We consider the one-to-one matching model with transfers of Choo and Siow (2006) and Galichon and Salanié (2015). When the analyst has data on one large market only, we study identification of the preference parameters without imposing parametric restrictions on the probability distribution of the agents’ unobserved characteristics. We provide a tractable characterisation of the sharp identified set for the preference parameters and discuss inference, under various classes of nonparametric distributional assumptions on the agents’ unobserved characteristics. Simulations suggest that the model can be informative about the sign and magnitude of the preference parameters. We use our methodology to empirically investigate if the variations in marriage matching patterns observed over time in the U.S. are caused by changes in the agents’ preferences for education assortativeness or by a shift in the proportion of educated women.

**Keywords**: One-to-One Matching, Transfers, Stability, Partial Identification, Nonparametric Identification, Linear Programming.

---

*We would like to thank Pierre-André Chiappori and Thierry Magnac for their useful comments. We would also like to thank the seminar and conference participants at Toulouse School of Economics, the University of Warwick, IAAE 2019, and EEA-ESEM 2019. Rossi Abi-Rafeh provided excellent research assistance. We acknowledge funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d’Avenir) program, grant ANR-17-EURE-0010.*

†Email: cristina.gualdani@tse-fr.eu, Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France.

‡Email: shruti.sinha@tse-fr.eu, Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France.
1 Introduction

Matching markets are two-sided markets, where agents on each side of the market have preferences over matching with agents on the other side. For example, students are allocated to schools, venture capitalists choose which start-ups to fund, social interactions lead individuals to find marital partners, production tasks are assigned to workers, kidneys are matched with dialysis patients, and auctions sort buyers with sellers. While the economic theory of matching models has been around for more than five decades, it is only recently that there has been a growing interest in the empirical models of matching (Chiappori and Salanié, 2016).

One of the predominant strands of the literature focuses on the econometrics of one-to-one matching models with transfers when the analyst has data on one large market only (e.g., Choo and Siow, 2006; Graham, 2013; Dupuy and Galichon, 2014; Galichon and Salanié, 2015; Fox, 2018). Many empirical applications can be studied in this framework, e.g., sorting of CEOs to firms (Gabaix and Landier, 2008; Tervio, 2008), sorting of job openings to workers, and the marriage market (Choo and Siow, 2006; Hitsch, Hortaçsu and Ariely, 2009; Chiappori, Salanié, and Weiss, 2017). In particular, Choo and Siow (2006) and Galichon and Salanié (2015) show that, if the agents belong to a finite number of observed types, point identification of the systematic components of the agents’ preferences (hereafter, preference parameters) is achieved under the assumption that the probability distribution of the agents’ unobserved characteristics (alternatively called taste shocks) is completely known by the researcher. In applied work, this typically amounts to fixing a parametric family for the taste shocks together with numerical values for all of its parameters. However, such an approach may raise concerns because parametric distributional assumptions on the taste shocks could overshadow the informativeness of the data. Therefore, our objective is to investigate identification and inference of the preference parameters when the researcher does not impose parametric restrictions on the probability distribution of the taste shocks. Instead, we allow for the inclusion of nonparametric distributional assumptions, like symmetric marginals, stochastic independence of the agents’ types, and much more. We believe that performing this robust identification exercise is important because it permits the researcher to epistemologically evaluate to what extent the empirical content of the model considered depends on the way in which the analyst has parameterised the agents’ unobserved heterogeneity.

Other papers in the literature examine identification in one-to-one matching models with transfers without incorporating parametric distributional assumptions on the taste shocks (e.g., Fox, 2010; Fox, Yang, and Hsu, 2018; Sinha, 2018). Their arguments exploit

---

1 One-to-one matching refers to the case where each agent on each side can either be matched with exactly one agent on the other side, or remain unmatched. In models with transfers agents can transfer part of their utility to their matched partner without frictions.

2 See Fox (2009) for a review.
variation across many i.i.d. markets. As data on many i.i.d. markets may not be always available, here we let the analyst collect observations from one large market only.

When the researcher has data on one large market only, not specifying the probability distribution of the taste shocks raises the possibility of partial identification of the preference parameters. Consequently, this poses the challenge of tractably characterising the region of preference parameter values that exhausts all the implications of the model and data (i.e., the sharp identified set for the preference parameters), under various classes of nonparametric distributional assumptions on the taste shocks. Our study aims to answer this methodological question.

Constructing the sharp identified set for the preference parameters requires facing two problems. First, for a given value of the preference parameters, the analyst has to find a probability distribution of the taste shocks that yields that hypothesised value. This amounts to solving an infinite-dimensional existence problem because every possible probability distribution of the taste shocks is not finitely parameterised. Second, the analyst should repeat the first step for every admissible value of the preference parameters. Usually, this is done in the partial identification literature by generating a grid of points to approximate the preference parameter space and then repeating the exercise of interest for each grid point. The difficulty of implementing such an approach increases with the size of the grid, which in turn, increases exponentially with the number of the agents’ types, hence leading quickly to a computational bottleneck.

We address the first issue by using Theorem 1 in Torgovitsky (2018) (also known as PIES\textsuperscript{3}) for both sides of the market. The theorem provides sufficient and necessary conditions for the existence of a probability distribution of the taste shocks that yields a given hypothesised value of the preference parameters. The sufficient and necessary conditions constitute a linear system of equalities and inequalities, which is a tractable and well-understood problem. We address the second issue by showing that the preference parameter space can be ex-ante partitioned into a finite number of convex subsets such that, for each subset, every value belonging to that subset gives rise to the same linear programming problem defined in the first step. Therefore, the researcher has to solve the linear programming problem only once for each subset. Overall, the procedure designed allows to feasibly construct the sharp identified set for the preference parameters under several classes of nonparametric distributional assumptions on the taste shocks. Moreover, after having reformulated the identifying restrictions as moment equalities, we explain how inference on the sharp identified set for the preference parameters can be conducted by applying a computationally tractable version of the profiled subsampling technique illustrated by Politis and Romano (1994), Romano and Shaikh (2008; 2012).

Simulations suggest that the model is informative about the sign of the preference parameters, under various classes of nonparametric distributional restrictions on the taste shocks.

\textsuperscript{3}That is, Partial Identification by Extending Subdistributions.
shocks. In addition, under some such classes, we obtain relatively tight bounds. Simulations also highlight that our procedure is useful to compare and contrast the identification power of several nonparametric distributional assumptions across different data generating processes.

Further, the methodology developed enables the researcher to answer certain relevant empirical questions. For example, we investigate if the variations in marriage matching patterns observed over time in the U.S. are caused by changes in the agents’ preferences for education assortativeness, or by a shift in the proportion of educated women. The application framework and the data are taken from Chiappori, Salanié, and Weiss (2017). In contrast with that study, we proceed without relying on parametric distributional assumptions for the taste shocks, and thus we get conclusions that are possibly more robust. Under a symmetry restriction on the marginals of the taste shocks, we are unable to reject the null hypothesis that the matching preferences have been invariant over time. Instead, when we impose stronger nonparametric distributional assumptions on the taste shocks, we reject such a null.

This paper is related to the literature on the econometrics of matching models. The literature is split into several strands depending on the preference structures of the agents (transferable utility (TU) models, non-transferable utility (NTU) models, or imperfectly-transferable utility (ITU) models); the maximum number of links an agent is permitted to form across sides (one-to-one, one-to-many, many-to-many); and/or the assignment allocation is centralised or decentralised. Seminal papers on TU matching models have been cited earlier. Identification in the NTU framework has been studied by Sørensen (2007), Dagsvik (2000), Menzel (2015), and Agarwal and Diamond (2017). The ITU framework has been introduced by Galichon, Kominers, and Weber (2018).

Moreover, the matching model we consider can be equivalently rewritten as two one-sided multinomial choice models linked via market-clearing transfers (Galichon and Salanié, 2015). Therefore, our project is related to the literature on nonparametric and semiparametric identification of multinomial choice models, e.g., Manski (1975), Matzkin (1993), and Fox (2007). Fox (2018) establishes point identification of the preference parameters in two-sided markets under the assumption that the taste shocks are exchangeable and the availability of a continuous regressor with large support. Furthermore, in the absence of a special regressor, Fox (2018) characterises an identified region (possibly, not sharp) that can be estimated. With respect to Fox (2018), we focus on the sharp identified set for the preference parameters and thus obtain the tightest possible bounds implied by the model. In addition, our procedure does not require exchangeability and applies to various classes of nonparametric distributional assumptions on the taste shocks.

Finally, this paper is also related to the literature on partial identification in applied

---

4For example, positive assortativeness means that more (less) educated men want to match with more (less) educated women.
research (see Ho and Rosen, 2017 for a review) and to the literature on assessing the sensitivity of empirical conclusions (e.g., Kline and Santos, 2013).

In what follows, Section 2 introduces the matching model. Section 3 develops identification arguments. Sections 4 illustrates the construction of the sharp identified set through simulations. Section 5 discusses inference. Section 6 presents the empirical application. Section 7 concludes.

2 The model

We focus on the empirical one-to-one matching model with transferable utilities, as it has been previously studied in Choo and Siow (2006) and Galichon and Salanié (2015). Specifically, we consider one market composed of two sides. On each side there is a continuum of agents. Every agent on each side has preferences over the set of all agents on the other side. Further, every agent on each side can at most be matched with one agent on the other side. We assume that utilities are transferable. Transfers act as prices which are determined in equilibrium simultaneously with the match assignment such that each agent maximises her own payoff and the market clears. The two sides of the market are stochastically independent and the matching is frictionless. Many empirical applications can be studied in this framework, e.g., sorting of CEOs to firms, sorting of job openings to workers, and the marriage market.

For simplicity of exposition, we refer to the agents in the first side of the market as men and to the agents in the second side of the market as women, but our results are clearly not restricted to the marriage market. Let $I$ be the set of men and $J$ be the set of women. We index individual men by $i \in I$ and individual women by $j \in J$. The outside option to remain unmatched is indicated by “0”, so that single agents are represented as being matched with “0”. Lastly, we define $I_0 \equiv I \cup \{0\}$ and $J_0 \equiv J \cup \{0\}$.

Each man $i \in I$ is endowed with some characteristics, $X_i$, whose probability distribution is denoted by $P_X$. Similarly, each woman $j \in J$ is endowed with some characteristics, $Y_j$, whose probability distribution is denoted by $P_Y$. The supports of $X_i$ and $Y_j$ are finite and indicated by $X$ and $Y$, respectively. As earlier, $X_0 \equiv X \cup \{0\}$ and $Y_0 \equiv Y \cup \{0\}$. $X_i$ and $Y_j$ are typically called man $i$’s type and woman $j$’s type, respectively. The realisations of $X_i$ and $Y_j$ are observed by the researcher.

Let $\tilde{\Phi}_{ij}$ be the match surplus generated when the couple $(i, j) \in I \times J$ is formed. If the couple $(i, j)$ is of type $(x, y) \in X \times Y$, then the match surplus is defined as

$$\tilde{\Phi}_{ij} \equiv \Phi_{xy} + \epsilon_{iy} + \eta_{xj},$$

(1)

where $\Phi_{xy}$ is the type-specific match surplus and $\{\epsilon_{iy}, \eta_{xj}\}$ are continuously distributed.

---

5This restriction is not crucial for our discussion and can be relaxed. See Footnote 12 in Section 3.1.
taste shocks assigned to the agents by nature and whose realisations are unobserved by the researcher. The payoffs that man $i \in I$ of type $x \in X$ and woman $j \in J$ of type $y \in Y$ get when staying single are defined as

$$\tilde{\Phi}_{i0} \equiv \Phi_{x0} + \epsilon_{i0},$$

and

$$\tilde{\Phi}_{0j} \equiv \Phi_{0y} + \eta_{0j}. \quad (3)$$

We denote the probability distribution of $\epsilon_i \equiv (\epsilon_{iy} \forall y \in Y_0)$ conditional on $X_i$ and the probability distribution of $\eta_j \equiv (\eta_{xj} \forall x \in X_0)$ conditional on $Y_j$ by $P_{\epsilon|X}$ and $P_{\eta|Y}$, respectively. \{\epsilon_i, \eta_j\} are introduced to reconcile the mechanical predictions of theory and the fuzziness of the actual data. For instance, with a supermodular match surplus, the matching assignment should be exactly assortative, which is never observed in practice.

As in Dagsvik (2000), Choo and Siow (2006), and Galichon and Salanié (2015), the latent heterogeneity in equation (1) does not consist of an $ij$-indexed term, say $\nu_{ij}$. Instead, it is modelled through an additively separable term, $\epsilon_{iy} + \eta_{xj}$ (hereafter referred to as separability). This means that when the researcher observes man $i \in I$ of type $x \in X$ matched with woman $j \in J$ of type $y \in Y$, it could be because $j$ has a strong unobserved preference for men of type $x$, or because $i$ has a strong unobserved preference for women of type $y$. It could also occur because $j$ has unobserved features that attract men of type $x$, or because $i$ has unobserved features that attract women of type $y$. However, simultaneous sorting on unobservables is ruled out. For example, it cannot be that $j$ has some unobserved preference for unobserved features of $i$ or vice versa. Ultimately, note that the latent terms \{\epsilon_{iy}, \eta_{xj}\} contribute to the match surplus and not necessarily to the preferences of a particular side of the market.

There are three main reasons for imposing separability. First, in the presence of $ij$-indexed latent variables without bounded support, the model could predict strange equilibrium outcomes, such as payoffs tending to infinity, no singles, and a match assignment almost based exclusively on the realisation of the taste shocks.\footnote{Menzel (2015) provides restrictions on the structure of the agents’ preferences that address similar concerns arising in non-transferable utility matching models. However, it is not obvious that the same insights can be applied to transferable utility matching models. Exploring this is outside the scope of our current study.} Second, if the agents had to choose from a continuum of potential partners, then the frictionless assumption would be hard to justify (Section 4.1.2 in Chiappori, 2017). Third, from a technical point of view, separability allows to transform the present framework into two one-sided multinomial choice models linked via market-clearing transfers, as shown by Proposition 1 in Galichon and Salanié (2015). This alternative representation of the problem guides the identification analysis, as explained in Section 3.

A competitive equilibrium of the model is characterised by a match assignment and a
match surplus sharing rule. A match assignment is a description of who is matched with whom, and a match surplus sharing rule tells us how the total match surplus is divided among the matched agents. Such division of surplus relies on endogenously determined transfers ensuring that every agent maximises her utility and the market clears.\(^7\) As per Shapley and Shubik (1972) in one-to-one matching models with transfers, a competitive equilibrium coincides with a stable matching. That is, a competitive equilibrium is such that no agent has an incentive to deviate from her current match.

In order to properly describe the equilibrium concept, we introduce the formal definitions of match assignment and match surplus sharing rule. Let \(\tilde{\mu}_{ij}\) be equal to 1 if man \(i \in I\) and woman \(j \in J\) are matched and zero otherwise. Let \(\tilde{\mu}_{i0}\) be equal to 1 if man \(i \in I\) is single and 0 otherwise. Let \(\tilde{\mu}_{0j}\) be equal to 1 if woman \(j \in J\) is single and 0 otherwise. The vector
\[
\tilde{\mu} \equiv (\tilde{\mu}_{ij} \forall (i,j) \in I_0 \times J_0 \setminus \{(0,0)\}),
\]
represents a match assignment. Let \(\tilde{U}_i\) and \(\tilde{V}_j\) be the payoffs gained by man \(i \in I\) and woman \(j \in J\) under the match assignment \(\tilde{\mu}\). The vectors
\[
\tilde{U} \equiv (\tilde{U}_i \forall i \in I), \quad \tilde{V} \equiv (\tilde{V}_j \forall j \in J),
\]
represent a match surplus sharing rule and they implicitly embed transfers across the two sides of the market.

**Definition 1.** *(Stable matching)* \((\tilde{\mu}, \tilde{U}, \tilde{V})\) is a stable matching if it satisfies three properties:

1. \[\int_{I_0} \tilde{\mu}_{ij} dP_W = 1 \forall i \in I \text{ and } \int_{J_0} \tilde{\mu}_{ij} dP_M = 1 \forall j \in J,\] where \(P_M\) and \(P_W\) denote the total measure of men and the total measure of women, respectively.
2. \[\tilde{U}_i + \tilde{V}_j \geq \tilde{\Phi}_{ij} \forall (i,j) \in I \times J.\]
3. \[\tilde{U}_i \geq \tilde{\Phi}_{i0} \forall i \in I \text{ and } \tilde{V}_j \geq \tilde{\Phi}_{0j} \forall j \in J.\]

\(\diamond\)

The first part of Definition 1 states that the match assignment is feasible in the sense of one-to-one, that is, every agent on each side can either be matched with exactly one agent on the other side, or remain unmatched. The second and third parts state that there is no man and woman that can get a strictly higher match surplus by deviating from their matches under \(\tilde{\mu}\). Moreover, a stable matching exists and is generically unique.\(^8\) Hence, from now on we will refer to it as *the* stable matching.

---

\(^7\)A competitive equilibrium can be equivalently characterised by a match assignment and a transfer scheme.

\(^8\)For more details on existence see Shapley and Shubik (1972). Further, in a large market with a continuum of agents on both sides, Gretsky, Ostroy and Zame (1992) show generic uniqueness.
3 Identification

We assume that the market has already reached the stable matching. In other words, as the researcher collects more data, the asymptotic fiction is that the researcher learns more about the already established stable matching without altering it.

Let \( Q_{M,i} \) and \( Q_{W,j} \) represent the type of woman matched with man \( i \in I \) and the type of man matched with woman \( j \in J \), respectively, at the equilibrium. Let \( P_{Q_M|X} \) denote the probability distribution of \( Q_{M,i} \) conditional on \( X_i \). Let \( P_{Q_W|Y} \) denote the probability distribution of \( Q_{W,j} \) conditional on \( Y_j \). We require the probability distributions \( \{ P_{Q_M|X}, P_{Q_W|Y}, P_X, P_Y \} \) to be nonparametrically identified by the sampling process. Hence, we expect the researcher to collect aggregate data on matching patterns in one large market. For example, it is sufficient to have a random sampling scheme in which the researcher draws agents from each side of the market and records the matched types. Lastly, we assume that the researcher does not observe equilibrium transfers. However, our discussion remains valid when data on equilibrium transfers are available, in which case we have additional information to incorporate in the analysis.

The parameter of interest for identification is the vector of type-specific match surpluses,

\[
\Phi \equiv (\Phi_{xy} \forall (x,y) \in \mathcal{X} \times \mathcal{Y} \setminus \{(0,0)\}).
\]

Indeed, identifying and conducting inference on \( \Phi \) is essential to understand the relative importance of the agents’ observed characteristics in determining the match preferences. This can be useful, for example, to test if the agents’ preferences for assortativeness change over time.

A critical result that our identification arguments rely on is Proposition 1 in Galichon and Salanié (2015). This proposition is based on the separability assumption of the latent terms introduced earlier.

**Proposition 1.** (Galichon and Salanié, 2015) Given the collection of primitives \( \{ \Phi, P_X, P_Y, P_{\epsilon|X}, P_{\eta|Y} \} \) generating the stable matching \( (\tilde{\mu}, \tilde{U}, \tilde{V}) \), there exists one and only one pair of vectors

\[
U \equiv (U_{xy} \forall (x,y) \in \mathcal{X} \times \mathcal{Y}), \quad V \equiv (V_{xy} \forall (x,y) \in \mathcal{X} \times \mathcal{Y}),
\]

such that

\[
\tilde{U}_i = \max_{y \in \mathcal{Y}_0}(U_{xy} + \epsilon_{iy}) \quad \forall i \in I \text{ of type } x \in \mathcal{X}, \forall x \in \mathcal{X},
\]

\[
\tilde{V}_j = \max_{x \in \mathcal{X}_0}(V_{xy} + \eta_{xj}) \quad \forall j \in J \text{ of type } y \in \mathcal{Y}, \forall y \in \mathcal{Y},
\]

\[
U_{xy} + V_{xy} = \Phi_{xy}, \quad U_{x0} = \Phi_{x0}, \quad V_{0y} = \Phi_{0y} \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}.
\]

\[\diamondsuit\]
Proposition 1 allows to rewrite the framework of Section 2 as two separate one-sided multinomial choice models linked by market-clearing transfers implicitly embedded into the vectors $U$ and $V$. Indeed, Proposition 1 claims that the probability distributions \( \{ P_{Q_M|X}, P_{Q_W|Y} \} \) are as if generated by the following model:

\[
\begin{align*}
Q_{M,i} &= \arg\max_{y \in Y} (U_{xy} + \epsilon_{iy}) \quad \forall i \in I \text{ of type } x \in X, \ \forall x \in X, \quad (4) \\
Q_{W,j} &= \arg\max_{x \in X} (V_{xy} + \eta_{xj}) \quad \forall j \in J \text{ of type } y \in Y, \ \forall y \in Y, \quad (5) \\
U_{xy} + V_{xy} &= \Phi_{xy}, \ \forall (x, y) \in X \times Y, \quad (6) \\
\epsilon_i|X_i \sim P_{\epsilon|X}, \ \eta_j|Y_j \sim P_{\eta|Y} \quad \forall i \in I, \forall j \in J, \quad (7)
\end{align*}
\]

where “\( \sim \)” denotes “distributed as” and the equality sign in front of the argmax operator is because zero probability events. Such alternative representation of the problem is useful as it immediately suggests a way to investigate identification of $\Phi$: the researcher can study identification of $U$ and $V$ from (4) and (5) using various restrictions on $P_{\epsilon|X}$ and $P_{\eta|Y}$ in (7), and then obtain identification results for $\Phi$ through (6).

Along these lines, Choo and Siow (2006) show that if $\Phi_{x0} = \Phi_{0y} = 0 \ \forall (x, y) \in X \times Y^9$, the taste shocks are i.i.d. Gumbel with scale 1 and location 0, and the taste shocks are independent of the agents’ types, then $\Phi$ is point identified via standard Logit arguments applied to each side of the market. The result is generalised by Galichon and Salanié (2015) who prove that if $P_{\epsilon|X}$ and $P_{\eta|Y}$ are fully known, then $\Phi$ is point identified.

Choo and Siow (2006) and Galichon and Salanié (2015) highlight that point identification of $\Phi$ crucially relies on the assumption that the conditional probability distribution of the taste shocks is completely known by the researcher, \emph{a priori}. It typically amounts to fixing a parametric family for $P_{\epsilon|X}$ and $P_{\eta|Y}$ together with numerical values for all of its parameters. This leads to obvious concerns because wrong specifications can induce inconsistent empirical results, as widely documented by the econometric literature on binary and multinomial choice models (e.g., Manski, 1975, 1985, 1988; Matzkin, 1992, 1993; Fox, 2007). Therefore, our objective is to investigate identification and inference on $\Phi$ when the researcher does not impose parametric restrictions on $P_{\epsilon|X}$ and $P_{\eta|Y}$. At the same time, we allow for the possibility to include nonparametric distributional assumptions like symmetric marginals, stochastic independence of the agents’ types, and much more.

We believe that performing such an exercise (hereafter referred to as the \emph{nonparametric exercise}) is important because it permits to obtain more robust conclusions on $\Phi$. Moreover, it helps us to do an epistemological evaluation of how the empirical content of the model depends on the way in which the analyst has parameterised the conditional probability distribution of the latent variables.

\footnote{Following the multinomial choice literature, these restrictions are location normalisations.}
Implementing the nonparametric exercise is non-trivial. Indeed, avoiding parametric restrictions on $P_{i|X}$ and $P_{i|Y}$ raises the possibility of partial identification of $\Phi$. Consequently, this poses the issue of tractably characterising and conducting inference on the sharp identified set for $\Phi$ (that is, the region of parameter values that exhausts all the implications of the model and data) under various classes of nonparametric distributional assumptions on the taste shocks. The next Sections 3.1 and 3.2 address the identification challenge, while Section 5 discusses inference.

There are other seminal papers in the literature considering related econometric questions when the analyst has data on one large market. A first group of works focuses on one-sided markets. For example, Manski (1975) and Matzkin (1993) show point identification of the payoff parameters in a multinomial choice model by combining in different ways the conditional i.i.d.-ness of the taste shocks, the presence of a continuous regressor with large support, and an index structure or shape restrictions on the systematic payoff components. Note that the model in Section 2 does not contemplate the possibility of having a continuous regressor and does not impose an index structure or shape restrictions on the systematic payoff components. Fox (2007) proves that the conditional exchangeability of the taste shocks can replace the conditional i.i.d.-ness because it is sufficient for the rank property to hold. More recently, researchers have started to extend these results to two-sided markets, e.g., Fox (2018).

### 3.1 The sharp identified set for $\Phi$

We start with introducing additional notation and then we define the sharp identified set for $\Phi$. Without loss of generality and to keep the exposition readable, let $\mathcal{X} = \mathcal{Y} = \{1, ..., r\}$ with $r \in \mathbb{N}$ and let $\mathbb{R}^{r+1}$ be the support of $\epsilon_i$ conditional on $X_i$ and of $\eta_j$ conditional on $Y_j$. Bearing in mind that in multinomial choice models what matters are differences in utilities, let

$$
\Delta \epsilon_i \equiv (\epsilon_{i1} - \epsilon_{i0}, ..., \epsilon_{ir} - \epsilon_{i0}, \epsilon_{i1} - \epsilon_{i2}, ..., \epsilon_{i1} - \epsilon_{ir}, \epsilon_{i2} - \epsilon_{i3}, ..., \epsilon_{i2} - \epsilon_{ir}, ..., \epsilon_{ir-1} - \epsilon_{ir}),
$$

and

$$
\Delta \eta_j \equiv (\eta_{1j} - \eta_{0j}, ..., \eta_{rj} - \eta_{0j}, \eta_{1j} - \eta_{2j}, ..., \eta_{1j} - \eta_{rj}, \eta_{2j} - \eta_{3j}, ..., \eta_{2j} - \eta_{rj}, ..., \eta_{r-1j} - \eta_{rj}),
$$

be the vectors of differences between every pairs of taste shocks for each side of the market, with length $d \equiv \binom{r+1}{2}$. Note that the first $r$ components of $\Delta \epsilon_i$ and $\Delta \eta_j$ can be arbitrary, while the remaining $(d - r)$ components are linear combination of the first $r$
components. Hence, $\Delta \epsilon_i$ and $\Delta \eta_j$ take values in the region

$$
\mathcal{B} \equiv \{(b_1, b_2, ..., b_d) \in \mathbb{R}^d : \ b_{r+1} = b_1 - b_2, b_{r+2} = b_1 - b_3, ..., b_{2r-1} = b_1 - b_r, \\
\quad \quad \quad \quad \quad \quad b_{2r} = b_2 - b_3, ..., b_{3r-3} = b_2 - b_r, ..., \\
\quad \quad \quad \quad \quad \quad b_d = b_{r-1} - b_r \}.
$$

The definition of the sharp identified set for $\Phi$ follows naturally by exploiting Proposition 1.

**Definition 2.** (Sharp identified set) Let $\Theta^\dagger$, $U^\dagger$, and $V^\dagger$ be the sets of admissible values of $\Phi$, $U$, and $V$, respectively. Let $\mathcal{P}_{\Delta \epsilon | X}^\dagger$ and $\mathcal{P}_{\Delta \eta | Y}^\dagger$ be the function spaces of admissible $d$-dimensional conditional probability distributions of the taste shock differences, $P_{\Delta \epsilon | X}$ and $P_{\Delta \eta | Y}$, respectively, which can include parametric and/or nonparametric restrictions.

The sharp identified set for $\Phi$ is

$$
\Theta^* \equiv \{ \Phi \in \Theta^\dagger : \exists U \in U^\dagger, V \in V^\dagger, P_{\Delta \epsilon | X} \in \mathcal{P}_{\Delta \epsilon}^\dagger, P_{\Delta \eta | Y} \in \mathcal{P}_{\Delta \eta}^\dagger \text{ s.t.} \\
P_{Q_M | x}(y) = \omega_{M,y | x}(U, P_{\Delta \epsilon | x}) \quad \forall (x, y) \in X \times Y_0 \\
P_{Q_W | y}(x) = \omega_{W,x | y}(V, P_{\Delta \eta | y}) \quad \forall (x, y) \in X_0 \times Y \\
U_{xy} + V_{xy} = \Phi_{xy}, \quad U_{x0} = \Phi_{x0}, \quad V_{0y} = \Phi_{0y} \quad \forall (x, y) \in X \times Y \\
P_{\Delta \epsilon | x}(\mathcal{B}) = 1, \quad P_{\Delta \eta | y}(\mathcal{B}) = 1 \quad \forall (x, y) \in X \times Y \},
$$

(9)

where $\omega_{M,y | x}$ and $\omega_{W,x | y}$ are known functions derived from (4) and (5).

Equivalently,

$$
\Theta^* = \{ \Phi \in \Theta^\dagger : \exists U \in U^*, V \in V^* \text{ s.t.} \\
U_{xy} + V_{xy} = \Phi_{xy}, \quad U_{x0} = \Phi_{x0}, \quad V_{0y} = \Phi_{0y} \forall (x, y) \in X \times Y \},
$$

(10)

where

$$
U^* \equiv \{ U \in U^\dagger : \exists P_{\Delta \epsilon | X} \in \mathcal{P}_{\Delta \epsilon}^\dagger \text{ s.t.} \\
P_{Q_M | x}(y) = \omega_{M,y | x}(U, P_{\Delta \epsilon | x}) \text{ and } P_{\Delta \epsilon | x}(\mathcal{B}) = 1 \forall (x, y) \in X \times Y_0 \},
$$

$U^\dagger$ and $V^\dagger$ can incorporate scale and location normalisations to ensure that the volume of the sharp identified set for $\Phi$ is not improperly inflated relative the the point identified case. See Section 4 for more on this.
and

\[ V^* = \left\{ V \in V^T : \exists P_{\Delta \eta|Y} \in \mathcal{P}_{\Delta \eta} \text{ s.t.} \right. \\
\left. P_{Q_{W|Y}}(x) = \omega_{W,x|y}(V, P_{\Delta \eta|y}) \quad \text{and} \quad P_{\Delta \eta|y}(B) = 1 \forall (x, y) \in \mathcal{X}_0 \times \mathcal{Y} \right\}. \]

Definition 2 uses (4), (5), and (6) to define the sharp identified set for \( \Phi \). In particular, for a given \( \{U, V, P_{\Delta x|X}, P_{\Delta \eta|Y}\} \), the first and second equations of (9) impose that the probability distributions \( \{P_{Q_{M|X}}, P_{Q_{W|Y}}\} \) coincide with the conditional match type probabilities as derived from the maximisation problems, (4) and (5). The third equation of (9) mimics (6). Finally, the fourth equation of (9) requires the conditional probability distributions of the taste shock differences to be concentrated on the region \( B \) (hereafter referred to as the degeneracy condition).\(^\dagger\) Equivalently, we can consider identification separately on each side of the market and then back out \( \Phi \) using market-clearing transfer conditions, as done in (10). One may think of many other ways to define \( \Theta^* \). We have provided two representations that are pedagogical for the procedure described below.\(^\ddagger\)

To clarify Definition 2, Example 1 provides an explicit characterisation of \( \Theta^* \) when \( r = 2 \) (\( d = 3 \)).

**Example 1.** When \( r = 2 \) (\( d = 3 \)), the vectors of taste shock differences, \( \Delta \epsilon_i \) and \( \Delta \eta_j \), are

\[ \Delta \epsilon_i \equiv (\epsilon_{i1} - \epsilon_{i0}, \epsilon_{i2} - \epsilon_{i0}, \epsilon_{i1} - \epsilon_{i2}), \]

\[ \Delta \eta_j \equiv (\eta_{ij} - \eta_{0j}, \eta_{2j} - \eta_{0j}, \eta_{1j} - \eta_{2j}), \]

and the region \( B \) is given by

\[ B \equiv \{(b_1, b_2, b_3) \in \mathbb{R}^3 : b_3 = b_1 - b_2\} . \]

Moreover, from (4), we can derive the probability that a man of type \( x \in \mathcal{X} \) chooses a woman of type 1,

\[ \Pr(U_{x1} + \epsilon_{x1} \geq U_{x0} + \epsilon_{x0}, U_{x1} + \epsilon_{x1} \geq U_{x2} + \epsilon_{x2} | X = x) \]
\[ = \Pr(\epsilon_{x1} - \epsilon_{x0} \geq U_{x0} - U_{x1}, \epsilon_{x1} - \epsilon_{x2} \geq U_{x2} - U_{x1} | X = x) \]
\[ = \Pr(\epsilon_{x1} - \epsilon_{x0} \geq U_{x0} - U_{x1}, \epsilon_{x2} - \epsilon_{x0} \geq -\infty, \epsilon_{x1} - \epsilon_{x2} \geq U_{x2} - U_{x1} | X = x) \]
\[ = P_{\Delta \epsilon|z}([U_{x0} - U_{x1}, \infty] \times [-\infty, \infty] \times [U_{x2} - U_{x1}, \infty]). \]

\(^\dagger\)Note that the degeneracy condition is trivially satisfied when \( r = 1 \).

\(^\ddagger\)These two representations of the sharp identified set for \( \Phi \) are derived under the assumption that the two sides of the market are stochastically independent. In the absence of stochastic independence, the definition of the sharp identified set for \( \Phi \) will be based on the joint probability distribution of \((\Delta \epsilon_i, \Delta \eta_j)\) conditional on \((X_i, Y_j)\).
Then, \( \forall (x, y) \in \mathcal{X} \times \mathcal{Y} \), we have that

\[
P_{Q_{\Delta|x}}(1) = P_{\Delta|x}([U_{x0} - U_{x1}, \infty] \times [-\infty, \infty] \times [U_{x2} - U_{x1}, \infty]),
\]

\[
P_{Q_{\Delta|x}}(2) = P_{\Delta|x}([-\infty, \infty] \times [U_{x0} - U_{x2}, \infty] \times [-\infty, U_{x2} - U_{x1}]),
\]

\[
P_{Q_{\Delta|x}}(0) = P_{\Delta|x}([-\infty, U_{x0} - U_{x1}] \times [-\infty, U_{x0} - U_{x2}] \times [-\infty, \infty]),
\]

(11)

and

\[
P_{Q_{\Delta|y}}(1) = P_{\Delta|y}([V_{0y} - V_{1y}, \infty] \times [-\infty, \infty] \times [V_{2y} - V_{1y}, \infty]),
\]

\[
P_{Q_{\Delta|y}}(2) = P_{\Delta|y}([-\infty, \infty] \times [V_{0y} - V_{2y}, \infty] \times [-\infty, V_{2y} - V_{1y}]),
\]

\[
P_{Q_{\Delta|y}}(0) = P_{\Delta|y}([-\infty, V_{0y} - V_{1y}] \times [-\infty, V_{0y} - V_{2y}] \times [-\infty, \infty]).
\]

(12)

Therefore,

\[
\mathcal{U}^* \equiv \left\{ U \in \mathcal{U}^\uparrow : \exists P_{\Delta|x} \in \mathcal{P}_{\Delta|x}^\uparrow \text{ s.t. } \forall x \in \mathcal{X} \text{ (11) is satisfied and } P_{\Delta|x}(\mathcal{B}) = 1 \right\},
\]

(13)

\[
\mathcal{V}^* \equiv \left\{ V \in \mathcal{V}^\uparrow : \exists P_{\Delta|y} \in \mathcal{P}_{\Delta|y}^\uparrow \text{ s.t. } \forall y \in \mathcal{Y} \text{ (12) is satisfied and } P_{\Delta|y}(\mathcal{B}) = 1 \right\},
\]

(14)

and \( \Theta^* \) is as in (10).

\[\Box\]

### 3.2 Constructing the sharp identified set

Performing the nonparametric exercise requires being able to construct \( \Theta^* \) without incorporating parametric restrictions into \( \mathcal{P}_{\Delta|x}^\uparrow \) and \( \mathcal{P}_{\Delta|y}^\uparrow \). This involves two methodological challenges on each side of the market. Specifically, on the men’s side, we first have to find whether, for a given \( U \in \mathcal{U}^\uparrow \), there exists \( P_{\Delta|x} \in \mathcal{P}_{\Delta|x}^\uparrow \) such that \( P_{Q_{\Delta|x}}(y) = \omega_{M,y|x}(U, P_{\Delta|x}) \) and \( P_{\Delta|x}(\mathcal{B}) = 1 \ \forall (x, y) \in \mathcal{X} \times \mathcal{Y}_0 \). Without parametric restrictions on the conditional probability distribution of the taste shocks, this corresponds to solving an infinite-dimensional existence problem because each \( P_{\Delta|x} \in \mathcal{P}_{\Delta|x}^\uparrow \) is an infinite-dimensional object. Second, such an infinite-dimensional existence problem has to be solved for every \( U \in \mathcal{U}^\uparrow \). Typically, this is done in the partial identification literature by constructing a grid of points roughly resembling \( \mathcal{U}^\uparrow \) and then repeating the exercise of interest for each grid point. The difficulty of implementing that approach increases with the size of the grid, which in turn, increases exponentially with the number of the agents’ types, hence quickly leading to a computational bottleneck. Similar issues are faced on the women’s side.

In what follows we design a procedure that ameliorates the methodological challenges just described. We organise the discussion in three steps, which are identical for each side of the market. Without loss of generality, let us consider the men’s side. In the
first step, Section 3.2.1 establishes that, for a given $U \in \mathcal{U}^\dagger$, determining whether there exists $P_{\Delta|x} \in \mathcal{P}_\Delta$ such that $P_{Q_{M|x}}(y) = \omega_{M,y|x}(U, P_{\Delta|x}) \forall (x,y) \in \mathcal{X} \times \mathcal{Y}_0$ is equivalent to finding whether a linear system has a solution, which is a well studied problem. This result uses Theorem 1 in Torgovitsky (2018) (also known as PIES) under various classes of nonparametric restrictions on the conditional probability distribution of the taste shock differences.\footnote{A necessary condition to apply PIES is completeness in the sense of Tamer (2003), which our model satisfies.}

In the second step, Section 3.2.2 shows that $\mathcal{U}^\dagger$ can be ex-ante partitioned into a finite number of convex subsets such that, for each subset, every value of $U$ belonging to that subset gives rise to the same linear programming problem of the first step. Therefore, the researcher has to solve such a linear programming problem once for each subset. Note that picking any one value of $U$ from each subset amounts to solving another linear programming problem because the subsets are convex.

In the third step, Section 3.2.3 discusses a way to approximate, for a given $P_{\Delta|x} \in \mathcal{P}_\Delta$, the degeneracy condition, $P_{\Delta|x}(B) = 1 \forall x \in \mathcal{X}$, by a system of linear equalities which can be easily added to the linear programming problem of the first step.

Specular considerations can be made for the women’s side. We now explain the three steps in detail by focusing on the men’s side.

### 3.2.1 A linear programming problem

As part of the construction of $\Theta^*$, the analyst has to find whether, for a given $U \in \mathcal{U}^\dagger$, there exists $P_{\Delta|x} \in \mathcal{P}_\Delta$ such that $P_{Q_{M|x}}(y) = \omega_{M,y|x}(U, P_{\Delta|x}) \forall (x,y) \in \mathcal{X} \times \mathcal{Y}_0$. Without parametric restrictions on the conditional probability distribution of the taste shocks, this corresponds to solving an infinite-dimensional existence problem because each $P_{\Delta|x} \in \mathcal{P}_\Delta$ is an infinite-dimensional object. We exploit Theorem 1 in Torgovitsky (2018) to transform such an infinite-dimensional existence problem into a linear programming problem.

**Proposition 2.** (Torgovitsky, 2018) Under various classes of nonparametric restrictions possibly incorporated into $\mathcal{P}_\Delta$, determining whether, for a given $U \in \mathcal{U}^\dagger$, there exists $P_{\Delta|x} \in \mathcal{P}_\Delta$ such that $P_{Q_{M|x}}(y) = \omega_{M,y|x}(U, P_{\Delta|x}) \forall (x,y) \in \mathcal{X} \times \mathcal{Y}_0$ is equivalent to finding whether a system of linear equalities and inequalities has at least one solution. ♦

We refer the reader to Theorem 1 in Torgovitsky (2018) for a notationally more precise and detailed statement of the result together with its proof. Here we illustrate the main intuition with an example. Consider the case $r = 2$ ($d = 3$) of Example 1. For each $P_{\Delta|x} \in \mathcal{P}_\Delta$, let $G_{\Delta|x} \in \mathcal{G}_\Delta$ denote the corresponding conditional CDF, where $\mathcal{G}_\Delta$ is the function space of all admissible $d$-dimensional conditional CDFs. To keep things simple suppose for the moment that the conditional probability distribution of the taste shock
differences is left completely unrestricted, i.e., $\mathcal{P}^d_{\Delta c} \equiv \mathcal{P}$ (and, hence, $\mathcal{G}^d_{\Delta c} \equiv \mathcal{G}$), where $\mathcal{P}$ is the function space of all possible $d$-dimensional conditional probability distributions (and, similarly, $\mathcal{G}$ is the function space of all possible $d$-dimensional conditional CDFs). We will discuss later how to incorporate nonparametric distributional assumptions on the taste shock differences. From (13), we have the following infinite-dimensional existence problem for a given $U \in \mathcal{U}^d$:

\begin{align}
\text{Find if there exists } P_{\Delta c|\epsilon} \in \mathcal{P} \text{ s.t. } \forall x \in X;
\end{align}

\begin{align}
P_{\Delta c|\epsilon}(1) &= P_{\Delta c|\epsilon}((\epsilon_{x_{0}} - \epsilon_{x_{1}}, \infty) \times (\infty, \infty) \times (\epsilon_{x_{2}} - \epsilon_{x_{1}}, \infty)), \\
P_{\Delta c|\epsilon}(2) &= P_{\Delta c|\epsilon}((-\infty, \infty) \times (\epsilon_{x_{0}} - \epsilon_{x_{2}}, \infty) \times (\epsilon_{x_{2}} - \epsilon_{x_{1}}, \infty)), \\
P_{\Delta c|\epsilon}(0) &= P_{\Delta c|\epsilon}((-\infty, \epsilon_{x_{0}} - \epsilon_{x_{1}}] \times (-\infty, \epsilon_{x_{0}} - \epsilon_{x_{2}}) \times (-\infty, \infty)).
\end{align}

Using $G_{\Delta c|\epsilon} \in \mathcal{G}$, (15) can be equivalently written as,\(^{14}\)

\begin{align}
\text{Find if there exists } G_{\Delta c|\epsilon} \in \mathcal{G} \text{ s.t. } \forall x \in X;
\end{align}

\begin{align}
P_{\Delta c|\epsilon}(1) &= 1 + G_{\Delta c|\epsilon}(\epsilon_{x_{0}} - \epsilon_{x_{1}}, \infty, \epsilon_{x_{2}} - \epsilon_{x_{1}}) - G_{\Delta c|\epsilon}(\epsilon_{x_{0}} - \epsilon_{x_{1}}, \infty, \epsilon_{x_{2}} - \epsilon_{x_{1}}) - G_{\Delta c|\epsilon}(\epsilon_{x_{0}} - \epsilon_{x_{1}}, \infty, \infty), \\
P_{\Delta c|\epsilon}(2) &= G_{\Delta c|\epsilon}(\epsilon_{x_{0}} - \epsilon_{x_{1}}, \infty, \epsilon_{x_{0}} - \epsilon_{x_{2}, \epsilon_{x_{2}} - \epsilon_{x_{1}}}) - G_{\Delta c|\epsilon}(\epsilon_{x_{0}} - \epsilon_{x_{2}}, \epsilon_{x_{2}} - \epsilon_{x_{1}}), \\
P_{\Delta c|\epsilon}(0) &= G_{\Delta c|\epsilon}(\epsilon_{x_{0}} - \epsilon_{x_{1}}, \epsilon_{x_{0}} - \epsilon_{x_{2}}, \epsilon_{x_{0}} - \epsilon_{x_{2}}). \\
\end{align}

The system of equations in (16) depends on the values of $G_{\Delta c|\epsilon}$ at a finite number of 3-tuples, $\forall x \in X$. We thus introduce three finite sets

\begin{align}
A_{x,1}(U) &\equiv \{U_{x_{0}} - U_{x_{1}}, \infty, -\infty\}, \\
A_{x,2}(U) &\equiv \{U_{x_{0}} - U_{x_{2}}, \infty, -\infty\}, \\
A_{x,3}(U) &\equiv \{U_{x_{2}} - U_{x_{1}}, \infty, -\infty\},
\end{align}

$\forall x \in X$, where $A_{x,1}(U)$ collects the elements at which $G_{\Delta c|\epsilon}$ is evaluated along the first dimension, $A_{x,2}(U)$ collects the elements at which $G_{\Delta c|\epsilon}$ is evaluated along the second dimension, and $A_{x,3}(U)$ collects the elements at which $G_{\Delta c|\epsilon}$ is evaluated along the third dimension. We add $-\infty$ to each set because the value of one-dimensional CDFs at $-\infty$ is known and equal to 0 by definition. Lastly, we define $A_{x}(U) \equiv A_{x,1}(U) \times A_{x,2}(U) \times A_{x,3}(U)$, where “$\times$” denotes the Cartesian product operator. Therefore, (16) can be equivalently written as

\(^{14}\)Recall that the volume of a cube can be written in terms of the CDF. Hence, $\forall x \in X$, we have that

\begin{align}
P_{\Delta c|\epsilon}([a_{1}, b_{1}] \times [a_{2}, b_{2}] \times [a_{3}, b_{3}]) &= \\
\begin{align}
&- G_{\Delta c|\epsilon}(a_{1}, a_{2}, a_{3}) + G_{\Delta c|\epsilon}(b_{1}, a_{2}, a_{3}) + G_{\Delta c|\epsilon}(a_{1}, b_{2}, a_{3}) - G_{\Delta c|\epsilon}(b_{1}, b_{2}, a_{3}) \\
+ G_{\Delta c|\epsilon}(a_{1}, a_{2}, b_{3}) - G_{\Delta c|\epsilon}(b_{1}, a_{2}, b_{3}) - G_{\Delta c|\epsilon}(a_{1}, b_{2}, b_{3}) + G_{\Delta c|\epsilon}(b_{1}, b_{2}, b_{3}).
\end{align}
\]
\( \forall x \in \mathcal{X} \), find if there exists \( \tilde{G}^U_{\Delta|x} : \mathcal{A}_x(U) \to \mathbb{R} \) s.t.

\[
P_{Q_3|x}(1) = 1 + \tilde{G}^U_{\Delta|x}(U_{x0} - U_{x1}, \infty, U_{x2} - U_{x1}) - \tilde{G}^U_{\Delta|x}(\infty, \infty, U_{x2} - U_{x1}) - \tilde{G}^U_{\Delta|x}(U_{x0} - U_{x1}, \infty, \infty),
\]

\[
P_{Q_3|x}(2) = \tilde{G}^U_{\Delta|x}(\infty, \infty, U_{x2} - U_{x1}) - \tilde{G}^U_{\Delta|x}(\infty, U_{x2} - U_{x1}),
\]

\[
P_{Q_3|x}(0) = \tilde{G}^U_{\Delta|x}(U_{x0} - U_{x1}, U_{x0} - U_{x2}, \infty),
\]

and \( \tilde{G}^U_{\Delta|x} \) can be extended to a conditional CDF in \( \mathcal{G} \).

(17)

(17) states that the existence problem (15) is equivalent to first finding whether a system of linear equalities has a solution, and second ensuring that such a solution solves an extension exercise. Note that the system has a finite number of equalities because the domain of the function \( \tilde{G}^U_{\Delta|x} \) is finite. With regard to the extension exercise, using fundamental results in copula theory, in particular Sklar’s Theorem, Torgovitsky (2018) shows that verifying whether \( \tilde{G}^U_{\Delta|x} \) can be extended to a conditional CDF in \( \mathcal{G} \) amounts to checking if it satisfies the following system of linear equalities and inequalities \( \forall x \in \mathcal{X} \):

\[
\begin{align*}
\tilde{G}^U_{\Delta|x}(\infty, \infty, \infty) &= 1, \\
0 &\leq \tilde{G}^U_{\Delta|x}(t, q, r) \leq 1, \\
\forall (t, q) &\in \mathcal{A}_{x,2}(U) \times \mathcal{A}_{x,3}(U), \\
\forall (t, q) &\in \mathcal{A}_{x,1}(U) \times \mathcal{A}_{x,3}(U), \\
\forall (t, q) &\in \mathcal{A}_{x,1}(U) \times \mathcal{A}_{x,2}(U), \\
\forall (t, q, r) &\in \mathcal{A}_{x}(U), \\
\forall (t, q, r) &\in \mathcal{A}_{x}(U) \\
\text{s.t. } (t, q, r) &\leq (t', q', r').
\end{align*}
\]

Specifically, bearing in mind the properties defining CDFs, the first four lines in (18) ensure that \( \tilde{G}^U_{\Delta|x} \) is equal to 0 when at least one of its arguments is \( -\infty \) and equal to 1 when all its arguments are \( \infty \). The fifth line in (18) guarantees that the range of \( \tilde{G}^U_{\Delta|x} \) is a subset of \([0, 1]\). The last line in (18) requires \( \tilde{G}^U_{\Delta|x} \) to be a 3-increasing function, i.e., for each pair of 3-tuples in \( \mathcal{A}_x(U) \) which are comparable, \((t, q, r)\) and \((t', q', r')\), the volume of the 3-dimensional box with vertices from \( \{t, t'\} \times \{q, q'\} \times \{r, r'\} \) is positive.\(^{15}\)

\(^{15}\)For a given \( x \in \mathcal{X} \), from Definition 1 in Torgovitsky (2018), \( \tilde{G}^U_{\Delta|x} : \mathcal{A}_x(U) \rightarrow \mathbb{R} \) is \( d \)-increasing if

\[
\text{Vol}_{\tilde{G}^U_{\Delta|x}}(u', u'') \equiv \sum_{u \in \text{Vrt}(u', u'')} \text{sgn}(u', u'') \text{sgn}(u, u'') \geq 0 \quad \forall u', u'' \in \mathcal{A}_x(U) \text{ s.t. } u' \leq u'',
\]

where \( \text{Vrt}(u', u'') \equiv \{ u \in \mathcal{A}_x(U) : u_l \in [u'_{l}, u''_{l}] \ \forall l \in \{1, ..., d\} \} \) and

\[
\text{sgn}(u', u'')(u) \equiv \begin{cases} 1 & \text{if } u_l = u'_{l} \text{ for an even number of } l \in \{1, ..., d\} \\ -1 & \text{if } u_l = u''_{l} \text{ for an odd number of } l \in \{1, ..., d\}. \end{cases}
\]

\( \text{Vol}_{\tilde{G}^U_{\Delta|x}}(u', u'') \) is the volume of the \( d \)-dimensional box \([u'_{1}, u''_{1}] \times [u'_{2}, u''_{2}] \times ... \times [u'_{L}, u''_{L}] \) and \( \text{Vrt}(u', u'') \) is the collection of the vertices of this box. Note that \( d \)-increasing reduces to the standard definition of
By merging (17) and (18), an easy-to-solve linear programming problem is obtained.

The procedure described allows to incorporate into \( P_{\Delta} \) many classes of nonparametric restrictions. This is useful in partial identification analysis to see how the bounds on \( \Phi \) shrink under more or less stringent assumptions. Indeed, Theorem 1 in Torgovitsky (2018) shows that such constraints are simply added to (18) as linear equalities and inequalities. For example, one can impose that the conditional probability distribution of the taste shock differences is characterised by identical marginals, i.e., \( P_{\Delta|x}^k = P_{\Delta|x}^{k'} \) for all \( k, k' \) \( \subseteq \{1, \ldots, d\} \), where \( P_{\Delta|x}^k \) denotes the \( k \)th marginal of \( P_{\Delta|x} \). We can include that the marginals are symmetric about zero, i.e., \( P_{\Delta|x}^k((\infty, a]) = 1 - P_{\Delta|x}^k((\infty, -a]) \) for all \( a \in \mathbb{R} \) and \( \forall k \in \{1, \ldots, d\} \). Joint or marginal independence of the agents’ types can be imposed by using \( P_{\Delta|x} = P_{\Delta|x'} \) or \( P_{\Delta|x}^k = P_{\Delta|x'}^k \) for all \( x, x' \) \( \subseteq \mathcal{X} \) and \( \forall k \in \{1, \ldots, d\} \), respectively. Further, any quantile of \( P_{\Delta|x} \) or \( \{P_{\Delta|x}^k\}_{k=1}^d \) can be set equal to known values. We refer the reader to Assumption A in Torgovitsky (2018) for an accurate taxonomy of the nonparametric distributional assumptions on the taste shock differences that can be accommodated.

To give an idea of how the linear programming problem should be modified when nonparametric distributional assumptions on the taste shock differences are included, consider imposing that the conditional marginal probability distributions of the taste shock differences are symmetric about zero. Then, we have

\[
\begin{align*}
A_{x1}(U) &\equiv \{U_{x0} - U_{x1}, -U_{x0} + U_{x1}, \infty, -\infty\}, \\
A_{x2}(U) &\equiv \{U_{x0} - U_{x2}, -U_{x0} + U_{x2}, \infty, -\infty\}, \\
A_{x3}(U) &\equiv \{U_{x2} - U_{x1}, -U_{x2} + U_{x1}, \infty, -\infty\},
\end{align*}
\]

\( \forall x \in \mathcal{X}. \)

The linear programming problem to solve becomes

\[
\begin{align*}
&\forall x \in \mathcal{X}, \text{ find if there exists } \bar{G}^U_{\Delta|x} : \mathcal{A}_x(U) \rightarrow \mathbb{R} \text{ s.t.} \\
&P_{\Delta|x}(1) = 1 + \bar{G}^U_{\Delta|x}(U_{x0} - U_{x1}, \infty, U_{x2} - U_{x1}) - \bar{G}^U_{\Delta|x}(\infty, \infty, U_{x2} - U_{x1}) - \bar{G}^U_{\Delta|x}(U_{x0} - U_{x1}, \infty, \infty), \\
&P_{\Delta|x}(2) = \bar{G}^U_{\Delta|x}(\infty, \infty, U_{x2} - U_{x1}) - \bar{G}^U_{\Delta|x}(\infty, U_{x0} - U_{x1}, U_{x2} - U_{x1}), \\
&P_{\Delta|x}(0) = \bar{G}^U_{\Delta|x}(U_{x0} - U_{x1}, U_{x0} - U_{x2}, \infty), \\
\end{align*}
\]

(18) is satisfied,

\[
\begin{align*}
\bar{G}^U_{\Delta|x}(U_{x0} - U_{x1}, \infty, \infty) &= 1 - \bar{G}^U_{\Delta|x}(-U_{x0} + U_{x1}, \infty, \infty), \\
\bar{G}^U_{\Delta|x}(\infty, U_{x0} - U_{x2}, \infty) &= 1 - \bar{G}^U_{\Delta|x}(\infty, -U_{x0} + U_{x2}, -\infty), \\
\bar{G}^U_{\Delta|x}(\infty, \infty, U_{x2} - U_{x1}) &= 1 - \bar{G}^U_{\Delta|x}(\infty, \infty, -U_{x2} + U_{x1}).
\end{align*}
\]

(19)

Note that the first three equations above are identical to (16), (18) refers to the extension exercise, and the last three equations impose symmetry of the marginals about zero.

\footnote{Note that every time we add restrictions to the linear programming problem the set \( \mathcal{A}_x(U) \) can potentially change.}

weakly increasing when \( r = 1 \).
Before concluding, we remark that the methodology illustrated does not allow to incorporate nonparametric assumptions on the conditional probability distributions of the original taste shocks, that is on $P_{t|X}$ and $P_{\eta|Y}$. Given that in multinomial choice models what matters are differences in utilities (and not absolute levels), this is the obvious price to pay for having a flexible nonparametric approach that permits us to compare the empirical content of the model under different scenarios.

3.2.2 Partitioning the parameter space, $\mathcal{U}^1$

As part of the construction of $\Theta^*$, the analyst has to solve the linear programming problem described in Section 3.2.1 for every $U \in \mathcal{U}^1$. Typically, this is done in the partial identification literature by constructing a grid of points approximating $\mathcal{U}^1$ and then repeating the exercise of interest for each grid point. The difficulty of implementing such an approach increases with the size of the grid, which in turn, increases exponentially with the number of the agents’ types, hence quickly leading to a computational bottleneck. In what follows we give a characterisation of $\mathcal{U}^1$ so that the issue of solving the linear programming problem for every $U \in \mathcal{U}^1$ is reduced to solving it for a handful of $U \in \mathcal{U}^1$.

Proposition 3. (Partitioning $\mathcal{U}^1$) Let $\mathcal{P}_{\Delta|X}^1$ include any nonparametric restrictions on the conditional probability distribution of the taste shock differences considered in Proposition 2. Then, for every $x \in \mathcal{X}$, $\mathcal{U}^1$ can be ex-ante partitioned into a finite number, $K_x$, of convex subsets, $\mathcal{U}^1_{1,x}, \ldots, \mathcal{U}^1_{K_x,x}$, such that $\forall k \in \{1, \ldots, K_x\}$ and $\forall \{U, \tilde{U}\} \subseteq \mathcal{U}^1_{k,x}$,

$$\exists P_{\Delta|X} \in \mathcal{P}_{\Delta|X}^1 \text{ such that } P_{s_{M|x}}(y) = \omega_{M,y|x}(U, P_{\Delta|X}) \forall y \in \mathcal{Y}_0$$

if and only if

$$\exists P_{\Delta|X} \in \mathcal{P}_{\Delta|X}^1 \text{ such that } P_{s_{M|x}}(y) = \omega_{M,y|x}(\tilde{U}, P_{\Delta|X}) \forall y \in \mathcal{Y}_0.$$  

\[\blacksquare\]

As a consequence of Proposition 3, given $k \in \{1, \ldots, K_x\}$ and $x \in \mathcal{X}$, the analyst has to pick any one value of $U$ from the partitioning subset $\mathcal{U}^1_{k,x}$, and solve the linear programming problem of Section 3.2.1 at that value. This should be done for every $k \in \{1, \ldots, K_x\}$ and $x \in \mathcal{X}$ in order to span the entire parameter space, $\mathcal{U}^1$. Hence, the linear programming problem of Section 3.2.1 is solved only $\sum_{x \in \mathcal{X}} K_x$ times.

We now sketch the intuition behind Proposition 3 continuing from the linear programming problem (19), which refers to the case with $r = 2$ ($d = 3$). Given $x \in \mathcal{X}$, the only piece of (19) that can potentially induce different solutions for different values of $U$ is the one requiring the function $\tilde{G}_{\Delta|X}^U : \mathcal{A}_x(U) \to \mathbb{R}$ to be 3-increasing. That is, for each pair of 3-tuples in $\mathcal{A}_x(U)$ which are comparable, $(t, q, r)$ and $(t', q', r')$, the volume of the
3-dimensional box with vertices from \( \{ t, t' \} \times \{ q, q' \} \times \{ r, r' \} \) is positive:

\[
- \vec{G}^U_{\Delta|x}(t, q, r) + \vec{G}^U_{\Delta|x}(t', q, r) + \vec{G}^U_{\Delta|x}(t, q', r) - \vec{G}^U_{\Delta|x}(t', q', r) \\
+ \vec{G}^U_{\Delta|x}(t, q, r') - \vec{G}^U_{\Delta|x}(t', q, r') - \vec{G}^U_{\Delta|x}(t, q', r') + \vec{G}^U_{\Delta|x}(t', q', r') \geq 0
\]

(20) can induce different solutions for different values of \( U \) because different values of \( U \) can generate different pairs of comparable 3-tuples in \( A_x(U) \). This implies that the 3-dimensional boxes for which we have to ensure a positive volume may have vertices located at different positions (in other words, “differently ordered”) in \( \mathbb{R}^d \) for different values of \( U \). Such observation immediately suggests that if two values of \( U \) generate pairs of 3-tuples in \( A_x(U) \) that are “similarly ordered” in \( \mathbb{R}^d \) (as formalised in the proof in Appendix A), then they should lead to the same constraints on \( \vec{G}^U_{\Delta|x} \’s \) image set of the type (20). In turn, they will have the same solution to (19) at \( x \in X \).

For every \( x \in X \), an important property of the partitioning subsets, \( U^1_{x,1}, \ldots, U^d_{x,d} \), is convexity. This follows from the fact that the elements of the set \( A_{x,j}(U) \) are linear functionals of the elements of the vector \( U \), for each \( j \in \{1, \ldots, d\} \). Therefore, ordering the elements of the set \( A_{x,j}(U) \) preserves linearity, for each \( j \in \{1, \ldots, d\} \). The convexity of the partitioning subsets enables us to formulate the problem of picking a value of \( U \) from \( U^j_{k,x} \) as a linear programming problem, for every \( k \in \{1, \ldots, K_x\} \).

Lastly, note that for every \( x \in X \) the number of partitioning subsets, \( K_x \), is finite. Providing a general formula to compute \( K_x \) that is valid under any class of nonparametric distributional assumptions on the taste shock differences contemplated by Proposition 2 does not seem to be viable. Nevertheless, we can certainly say that \( K_x \) increases with the cardinality of \( Y \) (the higher is the cardinality of \( Y \), the higher is \( d \)) and with the number of nonparametric distributional assumptions on the taste shock differences (a greater number of assumptions may imply higher cardinalities of \( A_x(U) \)).

### 3.2.3 The degeneracy condition

Recall that the conditional probability distribution of the vector of the original (i.e., not differenced) taste shocks, \( \epsilon_i \), is denoted by \( P_{i|X} \). When we consider the vector of taste shock differences, \( \Delta \epsilon_i \), its conditional probability distribution, \( P_{\Delta \epsilon|X} \), depends on \( P_{\epsilon|X} \). Specifically, \( P_{\Delta \epsilon|X} \) has to be concentrated on the region \( B \) defined by (8). Hence, as part of the construction of \( \Theta^* \) the researcher should ensure that the conditional probability distribution, \( P_{\Delta \epsilon|X} \), solving the linear programming problem of Section 3.2.1 satisfies the degeneracy condition, \( P_{\Delta \epsilon|X}(B) = 1 \). Without incorporating the degeneracy condition one does not exploit all the information coming from the model and thus obtains an outer set of \( \Theta^* \).

In what follows we propose a computationally tractable approximation of the degener-
eracy condition. Specifically, we suggest to approximate the degeneracy condition by a finite collection of equalities that are linear in the CDF $G_{\Delta|X}$ and therefore, linear in the function $\bar{G}^U_{\Delta|X}$ introduced in (17). Hence, these equalities can be easily added to the linear programming problem of Section 3.2.1.

More formally, let us first state Proposition 4 which provides an equivalent characterisation of the degeneracy condition in terms of zero probability measure conditions on boxes in $Q^d$, where $Q \subset \mathbb{R}$ is the set of rational numbers.

**Proposition 4.** (Degeneracy condition) For any $(\hat{b}, \tilde{b}) \in \mathbb{R}^2$, consider the $d$-dimensional boxes in $\mathbb{R}^d$

$$B_{t,p,q}(\hat{b}, \tilde{b}) \equiv \{(z_1, ..., z_d) \in \mathbb{R}^d: z_t > \hat{b} + \tilde{b}, z_p \leq \hat{b}, z_q \leq \tilde{b}\},$$

and

$$Q_{t,p,q}(\hat{b}, \tilde{b}) \equiv \{(z_1, ..., z_d) \in \mathbb{R}^d: z_t \leq \hat{b} + \tilde{b}, z_p > \hat{b}, z_q > \tilde{b}\},$$

$\forall t \in \{1, ..., r - 1\}$ and $\forall (p, q) \in \{(t + 1, r + 1), (t + 2, r + 2), ..., (r, d)\}$.

For each $P_{\Delta|X} \in \mathcal{P}_{\Delta}^1$,

$$P_{\Delta|X}(B_{t,p,q}(\hat{b}, \tilde{b})) = P_{\Delta|X}(Q_{t,p,q}(\hat{b}, \tilde{b})) = 0$$

$\forall t \in \{1, ..., r - 1\}$, $\forall (p, q) \in \{(t + 1, r + 1), (t + 2, r + 2), ..., (r, d)\}$, $\forall (b, \tilde{b}) \in Q^2$.

if and only if

$$P_{\Delta|X}(\mathcal{B}) = 1.$$  \hfill (21)

The idea underlying Proposition 4 is that, for a given $P_{\Delta|X} \in \mathcal{P}_{\Delta}^1$, if $P_{\Delta|X}(\mathcal{B}) = 1$, then any $d$-dimensional box in $\mathbb{R}^d$ not intersecting the region $\mathcal{B}$ has probability measure zero. Conversely, it is sufficient to impose such a zero probability measure condition for all the $d$-dimensional boxes in $Q^d$ not intersecting the region $\mathcal{B}$ to satisfy $P_{\Delta|X}(\mathcal{B}) = 1$.

Proposition 4 suggests that by imposing (21) at a finite grid of 2-tuples from $Q^2$ we can approximate the degeneracy condition. The finer is the grid, the better is the approximation. Furthermore, (21) is linear in the CDF $G_{\Delta|X}$ and hence, can be incorporated in the linear programming problem of Section 3.2.1 for every $U \in \mathcal{U}^1$ after having replaced $G_{\Delta|X}$ with $\bar{G}^U_{\Delta|X}$.

To clarify how the linear programming problem of Section 3.2.1 should be modified, consider assuming that the conditional marginal probability distributions of the taste shock differences are symmetric about zero. For simplicity, select two 2-tuples from $Q^2$,
for every \( (\hat{b}, \check{b}) \in \mathbb{Q}^2 \). Then, given \( U \in \mathcal{U} \), we have

\[
A_{x,1}(U) \equiv \{ U_{x_0} - U_{x_1}, -U_{x_0} + U_{x_1}, \hat{b}^1 + \check{b}^1, -\hat{b}^1 - \check{b}^1, \hat{b}^2 + \check{b}^2, -\hat{b}^2 - \check{b}^2, \infty, -\infty \},
\]

\[
A_{x,2}(U) \equiv \{ U_{x_0} - U_{x_2}, -U_{x_0} + U_{x_2}, \hat{b}^1, -\hat{b}^1, \hat{b}^2, -\hat{b}^2, \infty, -\infty \},
\]

\[
A_{x,3}(U) \equiv \{ U_{x_2} - U_{x_1}, -U_{x_2} + U_{x_1}, \hat{b}^1, -\hat{b}^1, \hat{b}^2, -\hat{b}^2, \infty, -\infty \},
\]

for every \( x \in \mathcal{X} \). The linear programming problem to solve becomes

\[
\forall x \in \mathcal{X}, \text{ find if there exists } \hat{G}_{\Delta|x}(U) : A_x(U) \to \mathbb{R} \text{ s.t.}
\]

\[
P_{\mathcal{Q}|x|}(1) = 1 + \hat{G}_{\Delta|x}(U_{x_0} - U_{x_1}, \infty, U_{x_2} - U_{x_1}) - \hat{G}_{\Delta|x}(\infty, U_{x_2} - U_{x_1}) - \hat{G}_{\Delta|x}(U_{x_0} - U_{x_1}, \infty, \infty),
\]

\[
P_{\mathcal{Q}|x|}(2) = \check{G}_{\Delta|x}(\infty, \infty, U_{x_2} - U_{x_1}) - \check{G}_{\Delta|x}(\infty, U_{x_0} - U_{x_2}, U_{x_2} - U_{x_1}),
\]

\[
P_{\mathcal{Q}|x|}(0) = \check{G}_{\Delta|x}(U_{x_0} - U_{x_1}, U_{x_0} - U_{x_2}, \infty),
\]

(18) is satisfied,

\[
\hat{G}_{\Delta|x}(U_{x_0} - U_{x_1}, \infty, \infty) = 1 - \hat{G}_{\Delta|x}(-U_{x_0} + U_{x_1}, \infty, \infty),
\]

\[
\hat{G}_{\Delta|x}(\hat{b}^1 + \hat{b}^1, \infty, \infty) = 1 - \hat{G}_{\Delta|x}(-\hat{b}^1 - \check{b}^1, \infty, \infty),
\]

\[
\check{G}_{\Delta|x}(\check{b}^2 + \check{b}^2, \infty, \infty) = 1 - \check{G}_{\Delta|x}(-\hat{b}^2 - \check{b}^2, \infty, \infty),
\]

\[
\hat{G}_{\Delta|x}(\infty, U_{x_0} - U_{x_2}, \infty) = 1 - \hat{G}_{\Delta|x}(\infty, -U_{x_0} + U_{x_2}, -\infty),
\]

\[
\check{G}_{\Delta|x}(\infty, \hat{b}_1, \infty) = 1 - \check{G}_{\Delta|x}(\infty, -\hat{b}_1, -\infty),
\]

\[
\hat{G}_{\Delta|x}(\infty, \check{b}_2, \infty) = 1 - \hat{G}_{\Delta|x}(\infty, -\check{b}_2, -\infty),
\]

\[
\check{G}_{\Delta|x}(\infty, \infty, U_{x_1} - U_{x_2}) = 1 - \check{G}_{\Delta|x}(\infty, \infty, -U_{x_1} + U_{x_2}),
\]

\[
\hat{G}_{\Delta|x}(\infty, \infty, \hat{b}_1) = 1 - \hat{G}_{\Delta|x}(\infty, \infty, -\hat{b}_1),
\]

\[
\check{G}_{\Delta|x}(\infty, \infty, \check{b}_2) = 1 - \check{G}_{\Delta|x}(\infty, \infty, -\check{b}_2),
\]

\[
\hat{G}_{\Delta|x}(\infty, \hat{b}_1, \check{b}_1) - \hat{G}_{\Delta|x}(\check{b}_1 + \check{b}_1, \check{b}_1, \check{b}_1) = 0,
\]

\[
\hat{G}_{\Delta|x}(\hat{b}_1 + \hat{b}_1, \infty, \infty) - \hat{G}_{\Delta|x}(\hat{b}_1 + \check{b}_1, \check{b}_1, \check{b}_1) - \hat{G}_{\Delta|x}(\check{b}_1 + \check{b}_1, \check{b}_1, \check{b}_1) + \hat{G}_{\Delta|x}(\hat{b}_1 + \check{b}_1, \check{b}_1, \check{b}_1) = 0,
\]

\[
\check{G}_{\Delta|x}(\infty, \check{b}_2, \check{b}_2) - \check{G}_{\Delta|x}(\check{b}_2 + \check{b}_2, \check{b}_2, \check{b}_2) = 0,
\]

\[
\check{G}_{\Delta|x}(\check{b}_2 + \check{b}_2, \infty, \infty) - \check{G}_{\Delta|x}(\check{b}_2 + \check{b}_2, \check{b}_2, \check{b}_2) - \check{G}_{\Delta|x}(\check{b}_2 + \check{b}_2, \check{b}_2, \check{b}_2) + \check{G}_{\Delta|x}(\check{b}_2 + \check{b}_2, \check{b}_2, \check{b}_2) = 0,
\]

where the last four equations impose that the approximated complement of \( \mathcal{B} \) has probability measure zero.

## 4 Monte Carlo simulations

This section investigates via simulations how the shape of the sharp identified set for \( \Phi, \Theta^* \), varies under different data generating processes. More precisely, we construct \( \Theta^* \)
when \( r = 2 \) \((d = 3)\) as in Example 1 under two data generating processes. According to
the first data generating process (hereafter, DGP1), the taste shocks on each side of the
market are distributed independently of the agents’ types as i.i.d. Gumbel with scale 1
and location 0, as in Choo and Siow (2006). Additionally, \( P_X(1) = P_Y(1) = \frac{1}{2} \).
According to the second data generating process (hereafter, DGP2), the 3-dimensional vectors of
taste shocks on each side of the market are distributed independently of the agents’ types
as equiprobable Gaussian mixtures of 4 components,

\[
\mathcal{N}\left(\begin{bmatrix} -1 \\ -1 \\ -2 \\ -2 \\ -2 \\ -2 \end{bmatrix}, \begin{bmatrix} 8.7 & 2.1 & 2.1 \\ 2.1 & 8.7 & 2.1 \\ 2.1 & 2.1 & 8.7 \\ 4.5 & -1.3 & -1.3 \\ -1.3 & 4.5 & -1.3 \\ -1.3 & -1.3 & 4.5 \end{bmatrix}\right) \cdot \mathcal{N}\left(\begin{bmatrix} 3.5 \\ 3.5 \\ 4.6 \\ 4.6 \\ 4.6 \\ 4.6 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \\ 1 & -0.4 & -0.4 \\ -0.4 & 1 & -0.4 \\ -0.4 & -0.4 & 1 \end{bmatrix}\right),
\]

Moreover, \( P_X(1) = \frac{1}{6} \) and \( P_Y(1) = \frac{1}{5} \). In both data generating processes, \( \Phi_{11} = \Phi_{22} = 3 \)
and \( \Phi_{12} = \Phi_{21} = 2 \), i.e., the systematic component of the match surplus is higher when
the two partners are of the same type. Lastly, \( \Phi_{0y} = \Phi_{x0} = 0 \ \forall (x, y) \in X \times Y \).

In order to construct \( \Theta^* \) we include scale and location normalisations into the pa-
rameter spaces, \( U^\dagger \) and \( V^\dagger \). Such normalisations ensure that the volume of \( \Theta^* \) is not improperly inflated relative to the point identified case. We impose different normalisations depending on whether the function spaces of admissible conditional probability
distributions for the taste shock differences, \( \mathcal{P}_{\Delta x}^\dagger \) and \( \mathcal{P}_{\Delta y}^\dagger \), incorporate or not indepen-
dence between the taste shock differences and the agents’ types. Specifically, when \( \mathcal{P}_{\Delta x}^\dagger \)
and \( \mathcal{P}_{\Delta y}^\dagger \) do not incorporate independence, we define

\[
U^\dagger \equiv \left\{(U_{xy} \ \forall (x, y) \in X \times Y_0) \in \mathbb{R}^6 : \begin{align*}
\text{[location normalisation]} & \quad U_{x0} = 0 \ \forall x \in X, \\
\text{[scale normalisation]} & \quad U_{x1} \in \{-1, 1\} \ \forall x \in X, 
\end{align*} \right\}, \tag{22}
\]

\[
V^\dagger \equiv \left\{(V_{xy} \ \forall (x, y) \in X_0 \times Y) \in \mathbb{R}^6 : \begin{align*}
\text{[location normalisation]} & \quad V_{0y} = 0 \ \forall y \in Y, \\
\text{[scale normalisation]} & \quad V_{1y} \in \{-1, 1\} \ \forall y \in Y, 
\end{align*} \right\}, \tag{23}
\]

where the first condition exactly mimics Choo and Siow (2006) as a location normalisa-
incorporate independence, we define

\[ U^\dagger \equiv \left\{ (U_{xy} \forall (x, y) \in \mathcal{X} \times \mathcal{Y}_0) \in \mathbb{R}^6: \right. \\
\left. \begin{array}{l}
\text{[location normalisation]} \\
U_{x0} = 0 \forall x \in \mathcal{X} \\
\text{[scale normalisation]} \\
U_{11} \in \{ -1, 1 \} 
\end{array} \right\}. \tag{24}
\]

\[ V^\dagger \equiv \left\{ (V_{xy} \forall (x, y) \in \mathcal{X}_0 \times \mathcal{Y}) \in \mathbb{R}^6: \right. \\
\left. \begin{array}{l}
\text{[location normalisation]} \\
V_{0y} = 0 \forall y \in \mathcal{Y} \\
\text{[scale normalisation]} \\
V_{11} \in \{ -1, 1 \} 
\end{array} \right\}. \tag{25}
\]

Note that when \( \mathcal{P}_{\Delta}^\dagger \) and \( \mathcal{P}_{\Delta_y}^\dagger \) do not incorporate independence we impose more scale normalisations than in the independence case. In particular, when \( \mathcal{P}_{\Delta}^\dagger \) and \( \mathcal{P}_{\Delta_y}^\dagger \) do not incorporate independence, we impose one scale normalisation for each \( x \) on the men’s side and one scale normalisation for each \( y \) in \( \mathcal{Y} \) on the women’s side. This is because determining whether \( U \) belongs to \( U^* \) requires recovering \(|\mathcal{X}|\) admissible conditional probability distributions, \( \{ P_{\Delta|x} \forall x \in \mathcal{X} \} \). Similarly, determining whether \( V \) belongs to \( V^* \) requires recovering \(|\mathcal{Y}|\) admissible conditional probability distributions, \( \{ P_{\Delta|y} \forall y \in \mathcal{Y} \} \). In fact, if \( U \) and \( V \) belong to \( U^* \) and \( V^* \) for some \( \{ P_{\Delta|x} \forall x \in \mathcal{X} \} \) and \( \{ P_{\Delta|y} \forall y \in \mathcal{Y} \} \), then any rescaled version of \( \{ P_{\Delta|x} \forall x \in \mathcal{X} \} \) and \( \{ P_{\Delta|y} \forall y \in \mathcal{Y} \} \) induces some scalar multiples of \( U \) and \( V \) to also belong to \( U^* \) and \( V^* \), respectively. Hence, the number of scale normalisations to impose on \( U^\dagger \) and \( V^\dagger \) is equal to the number of conditional probability distributions to recover, that is \(|\mathcal{X}| + |\mathcal{Y}|\). Instead, when \( \mathcal{P}_{\Delta}^\dagger \) and \( \mathcal{P}_{\Delta_y}^\dagger \) incorporate independence, we impose just one scale normalisation on the men’s side and one scale normalisation on the women’s side. This is because determining whether \( U \) belongs to \( U^* \) requires recovering only one admissible conditional probability distribution, \( P_{\Delta} \), where \( P_{\Delta} \equiv P_{\Delta|x} \forall x \in \mathcal{X} \). Similarly, determining whether \( V \) belongs to \( V^* \) requires recovering only one admissible conditional probability distribution, \( P_{\Delta_y} \), where \( P_{\Delta_y} \equiv P_{\Delta|y} \forall y \in \mathcal{Y} \).

Finally, note that the scale and location normalisations just discussed are imposed on the vectors \( U \) and \( V \) which are equilibrium objects (and, hence, they can vary with \( P_X, P_Y, P_{\Delta|x}, P_{\Delta|y} \), even if \( \Phi \) remains the same), and not primitive parameters. Consequently, the shapes of the normalised sharp identified sets based on DGP1 and DGP2

---

17In other words, we consider vectors like

\[
\begin{pmatrix}
U_{10} - U_{10} & U_{11} - U_{10} & U_{12} - U_{10} & U_{20} - U_{20} & U_{21} - U_{20} & U_{22} - U_{20} \\
\frac{|U_{10}|}{|U_{11}|} & \frac{|U_{11}|}{|U_{11}|} & \frac{|U_{12}|}{|U_{11}|} & \frac{|U_{20}|}{|U_{21}|} & \frac{|U_{21}|}{|U_{21}|} & \frac{|U_{22}|}{|U_{21}|}
\end{pmatrix},
\]

\[
\begin{pmatrix}
V_{01} - V_{01} & V_{11} - V_{01} & V_{21} - V_{01} & V_{02} - V_{02} & V_{12} - V_{02} & V_{22} - V_{02} \\
\frac{|V_{01}|}{|V_{11}|} & \frac{|V_{11}|}{|V_{11}|} & \frac{|V_{21}|}{|V_{11}|} & \frac{|V_{02}|}{|V_{12}|} & \frac{|V_{12}|}{|V_{12}|} & \frac{|V_{22}|}{|V_{12}|}
\end{pmatrix}.
\]
are not directly comparable.

The figures below report $\Theta^*$ under various classes of nonparametric distributional assumptions on the taste shock differences. Each figure is composed of six sub-figures where we project $\Theta^*$ along every two of its dimensions. All the linear programming problems have been solved by calling Gurobi in Matlab.

Figures 1, 2, 3, 4, 5, and 6 are based on DGP1. The black regions represent the projections of $\Theta^*$. The red dots represent the projections of the normalised true value of $\Phi$. Figure 1 reports the projections of $\Theta^*$ when no restriction on the conditional probability distributions of the latent variables is imposed, i.e., $P^1_{\Delta \epsilon} = P^1_{\Delta \eta} = P$. As expected, the black regions are completely uninformative, i.e., for any value of $\Phi$ one can find some $P_{\Delta \epsilon | X} \in P$ and $P_{\Delta \eta | Y} \in P$ that can reproduce the equilibrium conditional match type probabilities, $P_{Y | X}$ and $P_{X | Y}$. Note that the vertical black lines appearing in some sub-figures are due to the scale normalisations discussed in (22) and (23).

We continue the analysis by incorporating into $P^1_{\Delta \epsilon}$ and $P^1_{\Delta \eta}$ increasingly restrictive nonparametric distributional restrictions to see how the empirical content of the model varies under different scenarios. Figure 2 reports the projections of $\Theta^*$ when the analyst assumes that the conditional probability distributions of the taste shock differences have marginals symmetric about zero. Such a restriction allows to reject significant portions of the parameter space and to identify the signs of $\Phi_{22}$ and $\Phi_{11}$. Further, imposing that the conditional probability distributions of the taste shock differences are characterised by identical marginals (Figure 3) does not seem to induce noticeable improvements. Figure 4 reports the projections of $\Theta^*$ when the analyst assumes that the taste shock differences are independent of the agents’ types and that their marginal probability distributions are symmetric about zero and identical. As before, these restrictions are sufficient to identify the signs of $\Phi_{22}$ and $\Phi_{11}$.

Figure 5 adds to the scenario of Figure 3 the assumption that the 2-dimensional vectors of taste shock differences which are relevant for each type choice are identically distributed conditional on the agents’ types. This enables us to identify the sign of all the components of $\Phi$. Moreover, the bounds on $\Phi_{12}$ and $\Phi_{21}$ are very tight. Further, imposing that the taste shock differences are independent of the agents’ types (Figure 6) permits us to identify the sign of all the components of $\Phi$ and to obtain a narrow projection for $\Phi_{12}$. This last case is as close as one can get to the framework of Choo

---

18 Note that the black regions in Figure 4 are not expected to be tighter than those in Figure 3. This is because, as explained earlier, we impose less scale normalisations into $U^1$ and $V^1$ when $P^1_{\Delta \epsilon}$ and $P^1_{\Delta \eta}$ incorporate independence between the taste shock differences and the agents’ types.

19 Remember that, as highlighted by Example 1, when man $i \in I$ of type $x \in X$ decides whether to choose a woman of type $y \in Y_i$, he compares $U_{x y} + \epsilon_{iy}$ with $U_{x 0} + \epsilon_{i0}$ and $U_{x y} + \epsilon_{ig}$ for $\tilde{y} \neq y$. Hence, the vector of taste shock differences that are relevant for such a type choice is $(\epsilon_{iy} - \epsilon_{i0}, \epsilon_{iy} - \epsilon_{ig})$ with size $2 \times 1$. Similar considerations can be made for the women’s side.

20 Note that the black regions in Figure 6 are not expected to be tighter than those in Figure 5. See footnote 18 for an explanation.
and Siow (2006) without assuming mutually independent and Gumbel distributed taste shocks. Hence, the sizes of the black regions can be interpreted as the cost of removing those restrictions on the latent variables.\footnote{It is worthwhile to point out here that it is \textit{not} possible to test the Gumbel assumption or any other probability distribution of the taste shocks. This is because Galichon and Salanié (2015) prove that if $P_{\epsilon|X}$ and $P_{\eta|Y}$ are fully known, then there always exists one (and only one) value of $\Phi$ that can reproduce the empirical conditional match types probabilities.}

Figures 7, 8, 9, and 10 are based on DGP2. The figures do not include the projections of the normalised true value of $\Phi$ (red dots) because their calculations are very complicated under DGP2.\footnote{The formulas are provided by Proposition 2 in Galichon and Salanié (2015) based on the Envelope Theorem.} The black regions look overall less informative than under DGP1. For example, under all the scenarios considered, we obtain unbounded intervals for every component of $\Phi$. This suggests that, when the underlying data generating process is characterised by very correlated taste shocks, a substantial part of the empirical content of the model is determined by the assumed parametric specification of the conditional probability distribution of the latent variables together with the numerical values assigned to its parameters.
Figure 1: The figure is based on DGP1. The black regions represent the projections of $\Theta^\star$ along every two of its dimensions when no restriction on the conditional probability distributions of the taste shocks are incorporated. The red dots represent the projections of the true value of $\Phi$. The location and scale normalisations discussed in (22) and (23) are imposed.

Figure 2: The figure is based on DGP1. The black regions represent the projections of $\Theta^\star$ along every two of its dimensions when the researcher assumes that the conditional probability distributions of the taste shock differences have marginals symmetric about zero. The red dots represent the projections of the true value of $\Phi$. The location and scale normalisations discussed in (22) and (23) are imposed.
Figure 3: The figure is based on DGP1. The black regions represent the projections of $\Theta^*$ along every two of its dimensions when the researcher assumes that the conditional probability distributions of the taste shock differences have marginals symmetric about zero and identical. The red dots represent the projections of the true value of $\Phi$. The location and scale normalisations discussed in (22) and (23) are imposed. Imposing in addition to the scenario of Figure 2 that the conditional probability distributions of the taste shock differences are characterised by identical marginals does not noticeably improve the identifying power of the model.

Figure 4: The figure is based on DGP1. The black regions represent the projections of $\Theta^*$ along every two of its dimensions when the researcher assumes that the taste shock differences are independent of the agents’ types with marginal probability distributions identical and symmetric about zero. The red dots represent the projections of the true value of $\Phi$. The location and scale normalisations discussed in (24) and (25) are imposed.
Figure 5: The figure is based on DGP1. The black regions represent the projections of $\Theta^*$ along every two of its dimensions when the researcher assumes that (i) the conditional probability distributions of the taste shock differences have marginals identical and symmetric about zero, and (ii) the 2-dimensional vectors of taste shock differences which are relevant for each type choice are identically distributed conditional on the agents’ types. The red dots represent the projections of the true value of $\Phi$. The location and scale normalisations discussed in (22) and (23) are imposed.

Figure 6: The figure is based on DGP1. The black regions represent the projections of $\Theta^*$ along every two of its dimensions when the researcher assumes that (i) the taste shock differences are independent of the agents’ types with marginal probability distributions identical and symmetric about zero, and (ii) the 2-dimensional vectors of taste shock differences which are relevant for each type choice are identically distributed. The red dots represent the projections of the true value of $\Phi$. The location and scale normalisations discussed in (24) and (25) are imposed. This is as close as one can get to the framework of Choo and Siow (2006) without assuming mutually independent and Gumbel distributed taste shocks.
Figure 7: The figure is based on DGP2. The black regions represent the projections of $\Theta^\star$ along every two of its dimensions when the researcher assumes that the conditional probability distributions of the taste shock differences have marginals symmetric about zero. The location and scale normalisations discussed in (22) and (23) are imposed.

Figure 8: The figure is based on DGP2. The black regions represent the projections of $\Theta^\star$ along every two of its dimensions when the researcher assumes that the taste shock differences are independent of the agents’ types with marginal probability distributions identical and symmetric about zero. The location and scale normalisations discussed in (24) and (25) are imposed.
Figure 9: The figure is based on DGP2. The black regions represent the projections of $\Theta^*$ along every two of its dimensions when the researcher assumes that (i) the conditional probability distributions of the taste shock differences have marginals identical and symmetric about zero, and (ii) the 2-dimensional vectors of taste shock differences which are relevant for each type choice are identically distributed conditional on the agents’ types. The location and scale normalisations discussed in (22) and (23) are imposed.

Figure 10: The figure is based on DGP2. The black regions represent the projections of $\Theta^*$ along every two of its dimensions when the researcher assumes that (i) the taste shock differences are independent of the agents’ types with marginal probability distributions identical and symmetric about zero, and (ii) the 2-dimensional vectors of taste shock differences which are relevant for each type choice are identically distributed. The location and scale normalisations discussed in (24) and (25) are imposed.
5 Inference

Section 3 studies identification of the vector of type-specific match surpluses, $\Phi$, by relying on the assumption that the probability distributions $\{P_{Q_{M,i}}|X, P_{Q_{W,j}}|Y, P_X, P_Y\}$ are known by the analyst. However, when doing an empirical analysis, the analyst should replace these probability distributions with their sample analogues as resulting from having i.i.d. observations $\{Q_{M,i}, X_i, Q_{W,j}, Y_j\}_{i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\}}$. In what follows we illustrate how to construct a $(1 - \alpha)$ confidence region for each $\Phi \in \Theta^\star$, with $\alpha \in (0, 1)$.

We reformulate the characterization of $\Theta^\star$ in terms of unconditional moment equalities and then apply results from the recent literature on inference in unconditional moment (in)equality models. In particular, we suggest to use a computationally tractable version of the profiled subsampling technique illustrated by Politis and Romano (1994), Romano and Shaikh (2008; 2012).

To keep the exposition readable, we continue the discussion by focusing on the case $r = 2$ ($d = 3$) of Example 1. Given $U \in U^1$ and $V \in V^1$, consider the linear equalities in (17) for each side of the market

$$
P_{Q_{M,i}}(1) = 1 + G_{\Delta|x}^U(U_{x_0} - U_{x_1}, \infty, U_{x_2} - U_{x_1}) - G_{\Delta|x}^U(\infty, \infty, U_{x_2} - U_{x_1}) - G_{\Delta|x}^U(U_{x_0} - U_{x_1}, \infty, \infty),$$

$$P_{Q_{M,i}}(2) = G_{\Delta|x}^U(\infty, \infty, U_{x_2} - U_{x_1}) - G_{\Delta|x}^U(\infty, U_{x_0} - U_{x_2}, U_{x_2} - U_{x_1}),$$

$$P_{Q_{M,i}}(0) = G_{\Delta|x}^U(U_{x_0} - U_{x_1}, U_{x_0} - U_{x_2}, \infty),$$

$$P_{Q_{W,j}}(1) = 1 + G_{\Delta|y}^V(V_{y_0} - V_{y_1}, \infty, V_{y_2} - V_{y_1}) - G_{\Delta|y}^V(\infty, \infty, V_{y_2} - V_{y_1}) - G_{\Delta|y}^V(V_{y_0} - V_{y_1}, \infty, \infty),$$

$$P_{Q_{W,j}}(2) = G_{\Delta|y}^V(\infty, \infty, V_{y_2} - V_{y_1}) - G_{\Delta|y}^V(\infty, V_{y_0} - V_{y_2}, V_{y_2} - V_{y_1}),$$

$$P_{Q_{W,j}}(0) = G_{\Delta|y}^V(V_{y_0} - V_{y_1}, V_{y_0} - V_{y_2}, \infty),$$

$$\forall(x, y) \in \mathcal{X} \times \mathcal{Y}, \quad \text{(26)}$$

where the functions $G_{\Delta|x}^U : \mathcal{A}_x(U) \to \mathbb{R}$ and $G_{\Delta|y}^V : \mathcal{A}_y(V) \to \mathbb{R}$ are constructed, as explained earlier, according to the nonparametric distributional assumptions of interest (Section 3.2.1) and the approximated degeneracy condition (Section 3.2.3).

Let $\tilde{\omega}_{M,y|x}(G_{\Delta|x}^U)$ denote the right hand side of the equation for $P_{Q_{M,i}}(y)$ in (26), $\forall(x, y) \in \mathcal{X} \times \wp_0$. Let $\tilde{\omega}_{W,x|y}(G_{\Delta|y}^V)$ denote the right hand side of the equation for $P_{Q_{W,j}}(x)$ in (26), $\forall(x, y) \in \mathcal{X}_0 \times \mathcal{Y}$. Define

$$m_{M,i,x,y}(G_{\Delta|x}^U) \equiv 1\{Q_{M,i} = y, X_i = x\} - \tilde{\omega}_{M,y|x}(G_{\Delta|x}^U)1\{X_i = x\},$$

$$m_{W,j,x,y}(G_{\Delta|y}^V) \equiv 1\{Q_{W,j} = x, Y_j = y\} - \tilde{\omega}_{W,x|y}(G_{\Delta|y}^V)1\{Y_j = y\}.$$

23For simplicity of notation, we assume that the analyst has the same number of observations, $n$, from each side of the market.
Hence, (26) is equivalent to the system of unconditional moment equalities
\[
\begin{align*}
\mathbb{E}(m_{M,i,x,y}(\hat{G}^U_{\Delta|x})) &= 0 \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \\
\mathbb{E}(m_{W,j,x,y}(\hat{G}^V_{\Delta|y})) &= 0 \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}.
\end{align*}
\]
(27)

Using the unconditional moment equalities in (27), we now describe a test at level \( \alpha \in (0, 1) \) based on Romano and Shaikh (2008) for the null hypothesis \( H_0 : \Phi = \Phi_0 \). Then, a \((1 - \alpha)\) confidence region for each \( \Phi \in \Theta^* \) can be constructed by inverting the test, i.e., by collecting all the values \( \Phi_0 \) for which the test does not reject at level \( \alpha \).

Given \((x, y) \in \mathcal{X} \times \mathcal{Y}\), let \( S_{U,V,x,y} \) be the collection of pairs of functions \( \hat{G}^U_{\Delta|x} : \mathcal{A}_x(U) \to \mathbb{R} \) and \( \hat{G}^V_{\Delta|y} : \mathcal{A}_y(V) \to \mathbb{R} \) which satisfy the constraints guaranteeing that \( \hat{G}^U_{\Delta|x} \) is extendable to a conditional CDF in \( G^U_{\Delta|x} \), \( \hat{G}^V_{\Delta|y} \) is extendable to a conditional CDF in \( G^V_{\Delta|y} \), and such conditional CDFs are concentrated on the region \( \mathcal{B} \), as discussed in Sections 3.2.1 and 3.2.3. We propose the following test statistic.

\[
TS_\alpha(\Phi_0) \equiv \inf_{U^\dagger, V^\dagger} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left( \sqrt{n} \hat{m}_{M,n,x,y}(\hat{G}^U_{\Delta|x}) \right)^2 + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left( \sqrt{n} \hat{m}_{W,n,x,y}(\hat{G}^V_{\Delta|y}) \right)^2
\]
\[
s.t. \quad U \in U^\dagger, V \in V^\dagger, \quad \{ \hat{G}^U_{\Delta|x}, \hat{G}^V_{\Delta|y} \} \in S_{U,V,x,y} \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \quad U + V = \Phi_0,
\]
(28)

where \( \hat{m}_{M,n,x,y}(\hat{G}^U_{\Delta|x}) \) and \( \hat{m}_{W,n,x,y}(\hat{G}^V_{\Delta|y}) \) are the empirical counterparts of the moments \( \mathbb{E}(m_{M,i,x,y}(G^U_{\Delta|x})) \) and \( \mathbb{E}(m_{W,j,x,y}(G^V_{\Delta|y})) \), respectively. We show in Appendix B that (28) can be equivalently rewritten as a mixed integer quadratic programming, which can be solved in Gurobi by using pre-built packages.\(^{24}\) The quadratic feature is because \( \hat{m}_{M,n,x,y}(\hat{G}^U_{\Delta|x}) \) and \( \hat{m}_{W,n,x,y}(\hat{G}^V_{\Delta|y}) \) are linear functions of \( \hat{G}^U_{\Delta|x} \) and \( \hat{G}^V_{\Delta|y} \),\(^{25}\) and, thus, the objective function is quadratic. The mixed integer feature is because the constraints in \( S_{U,V,x,y} \) requiring \( \hat{G}^U_{\Delta|x} \) and \( \hat{G}^V_{\Delta|y} \) to be 3-increasing functions are relevant only for comparable 3-tuples in \( \mathcal{A}_x(U) \) and \( \mathcal{A}_y(V) \), and therefore, they are nonlinear in \( U \) and \( V \).\(^{26}\) Such nonlinear constraints can be incorporated into the problem by using auxiliary

\(^{24}\)For the validity of the procedure, it is important that the Gurobi solver finds the global minimum in (28) and does not get stuck in a local minimum. We reduce this risk by choosing a very small tolerance level, among the tuning parameters of the solver.

\(^{25}\)To preserve this linearity we have decided not to rescale the moment equalities by their standard deviations. The cost is losing the scale invariance property of the test statistic. Some of the earlier papers on inference for partially identified parameters, such as Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), and Ciliberto and Tamer (2009), consider modified methods of moments estimators that are not scale invariant. Andrews and Soares (2010) discuss that this may lead to poor power performances.

\(^{26}\)Recall that the programming discussed in Section 3.2.1 is linear in \( G^U_{\Delta|x} \) and \( G^V_{\Delta|y} \) and not in \( U \) and \( V \).
binary variables via the big-M modelling approach (e.g., Williams, 2013).

In order to obtain a critical value, we draw without replacement $B_n$ subsamples of size $b_n$ from the original sample\footnote{As highlighted in Romano and Shaikh (2008), $B_n$ and $b_n$ should be set such that $\lim_{n \to \infty} B_n = \infty$ and $\lim_{n \to \infty} \frac{b_n}{B_n} = 0$.} and compute $TS_{b_n,k}(\Phi_0)$, which is the test statistic defined in (28) using the $k$-th subsample, $\forall k \in \{1, ..., B_n\}$. Hence, the critical value $\hat{c}_{n,1-\alpha}(\Phi_0)$ corresponds to the $(1-\alpha)$ quantile of the subsampling distribution

$$L_n(t, \Phi_0) = \frac{1}{B_n} \sum_{k=1}^{B_n} 1\{TS_{b_n,k}(\Phi_0) \leq t\}.$$  

The test rejects if $TS_n(\Phi_0) > \hat{c}_{n,1-\alpha}(\Phi_0)$. In turn, the $(1-\alpha)$ confidence region for each $\Phi \in \Theta^*$ is

$$C_{n,1-\alpha} \equiv \{\Phi \in \Theta^*: TS_n(\Phi) \leq \hat{c}_{n,1-\alpha}(\Phi)\}. \quad (29)$$

6 Empirical illustration

The methodology developed can be used to empirically examine if the variations in matching patterns over time are caused by changes in the marginal probability distributions of the agents’ types, or by changes in the agents’ preferences for assortativeness in a sense that we define precisely below. In what follows we address this issue within an application in household economics. In particular, we investigate the evolution of the marriage market in the U.S. by education. The application framework and the data are taken from Chiappori, Salanié, and Weiss (2017) (hereafter, CSW). We distinguish from CSW in that we proceed without relying on parametric distributional assumptions for the latent variables, and thus we get conclusions that are possibly more robust.

The data are extracted by CSW from the American Community Survey, that is a representative part of the census. For shortness of exposition, we look at whites only, and not at whites and blacks in two distinct studies as done by CSW. We consider individuals born between 1940 and 1968. Using the definition of marriage adopted by CSW, we obtain a sample of 1,502,157 couples and 542,677 singles. We observe the latest school degree of each individual in the sample. To simplify the analysis, we regroup the five education categories specified by CSW into two levels, that are “Four-year college graduates or graduate degrees” (hereafter, CD) and “High school dropouts, high school graduates, or some college” (hereafter, NCD). We refer the reader to section 1.A in CSW for a more accurate description of the data.

By extending on Chiappori, Iyigun, and Weiss (2009), CSW combine a collective household model with a matching framework to formally explain the gender-asymmetric changes in the demand for higher education observed after 1950 in the U.S. Among the various results, CSW show that the matching assignment is positively assortative.
on education, i.e., individuals prefer to pair up with partners of the same education category, in line with Becker (1973). They further prove that the agents’ preferences for assortativeness, i.e., the extra-surplus generated when individuals pair up with partners of the same education category, has increased over time.

Providing an empirical confirmation of these theoretical findings is challenging. For example, with respect to the predicted change in the agents’ preferences for assortativeness over time, the proportion of couples in the U.S. in which both partners have a college degree has certainly increased in recent decades (Figure 11).

![Figure 11](image1.png)

**Figure 11:** Proportion of couples in the U.S. in which both partners have a college degree.

However, when examining the sources of this trend, the researcher should disentangle the part that is potentially due to an increase in the returns from homogamy (i.e., a change in $\Phi$), as conjectured by CSW’s model, from the part that is mechanically due to an increase in the proportion of educated women (i.e., a change in $P_Y$), as depicted in Figure 12.

![Figure 12](image2.png)

**Figure 12:** Proportion of individuals with and without a college degree by gender.
In the second part of the paper, CSW address such a question by using the matching framework of Section 2, under the assumption that the taste shocks on each side of the market are distributed independently of the agents’ types as i.i.d. Gumbel with scale 1 and location 0. We want to investigate the same issue without relying on parametric distributional assumptions for the latent variables, but just on nonparametric restrictions.

Let \( \mathcal{X} = \mathcal{Y} \equiv \{1, 2\} \) (hence, \( r = 2 \) and \( d = 3 \) as in Example 1), where “2” denotes education category CD and “1” denotes education category NCD. As per CSW, the agents’ preferences for assortativeness are measured by the supermodular core of \( \Phi \),

\[
D(\Phi) \equiv \Phi_{22} + \Phi_{11} - \Phi_{12} - \Phi_{21}.
\]

If \( D(\Phi) \geq 0 \) (\( D(\Phi) < 0 \)), then the matching process shows a tendency towards positive (negative) assortativeness, because agents in the same education category act as complements (substitutes) and produce a higher (lower) surplus when pairing up.

In order to test whether there have been changes in the agents’ preferences for assortativeness over time, one has to define and separate “matching markets”. As per CSW, matching market \( t \) involves the men whose year of birth is \( t \) and the women whose year of birth is \( t + 1 \), for a total of 28 matching markets. While CSW compare every two subsequent matching markets, here for an illustration of our method we focus on the matching market 1940 (hereafter, matching market \( t_1 \)) and the matching market 1967 (hereafter, matching market \( t_2 \)).

We assume that a match process featuring the characteristics described in Section 2 takes place in markets \( t_1 \) and \( t_2 \) with primitives denoted by \( \{\Phi_{t_1}, P^t_{
abla X}, P^t_{\eta|Y}, P^t_{\epsilon|X}, P^t_{\eta|Y}\} \) and \( \{\Phi_{t_2}, P^{t_2}_{\nabla X}, P^{t_2}_{\eta|Y}, P^{t_2}_{\epsilon|X}, P^{t_2}_{\eta|Y}\} \), respectively. To determine whether the changes in the matching patterns are due to changes in the marginal probability distributions of the agents’ types or changes in the surplus parameters, we abstract from variations in the conditional probability distributions of the latent variables. In particular, we impose that

\[
P^t_{\epsilon|X} = P^{t_2}_{\epsilon|X} \equiv P_{\epsilon|X} \quad \text{and} \quad P^t_{\eta|Y} = P^{t_2}_{\eta|Y} \equiv P_{\eta|Y}.
\]

The null of the test is

\[
H_0 : D(\Phi_{t_1}) - D(\Phi_{t_2}) = 0.
\]

It maintains that there have been no changes in the agents’ preferences for assortativeness across the two matching markets considered. Thus, it hypothesises that the observed changes in matching patterns over time are simply caused by changes in the marginal probability distributions of the agents across education categories.\(^{28}\)

Mimicking the profiled subsampling procedure illustrated by Romano and Shaikh

\(^{28}\) We refer the reader to Section III of CSW for a detailed explanation on how (30) allows \( \Phi_{t_1} \) and \( \Phi_{t_2} \) to be related through additively separable time-specific drifts.
(2008) and discussed in Section 5, we propose the following test statistic:

$$\text{TS}_{n_{t1}, n_{t2}} \equiv \inf_{U_{t1}, V_{t1}, U_{t2}, V_{t2}} \left\{ \sum_{t \in \{t_1, t_2\}} n_t \times \left[ \sum_{(x,y) \in X \times Y} \left( \hat{m}_{M,n,x,y}^0(G_{\Delta x}^{U_{t1}, U_{t2}}) \right)^2 + \sum_{(x,y) \in X \times Y} \left( \hat{m}_{W,n,x,y}^0(G_{\Delta y}^{V_{t1}, V_{t2}}) \right)^2 \right] \right\}$$

s.t. $U_{t1} \in U^1, U_{t2} \in U^2, V_{t1} \in V^1, V_{t2} \in V^2,$

$$\{G_{\Delta x}^{U_{t1}, U_{t2}}(\varphi), G_{\Delta y}^{V_{t1}, V_{t2}}(\varphi)\} \in S_{U_{t1}, U_{t2}, V_{t1}, V_{t2}, x, y} \ \forall (x, y) \in X \times Y,$$

$$U_{t1} + V_{t1} = \Phi_{t1}, U_{t2} + V_{t2} = \Phi_{t2},$$

$$D(\Phi_{t1}) - D(\Phi_{t2}) = 0,$$

under the null hypothesis. We use subsampling in order to obtain a critical value, $\hat{c}_{n_{t1}, n_{t2}, 1 - \alpha},$ with $\alpha \in (0, 1).$ The test rejects at level $\alpha$ if $\text{TS}_{n_{t1}, n_{t2}} > \hat{c}_{n_{t1}, n_{t2}, 1 - \alpha}.$

We first test $H_0$ under the assumption that the conditional probability distributions of the taste shock differences have marginals symmetric about zero. Also, we impose the location normalisations in (22) and (23), but proceed without scale normalisations. Indeed, we remind the reader that the location and scale normalisations in (22) and (23) are imposed on the vectors $U_{t1}, V_{t1}, U_{t2}, V_{t2}$ which are equilibrium objects (and, hence, they can vary across markets when $P_{X}^{t1} \neq P_{X}^{t2}$ and/or $P_{Y}^{t1} \neq P_{Y}^{t2},$ even if $\Phi_{t1} = \Phi_{t2}),$ and not primitive parameters. While the location normalisations are differentiated out in (30), the scale normalisations may induce the analyst to wrongly reject $H_0,$ for example when $\Phi_{t1} = \Phi_{t2}$ but $P_{X}^{t1} \neq P_{X}^{t2}$ and/or $P_{Y}^{t1} \neq P_{Y}^{t2}.$ Avoiding scale normalisations reduces the computational speed of the testing procedure by increasing the cardinality of $U^1$ and $V^1,$ but the test remains valid. The test does not reject $H_0$ at level $\alpha = 0.95.$ This means that there might have been no changes in the agents’ preferences for assortativeness across the two matching markets considered. Hence, the observed changes in matching patterns over time could be simply caused by changes in the marginal probability distributions of the agents across education categories. Such a conclusion is robust to various number of subsamples ($B_{n_t} = 50, 100, 300$ for each matching market $t \in \{t_1, t_2\}$) and various subsample sizes ($b_{n_t} = \left\lceil \frac{5}{6} n_t \right\rceil, \left\lceil \frac{2}{3} n_t \right\rceil, \left\lceil \frac{2}{3} n_t \right\rceil, \left\lceil \frac{2}{3} n_t \right\rceil, \left\lceil \frac{1}{6} n_t \right\rceil$ for each matching market $t \in \{t_1, t_2\}$), with subsamples drawn with or without replacements. In particular, the test statistic is equal to 175.041, while the critical value takes value between 487.517 and 1452.461 across the various subsampling schemes attempted. The results just discussed differ from CSW which reject $H_0$ at level $\alpha = 0.95$ under the Gumbel assumption.

To see whether a rejection outcome can be obtained in our setting, we test $H_0$ under stronger nonparametric distributional restrictions on the taste shocks differences. Specifically, we assume that (i) the conditional probability distributions of the taste

---

29 For each matching market $t \in \{t_1, t_2\},$ for simplicity of notation we assume that the analyst has the same number of observations, $n_t,$ from each side of the market.
shock differences have marginals identical and symmetric about zero, and (ii) the 2-
dimensional vectors of taste shock differences which are relevant for each type choice are
identically distributed conditional on the agents’ types. As discussed above, we impose
the location normalisations in (22) and (23), but proceed without scale normalisations.
The test now rejects $H_0$ at level $\alpha = 0.95$, in line with CSW. This means that there
have been changes in the agents’ preferences for assortativeness across the two match-
ing markets considered. Such a conclusion is robust to various number of subsamples
$(B_n = 50, 100, 300$ for each matching market $t \in \{t_1, t_2\})$ and various subsample sizes
$(b_n = \left\lceil \frac{n}{6} \right\rceil, \left\lceil \frac{n}{2} \right\rceil, \left\lceil \frac{n}{3} \right\rceil, \left\lceil \frac{n}{4} \right\rceil, \left\lceil \frac{n}{6} \right\rceil$ for each matching market $t \in \{t_1, t_2\}$), with subsamples
drawn with or without replacements. In particular, the test statistic is equal to 8195.558,
while the critical value takes value between 2253.917 and 6653.370 across the various
subsampling schemes attempted.

7 Conclusions

In recent times, there has been a surge in the application of empirical matching models.
Thus, it is crucial to understand what drives their identification, as robustly as possible.
This paper is an attempt in that direction. In particular, we focus on the one-to-one TU
matching model of Choo and Siow (2006) and Galichon and Salanié (2015). When the
analyst has data on one large market only, we study identification of the preference pa-
rameters without imposing parametric restrictions on the probability distribution of the
agents’ unobserved characteristics. We provide a tractable procedure to characterise and
conduct inference on the sharp identified set for the preference parameters, under various
classes of nonparametric distributional assumptions on the agents’ unobserved charac-
teristics. Simulations suggest that the model is informative about the preference pa-
rameters under various classes of nonparametric distributional restrictions on the agents’
unobserved characteristics. Lastly, we use our methodology to empirically investigate if
the variations in marriage matching patterns observed over time in the U.S. are caused
by changes in the agents’ preferences for education assortativeness, or by changes in the
marginal probability distributions of the agents’ types. Under a symmetry restriction on
the marginals of the taste shocks, we are unable to reject the null hypothesis that the
matching preferences have been invariant over time. Instead, when we impose stronger
nonparametric distributional assumptions on the latent variables, we reject such a null.
We leave to future work similar robust identification and inference exercises applied to
other matching frameworks.
References


A Proofs

Proof of Proposition 1 As mentioned in the main text, we refer the reader to Galichon and Salanié (2015) for the proof of Proposition 1.

Proof of Proposition 2 As mentioned in the main text, we refer the reader to Torgovitsky (2018) for the proof of Proposition 2 and to Assumption A in Torgovitsky (2018) for an accurate list of the nonparametric distributional assumptions on the taste shock differences that can be accommodated.

Proof of Proposition 3 The proof is organised in the following steps. In step 0 we introduce some new notation which is helpful to formalise our arguments. In step 1 we present the notion of equivalence class for every $U \in \mathcal{U}^i$ and prove that if $\tilde{U}, \hat{U} \in \mathcal{U}^i$ belong to the same equivalence class, then they have the same solution to the linear programming problem (19). The proof for a generic case follows exactly the same

For simplicity of exposition, we provide the proof of Proposition 3 referring to the linear programming problem (19). The proof for a generic case follows exactly the same steps, but becomes notationally more complicated.

Step 0 Let $x \in \mathcal{X}$ and $U \in \mathcal{U}^i$. List the 4 elements of the set $A_{x,j}(U)$ in a row vector, $\alpha_{x,U,j}$, for $j = 1, 2, 3$.\(^{30}\)

Let

$$\Pi_1 \equiv \{(i, j, k, l) : (i, j, k, l) \in \{1, 2, 3, 4\}^4, i \neq j \neq k \neq l\},$$

$$\Pi_2 \equiv \{(<, <, <), (<, <, =), (<, =, <), (<, =, =), (>, <, <), (>, <, =), (>, =, <), (>, =, =)\}.$$

Let $\pi : \mathbb{R}^4 \to \Pi_1 \times \Pi_2$ where $\pi(\alpha) \equiv (\pi_1(\alpha), \pi_2(\alpha))$. $\pi_1(\alpha)$ sorts the 4 elements of $\alpha$ from smallest to largest and reports their positions in the original vector. $\pi_2(\alpha)$ reports the relational operators, $<$ or $=$, among the sorted elements of $\alpha$. When $\alpha$ contains multiple elements with the same value, then we adopt any convention on which element to sort first. E.g., $\pi(100, 99, \infty, -\infty) = \{(4, 2, 1, 3), (<, <, <)\}$ and $\pi(5, 5, \infty, -\infty) = \{(4, 1, 2, 3), (<, =, <)\}$. Call $\pi(\alpha)$ the “$\pi$-ordering” of $\alpha$.

Consider the function $\tilde{G}_{\Delta|x}^U : A_x(U) \to \mathbb{R}$ introduced in Section 3.2.1. Let $\tilde{G}_{\Delta|x}^U(A_x(U))$ denote the image set of this function. Recall that $\tilde{G}_{\Delta|x}^U(A_x(U))$ is a finite set because $A_x(U)$ is a finite set.\(^{31}\) List the 4\(^3\) elements of $\tilde{G}_{\Delta|x}^U(A_x(U))$ in a column vector, $g_{x,U}$.

\(^{30}\)Recall that $4^3$ is the cardinality of $A_x(U)$ in the linear programming problem (19) when also possibly repeated elements in $A_{x,j}(U)$ for $j = 1, 2, 3$ are counted.

\(^{31}\)As discussed for $A_x(U)$ in Footnote 30, the cardinality of $\tilde{G}_{\Delta|x}^U(A_x(U))$ is also $4^3$. 

Observe that solving the linear programming problem (19) can be equivalently viewed as finding $g_{x,U}$ whose elements satisfy a collection of linear constraints.

Define a function $\iota : C_{A_x(U)}^{1,2} \rightarrow \{1, 2, \ldots, 4^3\}$, where $\iota(k)$ is the row index of scalar $k$ in the column vector $g_{x,U}$.

List the $4^3$ elements of the set $A_x(U)$ in a column vector of 3-tuples, $a_{x,U}$. $a_{x,U}$ can be viewed as a $4^3 \times 3$ matrix. Reorder the matrix $a_{x,U}$ lexicographically by row and call the reordered matrix $a_{x,U}$.\(^{32}\)

Define a function $\tau : A_x(U) \rightarrow \{1, 2, \ldots, 4^3\}$, where $\tau(k)$ is the row index of 3-tuple $k$ in the matrix $a_{x,U}^L$.

**Step 1** Let $x \in X$ and $\hat{U}, \hat{U} \in U^l$. Define

$$C_x(\hat{U}) \equiv \left\{ \{ (\hat{i}, \hat{q}, \hat{r}), (\hat{\ell}, \hat{q'}, \hat{r'}) \} : (\hat{i}, \hat{q}, \hat{r}), (\hat{\ell}, \hat{q'}, \hat{r'}) \in A_x(\hat{U}), (\hat{i}, \hat{q}, \hat{r}) \leq (\hat{\ell}, \hat{q'}, \hat{r'}) \right\}.$$  

$$C_x^l(\hat{U}) \equiv \left\{ \{ (i, q, r), (\ell, q', r') \} : (i, q, r), (\ell, q', r') \in A_x(\hat{U}), (i, q, r) \leq (\ell, q', r') \right\}.$$  

**Definition 3.** (Equivalence class) Let $\hat{U}, \hat{U} \in U^l$. $\hat{U}$ is said to belong to the equivalence class of $\hat{U}$ at $x \in X$ if the following two conditions hold:

1. For every $\{ (\hat{i}, \hat{q}, \hat{r}), (\hat{\ell}, \hat{q'}, \hat{r'}) \} \in C_x(\hat{U})$, there exists $\{ (i, q, r), (\ell, q', r') \} \in C_x^l(\hat{U})$ such that

   $$\iota(G_{\Delta|X}^{\ell}(\hat{i}, \hat{q}, \hat{r})) = \iota(G_{\Delta|X}^{\ell}(i, q, r)),$$

   $$\iota(G_{\Delta|X}^{\ell}(\hat{\ell}, \hat{q}, \hat{r})) = \iota(G_{\Delta|X}^{\ell}(\ell, q', r)),$$

   $$\iota(G_{\Delta|X}^{\ell}(\hat{i}, \hat{q'}, \hat{r})) = \iota(G_{\Delta|X}^{\ell}(\hat{i}, \hat{q'}, \hat{r})),$$

   $$\iota(G_{\Delta|X}^{\ell}(\hat{i}, \hat{q'}, \hat{r})) = \iota(G_{\Delta|X}^{\ell}(\hat{i}, \hat{q'}, \hat{r})),$$

   $$\iota(G_{\Delta|X}^{\ell}(\hat{i}, \hat{q'}, \hat{r})) = \iota(G_{\Delta|X}^{\ell}(\hat{i}, \hat{q'}, \hat{r})),$$

   and vice versa.

2. $\pi_2(\alpha_{x,U,j}) = \pi_2(\alpha_{x,U,j})$ for $j = 1, 2, 3$.

Let $[\hat{U}]_x$ denote the equivalence class of $\hat{U}$ at $x \in X$. \(\diamond\)

\(^{32}\)For example, suppose that $A_x(U)$ contains three 3-tuples (instead of $4^3$ 3-tuples, for shortness of exposition): $(3, 1, 2), (2, 3, 4), (2, 1, 3)$. Then,

$$a_{x,U}^L \equiv \begin{pmatrix} 2 & 1 & 3 \\ 2 & 3 & 4 \\ 3 & 1 & 2 \end{pmatrix}.$$
Lemma A.1. Let \( x \in X \) and \( \bar{U}, \hat{U} \in \mathcal{U}^t \). If \( \bar{U} \in [\hat{U}]_x \), then \( \bar{U} \) and \( \hat{U} \) induce the same solution to the linear programming problem (19) at \( x \in X \).

Proof. Let \( x \in X \) and \( \bar{U}, \hat{U} \in \mathcal{U}^t \). Suppose \( \bar{U} \in [\hat{U}]_x \). Take any \( \{(\hat{t}, \hat{q}, \hat{r}), (\hat{t}', \hat{q}', \hat{r}')\} \in \mathcal{C}_x(\hat{U}) \) and a corresponding \( \{(\bar{t}, \bar{q}, \bar{r}), (\bar{t}', \bar{q}', \bar{r}')\} \in \mathcal{C}_x(\bar{U}) \) such that conditions 1 and 2 of Definition 3 hold. Consider constraint (20) at \( \{(\bar{t}, \bar{q}, \bar{r}), (\bar{t}', \bar{q}', \bar{r}')\} \), where the terms of the form \( G_{\Delta|}\cdot(\cdot) \) are unknown parameters to be determined. Relabel them as \( \theta_t(G_{\Delta|}(\cdot)) \).

Then, restate (20) as

\[
-\theta_t(G_{\Delta|}(\bar{t}, \bar{q}, \bar{r})) + \theta_t(G_{\Delta|}(\bar{t}', \bar{q}', \bar{r})) + \theta_t(G_{\Delta|}(\bar{t}, \bar{q}, \bar{r}')) - \theta_t(G_{\Delta|}(\bar{t}', \bar{q}', \bar{r}')) \\
+ \theta_t(G_{\Delta|}(\bar{t}, \bar{q}, \bar{r})) - \theta_t(G_{\Delta|}(\bar{t}', \bar{q}, \bar{r})) - \theta_t(G_{\Delta|}(\bar{t}, \bar{q}', \bar{r}')) + \theta_t(G_{\Delta|}(\bar{t}', \bar{q}', \bar{r}')) \geq 0,
\]

(A.1)

where \( \theta \) is a \( 4^3 \times 1 \) vector of unknowns and \( \theta_h \) denotes the \( h \)-th element of \( \theta \). Similarly, consider the following relabelled constraint corresponding to \( \bar{U} \),

\[
-\theta_t(G_{\Delta|}(\bar{t}, \bar{q}, \bar{r})) + \theta_t(G_{\Delta|}(\bar{t}', \bar{q}, \bar{r})) + \theta_t(G_{\Delta|}(\bar{t}, \bar{q}', \bar{r}')) - \theta_t(G_{\Delta|}(\bar{t}', \bar{q}', \bar{r}')) \\
+ \theta_t(G_{\Delta|}(\bar{t}, \bar{q}, \bar{r})) - \theta_t(G_{\Delta|}(\bar{t}', \bar{q}', \bar{r})) - \theta_t(G_{\Delta|}(\bar{t}, \bar{q}', \bar{r})) + \theta_t(G_{\Delta|}(\bar{t}', \bar{q}', \bar{r})) \geq 0.
\]

(A.2)

By condition 1 of Definition 3, the subscripts of \( \theta \) in (A.1) and (A.2) are identical. Further, if some or all of the components of \((\bar{t}, \bar{q}, \bar{r})\) are equal to \((\hat{t}', \hat{q}', \hat{r}')\), then condition 2 of Definition 3 ensures that the same is true across \((\bar{t}, \bar{q}, \bar{r})\) and \((\hat{t}', \hat{q}', \hat{r}')\). Therefore, (A.1) and (A.2) are identical. Similar arguments can be repeated for every \( \{(\hat{t}, \hat{q}, \hat{r}), (\hat{t}', \hat{q}', \hat{r}')\} \in \mathcal{C}_x(\hat{U}) \) and \( \{(\bar{t}, \bar{q}, \bar{r}), (\bar{t}', \bar{q}', \bar{r}')\} \in \mathcal{C}_x(\bar{U}) \) so that \( \bar{U} \) and \( \hat{U} \) generate the same constraints of the type (20).

If \( \bar{U} \) and \( \hat{U} \) generate the same constraints of the type (20), then they induce the same solution to the linear programming problem (19) at \( x \in X \). This is because the only piece of (19) that can potentially generate different solutions for different values of \( U \) is the one requiring the function \( \bar{G}_{\Delta|} : \mathcal{A}_x(U) \to \mathbb{R} \) to be 3-increasing.

\[\square\]

Step 2

Lemma A.2. Let \( x \in X \) and \( \bar{U}, \hat{U} \in \mathcal{U}^t \). If

i. \( \pi_1(\alpha_{x,\bar{U},j}) = \pi_1(\alpha_{x,\hat{U},j}) \) for \( j = 1, 2, 3 \),

ii. \( \pi_2(\alpha_{x,\bar{U},j}) = \pi_2(\alpha_{x,\hat{U},j}) \) for \( j = 1, 2, 3 \),

then \( \hat{U} \in [\bar{U}]_x \).

Proof. Condition ii of Lemma A.2 coincides with condition 2 of Definition 3. Therefore, in what follows we show that conditions i and ii of Lemma A.2 imply condition 1 of Definition 3.

\[\text{Ref to equation in the main text.}\]
Let \( x \in \mathcal{X} \). Let \( \hat{U}, \hat{\hat{U}} \in \mathcal{U}^\dagger \) such that \( \pi(\alpha_{x,\hat{U}_j}) = \pi(\alpha_{x,\hat{\hat{U}}_j}) \) for \( j = 1, 2, 3 \). Take any \( \{(\hat{t}, \hat{q}, \hat{r}), (\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})\} \in \mathcal{C}_x(\hat{U}) \), i.e., a comparable pair of 3-tuples from \( \mathcal{A}_x(\hat{U}) \). Pick \( (\hat{t}, \hat{q}, \hat{r}), (\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}}) \in \mathcal{A}_x(\hat{U}) \) such that \( \tau((\hat{t}, \hat{q}, \hat{r})) = \tau((\hat{t}, \hat{q}, \hat{r})) \) and \( \tau((\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})) = \tau((\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})) \). Given that \( \pi_1(\alpha_{x,\hat{U}_j}) = \pi_1(\alpha_{x,\hat{\hat{U}}_j}) \) for \( j = 1, 2, 3 \), it should be that \( \{(\hat{t}, \hat{q}, \hat{r}), (\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})\} \in \mathcal{C}_x(\hat{U}) \). That is, by construction, \( (\hat{t}, \hat{q}, \hat{r}), (\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}}) \) should be a comparable pair of 3-tuples from \( \mathcal{A}_x(\hat{U}) \). Moreover, given \( \pi_1(\alpha_{x,\hat{U}_j}) = \pi_1(\alpha_{x,\hat{\hat{U}}_j}) \) for \( j = 1, 2, 3 \), it should be that

\[
\begin{align*}
\tau((\hat{t}, \hat{q}, \hat{r})) &= \tau((\hat{t}, \hat{q}, \hat{r})), \\
\tau((\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})) &= \tau((\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})), \\
\tau((\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})) &= \tau((\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})), \\
\tau((\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})) &= \tau((\hat{\hat{t}}, \hat{\hat{q}}, \hat{\hat{r}})).
\end{align*}
\]

Construct a \( 4^3 \times 3 \) new matrix, where the first, second, and third columns are defined as

\[
\begin{align*}
\mathbf{1}_{4^0} \otimes \left( (\pi_1(\alpha_{x,\hat{U}_1}))^T \otimes \mathbf{1}_{4^2} \right), \\
\mathbf{1}_{4^1} \otimes \left( (\pi_1(\alpha_{x,\hat{U}_2}))^T \otimes \mathbf{1}_{4^1} \right), \\
\mathbf{1}_{4^2} \otimes \left( (\pi_1(\alpha_{x,\hat{U}_3}))^T \otimes \mathbf{1}_{4^0} \right),
\end{align*}
\]

respectively, with \( \mathbf{1}_d \) denoting the \( d \)-dimensional vector of ones. Reorder this new matrix lexicographically by row and call the reordered matrix \( b_{x,\hat{U}} \). Observe that \( b_{x,\hat{U}} \) is a standardised relabelling of the matrix \( a_{x,\hat{U}}^L \).\(^{34}\) Construct the matrix \( b_{x,\hat{\hat{U}}} \) in a similar way.

Note that \( \pi_1(\alpha_{x,\hat{U}_j}) = \pi_1(\alpha_{x,\hat{\hat{U}}_j}) \) for \( j = 1, 2, 3 \) implies \( b_{x,\hat{U}} = b_{x,\hat{\hat{U}}} \). Further, \( \pi_2(\alpha_{x,\hat{U}_j}) = \pi_2(\alpha_{x,\hat{\hat{U}}_j}) \) for \( j = 1, 2, 3 \) ensures that \( b_{x,\hat{U}} \) and \( b_{x,\hat{\hat{U}}} \) are also logically equivalent. Therefore, it must be that condition 1 of Definition 3 is satisfied.

\[ \square \]

Let \( x \in \mathcal{X} \) and \( \hat{U}, \hat{\hat{U}} \in \mathcal{U}^\dagger \). Lemmas A.1 and A.2 imply that if \( \alpha_{x,\hat{U}_j} \) and \( \alpha_{x,\hat{\hat{U}}_j} \) have the same \( \pi \)-ordering for \( j = 1, 2, 3 \) (or, equivalently, if \( \hat{U} \) and \( \hat{\hat{U}} \) induce the same \( \pi \)-ordering), then \( \hat{U} \) and \( \hat{\hat{U}} \) generate the same solution to the linear programming problem (19) at \( x \in \mathcal{X} \).

**Step 3** Let \( x \in \mathcal{X} \). The equivalence classes introduced in Definition 3 satisfy the following properties:

1. *(Reflexivity)* For any \( U \in \mathcal{U}^\dagger \), \( U \in [U]_x \).

\(^{34}\)The matrix \( b_{x,\hat{U}} \) abstracts from the magnitudes of the elements of \( a_{x,\hat{U}}^L \) and captures their relative positions.
II. (Symmetry) For any $\hat{U}, \hat{U} \in U^l$, if $\hat{U} \in [\hat{U}]_x$, then $\hat{U} \in [\hat{U}]_x$.

III. (Transitivity) For any $\hat{U}, \hat{U}, U \in U^l$, if $\hat{U} \in [\hat{U}]_x$ and $\hat{U} \in [U]_x$, then $\hat{U} \in [U]_x$.

Therefore, the equivalence classes introduced in Definition 3 partition $U^l$ in mutually disjoint subsets (Theorem 1.1.1, Herstein, 1975).

This partition of $U^l$ is the most efficient (or, the coarsest) partition, in the sense that it allows us to run the linear programming problem (19) at disjoint subsets (Theorem 1.1.1, Herstein, 1975). Therefore, the equivalence classes introduced in Definition 3 partition $U^l$ in mutually disjoint subsets (Theorem 1.1.1 in Herstein (1975). This is the finer partition of $U^l$ we refer to above.

The subsets of $U^l$ inducing the same $\pi$-ordering will be called “partitioning subsets” of $U^l$ henceforth. Let $K_x$ denote the number of possible $\pi$-orderings and, hence, the number of partitioning subsets of $U^l$. We denote the partitioning subsets of $U^l$ by $U^l_{1,x}, \ldots, U^l_{K_x,x}$.

**Step 4** Let $x \in X$. Let $l_j$ be the length of the row vector $\alpha_{x,U,j}$ for $j = 1, \ldots, d$.\(^{35}\) Note that

$$K_x \leq \prod_{j=1}^{d} \binom{(l_j - 2)! \times 2^{l_j - 1 - 2}}{\infty},$$

where $(l_j - 2)!$ is the number of permutations of the elements of $\alpha_{x,U,j}$ without considering $\infty, -\infty$ whose sorted position is fixed, and $2^{l_j - 1 - 2}$ is the number of possible relational operators between the elements of $\alpha_{x,U,j}$ accounting for the fact that $-\infty$ is always followed by $<$ and $\infty$ is always preceded by $. Therefore, $K_x$ is finite.

**Step 5** Let $x \in X$. We now prove that $U^l_{k,x}$ is convex $\forall k \in \{1, \ldots, K_x\}$. Let $\hat{U}$ and $\hat{U}$ belong to the same partitioning subset, $U^l_{k,x}$, for some $k \in \{1, \ldots, K_x\}$. Note that, for any $U \in U^l$, the elements of $A_{x,j}(U)$ are linear functionals of the elements of $U$ for $j = 1, 2, 3$. Denote any two such linear functionals by $T$ and $T'$. Hence, without loss of generality, if $T(\hat{U}) < T'(\hat{U})$ and $T(\hat{U}) < T'(\hat{U})$, then

$$T(\beta\hat{U} + (1 - \beta)\hat{U}) = \beta T(\hat{U}) + (1 - \beta)T(\hat{U}) < \beta T'(\hat{U}) + (1 - \beta)T'(\hat{U}) = T'(\beta\hat{U} + (1 - \beta)\hat{U}),$$

\(^{35}\)Recall that $l_j = 4$ and $d = 3$ in the linear programming problem (19). Also, note that he length of the row vector $\alpha_{x,U,j}$ does not vary with $U, x$.  

45
for any $\beta \in (0, 1)$. Thus,

$$[\beta \hat{U} + (1 - \beta)\hat{U}] \in \mathcal{U}^3_{k,x},$$

for any $\beta \in (0, 1)$. Therefore, $\mathcal{U}^3_{k,x}$ is convex $\forall k \in \{1, \ldots, K_x\}$.

**Proof of Proposition 4** For simplicity of exposition, we provide the proof of Proposition 4 when $r = 2$ ($d = 3$). The proof for a generic $r$ follows exactly the same steps, but becomes notationally more complicated.

**Step 1** As highlighted in the main text, we should firstly observe that

$$\mathcal{B} \equiv \{(b_1, b_2, b_3) \in \mathbb{R}^3 : b_3 = b_1 - b_2\} = \{(b_1, b_2, b_3) \in \mathbb{R}^3 : b_1 = b_2 + b_3\}.$$ 

Accordingly, the relevant 3-dimensional boxes defined in Proposition 4 are

$$B_{1,2,3}(\hat{b}, \tilde{b}) \equiv \{(x, y, z) \in \mathbb{R}^3 : x > \hat{b} + \hat{b}, y \leq \hat{b}, z \leq \tilde{b}\},$$

$$Q_{1,2,3}(\hat{b}, \tilde{b}) \equiv \{(x, y, z) \in \mathbb{R}^3 : x \leq \hat{b} + \hat{b}, y > \hat{b}, z > \tilde{b}\},$$

for any $(\hat{b}, \tilde{b}) \in \mathbb{R}^2$.

**Step 2** Let $\mathcal{Q} \subset \mathbb{R}$ denote the set of rational numbers. We now show that

$$\bigcup_{(\hat{b}, \tilde{b}) \in \mathcal{Q}^2} B_{1,2,3}(\hat{b}, \tilde{b}) = \{(x, y, z) \in \mathbb{R}^3 : x > y + z\} \equiv A_1.$$ 

It is clear that $\bigcup_{(\hat{b}, \tilde{b}) \in \mathcal{Q}^2} B_{1,2,3}(\hat{b}, \tilde{b}) \subseteq A_1$. To prove the reverse, take any $(x, y, z) \in A_1$ and $\epsilon \equiv x - (y + z) > 0$. Since $\mathcal{Q}$ is dense in $\mathbb{R}$, there exists $p \in [y, y + \frac{\epsilon}{2}] \cap \mathcal{Q}$ and $q \in [z, z + \frac{\epsilon}{2}] \cap \mathcal{Q}$. Therefore, $x = y + z + \epsilon > p + q$ and, hence, $(x, y, z) \in B_{1,2,3}(p, q)$. Thus, $A_1 \subseteq \bigcup_{(\hat{b}, \tilde{b}) \in \mathcal{Q}^2} B_{1,2,3}(\hat{b}, \tilde{b})$.

**Step 3** By following the same arguments of step 2, we can show that

$$\bigcup_{(\hat{b}, \tilde{b}) \in \mathcal{Q}^2} Q_{1,2,3}(\hat{b}, \tilde{b}) = \{(x, y, z) \in \mathbb{R}^3 : x < y + z\} \equiv A_2.$$ 

**Step 4** Assume $P_{\Delta\mid X}(B_{1,2,3}(\hat{b}, \tilde{b})) = P_{\Delta\mid X}(Q_{1,2,3}(\hat{b}, \tilde{b})) = 0 \forall (\hat{b}, \tilde{b}) \in \mathcal{Q}^2$. Hence, by step 3, $A_1$ and $A_2$ are disjoint and infinitely countable unions of zero probability measure sets. Note that $\mathcal{B}^c = A_1 \cup A_2$, where $\mathcal{B}^c$ denotes the complement of the region $\mathcal{B}$ in $\mathbb{R}^3$. Therefore, $P_{\Delta\mid X}(\mathcal{B}^c) = P_{\Delta\mid X}(A_1 \cup A_2) = 0$, which is equivalent to $P_{\Delta\mid X}(\mathcal{B}) = 1$.

**Step 5** Conversely, assume $P_{\Delta\mid X}(\mathcal{B}) = 1$. This implies $P_{\Delta\mid X}(\bigcup_{(\hat{b}, \tilde{b}) \in \mathcal{Q}^2} B_{1,2,3}(\hat{b}, \tilde{b})) = 0$ and $P_{\Delta\mid X}(\bigcup_{(\hat{b}, \tilde{b}) \in \mathcal{Q}^2} Q_{1,2,3}(\hat{b}, \tilde{b})) = 0$. In turn, $P_{\Delta\mid X}(B_{1,2,3}(\hat{b}, \tilde{b})) = P_{\Delta\mid X}(Q_{1,2,3}(\hat{b}, \tilde{b})) = 0 \forall (\hat{b}, \tilde{b}) \in \mathcal{Q}^2$. 

46
B More on inference

This section illustrates how the optimisation problem (28) can be rewritten as a mixed integer quadratic programming. To keep the exposition readable, we continue focusing on the case $r = 2$ ($d = 3$) of Example 1. All the arguments can be immediately generalised to any $r$.

As explained in the main text, the mixed integer feature is because, $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}$, the constraints in $\mathcal{S}_{U,V,x,y}$ requiring $\vec{G}^U_{\Delta|x}$ and $\vec{G}^V_{\Delta|y}$ to be 3-increasing functions are relevant only for comparable 3-tuples in $\mathcal{A}_x(U)$ and $\mathcal{A}_y(V)$, and therefore, they are nonlinear in $U$ and $V$. Such nonlinear constraints can be incorporated into the problem using auxiliary binary variables via the big-M modelling approach (e.g., Williams, 2013).

More precisely, consider the men’s side and take the collection of constraints in (19) requiring $\vec{G}^U_{\Delta|x}$ to be a 3-increasing function for a given $x \in \mathcal{X}$:

$$
- \vec{G}^U_{\Delta|x}(t, q, r) + \vec{G}^U_{\Delta|x}(t', q, r) + \vec{G}^U_{\Delta|x}(t, q', r) - \vec{G}^U_{\Delta|x}(t', q', r) \\
+ \vec{G}^U_{\Delta|x}(t, q, r') - \vec{G}^U_{\Delta|x}(t', q, r') - \vec{G}^U_{\Delta|x}(t', q', r') + \vec{G}^U_{\Delta|x}(t', q', r') \geq 0 \\
\forall (t, q, r), (t', q', r') \in \mathcal{A}_x(U) \text{ s.t. } (t, q, r) \leq (t', q', r').
$$

(B.1) can be rewritten as the following collection of constraints:

$$
\forall (t, q, r), (t', q', r') \in \mathcal{A}_x(U),
- (t, q, r) + (t', q', r') \geq (0, 0, 0) \Rightarrow -\vec{G}^U_{\Delta|x}(t, q, r) + \vec{G}^U_{\Delta|x}(t', q, r) + \vec{G}^U_{\Delta|x}(t, q', r) - \vec{G}^U_{\Delta|x}(t', q', r) \\
+ \vec{G}^U_{\Delta|x}(t, q, r') - \vec{G}^U_{\Delta|x}(t', q, r') - \vec{G}^U_{\Delta|x}(t', q', r') + \vec{G}^U_{\Delta|x}(t', q', r') \geq 0.
$$

(B.2)

As per Williams (2013), (B.2) is equivalent to

$$
\forall (t, q, r), (t', q', r') \in \mathcal{A}_x(U),
(I) \ - (t, q, r) + (t', q', r') \leq M_{(t,q,r),(t',q',r')} \times (\lambda_1(t,q,r),(t',q',r'),\lambda_2(t,q,r),(t',q',r'),\lambda_3(t,q,r),(t',q',r')),
(II) \ \delta(t,q,r),(t',q',r') \geq 1 + \lambda_1(t,q,r),(t',q',r') + \lambda_2(t,q,r),(t',q',r') + \lambda_3(t,q,r),(t',q',r') - 3,
(III) \ - \vec{G}^U_{\Delta|x}(t, q, r) + \vec{G}^U_{\Delta|x}(t', q, r) + \vec{G}^U_{\Delta|x}(t, q', r) - \vec{G}^U_{\Delta|x}(t', q', r) \\
+ \vec{G}^U_{\Delta|x}(t, q, r') - \vec{G}^U_{\Delta|x}(t', q, r') - \vec{G}^U_{\Delta|x}(t', q', r') + \vec{G}^U_{\Delta|x}(t', q', r') \geq -M_{(t,q,r),(t',q',r')} (1 - \delta(t,q,r),(t',q',r')),
$$

(B.3)

where $\lambda_1(t,q,r),(t',q',r'),\lambda_2(t,q,r),(t',q',r'),\lambda_3(t,q,r),(t',q',r'),\delta(t,q,r),(t',q',r')$ are binary variables and $M_{(t,q,r),(t',q',r')}$ is chosen as small as possible\(^{36}\) but such that

$$
M_{(t,q,r),(t',q',r')} \geq 4, \quad M_{(t,q,r),(t',q',r')} \leq M_{(t,q,r),(t',q',r')} \forall U \in U^U.
$$

\(^{36}\) $M_{(t,q,r),(t',q',r')}$ unnecessarily large induces bad numerics in solvers and makes harder to solve the integer program.
To see why (B.2) is equivalent to (B.3), note that

\[- (t, q, r) + (t', q', r') > 0\]

\[\lambda_{1,(t,q,r),(t',q',r')} = \lambda_{2,(t,q,r),(t',q',r')} = \lambda_{3,(t,q,r),(t',q',r')} = 1\] so that (I) is satisfied by (B.5)

\[\delta_{(t,q,r),(t',q',r')} = 1\] so that (II) is satisfied

Desired constraint on the image set of \(\bar{G}_{\Delta_{x|x}}\) is activated through (III),

\[- (t, q, r) + (t', q', r') \leq 0\]

\[\lambda_{1,(t,q,r),(t',q',r')}, \lambda_{2,(t,q,r),(t',q',r')}, \lambda_{3,(t,q,r),(t',q',r')}\] can be 1 or 0 and, in any case, (I) is satisfied by (B.5)

\[\delta_{(t,q,r),(t',q',r')}\] can take value 1 or 0

Desired constraint on the image set of \(\bar{G}_{\Delta_{y|y}}\) are activated when \(\delta_{(t,q,r),(t',q',r')} = 1\); otherwise (III) satisfied by (B.4).

The number “4” on the right hand side of (B.4) reflects the maximum value that the left hand side of (III) in (B.3) can take. Similar steps can be replicated for the women’s side.

By using (B.3) in place of (B.1), one can rewrite the optimisation problem (28) as a mixed integer quadratic programming. Let \(\Lambda_M, \Lambda_W\) be the vectors of all the newly introduced dummy variables for each side of the market. For each \((x, y) \in \mathcal{X} \times \mathcal{Y}\), let \(\mathcal{S}_{U,V,\Lambda_M,\Lambda_W,x,y}\) be the collection of pairs of functions \(\bar{G}_{\Delta_{x|x}}^U : \mathcal{A}_x(U) \to \mathbb{R}\) and \(\bar{G}_{\Delta_{y|y}}^V : \mathcal{A}_y(V) \to \mathbb{R}\) which satisfy the constraints guaranteeing that \(\bar{G}_{\Delta_{x|x}}^U\) is extendable to a conditional CDF in \(\mathcal{G}_{\Delta_{x|x}}^U\); \(\bar{G}_{\Delta_{y|y}}^V\) is extendable to a conditional CDF in \(\mathcal{G}_{\Delta_{y|y}}^V\); and such conditional CDFs are concentrated on the region \(\mathcal{B}\), as discussed in Sections 3.2.1 and 3.2.3. Then, (28) is equivalent to

\[
\text{TS}_n(\Phi_0) \equiv \inf_{U,V} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left( \sqrt{n} \hat{m}_{M,n,x,y}(\bar{G}_{\Delta_{x|x}}^U) \right)^2 + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left( \sqrt{n} \hat{m}_{W,n,x,y}(\bar{G}_{\Delta_{y|y}}^V) \right)^2
\]

s.t. \(U \in \mathcal{U}^1, V \in \mathcal{V}^1\),

\(\{\bar{G}_{\Delta_{x|x}}^U, \bar{G}_{\Delta_{y|y}}^V\} \in \mathcal{S}_{U,V,\Lambda_M,\Lambda_W,x,y} \quad \forall (x,y) \in \mathcal{X} \times \mathcal{Y}, \quad U + V = \Phi_0\),

which corresponds to a mixed integer quadratic programming.