IDENTIFYING A SCREENING MODEL WITH MULTIDIMENSIONAL PRIVATE INFORMATION*

GAURAB ARYAL†

Abstract. In this paper I study the nonparametric identification of screening (price discrimination) models when consumers have multidimensional private information about their taste for product characteristics. In particular, I consider the model developed by Rochet and Choné (1998) and determine conditions to identify the cost function, the joint density of taste and the utility functions, from individual level data (on demand and prices). When the utility function is nonlinear, exogenous binary cost shifter is sufficient for identification. Moreover, I show that if there are some consumer covariates and the utility is nonlinear the model is over identified. I also characterize all testable restrictions of the model on the data.

Keywords: multidimensional screening, multiproduct nonlinear pricing, identification.

Date: First Version: November 1, 2014. This Version: Jan 7, 2015.
∗ Thanks to be added.

†University of Chicago. e-mail: aryalg@uchicago.edu.

Contents

1. Introduction

In this paper I study the identification of a static screening (nonlinear pricing) model where consumers have multidimensional private information about their preferences. Using both demand and supply side optimality conditions, within a framework of Lancaster (1971) with asymmetric information, I determine conditions under which we can (or cannot) identify the cost function, the joint density of consumers' multivariate types (taste of reach characteristics) and the (nonlinear) utility function. In so doing, I also explore the limit of identification from individual level data from *only one* market. I show that from that data we can identify the (truncated) joint density of consumer types, the cost function, however, it is insufficient to identify the nonlinear utility. To identify the utility we need some exogenous binary variation in cost, and hence the relative prices. This variation could any cost shifter over two time periods or over two markets served by the same seller. I also show that the model is *over-identified* if the utility function depends on consumers' observed characteristics. Over-identification can be used in a specification test to check the validity of optimal nonlinear pricing. Furthermore, I also determine the empirical content of the model by characterizing all testable implication of the model on the data, which consist of the choices and price for each consumer. I also show that while the identification is robust with respect to (separable and independent) measurement error in prices and some factor that affects the consumer taste for the product, it fails if the choices are measured with error or if some product characteristics are missing in the data but correlated with the observed characteristics.

In economics, at least since Akerlof (1970); Spence (1973); Rothschild and Stiglitz (1976), it has been believed that information asymmetry is a universal phenomenon, and that it always leads to significant loss of (social) welfare. This belief has been a guide to shape policy and regulatory choices, see Baron (1989); Joskow and Rose (1989); Laffont (1994). The empirical support for this belief, however, has been been mixed at best. Either there is no evidence of asymmetric information Chiappori and Salanié (2000) or the welfare loss is insignificant Einav, Finkelstein, and Cullen (2010). A probable reason could be that almost all models assume (only for tractability and simplicity) that the information asymmetry can be captured by a single dimensional parameter. But, in many environments consumers have multiple unobserved characteristics or types that cannot be sorted out in a satisfactory manner according to only one of these characteristics.

For motivation lets consider a static insurance market that has been studied the most in the literature and, as a result, the one we understand the most. Using French automobile insurance data Chiappori and Salanié (2000) find that insurees with high coverage do not file nearly as many claims as the theory (with one dimensional) adverse selection would suggest. This could be because insurees differ in both risk and risk aversion (see Finkelstein and McGarry, 2006; Cohen and Einav, 2007) and both are unobserved by the insurance company. Then those who buy higher coverage could be the ones with high risk aversion but low risk, and hence they file fewer claims. Furthermore, if risk and risk aversion are negatively correlated then it could lead to advantageous selection and not adverse selection, because those who want to buy better coverages are also the ones who have lower risk. Therefore, to get a full empirical picture we must allow multidimensional private information and model the optimality of both demand and the supply side while leaving the dependence between the types (in this example, risk and risk preference) unspecified. Aryal, Perrigne, and Vuong (2010) propose a way to nonparametrically identify the joint density of risk and risk aversion in an automobile insurance market. But beyond insurance market, we know very little about the role and importance of multidimensional private information.

Crawford and Yurukoglu (2012) study cable television market, where consumers are heterogenous with respect to their taste for each channel and they choose from multiple packages (of channels). So each household can be characterized by a vector of (privately known) taste parameters. Furthermore, since the observed packages (menu) will depend on the joint distribution of the taste profile, it is equally important to model the supply side as well. The profit (for the seller) is presumably highest for the products/contracts that are designed for high-type consumers, those who have higher willingness to pay or value the contract more. These high-types, however, cannot be prevented from choosing products that are meant for the medium-types or the low-types, and higher profit can only be realized if the seller distorts the products meant for the latter types in the direction that makes them relatively unattractive for the high-types. This distortion is what leads to inefficiency and hence loss of welfare, but to understand (and to estimate) the level of distortion we have to model the supply side. Rochet and Choné (1998) (henceforth, Rochet-Choné) shows that equilibrium with multidimensional private information is qualitatively very different from the equilibrium with one dimension. Thus, it is imperative that we allow multidimensional private information (type) and model the profit maximizing seller's problem of determining the product line and pricing function.[1]

In this paper I determine conditions that will allow us to use individual consumer choice (bundle and price) and characteristics data to identify the utility function, cost function and the joint density of consumer types, under the assumption that consumer choices are generated from Rochet-Choné equilibrium. The paper contributes to the research on empirical mechanism design, as articulated by Chiappori and Salanié (2003), by showing how we can make use of a theory of nonlinear pricing to improve our understanding of an environment where a monopoly seller sells a product which is characterized by multiple attributes to consumers with heterogenous tastes for each attribute.[2] This paper also contributes to the literature on demand estimation with endogenous product characteristics and prices, but where neither the seller nor the econometrician observes the consumer's willingness to pay for the characteristics, modulo the assumption about the single seller. Such an exercise is crucial when we want to study the effect of product repositioning on welfare similar to Fan (2013) who considers products (newspapers) that are a bundle of endogenously chosen continuous three characteristics (the news content quality index, local news ratio, and variety). While she considers a complete information game where the newspaper sellers know the consumer tastes, in my environment they do not. As mentioned above, one shortcoming of my exercise is that I consider an environment with only single seller. I chose this route because introducing competition with asymmetric information about multidimensional tastes is considerably difficult. Even though one seller could be restrictive in some market, there are some markets served either by a (local) monopoly or markets with multiple sellers with one large seller so that a treating that market as a monopoly could be a good first-order approximation, not to mention the fact that understanding identification with one seller can be a necessary first-step in understanding identification with multiple sellers. Identification in an competitive market is beyond the scope of this paper and is addressed elsewhere; (see Aryal, 2015).

[1] Multidimensionality also arises when consumers buy from multiple sellers, Aryal (2013).

[2] Since those tastes are private information of the consumer determining optimal (nonlinear) pricing and product varieties leads to multidimensional screening.

Nonetheless, with appropriate changes in notations and interpretation, the analysis and the results presented here apply among others to: a) multi-product monopoly seller where consumers have different taste for each product; b) an employer contracts on multiple tasks with several ex-ante identical workers with task specific skills; c) a regulation model where the regulated firm has private information about fixed cost and marginal cost; d) to the problem of designing optimal taxation for couples, where under assortative mating, Becker (1973, 1974); Siow (Forthcoming), a household is characterized by a two-dimensional productivity parameter that are correlated and known only by the couple, (see Kleven, Kreiner, and Saez, 2009);

In a multidimensional screening model, a seller knows its cost function offers a menu – pair of available qualities and prices – to consumers with (unobserved) heterogenous taste for each quality that maximizes the seller's expected profit with respect to a known joint density of taste parameters. Any observed consumer (socioeconomic) characteristics are conditioned upon (third-degree price discrimination). Rochet-Choné shows that in equilibrium, it is optimal for the seller to divide the type space into three sub-groups: a) the high-types, who are perfectly screened – each type is offered a unique bundle of qualities; b) the medium-types, who are further divided into different sub-categories such that everyone in the same index is offered the same bundle; and c) the low-types who are always excluded from the market or offered the outside option. I consider each of these sub-groups separately, and each with linear, bi-linear and nonlinear utility. Considering progressively more general form of utility allows me to highlight the role each additional assumption plays in identification.

The identification follows the following logic. When the utility is linear, the high types (who are allocated unique bundles) can be identified from (they are equal to) the price gradient because for them marginal utility from each characteristics is equal to the marginal prices and the former is equal to the types. The cost function can be identified, on an open and convex subset, from the equilibrium allocation rule. Furthermore, if the cost function is a real analytic (which includes all polynomials, exponential and trigonometric functions), it is identified everywhere. This identification strategy fails for medium-types because of bunching. Nonetheless, we might be interested in learning more about these types, for which I show we can exploit variation in consumer characteristics. If the utility function is bi-linear with as many covariates as types, each independent of types then the joint density of medium-types can be identified.

For similar identification strategy for random coefficients model (see Gautier and Hoderlein, 2012; Gautier and Kitamura, 2013; Hoderlein, Nesheim, and Simoni, 2013).

When, however, the utility is nonlinear we lose identification even for the high types because now the marginal utility depends on the type and the utility function. However, if there is an exogenous *binary* cost shifter that affects the prices and hence the choices, but not the utility or the type density, the model can be identified. Such cost shifters can be wide ranging from either different advertisement costs (coupons, say) across two markets, or changes in regulation that affects the cost of production, or even the same market over two periods. Exclusion restriction implies that the multivariate quantiles, Koltchinskii (1997), of demand under two cost regimes are the same. Furthermore, among the high-types, for any two consumers who buy the same bundle but at different (marginal) prices under two cost regimes must be such that the ratio of their types is equal to the ratio of the marginal prices. I show that these two restrictions can be used to identify the (multivariate) quantile function of type after normalizing a location.[3]

If the utility also depends on consumer characteristics, then the model is *over-identified*. This over-identification result uses a result from optimal mass transportation problem; see Brenier (1991); McCann (1995). Over-identification can be used in a specification test to check the validity of optimal nonlinear pricing, which to the best of my knowledge is new. I also show that while the identification is robust with respect to the classical measurement error in prices and unobserved, but independent, taste shifter, it fails when either the consumers' choices are measured with errors or if there is an unobserved product characteristics that is correlated with the observed types. Furthermore, I also determine the empirical content – testable implications – of the of the model (linear, bilinear and nonlinear utility) on the data, which will allow us to reject the Rochet-Choné model.

This paper also is closely related to Perrigne and Vuong (2011) who studied identification of contract models with adverse selection and moral hazard, and to Gayle and Miller (2014) who study identification and empirical content of the pure moral hazard and hybrid moral hazard principal-agent models; and to Aryal, Perrigne, and Vuong (2010) who show how we can nonparametrically

[3] This identification strategy of equating quantiles is similar to identification of utility function in first-price auction Guerre, Perrigne, and Vuong (2009) and a triangular system with discrete instrument variables by D'Haultfœuille and Février (2014); Torgovitsky (2014).

identify the joint distribution of risk and risk aversion in an automobile insurance market. In another related paper Luo, Perrigne, and Vuong (2012) use the model proposed by Armstrong (1996) to study telecommunication data, and Ivaldi and Martimort (1994); Aryal (2013) estimate consumer heterogeneity using nonlinear pricing with competition and multidimensional taste parameters. Since the identification argument uses invertibility of an equilibrium allocation rule (or equivalently a demand function) the paper is also related to the extensive and important research on invertibility of demand system; see Berry (1994); Berry, Gandhi, and Haile (2013).

The remainder of the paper is organized as follows. I begin by introducing notations in section 2. Section 3 describes the model of Rochet-Choné ; section 4 studies the identification, while section 5 provides the rationalizability lemmas for the models. Finally section 6 extends identification arguments with measurement error and unobserved heterogeneity in the data.

2. Notations and Definitions

For any variable ζ, I will use the notation $\zeta \in \mathcal{S}_\zeta \subset \mathbb{R}^{d_\zeta}$ to mean that it is d_ζ dimensional vector that can take value in the set $\mathcal{S}_\zeta$. The boundary values for a set $\mathcal{S}_\zeta$ is denoted by $\partial\mathcal{S}_\zeta$. If γ is another d_ω-dimensional vector then $\zeta\cdot\gamma$ denotes the inner product that is equal to $\sum_{j=1}^{d_\zeta}\zeta_j\gamma_j$, where ζ_j is the j^{th} element of the vector ζ. The function $\kappa:\mathcal{S}_\zeta\to\mathbb{R}^{d_\kappa}$ defines a d_κ dimensional vector of functions, i.e.,

$$\kappa(\zeta)=\begin{pmatrix}\kappa_1(\zeta_1,\ldots,\zeta_{d_\zeta})\\ \vdots\\ \kappa_{d_\kappa}(\zeta_1,\ldots,\zeta_{d_\zeta})\end{pmatrix}.$$

Consider the scalar function $\kappa(\zeta_1,\ldots,\zeta_{d_\zeta})\in\mathbb{R}$, $\nabla\kappa=\left(\frac{\partial\kappa}{\partial\zeta_1},\ldots,\frac{\partial\kappa}{\partial\zeta_{d_\zeta}}\right)$ denotes the gradient of the function $\kappa(\cdot)$, such that $\nabla_j\kappa(\cdot)$ is the j^{th} element of the gradient vector. The divergence of a scalar function $\kappa(\zeta)$ is defined as $\mathbf{div}\kappa=\sum_{j=1}^{d_\zeta}\frac{\partial\kappa(\zeta)}{\partial\zeta_j}$; Lang (1973).

Definition 2.1. *A scalar function $\kappa:\mathbb{R}^{d_\zeta}\to\mathbb{R}$ is a real analytic function at $\check{\zeta}\in\mathbb{R}^{d_\zeta}$ if there $\exists\delta>0$ and open ball $B(\check{\zeta},\delta)\subset\mathbb{R}^{d_\zeta}, 0\le r<\delta$ with $\sum_{k_1,\ldots,k_J}|a_{k_1,\ldots,a_{K_J}}|r^{k_1+\cdots+k_J}<\infty$ such that*

$$\kappa(\check{\zeta})=\sum_{k_1=0}^{\infty}\cdots\sum_{k_J=0}^{\infty}a_{k_1,\ldots,a_{K_J}}(\check{\zeta}_1-\zeta_1)^{k_1}\cdots(\check{\zeta}_J-\zeta_J)^{k_J},\zeta\in B(\check{\zeta},\delta).$$

One of the properties of analytic functions is that if two real analytic convex functions coincide on an open set then they coincide on any connected open subset of $\mathbb{R}^{d_\varsigma}$.

3. The Model

Consider a multidimensional screening environment analyzed by Rochet-Choné where a seller offers a product line $Q \subseteq \mathbb{R}_+^{d_q}$ of multiple products with a pricing function $P : Q \to \mathbb{R}_+$, together known as a menu, to agents (or consumers) who have multidimensional taste for the products. Let $\theta \in \mathcal{S}_\theta \subseteq \mathbb{R}_+^{d_\theta}$ denote the taste (or type), and each agent draws his type independently and identically (across agents) from a cumulative distribution function $F_\theta(\cdot)$. Agents also have some observed socioeconomic and/or demographic characteristics $X \in \mathcal{S}_x \subseteq \mathbb{R}^{d_x}$. Once the menu is offered, and a type θ agent chooses $\mathbf{q} \in Q$ he transfers $P(\mathbf{q})$ to the principal. I assume that his net payoff/utility is quasilinear in transfer, and is given by

$$V(\mathbf{q}; \theta, X) := u(\mathbf{q}, \theta, X) - P(\mathbf{q}).$$

Let $C : \mathbb{R}_+^{d_q} \to \mathbb{R}_+$ be the cost function with the interpretation that $C(\mathbf{q})$ is the cost of producing $\mathbf{q}$. The objective of the principal is to choose a (convex) set of product varieties Q and a transfer function $P(\cdot)$ that maximizes her expected profit when the cost function is given by $C(\cdot)$ and the CDF $F_\theta(\cdot)$. For notational simplification I suppress the dependence of Q and $P(\cdot)$ on X and Z. I begin with the following assumptions:

Assumption 1. *Let*
(i) $d_\theta = d_q = J$.
(ii) $\theta \overset{i.i.d}{\sim} F_\theta(\cdot)$ *which has a square integrable density* $f(\cdot) > 0$ *a.e. on* $\mathcal{S}_\theta$.
(iii) The net utility be an element of a Sobolev space

$$V(\mathbf{q}; \cdot, X) \in \mathcal{V}(\mathcal{S}_\theta) = \{V(\mathbf{q}, \cdot, X) | \int_{\mathcal{S}_\theta} V^2(\theta)\mathrm{d}\theta < \infty, \int_{\mathcal{S}_\theta} (\nabla V(\mathbf{q}, \theta, X))^2 \mathrm{d}\theta < \infty\},$$

with the norm $|V| = \left(\int_{\mathcal{S}_\theta} \{V^2(\theta) + ||\nabla V(\theta)||^2\}\mathrm{d}\theta\right)^{\frac{1}{2}}$.
(iv) The gross utility $u(\mathbf{q}, \theta, X) := \theta \cdot \mathbf{v}(\mathbf{q}, X)$, *where* $\mathbf{v}(\cdot, X) : \mathbb{R}_+^J \to \mathbb{R}_+^J$ *is a vector* $(v_1(\cdot), \ldots, v_J(\cdot))$ *where each function is differentiable and strictly increasing. Therefore* $u(\mathbf{q}; \theta, X) = \sum_{j=1}^J \theta_j v_j(q_j)$, *such that each* $v_j(\cdot, \cdot)$ *is either:*
*(iv-a) (***linear utility***):* $v_j(q_j, X) = q_j$.
*(iv-b) (***bilinear utility***):* $X \equiv (X_1, X_2) \in \mathbb{R}^{d_{x_1} + d_{x_2}}$ *such that* $d_{x_1} = J, d_{x_1} + d_{x_2} = d_x$ *and* $v_j(q_j, X) = q_j \cdot X_{1_j}$ *for* $j = 1, \ldots, J$.

(iv-c) (**nonlinear utility**)*:* $v_j(q_j, X) = X_{1j} \cdot v_j(q_j, X_2)$ *such that* $v_j(\cdot, X_2)$ *is twice continuously differentiable and strictly quasi concave, with full rank Jacobian matrix* $Dv(\mathbf{q}; X_2)$ *for all* $\mathbf{q} \in \mathbb{R}_+^J, v_j(0; \cdot) = 0$ *and* $\lim_{q\to\infty} v_j(q) = \infty$*.*
(v) $C(\cdot)$ *be a strongly convex function with parameter* ϵ'*, i.e. the minimum eigenvalues of the Hessian matrix is* ϵ'*.*

Assumption 1-(i) assumes that agents differ in as many dimensions as the attributes of a contract. It means that consumers' unobserved preference heterogeneity is exactly as rich as the number of products sold by the seller. When $d_\theta \neq d_q$ the model is very different, for instance if $d_\theta > d_q$ perfect screening is not possible and in equilibrium all agents are bunched.Assumption 1-(iv) is very important for our analysis. The first part suggests that the utility can be multiplicatively separated from the unobserved type θ and the base utility from $\mathbf{q}$ given the consumer characteristics is X. The second part of the assumption considers three progressively more general, forms for the base utility function. I begin with the most basic form, which is primarily what is considered in the theory literature. The second form allows me to interact the observed characteristics with unobserved types, and in writing the utility as a inner product I am implicitly assuming that there be as many observed characteristics as products, i.e. $d_x = J$. The third form is the most general form and allows nonlinear interaction between the bundles and consumer characteristics. Nonetheless these assumptions are important and affect identification–more on this later. Assumption 1-(v) assumes that the cost function is strictly increasing and convex.

A menu (or nonlinear pricing) $\{Q, P\}$ is feasible if there exists an allocation rule $\rho : \mathcal{S}_\theta \to Q$ that satisfies incentive compatibility (IC) condition, i.e.

$$\forall\theta \in \mathcal{S}_\theta, V(\rho(\theta), \theta) = \max_{\tilde{\mathbf{q}}\in Q} \{\theta \cdot \mathbf{v}(\tilde{\mathbf{q}}) - P(\tilde{\mathbf{q}})\} \equiv U(\theta), \tag{1}$$

and individual rationality (IR) condition: $U(\theta) \geq U_0 := \theta \cdot \mathbf{v}(\mathbf{q}_0) - P_0$. Here, $\{\mathbf{q}_0\}$ is the outside option available to all types at price P_0. To ensure the principal's optimization problem is convex, we assume that $P_0 \geq C(\mathbf{q}_0)$, so that the principal will always offer $\mathbf{q}_0$, i.e. $Q \ni \mathbf{q}_0$.[4] The principal chooses a feasible menu $(Q, \rho(\cdot), P(\mathbf{q}))$, that maximizes expected profit

$$\mathbb{E}\Pi = \int_{\mathcal{S}_\theta} \pi(\theta)\mathrm{d}F(\theta) := \int_{\mathcal{S}_\theta} \mathbb{1}(U(\theta) \geq U_0)\{P(\mathbf{q}(\theta)) - C(\mathbf{q}(\theta))\}\mathrm{d}F(\theta), \tag{2}$$

[4] This condition is violated in some instances, such as in the yellow pages advertisement market studied by Aryal (2013), where $\mathbf{q}_0$ was available for free.

where $\mathbb{1}\{A\}$ is a logical operator that equal to one if and only if $\{A\}$ is true. Let $S(\rho(\theta), \theta)$ be the social surplus when θ type is allocated $\mathbf{q}(\theta)$, so either

$$\begin{aligned} S(\rho(\theta), \theta) &= U(\theta) + \pi(\theta), \\ or, \quad S(\rho(\theta), \theta) &= \{\theta \cdot \mathbf{v}(\rho(\theta)) - P(\rho(\theta))\} + \{P(\rho(\theta)) - C(\rho(\theta))\}. \end{aligned}$$

Equating the two definitions, allows us to express the type (θ) specific profit as

$$\pi(\theta) = \theta \cdot \mathbf{v}(\rho(\theta)) - C(\rho(\theta)) - U(\theta).$$

In an important paper Rochet (1987) showed that under Assumption 1, a menu $\{Q, \rho(\cdot), P(\cdot)\}$ is such that $U(\theta)$ solves Equation (1) (satisfies IC) if and only if: (i) $\rho(\theta) = \mathbf{v}^{-1}(\nabla U(\theta))$; and (ii) $U(\cdot)$ is convex on Θ. This means that choosing an optimal contract $\{Q, \rho(\cdot), P(\cdot)\}$ is equivalent to determining the net utility (or the information rent) $U(\theta)$ that each θ gets by participating, because from $U(\theta)$ we can determined the optimal allocation as $\rho(\theta) = \mathbf{v}^{-1}(\nabla U(\theta))$. So the principal chooses $U(\theta) \in H^1(\mathcal{S}_\theta)$ to maximizes

$$\mathbb{E}\Pi(U) = \int_{\mathcal{S}_\theta} \{\theta \cdot \nabla U(\theta) - U(\theta) - C(\mathbf{v}^{-1}(\nabla U(\theta)))\} \mathrm{d}F(\theta),$$

subject to IC and IR constraints.

The global IC constraint is equivalent to convexity of $U(\cdot)$, i.e. $D^2U(\theta) \geq 0$, and IR is equivalent to $U(\theta) \geq U_0(\theta)$ for all $\theta \in \mathcal{S}_\theta$. Rochet-Choné showed that Assumption 1 is sufficient to guarantee existence of a unique maximizer $U^*(\cdot)$. In what follows we will characterize some key properties of the solution. This, however, requires us to solve the variational problem with inequality constraints that is known to be difficult. When $J = 1$, we can ignore the inequality constraints to find an unconstrained maximizer, and only then verify that these inequality (IC) constraints are satisfied, under the assumption that the type distribution is regular (the inverse hazard rate $[1 - F(\cdot)]/[f(\cdot)]$ is strictly decreasing). When $J > 1$, however, there are not any such "regularity" conditions that are easy to verify, except the ones in Armstrong (1996) and Wilson (1993), where they proposed two alternative methods that require very strong and non testable restrictions on $F_\theta(\cdot)$. Moreover, Rochet-Choné have shown that those assumptions are restrictive and are seldom satisfied. Therefore imposing such restrictions to simplify the problem, is at odds with the nonparametric objective of this paper.

One of the main insights from Rochet-Choné is that with multidimensional type, the principal will always find it profitable not to perfectly screen agents, even when like us $d_\theta = d_q$. In other words, bunching is robust outcome in an environment with multidimensional type (adverse selection). Even then, determining the types that get bunched and the types that are not would depend on the model parameters. Since, under bunching two distinct types of agents choose the same option, or the optimal allocation rule $\rho(\cdot)$ is not injective everywhere in its domain, it affects the identification strategy; see Aryal (2013) and Aryal, Perrigne, and Vuong (2010). Rochet-Choné showed that agents can be divided into three types: the lowest-types $\mathcal{S}_\theta^0$ who are screened out and offered only $\{\mathbf{q}_0\}$, the medium-types $\mathcal{S}_\theta^1$ who are bunched and offered "medium type" of bundles and the high-types $\mathcal{S}_\theta^2$ who are perfectly screened. So, $\rho(\cdot)$ is injective only when restricted to $\mathcal{S}_\theta^2$.

If an indirect utility function $U^*(\cdot)$ is optimal then offering any other feasible function $(U^*+h)(\cdot)$, where h is non-negative and convex, must lower expected profit for the principal, i.e., $\mathbb{E}\Pi(U^*) \geq \mathbb{E}\Pi(U^*+h)$. This means the directional derivative of the expected profit, in the direction of h, must be nonnegative so $U^*(\cdot)$ is the solution iff: (a) $U^*(\cdot)$ is convex function and for all convex, non negative function h, $\mathbb{E}\Pi'(U^*)h \geq 0$; and (b) $\mathbb{E}\Pi'(U^*)(U^*-U_0) = 0$ with $(U^*-U_0) \geq 0$. The Euler-Lagrange condition for the (unconstrained) problem is

$$\frac{\partial \pi}{\partial U^*} - \sum_{j=1}^{J} \frac{\partial}{\partial \theta_j}\left[\frac{\partial \pi}{\partial(\nabla_j U^*)}\right] = 0,$$

which can be written succinctly using divergence (**div**) as

$$\alpha(\theta) := -[f(\theta) + \mathbf{div}\,\{f(\theta)(\theta - \nabla C(\nabla U^*))\}] = 0. \quad (3)$$

Intuitively, $\alpha(\theta)$ measures the marginal loss of the principal when the indirect utility (information rent) of type θ is increased marginally from U^* to U^*+h. Alternatively, define $\nu(\theta) := \frac{\partial S(\theta,\mathbf{q}(\theta))}{\partial \mathbf{q}}$ –the marginal distortion vector, then $\alpha(\theta) = 0$ is equivalent to $\mathbf{div}(\nu(\theta)) = -f(\theta)$, which is the optimal tradeoff between distortion and information rent. Let $L(h) = -\mathbb{E}\Pi'(U^*)h$ be the loss of the principal at U^* for the variation h. So if the principal increases U^* in the direction of some h then the seller's marginal loss can be expressed as

$$L(h) = \int_{\mathcal{S}_\theta} h(\theta)\alpha(\theta)\mathrm{d}\theta + \int_{\partial\mathcal{S}_\theta} h(\theta)\underbrace{(-\nu(\theta)\cdot\hat{n}(\theta))}_{=\beta(\theta)}\mathrm{d}\sigma(\theta) := \int_{\mathcal{S}_\theta} h(\theta)\mathrm{d}\mu(\theta), \quad (4)$$

where $\mathrm{d}\sigma(\theta)$ is the Lebesgue measure on the boundary $\partial\mathcal{S}_\theta$, $\hat{n}(\theta)$ is an outward normal and $\mathrm{d}\mu(\theta) := \alpha(\theta)\mathrm{d}\theta + \beta(\theta)\mathrm{d}\sigma(\theta)$. If we consider the types who participate, i.e. $U^*(\theta) \geq U_0(\theta)$, this marginal loss $L(h)$ must be zero. Since $h \geq 0$ it means $\mu(\theta) = 0$, so that both $\alpha(\theta)$ and $\beta(\theta) := -\nu(\theta)\cdot\hat{n}(\theta)$ must be equal to zero. For those who do not participate, it must mean the loss is positive, i.e. $L(h) > 0$; see Proposition 4 in Rochet-Choné for more.

Lemma 1. *The optimal indirect utility U^* is such that*

$$\forall \theta \in \mathcal{S}_\theta, \qquad \alpha(\theta) \begin{cases} > 0, & U^*(\theta) \leq U_0(\theta) \\ = 0, & U^*(\theta) > U_0(\theta) \end{cases}, \tag{5}$$

$$\text{and } \forall \theta \in \partial\mathcal{S}_\theta, \qquad \beta(\theta) \begin{cases} > 0, & U^*(\theta) \leq U_0(\theta) \\ = 0, & U^*(\theta) > U_0(\theta) \end{cases}. \tag{6}$$

The global incentive compatibility condition is important because it determines the optimal bunching (if any) in the equilibrium by requiring $(U^* - U_0)(\theta)$ be convex. This corresponds to determination of the subset $\mathcal{S}_\theta^1$ where the optimal allocation rule $\rho(\cdot)$ will be such that some types are allotted same quantity q. Let $\mathcal{S}_\theta^1(q)$ be the types that gets the same q, i.e. $\mathcal{S}_\theta^1(q) = \{\theta \in \Theta : \rho(\theta) = q\} = \{\theta \in \Theta : U^*(\theta) = \theta \cdot q - P(q)\}$. If $U^*(\theta)$ is convex for all θ, that is if the global incentive compatibility constraint is satisfied then there is no bunching, in which case $\mathcal{S}_\theta^1$ would be an empty set. In most of the cases, however, the convexity condition fails and hence there will be non-trivial bunching. So, U^* is affine on all the bunches, and the incentive compatibility constraint is binding for any two types θ', θ if and only if they both belong to $\mathcal{S}_\theta^1(\mathbf{q})$, i.e. if $\theta' \notin \mathcal{S}_\theta^1(\mathbf{q})$ but $\theta \in \mathcal{S}_\theta^1(\mathbf{q})$ then $U^*(\theta') > U^*(\theta) + (\theta - \theta')^T\mathbf{q}$.

Theorem 3.1. *Under the Assumptions 1-(i)–(iv-a) and (v) the optimal solution U^* to the problem is characterized by three subsets $\mathcal{S}_\theta^0, \mathcal{S}_\theta^1$ and $\mathcal{S}_\theta^2$ such that:*

1. *A positive mass of types $\mathcal{S}_\theta^0$ do not participate because $U^*(\theta) = U_0(\theta)$. This set is characterized by $\mu(\mathcal{S}_\theta^0) = 1$, i.e. $\int_{\mathcal{S}_\theta^0} \alpha(\theta)\mathrm{d}\theta + \int_{\partial\mathcal{S}_\theta^0} \beta(\theta)\mathrm{d}\theta = 1$.*
2. *$\mathcal{S}_\theta^1$ is a set of "medium types" known as the bunching region, which is further subdivided into subset $\mathcal{S}_\theta^1(\mathbf{q})$ such that all types in this subset get one type q, U^* is affine. μ restricted to $\mathcal{S}_\theta^1(\mathbf{q})$ satisfies: $\int_{\mathcal{S}_\theta^1(\mathbf{q})} \mathrm{d}\mu(\theta) = 0$ and $\int_{\mathcal{S}_\theta^1(\mathbf{q})} \theta\mathrm{d}\mu(\theta) = 0$.*
3. *$\mathcal{S}_\theta^2$ is the perfect screening region where U^* satisfies the Euler condition $\alpha(\theta) = 0$, or equivalently $\mathbf{div}(\nu(\theta)) = -f(\theta)$, for all $\theta \in \mathcal{S}_\theta^1 \cap \mathcal{S}_\theta$, and there is no distortion in the optimal allocation on the boundary, i.e. $\beta(\theta) = 0$ on $\mathcal{S}_\theta^1 \cap \partial\mathcal{S}_\theta$.*

In summary: the type space is (endogenously) divided into three parts: those who are excluded $\mathcal{S}_\theta^0$ and get the outside option q_0; those who are bunched $\mathcal{S}_\theta^1$ and are allocated some intermediate quality $\mathbf{q} \in Q_1$ such that all $\theta \in \mathcal{S}_\theta^1(\mathbf{q})$ get the same quantity q; and finally those who are perfectly screened $\mathcal{S}_\theta^2$ and are allocated some unique (customized) $\mathbf{q} \in Q_2$. An example shown in Fig. 1. It is also important to note that the allocation rule $\rho(\cdot)$ is continuous.

Corollary 1. *$\rho(\cdot)$ is continuous for all and $\frac{\partial \rho(\theta_j,\theta_{-j})}{\partial \theta_j} > 0, \forall(\theta_j, \theta_{-j}) \in \Theta_2$.*

Proof. For $\theta \in \Theta_2$, since $D^2U^*(\theta) > 0$ and because $\rho(\theta) = \nabla U^*(\theta)$ it is also continuous. Likewise, for all $\theta \in \Theta_0, \rho(\theta) = \mathbf{q}_0$ and hence continuous. Similar arguments show that $\rho(\theta)$ is continuous for all $\theta \in \mathcal{S}_\theta^1$. □

4. Identification

In this section we study identification of the distribution of types $F_\theta(\cdot)$ and the cost function $C(\cdot)$ under the Assumption 1-(iv-a), (linear utility) from the observables that include the triplet $\{X_i, \mathbf{q}_i, p_i\}$ for each agent $i \in [N] := \{1, \ldots, N\}$.[5] I use the lower case p_i to refer to the price paid by consumer i and the upper case $P(\cdot)$ to denote the pricing function. On the principal side, we observe the menu of contracts $\{Q_X, P_X(\cdot)\}$ offered to each consumer with characteristics X. I assume that these observables are distributed *i.i.d* with respect to $\Psi_{p,\mathbf{q},X,Z}(\cdot,\cdot,\cdot,\cdot) = \Psi_{p,\mathbf{q}|X}(\cdot,\cdot|\cdot) \times \psi_X(\cdot)$, where $\psi_X(\cdot)$ is identified (or estimated) from the data, and hence known.

The seller offers $(Q_{(x)}, P_{(x)}(\cdot))$ to agent i with observed characteristics $X_i = x \sim \Psi_X(\cdot)$. I use the upper case to denote the random variable and lower case to denote a realization of the random variable. Each agent i draws $\theta_i \sim F_\theta(\cdot)$ and selects $\mathbf{q}_i \in Q_{(x)}$ and pays p_i, so as to maximize the net utility. The seller chooses the menu optimally, which from **revelation principal** is equivalent to saying that there exists a direct mechanism, a unique pair of allocation rule $\rho(\cdot) : \mathcal{S}_\theta \mapsto Q_{(x,z)}$ and pricing function $P_x(\cdot) : Q_{(x)} \mapsto \mathbb{R}_+$, such that $\mathbf{q}_i = \rho(\theta_i)$ and $p_i = P_x(\rho(\theta_i))$. Henceforth, $\rho(\cdot)$ will stand for optimal allocation rule. Thus assuming that : a) consumers have private information about θ; b) the seller only knows the $F_\theta(\cdot)$ and $C(\cdot)$, and designs a $\{Q, P(\cdot)\}$ to maximize profit; and c) consumers optimize, leads to the following model:

$$\begin{aligned} p_i &= P[\mathbf{q}_i, F_\theta(\cdot), C(\cdot); X] \\ \mathbf{q}_i &= \rho[\theta_i, F_\theta(\cdot), C(\cdot); X], \quad i \in [N], k = 1, 2. \end{aligned} \tag{7}$$

The model parameters $[F_\theta(\cdot), C(\cdot)]$ are identified if for any different parameters $[\tilde{F}_\theta(\cdot), \tilde{C}(\cdot)]$, the implied data distributions are also different: $\Psi_{p,\mathbf{q},X}(\cdot,\cdot,\cdot) \neq$

[5] The focus on linear utility is simply to streamline notations. Extending it to bilinear or nonlinear utility in the definition of identification is straightforward.

$\tilde{\Psi}_{p,\mathbf{q},X}(\cdot,\cdot,\cdot)$. The model has unique equilibrium, so for every model parameter there is unique $\Psi_{p,\mathbf{q},X}(\cdot,\cdot,\cdot)$, identification boils down to finding conditions under which Equation (7) are invertible. In particular, I am interested in determining some low-level assumptions under which we achieve global identification, which is distinct from the notion of local identification as articulated by Rothenberg (1971) and extended to nonparametric settings by Chen, Chernozhukov, Lee, and Newey (2014). See also Carrasco, Florens, and Renault (2007).

Following the equilibrium characterization, I consider the three subsets of types separately. For every $X = z$ let $Q^j_{(x)}$ be the set of choices made by consumers with type $\theta \in \mathcal{S}^j_\theta$ for $j = 0, 1, 2$, respectively. Since $\mathbf{q}(\cdot)$ is continuous (Corollary 1), these sets are well defined. In what follows, I will use data from $Q^k_{(x)}$ to identify the model parameters restricted to $\mathcal{S}^j_\theta$, beginning with the subset $Q^2_{(x)}$. The allocation rule $\rho(\cdot)$ is one-to-one when restricted to $\mathcal{S}^2_\theta$, and hence its inverse $\rho^{-1}(\cdot)$ exists on $Q^2_{(x)}$, but not when restricted to $\mathcal{S}^1_\theta$ because of bunching, and as a consequence the identification strategies are different.[6] In what follows, I suppress the dependence on X, until it is relevant.

Let $M(\cdot)$ and $m(\cdot)$ be the distribution and density of $\mathbf{q}$, respectively. Since the equilibrium indirect utility function U^* is unique, it implies that there is a unique distribution $M(\cdot)$ that corresponds to the model structure $[F_\theta(\cdot), C(\cdot)]$. Thus the structure is said to be identified if for given $m(\cdot)$ there exists a (unique) pair $[F_\theta(\cdot), C(\cdot)]$ that satisfies Equations (7). Let $\tilde{\theta}(\cdot) : Q \longrightarrow \mathcal{S}^2_\theta$ be the inverse of $\rho(\cdot)$ when restricted on Q^2, i.e. $\forall \mathbf{q} \in Q^2, \tilde{\theta}(\mathbf{q}) = \rho^{-1}(\mathbf{q})$. Similarly, let $M^*(\mathbf{q})$ and resp. $m^*(\mathbf{q})$ be the truncated distribution (resp. density) of $\mathbf{q} \in Q^2_{(x)}$, defined as:

$$\begin{aligned} M^*(\mathbf{q}) &:= \Pr(\tilde{\mathbf{q}} < \mathbf{q} | \mathbf{q} \in Q_2) = \Pr(\theta < \tilde{\theta}(\mathbf{q}) | \theta \in \mathcal{S}^2_\theta) \\ &= \int_{\mathcal{S}^2_\theta} \mathbb{1}\{\theta < \tilde{\theta}(\mathbf{q})\} f_\theta(\theta | \theta \in \mathcal{S}^2_\theta) \mathrm{d}\theta; \\ m^*(\mathbf{q}) &:= \frac{m(\mathbf{q})}{\int_{Q_2} m(\tilde{\mathbf{q}}) \mathrm{d}\tilde{\mathbf{q}}} = \frac{f(\tilde{\theta}(\mathbf{q}))}{\int_{Q_2} f(\tilde{\theta}(\tilde{\mathbf{q}})) \mathrm{d}\tilde{\mathbf{q}}} |det(D\tilde{\theta})(\mathbf{q})|, \end{aligned} \tag{8}$$

where $det(\cdot)$ is the determinant function. Then the bijection between high-type and high-qualities gives

$$M^*(\mathbf{q}) = \Pr(\rho(\theta, F_\theta, C) \leq \mathbf{q} | \mathbf{q} \in Q^2) = F_\theta \circ \rho^{-1}(\mathbf{q} | \mathbf{q} \in Q^2), \tag{9}$$

[6] A fundamental problem in Analysis is the existence and/or uniqueness of the solutions to the equation $\mathbf{q} = \rho(\theta)$ in the unknown θ. The existence of inverse $\rho^{-1}(\cdot)$ follows from the observation that $U^*(\theta)$ is convex on $\mathcal{S}^2_\theta$ and $DU^*(\theta) = \rho(\theta)$, see Kachurovksii (1960). Also see Parthasarathy (1983); Fujimoto and Herrero (2000) on global univalence.

which will be a key relationship for identification. Before moving on, I introduce new short-hand notations. Let $F_\theta(\cdot|j)$ be the CDF $F_\theta(\cdot)$ restricted to be in the set $\mathcal{S}_\theta^j$ and let N_j be the set of consumers who buy $\mathbf{q} \in Q^j$, for $j = 0, 1, 2$.

4.1. **Linear Utility.** I begin by showing that without any further restrictions $F_\theta(\cdot|2)$ can be identified and $C(\cdot)$ can be identified on Q^2. When the utility function is linear, for the high-types, the marginal utility θ is equal to the marginal prices, $\nabla P(\cdot)$, which is the gradient of price. Therefore the type that chooses $\mathbf{q} \in Q^2$ must satisfy the equality $\nabla P(\mathbf{q}) = \theta = \tilde{\theta}(\mathbf{q})$, thereby identifying $\theta_i = \tilde{\theta}(\mathbf{q}_i)$ for all $i \in [N_2]$. This identification is purely from the optimality of the demand side and the strong linear functional form assumption on the utility function. We lose this identification when the utility function is nonlinear (Subsection 4.3). As $\tilde{\theta}(\cdot)$ restricted to $\mathcal{S}_\theta^2$ is bijective we can identify

$$F_\theta(\xi|2) = \Pr(\mathbf{q} \le (\nabla P)^{-1}(\xi)|Q \in Q^2) = M^*((\nabla P)^{-1}(\xi)).$$

Next, I consider identification of the cost function. Recall that the equilibrium allocation condition (3) is $\alpha(\theta) = 0$, or

$$\mathbf{div}\,\{f_\theta(\theta)(\theta - \nabla C(\nabla U^*))\} = -f_\theta(\theta).$$

If we divide both sides by $\int_{\mathcal{S}_\theta^2} f_\theta(t)\mathrm{d}t$ we get

$$\begin{aligned}
\mathbf{div}\left\{\frac{f_\theta(\theta)}{\int_{\mathcal{S}_\theta^2} f_\theta(t)\mathrm{d}t}(\theta - \nabla C(\nabla U^*))\right\} &= -\frac{f_\theta(\theta)}{\int_{\mathcal{S}_\theta^2} f_\theta(t)\mathrm{d}t}\\
\mathbf{div}\left\{\frac{m^*(\mathbf{q})}{|det(D\tilde{\theta})(\mathbf{q})|}(\tilde{\theta}(\mathbf{q}) - \nabla C(\mathbf{q}))\right\} &= -\frac{m^*(\mathbf{q})}{|det(D\tilde{\theta})(\mathbf{q})|}\\
\mathbf{div}\left\{\frac{m^*(\mathbf{q})}{|det(D\nabla P(\mathbf{q}))|}(\nabla P(\mathbf{q}) - \nabla C(\mathbf{q}))\right\} &= -\frac{m^*(\mathbf{q})}{|det(D\nabla P(\mathbf{q}))|},
\end{aligned}$$

where the second equality follows from Equation (8) and the last equality follows from the definition of the curvature of the pricing function, i.e. $D\nabla P(\mathbf{q}) = D\tilde{\theta}(\mathbf{q})$. This means the cost function $C(\cdot)$ is the solution to the partial differential equation (PDE) with the following boundary condition (that follows from $\beta(\theta) = 0$ on ∂Q^2):

$$\frac{m^*(\mathbf{q})}{|det(D\nabla P(\mathbf{q}))|}(\nabla C(\mathbf{q}) - \nabla P(\mathbf{q})) \cdot \overrightarrow{n}(\nabla P(\mathbf{q})) = 0.$$

This PDE has a unique solution $C(\mathbf{q})$, see Evans (2010), and hence $C(\cdot)$ is identified on the convex set Q^2. To extend the function to the entire domain we need:

Assumption 2. *Cost function $C : Q \to \mathbb{R}$ is a real analytic function at $\mathbf{q} \in Q$.*

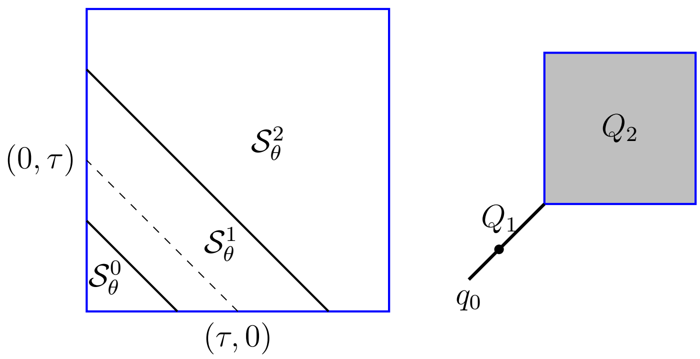

FIGURE 1. Optimal Product Line, from Rochet-Choné .

This assumption about $C(\cdot)$ being analytic is a technical assumption that assumes $C(\cdot)$ is infinitely differentiable and can be expressed (uniquely) as a Taylor series. Hence, it allows for any convex polynomial, trigonometric and exponential functions. Once the cost function is identified on an open convex set Q^2, analytic extension theorem implies that the function has a unique extension to the entire domain Q. Since the cost function is completely unspecified, besides convexity, the fact that we need analyticity is not surprising. Similar idea has been used in the previous literature. In the problem of instrumental variable estimation of nonparametric model, Newey and Powell (2003) restrict the conditional density to be of the exponential family, which has a unique extension property similar to analyticity; and to random coefficient in a logit model Fox, il Kim, Ryan, and Bajari (2012) and multinomial choice model Fox and Gandhi (2013), like here, assume the utility function is real analytic. This result is formalized below.

Theorem 4.1. *Under the Assumptions 1-(i)–(iv-a), (v), and 2, the model structure* $[F_\theta(\cdot|2), C(\cdot)]$ *is nonparametrically identified.*

It is clear that the monotonicity of $\rho(\cdot)$ is the key to identification, and since we lose monotonicity on $\mathcal{S}_\theta^1$ we lose identification, as shown in the example below which is taken from Rochet-Choné .

Example 4.1. Let $J = 2$ and the cost function be $C(\mathbf{q}) = c/2(q_1^2 + q_2^2)$ and types are independent and uniformly distributed on $\mathcal{S}_\theta = [0,1]^2$ and $\mathbf{q}_0 = 0$ and $P_0 = 0$. Then, the optimal indirect utility function U^* has different shapes in the three regions: (i) in the non participation region $\mathcal{S}_\theta^0, U^*(\theta) = 0$; (ii) in the bunching region $\mathcal{S}_\theta^1$, U^* depends only on $\theta_1 + \theta_2$; and (iii) in the perfect screening region $\mathcal{S}_\theta^2$, U^* is strictly convex.

On $\mathcal{S}_\theta^0, \rho(\theta) = 0$, which means $\alpha(\theta) = \mathbf{div}(\theta f(\theta)) + f(\theta) = 3$ and $\beta(\theta) = a$ on $\partial\mathcal{S}_\theta^0$. The boundary that separates $\mathcal{S}_\theta^0$ and $\mathcal{S}_\theta^1$ is a linear line $\tau_0 = \theta_1 + \theta_2$, where $\tau_0 = \frac{\sqrt{6}}{3}$. On $\mathcal{S}_\theta^1, \rho(\theta) = (\rho_1(\theta), \rho_2(\theta)) = (\rho_b(\tau), \rho_b(\tau))$, with $\theta_1 + \theta_2 = \tau$. In other words, all consumers with type $\theta_1 + \theta_2 = \tau$ are treated the same and they get the same $\rho_1(\tau) = \rho_2(\tau) = \rho_b(\tau)$. So $\alpha(\theta) = 3 - 2cq_b'(\tau)$ and on $\partial\mathcal{S}_\theta^1, \beta(\theta) = (c\rho(\theta) - \theta) \cdot \hat{n}(\theta) = -c\rho_b(\tau)$. Sweeping conditions are satisfied if $\alpha(\theta) \geq 0$ and $\beta(\theta) \geq 0$ and on each bunch

$$\int_0^\tau \alpha(\theta_1, \tau - \theta_1)\mathrm{d}\theta_1 + \beta(0, \tau) + \beta(\tau, 0) = 0,$$

which can be used to solve for q_b as $\rho_b(\tau) = \frac{3\tau}{4c} - \frac{1}{2c\tau}$. Then $\mathcal{S}_\theta^1 = \{\theta : \tau_0 \leq \theta_1 + \theta_2 \leq \tau_1\}$ where τ_1 is determined by the continuity condition on $\mathcal{S}_\theta$ of $\rho(\cdot)$, i.e. $\rho_b(\tau_1) = 0$. Now, define $\tau = \rho_b^{-1}(\mathbf{q})$ as the inverse of the optimal (bunching) mechanism. Then identification is to determine the joint cdf of (θ_1, θ_2) from that of $\tau = \theta_1 + \theta_2$, which is not possible.

To summarize: the seller divides the agents into three categories and perfectly screens only the top ones. We can then use the distribution of their choices to determine their types and the cost function. To understand the welfare consequence of asymmetric information we might also want to understand the heterogeneity in preference of those in the medium categories that are not perfectly screened but they are not excluded from the market either. The example above shows that if we restrict the utility function to be linear and independent of the consumer characteristics then because the bunching is also linear we cannot identify the types.

This brings me to the next question. If the utility is also a function of observed characteristics X, then can we use the variation in those observed characteristics to identify the medium-types, the types that are bunched? In the following subsection I show that the answer is positive. Under the Assumption 1-(iv-b) that the utility is bilinear, if the observed characteristics X are (statistically) independent of the type θ and if the dimension of X is the same as the dimension of θ then we can identify $f_\theta(\cdot|1)$.

4.2. **Bilinear Utility.** In this subsection I assume that the base utility function satisfies Assumption 1-(iv-b) and X_1 is independent of θ.

Assumption 3. *The observed characteristics $X = (X_1, X_2)$ and θ is mutually independent, i.e., $X_1 \perp\!\!\!\perp X_2, X_1 \perp\!\!\!\perp \theta$ and $X_2 \perp\!\!\!\perp \theta$.*

In particular, suppose that the net utility of choosing $\mathbf{q}$ by an agent with characteristics X and unobserved θ is

$$V(\mathbf{q};\theta,X)=\sum_{j\in[J]}\theta_j X_{1j}\mathbf{q}_j-P(\mathbf{q}). \tag{10}$$

Since the utility function has changed, the optimal contracts will also change. However, once we note that we can change the measurement unit from $\mathbf{q}$ to $\tilde{\mathbf{q}}=X_1\cdot\mathbf{q}$, it is straightforward to see that the general characterization does not change. determining optimal contract is the same as before. Alternatively, the seller can condition on X_1 and choose the product line and prices appropriately and because X_1 and θ are independent nothing changes. This means the identification result from Theorem 4.1 is still applicable, because we can simply ignore the variation in X_1 and directly apply the theorem. If we exploit the variation in X_1, however, we have more information than necessary.

In the remaining of this subsection I will consider identifying the type density when restricted to the bunching region $\mathcal{S}^1_\theta$, denoted (after abuse of notation) as $f_\theta(\cdot)$. In the example above we saw that all agents with type such that $\tau=\sum_{j\in[J]}\theta$ selected the same $\mathbf{q}(\tau)$. Now, that the agents vary in X, agents are bunched according to $W=\sum_{j\in J}\theta_j X_{1j}$, in other words, all agents with the same W self select $\rho(W)$, i.e. $\rho(\theta)=(\rho_1(\theta),\ldots,\rho_J(\theta))=(\rho_1(W),\ldots,\rho_J(W))$ for all $\theta\in\mathcal{S}^1_\theta$. In other words, W acts as a sufficient statistics, and incentive compatibility requires that $\mathbf{q}(W)$ be monotonic in W and hence invertible. So from the observed $\mathbf{q}$ we can determine the index $W:=\rho^{-1}(\mathbf{q})$. Then, the identification problem is to recover $f_\theta(\cdot)$ from the the joint density $f_{W,X_1}(\cdot,\cdot)$ of (W,X_1) when

$$W=\theta_1X_{1_1}+\cdots+\theta_JX_{1_J}.$$

I begin by normalizing the equation above by multiplying both sides by $||X||^{-1}$. Let $D:=||X_1||^{-1}X_1\in\mathbb{S}_{J-1}$, and $B:=||X_1||^{-1}W\in\mathbb{R}$ where $\mathbb{S}_{J-1}=\{\omega\in\mathbb{R}^J:||\omega||=1\}$ is a $J-$ dimensional unit sphere, so that $B=\theta\cdot D$. Then the conditional density of B given D is

$$f_{B|D}(b|d)=\int_{\mathcal{S}^1_\theta}f_{B|D,\theta}(b|d,\theta)f_\theta(\theta)d\theta=\int_{\{\theta\cdot d=b\}}f_\theta(\theta)\mathrm{d}\sigma(\theta):=Rf_\theta(b,d),$$

where $Rf_\theta(b,d)$ stands for the Radon transform, see Helgason (1999), of $f_\theta(\cdot)$. So to identify $f_\theta(\cdot)$ we must show that $Rf_\theta(\cdot,\cdot)$ is invertible, for which we need sufficient variation in X. Suppose not, and suppose X is a vector of constants $(a_1,\ldots,a_J)$. Then we cannot identify $f_\theta(\cdot)$ from $B=a_1\theta_1+\cdots+a_J\theta_J$.

Let

$$Ch_{Rf}(\xi) := \int_{-\infty}^{\infty} e^{-2\pi \mathbf{i} b\xi} Rf(d,b)\mathrm{d}b,$$

be the Fourier transform of $Rf(d,b)$ that can be identified from $f_{B,D}(\cdot,\cdot)$, and let

$$Ch_f(\xi d) := \int_{-\infty}^{\infty} e^{-2\pi \mathbf{i}(\theta\cdot\xi d)} f(\theta)\mathrm{d}\theta$$

be the Fourier transform of $f_\theta(\cdot)$ evaluated at ξd, which we do not know. However, the Projection slice theorem implies that these two functions are the same for a fixed d, i.e. $Ch_f(\xi d) = Ch_{Rf}(\xi)$, and hence $f_\theta(\cdot)$ can be identified as the Fourier inverse:

$$f_\theta(\theta) = \int_{-\infty}^{\infty} e^{2\pi \mathbf{i}\theta\cdot\xi} \psi_{Rf}(\xi)\mathrm{d}\xi.$$

Theorem 4.2. *Under Assumptions 1-(i)–(iv-b), (v), and 2 and 3 the densities $f_\theta(\cdot|1)$ and $f_\theta(\cdot|2)$ and the cost function $C(\cdot)$ are nonparametrically identified.*

Intuitively, the identification exploits the fact that two consumers with same θ but different X_1 will face different menus and different choices. So if we consider the population with $X = x$, the variation in the choices must be due to the variation in θ. But as we change X_1 from x_1 to x_1', the choices change but variation in θ remains the same because $X_1 \perp\!\!\!\perp \theta$. So with continuous variation in X_1, we have infinitely many moment condition for θ, which leads to the (mixture) Radon transform of the $f_\theta(\cdot|1)$. This shows that even when bijectivity of equilibrium fails, we might be able to use variation in consumer socioeconomic and demographic characteristics X_1 for identification. Since the joint density of types in $\mathcal{S}_\theta^2$ (who were perfectly screened) was identified even without X_1 this result suggests that the model is over identified, which can then be used for specification testing. Even though this intuition is correct, we will postpone the discussion of over identification until the next subsection when I consider nonlinear utility function. I will show that when utility is nonlinear and if we have access to discrete cost shifter then to identify the model it is sufficient that the cost shifter causes the gradient of pricing functions to intersect.

Note: So far I have implicitly assumed that we can divide the observed choices $\{\mathbf{q}_i\}$ into three subsets. We know the outside option $Q^0 = \{\mathbf{q}_0\}$, so the only thing left is to determine the bunching set Q^1. As seen in the Figure 1, the product line Q^1 is congruent to one dimensional $\mathbb{R}_+$, which is the main characteristic of bunching. In higher dimension, the set Q^1 will consist of all products that is congruent with the positive real of lower than J dimension.

4.3. **Nonlinear Utility.** In this section I consider the model with nonlinear utility (Assumption 1-(iv-c)) that is the (gross) utility function is equal to $X_1 \cdot \mathbf{v}(\mathbf{q}; X_2)$. To keep the arguments clear, I will ignore X, which is tantamount to assuming that $d_x = 0$, and focus on the identification of $\mathbf{v}(\mathbf{q})$ on $\mathcal{S}^2_\theta$. Once we have understood the what variation in the data drives identification, we can introduce X and consider the possibility of over-identification.

I begin by first showing that the model $[F_\theta(\cdot|2), C(\cdot), \mathbf{v}(\cdot)]$ cannot be identified because the two optimality conditions Equations (1) and (9) are insufficient. Identification fails, because of the substitutability between the type θ and the curvature of the utility function $\mathbf{v}(\cdot)$ as shown below.

Lemma 2. *Under Assumptions 1-(i)–(iv-c) and (v) the model* $\{F_\theta(\cdot|2), C(\cdot), \mathbf{v}(\cdot)\}$, *where the domain of the cost and utility functions are restricted to be* Q^2 *and* $\mathcal{S}^2_\theta$, *respectively, are not identified.*

Proof. Since the optimality condition (3) is used to determine the cost function, we can treat the cost function as known. I will suppress the dependence on X_2 and let $J = 2$, so $V(\mathbf{q}; \theta) = \theta_1 v_1(q_1) + \theta_2 v_2(q_2) - P(q_1, q_2)$. Let the utility function be $v_j(q_j) = q_j^{\omega_j}, \omega_j \in (0, 1)$, and the distribution be $F_\theta(\cdot, \cdot|2)$ and density be $f_\theta(\cdot|2)$. Observed $\{q_j, p_j\}$ solve the first order condition

$$\theta_j \omega_j q_j^{\omega_j - 1} = \frac{\partial P(q_1, q_2)}{\partial q_j} = p_j, \quad j = 1, 2.$$

Using the change of variable, the joint (truncated) density of (q_1, q_j) is

$$m^*_{\mathbf{q}}(q_1, q_2) = f_\theta \left(\frac{P_1}{\omega_1 q_1^{\omega_1 - 1}}, \frac{P_2}{\omega_2 q_2^{\omega_2 - 1}} |2 \right) \frac{P_1 P_2 (1 - \omega_1)(1 - \omega_2)}{\omega_1 \omega_2 q_1^{\omega_1} q_2^{\omega_2}}.$$

Let $\check{\theta}_j \equiv \theta_j \times \omega_j \sim F_{\check{\theta}}(\cdot|2)$, where $F_{\check{\theta}}(\cdot|2) = F_\theta(\cdot/\omega|2)$ with $\omega \equiv (\omega_1, \omega_2)$ and $\check{\mathbf{v}}(q_j) = \mathbf{v}(q_j)/\omega = q_j^{\omega_j}/\omega_j$, be a new model. It is easy to check that $\{q_j, p_j\}$ solves the first-order condition implied by $[\check{\mathbf{v}}(\cdot), F_{\check{\theta}}(\cdot)]$, and the joint (truncated) density of (q_1, q_2) is

$$\begin{aligned} \check{m}^*_{\mathbf{q}}(q_1, q_2) &= f_{\check{\theta}} \left(\frac{p_1}{q_1^{\omega_1 - 1}}, \frac{p_2}{q_2^{\omega_2 - 1}} \middle| 2 \right) \frac{p_1 p_2 (1 - \omega_1)(1 - \omega_2)}{q_1^{\omega_1} q_2^{\omega_2}} \\ &= f_\theta \left(\frac{p_1}{\omega_1 q_1^{\omega_1 - 1}}, \frac{p_2}{\omega_2 q_2^{\omega_2 - 1}} \middle| 2 \right) \frac{p_1 p_2 (1 - \omega_1)(1 - \omega_2)}{\omega_1 \omega_2 q_1^{\omega_1} q_2^{\omega_2}} = m^*_{\mathbf{q}}(q_1, q_2). \end{aligned}$$

□

As we have seen here, we can increase the type to $\theta \cdot \omega$ and decrease the curvature of utility to $\mathbf{v}(\mathbf{q})/\omega$ without changing the observable choices. Therefore, data from only one market is not enough for identification. To break this substitutability we need an exogenous factor, like an exogenous cost shifter, call it Z that only affects the cost, now denoted as $C(\cdot; Z)$ and neither the utility function nor the type, i.e. $Z \perp\!\!\!\perp (\theta, \mathbf{v}(\cdot))$. For such a shifter to have identification power it must not only change cost but also change the relative prices – increase or decreasing the prices uniformly is insufficient. Such a cost shifter could be either in the form of some exogenous changes in law that affects prices over two period, or different tax or marketing expenses across two independent markets. Either way, it only suffices that the cost shifter takes two values, i.e. $\mathcal{S}_Z = \{z_1, z_2\}$. This exclusion restriction implies that at different values of the cost shifter: a), the ratio of the types will be equal to the ratio of the slope of the prices at different values of the cost shifter; and b) the (multivariate) quantiles of choices by the high-types are the same.

Lets consider the non-identification example. Henceforth, I will use the subscript $\ell \in \{1, 2\}$ in $\{P_\ell(\cdot), \rho_\ell(\cdot)\}$ to denote the price function and allocation rule when $Z = z_\ell$. As in Lemma 2, the utility function is $\mathbf{v}(q_1, q_2) = \begin{pmatrix} v_1(q_1) \\ v_2(q_2) \end{pmatrix} = \begin{pmatrix} (q_1)^{\omega_1} \\ (q_2)^{\omega_2} \end{pmatrix}$. As before let us focus only on the high-types $\mathcal{S}_\theta^2$ and let's further assume that Q^2 is also invariant to Z.

Then the demand side optimality implies that the marginal utility equals the marginal price can be written as

$$\begin{pmatrix} \nabla_1 P_\ell(\mathbf{q}) \\ \nabla_2 P_\ell(\mathbf{q}) \end{pmatrix} = \begin{pmatrix} \tilde{\theta}_{\ell 1}(\mathbf{q}) \cdot v_1'(q_1) \\ \tilde{\theta}_{\ell 2}(\mathbf{q}) \cdot v_2'(q_2) \end{pmatrix} = \begin{pmatrix} \tilde{\theta}_{\ell 1}(\mathbf{q}) \cdot \omega_1 (q_1)^{\omega_1 - 1} \\ \tilde{\theta}_{\ell 2}(\mathbf{q}) \cdot \omega_2 (q_2)^{\omega_2 - 1} \end{pmatrix}, \ell = 1, 2.$$

Solving for $\nabla v(q_j)$ for $\ell = 1, 2$ and equating the two gives

$$\begin{pmatrix} \tilde{\theta}_{11}(\mathbf{q})/\tilde{\theta}_{21}(\mathbf{q}) \\ \tilde{\theta}_{21}(\mathbf{q})/\tilde{\theta}_{22}(\mathbf{q}) \end{pmatrix} = \begin{pmatrix} \nabla_1 P_1(\mathbf{q})/\nabla_1 P_2(\mathbf{q}) \\ \nabla_2 P_1(\mathbf{q})/\nabla_2 P_2(\mathbf{q}) \end{pmatrix},$$

i.e., the ratio of types should equal the ratio of marginal prices, or equivalently

$$\begin{pmatrix} \tilde{\theta}_{11}(\mathbf{q}) \\ \tilde{\theta}_{21}(\mathbf{q}) \end{pmatrix} = \begin{pmatrix} \nabla_1 P_1(\mathbf{q})/\nabla_1 P_2(\mathbf{q}) \cdot \tilde{\theta}_{21}(\mathbf{q}) \\ \nabla_2 P_1(\mathbf{q})/\nabla_2 P_2(\mathbf{q}) \cdot \tilde{\theta}_{22}(\mathbf{q}) \end{pmatrix}. \tag{11}$$

Equation (11) captures the fact that a consumer who pays higher marginal price for a $\mathbf{q}$ when $Z = z_1$ than when $Z = z_2$, then she must have higher type $\tilde{\theta}_1(\mathbf{q})$ than $\tilde{\theta}_2(\mathbf{q})$. So, if we know θ's choice $\mathbf{q} = \mathbf{q}_1(\theta)$ when $Z = z_1$ then we can use the curvature of the pricing functions to determine the θ that chooses the same bundle $\mathbf{q}$ when $Z = z_2$.

Now, consider the supply side. The allocation rule for the high-types is monotonic (IC constraint) so we know:

$$\begin{aligned} F_\theta(t|2) = F_\theta(t_1, t_2|2) &= \Pr(\theta_1 \leq t_1, \theta_2 \leq t_2 | \mathcal{S}_\theta^2) = \Pr(\rho(\theta, z_\ell) \leq \rho(t, z_\ell)|Q^2) \\ &= \Pr(\mathbf{q} \leq \rho(t, z_\ell)) = \Pr(q_1 \leq \rho_1(t, z_\ell), q_2 \leq \rho_2(t, z_\ell)|Q^2) \\ &= M_\ell^*(\rho_\ell(t)), \ell = 1, 2, \end{aligned}$$

where the third equality follows from monotonicity of $\rho(\cdot, Z)$ and exogeneity of Z. This relationship is independent of Z, which gives the following equality

$$M_1^*(\rho_1(t)) = M_2^*(\rho_2(t)).$$

Hence, the (multivariate) quantiles of the choice distribution when $Z = z_1$ are equal to those when $Z = z_2$, i.e.[7]

$$\rho_1(t) = (M_1^*)^{-1}[M_2^*(\rho_2(t))], \tag{12}$$

and since $(M_1^*)^{-1} \circ M_2^*(\cdot)$ is identified, we can identify $\rho_2(\theta)$ if we know $\rho_2(\theta)$. Therefore, the difference, $((M_1^*)^{-1} \circ M_2^*(\rho(\tau)) - \rho(\tau))$, measures the change in $\mathbf{q}$ when Z moves from z_2 to z_1, while fixing the quantile of $\mathbf{q}$ at τ. This variation (12) together with (11) can be used to first identify $\tilde{\theta}(\cdot)$ and then $\nabla \mathbf{v}(\mathbf{q})$ as a (vector valued) function that solves $\nabla P(\mathbf{q}) = \tilde{\theta}(\mathbf{q}) \circ \nabla \mathbf{v}(\mathbf{q})$.

The intuition behind identification is as follows: Start with a normalization $\theta^0 \equiv \tilde{\theta}_2(\mathbf{q}^0)$ for some bundle $\mathbf{q}^0 = (q_1^0, q_2^0) \in Q^2$, and determine $\nabla P_1(\mathbf{q}^0), \nabla P_2(\mathbf{q}^0)$, the quantile $\tau = M_2^*(\mathbf{q}^0)$, and $\theta^1 \equiv \tilde{\theta}_1(\mathbf{q}^0)$ from (11).[8] Using (12) determine $\mathbf{q}^1$ with the same quantile τ under $Z = z_1$. Then, for $\mathbf{q}^1$ determine $\nabla P_1(\mathbf{q}^1)$ and $\nabla P_2(\mathbf{q}^1)$, which can determine $\theta^2 = \tilde{\theta}_2(\mathbf{q}^1) = \nabla P_2(\mathbf{q}^1) \circ (\nabla P_1(\mathbf{q}^1))^{-1} \circ \theta^1$ (inverse of (11)). Then iterating these steps we can identify a sequence $\{\theta^0, \theta^1, \ldots, \theta^L, \ldots\}$ and the corresponding quantile. If these sequence form a dense subset of Q^2 then the function $\tilde{\theta}(\cdot) : Q^2 \times \mathcal{S}_Z \to \mathcal{S}_\theta^2$ is identified everywhere. I formalize this intuition for $J \geq 2$ below, starting with the assumption about exclusion restriction.

[7] At this stage, it is worth pointing out that unlike with one dimensional random variable determining multivariate quantiles is not straight forward because of the lack of natural order. But to keep the discussion simple I defer the discussion until later.

[8] Here, the superscript is an index of the sequence of bundles, and should not be confused with the utility function $v_j(q_j) = (q_j)^{\omega_j}$, similarly for the superscript on θ.

Assumption 4. *Let $Z \in \mathcal{S}_Z = \{z_1, z_2\}$ be independent of θ and $\mathbf{v}(\mathbf{q})$.*

As before, consumer optimality implies $\nabla P_\ell(\mathbf{q}) = \tilde{\theta}_\ell(\mathbf{q}) \circ \nabla \mathbf{v}(\mathbf{q})$, and the general version of Equation (11) can be written as

$$\begin{aligned} \tilde{\theta}_\ell(\mathbf{q}) &= \nabla P_\ell(\mathbf{q}) \circ \tilde{\theta}_{\ell'}(\mathbf{q}) \circ (\nabla P_{\ell'}(\mathbf{q}))^{-1} \\ &\equiv r_{\ell',\ell}(\tilde{\theta}_{\ell'}(\mathbf{q}), \mathbf{q}) = \begin{pmatrix} r^1_{\ell',\ell}(\tilde{\theta}_{\ell'}(\mathbf{q}), \mathbf{q}) \\ \vdots \\ r^J_{\ell',\ell}(\tilde{\theta}_{\ell'}(\mathbf{q}), \mathbf{q}) \end{pmatrix}. \end{aligned} \tag{13}$$

Next, Assumption 4 and the incentive compatibility condition for high types imply $F_\theta(t|2) = M^*(\mathbf{q}(t; z_\ell); z_\ell), \ell = 1, 2$ and hence

$$M^*_\ell(\rho_\ell(t)) := M^*(\rho(t; z_\ell); z_\ell) = M^*(\rho(t; z_{\ell'}); z_{\ell'}) := M^*_{\ell'}(\rho_{\ell'}(t)). \tag{14}$$

Once we determine multivariate quantiles, (14) generalizes (12). Quantiles are the proper inverse of a distribution function, but defining multivariate quantiles is not straightforward because of the lack of a natural order in $\mathbb{R}^J, J \geq 2$. One way around this problem is to choose an order (or a rank) function, and define the quantiles with respect to that order. Even though there are numerous ways to define such order, I follow Koltchinskii (1997). He shows that if we choose a continuously differentiable convex function $g_M(\cdot)$, then we can define the quantile function as the inverse of some transformation (see Appendix A) of $g_M(\cdot)$, denoted as $(\partial g_M)^{-1}(\tau) \in \mathbb{R}^J$ for quantile $\tau \in [0, 1]$. For this procedure to make sense, it must be the case that, conditional on the choice of $g_M(\cdot)$, there is a one to one mapping between the quantile function and the joint distribution. In fact Koltchinskii (1997) shows that for any two distributions $M_1(\cdot)$ and $M_2(\cdot)$, the coresponding quantile functions are equal, $(\partial g_{M_1})^{-1}(\cdot) = (\partial g_{M_2})^{-1}(\cdot)$, if and only if $M_1(\cdot) = M_2(\cdot)$. Henceforth, I assume that such a function $g_M(\cdot)$ is chosen and fixed, then (14) and (20) imply

$$\rho_1(\tau) = (\partial g_{M^*_1})^{-1}(M^*_2(\rho_2(\tau))) := s_{2,1}(\rho_2(\tau)), \quad \tau \in (0, 1). \tag{15}$$

This means we can then use

$$\begin{aligned} \tilde{\theta}_\ell(\mathbf{q}) &= r_{\ell',\ell}(\tilde{\theta}_{\ell'}(\mathbf{q}), \mathbf{q}); \\ \rho_\ell(\tau) &= s_{\ell',\ell}(\mathbf{q}_{\ell'}(\tau)) \end{aligned}$$

to identify $\tilde{\theta}_\ell(\cdot)$, for either $\ell = 1$ or $\ell = 2$. Since, for a $\mathbf{q}$ the probability that $\{\theta \leq t | Z = z_{\ell'}\}$ is equal to the probability that $\{\theta \leq r_{\ell',\ell}(t, \mathbf{q}) | Z = z_\ell\}$, i.e.,

$$\Pr(\theta \leq t | Z = z_{\ell'}) = \Pr(\theta \leq r_{\ell',\ell}(t, \mathbf{q}) | Z = z_\ell),$$

it means

$$\rho_\ell(r_{\ell',\ell}(\theta, \mathbf{q})) = s_{\ell',\ell}(\rho_{\ell'}(\theta)); \tag{16}$$

so if we know $\rho_{\ell'}(\cdot)$ at some θ then we can identify $\rho_\ell(\cdot)$ at $r_{\ell',\ell}(\theta, \mathbf{q})$. As mentioned earlier, let us normalize $v(\mathbf{q}^0) = \mathbf{q}^0$ for some $\mathbf{q}^0 \in Q^2$ so that we know $\{\mathbf{q}^0, \theta^0 = \tilde{\theta}_1(\mathbf{q}^0)\}$.[9] Then this will allow us to identify $\{\mathbf{q}^1, \tilde{\theta}_1(\mathbf{q}^1)\}$ where $\mathbf{q}^1 = s_{1,2}(\mathbf{q}^0)$ and $\tilde{\theta}_1(\mathbf{q}^1) = r_{2,1}(\theta^0, \mathbf{q}^1)$, which further identifies $\{\mathbf{q}^2, \tilde{\theta}_1(\mathbf{q}^2)\}$ with $\mathbf{q}^2 = s_{1,2}(\mathbf{q}^1)$ and $\tilde{\theta}_1(\mathbf{q}^2) = r_{2,1}(\tilde{\theta}_1(\mathbf{q}^1), \mathbf{q}^2)$ and so on. To complete the identification it must be the case that we can begin with any quantile $\rho(\tau) \in Q^2$ and identify $\tilde{\theta}(\rho(\tau))$, possibly by constructing a sequence as above.

To do that we can exploit the Assumption 4, which implies that for some θ the difference $(\theta - r_{2,1}(\theta, \mathbf{q}))$ measures the resulting change in θ if we switch from z_2 to z_1 for a fixed $\mathbf{q}$ so that we can trace $\tilde{\theta}(\cdot)$ as we move back and forth between z_2 and z_1. But for identification it is important that this "tracing" steps come to a halt or equivalently for some (fixed point) $\hat{\mathbf{q}} \in Q^2$ the mapping $(\theta(\cdot) - r_{2,1}(\theta, \cdot)) = 0$. For this it is sufficient that the marginal prices at $\hat{\mathbf{q}}$ are equal $(\nabla P_1(\hat{\mathbf{q}}) = \nabla P_2(\hat{\mathbf{q}}))$. Since this is multidimensional problem, it is also important that the fixed point is attractive (stable), in other words the the slope of all J components in $r_{\ell,\ell'}(\cdot)$ require the fixed point Having fixed point is not enough, it is also important that this fixed point is attractive (stable), so we require that the slope of of all J components of $r_{\ell,\ell'}(\cdot)$ (see (13)) should depend only on whether $q_j > \hat{q}_j$ or not and should be independent of the index $j = 1, 2, \dots, J$.

Assumption 5. *There exist a* $\hat{\mathbf{q}} \in Q^2$ *such that* $r_{\ell,\ell'}(\theta(\mathbf{q}), \hat{\mathbf{q}}) = \theta(\mathbf{q})$ *and* $\mathrm{sgn}[(r^j_{\ell,\ell'}(q_j) - q_j)(q_j - \hat{q}_j)]$ *is independent of* $j \in \{1, \dots, J\}$.

Both the components of this assumptions are testable and can be easily verified in the data.[10] Without loss of generality I assume the initial normalization be the fixed point $\hat{\mathbf{q}}$, so that $\theta^0 = \tilde{\theta}_1(\hat{\mathbf{q}})$ is known. In other words θ^0 is such that $\mathbf{q}(\theta^0, z_1) = \hat{\mathbf{q}}$. And from Assumption 4 suppose $\nabla P_1(\mathbf{q}) \circ \nabla P_2(\mathbf{q})^{-1} << 1$,

[9] We can also normalize some quantile of $F_\theta(\cdot)$.

[10] D'Haultfœuille and Février (2014) use results from the theory of Group of Circle Diffeomorphisms, (see Navas, 2011), to propose sufficient conditions to identify a nonseparable model with discrete instrument and multivariate errors. The assumptions there is very similar to Assumption 5. I am thankful to Xavier D'Haultfœuille for pointing this out.

whenever $\mathbf{q} << \hat{\mathbf{q}}$. Then, for τ^{th} quantile $\mathbf{q}(\tau) < \hat{\mathbf{q}}$:

$$\begin{aligned}
\tilde{\theta}_1(\tau) &:= (\mathbf{q})^{-1}(\mathbf{q}(\tau); z_1) = \tilde{\theta}_1(\mathbf{q}^0) = r_{1,2}(\tilde{\theta}_1(\mathbf{q}^1), \mathbf{q}^1)\\
&= [\nabla P_2(\mathbf{q}^1) \circ \nabla P_1(\mathbf{q}^1)^{-1}] \circ \tilde{\theta}_1(\mathbf{q}^1)\\
&= [\nabla P_2(\mathbf{q}^1) \circ \nabla P_1(\mathbf{q}^1)^{-1}] \circ [r_{1,2}(\tilde{\theta}_1(\mathbf{q}^2), \mathbf{q}^2)]\\
&= [\nabla P_2(\mathbf{q}^1) \circ \nabla P_1(\mathbf{q}^1)^{-1}] \circ [\nabla P_2(\mathbf{q}^2) \circ \nabla P_1(\mathbf{q}^2)^{-1}] \circ [r_{1,2}(\tilde{\theta}_1(\mathbf{q}^3), \mathbf{q}^3)]\\
&\vdots\\
&= r^L[\tilde{\theta}_1(s^L_{1,2}(\mathbf{q}(\tau))), s^L_{1,2}(\mathbf{q}(\tau))]\\
&= \lim_{L\to\infty} [\nabla P_2(\mathbf{q}^1) \circ \nabla P_1(\mathbf{q}^1)^{-1}] \circ \cdots \circ [\nabla P_2(\mathbf{q}^L) \circ \nabla P_1(\mathbf{q}^L)^{-1}] \circ [r_{1,2}(\tilde{\theta}_1(\mathbf{q}^{L+1}), \mathbf{q}^{L+1})]\\
&= \left\{\prod_{L=1}^{\infty} \nabla P_2(\mathbf{q}^L) \circ \nabla P_1(\mathbf{q}^L)^{-1}\right\} \lim_{L\to\infty} \tilde{\theta}_1(s_{1,2}(\mathbf{q}^{L+1}))\\
&= \left\{\prod_{L=1}^{\infty} \nabla P_2(\mathbf{q}^L) \circ \nabla P_1(\mathbf{q}^L)^{-1}\right\} \lim_{L\to\infty} \theta_0, \qquad (17)
\end{aligned}$$

where the first equality is simply the definition, the second equality is the normalization, the third equality follows from (16) with $\mathbf{q}^1 := s_{1,2}(\mathbf{q}^0 = \rho(\tau))$ so that $\tilde{\theta}_1(\mathbf{q}^0) = r_{1,2}(\tilde{\theta}_1(\mathbf{q}^1), \mathbf{q}^1)$ and the fourth equality follows from (13). Repeating this procedure L times leads to the seventh equality. The last equality uses the following facts: a) $\mathbf{q}^L = s_{1,2}(\mathbf{q}^{L-1})$; b) $\mathbf{q}(\tau) < \hat{\mathbf{q}}$; c) $s_{1,2}(\cdot)$ is an increasing continuous function so $\lim_{L\to\infty} s_{1,2}(\mathbf{q}^L) = s_{1,2}(\mathbf{q}^\infty) = s_{1,2}(\hat{\mathbf{q}})$; and d) $\tilde{\theta}_1(\hat{\mathbf{q}}) = \theta_0$. Since the quantile τ was arbitrary, we identify $\tilde{\theta}_1(\cdot)$.[11]

Once the quantile function of θ is identified we can identify $C(\cdot, Z)$ as before. The optimality condition $\alpha(\theta) = 0$ (from Equation (6)) and Equation (8) give

$$\mathbf{div}\left\{\frac{m_k^*(\mathbf{q})}{|det(D\tilde{\theta}_k)(\mathbf{q})|}(\tilde{\theta}_k(\mathbf{q})\nabla \mathbf{v}(\mathbf{q}) - \nabla C(\mathbf{q}; z_k))\right\} = -\frac{m^*(\mathbf{q})}{|det(D\tilde{\theta})(\mathbf{q})|}.$$

Differentiating $\theta_k \circ \nabla \mathbf{v}(\mathbf{q}) = \nabla P_k(\mathbf{q})$ with respect to $\mathbf{q}$ gives

$$\begin{aligned}
D\nabla P_k(\mathbf{q}) &= D\tilde{\theta}_k(\mathbf{q}) \circ \nabla \mathbf{v}(\mathbf{q}) + \tilde{\theta}_k(\mathbf{q}) \circ D\nabla \mathbf{v}(\mathbf{q})\\
D\tilde{\theta}_k(\mathbf{q}) \circ \nabla \mathbf{v}(\mathbf{q}) &= D\nabla P_k(\mathbf{q}) - \tilde{\theta}_k(\mathbf{q}) \circ (\nabla \mathbf{v}(\mathbf{q})) \circ (\nabla \mathbf{v}(\mathbf{q}))^{-1} \circ D\nabla \mathbf{v}(\mathbf{q})\\
D\tilde{\theta}_k(\mathbf{q}) &= D\nabla P_k(\mathbf{q}) \circ (\nabla \mathbf{v}(\mathbf{q}))^{-1} - \nabla P_k(\mathbf{q}) \circ (\nabla \mathbf{v}(\mathbf{q}))^{-2} \circ D\nabla \mathbf{v}(\mathbf{q}),
\end{aligned}$$

[11] Some other examples where similar constructive proof of identification that relies comparing ranks include but are not limited to Guerre, Perrigne, and Vuong (2009); Aryal and Kim (2014); Torgovitsky (2014); D'Haultfœuille and Février (2014).

which identifies $|det(D\tilde{\theta})(\mathbf{q})|$. Then substituting $|det(D\tilde{\theta})(\mathbf{q})|$ in above gives

$$\mathbf{div}\left\{\frac{m^*(\mathbf{q})}{|det(D\tilde{\theta})(\mathbf{q})|}(\nabla P(\mathbf{q})-\nabla C(\mathbf{q}))\right\}=-\frac{m^*(\mathbf{q})}{|det(D\tilde{\theta})(\mathbf{q})|},$$

(a partial differential equation for $C(\cdot,z_k)$), with boundary condition

$$\frac{m_k^*(\mathbf{q})}{|det(D\tilde{\theta}_k)(\mathbf{q})|}(\nabla C(\mathbf{q};z_k)-\nabla P_k(\mathbf{q}))\cdot\overrightarrow{n}(\nabla P_k(\mathbf{q}))=0,\forall\mathbf{q}\in\partial Q^2.$$

This PDE has a unique solution $C(\mathbf{q})$, and hence, we have the following result:

Theorem 4.3. *Under Assumptions 1-(i)–(iv-c) and (v) and Assumptions 2–5, $[F_\theta(\cdot|2),\mathbf{v}(\cdot),C(\cdot;Z)]$ are identified.*

To identify the density $f_\theta(\cdot|1)$ we can use Theorem 4.2, except now the gross utility function is $\sum_{j\in[J]}\theta_j X_{1j}v_j(q_j,X_2)$. Therefore to account for $\mathbf{v}(\cdot,X_2)$ we need to to be able to extend the utility function from Q^2 to $Q^2\cup Q^1$. For the identification strategy then if $\mathbf{v}(\cdot)$ is a real-analytic, like the cost function, then we can extend the domain of $\mathbf{v}(\cdot)$ to include Q^1.

Assumption 6. *Let the utility function $\mathbf{v}(\cdot,X_2)$ be a real analytic function.*

Then under Assumption 6, we can change the unit of measurement from $\mathbf{q}$ to $\tilde{\mathbf{q}}\equiv\mathbf{v}(\mathbf{q},X_2)$, then apply Theorem 4.2 with gross utility as $\sum_{j\in[J]}\theta_j X_j\tilde{q}_j$.

4.4. **Overidentification.** Now that we know identification depends on how many cost shifters we have and whether or not the gradient of the pricing function cross, the next step is analyze the effect of observed characteristics X on identification. Before we begin, let us assume that the nonlinear utility model is identified. Then I ask the following question: if the utility function depended on X, and X is independent of θ is the model *over identified*?

Lemma 3. *Consider the optimal allocation rule restricted for high types $\mathcal{S}_\theta^2$, where $\mathbf{q}=\rho(\theta,X,z_\ell):=\rho_\ell(\theta,X)$. Suppose $F_\theta(\cdot|2)$ and $M_{\mathbf{q}|X,Z}(\cdot|\cdot,\cdot)$ have finite second moments. Then the CDF $F_\theta(\cdot|2)$ is over identified.*

Proof. From the previous results $F_\theta(\cdot|2)$ and $M_{\mathbf{q}|X,Z}(\cdot|X)$ are nonparametrically identified. Since Z is observed, we can suppress the notation. We want to use the data $\{\mathbf{q},X\}$ and the knowledge of $F_\theta(\cdot|2)$ and the truncated distribution $M^*_{\mathbf{q}|X}(\cdot|X)$ to identify $\rho(\cdot,X)$. Let $\mathcal{L}(\mathcal{S}_\theta^2,Q^2)$ be the set of joint distribution defined as

$$\mathcal{L}(\mathcal{S}_\theta^2,Q^2)=\{L(\mathbf{q},\theta):\int_{\mathcal{S}_\theta^2}L(\mathbf{q},\theta)\mathrm{d}\theta=M^*_{\mathbf{q}|X}(\mathbf{q}|\cdot);\int_{Q_X^2}L(\mathbf{q},\theta)\mathrm{d}\mathbf{q}=F_\theta(\theta|2)\}.(18)$$

To that end consider the following optimization problem:

$$\min_{L(\mathbf{q},\theta)\in\mathcal{L}} \mathbb{E}(|\mathbf{q}-\theta|^2|X).$$

In other words, given two sets $\mathcal{S}_\theta^2$ and Q_X^2 of equal volume we want to find the optimal volume-preserving map between them, where optimality is measured against cost function $|\theta-\mathbf{q}|^2$. If the observed $\mathbf{q}\in Q^2$ were generated under equilibrium then the solution will map $\mathbf{q}$ to the right θ such that $\mathbf{q}=\rho(\theta;X)$, for a fixed X. The minimization problem is equivalent to

$$\max_{L(\mathbf{q},\theta)\in\mathcal{L}} \mathbb{E}(\theta\cdot\mathbf{q}|X),$$

such that the solution maximizes the (conditional) covariance between θ and $\mathbf{q}$. So either we minimize the quadratic distance or the covariance, our objective is to find an optimal way to "transport" $\mathbf{q}$ to θ. Let $\delta[\cdot]$ be a Dirac measure or a degenerate distribution. Brenier (1991); McCann (1995) show that that there exists a unique convex function $\Gamma(\mathbf{q},X)$ such that $\mathrm{d}L(\mathbf{q},\theta)=\mathrm{d}M^*_{\mathbf{q}|X}(\mathbf{q})\delta[\theta=\nabla_{\mathbf{q}}\Gamma(\mathbf{q},X)]$ is the solution. Therefore for all $\mathbf{q}\in Q_X^2$ we can determine its inverse $\theta=\nabla_{\mathbf{q}}\Gamma(\mathbf{q},X)$ which identifies $F_\theta(\cdot|2)$. □

This means, we can use $\Gamma(\mathbf{q},X)$ to test the validity of the supply side equilibrium. There are many ways to think of a "specification test." One way is by verifying that using $\nabla_{\mathbf{q}}\Gamma(\mathbf{q},X)$ (instead of θ) in Equation (3) leads to the same equilibrium $\rho(\theta;X)$. The result only guarantees that such a unique function $\Gamma(\mathbf{q},X)$ exists, computing it is significantly difficult; see Su, Zeng, Shi, Wang, Sun, and Gu (2013) for an example.

5. Model Restrictions

In this section I derive the restrictions imposed by the model on observables under the Assumption 1-(iv) –a, b and c, respectively. These restrictions can be used to test the model validity. For every agent we observe $[p_i,\mathbf{q}_i,X_i]$ and for the seller we observe $\{z_1,z_2\}$. From the model p_i and $\mathbf{q}_i$ are given by $p=P_\ell(\mathbf{q},z_\ell)$ and $\mathbf{q}=\rho_\ell(\theta,z_\ell)$. Specifically, suppose a researcher observes a sequence of price and quantity data, and some agents and cost characteristics. Does there exist any possibility to rationalize the data such that the underlying screening model is optimal when the utility function satisfies Assumption 1-(iv-a) (Model 1) or Assumption 1-(iv-b) (Model 2) or Assumption 1-(iv-c) (Model 3)? In all three models we ask, in the presence of multidimensional asymmetric information, what are the restrictions on the sequence of data $(Z,X_i,\{\mathbf{q}_i,p_i\})$ we can test if and only if it is generated by an optimal screening model, without knowing the cost function, the type distribution and for Model 3 the utility function. We say that a distribution of the observables is rationalized by a model if and only if it satisfies all the restrictions of the model. In other words, a distribution of the observables is rationalized if and only if there is a structure (not necessarily unique) in the model that generates such a distribution.

Let $D_1 = (\mathbf{q}, p), D_2 = (\mathbf{q}, p, X_1), D_3 = (\mathbf{q}, p, X, Z)$ distributed, respectively, as $\Psi_{D_\ell}(\cdot), \ell = 1, 2, 3$, and let

$$\begin{aligned}
\mathcal{M}_1 &= \{(F_\theta(\cdot), C(\cdot)) \in \mathcal{F} \times \mathcal{C} : \text{satisfy Assumption} 1 - (i) - (iv - a), (v)\} \\
\mathcal{M}_2 &= \{(F_\theta(\cdot), C(\cdot)) \in \mathcal{F} \times \mathcal{C} : \text{satisfy Assumption} 1 - -(i) - (iv - b), (v)\} \\
\mathcal{M}_3 &= \{(F_\theta(\cdot), C(\cdot, Z)) \in \mathcal{F} \times \mathcal{C}_Z : \text{satisfy Assumptions} 1 - -(i) - (iv - c), (v), 3 \text{ and } 4\}
\end{aligned}$$

Define the following conditions:
$C1$. $\Psi_{D_1}(\cdot) = \delta[p = P(\mathbf{q})] \times M(\mathbf{q})$, *with density* $m(\mathbf{q}) > 0$ *for all* $\mathbf{q} \in Q^1 \cup Q^2$.
$C2$. *There is a subset* $Q^1 \subsetneq Q$ *which is a* $J - 1$ *dimensional flat (hyperplane) in* $\mathbb{R}^J_+$.
$C3$. $p = P(\mathbf{q})$ *has non vanishing gradient and Hessian for all* $\mathbf{q} \in Q^2$.
$C4$. *Let* $\{W\} := \{\nabla P(\mathbf{q}) : \mathbf{q} \in Q^2\}$. *Then* $F_W(w) = \Pr(W \leq w) = M^*(\mathbf{q})$ *and let* $m^*(\cdot) > 0$ *be he density of* $M^*(\cdot)$
$C5$. *Let* $C(\cdot)$ *be the solution of the differential equation*

$$\mathbf{div}\left\{\frac{m^*(\mathbf{q})}{|det(D\nabla P(\mathbf{q}))|}(\nabla P(\mathbf{q}) - \nabla C(\mathbf{q}))\right\} = -\frac{m^*(\mathbf{q})}{|det(D\nabla P(\mathbf{q}))|}, \qquad (19)$$

with boundary conditions

$$\frac{m^*(\mathbf{q})}{|det(D\nabla P(\mathbf{q}))|}(\nabla C(\mathbf{q}) - \nabla P(\mathbf{q})) \cdot \overrightarrow{n}(\nabla P(\mathbf{q})) = 0.$$

5.1. **Linear Utility.** For every consumer we observe D_1 and the objective is to determine the necessary and sufficient on the joint distribution $\Psi_{D_1}(\cdot, \cdot)$ for it to be rationalized by model $\mathcal{M}_1$.

Lemma 5.1. *If* $\mathcal{M}_1$ *rationalizes* $\Psi_{D_1}(\cdot)$ *then* $\Psi_{D_1}(\cdot)$ *satisfies conditions* $C1. - C5$. *Conversely, if* $F_\theta(\cdot|0)$ *and* $F_\theta(\cdot|1)$ *are known and* $\Psi_{D_1}(\cdot)$ *satisfies the* $C1. - C5$ *then there is a model* $\mathcal{M}_1$ *that generates* D_1.

Proof. If. Since $F_\theta(\cdot)$ is such that the density $f_\theta(\cdot) > 0$ everywhere on $\mathcal{S}_\theta$ and the equilibrium allocation rule $\rho : \mathcal{S}_\theta \to Q$ is onto, and continuous, the CDF $M(\mathbf{q})$ is well defined and the density $m(\mathbf{q}) > 0$. Moreover, since the equilibrium allocation rule is deterministic, for every $\mathbf{q}$ there is only one price $P(\mathbf{q})$, hence the Dirac measure, which completes $C1$. Rochet-Choné shows that in equilibrium the bunching set Q^1 is nonempty, and hence $m(\mathbf{q}) > 0$ for all $\mathbf{q} \in Q^1$. Moreover the allocation rule $\rho : \mathcal{S}_\theta^1 \to Q^1$ is not bijective, and as a result Q^1 as a subset of $\mathbb{R}_+^J$ is flat, which completes $C2$. The optimality condition for the types that are perfectly screened is $\theta = \nabla P(\mathbf{q}) := \tilde{\theta}(\mathbf{q})$, and incentive compatibility implies the indirect utility function is convex and hence $P(\mathbf{q})$ has non vanishing gradient and Hessian, which completes $C3$. Then, $M^*(q) = \Pr(\mathbf{q} \le q) = \Pr(\nabla P(\mathbf{q}) \le \nabla P(q)) = \Pr(W \le w) = F_W(w)$, hence $C4$. Finally, if we use (8) to replace $f_\theta(\cdot)$ in $\alpha(\theta) = 0, \forall\theta \in \mathcal{S}_\theta^2$ with the boundary condition $\beta(\theta) = 0, \forall\theta \in \partial\mathcal{S}_\theta^2 \cap \partial\mathcal{S}_\theta$ we get $C5$.

Only if. Now, we show that if $\Psi_{D_1}(\cdot)$ satisfies all $C1. - C5.$ conditions listed above then we can determine a model $\mathcal{M}_1$ that rationalizes $\Psi_{D_1}(\cdot)$. Let $C(\cdot)$ be the (cost) function that satisfies $C5.$ that we can determine the cost function $C(\cdot)$. Moreover it is real analytic so it can be extend uniquely to all Q. From $C4.$ we can determine the vector W which is also the type θ and it satisfies the first order optimality condition. Thus the indirect utility of the type θ that corresponds to the choices $\mathbf{q} \in Q^2$ is convex and hence satisfies the incentive compatibility constraint. Moreover, since $m^*(\mathbf{q}) > 0$ the density $f_\theta(\cdot|2) > 0$ and $F_\theta(\cdot|2) = \int_{\theta\in\{W:=\nabla P(\mathbf{q}),\mathbf{q}\in Q^2\}} f_\theta(\theta|2)\mathrm{d}\theta$. As far as $F_\theta(\cdot|1)$ is concerned we can simply ignore bunching and define $F_\theta(\theta) = M(q|\mathbf{q} \in Q^1)$ where $\mathbf{q} \in Q^1$ is such that $\theta = \nabla P(\mathbf{q})$. □

5.2. **Bi-Linear Utility.** Now, I consider the case of bi-linear utility function. Since X_2 is redundant information, we can ignore it. The only difference between this and the previous model is now there is X_2 but everything else is the same. So to save more notations, I slightly abuse notations and use the same conditions $C1. - C5.$ except now they are understood with respect to D_2. For instance $C1.$ becomes $\Psi_{D_2}(\cdot) = \delta[p = P(\mathbf{q}; X_1)] \times M(\mathbf{q}) \times \Psi_{X_1}$.

Lemma 5.2. *If $\mathcal{M}_2$ rationalizes $\Psi_{D_2}(\cdot)$ then $\Psi_{D_2}(\cdot)$ satisfies conditions $C1. - C5.$ Conversely, if $F_\theta(\cdot|0)$ is known, $dim(X_1) = dim(\mathbf{q}) = J$, and $\Psi_{D_2}(\cdot)$ satisfies $C1. - C5$ then there is a model $\mathcal{M}_2$ that generates D_2.*

The proof of this lemma is very similar to that of Lemma 5.2, except in here the menu (allocation and prices) depend on X_1 but the cost function and the type CDF do not depend on, and the conditional density $f_\theta(\cdot|1)$ can be determined from the data. In view of the space I omit the proof.

5.3. **Nonlinear Utility.** Finally, I consider the case of nonlinear utility. Before I proceed, I introduce two more conditions.

$C4'$. If $\rho_\tau(X_2, Z)$ is the $\tau \in [0,1]$ quantile of $\mathbf{q} \in Q^2_{X_2,Z}$ then $\rho_\tau(\cdot, z_1) = \rho_\tau(\cdot, z_2)$.

$C6$. The truncated distribution of choices $M^*_{\mathbf{q}|X,Z}(\cdot|\cdot,\cdot)$ has finite second moment, and for a given $Z = z_\ell$ (henceforth suppressed) the solution of

$$\max_{L(\mathbf{q},\theta)\in\mathcal{L}(Q^2,\mathcal{S}^2_\theta)} \mathbb{E}(\theta\cdot\mathbf{q}|X),$$

where $\mathcal{L}(Q^2,\mathcal{S}^2_\theta)$ is defined in (18), be is given by a mapping $\theta = \nabla_{\mathbf{q}}\Gamma(\mathbf{q}, X)$ for some convex function $\Gamma(\mathbf{q}, X)$ such that it solves the optimality condition (3).

So with nonlinear utility, condition $C4'$. replaces condition $C4$. and as with the bi-linear utility the conditions should be interpreted as being conditioned on both X and Z, wherever appropriate.

Lemma 5.3. *Let $F_\theta(\cdot|2)$ have finite second moment. If $\mathcal{M}_3$ rationalizes $\Psi_{D_3}(\cdot)$ then $\Psi_{D_3}(\cdot)$ satisfies $C1.-C3., C4.'-C7$. Conversely, if $F_\theta(\cdot|0)$, and a quantile $\tilde{\theta}(\mathbf{q}_\tau)$ is known, $dim(X_1) = dim(\mathbf{q}) = J, Q^2_{X,z_k} = Q^2_{X,z_{k'}}$ (common support) and $\Psi_{D_3}(\cdot)$ satisfies $C1. - C3., C4.' - C6.$ then there exists a model $\mathcal{M}_3$ that rationalizes $\Psi_{D_3}(\cdot)$.*

Proof. If. The CDF is $F_\theta(\cdot)$ with non vanishing density $f_\theta(\cdot)$ everywhere on the support $\mathcal{S}_\theta$. Moreover, the equilibrium allocation rule $\rho : \mathcal{S}_\theta \times X \times Z \to Q$ is onto, and continuous for given (X, Z). Therefore the CDF $M_{\mathbf{q}|X,Z}(\cdot|\cdot,\cdot)$ is a push forward of $F_\theta(\cdot)$ given (X, Z). Since $Q = Q^2_{(X,Z)} \cup Q^1_{(X,Z)} \cup \{\mathbf{q}_0\}$ the (truncated) density $m_{\mathbf{q}|X,Z}(\mathbf{q}|\cdot,\cdot) > 0$ for all $\mathbf{q} \in Q^2_{(X,Z)} \cup Q^1_{(X,Z)}$. In equilibrium, for a given $(\mathbf{q}, X, Z)$ the pricing function is deterministic, therefore the distribution is degenerate at $p = P(\mathbf{q}; X, Z)$. Hence the Dirac measure. This completes $C1$. For $C2$. note that the allocation rule is not bijective, and as a result $\rho(\mathcal{S}^1_\theta; X, Z) = Q^1 \subsetneq \mathbb{R}^J_+$ is a hyperplane. For the high-types, optimality requires the marginal utility $\theta \cdot v(\mathbf{q}; X_2)$ is equal to the marginal price $P(\mathbf{q}; X, Z)$, and since $\mathbf{v}(\cdot; X_2)$ has non vanishing Hessian, $P(\cdot; X, Z)$ also has non vanishing gradient and non vanishing gradients $\nabla P(\cdot; X, Z)$ and Hessian, which completes $C3$. Since $Z \perp\!\!\!\perp \theta$, using Equation (14) gives $F_\theta(\xi|2) = M_{\mathbf{q}|X,Z}(\rho_1(\xi)|X, z_1) = M_{\mathbf{q}|X,Z}(\rho_2(\xi)|X, z_2)$, as desired for $C4'$. The condition $C5$. follows once we replace $m^*(\cdot)$ and $P(\mathbf{q})$ in (19) with $m^*_{\mathbf{q}|X,Z}(\cdot|\cdot,\cdot)$ and $P(\mathbf{q}; X, Z)$, respectively and observe that for any pair (X, Z) the equilibrium for high-type is given by $\alpha(\theta) = 0$. Since $F_\theta(\cdot|2)$ is known and $M^*_{\mathbf{q}|X,Z}(\cdot)$ is determined, condition $C6$. follows from Lemma 3.

Only if. We want to show that if $\Psi_{D_3}(\cdot)$ satisfies all conditions in the statement, then we can construct a model $\mathcal{M}_3$ that rationalizes $\Psi_{D_3}(\cdot)$. For $Z = z_k$, using condition $C6$. we can determine two cost functions $C(\cdot, z_1)$ and $C(\cdot, z_2)$. Since (19) is applicable only to $Q^2_{X,Z}$, we need to extend the domain of the cost function. Of many ways to extend the domain, the simplest is to assume that the cost is quadratic, i.e. $C(\mathbf{q}; X, Z) = 1/2 \sum_{j=1}^{J} q_j^2$ for all $\mathbf{q} \in Q^1_{X,Z} \cup \{\mathbf{q}_0\}$. Using the exclusion restriction and (17) for all $\mathbf{q} \in Q^2_{X,Z}$ we can determine the function $\tilde{\theta}(\mathbf{q}_\tau; Z = z_k)$ along a set $\tilde{Q}^2_{X,Z} \subseteq Q^2_{X,Z}$ for $k = 1, 2$. If the set $\tilde{Q}^2_{X,Z}$ is a dense subset then there is a unique extension of $\tilde{\theta}(\cdot;\cdot)$ over all $Q^2_{X,Z}$. If not, then, let us linearly extend the function to the entire domain of $Q^2_{X,Z}$. Then define $\mathbf{v}(\mathbf{q}; X_2) = \nabla P(\mathbf{q}; X, Z) \circ (\tilde{\theta}_k(\mathbf{q}))^{-1}$. Finally, to extend the function to Q we can assume that each function $v_j(q_j; X) = q_j^{1/2}, j = 1, \ldots, J$ for all $\mathbf{q} \in Q^1_{X,Z} \cup \{\mathbf{q}_0\}$. As far as $F_\theta(\cdot|1)$ is concerned we can simply ignore bunching and define $F_\theta(\theta) = M(q|\mathbf{q} \in Q^1)$ where $\mathbf{q} \in Q^1$ is such that $\theta = \nabla P(\mathbf{q}; X, Z) \circ (\mathbf{v}(\mathbf{q}; X_2))^{-1}$. Since the probability of $\mathbf{q} = \{\mathbf{q}_0\}, \mathbf{q} \in Q^1_{X,Z}$ and $\mathbf{q} \in Q^2_{X,Z}$ is equal to the probability of $\theta \in \mathcal{S}^0_\theta, \theta \in \mathcal{S}^1_\theta$ and $\theta \in \mathcal{S}^2_\theta$, respectively we can determine $F_\theta(\cdot)$. It is then straightforward to verify that the triplet thus constructed belong to $\mathcal{M}_3$. □

6. Extensions

6.1. **Unobserved Taste Shifter.** So far I have assumed that consumer's tastes are completely characterized by a vector θ. But suppose there is an unobserved market level taste shifter Y that scales the taste for all consumers, and as such it is observed by all consumers and the seller but not the econometrician.

Assumption 7. *Let*

(1) *The random variables (θ, Y) are distributed on $\mathcal{S}_\theta \times \mathbb{R}_{++}$ according to the CDF $F_{\theta,Y}(\cdot,\cdot)$ such that $\Pr(\theta \leq \theta_0, Y \leq y_0) = F_{\theta,Y}(\theta_0, y_0)$.*
(2) *Let $\theta^* := Y \times \theta$ be such that $\theta^* \sim F_{\theta^*|Y}(\cdot|y) = F_{\theta^*}(\cdot)$ and $\mathbb{E}(\log Y) = 0$.*

Let $\mathcal{S}^2_{\theta^*|Y}$ denote the types that are perfectly screened. Then under assumption 7 optimality of these types means $\theta^*_i = \nabla P(\mathbf{q}_i)$ and since $\theta^*_i = \theta_i y, i \in [N_2]$, we want to identify $F_\theta(\cdot)$ and $F_Y(\cdot)$ from above. Dividing $[N_2]$ into two parts and reindexing $\{1, \ldots, N_{21}\}$ and $\{1, \ldots, N_{22}\}$ and taking the log of the above we get

$$\log \theta^*_{i_j} \quad = \quad \log \theta_{i_j} + \log Y, \quad i_j = 1, \ldots N_{2j}, j = 1, 2.$$

Let $Ch(\cdot,\cdot)$ be the joint characteristic function of $(\log \theta_{i_1}, \log \theta_{i_2})$ and $Ch_1(\cdot,\cdot)$ be the partial derivative of this characteristic function with respect to the first component. Similalry, let $Ch_{\log Y}(\cdot)$ and $Ch_{\log \theta_j}(\cdot)$ denote characteristic functions of $\log Y$ and $\log \theta_j$, which is the short hand for $\theta_{i_j}, i_j \in [N_{2j}]$. Then from

Kotlarski (1966):

$$Ch_{\log Y}(\xi) \quad = \quad \exp\left(\int_0^{\xi}\frac{Ch_1(0,t)}{Ch(0,t)}dt\right) - \mathbf{i}t\mathbb{E}[\log\theta_1].$$

Then the characteristic function of $Ch_{\log\theta_1}(\xi) = \frac{Ch(\xi,0)}{Ch_{\log Y}(\xi)}$, which identifies $F_\theta(\cdot)$.

Lemma 6.1. *Under Assumption 7, the model $[F_\theta(\cdot), F_Y(\cdot), C(\cdot), \mathbf{v}(\cdot)]$ with unobserved heterogeneity is identified.*

6.2. **Measurement Errors.** So far we have assumed that the econometrician observes both the transfers and the contract characteristics without an error. Such an assumption could be strong in some environment. Sometimes it is hard to measure the transfers (wages, prices etc) and sometimes it is hard to measure different attributes of contracts. For instance a monopoly who sells differentiated products it is possible that some if not all of the attributes of the product are measured with error. In this subsection we allow data to be measured with error.

6.2.1. *Measurement Error in Prices.* I begin by considering the case when only the transfers are measured with error, and subsequently consider the case when only the contract choices are measured with error. If only the prices are measured with additive error, and if the error is independent of the true price then the model is still identified. The intuition behind this is simple. When choices $\{\mathbf{q}\}$ are observed without error, but only prices are observed with error, and if this error is additively separable and independent of the true prices, i.e.,

$$P^{\varepsilon}(\mathbf{q}) = P(\mathbf{q}) + \varepsilon, \qquad P(\mathbf{q})\perp\!\!\!\perp\varepsilon,$$

then the observed marginal prices $\nabla P^{\varepsilon}(\cdot)$ and the true marginal prices $\nabla P(\cdot)$ are the same, which means the previous identification arguments are still applicable.

Lemma 6.2. *If $\{P^{\varepsilon} = P + \varepsilon\}$ is observed, where P is the price and $P\perp\!\!\!\perp\varepsilon$ is the measurement error, then the model parameters $[F_\theta(\cdot), C(\cdot)]$ are identified.*

6.2.2. *Measurement Error in Choices.* Now consider a case where the choices $\mathbf{q}$'s are observed with error. Furthermore, lets assume that there is one (dimensional) error $\eta \in \mathbb{R}_+$ that affects all J characteristics. In other words, we envision a situation where there is one η for each consumer, and instead of the choice $\mathbf{q}_i$ we only measure $\mathbf{q}^{\eta} = \mathbf{q} + \eta\cdot\mathbf{1}$, where $\mathbf{1}$ is $J-$ dimensional vector of ones. Since there is no reason why each component $\mathbf{q}$ should have unique measurement error associated with it, having one error unique to each consumer choice seems more natural in this environment. We also assume that $\eta \perp\!\!\!\perp \mathbf{q}$ and $\eta \sim F_\eta(\cdot)$. The data is then $\{P, \mathbf{q}^\eta\}$ pair for every consumer with type $\theta \in \mathcal{S}_\theta^2$. Then $P = P(\mathbf{q}) = P(\mathbf{q}^\eta - \eta \cdot \mathbf{1})$ implies $\nabla P(\mathbf{q}) \neq \nabla P(\mathbf{q}^\eta)$, which means without correcting for η the taste parameter θ cannot be identified. Following the same logic as Lemma 6.1 we can identify $F_\eta(\cdot)$ but that still does not mean we can identify θ because we have $J+1$ unknowns and only J equation for each consumer.

Lemma 6.3. *If $\{\mathbf{q}^\eta = \mathbf{q} + \eta \cdot \mathbf{1}\}$ is observed, where $\eta \perp\!\!\!\perp \mathbf{q}$ is the measurement error then the model $[F_\theta(\cdot), C(\cdot)]$ cannot be identified.*

6.3. **Unobserved Product Characteristic.** In this section I extend the linear utility model to allow for unobserved product characteristic. Suppose we observe two bundles $\mathbf{q}$ and $\mathbf{q}'$ where the former dominates the later $\mathbf{q} >> \mathbf{q}'$ and $P(\mathbf{q}) \leq P(\mathbf{q}')$, with positive demand for both. This suggests either that the model is wrong, and hence we can reject the Rochet-Choné model as a good description of the data generating process, or that some product characteristic is missing in the data. Let $Y \in \mathbb{R}_+$ denote such characteristic that is observed only by the consumers and the seller such that the net utility when a type-$\theta := (\theta_1, \ldots, \theta_J, \theta_y)$ chooses $\{\mathbf{q}\} \cup \{Y\}$ bundle and pays P is

$$V(\mathbf{q}, Y; \theta) = \theta \cdot \mathbf{q} + \theta_y Y - P,$$

where $\theta_y \in \mathcal{S}_y$ is the consumer's taste for Y. The two bundles mentioned above are in fact $(\mathbf{q}, Y)$ and $(\mathbf{q}', Y')$ with $P(\mathbf{q}, Y) \leq P(\mathbf{q}', Y')$. This means our econometrics model can be written as

$$\begin{aligned} P &= P(\mathbf{q}, Y) \\ \begin{pmatrix} \mathbf{q} \\ Y \end{pmatrix} &= \rho(\theta_1, \ldots, \theta_J, \theta_y, F_\theta, C), \end{aligned}$$

where $(Y, P(\cdot,\cdot), F_{\theta,\theta_y}(\cdot,\cdot), C(\cdot))$ are unknown. Since the product characteristics are endogenous, the observed product characteristics are correlated with the unobserved characteristics Y and hence the model cannot be identified.[12]

7. Conclusion

In this paper I study the identification of a screening model studied by Rochet and Choné (1998) where consumers have multidimensional private information.

[12] Bajari and Benkard (2005) assume $\theta_y \equiv 1$ and assume Y is independent of $\mathbf{q}$ as one way to identify the model.

I show that if the utility is linear or bi-linear, as is often used in empirical industrial organization literature, then we can use the optimality of both supply side and the demand side to nonparmetrically identify the multidimensional unobserved consumer taste distribution and the cost function of the seller. The key to identification is to exploit equilibrium bijection between the unobserved types and observed choices and the fact that in equilibrium consumer will choose a bundle that equates marginal utility to marginal prices. When private information is multidimensional, however, the allocation rule need not be bijective for all types. For those medium-types who are bunched, I show that if we have information about consumers' socioeconomic and demographic characteristics that are independent of the types and if there are as many such characteristics as products, the joint density of types can be identified.

When utility is nonlinear, having a binary and exogenous cost shifter is sufficient for identification. I also show that with nonlinear utility if we have independent consumer characteristics then the model is over identified, which can be used to test the validity of supply side optimality. To the best of my knowledge, this is a first study that provides a way to test optimality of equilibrium in a principal-agent model. Furthermore, I characterize all testable restrictions of the model on the data, and extend the identification to consider measurement error and unobserved heterogeneity.

This paper complements the literature on structural analysis of data using principal-agent models. The next step in this line of research is estimating the models. Estimation of the linear or bilinear utility model is straightforward, but that of nonlinear utility is not obvious because of the way the utility has been identified. Having said that, it might be possible to extend the minimum distance estimation method proposed by Torgovitsky (2013) to estimate nonseparable models with univariate error to multivariate errors. One important caveat of this model is that it considers only a single seller, and allowing competition is an important area of research, addressed in Aryal (2015).

Appendix A. Multivariate Quantiles

Let $(\mathbb{S}, \mathcal{B}, L)$ be a probability space with probability measure L. Let $g : \mathbb{R}^J \times S \to \mathbb{R}$ be a function such that $g(\mathbf{q}, \cdot)$ is integrable function $L-$ almost everywhere and $g(\cdot, s)$ is strictly convex. Let

$$g_L(\mathbf{q}) := \int_{\mathbb{S}} g(\mathbf{q}, s)L(ds), \quad \mathbf{q} \in \mathbb{R}^J.$$

be an integral transform of L. Let the minimal point of the functional

$$g_{L,t}(\mathbf{q}) := g_L(\mathbf{q}) - < \mathbf{q}, t >, \quad \mathbf{q} \in \mathbb{R}^J$$

be called an $(M, t)-$ parameter of L with respect to g, where $< \cdot, \cdot >$ is the inner product in $\mathbb{R}^J$. The subdifferential of g at a point $s \in \mathbb{R}^J$ is denoted by $\partial g(s) = \{t \in \mathbb{R}^J | g(s') \geq g(s) + < s' - s, t >\}$. Since the kernel $g(\cdot, s)$ is strictly convex, g_L is convex and the subdifferential map ∂g_L is well defined. The inverse of this map $\partial g_L^{-1}(t)$ is the quantile function and is the set of all $(M, t)-$ parameters of L. Since g is strictly convex, ∂g_L^{-1} is a single-valued map, and hence we get a unique quantile.[13]

I can choose any kernel function g as long as it satisfies the conditions mentioned above to define multivariate quantile. Then from Proposition 2.6 and Corollary 2.9 in Koltchinskii (1997) we know that ∂g_L is a strictly monotone homeomorphism from $\mathbb{R}^J$ onto $\mathbb{R}^J$ and for any two probability measures L_1 and L_2, the equality $\partial g_{L_1} = \partial g_{L_2}$ implies $L_1 = L_2$. For this paper, we choose $g(\mathbf{q}; s) := |\mathbf{q} - s| - |s|$, so that $g_L(\mathbf{q}) = \int_{\mathbb{R}^J} (|\mathbf{q} - s| - |s|) L(ds), s \in \mathbb{R}^J$, and

$$\partial g_L(\mathbf{q}) := \int_{\{s \neq \mathbf{q}\}} \frac{(\mathbf{q} - s)}{|\mathbf{q} - s|} L(ds), \tag{20}$$

with the inverse $\partial g_L^{-1}(\cdot)$ as the (unique) quantile function.

References

Akerlof, G. A. (1970): "The Market for "lemons": Quality uncertainty and the market mechanism," *Quarterly Journal of Economics*, 84(3), 488–500.

Armstrong, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51–75.

Aryal, G. (2013): "An Empirical Analysis of Competitive Nonlinear Pricing," *Working Paper*.

——— (2015): "Identification of Preferences under Competition and Endogenous Product Characteristics.," *Mimeo*.

Aryal, G., and D.-H. Kim (2014): "Empirical Relevance of Ambiguity in First Price Auctions," *Mimeo*.

Aryal, G., I. Perrigne, and Q. H. Vuong (2010): "Identification of Insurance Models with Multidimensional Screening," *Working Paper*.

[13] For example, with one-dimensional case, for any $t \in (0, 1)$ the set of all t^{th} quantiles of a cdf M is exactly the set of all minimal points of $g_{L,t}(q) := 1/2 \int_{\mathbb{R}} (|q - s| - |s| + q) L(ds) - qt$.

Bajari, P., and C. L. Benkard (2005): "Demand Estimation with Heterogenous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, 113(6), 1239–1276.

Baron, D. (1989): "Design of Regulatory Mechanisms and Institutions," in *Handbook of Industrial Organization*, ed. by R. Schlamensee, and R. Willing, vol. II, pp. 1347–1448. Amsterdam-North Holland.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–846.

——— (1974): "A Theory of Marriage: Part II," *Journal of Political Economy*, 82(2), S11–S26.

Berry, S. T. (1994): "Estimation of Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25(2), 242–262.

Berry, S. T., A. Gandhi, and P. A. Haile (2013): "Connected Substitutes and Invertibility of Demand," *Econometrica*, 81(5), 2087–2111.

Brenier, Y. (1991): "Polar Factorization and Monotone Rearrangement of Vector-Valued Functions," *Communication on Pure and Applied Mathematics*, 44, 375–417.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6. Elsevier.

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, 82(2), 785–809.

Chiappori, P.-A., and B. Salanié (2000): "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108(1), 56–78.

——— (2003): "Testing Contract Theory: A Survey of Some Recent Work," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. P. Hansen, and S. Turnovksy, vol. 1. Cambridge University Press.

Cohen, A., and L. Einav (2007): "Estimating Risk Preferences from Deductible Choices," *American Economic Reveiw*, 97(3), 745–788.

Crawford, G. S., and A. Yurukoglu (2012): "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 102, 301–317.

D'Haultfœuille, X., and P. Février (2014): "Identification of Nonseprarable Triangular Models with Discrete Instruments," *Econometrica*, Forthcoming.

Einav, L., A. Finkelstein, and M. R. Cullen (2010): "Estimating Welfare in Insurance Markets using Variation in Prices," *Quarterly Journal of Economics*, 125(3), 877–921.

Evans, L. C. (2010): *Partial Differential Equations*. AMS, second edn.

Fan, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market," *American Economic Reveiw*, 103(5), 1598–1628.

Finkelstein, A., and K. McGarry (2006): "Multiple dimensions of private information: Evidence from the long-term care insurance market," *American Economic Reveiw*, 96(4), 938–958.

Fox, J. T., and A. Gandhi (2013): "Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models," *Mimeo*.

Fox, J. T., K. il Kim, S. P. Ryan, and P. Bajari (2012): "The Random Coefficients Logit Model is Identified," *Journal of Econometrics*, 166, 204–212.

Fujimoto, T., and C. Herrero (2000): "A Univalence Theorem for Nonlinear Mappings: An Elementary Approach," , 31(4), 277–283.

Gautier, E., and S. Hoderlein (2012): "A Triangular Treatment Effect Model with Random Coefficients in the Selection Equation," *Mimeo*.

Gautier, E., and Y. Kitamura (2013): "Nonparametric Estimation in Random Coefficients Binary Choice Models," *Econometrica*, 81(2), 581–607.

Gayle, G.-L., and R. A. Miller (2014): "Identifying and Testing Models of Managerial Compensation," *Review of Economic Studies*, forthcoming.

Guerre, E., I. Perrigne, and Q. Vuong (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77(4), 1193–1227.

Helgason, S. (1999): *The Radon Transform*. Birkhauser Boston, 2nd edn.

Hoderlein, S., L. Nesheim, and A. Simoni (2013): "Semiparametric Estimation of Random Coefficients in Structural Economic Models," *Mimeo*.

Ivaldi, M., and D. Martimort (1994): "Competition under Nonlinear Pricing," *Annales d'Économie et de Statistique*, 34, 71–114.

Joskow, N., and N. Rose (1989): "The Effects of Economic Regulation," in *Handbook of Industrial Organization*, ed. by R. Schlamensee, and R. Willing, vol. II, pp. 1449–1506.

Kachurovksii, I. R. (1960): "On Monotone Operators and Convex Functionals," *Uspekhi Mat. Nauk*, 15(4), 213–215.

Kleven, H. J., C. T. Kreiner, and E. Saez (2009): "The Optimal Income Taxation of Couples," *Econometrica*, 77(2), 537–560.

Koltchinskii, V. I. (1997): "M-Estimation, Convexity and Quantiles," *The Annals of Statistics*, 25(2), 435–477.

Kotlarski, I. (1966): "On Some Characterization of Probability Distributions in Hilbert Spaces," *Annali di Mathematica Pura ed Applicata*, 74(1), 129–134.

Laffont, J.-J. (1994): "The New Economics of Regulation Ten Years After," *Econometrica*, 62, 507–537.

Lancaster, K. J. (1971): *Consumer Demand: A New Approach*. Columbia University Press, New York.

Lang, S. (1973): *Calculus of Several Variables*. Addison-Wesley.

Luo, Y., I. Perrigne, and Q. Vuong (2012): "Multiproduct Nonlinear Pricing: Mobile Voice Service and SMS," *Mimeo*.

McCann, R. J. (1995): "Existence and Uniqueness of Monotone Measure-Preserving Maps," *Duke Mathematical Journal*, 80, 309–323.

Navas, A. (2011): *Groups of Circle Diffeomorphisms*, Chicago Lectures in Mathematics. The University of Chicago Press.

Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Parthasarathy, T. (1983): *On global univalence theorems*. Springer.

Perrigne, I., and Q. H. Vuong (2011): "Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard," *Econometrica*, 79(5), 1499–1539.

Rochet, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 16(2), 191–200.

Rochet, J.-C., and P. Choné (1998): "Ironing, Sweeping, and Multidimensional Screeing," *Econometrica*, 66(4), 783–826.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–591.

Rothschild, M., and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, 629–650.

Siow, A. (Forthcoming): "Testing Becker's Theory of Positive Assortative Matching," *Journal of Labor Economics*.

Spence, M. A. (1973): "Job Market Signalling," *Quarterly Journal of Economics*, 87(3), 355–374.

Su, Z., W. Zeng, R. Shi, Y. Wang, J. Sun, and X. Gu (2013): "Area Preserving Brain Mapping," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 2235–2242. IEEE.

Torgovitsky, A. (2013): "Minimum Distance from Independence Estimation of Nonseparable Instrumental Variables Models," *Mimeo.*

——— (2014): "Identification of Nonseparable Models Using Instruments with Small Support," *Forthcoming Econometrica.*

Wilson, R. (1993): *Nonlinear Pricing*. Oxford University Press.