J_l
- It can be estimated using MLE based on the following log-likelihood function:

$$\mathcal{L}(A) = \frac{1}{N} \sum_{n=1}^{N} \mu_n^A \tag{22}$$

  - where μ_n^A is given by the market clearing conditions for a given value of A

Saliency analysis

- Suppose that we have the estimate for the affinity matrix A in hand
- We now want to determine how important these different dimensions are for sorting
- We can perform a *saliency analysis* as introduced in Dupuy and Galichon (2014)
- This consists of a singular value decomposition of the affinity matrix and testing its rank

Saliency analysis (cont.)

- Think of Singular Value Decomposition (SVD) as a data-reduction tool
- Formally an SVD of a matrix A is:
$$A = U'\Lambda V \tag{23}$$
    - with Λ a diagonal matrix that captures the importance of each dimension
    - with U and V matrices of loadings that describe the *nature* of each dimension
- Technicalities are not the main point – focus is on interpretation

Example – Affinity matrix

Figure: Affinity Matrix – from Dupuy and Galichon (2014)

TABLE 3
ESTIMATES OF THE AFFINITY MATRIX: QUADRATIC SPECIFICATION (N = 1,158)

	Wives									
Husbands	Education	Height	BMI	Health	Conscientiousness	Extraversion	Agreeableness	Emotional Stability	Autonomy	Risk Aversion
Education	.56*	.02	−.08	.02	−.04	−.01	−.03	−.04	.05	−.02
Height	.01	.18*	.04	−.01	−.04	.05	.02	.02	.02	.02
BMI	−.05	.05	.21*	.01	.06	.00	−.04	.04	−.01	−.01
Health	−.07	.00	−.06	.14*	−.04	.05	−.04	.04	.02	.00
Conscientiousness	−.06	−.03	.07	.00	.14*	.07	.04	.06	−.02	−.01
Extraversion	.01	−.02	.05	.02	−.06	.02	−.02	−.01	−.03	−.05
Agreeableness	.00	.01	−.08	.02	.13*	−.14*	.02	.11	−.09	−.04
Emotional stability	.03	.00	.12*	.04	.21*	.05	−.03	−.04	.08	.01
Autonomy	.02	.00	.00	.01	−.11*	.11*	−.04	.03	−.09	.01
Risk aversion	.00	.02	−.03	.02	.01	−.01	−.01	−.05	.05	.11*

* Significant at the 5 percent level.

Example – Saliency analysis

Figure: Saliency Analysis – from Chiappori *et al.* (2024)

Table 4: Saliency analysis (Sample 1)

	Men		Women	
	Index 1	Index 2	Index 1	Index 2
Education	**0.21**	**0.93**	**0.12**	**0.92**
	(0.02)	(0.02)	(0.02)	(0.02)
Age	**0.97**	**−0.23**	**0.99**	**−0.12**
	(0.01)	(0.02)	(0.00)	(0.02)
Height	**0.12**	**0.28**	**0.08**	0.14
	(0.03)	(0.06)	(0.03)	(0.07)
BMI	**0.06**	−0.04	0.01	**−0.36**
	(0.02)	(0.05)	(0.02)	(0.05)
Index share	**0.74**	**0.17**	**0.74**	**0.17**
	(0.07)	(0.02)	(0.07)	(0.02)

Notes. The table reports men's and women's singular vectors, V and U respectively, and singular values, $diag(\Lambda)$, from the singular value decomposition of $\hat{A} = U'\Lambda V$. We report standard errors in parentheses; they are obtained with 1,000 bootstrap replications (Milan and Whittaker, 1995). Boldfaced estimates are significant at the 5% level. In the last line, each value of $diag(\Lambda)$ can be interpreted as the relative importance of each sorting dimension.

Bibliography I

Chiappori, P.-A., Ciscato, E. and Guerriero, C. (2024). Analyzing matching patterns in marriage: Theory and application to Italian data. *Quantitative Economics*, **15** (3), 737–781.

Choo, E. and Siow, A. (2006). Who marries whom and why. *Journal of Political Economy*, **114** (1), 175–201.

Dupuy, A. and Galichon, A. (2014). Personality traits and the marriage market. *Journal of Political Economy*, **122** (6), 1271–1319.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, **3** (4), 303–328.