**PARTIAL IDENTIFICATION IN NONPARAMETRIC ONE-TO-ONE MATCHING MODELS**

Cristina Gualdani*  Shruti Sinha†

February 2019

**Abstract**

We consider the one-to-one matching models 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 systematic components of the agents’ preferences without imposing parametric restrictions on the probability distribution of the latent variables. Specifically, we provide a tractable characterisation of the region of parameter values that exhausts all the implications of the model and data (the sharp identified set), under various classes of non-parametric distributional assumptions on the unobserved terms. We discuss a way to conduct inference on the sharp identified set and conclude with Monte Carlo simulations.

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

---

*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 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\(^1\) 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 (Tervio, 2008; Gabaix and Landier, 2008), sorting of job openings to workers, and the marriage market (see Fox, 2009 for a review). In particular, Choo and Siow (2006) and Galichon and Salanié (2015) show that, if the agents belong to a finite number of observed types (as it is commonly assumed in this strand of literature), we get point identification of the systematic components of the agents’ preferences, under the assumption that the probability distribution of the agents’ unobserved characteristics (or, taste shocks) is completely known by the researcher. This, in turn, typically amounts to fixing a parametric family together with numerical values for all of its parameters. However, such an approach may raise concerns because wrong specifications of the probability distribution of the latent terms can induce inconsistent empirical results and misleading counterfactual outcomes. Our objective is to investigate identification of the systematic components of the agents’ preferences when the researcher does not impose parametric restrictions on the probability distribution of the taste shocks. We instead allow the inclusion of nonparametric distributional assumptions like symmetric marginals, stochastic independence of types, etc. We believe that performing this exercise is important because it permits to 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 variation across many i.i.d. markets. As data on many i.i.d. markets may not be always available, we consider instead a framework where the analyst collects observations from

\(^1\)One-to-One matching refers to the case where each agent 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.
one large market only.

When the researcher has data on one large market only, avoiding parametric restrictions on the probability distribution of the latent variables raises the possibility of partial identification. Consequently, this poses the challenge of characterising, in a tractable way, the region of parameter values that exhausts all the implications of the model and data (the sharp identified set), under various classes of nonparametric distributional assumptions on the unobserved terms. Our study aims to answer this methodological question.

Specifically, constructing the sharp identified set for the systematic components of the agents’ preferences requires facing two problems. First, the analyst has to find, for a given value of the systematic components of the agents’ preferences, a probability distribution of the latent variables that yields that hypothesised value. This amounts to solving an infinite-dimensional existence problem because every possible probability distribution of the latent variables is not finitely parameterised and thus represents an infinite-dimensional object. Second, the analyst should repeat the first step for every admissible value of the systematic components of the agents’ preferences. Usually, this is done in the partial identification literature by generating a more or less coarse grid of points and 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 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{2}) for both sides of the market. The theorem provides sufficient and necessary conditions for the existence of a probability distribution of the latent terms that yields the considered value of the systematic components of the agents’ preferences. 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 collection of admissible values of the systematic components of the agents’ preferences can be ex-ante partitioned into a finite number of 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 once for each subset. Overall, the procedure designed allows to feasibly construct the sharp identified set for the systematic components of the agents’ preferences under several classes of nonparametric distributional assumptions on the taste shocks. Furthermore, inference on the sharp identified set can be conducted by applying any of the available methods for parameters defined by moment inequalities and equalities (see Canay and Shaikh, 2017 for a review).

Simulations suggest that, even in a very simple setting, with limited heterogeneity in types, one market is not sufficiently informative about the systematic components of the

\textsuperscript{2}That is, partial identification by extending subdistributions.
agents’ preferences, under nonparametric distributional restrictions on the taste shocks. This is striking and illustrates the crucial role played by functional form assumptions on the joint distribution of the latent variables in delivering point identification of the payoff parameters in one market.

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 if the assignment allocation is centralised or decentralised. Seminal papers on TU matching models have been cited earlier. Important works within the NTU framework are e.g., Sørensen (2007), Dagsvik (2000), Menzel (2015), and Agarwal and Diamond (2017). The ITU framework has been introduced by Galichon, Kominers, and Weber (2018).

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 identification of multinomial choice models. More precisely, Manski (1975) and Matzkin (1993) show point identification of the payoff parameters in a multinomial choice model under two main ingredients: conditional i.i.d.-ness of the taste shocks and the presence of a continuous regressor having large support. Additionally, they put assumptions on the systematic part of payoffs, for example, an index structure or nonparametric utilities with shape restrictions. Fox (2007) proves that conditional exchangeability of the taste shocks can replace conditional i.i.d.-ness, because it is sufficient for the rank property to hold. More recently, researchers have started to extend those results to two-sided markets. In particular, Fox (2018) derives point identification in a matching framework under conditional exchangeability of the taste shocks, an index structure of the systematic part of payoffs, and the availability of a continuous regressor with large support. Furthermore, in the absence of such a special regressor, he characterises an identified region (not shown to be sharp nor studied in relation to the sharp identified set) that can be estimated. In the humblest possible way, we view our work as complementary to these papers by providing a procedure that allows to tractably and fully explore the empirical content of the model discussed above under various classes of nonparametric distributional assumptions on the latent variables.

Ultimately, this paper is also related to the literature on partial identification in applied research (see Ho and Rosen, 2017 for a review).

In what follows, Section 2 illustrates the model, Section 3 develops identification arguments, Sections 4 illustrates the construction of the sharp identified set through simulations, Section 5 discusses inference, and Section 6 provides conclusions and directions for future research.
2 The model

We focus on the one-to-one matching model with transfers of 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 and can either be matched with exactly one agent on the other side, or remain unmatched. We allow preferences to include transferable payoffs. 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. We assume that the two sides of the market are stochastically independent and that 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.

In what follows we describe the formal model. 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 $\mathcal{I}$ be the set of all men and $\mathcal{J}$ be the set of all women. We index individual men by $i \in \mathcal{I}$ and individual women by $j \in \mathcal{J}$. We indicate the outside option to remain unmatched by “0”, so that single agents are represented as being matched with “0”. We further define $\mathcal{I}_0 \equiv \mathcal{I} \cup \{0\}$ and $\mathcal{J}_0 \equiv \mathcal{J} \cup \{0\}$.

Each man $i \in \mathcal{I}$ and each woman $j \in \mathcal{J}$ are endowed with some characteristics $X_i$ and $Y_j$, whose probability distributions are denoted by $P_X$ and $P_Y$. The supports of $X_i$ and $Y_j$ are finite and indicated by $\mathcal{X}$ and $\mathcal{Y}$. As earlier, $\mathcal{X}_0 \equiv \mathcal{X} \cup \{0\}$ and $\mathcal{Y}_0 \equiv \mathcal{Y} \cup \{0\}$. $X_i$ and $Y_j$ are typically referred to as man $i$’s type and woman $j$’s type. The realisations of $X_i$ and $Y_j$ are observed by the researcher.

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

$$\tilde{\Phi}_{ij} \equiv \Phi_{xy} + \epsilon_{iy} + \eta_{xj}, \quad (1)$$

where $\Phi_{xy}$ is the type-specific match surplus and $\{\epsilon_{iy}, \eta_{xj}\}$ are continuously distributed taste shocks assigned to the agents by nature and whose realisations are unobserved by the researcher. Similarly, the payoffs that man $i \in \mathcal{I}$ of type $x \in \mathcal{X}$ and woman $j \in \mathcal{J}$ of type $y \in \mathcal{Y}$ get when staying single are

$$\tilde{\Phi}_{i0} \equiv \Phi_{x0} + \epsilon_{i0}, \quad (2)$$

and

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

\footnote{This restriction is not crucial for our discussion and can be relaxed.}
We denote by $P_{\epsilon|X}$ and $P_{\eta|Y}$ 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 $Y_j$.

It is not necessary to require that $\epsilon_i$ is the vector of taste shocks assigned by nature to man $i \in I$ and that $\eta_j$ is the vector of taste shocks assigned by nature to woman $j \in J$. What is crucial instead is that (1) does not contain a term simultaneously indexed by $i$ and $j$. In Galichon and Salanié (2015) this is referred to as the separability assumption. Imposing separability 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 $y$, or 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, i.e., it cannot be that $j$ has some unobserved preference for unobserved features of $i$ or vice versa. In other words, agents have preferences over the finite types of agents on the other side of the market and not over the individual agents. Hence, from a man’s (woman’s) point of view, women (men) of a given type are perfect substitutes. It is worth noting here, though, that we still maintain heterogeneity in preferences, as men (or, women) of the same type are allowed to have different preferences over the types of agents on the other side of the market.

Restrictions similar to separability are often introduced in models of large games with infinitely many agents where individual identities are unknown or irrelevant to the researcher (e.g., De Paula, Richards-Shubik, and Tamer, 2018). 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). Such 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 split. A match assignment is a description of who is matched with whom, whereas a match surplus split 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 that the markets clear.\(^4\) 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 break her current match and pair up with other individuals.

In order to formally describe the equilibrium notion, we introduce some additional notation. Firstly, we define a match assignment. Let $\hat{\mu}_{ij}$ be equal to 1 if man $i \in I$ and woman $j \in J$ are matched and zero otherwise. Let $\hat{\mu}_{i0}$ be equal to 1 if man $i \in I$ is single and 0 otherwise. Let $\hat{\mu}_{0j}$ be equal to 1 if woman $j \in J$ is single and 0 otherwise.

\(^4\)A competitive equilibrium can equivalently be characterised by a match assignment and transfer scheme.
The vector

$$\tilde{\mu} \equiv (\tilde{\mu}_{ij} \ \forall (i, j) \in I_0 \times J_0 \setminus \{(0, 0)\}),$$

represents a match assignment. Secondly, we define a match surplus split. 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

$$\tilde{U} \equiv (\tilde{U}_i \ \forall i \in I), \ \tilde{V} \equiv (\tilde{V}_j \ \forall j \in J),$$

represent a match surplus split and they implicitly embed transfers across the two sides of the market.

$(\tilde{\mu}, \tilde{U}, \tilde{V})$, is a stable matching if it satisfies three properties. First, the match assignment is feasible in the sense of one-to-one, i.e.,

$$\int_{J_0} \tilde{\mu}_{ij} dP_W = 1 \ \forall i \in I, \ \int_{I_0} \tilde{\mu}_{ij} dP_M = 1 \ \forall j \in J,$$

where $P_M$ and $P_W$ denote the total measure of men and the total measure of women. Second, there is no man and woman that can get a strictly higher match surplus by breaking their matches under $\tilde{\mu}$ and pairing up together, i.e.,

$$\tilde{U}_i + \tilde{V}_j \geq \tilde{\Phi}_{ij} \ \forall (i, j) \in I \times J.$$

Third, there is no agent that can get a strictly higher payoff by breaking her match under $\tilde{\mu}$ and remaining single, i.e.,

$$\tilde{U}_i \geq \tilde{\Phi}_{i0} \ \forall i \in I, \ \tilde{V}_j \geq \tilde{\Phi}_{0j} \ \forall j \in J.$$

Moreover, a stable matching exists and is generically unique.\(^5\) Hence, from now on we will refer to it as the stable matching.

### 3 Identification

The identification analysis opens with recalling which data are considered available to the researcher and which parameters the researcher aims to back out using the information at her disposal.

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 is learning 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

\(^5\)In particular, uniqueness relies on the fact that the taste shocks are continuously distributed conditional on types. For more details see Shapley and Shubik (1972) and Gretsky, Ostroy and Zame (1992).
of man matched with woman $j \in J$ at the equilibrium. Let $P_{Q_M|X}$ and $P_{Q_W|Y}$ denote the probability distribution of $Q_{M,i}$ conditional on $X_i$ and 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 analyst to have aggregate data on matches in one market. For example, it is sufficient to have a random sampling scheme in which the analyst draws agents from each side of the market and records the matched types.

The parameter we are after is the vector of type-specific match surpluses

$$\Phi \equiv (\Phi_{xy} \forall (x, y) \in \mathcal{X}_0 \times \mathcal{Y}_0 \setminus \{(0, 0)\}).$$

Indeed, knowing $\Phi$ is essential to understand the relative importance of different characteristics of agents, to estimate the gains to matching, or to run counterfactuals and evaluate policy interventions aiming to modify existing matching outcomes in favour of more desirable allocations.

Throughout the analysis, we will assume that the researcher does not observe equilibrium transfers. However, our discussion remains valid when data on equilibrium transfers are available. Furthermore, in that case one can also identify pre-transfers systematic payoff components for each side of the market.

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 introduced earlier.

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

$$U \equiv (U_{xy} \forall (x, y) \in \mathcal{X} \times \mathcal{Y}), \quad V \equiv (V_{xy} \forall (x, y) \in \mathcal{X}_0 \times \mathcal{Y}),$$

such that

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

Proposition 1 allows to rewrite the framework of Section 2 as two separate one-sided multinomial choice models with utility maximising agents. These two models are linked by market-clearing transfers implicitly embedded into $U$ 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:

\[ Q_{M,i} = \arg\max_{y \in Y_0} (U_{xy} + \epsilon_{iy}) \quad \forall i \in \mathcal{I} \text{ of type } x \in \mathcal{X}, \forall x \in \mathcal{X}, \quad (4) \]

\[ Q_{W,j} = \arg\max_{x \in X_0} (V_{xy} + \eta_{jx}) \quad \forall j \in \mathcal{J} \text{ of type } y \in \mathcal{Y}, \forall y \in \mathcal{Y}, \quad (5) \]

\[ U_{xy} + V_{xy} = \Phi_{xy}, \ U_{x0} = \Phi_{x0}, \ V_{0y} = \Phi_{0y} \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \quad (6) \]

\[ \epsilon_{i|X_i} \sim P_{\epsilon|X} \ | \ \eta_{j|Y_j} \sim P_{\eta|Y} \quad \forall i \in \mathcal{I}, \forall j \in \mathcal{J}, \quad (7) \]

where “\( \sim \)” denotes “distributed as” and the equality sign in front of the argmax operator is because ties are zero probability events. Such alternative representation of the problem is useful because 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 the taste shocks are independent of types and independently and identically distributed as Gumbel with scale 1 and location 0, then \( \Phi \) is point identified via standard Logit model arguments applied to each side of the market. Such conclusion 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 distributions of the taste shocks are completely known by the researcher, which, in turn, 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 and misleading counterfactual outcomes, as widely documented by the econometric literature on binary and multinomial choice models (e.g., Manski, 1975, 1985, 1988; Matzkin, 1992, 1993; Fox, 2007). Our objective is therefore to investigate identification of \( \Phi \) when the researcher does not impose parametric restrictions on \( P_{\epsilon|X} \) and \( P_{\eta|Y} \), but, possibly, includes nonparametric distributional assumptions like symmetric marginals, stochastic independence of types, and many more. We believe that performing such an exercise (hereafter referred to as the nonparametric exercise) is important because it allows to evaluate to what extent the empirical content of the model illustrated in Section 2 depends on the way in which the analyst has parameterised the conditional probability distributions of the latent variables.

Avoiding parametric restrictions on \( P_{\epsilon|X} \) and \( P_{\eta|Y} \) raises the possibility of partial identification and the consequent difficulty of characterising in a tractable way the region of parameter values that exhausts all the implications of the model and data (the sharp identified set), under various classes of nonparametric distributional assumptions on the taste shocks. Our study aims to address this methodological challenge.

There are other seminal papers in the literature considering close research questions when the analyst has data on one market. A first group of works focuses on one-sided
markets. More precisely, Manski (1975) and Matzkin (1993) show point identification of the payoff parameters in a multinomial choice model under two main ingredients: conditional i.i.d.-ness of the taste shocks and the presence of a continuous regressor having large support. Additionally, they put assumptions on the systematic part of payoffs, for example, an index structure or nonparametric utilities with shape restrictions. Fox (2007) proves that conditional exchangeability of the taste shocks can replace conditional i.i.d.-ness, because it is sufficient for the rank property to hold. More recently, researchers have started to extend those results to two-sided markets. In particular, Fox (2018) derives point identification in a matching framework under conditional exchangeability of the taste shocks, an index structure of the systematic part of payoffs, and the availability of a continuous regressor with large support. Furthermore, in the absence of such a special regressor, he characterises an identified region (not shown to be sharp nor studied in relation to the sharp identified set) that can be estimated. In the humblest possible way, we view our work as complementary to these papers by providing a procedure that allows to tractably and fully explore the empirical content of the model discussed above under various classes of nonparametric distributional assumptions on the latent variables.

We now explain in detail why performing the nonparametric exercise is methodologically challenging.

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

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

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

and

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

be the vectors of differences between every pair 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 $(d - r)$ elements are linear combination of the first $r$ components.
Hence, $\Delta \epsilon_i$ and $\Delta \eta_j$ take values in the region

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

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

**Definition 1.** *(Sharp identified set)* Let $\Theta^\dagger$, $U^\dagger$, and $V^\dagger$ be the set of admissible values of $\Phi$, $U$, and $V$. Let $P_{\Delta \epsilon}$ and $P_{\Delta \eta}$ be the function spaces of admissible $d$-dimensional conditional probability distributions of the taste shock differences, which can include parametric and/or nonparametric restrictions. The sharp identified set for $\Phi$ is

$$\Theta^* \equiv \{ \Phi \in \Theta^\dagger : \exists U \in U^\dagger, V \in V^\dagger, P_{\Delta \epsilon|x} \in P_{\Delta \epsilon}, P_{\Delta \eta|y} \in P_{\Delta \eta} \text{ s.t.}$$

$$P_{Q_{\Delta \epsilon}|x}(y) = \omega_{M,y|x}(U, P_{\Delta \epsilon|x}) \quad \forall (x, y) \in X \times Y_0$$

$$P_{Q_{\Delta \eta}|y}(x) = \omega_{W,x|y}(V, P_{\Delta \eta|y}) \quad \forall (x, y) \in X_0 \times Y$$

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

$$P_{\Delta \epsilon|x}(B) = 1, \quad P_{\Delta \eta|y}(B) = 1 \quad \forall (x, y) \in X \times Y \},$$

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

$$U^* \equiv \{ U \in U^\dagger : \exists P_{\Delta \epsilon|x} \in P_{\Delta \epsilon} \text{ s.t.}$$

$$P_{Q_{\Delta \epsilon}|x}(y) = \omega_{M,y|x}(U, P_{\Delta \epsilon|x}) \text{ and } P_{\Delta \epsilon|x}(B) = 1 \forall (x, y) \in X \times Y_0 \},$$

$$V^* \equiv \{ V \in V^\dagger : \exists P_{\Delta \eta|y} \in P_{\Delta \eta} \text{ s.t.}$$

$$P_{Q_{\Delta \eta}|y}(x) = \omega_{W,x|y}(V, P_{\Delta \eta|y}) \text{ and } P_{\Delta \eta|y}(B) = 1 \forall (x, y) \in X_0 \times Y \},$$

such that

$$\Theta^* = \{ \Phi \in \Theta^\dagger : \exists U \in U^*, V \in V^* \text{ s.t.}$$

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

One may think of many other equivalent ways to define $\Theta^*$. We have provided a
representation that is pedagogical for the procedure described below. In particular, for a given \( \{U, V, P_{\Delta|x}, P_{\Delta|y}\} \), the first and second equations of (8) impose that the probability distributions \( \{P_{Q_{x}|x}, P_{Q_{y}|y}\} \) coincide with the probability distributions of the observables as predicted by the model; the third equation mimics (6); the fourth equation requires the conditional probability distributions of the taste shock differences to be concentrated on the region \( B \) (hereafter referred to as the degeneracy condition). Notice that the degeneracy condition is trivially satisfied when \( r = 1 \).

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

**Example 1.** When \( r = 2 \) \((d = 3)\), the region \( B \) is given by

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

and the vectors of taste shock differences, \( \Delta \epsilon_i \) and \( \Delta \eta_j \), are

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

\[
\Delta \eta_j \equiv (\eta_{j1} - \eta_{j0}, \eta_{j2} - \eta_{j0}, \eta_{j1} - \eta_{j2}).
\]

Moreover, from (4) and (5), we have that \( \forall (x, y) \in \mathcal{X} \times \mathcal{Y} \)

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

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

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

and

\[
P_{Q_{y}|y}(1) = P_{\Delta|y}([V_{y0} - V_{y1}, \infty] \times [-\infty, \infty] \times [V_{y2} - V_{y1}, \infty]),
\]

\[
P_{Q_{y}|y}(2) = P_{\Delta|y}([-\infty, \infty] \times [V_{y0} - V_{y2}, \infty] \times [-\infty, V_{y2} - V_{y1}]),
\]

\[
P_{Q_{y}|y}(0) = P_{\Delta|y}([-\infty, V_{y0} - V_{y1}] \times [-\infty, V_{y0} - V_{y2}] \times [-\infty, \infty]).
\]

Therefore,

\[
\mathcal{U}^* \equiv \left\{ U \in \mathcal{U}^1 : \exists P_{\Delta|\mathcal{X}} \in \mathcal{P}_{\Delta^*} \text{ s.t. } \forall x \in \mathcal{X} \text{ (10) is satisfied and } P_{\Delta|x}(B) = 1 \right\},
\]

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

and \( \Theta^* \) is as in (9).

Performing the nonparametric exercise means constructing \( \Theta^* \) without incorporating parametric restrictions into \( \mathcal{P}_{\Delta^*} \) and \( \mathcal{P}_{\Delta^*}^1 \). 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 $U \in \mathcal{U}^\dagger$, there exists $P_{\Delta|x} \in \mathcal{P}^\dagger_{\Delta}$ such that $P_{Q_{\Delta|x}|x}(y) = \omega_{M,y|x}(U, P_{\Delta|x})$ and $P_{\Delta|x}(B) = 1 \ \forall (x, y) \in X \times Y_0$. Without parametric restrictions on the conditional probability distributions of the taste shocks, this corresponds to solving an infinite-dimensional existence problem because each $P_{\Delta|x} \in \mathcal{P}^\dagger_{\Delta}$ is an infinite-dimensional object. Second, such infinite-dimensional existence problem has to be solved for every $U \in \mathcal{U}^\dagger$. Typically, this is done in the partial identification literature by constructing a grid roughly resembling $\mathcal{U}^\dagger$ and 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 types, hence leading quickly to a computational bottleneck. Similar issues are faced on the women’s side.

3.2 Constructing the sharp identified set

In what follows we design a procedure that ameliorates the challenges described above, and thus enables us to tractably approximate $\Theta^*$. We organise the discussion of the methodology in three steps, which are identical for each side of the market. Without loss of generality, let us consider the men’s side. First, Section 3.2.1 establishes that determining, for a given $U \in \mathcal{U}^\dagger$, whether there exists $P_{\Delta|x} \in \mathcal{P}^\dagger_{\Delta}$ such that $P_{Q_{\Delta|x}|x}(y) = \omega_{M,y|x}(U, P_{\Delta|x})$ and $P_{\Delta|x}(B) = 1 \ \forall (x, y) \in X \times Y_0$ is equivalent to finding whether a system of linear equalities and inequalities has a solution, which is a manageable and well-understood problem. This result uses Theorem 1 by Torgovitsky (2018) (also known as PIES) under various classes of nonparametric restrictions on the conditional probability distributions of the taste shock differences.

Second, Section 3.2.2 shows that $\mathcal{U}^\dagger$ can be ex-ante partitioned into a finite number of subsets such that, for each subset, every value of $U$ belonging to that subset gives rise to the same linear programming problem defined in the first step. Therefore, the researcher has to solve such a linear programming problem once for each subset.

Third, given $P_{\Delta|x} \in \mathcal{P}^\dagger_{\Delta}$, Section 3.2.3 discusses a way to approximate the degeneracy condition $P_{\Delta|x}(B) = 1 \ \forall x \in X$ as a system of linear equalities which can be easily added to the linear programming problem of the first step. This approximation of the degeneracy condition is the reason why our approach allows to approximate $\Theta^*$ rather than exactly construct it.\footnote{As mentioned before, the degeneracy condition is trivially satisfied when $r = 1$. Therefore, in that case, our approach allows to exactly construct $\Theta^*$.}

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

**Proposition 2.** (Torgovitsky, 2018) Under various classes of nonparametric restrictions possibly incorporated into \( P_{\Delta \epsilon \mid X} \), determining whether, for a given \( U \in \mathcal{U} \), there exists \( P_{\Delta \epsilon |X} \in \mathcal{P}_{\Delta \epsilon} \) such that \( P_{Q_M|y|x}(y) = \omega_{M,g|x}(U, P_{\Delta \epsilon |x}) \forall (x, y) \in \mathcal{X} \times \mathcal{Y}_0 \) 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 \( P_{\Delta \epsilon |X} \in \mathcal{P}_{\Delta \epsilon} \), let \( G_{\Delta \epsilon |X} \in \mathcal{G}_{\Delta \epsilon} \) denote the associated conditional CDF, where \( \mathcal{G}_{\Delta \epsilon} \) is the function space of all admissible \( d \)-dimensional conditional CDFs. To keep things simple suppose for the moment that the conditional probability distributions of the taste shock differences are left completely unrestricted, i.e., \( \mathcal{P}_{\Delta \epsilon} \equiv \mathcal{P} \) (and, hence, \( \mathcal{G}_{\Delta \epsilon} \equiv \mathcal{G} \)), 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).

Later we will discuss how to incorporate nonparametric assumptions into \( \mathcal{P}_{\Delta \epsilon} \). From (12), we have the following infinite-dimensional existence problem for a given \( U \in \mathcal{U} \):

Find if there exists \( P_{\Delta \epsilon |X} \in \mathcal{P} \) s.t. \( \forall x \in \mathcal{X} \)

\[
\begin{align*}
P_{Q_M|x}(1) &= P_{\Delta \epsilon |x}((U_{x0} - U_{x1}, \infty) \times (-\infty, \infty) \times [U_{x2} - U_{x1}, \infty)), \\
P_{Q_M|x}(2) &= P_{\Delta \epsilon |x}((-\infty, \infty) \times [U_{x0} - U_{x2}, \infty) \times (-\infty, U_{x2} - U_{x1})], \\
P_{Q_M|x}(0) &= P_{\Delta \epsilon |x}((-\infty, U_{x0} - U_{x1}) \times (-\infty, U_{x0} - U_{x2}]) \times (-\infty, \infty)).
\end{align*}
\]
Using $G_{\Delta|x} \in \mathcal{G}$, (14) can be rewritten as:

Find if there exists $G_{\Delta|x} \in \mathcal{G}$ s.t. $\forall x \in \mathcal{X}$

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

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

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

(15)

The system of equations in (15) depends on the values of $G_{\Delta|x}$ at a finite number of 3-tuples, $\forall x \in \mathcal{X}$. We thus define three finite sets (sometimes referred to below as the $\mathcal{A}$-sets)

$$\mathcal{A}_{x,1}(U) \equiv \{U_{x_0} - U_{x_1}, \infty, -\infty\},$$

$$\mathcal{A}_{x,2}(U) \equiv \{U_{x_0} - U_{x_2}, \infty, -\infty\},$$

$$\mathcal{A}_{x,3}(U) \equiv \{U_{x_1} - U_{x_2}, \infty, -\infty\},$$

$\forall x \in \mathcal{X}$, where $\mathcal{A}_{x,1}(U)$ collects the elements at which $G_{\Delta|x}$ is evaluated along the first dimension, $\mathcal{A}_{x,2}(U)$ collects the elements at which $G_{\Delta|x}$ is evaluated along the second dimension, and $\mathcal{A}_{x,3}(U)$ collects the elements at which $G_{\Delta|x}$ is evaluated in 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 set $\mathcal{A}_{x}(U) \equiv \mathcal{A}_{x,1}(U) \times \mathcal{A}_{x,2}(U) \times \mathcal{A}_{x,3}(U)$, where “$\times$” denotes the Cartesian product operator. Therefore, (15) can be rewritten as:

$\forall x \in \mathcal{X}$, find if there exists $\tilde{G}^U_{\Delta|x} : \mathcal{A}_{x}(U) \rightarrow \mathbb{R}$ s.t.

$$P_{Q_{x}}(1) = 1 + \tilde{G}^U_{\Delta|x}(U_{x_0} - U_{x_1}, \infty, U_{x_2} - U_{x_1}) - \tilde{G}^U_{\Delta|x}(\infty, U_{x_2} - U_{x_1}) - \tilde{G}^U_{\Delta|x}(U_{x_0} - U_{x_1}, \infty, \infty),$$

$$P_{Q_{x}}(2) = \tilde{G}^U_{\Delta|x}(\infty, \infty, U_{x_2} - U_{x_1}) - \tilde{G}^U_{\Delta|x}(\infty, U_{x_0} - U_{x_2}, U_{x_2} - U_{x_1}),$$

$$P_{Q_{x}}(0) = \tilde{G}^U_{\Delta|x}(U_{x_0} - U_{x_1}, U_{x_0} - U_{x_2}, \infty),$$

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

(16)

(16) highlights that the existence problem (14) is equivalent to first finding whether a system of linear equalities has a solution and second ensuring that such a solution additionally solves an extension exercise. With regard to the latter, 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

---

8Recall that the conditional CDF of a cube corresponds to its volume under the conditional probability distribution. Hence, $\forall x \in \mathcal{X}$ and under the conditional CDF $\mathcal{G}_{\Delta|x}$, we have that

$$P_{\Delta|x}([a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]) =$$

$$- \mathcal{G}_{\Delta|x} (a_1, a_2, a_3) + \mathcal{G}_{\Delta|x} (b_1, a_2, a_3) + \mathcal{G}_{\Delta|x} (a_1, b_2, a_3) - \mathcal{G}_{\Delta|x} (b_1, b_2, a_3)$$

$$+ \mathcal{G}_{\Delta|x} (a_1, a_2, b_3) - \mathcal{G}_{\Delta|x} (b_1, a_2, b_3) - \mathcal{G}_{\Delta|x} (a_1, b_2, b_3) + \mathcal{G}_{\Delta|x} (b_1, b_2, b_3).$$
if it satisfies the following system of linear equalities and inequalities \( \forall x \in \mathcal{X} \):

\[
\begin{align*}
\bar{G}^U_{\Delta|x}(\infty, t, q) &= 0 & \forall (t, q) \in \mathcal{A}_{x,2}(U) \times \mathcal{A}_{x,3}(U), \\
\bar{G}^U_{\Delta|x}(t, \infty, q) &= 0 & \forall (t, q) \in \mathcal{A}_{x,1}(U) \times \mathcal{A}_{x,3}(U), \\
\bar{G}^U_{\Delta|x}(t, q, \infty) &= 0 & \forall (t, q) \in \mathcal{A}_{x,1}(U) \times \mathcal{A}_{x,2}(U), \\
\bar{G}^U_{\Delta|x}(\infty, \infty, \infty) &= 1
\end{align*}
\]

\[
0 \leq \bar{G}^U_{\Delta|x}(t, q, r) \leq 1 & \forall (t, q, r) \in \mathcal{A}_x(U), \\
-\bar{G}^U_{\Delta|x}(t, q, r) + \bar{G}^U_{\Delta|x}(t, q', r) + \bar{G}^U_{\Delta|x}(t, q, r') - \bar{G}^U_{\Delta|x}(t, q', r') &+ \bar{G}^U_{\Delta|x}(t, q, r') - \bar{G}^U_{\Delta|x}(t, q', r') - \bar{G}^U_{\Delta|x}(t, q', r') \geq 0 & \forall (t, q, r), (t', q', r') \in \mathcal{A}_x(U) \text{ s.t. } (t, q, r) \leq (t', q', r').
\]

Specifically, bearing in mind the properties defining CDFs, the first four lines in (17) ensure that \( \bar{G}^U_{\Delta|x} \) 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 guarantees that the range of \( \bar{G}^U_{\Delta|x} \) is a subset of \([0, 1]\); the last line requires \( \bar{G}^U_{\Delta|x} \) to be a 3-increasing function, i.e., for each pair of 3-tuples in the domain of \( \bar{G}^U_{\Delta|x} \) which are comparable component-wise, the volume of the 3-dimensional box with vertices \([t, t'] \times [q, q'] \times [r, r'] \) is positive. By merging (16) and (17), an easy-to-solve linear programming problem is obtained.

The procedure described allows to include into \( P^I_{\Delta} \) 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. Theorem 1 in Torgovitsky (2018) shows that, indeed, such constraints are simply added to (17) as linear equalities and inequalities. For example, one can impose that the conditional probability distributions of the taste shock differences are characterised by identical marginals, i.e., \( P^{(k)}_{\Delta|x} = P^{(k')}_{\Delta|x} \) \( \forall \{k, k'\} \subseteq \{1, \ldots, d\} \) where \( P^{(k)}_{\Delta|x} \) denotes the kth marginal of \( P_{\Delta|x} \), by marginals symmetric about zero, i.e., \( P^{(k)}_{\Delta|x}((\infty, a]) = 1 - P^{(k)}_{\Delta|x}((\infty, -a]) \) \( \forall a \in \mathbb{R} \) and \( \forall k \in \{1, \ldots, d\} \), and by joint or marginal independence of types, i.e., \( P_{\Delta|x} = P_{\Delta|x'} \) or \( P^{(k)}_{\Delta|x} = P^{(k')}_{\Delta|x'} \) \( \forall \{x, x'\} \subseteq \mathcal{X} \) and \( \forall k \in \{1, \ldots, d\} \). Additionally, any quantile of \( P_{\Delta|x} \) and \( P^{(k)}_{\Delta|x} \) \( \forall k \in \{1, \ldots, d\} \), e.g., the median, can be set equal to known values. Other types of restrictions are possible and 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 on how the linear programming problem should be modified when nonparametric distributional assumptions on the taste shock differences are added, consider imposing that the conditional marginal probability distributions of the taste shock

\[\text{See Definition 1 in Torgovitsky (2018) or the proof of Proposition 3 for a general definition of } d\text{-increasingness. Notice also that } d\text{-increasingness reduces to the usual definition of weakly increasingness when } r = 1.\]
differences are symmetric. Then, the $A$-sets are

$$A_{x,1}(U) \equiv \{U_{x0} - U_{x1}, -U_{x0} + U_{x1}, \infty, -\infty\},$$

$$A_{x,2}(U) \equiv \{U_{x0} - U_{x2}, -U_{x0} + U_{x2}, \infty, -\infty\},$$

$$A_{x,3}(U) \equiv \{U_{x1} - U_{x2}, -U_{x1} + U_{x2}, \infty, -\infty\}.$$ 

$\forall x \in \mathcal{X}$. The linear programming problem to solve becomes

$$\forall x \in \mathcal{X}, \text{ find if there exists } \bar{G}^U_{\Delta|x}: A_x(U) \to \mathbb{R} \text{ s.t.}$$

$$P_{Q_{x|x}}(1) = 1 + \bar{G}^U_{\Delta|x}(U_{x0} - U_{x1}, \infty, U_{x2} - U_{x1}) - \bar{G}^U_{\Delta|x}(\infty, \infty, U_{x2} - U_{x1}) - \bar{G}^U_{\Delta|x}(U_{x0} - U_{x1}, \infty, \infty),$$

$$P_{Q_{x|x}}(2) = \bar{G}^U_{\Delta|x}(\infty, \infty, U_{x2} - U_{x1}) - \bar{G}^U_{\Delta|x}(\infty, U_{x0} - U_{x2}, U_{x2} - U_{x1}),$$

$$P_{Q_{x|x}}(0) = \bar{G}^U_{\Delta|x}(U_{x0} - U_{x1}, U_{x0} - U_{x2}, \infty),$$

(17) is satisfied,

$$\bar{G}^U_{\Delta|x}(U_{x0} - U_{x1}, \infty, \infty) = 1 - \bar{G}^U_{\Delta|x}(-U_{x0} + U_{x1}, \infty, \infty),$$

$$\bar{G}^U_{\Delta|x}(\infty, U_{x0} - U_{x2}, \infty) = 1 - \bar{G}^U_{\Delta|x}(\infty, -U_{x0} + U_{x2}, -\infty),$$

$$\bar{G}^U_{\Delta|x}(\infty, \infty, U_{x1} - U_{x2}) = 1 - \bar{G}^U_{\Delta|x}(\infty, \infty, -U_{x1} + U_{x2}).$$

(18)

Before concluding, we remark that the methodology illustrated does not allow to incorporate nonparametric assumptions on the conditional probability distributions of the original, i.e., not differenced, taste shocks. This is the obvious price to pay for having an approach that, on the other hand, permits us to be very flexible regarding the classes of nonparametric distributional assumptions considered and to compare, in turn, the empirical content of the model in Section 2 under different scenarios. Given that in multinomial choice models what matters are differences in utilities, providing a tractable characterisation of the sharp identified set which remains valid under various classes of nonparametric assumptions on the conditional probability distributions of the original latent variables remains an open question to the best of our knowledge.

### 3.2.2 Partitioning the parameter space, $U^I$

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

---

17
Proposition 3. (Partitioning $U^\uparrow$) For every $x \in \mathcal{X}$, under the classes of nonparametric restrictions on the conditional probability distributions of the taste shock differences considered in Proposition 2, $U^\uparrow$ can be ex-ante partitioned into a finite number, $K_x$, of subsets \( \{U_{k,x}^\uparrow \}_{k=1}^{K_x} \) such that \( \forall k \in \{1, ..., K_x \} \) and \( \forall \{U, \bar{U}\} \subseteq U_{k,x}^\uparrow \),

\[
\exists P_{\Delta|x} \in \mathcal{P}_{\Delta} \text{ such that } P_{Q_{\Delta|x}}(y) = \omega_{M,y|x}(U, P_{\Delta|x}) \quad \forall y \in \mathcal{Y}_0 \\
\text{if and only if} \\
\exists P_{\Delta|x} \in \mathcal{P}_{\Delta} \text{ such that } P_{Q_{\Delta|x}}(y) = \omega_{M,y|x}((U, \bar{U}), P_{\Delta|x}) \quad \forall y \in \mathcal{Y}_0.
\]

Moreover, \( \forall k \in \{1, ..., K_x \} \) and \( \forall x \in \mathcal{X} \), finding any \( U \in U_{k,x}^\uparrow \) amounts to solving a linear programming problem.

As a consequence of Proposition 3, picking any values of $U$ from $U_{k,x}^\uparrow$ \( \forall k \in \{1, ..., K_x \} \) and \( \forall x \in \mathcal{X} \), and then solving the linear programming problem of Section 3.2.1 at the selected \( \Sigma_{x \in \mathcal{X}} K_x \) values of $U$ is sufficient to span the entire $U^\uparrow$.

We now sketch the intuition behind Proposition 3 continuing from the linear programming problem (18) which arises in the case $r = 2$ ($d = 3$), when the analyst imposes that the conditional marginal probability distributions of the taste shock differences are symmetric about zero. First, note that for a given $x \in \mathcal{X}$, the only chunk of (18) that can potentially generate different solutions for different values of $U$ is the one requiring $G_{\Delta|x}^U$ to be a 3-increasing function, i.e.,

\[
-G_{\Delta|x}^U(t, q, r) + G_{\Delta|x}^U(t', q, r) + G_{\Delta|x}^U(t, q', r) - G_{\Delta|x}^U(t', q', r) \\
+ G_{\Delta|x}^U(t, q, r') - G_{\Delta|x}^U(t', q, r') + G_{\Delta|x}^U(t, q', r') + G_{\Delta|x}^U(t', q', r') \geq 0 \quad (19)
\]

\( \forall (t, q, r), (t', q, r') \in \mathcal{A}_x(U) \text{ s.t. } (t, q, r) \leq (t', q, r'). \)

This is because different values of $U$ can give rise to different collections of pairs of 3-tuples from $\mathcal{A}_x(U)$ that are comparable component-wise and, in turn, to different constraints on the image set of $G_{\Delta|x}^U$ of the type (19). Such observation immediately suggests that if two values of $U$ induce component-wise comparisons featuring the same “coordinates” in $\mathcal{A}_x(U)$ (in a sense that will be specified below), then they should generate the same constraints on the image set of $G_{\Delta|x}^U$ of the type (19) and, hence, have the same solution to (18) at $x \in \mathcal{X}$.

More formally, for a given $x \in \mathcal{X}$, fix any order of the $\mathcal{A}$-sets so that they become 4-tuples and apply such ordering choice to the $\mathcal{A}$-sets for each $U \in U^\uparrow$. Denote the resulting 4-tuples by $a_{x,1}(U), a_{x,2}(U), a_{x,3}(U)$. For example,

\[
a_{x,1}(U) \equiv (U_{x0} - U_{x1}, -U_{x0} + U_{x1}, \infty, -\infty), \\
a_{x,2}(U) \equiv (U_{x0} - U_{x2}, -U_{x0} + U_{x2}, \infty, -\infty), \\
a_{x,3}(U) \equiv (U_{x1} - U_{x2}, -U_{x1} + U_{x2}, \infty, -\infty).
\]
If applied to a finite tuple of real numbers, let \( \pi \) be a function delivering the position of the elements in that tuple when re-arranged from the smallest to the largest and the relational operators, \(<\) or \(=\), among the re-arranged elements. When the tuple considered contains multiple elements with the same value, then any convention on which element should be listed first can be adopted. E.g., \( \pi(100, 99, \infty, -\infty) = \{(4, 2, 1, 3), (<, <, <)\} \) and \( \pi(5, 5, \infty, -\infty) = \{(4, 1, 2, 3), (<, =, <)\} \). We call the collection \( \{\pi(a_{\cdot,1}(U)), \pi(a_{\cdot,2}(U)), \pi(a_{\cdot,3}(U))\} \) as the \( x \)-coordinates induced by \( U \in \mathcal{U}^\dagger \).

Take now \( U, \tilde{U} \in \mathcal{U}^\dagger \). Three facts are immediate to see. First, in the light of our initial observation, if \( U \) and \( \tilde{U} \) induce the same \( x \)-coordinates, then they are characterised by the same solution to (18) at \( x \in \mathcal{X} \). Second, the number of feasible \( x \)-coordinates is 26.\(^{10}\) The first and the second facts combined imply that \( \mathcal{U}^\dagger \) can be ex-ante partitioned into \( K_x = 26 \) subsets (one for each feasible \( x \)-coordinates) such that, for every subset, each value of \( U \) belonging to that subset features the same solution to (18) at \( x \in \mathcal{X} \). Therefore, one has to pick any value of \( U \) from each of the 26 subsets and solve (18) at \( x \in \mathcal{X} \) only for the selected 26 values of \( U \). Repeating this at every \( x \in \mathcal{X} \) is sufficient to span the entire \( \mathcal{U}^\dagger \). Third, picking a value of \( U \) from each of the 26 subsets amounts to work out simple linear equalities and inequalities because the \( \pi \) function establishes linear relations among the \( A \)-sets’ elements.

The same approach is replicable under any class of nonparametric distributional assumptions on the taste shock differences contemplated by Proposition 2. Providing a generic expression to compute \( K_x \) that is valid under any such a class is not viable, because of the impossibility of expressing the cardinalities of the \( A \)-sets as functions of the type of nonparametric distributional assumptions imposed. Nevertheless, we expect \( K_x \) to increase in the cardinality of \( \mathcal{Y} \) (the higher is the cardinality of \( \mathcal{Y} \), the higher is the number of \( A \)-sets) and in the number of nonparametric distributional assumptions (the more are the nonparametric distributional assumptions, the higher are the cardinalities of the \( A \)-sets).

### 3.2.3 The degeneracy condition

The conditional probability distribution of the taste shocks, \( \epsilon \), is given by \( P_{\epsilon|X} \). When we consider the vector of taste shock differences, \( \Delta \epsilon_t \), its conditional probability distribution
\( P_{\Delta|X} \) depends on \( P_{\epsilon|X} \). In particular, \( P_{\Delta|X} \) is concentrated on the region \( B \), i.e., the degeneracy condition, \( P_{\Delta|X}(B) = 1 \), holds. Hence, as part of the nonparametric exercise, the researcher has to ensure that \( P_{\Delta|X} \in \mathcal{P}_{\Delta}^\dagger \) possibly solving the linear programming problem of Section 3.2.1 for a given \( U \in \mathcal{U}^\dagger \) satisfies the degeneracy condition. Without incorporating the degeneracy condition, one does not exploit all the information coming from the primitive assumptions of the model and hence characterises just an outer set of the sharp identified set for \( \Phi \).

In what follows, we propose a way to approximate the degeneracy condition as a finite collection of equalities that are linear in \( G_{\Delta|X} \), and therefore, linear in \( \tilde{G}_{\Delta|X}^U \), as introduced in Equation (16). These equalities require the (approximate) complement of \( B \) to have probability mass zero. More formally, Proposition 4 claims that imposing “enough” of the \( d \)-dimensional boxes in \( \mathbb{R}^d \) not intersecting the region \( B \) to have zero probability is sufficient for the degeneracy condition to hold.

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

\[
B_{t,p,q}(\bar{b}, \bar{b}) \equiv \{(z_1, \ldots, z_d) \in \mathbb{R}^d: z_p \leq \bar{b}, \ z_q \leq \bar{b}, \ z_t > \bar{b} + \bar{b}\},
\]

and

\[
Q_{t,p,q}(\bar{b}, \bar{b}) \equiv \{(z_1, \ldots, z_d) \in \mathbb{R}^d: z_p > \bar{b}, \ z_q > \bar{b}, \ z_t \leq \bar{b} + \bar{b}\},
\]

\( \forall t \in \{1,\ldots, r - 1\} \) and \( \forall (p, q) \in \{(t + 1, r), (t + 2, r + 1), \ldots, (r, d)\} \). Let \( \mathbb{Q} \) denote the set of rational numbers. For each \( P_{\Delta|X} \in \mathcal{P}_{\Delta}^1 \), if

\[
P_{\Delta|X}(B_{t,p,q}(\bar{b}, \bar{b})) = P_{\Delta|X}(Q_{t,p,q}(\bar{b}, \bar{b})) = 0 \quad \forall t \in \{1,\ldots, r - 1\}, \forall (p, q) \in \{(t + 1, r), (t + 2, r + 1), \ldots, (r, d)\}, \quad (20)
\]

\( \forall (\bar{b}, \bar{b}) \in \mathbb{Q}^2 \), then \( P_{\Delta|X}(B) = 1 \).

Proposition 4 thus suggests that by imposing (20) at a finite grid of 2-tuples \((\bar{b}, \bar{b}) \in \mathbb{Q}^2\), one approximates the degeneracy condition. Furthermore, (20) is linear in \( G_{\Delta|X} \) and, hence, can be added to the linear programming problem of Section 3.2.1 \( \forall U \in \mathcal{U}^\dagger \) after having replaced \( G_{\Delta|X} \) with \( \tilde{G}_{\Delta|X}^U \). This step concludes our construction of \( \Theta^\dagger \).

We now sketch the intuition behind Proposition 4 continuing the case \( r = 2 \ (d = 3) \) of Example 1. The underlying idea is that, for a given \( P_{\Delta|X} \in \mathcal{P}_{\Delta}^1 \), if \( P_{\Delta|X}(B) = 1 \), then any 3-dimensional box in \( \mathbb{R}^3 \) not intersecting the region \( B \) has probability zero. Vice versa, if one imposes such a zero probability condition for “enough” of the 3-dimensional boxes in \( \mathbb{R}^3 \) not intersecting the region \( B \), then \( P_{\Delta|X}(B) = 1 \) should be satisfied. More formally, firstly notice that

\[
B \equiv \{(b_1, b_2, b_3) \in \mathbb{R}^3: b_3 = b_1 - b_2\} = \{(b_1, b_2, b_3) \in \mathbb{R}^3: b_1 = b_2 + b_3\}.
\]
Accordingly, the relevant 3-dimensional boxes as defined in Proposition 4 are

\[ B_{1,2,3}(\tilde{b}, \tilde{b}) \equiv \{(z_1, z_2, z_3) \in \mathbb{R}^3 : z_1 > \tilde{b} \}, \]

\[ Q_{1,2,3}(\tilde{b}, \tilde{b}) \equiv \{(z_1, z_2, z_3) \in \mathbb{R}^3 : z_1 < \tilde{b} \}, \]

for any \((\tilde{b}, \tilde{b}) \in \mathbb{R}^2\). We can show that the complement of the region \(B\) in \(\mathbb{R}^3\) is equivalent to the union of the countably infinite sets \(\bigcup_{(\tilde{b}, \tilde{b}) \in \mathbb{Q}^2} B_{1,2,3}(\tilde{b}, \tilde{b})\) and \(\bigcup_{(\tilde{b}, \tilde{b}) \in \mathbb{Q}^2} Q_{1,2,3}(\tilde{b}, \tilde{b})\). Therefore, by requiring that such a union has probability measure zero, one gets \(P_{\Delta|X}(B) = 1\).

Moreover, as noticed above, the condition \(P_{\Delta|X}(B_{1,2,3}(\tilde{b}, \tilde{b})) = P_{\Delta|X}(Q_{1,2,3}(\tilde{b}, \tilde{b})) = 0\) is linear in \(G_{\Delta|X}\) because

\[ P_{\Delta|X}(B_{1,2,3}(\tilde{b}, \tilde{b})) = G_{\Delta|X}(\infty, \tilde{b}, \tilde{b}) - G_{\Delta|X}(\tilde{b} + \tilde{b}, \tilde{b}, \tilde{b}), \]

\[ P_{\Delta|X}(Q_{1,2,3}(\tilde{b}, \tilde{b})) = G_{\Delta|X}(\tilde{b} + \tilde{b}, \infty, \infty) - G_{\Delta|X}(\tilde{b} + \tilde{b}, \tilde{b}, \infty) - G_{\Delta|X}(\tilde{b} + \tilde{b}, \tilde{b}, \tilde{b}), \]

for any \((\tilde{b}, \tilde{b}) \in \mathbb{Q}^2\).

To give an idea on how the linear programming problem illustrated in step 1 should be modified, consider assuming that the conditional marginal probability distribution of the taste shock differences are symmetric about zero and selecting, just for simplicity, two 2-tuples \((\tilde{b}_1, \tilde{b}_1), (\tilde{b}_2, \tilde{b}_2)\) from \(\mathbb{Q}^2\) to approximate the degeneracy condition. Then, for a given \(U \in \mathcal{U}^1\), the \(\mathcal{A}\)-sets are

\[ \mathcal{A}_{x,1}(U) \equiv \{U_{x_0} - U_{x_1}, -U_{x_0} + U_{x_1}, \tilde{b}_1 + \tilde{b}_1, -\tilde{b}_1 - \tilde{b}_1, \tilde{b}_2 + \tilde{b}_2, -\tilde{b}_2 - \tilde{b}_2, \infty, -\infty\}, \]

\[ \mathcal{A}_{x,2}(U) \equiv \{U_{x_0} - U_{x_2}, -U_{x_0} + U_{x_2}, \tilde{b}_1, -\tilde{b}_1, \tilde{b}_2, -\tilde{b}_2, \infty, -\infty\}, \]

\[ \mathcal{A}_{x,3}(U) \equiv \{U_{x_1} - U_{x_2}, -U_{x_1} + U_{x_2}, \tilde{b}_1, -\tilde{b}_1, \tilde{b}_2, -\tilde{b}_2 \infty, -\infty\}, \]
\( \forall x \in \mathcal{X} \). The linear programming problem to solve becomes

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

\[
P_{Q(x)}(2) = \tilde{G}_{\Delta|x}^U(\infty, \infty, U_{x2} - U_{x1}) - \tilde{G}_{\Delta|x}^U(\infty, U_{x0} - U_{x2}, U_{x2} - U_{x1}),
\]

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

(17) is satisfied,

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

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

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

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

\[
\tilde{G}_{\Delta|x}^U(\infty, b_1^2, \infty) = 1 - \tilde{G}_{\Delta|x}^U(\infty, -b_2^2, -\infty),
\]

\[
\tilde{G}_{\Delta|x}^U(\infty, b_2^2, \infty) = 1 - \tilde{G}_{\Delta|x}^U(\infty, -b_2^2, -\infty),
\]

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

\[
\tilde{G}_{\Delta|x}^U(\infty, \infty, b_1^2) = 1 - \tilde{G}_{\Delta|x}^U(\infty, \infty, -b_2^2),
\]

\[
\tilde{G}_{\Delta|x}^U(\infty, \infty, b_2^2) = 1 - \tilde{G}_{\Delta|x}^U(\infty, \infty, -b_2^2),
\]

\[
\tilde{G}_{\Delta|x}^U(\infty, \bar{b}^1, \bar{b}^1) - \tilde{G}_{\Delta|x}^U(\bar{b}^1 + \bar{b}^1, \bar{b}^1, \bar{b}^1) = 0,
\]

\[
\tilde{G}_{\Delta|x}^U(\bar{b}^1 + \bar{b}^1, \infty, \infty) - \tilde{G}_{\Delta|x}^U(\bar{b}^1 + \bar{b}^1, \bar{b}^1, \infty) - \tilde{G}_{\Delta|x}^U(\bar{b}^1 + \bar{b}^1, \infty, \bar{b}^1) + \tilde{G}_{\Delta|x}^U(\bar{b}^1 + \bar{b}^1, \bar{b}^1, \bar{b}^1) = 0,
\]

\[
\tilde{G}_{\Delta|x}^U(\infty, \bar{b}^2, \bar{b}^2) - \tilde{G}_{\Delta|x}^U(\bar{b}^2 + \bar{b}^2, \bar{b}^2, \bar{b}^2) = 0,
\]

\[
\tilde{G}_{\Delta|x}^U(\bar{b}^2 + \bar{b}^2, \infty, \infty) - \tilde{G}_{\Delta|x}^U(\bar{b}^2 + \bar{b}^2, \bar{b}^2, \infty) - \tilde{G}_{\Delta|x}^U(\bar{b}^2 + \bar{b}^2, \infty, \bar{b}^2) + \tilde{G}_{\Delta|x}^U(\bar{b}^2 + \bar{b}^2, \bar{b}^2, \bar{b}^2) = 0.
\]

(21)

Before concluding the discussion, we note that in this last example the value of the parameter \( K_x \) of Proposition 3 depends on the choice of the 2-tuples \( (\bar{b}^1, \bar{b}^1) \) and \( (\bar{b}^2, \bar{b}^2) \). For example, if the 2-tuples are \( (3, 2) \) and \( (-1, -5) \), then \( K_x = 84 \forall x \in \mathcal{X} \).

### 4 Monte Carlo experiments

In this section we construct the sharp identified set for the vector of type-specific match surpluses, \( \Phi \), when \( r = 2 \) (\( d = 3 \)) as in Example 1 under two different data generating processes. According to the first data generating process (hereafter DGP1), the taste shocks on each side of the market are i.i.d. Gumbel, independent of types, with scale 0 and location 1, as in Choo and Siow (2006)\(^\text{12}\). Moreover, \( P_X(x) = P_Y(y) = \frac{1}{2} \forall (x, y) \in \mathcal{X} \times \mathcal{Y} \). 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 types as equiprobable Gaussian mixtures of 4 components,

\(^{11}\)4 has been computed without considering equal elements inside the \( \mathcal{A} \)-sets and without incorporating scale or location normalisations into \( U \).

\(^{12}\)Hence, the taste shock differences are distributed as Logistic.
\begin{align*}
\mathcal{N}\left(\begin{bmatrix}
-1 & 8.7 & 2.1 & 2.1 \\
-1 & 2.1 & 8.7 & 2.1 \\
-1 & 2.1 & 2.1 & 8.7
\end{bmatrix},
\mathcal{N}\left(\begin{bmatrix}
3.5 & 3.5 & 3.5 \\
0.9 & 0.9 & 0.9
\end{bmatrix}\right)\right),
\mathcal{N}\left(\begin{bmatrix}
-2 & 4.5 & -1.3 & -1.3 \\
-2 & -1.3 & 4.5 & -1.3 \\
-2 & -1.3 & -1.3 & 4.5
\end{bmatrix},
\mathcal{N}\left(\begin{bmatrix}
4.6 & 4.6 & 4.6 \\
-0.4 & -0.4 & 0.4
\end{bmatrix}\right)\right).
\end{align*}

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., agents prefer to be matched with individuals of the same type.

Lastly, \( \Phi_{0y} = \Phi_{x0} = 0 \) \( \forall (x,y) \in \mathcal{X} \times \mathcal{Y} \).

In order to construct the sharp identified set for \( \Phi \) we include scale and location normalisations into the parameter spaces, \( \mathcal{U}^f \) and \( \mathcal{V}^f \). Such normalisations ensure that the volume of the sharp identified set is not improperly inflated relative to the point identified case. Specifically, when the function spaces of admissible conditional probability distributions for the taste shock differences, \( \mathcal{P}^f_{\Delta\epsilon} \) and \( \mathcal{P}^f_{\Delta\eta} \), do not incorporate independence of the taste shock differences on types, we define

\begin{align}
\mathcal{U}^f &\equiv \{(U_{xy} \forall (x,y) \in \mathcal{X} \times \mathcal{Y}_0) \in \mathbb{R}^6: \\
&\text{[location normalisation]} \quad U_{x0} = 0 \ \forall x \in \mathcal{X} \\
&\text{[scale normalisation]} \quad U_{1y} = \frac{U_{1y}}{|U_{11}|} \text{ and } U_{2y} = \frac{U_{2y}}{|U_{21}|} \ \forall y \in \mathcal{Y}\}, \quad (22) \\
\mathcal{V}^f &\equiv \{(V_{xy} \forall (x,y) \in \mathcal{X}_0 \times \mathcal{Y}) \in \mathbb{R}^6: \\
&\text{[location normalisation]} \quad V_{0y} = 0 \ \forall y \in \mathcal{Y} \\
&\text{[scale normalisation]} \quad V_{x1} = \frac{V_{x1}}{|V_{11}|} \text{ and } V_{x2} = \frac{V_{x2}}{|V_{12}|} \ \forall x \in \mathcal{X}\}, \quad (23)
\end{align}

where the first condition exactly mimics Choo and Siow (2006) as a location normalisation, and the second condition is a scale normalisation. Instead, when \( \mathcal{P}^f_{\Delta\epsilon} \) and \( \mathcal{P}^f_{\Delta\eta} \) incorporate independence of the taste shock differences on types, we define

\begin{align}
\mathcal{U}^f &\equiv \{(U_{xy} \forall (x,y) \in \mathcal{X} \times \mathcal{Y}_0) \in \mathbb{R}^6: \\
&\text{[location normalisation]} \quad U_{x0} = 0 \ \forall x \in \mathcal{X} \\
&\text{[scale normalisation]} \quad U_{xy} = \frac{U_{xy}}{|U_{11}|} \ \forall (x,y) \in \mathcal{X} \times \mathcal{Y}\}, \quad (24) \\
\mathcal{V}^f &\equiv \{(V_{xy} \forall (x,y) \in \mathcal{X}_0 \times \mathcal{Y}) \in \mathbb{R}^6: \\
&\text{[location normalisation]} \quad V_{0y} = 0 \ \forall y \in \mathcal{Y} \\
&\text{[scale normalisation]} \quad V_{xy} = \frac{V_{xy}}{|V_{11}|} \ \forall (x,y) \in \mathcal{X} \times \mathcal{Y}\}. \quad (25)
\end{align}
Notice that when $P^{†}_\Delta \epsilon$ and $P^{†}_\Delta \eta$ do not incorporate independence of the taste shock differences on types we impose more scale normalisations than in the independence case. Specifically, when $P^{†}_\Delta \epsilon$ and $P^{†}_\Delta \eta$ do not incorporate independence of the taste shock differences on types, we impose one scale normalisation for each $x \in X$ on the men’s side and one scale normalisation for each $y \in Y$ on the women’s side. This is because the linear programming problem discussed in Sections 3.2.1 and 3.2.3 has to be solved separately for each $x \in X$ on the men’s side and for each $y \in Y$ on the women’s side. Instead, when $P^{†}_\Delta \epsilon$ and $P^{†}_\Delta \eta$ incorporate independence of the taste shock differences on types, we impose just one scale normalisation on the men’s side and one scale normalisation on the women’s side. This is because the linear programming problem discussed in Sections 3.2.1 and 3.2.3 has to be solved once on the men’s side and once on the women’s side.

The figures below report the sharp identified set for $\Phi$ under various classes of nonparametric distributional assumptions on the taste shock differences. Each figure is composed of six sub-figures where we project the sharp identified set for $\Phi$ 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 the data generated according to DGP1. The blue regions represent the projections of the sharp identified set for $\Phi$. The red dots represent the projections of the true value of $\Phi$ under the location and scale normalisations discussed earlier. Figure 1 reports the projections of the sharp identified set for $\Phi$ when no restriction on the conditional probability distributions of the latent variables is imposed, i.e., $P^{†}_\Delta \epsilon = P^{†}_\Delta \eta = P$. As expected, the blue regions are completely uninformative, i.e., for any value of $\Phi$ one can find some $\{P_{\Delta|X}, P_{\Delta|Y}\} \subseteq P$ that can reproduce the equilibrium matched type shares $\{P_{Y|X}, P_{X|Y}\}$.

We continue the analysis by incorporating into $P^{†}_\Delta \epsilon$ and $P^{†}_\Delta \eta$ increasingly restrictive nonparametric restrictions to see how the empirical content of the model varies under different scenarios. Figure 2 reports the projections of the sharp identified set for $\Phi$ when the analyst assumes that the conditional probability distributions of the taste shock differences have marginals symmetric about zero. Such a restriction allows to identify the signs of $\Phi_{22}$ and $\Phi_{11}$. Imposing in addition that the conditional probability distributions of the taste shock differences are independent of types and that their marginal probability distributions are symmetric about zero and identical does not seem to noticeably improve the identifying power of the model. Figure 4 reports the projections of the sharp identified set for $\Phi$ when the analyst assumes that the taste shock differences are independent of types and that their marginal probability distributions are symmetric about zero and identical. As before, this restriction seems sufficient to identify the signs of $\Phi_{22}$ and $\Phi_{11}$, but all the blue regions remain partially unbounded.

\[\text{13}\] Notice that the blue regions are not tighter than in Figure 3. This is because, as explained earlier, we impose less scale normalisations into $U^i$ and $V^i$ when $P^{†}_\Delta \epsilon$ and $P^{†}_\Delta \eta$ incorporate independence of the taste shock differences on types.
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 types. This enables to identify the sign of all the \( \Phi \)'s components. Moreover, the bounds on \( \Phi_{12} \) and \( \Phi_{21} \) are tight and instructive about their magnitudes. Imposing also that the taste shock differences are independent of types (Figure 6) permits to identify the sign of all the \( \Phi \)'s components and to obtain a narrow projection for \( \Phi_{12} \). Such last case is about as close as one can get to the framework of Choo and Siow (2006) without assuming mutually independent and Gumbel distributed taste shocks. Hence, the sizes of the blue regions can be interpreted as the cost of removing those restrictions on the latent variables. At least in DGP1, requiring the taste shocks to be mutually independent and Gumbel distributed seems to drive most of the identifying power of the model.

Figures 7, 8, 9, 10, 11, and 12 are based on the data generated according to DGP2. The blue regions represent the projections of the sharp identified set for \( \Phi \). The figures do not include the projections of the true value of \( \Phi \) under the location and scale normalisations discussed earlier because their expressions are very complicated to obtain under DGP2. We consider the same nonparametric distributional assumptions on the taste shock differences as done for DGP1. The blue regions look overall less informative than in DGP1. For example, under none of the nonparametric distributional assumptions considered, we obtain bounded intervals for some of the components of \( \Phi \). This suggests that, when the underlying data generating process is characterised by correlated taste shocks, the empirical content of the model is essentially determined by the parametric specification of the conditional probability distribution of the latent variables together with the numerical values assigned to its parameters.

---

14Remember that, as highlighted by Example 1, when man \( i \in \mathcal{I} \) of type \( x \in \mathcal{X} \) decides whether to choose a woman of type \( y \in \mathcal{Y}_0 \), he compares \( U_{xy} + \epsilon_{iy} \) with \( U_{x0} + \epsilon_{i0} \) and \( U_{xy} + \epsilon_{iy} \). Hence, the vector of taste shock differences that are relevant for such a type choice is \( (\epsilon_{iy} - \epsilon_{i0}, \epsilon_{iy} - \epsilon_{i\bar{y}}) \) with dimension 2. Similar considerations can be made for the women’s side.

15As before, notice that the blue regions are not tighter than in Figure 5. This is because, as explained earlier, we impose less scale normalisations into \( \mathcal{U}^i \) and \( \mathcal{V}^i \) when \( \mathcal{P}_{\Delta_U}^1 \) and \( \mathcal{P}_{\Delta_V}^1 \) incorporate independence of the taste shock differences on types.

16Their formulas are provided by Proposition 2 in Galichon and Salanié (2015).
Figure 1: The figure is based on the data generated according to DGP1 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ along every two of its dimensions when no restriction on the conditional probability distributions of the taste shocks is 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 the data generated according to DGP1 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ 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 the data generated according to DGP1 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ along every two of its dimensions when the researcher assumes that the conditional probability distributions of the taste shock differences have marginals identical and 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 4: The figure is based on the data generated according to DGP1 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ along every two of its dimensions when the researcher assumes that the taste shock differences are independent of 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 the data generated according to DGP1 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ 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 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 the data generated according to DGP1 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ along every two of its dimensions when the researcher assumes that (i) the taste shock differences are independent of 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.
Figure 7: The figure is based on the data generated according to DGP2 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ along every two of its dimensions when no restriction on the conditional probability distributions of the taste shocks is incorporated. The location and scale normalisations discussed in (22) and (23) are imposed.
Figure 8: The figure is based on the data generated according to DGP2 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ 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 9: The figure is based on the data generated according to DGP2 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ along every two of its dimensions when the researcher assumes that the conditional probability distributions of the taste shock differences have marginals identical and symmetric about zero. The location and scale normalisations discussed in (22) and (23) are imposed.
Figure 10: The figure is based on the data generated according to DGP2 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ along every two of its dimensions when the researcher assumes that the taste shock differences are independent of types with marginal probability distributions identical and symmetric about zero. The location and scale normalisations discussed in (24) and (25) are imposed.
Figure 11: The figure is based on the data generated according to DGP2 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ 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 types. The location and scale normalisations discussed in (22) and (23) are imposed.
Figure 12: The figure is based on the data generated according to DGP2 as described in the main text. The blue regions in the 6 sub-figures represent the projections of the sharp identified set for $\Phi$ along every two of its dimensions when the researcher assumes that (i) the taste shock differences are independent of 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 \( \{P_{Q,q|x}, P_{Q,\nu|y}, P_X, P_{\nu}\} \) are known. When doing an empirical analysis, the analyst should replace these probability distributions with their sample analogues. In what follows we illustrate how to construct a confidence region for \( \Phi \). Before proceeding, we remark that obtaining confidence regions for partially identified parameters that are uniformly asymptotically valid over a large class of probability distributions is a difficult problem and outside the scope of our work. Here we simply provide a description of how an existing method from the literature can practically be applied to the present setting. Specifically, we suggest to use a computationally tractable version of the profiled subsampling procedure illustrated by Romano and Shaikh (2008) and further investigated by Politis and Romano (1994) and Romano and Shaikh (2012). To keep the exposition readable, we continue the discussion by focusing on the case \( r = 2 \) (\( d = 3 \)) of Example 1.

Given \( U \in U^i \) and \( V \in \mathcal{V}^i \), consider firstly the linear equalities in (16) for each side of the market

\[
P_{Q,q|x}(1) = 1 + G_{q|x}^U(U_{x0} - U_{x1}, \infty, U_{x2} - U_{x1}) - \bar{G}_{q|x}^U(\infty, \infty, U_{x2} - U_{x1}) - \bar{G}_{q|x}^U(U_{x0} - U_{x1}, \infty, \infty),
\]

\[
P_{Q,q|x}(2) = \bar{G}_{q|x}^U(\infty, \infty, U_{x2} - U_{x1}) - \bar{G}_{q|x}^U(U_{x0} - U_{x2}, U_{x2} - U_{x1}),
\]

\[
P_{Q,q|x}(0) = \bar{G}_{q|x}^U(U_{x0} - U_{x1}, U_{x0} - U_{x2}, \infty),
\]

\[
P_{Q,\nu|y}(1) = 1 + G_{\nu|y}^V(V_{y0} - V_{y1}, \infty, V_{y2} - V_{y1}) - \bar{G}_{\nu|y}^V(\infty, \infty, V_{y2} - V_{y1}) - \bar{G}_{\nu|y}^V(V_{y0} - V_{y1}, \infty, \infty),
\]

\[
P_{Q,\nu|y}(2) = \bar{G}_{\nu|y}^V(\infty, \infty, V_{y2} - V_{y1}) - \bar{G}_{\nu|y}^V(V_{y0} - V_{y2}, V_{y2} - V_{y1}),
\]

\[
P_{Q,\nu|y}(0) = \bar{G}_{\nu|y}^V(V_{y0} - V_{y1}, V_{y0} - V_{y2}, \infty),
\]

\[
\forall (x, y) \in \mathcal{X} \times \mathcal{Y},
\]

(26)

where \( \forall (x, y) \in \mathcal{X} \times \mathcal{Y} \), the functions \( \bar{G}_{q|x}^U: \mathcal{A}(U) \to \mathbb{R} \) and \( \bar{G}_{\nu|y}^V: \mathcal{A}(V) \to \mathbb{R} \), and the \( \mathcal{A} \)-sets are constructed as explained above according to the nonparametric distributional assumptions of interest on the taste shock differences (Section 3.2.1) and on the grid points selected to approximate the degeneracy condition (Section 3.2.3). Let \( \bar{\omega}_{M,y|z}(U, \bar{G}_{q|x}^U) \) denote the right hand side of the equation for \( P_{Q,q|x}(y) \) in (26), \( \forall (x, y) \in \mathcal{X} \times \mathcal{Y}_0 \). Let \( \bar{\omega}_{W,x|y}(V, \bar{G}_{\nu|y}^V) \) denote the right hand side of the equation for \( P_{Q,\nu|y}(x) \) in (26), \( \forall (x, y) \in \mathcal{X}_0 \times \mathcal{Y} \). Hence, (26) is equivalent to the following system of unconditional moment equalities:

\[
E(m_{M,i,x,y}(U, \bar{G}_{q|x}^U)) = E(1 \{Q_{M,i} = y, X_i = x\} - \bar{\omega}_{M,y|z}(U, \bar{G}_{q|x}^U)1\{X_i = x\}) = 0 \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}_0,
\]

(27)

\[
E(m_{W,j,x,y}(V, \bar{G}_{\nu|y}^V)) = E(1 \{Q_{W,j} = x, Y_j = y\} - \bar{\omega}_{W,x|y}(V, \bar{G}_{\nu|y}^V)1\{Y_j = y\}) = 0 \quad \forall (x, y) \in \mathcal{X}_0 \times \mathcal{Y}.
\]
Next, given \((x, y) \in \mathcal{X} \times \mathcal{Y}\), let \(S_{U,V,x,y}\) be the collection of functions \(G_{\Delta|x}^U : \mathcal{A}_x(U) \to \mathbb{R}\) and \(G_{\Delta|y}^V : \mathcal{A}_y(V) \to \mathbb{R}\) which satisfy the constraints guaranteeing that \(G_{\Delta|x}^U\) is extendable to a conditional CDF in \(G_{\Delta|x}^U\), \(G_{\Delta|y}^V\) is extendable to a conditional CDF in \(G_{\Delta|y}^V\), and such conditional CDFs are concentrated on the region \(\mathcal{B}\), as discussed in Sections 3.2.1 and 3.2.3.

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 \(H_0 : \Phi = \Phi_0\). A \((1 - \alpha)\) confidence region for each \(\Phi \in \Theta^\dagger\) can then be constructed by inverting the test, i.e., by collecting all the values \(\Phi_0\) for which the test does not reject at level \(\alpha\). We assume that the analyst has a sample of i.i.d. observations \(\{Q_{M,i}, X_i, Q_{W,j}, Y_j\}_{i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\}}\). We propose the following test statistic.

\[
TS_n(\Phi_0) \equiv \inf_{U,V} \left\{ \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}_0} \left( \sqrt{\hat{m}_{M,n,x,y}(U, \hat{G}_{\Delta|x}^U)} \right)^2 + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}_0} \left( \sqrt{\hat{m}_{W,n,x,y}(V, \hat{G}_{\Delta|y}^V)} \right)^2 \right\}
\]

\[
\text{s.t. } U \in \mathcal{U}^\dagger, V \in \mathcal{V}^\dagger,
\]

\[
\{\hat{G}_{\Delta|x}^U, \hat{G}_{\Delta|y}^V\} \subseteq S_{U,V,x,y} \quad \forall (x,y) \in \mathcal{X} \times \mathcal{Y},
\]

\[
U + V = \Phi_0,
\]

where \(\hat{m}_{M,n,x,y}(U, \hat{G}_{\Delta|x}^U)\) and \(\hat{m}_{W,n,x,y}(U, \hat{G}_{\Delta|y}^V)\) are the empirical counterparts of the moments \(E(m_{M,i,x,y}(U, G_{\Delta|x}^U))\) and \(E(m_{W,j,x,y}(V, G_{\Delta|y}^V))\) in (27). We show in Appendix B that (28) can be rewritten as a mixed integer quadratic programming. The quadratic feature is because \(\hat{m}_{M,n,x,y}(U, \hat{G}_{\Delta|x}^U)\) and \(\hat{m}_{W,n,x,y}(U, \hat{G}_{\Delta|y}^V)\) are linear functions of \(U\) and \(V\), and, thus, the objective function is quadratic. The mixed integer feature is because \(S_{U,V,x,y}\) is non-convex. Indeed, the constraints in \(S_{U,V,x,y}\) requiring \(\hat{G}_{\Delta|x}^U\) and \(\hat{G}_{\Delta|y}^V\) to be 3-increasing functions are activated only for 3-tuples in \(\mathcal{A}_x(U)\) and \(\mathcal{A}_y(V)\) that are comparable component-wise, and, therefore, they are non-linear in \(U\) and \(V\). Such non-linear constraints can be casted into the problem using auxiliary 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\(^\text{17}\) and compute \(TS_{b_n,k}(\Phi_0)\), which is the test statistic in (28) using the \(k\)th subsample, \(\forall k \in \{1, \ldots, B_n\}\). Hence, the critical value \(\hat{c}_{n,1-\alpha}(\Phi_0)\) is set equal to the \((1 - \alpha)\)-quantile of the subsampling distribution

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

The test rejects if \(TS_n(\Phi_0) > \hat{c}_{n,1-\alpha}(\Phi_0)\). The \((1 - \alpha)\) confidence region for each

\(^\text{17}\)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 \(\lim_{n \to \infty} \frac{b_n}{n} = 0\).
\( \Phi \in \Theta^* \) is in turn
\[
C_{n,1-\alpha} \equiv \{ \Phi \in \Theta^1 : TS_n(\Phi) \leq \hat{c}_{n,1-\alpha}(\Phi) \}.
\] (29)

Constructing the \((1 - \alpha)\) confidence region \(C_{n,1-\alpha}\) can be computationally costly if performed on a personal computer, but the procedure becomes tractable with the assistance from a computing cluster.

6 Conclusions

In this paper we focus on the one-to-one matching models with transfers of Choo and Siow (2006) and Galichon and Salanié (2015). When the analyst has data on one large market only, we study (partial) identification of the systematic components of the agents’ preferences without imposing parametric restrictions on the probability distribution of the latent variables. Specifically, we provide a tractable procedure to characterise the region of parameter values that exhausts all the implications of the model and data (the sharp identified set), under various classes of nonparametric distributional assumptions on the unobserved terms. We also discuss a possible way to conduct inference on the sharp identified set and conclude with Monte Carlo simulations. One of our conclusions from the simulations is that even in a very simple setting, with limited heterogeneity in types, one market is not sufficiently informative on the systematic match surplus, under nonparametric distributional assumptions on the taste shocks. This is striking and illustrates the crucial role played by functional form assumptions on joint distributions of taste shocks in delivering point identification in one market.

We are currently working on an empirical illustration of the methodology to sorting of CEOs to firms using U.S. firm-level data - Compustat.
References

Agarwal, N., and W. Diamond (2017): “Latent Indices In Assortative Matching Models,” *Quantitative Economics*, 8(3), 685–728.

Canay, I., and A.M. Shaikh (2017): “Practical and Theoretical Advances in Inference for Partially Identified Models,” in *Advances in Economics and Econometrics: Eleventh World Congress of the Econometric Society (Econometric Society Monographs)*, ed. by B. Honore, A. Pakes, M. Piazzesi and L. Samuelson, 2, 271–306.

Chesher, A., A.M. Rosen, and K. Smolinski (2013): “An Instrumental Variable Model of Multiple Discrete Choice,” *Quantitative Economics*, 4(2), 157–196.

Chiappori, P.-A., and B. Salanié (2016): “The Econometrics of Matching Models,” *Journal of Economic Literature*, 54(3), 832–861.

Choo, E., and A. Siow (2006): “Who Marries Whom and Why,” *Journal of Political Economy*, 114(1), 175–201.

Dagsvik, J.K. (2000): “Aggregation in Matching Markets”, *International Economic Review*, 41(1), 27–57.

de Paula, A., S. Richards-Shubik, and E. Tamer (2018): “Identifying Preferences in Networks with Bounded Degree”, *Econometrica*, 86(1), 263–288.

Dupuy, A.M, and A. Galichon (2014): “Personality Traits and the Marriage Market,” *Journal of Political Economy* 122 (6), 1271–1319.

Fox, J.T. (2007): “Semiparametric Estimation of Multinomial Discrete Choice Models Using a Subset of Choices,” *The Rand Journal of Economics*, 38(4), 1002-1019.

Fox, J.T. (2009): “Structural Empirical Work Using Matching Models,” in S.N. Durlauf and L.E. Blume (Eds.), *New Palgrave Dictionary of Economics* (Online ed.) 203–254.

Fox, J.T. (2010): “Identification in Matching Games,” *Quantitative Economics*, 1(2), 203–254.

Fox, J.T. (2018): “Estimating Matching Games with Transfers,” *Quantitative Economics*, 9(1), 1–38.

Fox, J.T., C. Yang, and D.H. Hsu (2018): “Unobserved Heterogeneity in Matching Games,” *Journal of Political Economy*, 126(4), 1339–1373.

Gabaix, X. and A. Landier (2008): “Why has CEO pay increased so much?,” *The Quarterly Journal of Economics*, 123(1), 49-100.
Galichon, A., Kominers, S. D., and Weber, S. (2018): “Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility," *Journal of Political Economy*, forthcoming.

Galichon, A., and B. Salanié (2015): “Cupid’s Invisible Hand: Social Surplus and Identification in Matching Models,” Columbia University Discussion Paper 1415-02.

Graham, B. (2013): “Comparative Static and Computational Methods for an Empirical One-to-One Transferable Utility Matching Model,” in *Advances in Econometrics*, ed. by E. Choo and M. Shum, 31, 153–181.

Gretsky, N., J. Ostroy, and W. Zame (1992): “The Nonatomic Assignment Model,” *Economic Theory*, 2(1), 103–127.

Ho, K., and A.M. Rosen (2017): “Partial Identification in Applied Research: Benefits and Challenges,” in *Advances in Economics and Econometrics: Eleventh World Congress of the Econometric Society (Econometric Society Monographs)*, ed. by B. Honore, A. Pakes, M. Piazzesi and L. Samuelson, 2, 307–359.

Manski, C.F. (1975): “Maximum Score Estimation of the Stochastic Utility Model of Choice,” *Journal of Econometrics*, 3(3), 205–228.

Manski, C.F. (1985): “Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator,” *Journal of Econometrics*, 27(3), 313–333.

Manski, C.F. (1988): “Identification of Binary Response Models,” *Journal of the American Statistical Association*, 83(403), 729–738.

Matzkin, R.L. (1992): “Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and The Binary Choice Models,” *Econometrica*, 60(2), 239–270.

Matzkin, R.L. (1993): “Nonparametric Identification and Estimation of Polychotomous Choice Models,” *Journal of Econometrics*, 58(1-2), 137–168.

Menzel, K. (2015): “Large Matching Markets as Two-Sided Demand Systems,” *Econometrica*, 83 (3), 897–941.

Politis, D. N., and J.P. Romano (1994): “Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions,” *The Annals of Statistics*, 22(4), 2031–2050.

Romano, J. P., and A.M. Shaikh (2008): “Inference for Identifiable Parameters in Partially Identified Econometric Models,” *Journal of Statistical Planning and Inference*, 138(9), 2786-2807.
Romano, J. P., and A.M. Shaikh (2012): “On the Uniform Asymptotic Validity of Subsampling and the Bootstrap,” *The Annals of Statistics*, 40(6), 2798–2822.

Shapley, L., and M. Shubik (1972): “The Assignment Game I: The Core”, *International Journal of Game Theory*, 1(1), 111–130.

Sinha, S. (2018): “Identification in One-to-One Matching Models with Nonparametric Unobservables,” TSE Working Paper 18-897.

Sørensen, M. (2007): “How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital,” *The Journal of Finance*, 62(6), 2725–2762.

Tervio, M. (2008): “The difference that CEOs make: An Assignment Model Approach,” *American Economic Review*, 98(3), 642–668.

Torgovitsky, A. (2018): “Partial Identification by Extending Subdistributions,” *Quantitative Economics*, 10(1), 105–144.

Williams, H.P. (2013): *Model Building in Mathematical Programming*, Fifth Edition, Wiley.
A Proofs

Proof of Proposition 2  As mentioned in the main text, we refer the reader to Theorem 1 in 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  Step 1: As claimed by Proposition 2, Theorem 1 in Torgovitsky (2018) shows that, under various classes of nonparametric restrictions possibly incorporated into $P_{\Delta|x}$, determining whether, for a given $U \in U^I$, there exists $P_{\Delta|x} \in P_{\Delta|x}$ such that $P_{Q_m|x}(y) = \omega_{M,y|x}(U, P_{\Delta|x}) \forall (x,y) \in \mathcal{X} \times \mathcal{Y}$ is equivalent to finding whether a system of linear equalities and inequalities has at least one solution.

In this linear programming problem, the only chunk potentially generating different solutions for different values of $U$ is the one requiring $\bar{G}^U_{\Delta|x}$ to be a $d$-increasing function $\forall x \in \mathcal{X}$. Specifically, for a given $x \in \mathcal{X}$, from Definition 1 in Torgovitsky (2018), $\bar{G}^U_{\Delta|x} : \mathcal{A}_x(U) \to \mathbb{R}$ is $d$-increasing if

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

(A.1)

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

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

$\text{Vol}_{\bar{G}^U_{\Delta|x}}(u', u'')$ is the volume of the $d$-dimensional box $[u'_1, u''_1] \times [u'_2, u''_2] \times \cdots \times [u'_L, u''_L]$ and $\text{Vrt}(u', u'')$ is the collection of the vertices of this box. Hence, different values of $U$ can induce different constraints on the image set of $\bar{G}^U_{\Delta|x}$ of the type (A.1) when they differ in the collections of $u', u'' \in \mathcal{A}_x(U)$ such that $u' \leq u''$.

Such observation immediately suggests that if two values of $U$ induce component-wise comparisons featuring the same “coordinates” in $\mathcal{A}_x(U)$ (in a sense that will be specified below), then they should generate the same constraints on the image set of $\bar{G}^U_{\Delta|x}$ of the type (A.1) and, hence, have the same solution to the linear programming problem of Proposition 2 at $x \in \mathcal{X}$.

Step 2: For a given $x \in \mathcal{X}$, we provide sufficient conditions such that any $U, \tilde{U} \in U^I$ induce the same collection of constraints on the image sets of $\bar{G}^U_{\Delta|x}$, $\bar{G}^{\tilde{U}}_{\Delta|x}$ of the type (A.1).

Fix any order of the $\mathcal{A}$-sets so that they become tuples and apply such ordering choice to the $\mathcal{A}$-sets for each $U \in U^I$. Denote the resulting tuples by $a_{x,1}(U), ..., a_{x,d}(U)$. If applied to a finite tuple of real numbers, let $\pi$ be a function delivering the position of the
elements in that tuple when re-arranged from the smallest to the largest and the relational operators, \(<\) or \(=\), among the re-arranged elements. When the tuple considered contains multiple elements with the same value, then any convention on which element should be listed first can be adopted. For example, \(\pi(100, 99, \infty, -\infty) = \{(4, 2, 1, 3), (<, <, <)\}\) and \(\pi(5, 5, \infty, -\infty) = \{(4, 1, 2, 3), (<, =, <)\}\). We call the collection
\[
\{\pi(a_{x,1}(U)), ..., \pi(a_{x,d}(U))\}
\]
as the \(x\)-coordinates induced by \(U \in \mathcal{U}^\dagger\).

Take \(U, \tilde{U} \in \mathcal{U}^\dagger\). It is easy to see that if \(\pi(a_{x,1}(U)) = \pi(a_{x,1}(U)), ..., \pi(a_{x,d}(U)) = \pi(a_{x,d}(U))\), then \(U\) and \(\tilde{U}\) induce the same collection of constraints on the image set of \(\bar{G}^\dagger_{\Delta|x}, \tilde{G}^\dagger_{\Delta|x}\) of the type (A.1).

**Step 3:** For each \(x \in \mathcal{X}\), the number of feasible \(x\)-coordinates, \(K_x\), is finite because the \(A\)-sets are finite and, hence, the range of \(\pi\) is finite.

**Step 4:** By combining steps 2 and 3, it follows that, \(\forall x \in \mathcal{X}\), \(\mathcal{U}^\dagger\) can be ex-ante partitioned into a finite number \(K_x\) of subsets \(\{\mathcal{U}_{1,x}, ..., \mathcal{U}_{K_x,x}\}\) such that, \(\forall k \in \{1, ..., K_x\}\) and \(\forall \{U, \tilde{U}\} \subseteq \mathcal{U}_{k,x}^\dagger\),

\[
\exists P_{\Delta|x} \in \mathcal{P}_{\Delta}^\dagger \text{ such that } P_{\Delta|x} = \omega_{M,y|x}(U, P_{\Delta|x}) \forall y \in \mathcal{Y}_0
\]

if and only if

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

**Step 5:** For every \(k \in \{1, ..., K_x\}\) and \(x \in \mathcal{X}\), finding any \(U \in \mathcal{U}_{k,x}^\dagger\) amounts to solve a linear programming problem because \(\pi\) simply establishes linear relations among the \(A\)-sets’ elements.

**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 as defined in Proposition 4 are

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

\[
Q_{1,2,3}(\bar{b}, \tilde{b}) \equiv \{(x, y, z) \in \mathbb{R}^3 : x \leq \bar{b} + \tilde{b}, y > \bar{b}, z > \tilde{b}\}.
\]
for any $(\bar{b}, \tilde{b}) \in \mathbb{R}^2$.

**Step 2:** We show that

$$\bigcup_{(\bar{b}, \tilde{b}) \in \mathbb{Q}^2} B_{1.2.3}(\bar{b}, \tilde{b}) = \{(x, y, z) \in \mathbb{R}^3 : x > y + z\} \equiv A_1.$$  

It is clear that $\bigcup_{(\bar{b}, \tilde{b}) \in \mathbb{Q}^2} B_{1.2.3}(\bar{b}, \tilde{b}) \subseteq A_1$. To prove the reverse, take any $(x, y, z) \in A_1$ and $\epsilon \equiv x - (y + z) > 0$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists $p \in [y, y + \frac{\epsilon}{2}] \cap \mathbb{Q}$ and $p \in [z, z + \frac{\epsilon}{2}] \cap \mathbb{Q}$. Therefore, $x = y + z + \epsilon > p + q$ and, hence, $(x, y, z) \in Q_{1.2.3}(p, q)$.

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

$$\bigcup_{(\bar{b}, \tilde{b}) \in \mathbb{Q}^2} Q_{1.2.3}(\bar{b}, \tilde{b}) = \{(x, y, z) \in \mathbb{R}^3 : x < y + z\} \equiv A_2.$$

**Step 4:** Notice that $B^c = A_1 \cup A_2$, where $B^c$ denotes the complement of $B$ in $\mathbb{R}^3$. Moreover, $A_1$ and $A_2$ are disjoint and infinitely countable unions of zero probability measure sets. Therefore, $P_{\Delta|x}(B^c) = 0$ which is equivalent to $P_{\Delta|x}(B) = 1$.

### 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 are immediately generalisable to any $r$.

As explained in the main text, the mixed integer feature is because, $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}$, $S_{U,V,x,y}$ is non-convex. Indeed, the constraints in $S_{U,V,x,y}$ requiring $G_{\Delta|\mathcal{X}}^U$ and $G_{\Delta|\mathcal{Y}}^U$ to be 3-increasing functions are activated only for 3-tuples in $A_x(U)$ and $A_y(V)$ that are comparable component-wise, and, therefore, they are non-linear in $U$ and $V$. Such non-linear constraints can be casted 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 chunk of (18) requiring $G_{\Delta|\mathcal{X}}^U$ to be a 3-increasing function for a given $x \in \mathcal{X}$:

$$- G_{\Delta|\mathcal{X}}^U(t, q, r) + G_{\Delta|\mathcal{X}}^U(t', q, r) + G_{\Delta|\mathcal{X}}^U(t, q', r) - G_{\Delta|\mathcal{X}}^U(t', q', r)$$

$$+ G_{\Delta|\mathcal{X}}^U(t, q, r') - G_{\Delta|\mathcal{X}}^U(t', q, r') + G_{\Delta|\mathcal{X}}^U(t, q', r') - G_{\Delta|\mathcal{X}}^U(t', q', r') \geq 0 \quad (B.1)$$

$$\forall (t, q, r), (t', q', r') \in A_x(U) \text{ s.t. } (t, q, r) \leq (t', q', r').$$

46
(B.1) can be rewritten as the following collection of constraints:

\[
\forall (t,q,r),(t',q',r') \in A_x(U),
- (t,q,r) + (t',q',r') \geq (0,0,0) \Rightarrow -\bar{G}^U_{\Delta \epsilon|z}(t,q,r) + \bar{G}^U_{\Delta \epsilon|z}(t',q',r) - \bar{G}^U_{\Delta \epsilon|z}(t,q,r) + \bar{G}^U_{\Delta \epsilon|z}(t',q',r) \geq 0.
\]

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

\[
\forall (t,q,r),(t',q',r') \in A_x(U),
(1) - (t,q,r) + (t',q',r') \leq M(t,q,r),(t',q',r') \times (\lambda_1(t,q,r),(t',q',r'), \lambda_2(t,q,r),(t',q',r'), \lambda_3(t,q,r),(t',q',r')),
(II) \delta(t,q,r),(t',q',r') \geq 1 + \lambda_1(t,q,r),(t',q',r') + \lambda_2(t,q,r),(t',q',r') + \lambda_3(t,q,r),(t',q',r') - 3,
(III) - \bar{G}^U_{\Delta \epsilon|z}(t,q,r) + \bar{G}^U_{\Delta \epsilon|z}(t',q',r) - \bar{G}^U_{\Delta \epsilon|z}(t,q,r) + \bar{G}^U_{\Delta \epsilon|z}(t',q',r) \geq -M(t,q,r),(t',q',r')(1 - \delta(t,q,r),(t',q',r'))
\]

where, \(\forall (t,q,r),(t',q',r') \in A_x(U), \lambda_1(t,q,r),(t',q',r'), \lambda_2(t,q,r),(t',q',r'), \lambda_3(t,q,r),(t',q',r'), \delta(t,q,r),(t',q',r')\) are binary variables and \(M(t,q,r),(t',q',r')\) is chosen as small as possible\(^{18}\) but such that

\[
M(t,q,r),(t',q',r') \geq 4,
- (t,q,r) + (t',q',r') \leq M(t,q,r),(t',q',r') \forall U \in U^\dagger.
\]

To see why (B.2) is equivalent to (B.3), notice that

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

\[
\downarrow
\]

\[
\lambda_1(t,q,r),(t',q',r') = \lambda_2(t,q,r),(t',q',r') = \lambda_3(t,q,r),(t',q',r') = 1 \text{ so that (I) is satisfied by (B.5)}
\]

\[
\downarrow
\]

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

\[
\downarrow
\]

Desired constraint on the image set of \(\bar{G}^U_{\Delta \epsilon|z}\) are activated through (III) and that

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

\[
\downarrow
\]

\[
\lambda_1(t,q,r),(t',q',r'), \lambda_2(t,q,r),(t',q',r'), \lambda_3(t,q,r),(t',q',r') \text{ can be 1 or 0 and, in any case, (I) is satisfied by (B.5)}
\]

\[
\downarrow
\]

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

\[
\downarrow
\]

Desired constraint on the image set of \(\bar{G}^U_{\Delta \epsilon|z}\) are activated when \(\delta(t,q,r),(t',q',r') = 1\); otherwise (III) satisfied by (B.4)

\[^{18}\]M(t,q,r),(t',q',r') unnecessarily large induces bad numerics in solvers and makes harder to solve the integer program.
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 functions $\bar{G}_{\Delta|x}^U : A_x(U) \to \mathbb{R}$ and $\bar{G}_{\Delta|y}^V : A_y(V) \to \mathbb{R}$ which satisfy the constraints guaranteeing that $\bar{G}_{\Delta|x}^U$ is extendable to a conditional CDF in $\mathcal{G}_{\Delta|x}^\dagger$, $\bar{G}_{\Delta|y}^V$ is extendable to a conditional CDF in $\mathcal{S}_{\Delta|y}^\dagger$, and such conditional CDFs are concentrated on the region $\mathcal{B}$, as discussed in Sections 3.2.1 and 3.2.3 but using (B.3) in place of (B.1) for each side of the market. Then, (28) becomes

$$TS_n(\Phi_0) \equiv \inf_{U,V,\Lambda_M,\Lambda_W} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left( \sqrt{n} \hat{m}_{M,n,x,y}(U,\bar{G}_{\Delta|x}^U) \right)^2 + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left( \sqrt{n} \hat{m}_{W,n,x,y}(V,\bar{G}_{\Delta|y}^V) \right)^2$$

subject to $U \in U^\dagger, V \in V^\dagger$,

$$\{\bar{G}_{\Delta|x}^U, \bar{G}_{\Delta|y}^V\} \subseteq \mathcal{S}_{U,V,\Lambda_M,\Lambda_W,x,y} \ \forall (x, y) \in \mathcal{X} \times \mathcal{Y},$$

$$U + V = \Phi_0,$$

which corresponds to a mixed integer quadratic programming.

(B.6)