

IDENTIFYING A SCREENING MODEL WITH MULTIDIMENSIONAL PRIVATE INFORMATION*

GAURAB ARYAL[†]

ABSTRACT. In this paper I study the nonparametric identification of screening models when consumers have multidimensional private information. In particular, I consider the model developed by [Rochet and Choné \(1998\)](#) and determine conditions to identify the cost function, the joint density of types and the utility functions with data (on demand and prices) from only one market. When the utility function is nonlinear the model cannot be identified, but with an exogenous binary cost shifter the model is identified. 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.

1. INTRODUCTION

In this paper I study the identification of a screening (nonlinear pricing) model where consumers have multidimensional private information. Using the optimality conditions for both demand and supply (optimal contract) I determine conditions under which we can (or cannot) identify the cost function, the joint density of consumers' multivariate types and the utility function. 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. I also determine the empirical content of the model by characterizing all testable implication of the model on the data. The identification result is robust with respect to the classical measurement error in prices and (type independent) unobserved heterogeneity.

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 substantial loss of (social) welfare.

Date: First Version: November 1, 2014. This Version: Jan 7, 2015.

* Thanks to be added.

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

This belief has guide shape policy and regulatory choices, see [Baron \(1989\)](#); [Joskow and Rose \(1989\)](#); [Laffont \(1994\)](#). The empirical support for this belief, however, has 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 (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/types that cannot be sorted out in a satisfactory manner according to only one of these characteristics. And, ignoring multidimensionality will lead to incorrect conclusion(s). For instance [Chiappori and Salanié \(2000\)](#) show insurees with high coverage do not file more claims as the theory would predict. This could be because the theory assumes that insurees differ only in their risk, and not risk preference, but insurees differ in both risk and risk aversion as in [Finkelstein and McGarry \(2006\)](#); [Cohen and Einav \(2007\)](#). Then those who buy higher coverage could include those with high risk aversion but low risk. Whether the effect of risk aversion prevails over the effect of risk will depend on their joint density, and [Aryal, Perrigne, and Vuong \(2010\)](#) show it can be nonparametrically identified. Thus, it stands to reason that ignoring multidimensionality will lead to misspecification error.

The focus of this paper is on other environment where multidimensional private information is equally important: a monopoly seller who sells multiple (differentiated) products to heterogeneous consumers. If consumers have different taste for each product and if those tastes are known only to them, most likely in reality, then it leads to multidimensional private information. Similarly, environments where a seller sells a differentiated product with multiple attributes or a principal contracts with several ex-ante identical applicants on multiple tasks, are characterized by multidimensional private information.

[Crawford and Yurukoglu \(2012\)](#) study cable television market, where consumers are heterogeneous with respect to their taste for each channel and they choose from multiple packages (of channels), which can be captured better by multidimensional preferences than one dimensional.¹ Furthermore, since the observed bundles (menu) will depend on the joint distribution of the taste profile, it is equally important to endogenize the supply side as well. Another reason

¹Multidimensionality also arises with competition, as in [Ivaldi and Martimort \(1994\)](#); [Aryal \(2013\)](#). For an excellent survey on this topic see [Rochet and Stole \(2003\)](#).

why we should consider supply side is as follows: 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 (multiproduct nonlinear pricing).

Before we can estimate such a model, we have to be sure that it is, or it is not, identifiable. In this paper I address that issue by determining 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. In so doing, the paper contributes to the research on empirical mechanism design as articulated by [Chiappori and Salanié \(2003\)](#). Since the model is also complicated, it is desirable to know what are the testable restrictions, if any, imposed by the model on the data. Such restrictions shed light on the empirical content of the model, which allow us to design specification tests that can then be used to falsify the model. Keeping this motivation in mind, I also derive all such testable restrictions.

Both these results – identification and testable restrictions– will also be useful for empirical analysis of optimal income taxation for couples. If marriages are assortative, as posited by [Becker \(1973, 1974\)](#) and verified by [Siow \(Forthcoming\)](#), then a unit of observation would be a household with (at least) two dimensional type (couple's productivity parameters), which makes the problem of optimal taxation that of multidimensional screening.² See [Kleven, Kreiner, and Saez \(2009\)](#) for one way to solve this problem. In this paper, however, I

² If preferences for children are heterogenous and private, and since child bearing and raising decisions interact with labor supply decision, we would need more than two dimensional type to index a household.

will consider the problem from the point of view of multiproduct monopolist, and leave the problem of couples' taxation for future research.

In a multidimensional screening model, a seller offers a menu (pair of bundles and prices) of multiple differentiated products to consumers with (unobserved) heterogeneous taste for each product that maximizes her expected profit with respect to a known joint density of taste parameters and the cost function. Any observed consumer (socioeconomic) covariates 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) high-types, who are perfectly screened – each type is offered a unique bundle; b) medium-types, who are further divided into different indices such that everyone with the same index is offered the same bundle, i.e., are bunched together according to this index; and c) low-types who are excluded from the market. I consider each of these sub-groups separately, and each with linear, bi-linear and nonlinear utility. Considering progressively more general form of utility provides an opportunity to highlight the role each additional assumption plays in identification.

The identification follows the following logic. First, consider the high-types with linear utility. In equilibrium the seller must satisfy the incentive compatibility condition, which implies that the allocation rule (mapping from type space to menu) is bijective, so that the marginal utility is equal to the marginal price. Since the marginal utility is type and marginal price is directional derivative of the (observed) pricing function the (truncated) joint density is identified. We can then change the variables to express the equilibrium – Euler – condition only in terms of observables and the unknown cost function, which is identified as the unique solution of this nonlinear partial differential equation. Furthermore, like in [Fox and Gandhi \(2013\)](#) if the cost function is a real analytic (which includes all polynomials, exponential and trigonometric functions), the cost function is identified everywhere.

This identification strategy fails for medium-types who are bunched when the utility is linear. Nonetheless, we are interested in learning more about these types. If the utility function is bi-linear in products and consumer characteristics that are independent of the unobserved taste parameters then I show that the conditional density of choices given consumer characteristics can be written as a Radon transform, (see [Helgason, 1999](#)), of the (truncated) joint density of (medium) types. Since the former is known (or estimable from the data) and

Radon transform is invertible, the later is identified. which is invertible.³ Next I consider the case with nonlinear utility functions.

First, I show that even with the high-types the utility function is not identified, because the curvature of the utility function and the density of the types are substitutable. However, if there is an exogenous *binary* cost shifter that affects the menus 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. Once the quantile function of types is identified, we can identify the utility function from consumers' optimality condition. The cost function is identified from the equilibrium Euler condition as before.

So far I have not used variation in consumer characteristics. 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 the identification is robust with respect to the classical measurement error in prices and unobserved heterogeneity that is independent of types. If on the other hand, consumers' choices are measured with error then the model cannot be identified. Furthermore, I also determine the empirical content of the model. In particular I characterize all testable implication of the model (linear, bilinear and nonlinear utility) separately on the data.

This paper also is closely related to [Perrigne and Vuong \(2011\)](#) who were the first to study 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

³ For other applications of Radon transformation see [Gautier and Hoderlein \(2012\)](#); [Hoderlein, Nesheim, and Simoni \(2013\)](#).

models; and to [Aryal, Perrigne, and Vuong \(2010\)](#) who show how we can non-parametrically 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 \(1994\)](#) and [Berry, Gandhi, and Haile \(2013\)](#).

The remainder of the paper is organized as follows. Section 2 describes the theoretical model; section 3 describes the identification of the all three models, while section 4 provides the rationalizability lemmas for the models. Finally section 5 extends identification arguments with measurement error and unobserved heterogeneity in the data.

2. 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 $P : Q \rightarrow \mathbb{R}_+$, 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 $C : \mathbb{R}_+^{d_q} \times \mathcal{S}_Z \rightarrow \mathbb{R}_+$ be the cost function with the interpretation that $C(\mathbf{q}, Z)$ is the cost of producing \mathbf{q} when the market is characterized by exogenous (cost shifter) $Z \in \mathcal{S}_Z := \{z_1, z_2\}$.⁴ The objective of the principal is to choose a set Q and a transfer function $P(\cdot)$ that maximizes her expected profit when the cost function is given by $C(\cdot, Z)$ and the CDF $F_\theta(\cdot)$. For notational simplification I

⁴ Z is as a binary seller or market specific technology Z that only affects the cost of production.

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) $\theta \stackrel{i.i.d.}{\sim} F_\theta(\cdot)$ which has a square integrable density $f(\cdot) > 0$ a.e. on \mathcal{S}_θ .
- (iii) The net utility be an element of a Sobolev space

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

with the norm $|V| = \left(\int_{\mathcal{S}_\theta} \{V^2(\theta) + \|\nabla V(\theta)\|^2\} d\theta \right)^{\frac{1}{2}}$.

(iv) The gross utility $u(\mathbf{q}, \theta, X) := \theta \cdot v(\mathbf{q}, X)$, where $v(\cdot, X) : \mathbb{R}_+^J \rightarrow \mathbb{R}_+^J$ is a vector $(v_1(\cdot), \dots, v_J(\cdot))$ 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) $v_j(q_j, X) = q_j$.
- (iv-b) $X = (X_1, X_2)$ such that $X_1 \in \mathcal{S}_x \subset \mathbb{R}^{d_{x_1}}$, $d_{x_1} = J$ and $v_j(q_j, X) = q_j \cdot X_{1j}$.
- (iv-c) $v_j(q_j, X) = X_{1j} \cdot v_j(q_j, X_2)$ 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 $\lim_{q \rightarrow \infty} v_j(q) = \infty$.
- (v) $C(\cdot, Z)$ 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.⁵ 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,

⁵ See Ivaldi and Martimort (1994); Aryal, Perrigne, and Vuong (2010); Aryal (2013) for identification and estimation of screening models with bunching.

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. Until the section on nonlinear utility, I will suppress the dependence of cost on Z .

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

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

and individual rationality (IR) condition: $U(\theta) \geq U_0 := \theta^T v(\mathbf{q}_0) - P_0$. 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. $Q \ni \mathbf{q}_0$.⁶ The principal chooses a feasible menu $(Q, \mathbf{q}(\cdot), P(\mathbf{q}))$, that maximizes expected profit

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

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

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

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

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

In an important paper [Rochet \(1987\)](#) showed that under Assumption 1, a menu $\{Q, \mathbf{q}(\cdot), P(\cdot)\}$ is such that $U(\theta)$ solves Equation (1) (satisfies IC) if and only if: (i) $\mathbf{q}(\theta) = v^{-1}(\nabla U(\theta))$; and (ii) $U(\cdot)$ is convex on Θ . This means that choosing an optimal contract $\{Q, \mathbf{q}(\cdot), P(\cdot)\}$ 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 $\mathbf{q}(\theta) = v^{-1}(\nabla U(\theta))$. So the principal chooses $U(\theta) \in H^1(\mathcal{S}_\theta)$ to maximizes

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

subject to IC and IR constraints.

⁶ 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.

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 $\mathbf{q}^*(\cdot)$ 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}_θ^0 who are screened out and offered only $\{\mathbf{q}_0\}$, the medium-types \mathcal{S}_θ^1 who are bunched and offered “medium type” of bundles and the high-types \mathcal{S}_θ^2 who are perfectly screened. So, $\mathbf{q}^*(\cdot)$ is injective only when restricted to \mathcal{S}_θ^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**)⁷ 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

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

where $d\sigma(\theta)$ is the Lebesgue measure on the boundary $\partial \mathcal{S}_\theta$, $\hat{n}(\theta)$ is an outward normal and $d\mu(\theta) := \alpha(\theta) d\theta + \beta(\theta) d\sigma(\theta)$. 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*

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

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

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}_θ^1 where the optimal allocation rule \mathbf{q}^* 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. $\mathcal{S}_\theta^1(q) = \{\theta \in \Theta : \mathbf{q}^*(\theta) = q\} = \{\theta \in \Theta : U^*(\theta) = \theta \cdot q - P(q)\}$. If $U^*(\theta)$ is convex for all θ , that is if the

⁷ For a vector field $W = (w_1, w_2)$ the divergence of W is defined as $\mathbf{div}W = D_1 w_1 + D_2 w_2$ which for $(x - y)$ coordinates is $(\mathbf{div}W)(x, y) = \frac{\partial w_1}{\partial x} + \frac{\partial w_2}{\partial y}$; see Lang (1973).

global incentive compatibility constraint is satisfied then there is no bunching, in which case \mathcal{S}_θ^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 2.1. *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}_θ^2 such that:*

- (1) *A positive mass of types \mathcal{S}_θ^0 do not participate because $U^*(\theta) = U_0(\theta)$. This set is characterized by $\mu(\mathcal{S}_\theta^0) = 1$, i.e. $\int_{\mathcal{S}_\theta^0} \alpha(\theta) d\theta + \int_{\partial \mathcal{S}_\theta^0} \beta(\theta) d\theta = 1$.*
- (2) *\mathcal{S}_θ^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 \mathbf{q} , U^* is affine. μ restricted to $\mathcal{S}_\theta^1(\mathbf{q})$ satisfies: $\int_{\mathcal{S}_\theta^1(\mathbf{q})} d\mu(\theta) = 0$ and $\int_{\mathcal{S}_\theta^1(\mathbf{q})} \theta d\mu(\theta) = 0$.*
- (3) *\mathcal{S}_θ^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}_θ^0 and get the outside option q_0 ; those who are bunched \mathcal{S}_θ^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}_θ^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 $\mathbf{q}^*(\cdot)$ is continuous.

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

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

3. IDENTIFICATION

In this section we study identification of the distribution of types $F_\theta(\cdot)$ and the cost function $C(\cdot; Z)$ 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 $i \in [N] := \{1, \dots, N\}$. On the principal side, we observe the menu of contracts $\{Q_{X,Z}, P_{X,Z}(\cdot)\}$

offered to each consumer with characteristics X and when the cost shifter is Z . I assume that these observables are distributed *i.i.d* with respect to $\Psi_{P,\mathbf{q},X,Z}(\cdot, \cdot, \cdot, \cdot)$. Since the consumer characteristics and the cost shifters are assumed to be mutually independent, i.e. $X \perp\!\!\!\perp Z$, the joint CDF $\Psi_{P,\mathbf{q},X,Z}(\cdot, \cdot, \cdot, \cdot) \equiv \Psi_{P,\mathbf{q}|X,Z}(\cdot, \cdot | \cdot, \cdot) \times \Psi_Z(\cdot) \times \psi_X(\cdot)$ is identified (or equally, estimated) from the data, and henceforth is treated as known.

The seller draws $Z \in \mathcal{S}_Z := \{z_1, z_2\} \sim \Psi_Z(\cdot)$ and offers $(Q_{(x,z)}, P_{(x,z)}(\cdot))$ 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 $\mathbf{q}_i \in Q_{(x,z)}$ 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 $\mathbf{q}^*(\cdot) : \mathcal{S}_\theta \mapsto Q_{(x,z)}$ and pricing function $P_{x,z}(\cdot) : Q_{(x,z)} \mapsto \mathbb{R}_+$, such that $\mathbf{q}_i = \mathbf{q}^*(\theta_i)$ and $P_i = P_{x,z}(\mathbf{q}^*(\theta_i))$. Henceforth, $\mathbf{q}(\cdot)$ 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 econometrics model:

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

The model parameters $[F_\theta(\cdot), C(\cdot, Z)]$ are identified if for any different parameters $[\tilde{F}_\theta(\cdot), \tilde{C}(\cdot, Z)]$, the implied data distributions are also different, that is $\Psi_{P,\mathbf{q},X,Z}(\cdot, \cdot, \cdot, \cdot) \neq \tilde{\Psi}_{P,\mathbf{q},X,Z}(\cdot, \cdot, \cdot, \cdot)$. Since $\Psi_Z(\cdot)$ and $\Psi_X(\cdot)$ are identified, identification boils down to finding conditions under which Equation (7) are invertible. Moreover, the model has unique equilibrium which means there is only one relevant $\Psi_{P,\mathbf{q},X,Z}(\cdot, \cdot, \cdot, \cdot)$.

Following the equilibrium characterization, I consider the three subsets of types separately. For every pair $(X = x, Z = z)$, let $Q_{(x,z)}^j$ be the set of choices made by consumers with type $\theta \in \mathcal{S}_\theta^j$ 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 $Q_{(x,z)}^k$ to identify the model parameters restricted to \mathcal{S}_θ^j , beginning with the subset $Q_{(x,z)}^2$. The allocation rule $\mathbf{q}(\cdot)$ is one-to-one when restricted to \mathcal{S}_θ^2 , and hence its inverse $(\mathbf{q})^{-1}(\cdot)$ exists on $Q_{(x,z)}^2$, but not when restricted to \mathcal{S}_θ^1 because of bunching, and as a consequence the identification strategies are

different.⁸ In what follows, I suppress the dependence on X and Z , except for the cost function, where Z enters as an argument, until otherwise.

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

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

where $det(\cdot)$ is the determinant function. Using the short hand notation C_z for $C(\cdot, Z)$ we get

$$M^*(\mathbf{q}) = \Pr(\mathbf{q}(\theta, F_\theta, C_z) \leq \mathbf{q}) = (F_\theta \circ (\mathbf{q}_2)_{C_z}^{-1})(\mathbf{q}), \quad (9)$$

which will be a key relationship for identification. Let \mathcal{A} and \mathcal{B} be two Banach spaces, that includes the model parameters and define a mapping $\Gamma : \mathcal{A} \rightarrow \mathcal{B}$ with $\Gamma = (\Gamma_1, \Gamma_2)$ to write Equations (1) and (9) succinctly as

$$\begin{aligned} \Gamma_1(F_\theta, C_z) &\equiv (\nabla P - \tilde{\theta})(\cdot) \\ \Gamma_2(F_\theta, C_z) &\equiv (M^* - F_\theta \circ (\mathbf{q}_2)_{C_z}^{-1})(\cdot). \end{aligned}$$

Written this way, it is immediate to see that $\Gamma(\cdot, \cdot) = 0$ is our (infinitely many) moment restrictions that can be used for identification. As mentioned earlier, we say that the model is identified if Γ is invertible at 0, for which we would have to verify the conditions of the Banach inverse function theorem, see [Luenberger \(1969\)](#). If we use $\zeta_0 = (F_\theta^0(\cdot, \cdot, \mathcal{S}_\theta^2), C^0(\cdot; Z)) \in \mathcal{A}$ to denote the true parameter such that $\|\Gamma(\zeta_0)\|_{\mathcal{B}} = 0$, then we say that ζ_0 is “locally identified” in a neighborhood $\mathcal{N} \subset \mathcal{A}$ if $\|\Gamma(\zeta)\|_{\mathcal{B}} > 0$ for all $\zeta \in \mathcal{N}$ such that $\zeta \neq \zeta_0$.

⁸ A fundamental problem in Analysis is the existence and/or uniqueness of the solutions to the equation $\mathbf{q} = \mathbf{q}(\theta)$ in the unknown θ . The existence of inverse $(\mathbf{q})^{-1}(\cdot)$ follows from the observation that $U^*(\theta)$ is convex on \mathcal{S}_θ^2 and $DU^*(\theta) = \mathbf{q}^*(\theta)$, see [Kachurovskii \(1960\)](#). Also see [Parthasarathy \(1983\)](#); [Fujimoto and Herrero \(2000\)](#) on global univalence.

If this local neighborhood is indeed the entire set \mathcal{A} , i.e. $\mathcal{N} = \mathcal{A}$, then we say that the model is globally identified. But to apply the Banach inversion theorem we need to assume that for all $\theta \in \mathcal{S}_\theta^2$, $\nabla_{\mathbf{q}_2}(\cdot; C_Z, F_\theta^0)$ is nonsingular. If the Fréchet derivative $D\Gamma$ is one-to-one, or if the null space of $D\Gamma$ is $\{\mathbf{0}\}$ vector around a neighborhood of ζ_0 , and has bounded inverse, then from Banach Inverse function theorem implies that Γ is invertible and hence the model is locally identified. [Chen, Chernozhukov, Lee, and Newey \(2014\)](#) showed that this injective condition is equivalent to the rank condition—there is a set \mathcal{N}' such that $\|D\Gamma(\zeta - \zeta_0)\|_{\mathcal{B}} > 0$ for all $\zeta \in \mathcal{N}'$ with $\zeta \neq \zeta_0$ —that is intuitive and widely used (at least for parametric model). In particular, if $\Gamma(\zeta)$ is Fréchet differentiable at ζ_0 , and the rank condition is satisfied for $\mathcal{N}' = \mathcal{N}_\epsilon$, for some $\epsilon > 0$, and $D\Gamma : \mathcal{A} \mapsto \mathcal{B}$ is **onto**, then ζ_0 is identified on $\mathcal{N}_{\epsilon'}$ with $0 < \epsilon' \leq \epsilon$.

There are three immediate difficulties in pursuing this line of identification. First, we know that $D\Gamma$ is onto if and only if it is continuous (Proposition 1 in Chapter 6, [Luenberger \(1969\)](#)), but we know that nonlinear integral equations are often ill-posed, [Carrasco, Florens, and Renault \(2007\)](#). Second, verifying that $D\Gamma$ is one-to-one is difficult in our environment because of the nonlinearity. Third, the assumption that $q(\cdot)$ is nonsingular is an untestable condition imposed on the equilibrium outcome and not on the model primitives. So, it is difficult to show that the problem is not ill-posed or at least only mildly ill-posed (Chapter 2, [Engl, Hanke, and Neubauer \(1996\)](#)) and find a simpler and transparent way to verify the mapping is injective, we abandon this intuitive and direct strategy of starting with local identification before extending it to global identification in favor of an indirect approach. Let $F_\theta(\cdot|j)$ be the CDF $F_\theta(\cdot)$ restricted to be in the set \mathcal{S}_θ^j and let N_j be the set of consumers who buy $\mathbf{q} \in Q^j$, for $j = 0, 1, 2$.

3.1. 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 3.3). As $\tilde{\theta}(\cdot)$ restricted to \mathcal{S}_θ^2 is bijective we can identify

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

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 $\int_{S_\theta^2} f_\theta(t) dt$ we get

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

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):

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

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. *The cost function $C : Q \rightarrow \mathbb{R}$ is a real analytic function at $\mathbf{q} \in Q$, i.e., $\exists \delta > 0$ and open ball $B(\mathbf{q}, \delta) \subset Q$, $0 \leq r < \delta$, $\sum_{k_1, \dots, k_J} |a_{k_1, \dots, k_J}| r^{k_1 + \dots + k_J} < \infty$*

$$C(\mathbf{q}) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_J=0}^{\infty} a_{k_1, \dots, k_J} (q_1 - p_1)^{k_1} \cdots (q_J - p_J)^{k_J}, \mathbf{q} \in B(\mathbf{q}, \delta).$$

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 .⁹ Since the cost function is completely unspecified, besides convexity, the fact that we need analyticity is not surprising. Similar idea has

⁹ If two real analytic convex functions coincide on an open set Q^2 , then they coincide on any connected open subset of Q which has nontrivial intersection with Q^2 .

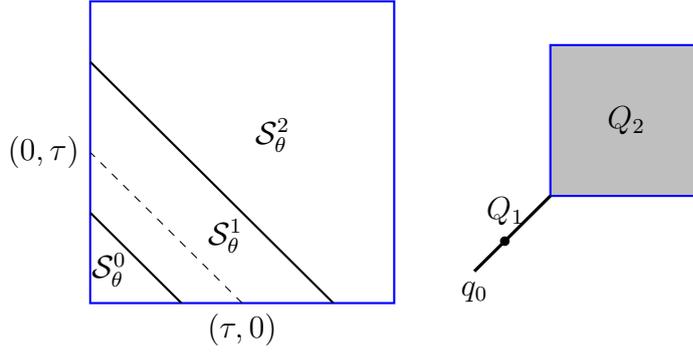


FIGURE 1. Optimal Product Line

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 just like analyticity; and to random coefficient in a multinomial choice model [Fox and Gandhi \(2013\)](#), like here, assume the utility function is real analytic. This result is formalized below.

Theorem 3.1. *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 $\mathbf{q}(\cdot)$ is the key to identification, and since we lose monotonicity on \mathcal{S}_θ^1 we lose identification, as shown in the example below.

Example 3.1. 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}_θ^0 , $U^*(\theta) = 0$; (ii) in the bunching region \mathcal{S}_θ^1 , U^* depends only on $\theta_1 + \theta_2$; and (iii) in the perfect screening region \mathcal{S}_θ^2 , U^* is strictly convex.

On \mathcal{S}_θ^0 , $\mathbf{q}^*(\theta) = 0$, 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}_θ^0 and \mathcal{S}_θ^1 is a linear line $\tau_0 = \theta_1 + \theta_2$, where $\tau_0 = \frac{\sqrt{6}}{3}$. On \mathcal{S}_θ^1 , $\mathbf{q}(\theta) = (q_1(\theta), q_2(\theta)) = (q_b(\tau), q_b(\tau))$, 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 $q_1(\tau) = q_2(\tau) = q_b(\tau)$. So $\alpha(\theta) = 3 - 2cq_b'(\tau)$ and on $\partial\mathcal{S}_\theta^1$, $\beta(\theta) = (cq^*(\theta) - \theta) \cdot \hat{n}(\theta) = -cq_b(\tau)$. Sweeping conditions are satisfied if $\alpha(\theta) \geq 0$ and $\beta(\theta) \geq 0$ and on each bunch

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

which can be used to solve for q_b as $q_b(\tau) = \frac{3\tau}{4c} - \frac{1}{2c\tau}$. 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}_θ of \mathbf{q}^* , i.e. $q_b(\tau_1) = 0$. Now, define $\tau = \mathbf{q}_b^{-1}(\mathbf{q})$ 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 $f_\theta(\cdot|1)$.

3.2. 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

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

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 3.1 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}_θ^1 , 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 $\mathbf{q}(W)$, i.e. $\mathbf{q}(\theta) = (q_1(\theta), \dots, q_J(\theta)) = (q_1(W), \dots, q_J(W))$ for all $\theta \in \mathcal{S}_\theta^1$. 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 $W := (\mathbf{q})^{-1}(\mathbf{q})$. Then, the identification problem is to recover $f_\theta(\cdot)$ from the the joint density $f_{W,X}(\cdot, \cdot)$ of (W, X) when

$$W = \theta_1 X_{11} + \dots + \theta_J X_{1J}.$$

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

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

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, \dots, a_J) . Then we cannot identify $f_\theta(\cdot)$ from $B = a_1 \theta_1 + \dots + a_J \theta_J$.

Let

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

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

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

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:

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

Theorem 3.2. *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}_θ^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.

3.3. Nonlinear Utility. In this section I consider the model with nonlinear utility, i.e. Assumption 1-(iv-c). The model parameter is the triplet $F_\theta(\cdot|\cdot)$, $C(\cdot; Z)$ and $v(\cdot, X_2)$. I will begin with the case where I fix X_2 at some arbitrary value x_2 . Essentially this is like assuming the utility function $X_1 \cdot v(\cdot, X_2) = v(\cdot)$ so X does not enter the utility.

Compare to previous two cases we have one additional (vector) function to identify, so the two optimality conditions Equations (1) and (9) are insufficient.

Lemma 2. *Under Assumptions 1-(i)-(iv-c) and (v) the model $\{F_\theta(\cdot|2), C(\cdot), v(\cdot)\}$, where the domain of the cost and utility functions are restricted to be Q^2 and \mathcal{S}_θ^2 , 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_2) is

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

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 $\check{v}(q_j) = v(q_j)/\omega = q_j^{\omega_j}/\omega_j$, be a new model. It is easy to check that $\{q_j, P_j\}$ solves the first-order condition implied by $[\check{v}(\cdot), F_{\check{\theta}}(\cdot)]$, and the joint (truncated) density of (q_1, q_2) is

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

□

Identification fails because the type θ (identified from the prices) and the curvature of the utility function are substitutable. This substitutability would break if there is a factor (an exogenous cost shifter) that only affects the cost (and hence the allocation rule and the prices) but not the type or the utility function. As this factor takes different (discrete) values the price function would change and hence the implied types, but the utility function would remain the same. 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. To highlight the intuition about how such discrete cost shifter is sufficient for identification I will begin by re-examining the example used in Lemma 2 before formalizing the proof.

Recall that the cost function $C(\cdot, Z)$ depends on a random variable Z , and now let $Z \in \mathcal{S}_Z \equiv \{z_1, z_2\}$. Henceforth, I will use the subscript $\ell \in \{1, 2\}$ in $\{P_\ell(\cdot), \mathbf{q}_\ell(\cdot)\}$ to denote the price function and allocation rule when $Z = z_\ell$. As in Lemma 2, the utility function is $v(q_1, q_2) = \begin{pmatrix} v_1(q_1) \\ v_2(q_2) \end{pmatrix} = \begin{pmatrix} (q_1)^{\omega_1} \\ (q_2)^{\omega_2} \end{pmatrix}$. As before let us focus only on the high-types \mathcal{S}_θ^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

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

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 $Z = z_2$.¹⁰

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

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

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

$$M_1^*(\mathbf{q}_1(t)) = M_2^*(\mathbf{q}_2(t)).$$

¹⁰ Note that this analysis is independent of the utility function, and uses the fact that Q^2 is invariant to Z .

Hence, the (multivariate) quantiles of the choice distribution when $Z = z_1$ are equal to those when $Z = z_2$, i.e.¹¹

$$\mathbf{q}_1(t) = (M_1^*)^{-1}[M_2^*(\mathbf{q}_2(t))], \quad (12)$$

and since $(M_1^*)^{-1} \circ M_2^*(\cdot)$ is identified, we can identify $\mathbf{q}_2(\theta)$ if we know $\mathbf{q}_2(\theta)$. Therefore, the difference, $((M_1^*)^{-1} \circ M_2^*(\mathbf{q}(\tau)) - \mathbf{q}(\tau))$, 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 $\nabla v(\mathbf{q})$ as a (vector valued) function that solves $\nabla P(\mathbf{q}) = \tilde{\theta}(\mathbf{q}) \circ \nabla v(\mathbf{q})$.

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).¹² 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 $\{\theta^0, \theta^1, \dots, \theta^L, \dots\}$ and the corresponding quantile. If these sequence form a dense subset of Q^2 then the function $\tilde{\theta}(\cdot) : Q^2 \times \mathcal{S}_Z \rightarrow \mathcal{S}_\theta^2$ is identified everywhere. I formalize this intuition for $J \geq 2$ below, starting with the assumption about exclusion restriction.

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

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

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

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

$$M_\ell^*(\mathbf{q}_\ell(t)) := M^*(\mathbf{q}(t; z_\ell); z_\ell) = M^*(\mathbf{q}(t; z_{\ell'}); z_{\ell'}) := M_{\ell'}^*(\mathbf{q}_{\ell'}(t)). \quad (14)$$

¹¹ 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.

¹² I use the superscript here to index the sequence of bundles, not be confused with the utility function $v_j(q_j) = (q_j)^{\omega_j}$, similarly for the superscript on θ .

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 corresponding 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

$$\mathbf{q}_1(\tau) = (\partial g_{M_1^*})^{-1}(M_2^*(\mathbf{q}_2(\tau))) := s_{2,1}(\mathbf{q}_2(\tau)), \quad \tau \in (0, 1). \quad (15)$$

This means we can then use

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

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

$$\mathbf{q}_k(r_{\ell,\ell}(\theta, \mathbf{q})) = s_{\ell,\ell}(\mathbf{q}_{\ell'}(\theta)); \quad (16)$$

so if we know $\mathbf{q}_{\ell'}(\cdot)$ at some θ then we can identify $\mathbf{q}_\ell(\cdot)$ 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 $\{\mathbf{q}^0, \theta^0 = \tilde{\theta}_1(\mathbf{q}^0)\}$.¹³ 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 $\mathbf{q}(\tau) \in Q^2$ and identify $\tilde{\theta}(\mathbf{q}(\tau))$, possibly by constructing a sequence as above.

¹³ We can also normalize some quantile of $F_\theta(\cdot)$.

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 $\text{sgn}[(r_{\ell,\ell'}^j(q_j) - q_j)(q_j - \hat{q}_j)]$ is independent of $j \in \{1, \dots, J\}$.*

Both the components of this assumptions are testable and can be easily verified in the data. 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 $\nabla P_1(\mathbf{q}) \circ \nabla P_2(\mathbf{q})^{-1} \ll 1$, whenever $\mathbf{q} \ll \hat{\mathbf{q}}$. Then, for τ^{th} quantile $\mathbf{q}(\tau) < \hat{\mathbf{q}}$:

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

where the first equality is simply the definition, the second equality is the normalization, the third equality follows from (16) with $\mathbf{q}^1 := s_{1,2}(\mathbf{q}^0 = \mathbf{q}(\tau))$ 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 $\lim_{L \rightarrow \infty} s_{1,2}(\mathbf{q}^L) = s_{1,2}(\mathbf{q}^\infty) = s_{1,2}(\hat{\mathbf{q}})$; and d) $\tilde{\theta}_1(\hat{\mathbf{q}}) = \theta_0$. Since the quantile τ was arbitrary, we identify $\tilde{\theta}_1(\cdot)$.¹⁴

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

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

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

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

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

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

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

Theorem 3.3. *Under Assumptions 1-(i)-(iv-c) and (v) and Assumptions 2-5, $[F_\theta(\cdot|2), v(\cdot), C(\cdot; Z)]$ are identified.*

To identify the density $f_\theta(\cdot|1)$ we can use Theorem 3.2, 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 $v(\cdot, X_2)$ we need to be able to extend the utility function from Q^2 to $Q^2 \cup Q^1$. For the identification strategy then if $v(\cdot)$ is a real-analytic, like the cost function, then we can extend the domain of $v(\cdot)$ to include Q^1 .

¹⁴ Some other examples where identification relies on constructing some sequence include Guerre, Perrigne, and Vuong (2009); Aryal and Kim (2014); Torgovitsky (2014), among others. In a related paper D'Haultfœuille and Février (2014) use powerful results from the theory of Group of Circle Diffeomorphisms, Navas (2011), to identify a nonseparable model with discrete instrument. As a consequence it is possible to view the constructive proof here as a special case in that paper. Again, I am thankful to Xavier D'Haultfœuille for pointing this out.

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

Then under Assumption 6, we can change the unit of measurement from \mathbf{q} to $\tilde{\mathbf{q}} \equiv v(\mathbf{q}, X_2)$, then apply Theorem 3.2 with gross utility as $\sum_{j \in [J]} \theta_j X_j \tilde{q}_j$.

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}_θ^2 , where $\mathbf{q} = \mathbf{q}(\theta, X, z_k) := \mathbf{q}_k(\theta, X)$. 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 $\mathbf{q}(\cdot, X)$. Let $\mathcal{L}(\mathcal{S}_\theta^2, Q^2)$ be the set of joint distribution defined as

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

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}_θ^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 $\mathbf{q} = \mathbf{q}(\theta; X)$, 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 $dL(\mathbf{q}, \theta) = dM_{\mathbf{q}|X}^*(\mathbf{q})\delta[\theta =$

$\nabla_{\mathbf{q}}\Gamma(\mathbf{q}, X)$] 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 $F_{\theta}(\cdot|2)$. \square

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 $\mathbf{q}(\theta; X)$. The result is only theoretical: it only guarantees that a unique function $\Gamma(\mathbf{q}, X)$ exists, but is silent about finding it.¹⁵

4. 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 $P = P_k(\mathbf{q}, z_k)$ and $\mathbf{q} = \mathbf{q}_k(\theta, z_k)$. 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

$$\mathcal{M}_1 = \{(F_{\theta}(\cdot), C(\cdot)) \in \mathcal{F} \times \mathcal{C} : \text{satisfy Assumption 1} - (i) - (iv - a), (v)\}$$

$$\mathcal{M}_2 = \{(F_{\theta}(\cdot), C(\cdot)) \in \mathcal{F} \times \mathcal{C} : \text{satisfy Assumption 1} - (i) - (iv - b), (v)\}$$

$$\mathcal{M}_3 = \{(F_{\theta}(\cdot), C(\cdot, Z)) \in \mathcal{F} \times \mathcal{C}_Z : \text{satisfy Assumptions 1} - (i) - (iv - c), (v), \mathbf{3} \text{ and } \mathbf{4}\}$$

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)

¹⁵ Su, Zeng, Shi, Wang, Sun, and Gu (2013) provide one way to compute the function.

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 the density of $M^*(\cdot)$

C5. Let $C(\cdot)$ be the solution of the differential equation

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

with boundary conditions

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

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 4.1. *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}_θ and the equilibrium allocation rule $\mathbf{q} : \mathcal{S}_\theta \rightarrow Q$ 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 $\mathbf{q} : \mathcal{S}_\theta^1 \rightarrow Q^1$ 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, $M^*(q) = \Pr(\mathbf{q} \leq q) = \Pr(\nabla P(\mathbf{q}) \leq \nabla P(q)) = \Pr(W \leq w) = F_W(w)$, 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 $F_\theta(\cdot|2) = \int_{\theta \in \{W := \nabla P(\mathbf{q}), \mathbf{q} \in Q^2\}} f_\theta(\theta|2) d\theta$. 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 $\theta = \nabla P(\mathbf{q})$. \square

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 4.2. *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 4.2, 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.

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

C4'. If $\mathbf{q}_\tau(X_2, Z)$ is the $\tau \in [0, 1]$ quantile of $\mathbf{q} \in Q_{X_2, Z}^2$ then $\mathbf{q}_\tau(\cdot, z_1) = \mathbf{q}_\tau(\cdot, z_2)$.
 C6. The truncated distribution of choices $M_{\mathbf{q}|X, Z}^*(\cdot|\cdot, \cdot)$ has finite second moment, and for a given $Z = z_k$ (henceforth suppressed) the solution of

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

where $\mathcal{L}(Q^2, \mathcal{S}_\theta^2)$ 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 4.3. *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, $\dim(X_1) = \dim(\mathbf{q}) = J$, $Q_{X,z_k}^2 = Q_{X,z_{k'}}^2$ (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}_θ . Moreover, the equilibrium allocation rule $\mathbf{q} : \mathcal{S}_\theta \times X \times Z \rightarrow Q$ 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 $Q = Q_{(X,Z)}^2 \cup Q_{(X,Z)}^1 \cup \{\mathbf{q}_0\}$ the (truncated) density $m_{\mathbf{q}|X,Z}(\mathbf{q}|\cdot, \cdot) > 0$ for all $\mathbf{q} \in Q_{(X,Z)}^2 \cup Q_{(X,Z)}^1$. 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 $\mathbf{q}(\mathcal{S}_\theta^1; X, Z) = Q^1 \subsetneq \mathbb{R}_+^J$ 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 $v(\cdot; X_2)$ 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 $Z \perp \theta$, using Equation (14) gives $F_\theta(\xi|2) = M_{\mathbf{q}|X,Z}(\mathbf{q}_k(\xi)|X, z_1) = M_{\mathbf{q}|X,Z}(\mathbf{q}_k(\xi)|X, z_2)$, 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 $F_\theta(\cdot)$ 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 $Q_{X,Z}^2$, 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 $\mathbf{q} \in Q_{X,Z}^1 \cup \{\mathbf{q}_0\}$. Using the exclusion restriction and (17) for all $\mathbf{q} \in Q_{X,Z}^2$ we can determine the function $\tilde{\theta}(\mathbf{q}_\tau; Z = z_k)$ along a set $\tilde{Q}_{X,Z}^2 \subseteq Q_{X,Z}^2$ for $k = 1, 2$. If the set $\tilde{Q}_{X,Z}^2$ is a dense subset then there is a unique extension of $\tilde{\theta}(\cdot; \cdot)$ over all $Q_{X,Z}^2$. If not, then, let us linearly extend the function to the entire domain of $Q_{X,Z}^2$. Then define $v(\mathbf{q}; X_2) = \nabla P(\mathbf{q}; X, Z) \circ (\tilde{\theta}_k(\mathbf{q}))^{-1}$. Finally, to extend the function to Q we can assume that each function $v_j(q_j; X) = q_j^{1/2}$, $j = 1, \dots, J$ for all $\mathbf{q} \in Q_{X,Z}^1 \cup \{\mathbf{q}_0\}$. 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 $\theta = \nabla P(\mathbf{q}; X, Z) \circ (v(\mathbf{q}; X_2))^{-1}$. Since the probability of $\mathbf{q} = \{\mathbf{q}_0\}$, $\mathbf{q} \in Q_{X,Z}^1$

and $\mathbf{q} \in Q_{X,Z}^2$ is equal to the probability of $\theta \in \mathcal{S}_\theta^0$, $\theta \in \mathcal{S}_\theta^1$ and $\theta \in \mathcal{S}_\theta^2$, respectively we can determine $F_\theta(\cdot)$. It is then straightforward to verify that the triplet thus constructed belong to \mathcal{M}_3 . \square

5. MEASUREMENT ERROR AND UNOBSERVED HETEROGENEITY

5.1. Measurement Error. 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.

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 transfers are measured with additive error, and if the error is independent of the true transfer 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.,

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

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

Now consider a case where the choices \mathbf{q} 's are observed with error. We observe $\mathbf{q}^\eta = \mathbf{q} + \eta \cdot \mathbf{1}$ and not \mathbf{q} , where $\eta \in \mathbb{R}_+$ is a measurement error and $\mathbf{1}$ is J -dimensional vector of ones that is also independent of \mathbf{q} . The data is $\{P, \mathbf{q}^\eta\}$ 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 $\nabla P(\mathbf{q}) \neq \nabla P(\mathbf{q}^\eta)$. Moreover, without the the knowledge of the density of η , even $M^*(\cdot)$ cannot be identified.

Lemma 5.2. *If $\{\mathbf{q}^\eta = \mathbf{q} + \eta \cdot \mathbf{1}\}$ is observed, where \mathbf{q} is the choice and $\eta \perp \mathbf{q}$ is the measurement error then the model $[F_\theta(\cdot), C(\cdot)]$ cannot be identified.*

5.2. Unobserved Heterogeneity. In this section I extend the linear utility model to allow for unobserved heterogeneity. Let $Y \in \mathbb{R}_+$ denote some factors that is relevant for the consumers and is observed by them, but is unobservable to the econometrician. There are many ways to introduce unobserved heterogeneity. Consider the following assumption:

Assumption 7. *Let the net utility when a θ type consumer chooses \mathbf{q} bundle be*

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

where $Y \sim F_Y(\cdot)$ defined over $\mathcal{S}_Y \subset \mathbb{R}_+$ and $Y \perp\!\!\!\perp \theta$.

Hence, Y affects the utility but not the price. As a result the model is observationally equivalent to a measurement error model discussed in section 5.1 where we observe $P^y(\mathbf{q}) := P(\mathbf{q}) - y$ instead of $P(\mathbf{q})$ when $Y = y$. Then identification follows from straightforward application of Lemma 5.1.

An alternative way to model unobserved heterogeneity is to interpret the type of each consumer as a product of two components: one (Y) is common and known to all consumer; the other (θ) is individual and the private information of consumer.

Assumption 8. *Let*

- (1) *The random variables (θ, Y) are distributed on $\mathcal{S}_\theta \times \mathcal{S}_Y$ 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 $\mathcal{S}_{\theta^*|Y}^2$ denote the types that are perfectly screened. Then under assumption 8 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 $\{1, \dots, N_{21}\}$ and $\{1, \dots, N_{22}\}$ and taking the log of the above we get

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

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. Similarly, 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):

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

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

Lemma 5.3. *Under Assumption 8, the model $[F_\theta(\cdot), F_Y(\cdot), C(\cdot), v(\cdot)]$ with unobserved heterogeneity is identified.*

6. CONCLUSION

In this paper I study the identification of a screening model studied by [Rochet and Choné \(1998\)](#) where consumers have multidimensional private information. 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 nonparametrically 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. One important caveat of this model is that it considers only a single seller. Extending it to an oligopoly set up is a very important area of research left for future.

APPENDIX A. MULTIVARIATE QUANTILES

Let $(\mathbb{S}, \mathcal{B}, L)$ be a probability space with probability measure L . Let $g : \mathbb{R}^J \times \mathbb{S} \rightarrow \mathbb{R}$ 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

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

be called an (M, t) - parameter of L with respect to g , where $\langle \cdot, \cdot \rangle$ is the inner product in \mathbb{R}^J . The subdifferential of g at a point $s \in \mathbb{R}^J$ is denoted by $\partial g(s) = \{t \in \mathbb{R}^J | g(s') \geq g(s) + \langle s' - s, t \rangle\}$. 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.¹⁶

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

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

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

¹⁶ 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 $g_{L,t}(q) := 1/2 \int_{\mathbb{R}} (|q - s| - |s| + q) L(ds) - qt$.

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*.
- 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*.
- 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. Turnovsky, vol. 1. Cambridge University Press.

- COHEN, A., AND L. EINAV (2007): “Estimating Risk Preferences from Deductible Choices,” *American Economic Review*, 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 Non-separable 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.
- ENGL, H. W., M. HANKE, AND A. NEUBAUER (1996): *Regularization of Inverse Problems*. Kluwer Academic Publisher, Dordrecht, Boston, London.
- EVANS, L. C. (2010): *Partial Differential Equations*. AMS, second edn.
- FINKELSTEIN, A., AND K. MCGARRY (2006): “Multiple dimensions of private information: Evidence from the long-term care insurance market,” *American Economic Review*, 96(4), 938–958.
- FOX, J. T., AND A. GANDHI (2013): “Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models,” *Mimeo*.
- FUJIMOTO, T., AND C. HERRERO (2000): “A Univalence Theorem for Non-linear 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*.
- 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.

- KACHUROVSKII, 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 Matematica Pura ed Applicata*, 74(1), 129–134.
- LAFFONT, J.-J. (1994): “The New Economics of Regulation Ten Years After,” *Econometrica*, 62, 507–537.
- LANG, S. (1973): *Calculus of Several Variables*. Addison-Wesley.
- LUENBERGER, D. G. (1969): *Optimization by Vector Space Methods*. John Wiley & Sons, Inc.
- 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.
- NEWBY, 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 Screening,” *Econometrica*, 66(4), 783–826.
- ROCHET, J.-C., AND L. A. STOLE (2003): “The Economics of Multidimensional Screening,” in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, chap. 5. Cambridge University Press.
- 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. (2014): “Identification of Nonseparable Models Using Instruments with Small Support,” *Forthcoming Econometrica*.

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