IDENTIFYING A MODEL OF SCREENING WITH MULTIDIMENSIONAL CONSUMER HETEROGENEITY∗

GAURAB ARYAL†

ABSTRACT. In this paper, we study the nonparametric identification of a model of price discrimination with multidimensional consumer heterogeneity from disaggregated data on consumers’ choices and characteristics. In particular, we consider the screening problem faced studied by Rochet and Choné (1998) where a seller of a product with multiple (and continuous) characteristics who only knows the joint density of consumer “taste” and the production cost and chooses a product “line”—endogenous product characteristics. We determine the data features and additional conditions that are sufficient to identify the joint density of consumer heterogeneity, the cost function, and the utility functions that are common across consumers. If the product characteristics enter the utility function linearly, data from only one market is enough for identification, but if they enter nonlinearly we need data from at least two markets, or over two periods, with exogenous differences in costs. We also derive all testable restrictions imposed by the model on the data, i.e., the empirical content of the model, and also explore identification when prices are mismeasured and a product characteristic is missing.

Keywords: identification, multidimensional private information, consumer heterogeneity, rationalizability.

JEL classification: L12, C57, D82

Date: February 15, 2017.

∗ Previously, the paper was circulated under the title, “Identifying Multidimensional Adverse Selection Models,” and “Identifying a Screening Model of Multidimensional Private Information.” I want to thank Daniel Ackerberg, Stéphane Bonhomme, Xavier D’Haultfœuille, Liran Einav, Jeremy Fox, Maria F. Gabrielli, Eric Gautier, Marc Henry, Ali Hortacsu, Yao Luo, Ariel Pakes, Alexander Torgovitsky, Ali Yurukoglu, seminar participants at JHU, Penn State, Stanford IO-Lunch, UBC, Michigan, USC, UVA, and conference participants at IIOC-2016, NASM-2016. All errors are mine.

†University of Virginia. e-mail: aryalg@virginia.edu.
1. Introduction

Information asymmetry and its effect on welfare is an important topic in economics.\(^1\) Understanding the nature and prevalence of information asymmetry and its effect on welfare and efficiency are crucial in designing policies that are meant to correct or minimize losses due to information-asymmetry-led market failures.\(^2\) However, in an important paper, Chiappori and Salanié (2000) did not find any evidence of information asymmetry in French auto insurance market, questioning the then widely accepted dictum that adverse selection is universal and leads to substantial welfare loss. More recently, Chiappori and Salanié (2003); Einav, Finkelstein, and Cullen (2010); and Einav, Finkelstein, and Schrimpf (2010), among others, have studied insurance and annuities markets and found mixed evidence of adverse selection and its effect on welfare.

One reason for this mismatch between the theory and the empirics is that the former assumes that consumer heterogeneity can be summarized in a one-dimensional parameter, but the reality is more likely to be multidimensional and cannot be sorted, in a satisfactory manner, along only one dimension. For example, Finkelstein and McGarry (2006) and Cohen and Einav (2007) have shown that private information about both risk and risk aversion is crucial in insurance markets, and any welfare/efficiency calculation depends on their mutual dependence. Also see Aryal, Perrigne, and Vuong (2016) for non-parametric identification of multidimensional information in insurance markets. However, we know very little about markets other than insurance and annuities, where multidimensional preference heterogeneity is presumably equally important, and where researchers do not observe cost (researchers in insurance market rely heavily on claims data).

The aim of this paper is to contribute to the literature on empirical multidimensional adverse selection by studying the identification of a model of non-linear pricing with multidimensional heterogenous taste, from individual level data on choices, prices, and socioeconomic characteristics, in a market with a single seller, and a continuous optimal menu. In particular, we consider the model of multidimensional screening of Rochet and Choné (1998) (henceforth, Rochet-Choné) where a seller offers an optimal (and continuous) menu that

---

1 See Akerlof (1970), Spence (1973); and Rothschild and Stiglitz (1976) for earlier results, and Laffont and Martimort (2001) for a more recent synthesis.

2 For regulation see Baron (1989); Baron and Myerson (1982); Joskow and Rose (1989); Laffont and Tirole (1993); Laffont (1994); for income taxation see Mirrlees (1971); and for price discrimination see Armstrong (1996); Wilson (1993).
consists of product characteristics and price function to consumers with heterogeneous taste for each of these characteristics. Only the consumers know their taste profile, the seller only knows the joint distribution of taste, and since the sellers chooses product characteristics and prices to maximize profit, the observed characteristics and prices are endogenous. In fact they will not only be a function of the cost function, but also of the joint distribution of consumers’ taste and any common utility function(s). We show that under the assumption that the data are an outcome of the Rochet-Choné model, and if some additional conditions (e.g., exclusion restrictions and independence) are satisfied, we can identify these model parameters. We also derive all testable restrictions the model imposes on the data that can be used to test model validity.

In equilibrium, consumers’ type space is divided into three subsets: the high-types who are perfectly screened (where each unique type is allocated a unique bundle); the medium-types who are bunched (where two distinct types of consumers choose the same bundle) with restricted options, and the low-types who are excluded. This is in sharp contrast to the model with one dimensional heterogeneity where every consumer type is allocated a unique bundle. Any welfare analysis that ignores multidimensionality or relies only on the demand side will be incorrect because it cannot account for the bunching. In this paper, we show how the supply side conditions can be useful for identification if we are interested in studying markets with multidimensional heterogeneity. Interestingly, besides accounting for bunching, we also find that supply side optimality conditions provide a tight link between the distribution of choices and unobserved cost which can be exploited to identify cost with minimal assumption. Therefore we can consider more general cost functions.

To study the identification we proceed in the following way. We first consider an environment where product characteristics enter the utility function linearly. There, we first study the identification of the joint density of taste for high-type consumers, before studying the identification of the cost function and the joint density of taste for medium-type consumers. Then we consider the identification when the product characteristics enter the gross utility function nonlinearly. Proceeding this way allows us to present the identification strategy in a transparent way, as we can then highlight the identification content of a data feature or an assumption.

---

3 We can replace multiple characteristics with multiple products, or optimal contract with multiple tasks and multidimensional task-specific skills, and our results will still apply.
In the linear specification, the observed price gradient (marginal prices) identifies the high-types corresponding to each choice. Intuitively, the incentive compatibility constraints ensures that the marginal prices must equal marginal utility for each type, which under linear specification is just the consumers’ type. This bijection between choices and types does not apply to the medium-types because they are not perfectly screened. Once the high-types are identified, the first-order condition that characterizes equilibrium allocation rule provides a link between the observed prices and the unobserved (nonlinear) marginal cost, which identifies the cost function over an appropriate domain as a solution to a partial differential equation.

Since the medium-types are (by definition) bunched the mapping between types and choices is no longer bijective. Thus the demand side optimality condition is insufficient and we need other source(s) of variation that affect the choices through utility specification and independently of the consumer tastes. One such source of variation is the consumer characteristics as long as there are as many characterstics as the dimension of consumer taste that are independent of each other. Then the conditional choice density given consumer characteristic can be written as a mixture of type distribution and choice function. Such a mixture is known as a Radon transform, and with sufficient variation in consumer characteristics this mixture is invertible (Helgason, 1999), thereby identifying the joint density of the medium-types.

Linear specification is less of a problem if we are studying multiproduct monopoly because then we use quantities which are well defined across all products. But in a pure characteristics environment, it is harder to compare two attributes of the same product. In such cases we would like to allow characteristics to enter utility through a nonlinear and unknown transformation. This widens the applicability of this model, but at a significant cost in terms of identification.

In particular, in this case data from only one market is insufficient for identification. This is because the marginal utility is now a product of the slope of the utility function and consumer taste, both of which are unknown. Therefore we need some form of exogenous “shifter” that affects one but not the other. We show that a discrete cost shifter (either over time for the same market or across two markets served by the same seller) that affects the marginal prices

---

4 This is analogous to the the identification of a random coefficient model with binary choice in Ichimura and Thompson (1998); Hoderlein, Klemelä, and Mammen (2010) and Gautier and Kitamura (2013) among others, where a key identifying assumption is that the regressors and the random coefficients are independent.
is sufficient for identification. For example, in telcom data, we can observe a representative sample of a market every few months, and where only the cost function changes over time. If cost changes (exogenously) then it will affect the marginal prices but not the slope of the functions. And since we consider only one seller, such shift in cost must be either over two independent markets or the same market over two periods.

Although for this strategy to work it is not important for us to observe what these shifters are it is important that these shifters are non trivial in the sense that they not only affect the cost and the price levels but also they should also affect the price gradients for all high-type bundles. We then show that under some appropriate normalization, we can exploit these exogenous changes in choices across two cost regimes to identify both the (vector valued) nonlinear functions and the joint density of high-types. Intuitively, the exogenous changes in cost affects the marginal prices, which then affect the choices, but not the types. A “median” or centerpoint (high-type) consumer will still buy the median bundle, irrespective of the cost. This allows us to match the multivariate quantiles across two cost regimes to identify the nonlinear functions. Matching these multivariate quantiles across two cost regimes allow us to identify the quantile function of the high-type consumers.\footnote{This idea of matching quantiles under exclusion restriction in one dimension owes much to Matzkin (2003) and Guerre, Perrigne, and Vuong (2009). The strategy we use is also similar to (the working paper) D’Haultfoeuille and Février (2011) who study triangular models with discrete instruments. For additional examples see Aryal, Grundl, Kim, and Zhu (2016); D’Haultfoeuille and Février (2015); and Torgovitsky (2015).} One caveat of this strategy is that due to the fact that there is no natural ordering in multivariate quantiles we have to choose a proper ordering, for which we can follow (Koltchinskii, 1997). Once we identify the quantile function for the consumer type, we can then use the consumer’s optimality condition to identify the joint density of high-types and use the seller’s optimality to identify the cost function as before.

On top of that, if we use variation in consumer characteristics then even in the nonlinear case the joint density of high-type consumer is over-identified. Over identification is important because it affects both the efficiency of an estimator and the refutability of the model (Koopmans and Riersol, 1950). Lastly, we also determine all testable restrictions imposed by the model on the data. In principle these restrictions form the basis of specification tests to check the validity of the model, and ensure non-emptiness of the identified set.
Motivating Example. An example is useful for illustration. Consider the optimal bundling problem faced by a dominant (monopoly) telecommunication company. Such a market is studied by Luo, Perrigne, and Vuong (2015). The company offers multiple cellphone plans with different rates for talk-time, data, and instant text messaging. Figure 1 shows a scatter plot of the usages and the payments made by the subscribers, which points to substantial heterogeneity among consumers. Consumers’ preference for talk time vs. data could vary in a general way that is hard to capture by a one dimensional index, i.e., consumer heterogeneity could be multidimensional.

Profit for the company is highest for the plans designed for the “high-type” consumers, who have higher willingness to pay. They, however, cannot be prevented from choosing plans that are meant for medium-type or low-type consumers. To realize higher profit the company must distort the product characteristics offered for lower types in the direction that make them relatively unattractive for the high-types. This distortion and its effect on welfare will

---

6 I thank Yao Luo for sharing these scatter plots with me.
depend on the distribution of consumer heterogeneity, the cost and the utility. This means that the observed product characteristics are endogenous and consumers’ preferences have multidimensional unobserved components.\footnote{Recently, Fan (2013) has extended the identification arguments in Berry (1994) and Berry, Levinsohn, and Pakes (1995) to allow for endogenous product characteristics albeit under perfect information.}

**Related Literature.** In terms of the subject matter, this paper is closest to Pioner (2009) and in terms of the objective it is closest to Aryal, Perrigne, and Vuong (2016) who study the nonparametric identification of the joint distribution of risk and risk preference in insurance markets. Pioner (2009) studies the semiparametric identification of the Rochet-Choné model, but only under the restriction that private information is two-dimensional and, more importantly, the data contain information about one of those types. If one of the dimensions were income, say, then it is possible for a researcher to use multiple data sources to piece together market level distribution of the said variable. But for his identification strategy to work we need to know the information at the individual level, which is hard to come by even for income. In general, between the seller and the researcher, it is most likely that the seller is who knows more, and we show that neither of these two assumptions are necessary for identification.

Aryal, Perrigne, and Vuong (2016) study the identification of multidimensional consumer heterogeneity, they focus on insurance markets and the results from that paper do not apply outside insurance. For instance one of the key sources of identification in insurance market is the claims data, which is informative about the risk involved on an insurance contract. It is unlikely that similar “extra” information is available in most of the cases.

Other papers that are similar to this paper in both spirit and approach is the paper by Perrigne and Vuong (2011) who study nonparametric identification of Laffont and Tirole (1986); Luo, Perrigne, and Vuong (2015), who study the telecommunication in China; Ivaldi and Martimort (1994) and Aryal (2016), who study competitive nonlinear pricing; and Ekedal, Heckman, and Nesheim (2002, 2004); Heckman, Matzkin, and Nesheim (2010); Chernozhukov, Galichon, Henry, and Pass (2014), who study hedonic models.

The remainder of the paper is organized as follows: Section 2 describes the model, Section 3 contains the identification results, and Section 4 provides the rationalizability lemmas. Section 5 considers some extensions before concluding. Appendix A contains all the necessary technical details.
2. The Model

In this section we present the model adapted from Rochet and Choné (1998). The objective is to introduce the environment and the model of multidimensional screening without regurgitating every detail of their paper. The reader interested in the details of the model should consult their paper.

A seller offers a menu that includes a product line \( Q \subseteq \mathbb{R}^d_{+} \) — set of all feasible characteristics — and a (nonlinear) price function \( P : Q \rightarrow \mathbb{R}_{+} \). Let \( \theta \in S_\theta \subseteq \mathbb{R}^d_{+} \) denote the consumer’s heterogeneity, i.e, taste profile (or simply type) that is independently and identically distributed (across them) as \( F_\theta(\cdot) \). The seller knows the distribution \( F_\theta(\cdot) \) and the cost function \( C : \mathbb{R}^d_{+} \rightarrow \mathbb{R}_{+} \).

**Assumption 1.** The utility from \( q \in Q \) for \( \theta \)-type is \( V(q; \theta) := u(q, \theta) - P(q) \).

Often we also observe socioeconomic or demographic characteristics of the consumers denoted by \( X \in S_x \subseteq \mathbb{R}^d_x \) which affect the utility function. To accommodate that let the utility be

\[
V(q; \theta, X) := u(q, \theta, X) - P(q, X).
\]

If \( X \) affects the utility, then the seller first conditions on \( X \) (third-degree price discrimination) and then for each \( X = x \) chooses the product line and price function, and hence \( P(\cdot, X) \) (second-degree price discrimination). Throughout the paper we maintain the quasi-linear preference assumption. We impose the following assumption to simplify the problem.

**Assumption 2.** Let

(i) \( d_\theta = d_q = J \).

(ii) \( \theta \overset{i.i.d.}{\sim} F_\theta(\cdot) \) which has a square integrable density \( f(\cdot) > 0 \) a.e. on \( S_\theta \).

(iii) The net utility is an element of a Sobolev space

\[
V(q; \cdot, X) \in \mathcal{V}(S_\theta) = \{V(q; \cdot, X)| \int_{S_\theta} V^2(\theta) d\theta < \infty, \int_{S_\theta} (\nabla V(q, \theta, X))^2 d\theta < \infty\},
\]

with the norm \( |V| := \left( \int_{S_\theta} \{V^2(\theta) + \|\nabla V(\theta)\|^2\} d\theta \right)^{\frac{1}{2}} \).

(iv) The gross utility function is multiplicative in \( \theta \):

\[
u(q, \theta, X) = \theta \cdot v(q, X) \equiv \sum_{j=1}^{J} \theta_j v_j(q_j; X),
\]

\[\text{For a scalar function } \kappa(\zeta_1, \ldots, \zeta_{d_\zeta}) \in \mathbb{R}, \nabla \kappa_j = \left( \frac{\partial \kappa}{\partial \zeta_1}, \ldots, \frac{\partial \kappa}{\partial \zeta_{d_\zeta}} \right) \text{ denotes the gradient of } \kappa(\cdot), \text{ and } \nabla_j \kappa(\cdot) \text{ denotes the } j^{th} \text{ element of the gradient vector.}\]
where each $v_j(\cdot; X)$ is differentiable and strictly increasing, and is either:

(iv-a) linear utility: $v_j(q_j, X) = q_j$.

(iv-b) bilinear utility: there are at least $J$ vectors in $X$, i.e., $d_x \geq J$. Furthermore $X = (X_1, X_2)$ such that $d_{x_1} = J$ and $d_{x_2} \geq 0$, where $X_1 = (X_{11}, \ldots, X_{1J})$ is the vector of consumer characteristics that interact multiplicatively with the product characteristics. So the gross utility from $q_j$, after relabeling if necessary, is given by $\theta_j \cdot v_j(q_j, X) = \theta_j(q_j \cdot X_{1j})$.

(iv-c) nonlinear utility: the gross utility from $q_j$ is given by $\theta_j \cdot v_j(q_j, X) = \theta_j(q_j \cdot X_{1j} \cdot X_{2j})$, where $X_{1j}$ and $X_{2j}$ are defined in part (iv-b) above; for every $j, v_j(\cdot, X_2)$ is a twice continuously differentiable and strictly quasi-concave function such that the Jacobian matrix $Dv(q_j; X_2)$ is of full rank for all $q_j$; and finally $v_j(0; \cdot) = 0$ and $\lim_{q_j \to \infty} v_j(q_j) = \infty$.

(v) $C(\cdot)$ is a strongly convex function with parameter $\epsilon'$, i.e. the minimum eigenvalue of the Hessian matrix is $\epsilon' > 0$.

Assumption 2-(i) assumes that agents differ in as many dimensions as the attributes of a product. Assumption 2-(ii)-(iii) are standard assumptions in the literature in mechanism design; for more see Rochet-Choné. Assumption 2-(v) is the standard convexity of cost assumption. Assumption 2-(iv), however, needs more explanation. The first part assumes that the utility function is multiplicatively separable in consumer type $\theta$ and some function of product characteristics. This multiplicative separability assumption is an important assumption and is used widely in the price discrimination literature (Wilson, 1993; Laffont and Martimort, 2001) while Carlier (2001) and Figalli, Kim, and McCann (2011) are recent exceptions.

The second part puts more structure on how $q$ and $X$ enter the utility function. Assumption 2-(iv-a) assumes that the product characteristics enter linearly and do not depend on consumer characteristics $X$. Assumption 2-(iv-b) assumes that some consumer characteristics might interact with the product characteristics. What is crucial for us is that there should be at least $J$ many vectors in $X$, i.e., $d_x \geq J$. In other words, we can allow $X_2$ to be a null set. This assumption is not important for the model but it is important for identification. The gross utility from a product characteristic $q_j$ is then $\theta_j \cdot v_j(q_j, X) = \theta_j \cdot q_j \cdot X_{1j}$. If we want to allow interaction between, say $X_{11}$ and $X_{12}$ to affect utility from $q_1$ (say) then we can simply define a new vector $X_{11} \cdot X_{12}$ and relabel it as $X_{1j}$. 

\footnote{This formulation is more general than the random coefficient models. If for example $X_{11}$ and $X_{12}$ affect the utility from $q_1$, ...}
Assumption 2-(iv-c) generalizes the previous assumption and allows product characteristics to enter the utility function nonlinearly and allows this function to also depend on $X_2$ as long as it is bilinear in $\theta$ and $X_1$.

It is important to note that for the theoretical model it suffices that the utility function $u(\theta, q; X)$ be multiplicatively separable in $\theta$. We can then redefine the units of measurement for the product characteristics and substitute $\tilde{q}$ for $q$. Since the seller observes $X$ and knows $v(\cdot, \cdot)$, selling $q$ is isomorphic to choosing $\tilde{q}$. The theory is silent about how the product characteristics enter the utility function, so we consider three (linear, bilinear, and nonlinear) cases and study their identification separately, starting with the simplest case (linear specification) assumption 2-(iv-a) before considering bilinear and nonlinear specifications assumption 2-(iv-b) and 2-(iv-c).

For notational ease suppress the dependence on $X$. A menu $\{Q, P\}$ is feasible if there exists an allocation rule $\rho : S_\theta \rightarrow Q$ that satisfies the incentive compatibility (IC) condition,

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

and the individual rationality (IR) condition, $\forall \theta \in S_\theta, U(\theta) \geq U_0 := \theta \cdot v(q_0) - P_0$, where $\{q_0\}$ denotes the outside option available to everyone at a fixed price $P_0$. So, we can view $U(\theta)$ as the information rent $\theta$ gets by virtue of knowing his/her type. To ensure the principal’s optimization problem is convex, we assume $P_0 \geq C(q_0)$, so that the seller always offers $q_0$, i.e., $Q \ni q_0$. The seller chooses a feasible menu $(Q, \rho(\cdot), P(q))$ that maximizes expected profit

$$\Pi = \int_{S_\theta} \pi(\theta) dF(\theta) := \int_{S_\theta \cap \{\theta : U(\theta) \geq U_0\}} (P(q(\theta)) - C(q(\theta))) dF(\theta). \tag{2}$$

Let $W(\rho(\theta), \theta)$ be the surplus when $\theta$ is allocated $q(\theta)$. Then $W(\rho(\theta), \theta) = U(\theta) + \pi(\theta)$, or $W(\rho(\theta), \theta) = \{\theta \cdot v(\rho(\theta)) - P(\rho(\theta))\} \cup \{P(\rho(\theta)) - C(\rho(\theta))\}$. Equating these two definitions gives $\pi(\theta) = \theta \cdot v(\rho(\theta)) - C(\rho(\theta)) - U(\theta)$. Under Assumption 2, a menu $\{Q, \rho(\cdot), P(\cdot)\}$ is such that $U(\theta)$ solves Equation (1) (i.e., satisfies IC) if and only if: (i) $\rho(\theta) = v^{-1}(\nabla U(\theta))$; and (ii) $U(\cdot)$ is convex on $S_\theta$, Rochet (1987). This means that choosing an optimal contract $\{Q, \rho(\cdot), P(\cdot)\}$ is equivalent to determining the net utility $U(\theta)$ each $\theta$ gets by participating. From $U(\theta)$, we can determine the optimal allocation rule as $\rho(\theta) = v^{-1}(\nabla U(\theta))$.

Thus the seller chooses $U(\theta) \in H^1(S_\theta)$ that maximizes the expected profit

$$\Pi(U) = \int_{S_\theta} \{\theta \cdot \nabla U(\theta) - U(\theta) - C(v^{-1}(\nabla U(\theta)))\} dF(\theta),$$
subject to the IC and the IR constraints. As mentioned above, the global IC
constraint is equivalent to convexity of $U(\cdot)$, i.e., $D^2U(\theta) \geq 0$, and IR is equiva-
lent to $U(\theta) \geq U_0(\theta)$ for all $\theta \in S_\theta$. Rochet-Choné showed that Assumption 2
is sufficient to guarantee the existence of a unique maximizer $U^*(\cdot)$.

**Note 1.** When $J = 1$, under the assumption that $[1 - F(\cdot)]/[f(\cdot)]$ is strictly
decreasing the equilibrium will always have perfect screening. So to determine
the solution we can ignore the IC constraint and verify it ex-post. When $J > 1$
Armstrong (1996) showed that low-type consumers would be excluded from the
market. Rochet-Choné showed another consequence of multidimensionality is
that some medium-type consumers will be bunched together, and this bunching
is a generic feature of any equilibrium. It is also important to note that the
“low-types” or the “medium-types” are determined endogenously as a function
of the cost function and the type distribution. Bunching means two distinct types
of consumers can choose the product configuration, which makes the problem of
identification difficult.

Rochet-Choné showed that in equilibrium the consumers will be divided into
three types: the lowest-types $S^0_\theta$ who are screened out and offered only $\{q_0\}$, the
medium-types $S^1_\theta$ who are bunched and offered “medium type” bundles and the
high-types $S^2_\theta$ who are perfectly screened. If $U^*(\cdot)$ is optimal then offering any
other feasible function $(U^* + h)(\cdot) - h$ is a non-negative and convex function—
must lower expected profit for the seller, i.e., $\mathbb{E}\Pi(U^*) \geq \mathbb{E}\Pi(U^* + h)$. This
inequality means that the directional derivative of the expected profit, in the
direction of $h$, must be non-negative so $U^*(\cdot)$ is the solution iff: (a) $U^*(\cdot)$ is a
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

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

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

Intuitively, $\alpha(\theta)$ measures the marginal loss of the seller when the indirect
utility (information rent) of type $\theta$ is increased marginally from $U^*$ to $U^* + h$.
Alternatively, define $\nu(\theta) := \frac{\partial W(\theta, q_\theta(\theta))}{\partial q_\theta}$, the marginal distortion vector, then
$\alpha(\theta) = 0$ is equivalent to $\text{div}(\nu(\theta)) = -f(\theta)$, which is the optimal trade-off
between distortion and information rent.

---

10 The divergence $\text{div}$ of a scalar function $\kappa(\zeta)$ is defined as $\text{div} \kappa = \sum_{j=1}^d \frac{\partial \kappa(\zeta)}{\partial \zeta_j}$. 
Let $L(h) = -\mathbb{E}[U^\prime(U^*)]h$ be the loss of the seller at $U^*$ for the variation $h$. So if the seller increases $U^*$ in the direction of some $h$, then the seller’s marginal loss can be expressed as

$$L(h) = \int_{S_0} h(\theta)\alpha(\theta)d\theta + \int_{\partial S_0} h(\theta)(-\nu(\theta) \cdot \overrightarrow{n}(\theta))d\sigma(\theta) := \int_{S_0} h(\theta)d\mu(\theta), \quad (4)$$

where $d\sigma(\theta)$ is the Lebesgue measure on the boundary $\partial S_0$, $\overrightarrow{n}(\theta)$ is an outward normal vector, and $d\mu(\theta) := \alpha(\theta)d\theta + \beta(\theta)d\sigma(\theta)$. If we consider those 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 $\beta(\theta) := -\nu(\theta) \cdot \overrightarrow{n}(\theta)$ must be equal to zero. For those who do not participate, it must mean that the loss is positive.

The global incentive compatibility condition is important because it determines optimal bunching in the equilibrium by requiring $(U^* - U_0)(\theta)$ to be convex, and also determines the subset $S^1_\theta$ where the optimal allocation rule $\rho(\cdot)$ is such that some types are allotted the same quantity $q$. Let $S^1_\theta(q)$ be the types that get the same $q$, i.e., $S^1_\theta(q) = \{\theta \in \Theta : \rho(\theta) = q\} = \{\theta \in \Theta : U^*(\theta) = \theta \cdot q - P(q)\}$. If $U^*(\theta)$ is convex for all $\theta$, i.e., if the global incentive compatibility constraint is satisfied, then there is no bunching and $S^1_\theta$ would be an empty set. In most 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 $\theta'$ and $\theta$ if and only if they both belong to $S^1_\theta(q)$, i.e., if $\theta' \not\in S^1_\theta(q)$ but $\theta \in S^1_\theta(q)$ then $U^*(\theta') > U^*(\theta) + (\theta - \theta')^Tq$.

Rochet-Choné Theorem 2': Under Assumptions 2-(i)–(iv-a) and (v) the optimal solution $U^*$ to the problem is characterized by a continuous allocation rule $\rho(\cdot)$ and three endogenously determined subsets $S^0_\theta, S^1_\theta$, and $S^2_\theta$ such that:

1. Those who belong to $S^0_\theta$ (low-types) do not participate because $U^*(\theta) = U_0(\theta)$.
2. Those who belong to $S^1_\theta$ (medium types) are bunched together and are allocated some intermediate quality $q \in Q^1$. They are further subdivided into subset $S^1_\theta(q)$ such that all types in this subset get one type $q$, and $U^*$ is affine.
3. Those who belong to $S^2_\theta$ (high-types) are perfectly screened, where each unique type is allocated a unique and customized bundle. For them, $U^*$ satisfies the Euler condition (3) and there is no distortion in the optimal allocation on the boundary.
For illustration of the geometry of the partition, see Fig. 2. As can be seen, Rochet-Choné imply strong restrictions on the support of the product characteristics. For instance, the characteristics region $Q^1$ corresponding to $S^1_\theta$ may be of low dimension, as can be seen in the figure. This prediction of the model is conditional on $X$; without conditioning the data can look like the one in Fig. 1-(c). For the identification, we exploit this feature, that $Q^1$ has lower dimension than $J$ hence the price function is also restricted – in particular the price gradients are restricted for all $q \in Q^1$. Since we observe prices, we can first estimate the price function and use it determine whether the observed product $q$ is in $Q^1$ or $Q^2$ (see the discussion in the next section).

3. Identification

In this section we study the identification of the model parameters –the type-distribution $F_\theta(\cdot)$, the cost function $C(\cdot)$ and the functions $v(\cdot)$– from the observables $\{q_i, p_i, X_i\}$ for $i = 1, \ldots, N$ consumers. Consider an example from Rochet-Choné, with two-dimensional product characteristics. This example, albeit stylized, illustrates some of the key predictions of the model.

**Example 3.1.** Let $J = 2$, the cost function be $C(q) = c/2(q_1^2 + q_2^2)$ and types be independent and uniformly distributed on $S_\theta = [0, 1]^2$ and $q_0 = 0$ and $P_0 = 0$. Then, the optimal indirect utility function $U^*$ has different shapes in the three regions: (i) in the non-participation region $S^0_\theta$, $U^*(\theta) = 0$; (ii) in the bunching region $S^1_\theta$, $U^*$ depends only on $\theta_1 + \theta_2$; and (iii) in the perfect screening region $S^2_\theta$, $U^*$ is strictly convex.

Heuristically, the optimal product line is given in Figure 2. On the left we have consumer type space, which is a unit square, and on the right we have the product line which is a subset in $\mathbb{R}^2$. All those who belong to the set $S^0_\theta$
choose the outside option \( Q^0 := \{ q_0 \} \). Those who are in \( S^1_\theta \) are bunched. For instance, all those with \( \theta = (\theta_1, \theta_2) \) such that \( \theta_1 + \theta_2 = \tau \) (along the dotted line between \((0, \tau)\) and \((\tau, 0)\)) buy the same bundle \( q_1 \), and those with different \( \tau \) buy a different bundle. The set \( Q^1 \) denotes these “medium” bundles and in equilibrium will be bought by consumers with \( \theta \in S^2_\theta \).

### 3.1. Identification

Let \((p, q, X) \overset{i.i.d.}{\sim} \Psi_{p,q,X}(\cdot, \cdot, \cdot) = \Psi_{p,q|X}(\cdot, \cdot) \times \psi_X(\cdot)\). The seller offers \((Q(x), P(x)(\cdot))\) to agent \( i \) with observed characteristics \( X_i = x \sim \Psi_X(\cdot) \) and unobserved type \( \theta_i \sim F_\theta(\cdot) \), who then chooses a utility maximizing \( q_i \in Q(x) \) and pays \( p_i = P(x)(q_i) \). The seller chooses the menu optimally, and the revelation principle implies that it is equivalent to saying that there exists a direct mechanism, a unique pair of (allocation rule) \( \rho(\cdot) : S_\theta \mapsto Q(x,z) \), and (pricing function) \( P_x(\cdot) : Q(x) \mapsto \mathbb{R}_+ \), such that \( q_i = \rho(\theta_i) \) and \( p_i = P(x)(\rho(\theta_i)) \). Henceforth, \( \rho(\cdot) \) will stand for the optimal allocation rule. Thus assuming that the observed data are an outcome of Rochet-Choné model gives us the econometric model:

\[
\begin{align*}
p_i &= P[q_i, F_\theta(\cdot), C(\cdot); X] \\
q_i &= \rho[\theta_i, F_\theta(\cdot), C(\cdot); X], \quad i \in [N], k = 1, 2. \tag{5}
\end{align*}
\]

The model parameters \([F_\theta(\cdot), C(\cdot)]\) are identified if for any different parameters \([\tilde{F}_\theta(\cdot), \tilde{C}(\cdot)]\) the implied data distributions are also different, i.e., \( \Psi_{p,q,X}(\cdot, \cdot, \cdot) \neq \tilde{\Psi}_{p,q,X}(\cdot, \cdot, \cdot) \). Since the equilibrium is unique, there is a unique distribution of the observable \( \Psi_{p,q,X}(\cdot, \cdot, \cdot) \) for every parameter (Jovanovic, 1989) so it is enough to determine conditions under which equations (5) are globally invertible.\(^{11}\)

Following the equilibrium characterization, consider the three subsets of types separately. For every \( X = x \), let \( Q^j(x) \) be the set of choices made by consumers with type \( \theta \in S^j_\theta \) for \( j = 0, 1, 2 \), respectively. Since \( q(\cdot) \) is continuous (because \( \rho(\cdot) \) is continuous), these sets are well defined. We can use data from \( Q^k(x) \) to identify the model parameters restricted to \( S^j_\theta \), beginning with the subset \( Q^2(x) \). The allocation rule \( \rho(\cdot) \) is one-to-one when restricted to \( S^j_\theta \), and hence its inverse \( \rho^{-1}(\cdot) \) exists on \( Q^2(x) \), but not when restricted to \( S^j_\theta \) due to bunching, and, as a consequence, the identification strategies are different.

**Note 2.** To exploit the bijection between the parameters and the data we need to classify each choice \( \{ q_i \} \) to be an element of one of the three sets: \( Q^j, j = 0, 1, 2 \). (Everything should be understood as conditioning on \( X \).) We know the outside

\(^{11}\) See Rothenberg (1971) and Chen, Chernozhukov, Lee, and Newey (2014) for local identification in parametric and (semi)nonparametric setups, respectively.
option $Q^0 = \{q_0\}$, so the only thing left is to determine the bunching set $Q^1$. For that we make use of the geometry of $Q^1$, that the product line $Q^1$ is congruent to one-dimensional $\mathbb{R}_+$, which is the main characteristic of bunching as shown in Fig. 2. In particular note that $Q^1$ (a line in Fig. 2) is of lower-dimension than $J$ but $Q^2$ (a plane) is of the same dimension as $J$. This further implies strong restrictions on the gradient of the associated price function. Using the observables $(q_i, P_i)$ for each consumer $i$ we can recover the price function $P(\cdot)$. When $J = 2$, for example, the fact that $Q^1$ is a line and $Q^2$ is a plane, see Fig. 2, means the directional derivative or the gradient of the price function with respect to $q$ is defined in all direction only when $q \in Q^2$. Therefore, the projection of the price on the $q_1 - q_2$ plane for $J = 2$ has a particular shape, see Fig. 3, where $\tilde{P}_1 = Q^1$ and $\tilde{P}_2 = Q^2$. This means we can use the estimated price function to determine the marginal prices: the smallest $q$ at which the marginal prices are defined in all direction must be at the boundary of $Q^1$ and $Q^2$, thereby demarcating $Q^2$ from $Q^1$.

Let $M(\cdot)$ and $m(\cdot)$, respectively, be the distribution and density of choices $q$. 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 $[F_\theta(\cdot), C(\cdot)]$. We say the structure is said to be identified if for given $m(\cdot)$ there exists a (unique) pair $[F_\theta(\cdot), C(\cdot)]$ that satisfies equations (5). Let $\tilde{\theta}(\cdot) : Q \to S^2_\theta$ be the inverse of $\rho(\cdot)$ when restricted on $Q^2$, i.e., $\forall q \in Q^2, \tilde{\theta}(q) = \rho^{-1}(q)$. Similarly, let $M^*(q)$ and $m^*(q)$ be the truncated distribution and density of
where $q \in Q^2$ defined, respectively, as:

$$M^*(q) := \Pr(\tilde{q} < q | q \in Q^2) = \Pr(\theta < \tilde{\theta}(q) | \theta \in S^2_\theta) = \int_{S^2_\theta \cap \{\theta, \tilde{\theta}(q) \leq \tilde{\theta}(q)\}} f_\theta(\theta | \theta \in S^2_\theta) d\theta;$$

$$m^*(q) := \frac{m(q)}{\int_{Q^2} m(q)d\tilde{q}} = \frac{f(\tilde{\theta}(q))}{\int_{Q^2} f(\theta(q))d\tilde{q}} |\text{det}(D\tilde{\theta})(q)|,$$

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

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

where the inverse $\rho^{-1}(q) = (\rho_1^{-1}(q), \ldots, \rho_j^{-1}(q))$ that solves $q = \rho(\theta)$ exists since $U^*(\theta)$ is convex on $S^2_\theta$; see Parthasarathy (1983) and Fujimoto and Herrero (2000). Let $F_\theta(\cdot | j)$ be the distribution function $F_\theta(\cdot)$ restricted to be in the set $S^2_\theta$, and let $N_j$ be the set of consumers who buy $q \in Q^j$, for $j = 0, 1, 2$.

3.2. Linear Utility. When the product characteristics enter the utility function linearly, as is usually assumed in the literature on demand estimation, both $F_\theta(\cdot | 2)$ and $C(\cdot)$ can be nonparametrically identified on $S^2_\theta$ and $Q^2$, respectively. We know that for the high-types the allocation rule $\rho(\cdot)$ is bijective, and from the incentive compatibility constraints (consumer maximization) the marginal (gross) utility $\theta$ is equal to the marginal price $\nabla P(\cdot)$ –the gradient of price function– evaluated at $q = \rho(\theta)$. Therefore the type that chooses $q \in Q^2$ must satisfy $\nabla P(q) = \theta = \tilde{\theta}(q)$, which identifies the (pseudo) type $\theta_i = \tilde{\theta}(q_i)$ for all $i \in [N_2]$, where $[N_k] := \{i \in [N] : i \in Q^k, k = 0, 1, 2\}$. As $\tilde{\theta}(\cdot)$ restricted to $S^2_\theta$ is bijective, the joint distribution of high-types can be identified as

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

Dividing both sides of supply side optimality condition (3) by $\int_{S^2_\theta} f_\theta(t) dt$ we get

$$\text{div} \left\{ \frac{f_\theta(\theta)}{\int_{S^2_\theta} f_\theta(t) dt} (\theta - \nabla C(\nabla U^*)) \right\} = -\frac{f_\theta(\theta)}{\int_{S^2_\theta} f_\theta(t) dt}$$

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

where the second equality uses Equation (6). Then using the curvature of the pricing function $D\nabla P(q) = D\tilde{\theta}(q)$ we can re-write the above equation as

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

12 This fails when the seller offers discrete/finite options; see the discussion in Section 6.
which is a partial differential equation in $C(\cdot)$ with the boundary condition $\beta(\theta) = 0$ on $\partial Q^2$, i.e.,

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

where $\vec{n}$ is the outward normal vector. The cost function $C(\cdot)$ is the only function unknown in Equation (8), everything else is known (or can be estimated). Moreover, there is a unique solution to this partial differential equation Evans (2010) which identifies $C(\cdot)$ albeit only on $Q^2 \subseteq Q$. The function is identified everywhere whenever $Q^1$ is an empty set.

Since there is no other information about the cost function that is available, we can extend the identification from $Q^2$ to $Q$ only under strong functional form assumption. For instance, if $J = 2$ and $C(q) = \frac{1}{2}(\alpha_1 q_1^2 + \alpha_2 q_2^2)$ then we can identify $\alpha_1$ and $\alpha_2$ by solving (8). Alternatively, we could also restrict the space of cost function to be such that it has a unique representation in terms of a series. In particular we can, follow the literature on the nonparametric identification of random coefficient Logit model (Fox, il Kim, Ryan, and Bajari, 2012) or random utility (Fox and Gandhi, 2013), and make the follow assumption:

**Assumption 3.** Cost function $C : Q \rightarrow \mathbb{R}$ is a real analytic function at $q \in Q$.\(^{13}\)

An implication of this assumption is that $C(\cdot)$ is infinitely differentiable and can be expressed (uniquely) as a Taylor series. Some examples of analytic functions include convex polynomials, trigonometric functions, and exponential functions. Once we have determined the coefficients of the Taylor series by solving (8), and since $Q^2$ is convex the $C(\cdot)$ can be extended uniquely everywhere. Although this is a fairly stringent assumption, as far as we know this is still one of the most general way to identify the cost function.

**Theorem 3.1.** Under the Assumptions 2-(i)–(iv-a), (v), and 3, the model structure $[F_\theta(\cdot|2), C(\cdot)]$ is nonparametrically identified.

---

\(^{13}\) A scalar function $\kappa : \mathbb{R}^{d_\xi} \rightarrow \mathbb{R}$ is a real analytic function at $\check{\zeta} \in \mathbb{R}^{d_\xi}$ if there $\exists \delta > 0$ and open ball $B(\check{\zeta}, \delta) \subset \mathbb{R}^{d_\xi}, 0 \leq r < \delta$ with $\sum_{k_1, \ldots, k_J} |a_{k_1, \ldots, a_{K_J}}| r^{k_1 + \cdots + k_J} < \infty$ such that

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

One of the properties of analytic functions is that if two real analytic convex functions coincide on an open set, then they coincide on any connected open subset of $\mathbb{R}^{d_\xi}$.
For illustration, consider the identification in nonparametric IV models by Newey and Powell (2003). There the authors assume that the conditional density belongs to the exponential family, which is analytic (see Liese and Miescke, 2008, Lemma 1.16), and hence the conditional moments take the form of Laplace transforms which are injective. Since we do not observe the conditional moment of cost everywhere, the identification is achieved through extrapolation.

When we move to $S^1_0$ the allocation rule $\rho(\cdot)$ is no longer bijective and the identification argument does not work. It is useful to consider the Example 3.1 further and focus on the medium-types.

**Example 3.2.** Consider the model defined in Example 3.1. On $S^0_\theta$, $\rho(\theta) = 0$, which means $\alpha(\theta) = \text{div}(\theta f(\theta)) + f(\theta) = 3$ and $\beta(\theta) = a$ on $\partial S^0_\theta$. The boundary that separates $S^0_\theta$ and $S^1_\theta$ is a line $\tau_0 = \theta_1 + \theta_2$, where $\tau_0 = \sqrt{3}$. On $S^1_\theta$, $\rho(\theta) = (\rho_1(\theta), \rho_2(\theta)) = (\rho_b(\tau), \rho_b(\tau))$, with $\theta_1 + \theta_2 = \tau$. In other words, all consumers with type $\theta_1 + \theta_2 = \tau$ are treated the same and they get the same $\rho_1(\tau) = \rho_2(\tau) = \rho_b(\tau)$. So $\alpha(\theta) = 3 - 2c\rho_b(\tau)$ and on $\partial S^1_\theta$,

$$
\beta(\theta) = (c\rho(\theta) - \theta) \cdot \hat{a}(\theta) = -c\rho_b(\tau).
$$

Rochet-Choné showed that for each bunch, the optimal bundle is given by $\alpha(\theta) \geq 0$ and $\beta(\theta) \geq 0$ so

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

which can be used to solve for $q_b$ as $\rho_b(\tau) = \frac{3\tau}{4\tau} - \frac{1}{2\tau}$. Then $S^1_\theta = \{\theta : \tau_0 \leq \theta_1 + \theta_2 \leq \tau_1\}$ where $\tau_1$ is determined by the continuity condition on $S_\theta$ of $\rho(\cdot)$, i.e. $\rho_b(\tau_1) = 0$. Now, define $\tau = \rho_b^{-1}(q)$ as the inverse of the equilibrium allocation. And since we cannot determine the joint density of $\theta_1, \theta_2$ uniquely from $\tau$ where $\tau = \theta_1 + \theta_2$, we cannot identify $f(\theta|1)$.

**Note 3.** If we had ignored the supply side and used only the demand side optimality condition then it would have led to a misspecification error. To see that it is sufficient to consider discrete types. Suppose there are only two characteristics $J = 2$ and let consumers’ type space be $S_\theta = \{\{\theta_{1,1}, \theta_{1,2}\}, \{\theta_{2,1}, \theta_{2,2}\}, \{\theta_{3,1}, \theta_{3,2}\}\}$ with probability $\{f_1, f_2, f_3\}$, respectively. Suppose the optimal allocation is such that $\theta_1$ is allocated a bundle $q_1 = (q_{1,1}, q_{1,2})$ but the remaining two types are bunched at $q_2 = (q_{2,1}, q_{2,2})$, and $P(\cdot, \cdot)$ is the pricing function and $\nabla_j P(\cdot, \cdot)$ is the partial derivative with respect to the $j$th argument.\(^{14}\) Suppose the data consists of choice-prices pairs $\{q_i, P_i : i = 1, \ldots, N\}$, and let $m_1$ and $m_2$ be

\(^{14}\) Abusing the notation and using partial derivatives to mean finite differences.
the fraction of consumers who choose \( q_1 \) and \( q_2 \), respectively. If we ignore the supply side and only use the demand side, then we use \( \theta_i = (\theta_j, 1, \theta_j, 2) = (\nabla_1 P(q_{j1}, q_{j2}), \nabla_2 P(q_{j1}, q_{j2})) \) whenever \( q_i = (q_{j1}, q_{j2}) \), \( j \in \{1, 2\} \). Thus there are only two types of consumer given by \( (\nabla_1 P(q_{j1}, q_{j2}), \nabla_2 P(q_{j1}, q_{j2})) \) for \( j = 1, 2 \) with probability \( m_1 = f_1, m_2 = f_2 + f_3 \), which is not the right parameter. Moreover, this error would also lead to misspecification error in the cost function.

Thus we need other sources of variations in the preferences that are independent of \( \theta \), but affect the offered choices. One such candidate is the consumer characteristics \( X \). When the utility function is bilinear in \( q \) and \( X \) (Assumption 2-(iv-b)), and if \( d_x \geq J \) and there is rich variation of \( X \) in the data, independently of \( \theta \), then \( f(\theta|1) \) can be nonparametrically identified.

3.3. Bilinear Utility. Suppose the base utility function satisfies Assumption 2-(iv-b) and \( X_1 \) is independent of \( \theta \). From Assumption 2-(iv), \( X_1 \) denotes those characteristics that interact with product characteristics, while \( X_2 \) does not.

**Assumption 4.** Let \( X = (X_1, X_2) \) with \( X_1 = (X_{11}, \ldots, X_{1J}) \) and \( X_{1j}\perp\theta_{j'} \) for all \( j, j' \in \{1, \ldots, J\} \) and that \( X_1 \) has variation in all dimensions.

In particular, suppose the net utility of choosing \( q \) by an agent with characteristics \( X \) and unobserved \( \theta \) is

\[
V(q; \theta, X) = \sum_{j=1}^{J} \theta_j X_{1j} q_j - P(q). \tag{9}
\]

Since \( X_1 \) affects the utility, it will also affect the product line and the price functions because they depend on \( X \) (third-degree price discrimination). Once we fix the value of \( X_1 = x_1 \) (which is observed by the seller) and change the unit of measurement of product quality from \( q \) to \( \tilde{q} := x_1 \cdot q \), we can apply Theorem 3.1 to identify \( f(\theta|2) \). We can also use exogenous variation in \( X_1 \) to identify the density \( f(\theta|1) \) over the bunching region \( S^1_\theta \). So, independent variation in \( X_1 \) is an important assumption for identification. We use the notation \( f_\theta(\cdot|1) \) to make it clear that it is the density of \( \theta \in S^1_\theta \).

In the example above we saw that all agents with type such that \( \tau = \sum_{j \in J} \theta_j \) selected the same \( q(\tau) \). Now that the agents vary in \( X \), agents are bunched according to \( W = \sum_{j \in J} \theta_j X_{1j} \); in other words, all agents with the same \( W \) self select \( \rho(W) \), i.e., \( \rho(\theta) = (\rho_1(\theta), \ldots, \rho_J(\theta)) = (\rho_1(W), \ldots, \rho_J(W)) \) for all \( \theta \in S^1_\theta \). In other words, \( W \) acts as a sufficient statistic, and incentive compatibility
requires that \( q(W) \) be monotonic in \( W \) and hence invertible. So from the observed \( q \) we can determine the index \( W := \rho^{-1}(q) \). Then, the problem is to identify \( f_{\theta}(|\cdot|) \) from the the joint density \( f_{W,X_1}(\cdot,\cdot) \) of \((W,X_1)\) where

\[
W = \theta_1 X_1 + \cdots + \theta_J X_{1J}.
\]

Our objective is to identify the joint density of \((\theta_1, \ldots, \theta_J)\) from the joint density of \( W \) and \( X_1 := (X_1, \ldots, X_{1J}) \) when \( X_1 \) and \( \theta \) are independent, that is identify \( f(\theta|1) \) from the subset of the \((J - 1)\)-dimensional orthogonal projections of \( \theta \).

Below we present an identification argument that uses the integral transform of such projections.

Multiplying both sides of \( W \) by \( ||X_1||^{-1} \) and using \( D := ||X_1||^{-1} X_1 \), which is an element of a \( J \)-dimensional unit sphere \( S_{J-1} := \{ \omega \in \mathbb{R}^J : ||\omega|| = 1 \} \), we get

\[
\widetilde{B} := \{ b = \theta d \in \mathbb{R} \}
\]

Then the conditional density of \( B \) given \( D = d \) is

\[
f_{B|D}(b|d) = \int_{S_{\theta}^b} f_{B|D,\theta}(b|d, \theta) f_{\theta}(\theta|1) d\theta = \int_{\{b = \theta d\}} f_{\theta}(\theta|1) d\sigma(\theta) := Rf_{\theta}(b, d).
\]

Since we observe \((W,X_1)\) and hence \((B,D)\) we can identify \( f_{B|D}(\cdot|\cdot) \) in the LHS, and hence the RHS \( Rf_{\theta}(b, d) \). This equation is known in the literature as the Radon transform of \( f_{\theta}(\cdot|1) \), which is invertible (Bracewell, 1990; Helgason, 1999) which identifies \( f_{\theta}(\cdot|1) \) as long as there is sufficient variation in \( X_1 \).\(^{15}\)

This argument is formalized below. Let

\[
Ch_{RF}(\xi) = \int_{-\infty}^{\infty} e^{-2\pi ib\xi} Rf(d, b) db; \quad Ch_{f}(\xi d) = \int_{-\infty}^{\infty} e^{-2\pi i(\theta|d)} f_{\theta}(\theta|1) d\theta,
\]

be the Fourier transforms of \( Rf(d, b) \) and \( f_{\theta}(\cdot|1) \) evaluated at \( \xi \) and \( \xi d \), for a fixed \( d \), respectively. Since \( f_{B,D}(\cdot,\cdot) \) is known we can treat \( Ch_{RF}(\xi) \) as known, but \( Ch_{f}(\cdot) \) is unknown. For a function of two variables \( h(y_1, y_2) \), its Fourier transform is

\[
Ch_{h}(t_1, t_2) = \int h(y_1, y_2) e^{-2\pi i(t_1 y_1 + t_2 y_2)} dy_1 dy_2,
\]

and let its projection of \( h(\cdot,\cdot) \) on to the \( y_1 \) axis be \( \tilde{p}(x) = \int h(y_1, y_2) dy_2 \) so that

\[
Ch_{h}(t_1, 0) = \int h(y_1, y_2) e^{-2\pi i t_1 y_2} dy_1 dy_2 = \int_{-\infty}^{\infty} \tilde{p}(y_1) e^{-2\pi i t_1 y_1} dy_1 = Ch_{\tilde{p}}(t_1).
\]

\(^{15}\) For instance suppose \( X_1 \) is a vector of constants \((a_1, \ldots, a_J) \in S_{J-1} \), then it is clear that we cannot identify \( f_{\theta}(\cdot|1) \) from \( B = a_1 \theta_1 + \cdots + a_J \theta_J \).
With a slight abuse of notation for the ease of exposition, ignoring the differences where the functions should be evaluated, note that $Ch_h(t_1, 0)$ is $Ch_{RF}(t_1)$ and $Ch_p(t_1)$ is $Ch_f(t_1)$, hence (the unknown) $Ch_f(\cdot)$ is equal to (the known) $Ch_{RF}(\cdot)$. Formally, this is the Projection Slice Theorem (Bracewell, 1956) which implies that for a fixed $d$, $Ch_f(\xi d) = Ch_{RF}(\xi)$. Therefore we can identify $f_\theta(\cdot|1)$ as the Fourier inverse of $Ch_f(\cdot)$:

$$f_\theta(\theta|1) = \int_{-\infty}^{\infty} e^{2\pi i \theta \xi} Ch_{RF}(\xi) d\xi.$$ 

Intuitively, the identification exploits the fact that two consumers with same $\theta$ but different $X_1$ face different menus and different choices. So if we consider the population with fixed $X_1$, the variation in the choices must be due to the variation in $\theta$. But as we change $X_1$ from $x_1$ to $x_1'$, the choices change but variation in $\theta$ remains the same, because $X_1 \parallel \theta$. So with continuous variation in $X_1$, we have infinitely many “moment conditions” for $\theta$, which allows us to express the conditional choice density given $X_1$ as a (mixture) Radon transform of $f_\theta(\cdot|1)$ with the marginal density of $X_1$ as the mixing density. Similar identification problem is addressed in Ichimura and Thompson (1998); Gautier and Kitamura (2013) and Dunker, Hoderlein, and Kaido (2015). These papers use two related methods for identification: either the Cramér-Wold device or the Radon transform, as we do here; for more see Belisle, Masse, and Ransford (1997) and Boman and Lindskog (2009).

**Theorem 3.2.** Under Assumptions 2-(i)–(iv-b), (v), and 3 and 4 the densities $f_\theta(\cdot|1), f_\theta(\cdot|2)$ and the cost function $C(\cdot)$ are nonparametrically identified.

3.4. **Nonlinear Utility.** In this section we consider the model with nonlinear utility (Assumption 2-(iv-c)). Here the (gross) utility is equal to $X_1 \cdot v(q; X_2)$. To keep the arguments clear, suppress the dependence on $X$ and focus on the identification of $v(q)$ on $S^2_{\theta}$. Once the variation in the data that drives identification is known we can introduce $X$ and consider the possibility of over-identification.

When the product characteristics enter nonlinearly, equations (1) and (7) are insufficient to identify $[F_\theta(\cdot|2), C(\cdot), v(\cdot)]$. Identification fails because of the substitutability between the type $\theta$ and the curvature of the utility function $v(\cdot)$: marginal utility could change either because of $\theta$ or because of $v(\cdot)$. 


Lemma 1. Under Assumptions 2-(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 \( S_\theta^2 \), respectively, are not identified.

Proof. Since the optimality condition (3) determines the cost function, we can treat the cost function as known. Let \( J = 2 \) and \( v_j(q_j) = q_j^{\omega_j}, \omega_j \in (0, 1) \) so that

\[
V(q; \theta) = \theta_1 v_1(q_1) + \theta_2 v_2(q_2) - P(q_1, q_2) = \theta_1 q_1^{\omega_1} + \theta_2 q_2^{\omega_2} - P(q_1, q_2)
\]

is the utility function and let \( F_\theta(\cdot|2) \) and \( f_\theta(\cdot|2) \), respectively, be the distribution and density of \( \theta \in S_\theta^2 \). Observe that \( \{q_j, p_j\} \) solves the first-order condition

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

After the change of variable, the joint (truncated) density of \( (q_1, q_2) \) is

\[
m_q^*(q_1, q_2) = f_\theta \left( \frac{P_1}{\omega_1 q_1^{\omega_1-1}, \omega_2 q_2^{\omega_2-1}} \right)^2 P_1 P_2 (1-\omega_1)(1-\omega_2) \omega_1\omega_2 q_1^{\omega_1} q_2^{\omega_2}.
\]

Let \( \theta_j \equiv \theta_j \times \omega_j \sim F_\theta(\cdot|2) \), where \( F_\theta(\cdot|2) = F_\theta(\cdot/\omega|2) \) with \( \omega \equiv (\omega_1, \omega_2) \) and \( \tilde{v}(q_j) = v(q_j)/\omega = q_j^{\omega_j}/\omega_j \), denote a set of new model parameters. It is easy to check that \( \{q_j, p_j\} \) solves the first-order condition implied by \( [v(\cdot), F_\theta(\cdot)] \), and these two models are observationally equivalent because the the joint (truncated) density of \( (q_1, q_2) \) are the same:

\[
\tilde{m}_q^*(q_1, q_2) = f_\theta \left( \frac{p_1}{q_1^{\omega_1-1}, q_2^{\omega_2-1}} \right)^2 p_1 p_2 (1-\omega_1)(1-\omega_2) \omega_1\omega_2 q_1^{\omega_1} q_2^{\omega_2} = m_q^*(q_1, q_2).
\]

As seen here, we can increase the type to \( \theta \cdot \omega \) and decrease the curvature of the utility to \( v(q)/\omega \) without changing the observable choices. Therefore, data from only one market is not sufficient for identification. However, if we have data from two markets (served by the same seller) or data over two periods that are separated by an exogenous change in cost that affects the marginal cost (and the prices) then we can identify the model. Such exogenous changes in cost could be in the form of some exogenous changes in law that affects costs over two periods, or different tax or marketing expenses across these two markets.

Let \( Z \in \mathcal{S}_Z = \{ z_1, z_2 \} \) be an exogenous, and binary cost shifter that only affects the cost function \( C(\cdot; Z) \) and is independent of the consumer type and the utility function, i.e., \( Z \perp (\theta, v(\cdot)) \). 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 will be the same. Without loss of generality assume that $Z \in \{1, 2\}$, and let $\{P_1(\cdot), \rho_1(\cdot), C_1(\cdot)\}$ denote the price function, allocation rule and cost, when $Z = z_1$.

To see how $Z$ can help in the identification, consider the non-identification example in Lemma 1. As in Lemma 1, the utility function is $v(q_1, q_2) = \begin{pmatrix} v_1(q_1) \\ v_2(q_2) \end{pmatrix} = \begin{pmatrix} (q_1)^{\omega_1} \\ (q_2)^{\omega_2} \end{pmatrix}$. Let us focus only on the high-types $S^2_\theta$ and further assume that $Q^2$ is also invariant to $Z$. Then the demand-side optimality (marginal utility equals the marginal price) can be written as

$$
\begin{pmatrix}
\nabla_1 P_1(q) \\
\nabla_2 P_1(q)
\end{pmatrix} = \begin{pmatrix}
\tilde{\theta}_{11}(q) \cdot v_1'(q_1) \\
\tilde{\theta}_{21}(q) \cdot v_2'(q_2)
\end{pmatrix} = \begin{pmatrix}
\tilde{\theta}_{11}(q) \cdot \omega_1(q_1)^{\omega_1-1} \\
\tilde{\theta}_{21}(q) \cdot \omega_2(q_2)^{\omega_2-1}
\end{pmatrix}, \ell = 1, 2.
$$

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

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

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

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

Equation (10) captures the fact that if two consumers buy the same $q$ (one when $Z = z_1$ and the other when $Z = z_2$) and if the marginal price for $q$ is higher (say) when $Z = z_1$ than when $Z = z_2$ then it must be the case that the first consumer’s type $\tilde{\theta}_1(q)$ must be higher than the second consumer’s type $\tilde{\theta}_2(q)$. So, if we know $\theta$’s choice $q = q_1(\theta)$ when $Z = z_1$ then we can use the curvature of the pricing functions to determine $\theta$ that chooses the same $q$ when $Z = z_2$.

Consider the supply side for the high-types where the IC constraint implies $\rho$ is bijective and hence:

$$
F_\rho(t|2) = F_\rho(t_1, t_2|2) = \Pr(\theta_1 \leq t_1, \theta_2 \leq t_2|S^2_\theta) = \Pr(\rho(\theta, z_\ell) \leq \rho(t, z_\ell)|Q^2)
$$

$$
= \Pr(q_1 \leq \rho(t, z_\ell)) = \Pr(q_1 \leq \rho_1(t, z_\ell), q_2 \leq \rho_2(t, z_\ell)|Q^2)
$$

$$
= M_\ell^*(\rho_\ell(t)), \ell = 1, 2,
$$

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

$$
M_1^*(\rho_1(t)) = M_2^*(\rho_2(t)).
Hence, the (multivariate) quantiles of the choice distribution when $Z = z_2$

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

(11)

and since $(M_1^*)^{-1} \circ M_2^*(\cdot)$ is identified, we can identify $\rho_1(\theta)$ if we know $\rho_2(\theta)$. Therefore, the difference, $((M_1^*)^{-1} \circ M_2^*(\rho(\tau)) - \rho(\tau))$, measures the change in $q$ when $Z$ moves from $z_2$ to $z_1$ while fixing the quantile of $q$ at $\tau$. This variation (11) together with (10) can be used to first identify $\tilde{v}$ with the utility function $Z$. Hence, the (multivariate) quantiles of the choice distribution when

$\theta$

when

Therefore, the difference, $((M_1^*)^{-1} \circ M_2^*(\rho(\tau)) - \rho(\tau))$, measures the change in $q$

The intuition behind identification is as follows: start with a normalization $\theta^0 \equiv \tilde{\theta}_2(q^0)$ for some bundle $q^0 = (q_0^0, q_2^0) \in Q_2$, and determine $\nabla P_1(q^0), \nabla P_2(q^0)$, the quantile $\tau = M_2^*(q^0)$, and $\theta^1 \equiv \tilde{\theta}_1(q^0)$ from (10). Using (11), determine $q^1$ with the same quantile $\tau$ under $Z = z_1$. Then, for $q^1$ determine $\nabla P_1(q^1)$ and $\nabla P_2(q^1)$, which can determine $\theta^2 = \tilde{\theta}_2(q^1) = \nabla P_2(q^1) \circ (\nabla P_1(q^1))^{-1} \circ \theta^1$ (inverse of (10)). Iterating these steps, we identify a sequence $\{\theta^0, \theta^1, \ldots, \theta^\ell, \ldots\}$ and the corresponding quantile. If these sequences form a dense subset of $Q_2$ then the function $\hat{\theta}(\cdot) : Q_2 \times S_Z \to S_\theta$ is identified everywhere. We formalize this intuition for $J \geq 2$ below, starting with the assumption about exclusion restriction.

**Assumption 5.** Let $Z \in S_Z = \{z_1, z_2\}$ be independent of $\theta$, and $Z$ not an argument in $v(q)$.

Consumer optimality implies $\nabla P_\ell(q) = \tilde{\theta}_\ell(q) \circ \nabla v(q)$, and the general version of Equation (10) can be written as

$$\tilde{\theta}_\ell(q) = \nabla P_\ell(q) \circ \tilde{\theta}_\ell(q) \circ (\nabla P_\ell(q))^{-1}$$

$$\equiv r_{\ell,\ell}(\tilde{\theta}_\ell(q), q) = \begin{pmatrix} r_{\ell,\ell}^1(\tilde{\theta}_\ell(q), q) \\ \vdots \\ r_{\ell,\ell}^\ell(\tilde{\theta}_\ell(q), q) \end{pmatrix}. \quad (12)$$

Next, Assumption 5 and the incentive compatibility condition for high-types imply $F_\ell(t|2) = M^*(q(t; z_\ell); z_\ell), \ell = 1, 2$ and hence

$$M_\ell^*(\rho_\ell(t)) := M^*(\rho(t; z_\ell); z_\ell) = M^*(\rho(t; z_\ell); z_\ell) := M_\ell^*(\rho_\ell(t)). \quad (13)$$

Once we determine multivariate quantiles, (13) generalizes (11). Quantiles are the proper inverse of a distribution function, but defining multivariate quantiles

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

For the definition of multivariate quantiles (Koltchinskii, 1997) see Appendix A. (Koltchinskii, 1997) showed 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 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 $M_1(\cdot) = M_2(\cdot)$.

Hence once $g_M(\cdot)$ is chosen, (13) and (20) imply
\[
\rho_1(\tau) = (\partial g_{M_1}^{-1}(M_2^*(\rho_2(\tau)))) = s_{2,1}(\rho_2(\tau)), \quad \tau \in (0, 1).
\]
This means we can then use
\[
\tilde{\theta}_\ell(q) = r_{\ell,\ell}(\tilde{\theta}_\ell(q), q);
\]
\[
\rho_\ell(\tau) = s_{\ell,\ell}(\rho_\ell(\tau))
\]
to identify $\tilde{\theta}_\ell(\cdot)$, for either $\ell = 1$ or $\ell = 2$. Since for a $q$ the probability that $\{\theta \leq t|Z = z_{\ell}\}$ is equal to the probability that $\{\theta \leq r_{\ell,\ell}(t, q)|Z = z_{\ell}\}$, i.e., $\Pr(\theta \leq t|Z = z_{\ell}) = \Pr(\theta \leq r_{\ell,\ell}(t, q)|Z = z_{\ell})$, it means
\[
\rho_\ell(r_{\ell,\ell}(\theta, q)) = s_{\ell,\ell}(\rho_\ell(\theta)).
\]
If we know $\rho_{\ell}(\cdot)$ at some $\theta$ then we can identify $\rho_\ell(\cdot)$ at $r_{\ell,\ell}(\theta, q)$. As mentioned earlier, we need a normalization, so let $v(q^0) = q^0$ for some $q^0 \in Q^2$ so that we know $\{q^0, \theta^0 = \tilde{\theta}_1(q^0)\}$.\footnote{We can also normalize some quantile of $F_\theta(\cdot)$.} This identifies $\{q^1, \tilde{\theta}_1(q^1)\}$ where $q^1 = s_{1,2}(q^0)$ and $\tilde{\theta}_1(q^1) = r_{2,1}(\theta^0, q^1)$, which further identifies $\{q^2, \tilde{\theta}_1(q^2)\}$ with $q^2 = s_{1,2}(q^1)$ and $\tilde{\theta}_1(q^2) = r_{2,1}(\tilde{\theta}_1(q^1), q^2)$ and so on. If for any choice of initial quantile $\rho(\tau) \in Q^2$ we can identify $\tilde{\theta}(\rho(\tau))$, possibly by constructing a sequence as above, then this completes the identification argument.

For that we can exploit Assumption 5, which implies that for some $\theta$ the difference $(\theta - r_{2,1}(\theta, q))$ measures the resulting change in $\theta$ if we switch from $z_2$ to $z_1$ for a fixed $q$. This allows us to trace $\tilde{\theta}(\cdot)$ as we move back and forth between $z_2$ and $z_1$. But for this “tracing” procedure to identify $\tilde{\theta}(\cdot)$ we want this procedure to eventually stop, which is the same thing as saying that there
be \( \hat{q} \in Q^2 \) (fixed point) such that \( (\theta(\cdot) - r_{2,1}(\theta, \cdot)) = 0 \). However, since \( \theta \) is multidimensional, we need the fixed point to be stable (or attractive).

**Assumption 6.** (a) There exist \( \hat{q} \in Q^2 \) such that \( r_{\ell, \ell'}(\theta(\hat{q}), \hat{q}) = \theta(\hat{q}) \); and (b) \( \text{sgn}[(r_{\ell, \ell'}(q_j) - q_j)(q_j - \hat{q}_j)] \) is independent of \( j \in \{1, \ldots, J\} \).

For Assumption 6-(a) to hold, it is sufficient that for some \( \hat{q} \) the marginal prices across two cost regimes are the same, i.e., \( \nabla P_1(\hat{q}) = \nabla P_2(\hat{q}) \), which is verifiable from the data. The second part of Assumption 6-(b) ensures that the fixed point is stable as log as the slopes of the \( j^{th} \) component of \( r_{\ell, \ell'}(\cdot) \) depends only on the difference \( (q_j - \hat{q}_j) \), irrespective of what \( j \) is. Since \( r_{\ell, \ell'}(\cdot) \) is a function of the marginal price, Equation (12), this condition implies that the slope of the marginal prices should depend only on the distance between \( q_j \) (for arbitrary \( j \)) and the fixed point \( \hat{q}_j \). Anytime there is an iterative procedure in a multidimensional setting, for the procedure to converge to the same fixed point irrespective of the starting point we always need such assumption.\(^{18}\)

Without loss of generality, assume the initial normalization to be the fixed point \( \hat{q} \), so that \( \theta^0 = \bar{\theta}(\hat{q}) \) is known. In other words, \( \theta^0 \) is such that \( \text{sgn}(\theta^0, z_1) = \hat{q} \). From Assumption 5, suppose \( \nabla P_1(q) \circ \nabla P_2(q)^{-1} << 1 \), whenever \( q << \hat{q} \). Then, for \( r^{th} \) quantile \( q(\tau) < \hat{q} \):

\[
\bar{\theta}_1(\tau) := (q(\tau))^{-1}(q(\tau); z_1) = \bar{\theta}_1(q^0) = r_{1,2}(\bar{\theta}_1(q^1), q^1)
\]

\[
= [\nabla P_2(q^1) \circ \nabla P_1(q^1)^{-1}] \circ \bar{\theta}_1(q^1)
\]

\[
= [\nabla P_2(q^1) \circ \nabla P_1(q^1)^{-1}] \circ [r_{1,2}(\bar{\theta}_1(q^2), q^2)]
\]

\[
= [\nabla P_2(q^1) \circ \nabla P_1(q^1)^{-1}] \circ [\nabla P_2(q^2) \circ \nabla P_1(q^2)^{-1}] \circ [r_{1,2}(\bar{\theta}_1(q^2), q^2)]
\]

\[
= \cdots
\]

\[
= \lim_{L \to \infty} [\nabla P_2(q^L) \circ \nabla P_1(q^L)^{-1}] \circ \cdots \circ [\nabla P_2(q^L) \circ \nabla P_1(q^L)^{-1}] \circ [r_{1,2}(\bar{\theta}_1(q^{L+1}), q^{L+1})]
\]

\[
= \left\{ \prod_{L=1}^\infty \nabla P_2(q^L) \circ \nabla P_1(q^L)^{-1} \right\} \lim_{L \to \infty} \bar{\theta}_1(s_{1,2}(q^{L+1}))
\]

\[
= \left\{ \prod_{L=1}^\infty \nabla P_2(q^L) \circ \nabla P_1(q^L)^{-1} \right\} \lim_{L \to \infty} \theta_0.
\]  \hspace{1cm} (16)

\(^{18}\) For example, in a recent paper D’Haultfouille and Février (2011) use the same stability condition to identify a nonseparable model with a discrete instrument and multivariate errors.

I want to thank Xavier D’Haultfouille for pointing out the connection to me.
where the first equality is simply the definition, the second equality is the normalization, the third equality follows from (15) with $q_1 := s_{1,2}(q^0 = \rho(\tau))$ so that $\hat{\theta}_1(q^1) = r_{1,2}(\hat{\theta}_1(q^1), q^1)$, and the fourth equality follows from (12). Repeating this procedure $L$ times leads to the seventh equality. The last equality follows from: a) $q^L = s_{1,2}(q^{L-1})$; b) $q(\tau) < \hat{q}$; c) $s_{1,2}(\cdot)$ is an increasing continuous function so $\lim_{L \to \infty} s_{1,2}(q^L) = s_{1,2}(q^\infty) = s_{1,2}(\hat{q})$; and d) $\hat{\theta}_1(q) = \theta_0$. Since the quantile $\tau$ was arbitrary, we identify $\hat{\theta}_1(\cdot)$.

Once the quantile function of $\theta$ is identified, we can identify $C(\cdot, Z)$ as before (with linear specification). The optimality condition $\alpha(\theta) = 0$ and Equation (6) give
\[
\text{div} \left\{ \frac{m^*_k(q)}{|\text{det}(D\theta_k)(q)|} (\hat{\theta}_k(q) \nabla v(q) - \nabla C(q; z_k)) \right\} = -\frac{m^*(q)}{|\text{det}(D\theta)(q)|}.
\]
Differentiating $\theta_k \circ \nabla v(q) = \nabla P_k(q)$ with respect to $q$ gives
\[
D\nabla P_k(q) = D\hat{\theta}_k(q) \circ \nabla v(q) + \hat{\theta}_k(q) \circ D\nabla v(q)
\]
\[
D\hat{\theta}_k(q) \circ \nabla v(q) = D\nabla P_k(q) - \hat{\theta}_k(q) \circ (\nabla v(q)) \circ (\nabla v(q))^{-1} \circ D\nabla v(q)
\]
\[
D\hat{\theta}_k(q) = D\nabla P_k(q) \circ (\nabla v(q))^{-1} - \nabla P_k(q) \circ (\nabla v(q))^{-2} \circ D\nabla v(q)
\]
which identifies $|\text{det}(D\hat{\theta})(q)|$. Then, substituting $|\text{det}(D\hat{\theta})(q)|$ in above gives
\[
\text{div} \left\{ \frac{m^*(q)}{|\text{det}(D\theta)(q)|} (\nabla P(q) - \nabla C(q)) \right\} = -\frac{m^*(q)}{|\text{det}(D\theta)(q)|},
\]
(a partial differential equation for $C(\cdot, z_k)$), with boundary condition
\[
\frac{m^*_k(q)}{|\text{det}(D\theta_k)(q)|} (\nabla C(q; z_k) - \nabla P_k(q)) \cdot \hat{n} \cdot (\nabla P_k(q)) = 0, \forall q \in \partial Q^2.
\]
This PDE has a unique solution $C(q)$.

**Theorem 3.3.** Under Assumptions 2-(i)-(iv-c) and (v) and Assumptions 3–6, $[F_\theta(\cdot|2), v(\cdot), C(\cdot; Z)]$ are identified, up to the convex function $g_M(\cdot)$ used to define multivariate quantiles.

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 extend the utility function from $Q^2$ to $Q^2 \cup Q^1$. For the identification strategy, if $v(\cdot)$ is a real analytic, like the cost function, then we can extend the domain of $v(\cdot)$ to include $Q^1$.

**Assumption 7.** Let the utility function $v(\cdot, X_2)$ be a real analytic function.

Then, under Assumption 7, we can change the unit of measurement from $q$ to $\bar{q} \equiv v(q, X_2)$, then apply Theorem 3.2 with gross utility $\sum_{j \in [J]} \theta_j X_{1j} \bar{q}_j$.  


3.5. Overidentification. Now that we know identification depends on how many cost shifters we have and whether or not the gradient of the pricing functions cross each other, the next step is to analyze the effect of observed characteristics $X$ on identification. Let us assume that the nonlinear utility model is identified. We are interested in determining if when the utility function depends on $X$, and if $X$ is independent of $\theta$, does this variation in $X$ imply the model is over identified.

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

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

$$
\int_{S^2_\theta} L(q, \theta) d\theta = M^*_{q;X}(q|\cdot); \quad \int_{Q^2_X} L(q, \theta) dq = F_\theta(\theta|2).
$$

To that end, consider the following optimization problem:

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

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

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

such that the solution maximizes the (conditional) covariance between $\theta$ and $q$. When we minimize the quadratic distance, or equivalently the covariance, our objective is to find an optimal way to “transport” $q$ to $\theta$. Let $\delta[\cdot]$ be a Dirac measure or a degenerate distribution. Brenier (1991) and McCann (1995) show that there exists a unique convex function $\Gamma(q, X)$ such that $dL(q, \theta) = dM^*_{q;X}(q) \delta[\theta = \nabla_q \Gamma(q, X)]$ is the solution. Therefore for all $q \in Q^2_X$ we can determine its inverse $\theta = \nabla_q \Gamma(q, X)$, which identifies $F_\theta(\cdot|2)$. \qed
Therefore $\Gamma(q, X)$ can be used to test the supply-side optimality. There are many ways to think of a “specification test.” One way is by verifying that $\nabla_q \Gamma(q, X)$ (instead of $\theta$) in Equation (3) leads to the same equilibrium $\rho(\theta; X)$.

4. Model Restrictions

In this section we derive the restrictions imposed by the model on observables under Assumption 2-(iv) –a, b, and c, respectively. These restrictions can be used to test the model validity. For every agent we observe $[p_i, q_i, X_i]$ and for the seller we observe $\{z_1, z_2\}$. From the model $p_i$ and $q_i$ are given by $p = P(\theta)\left(q, z_\ell\right)$ and $q = \rho(\theta, z_\ell)$. Specifically, suppose a researcher observes a sequence of price and quantity data, and some agents and cost characteristics. Does there exist any possibility to rationalize the data such that the underlying screening model is optimal when the utility function satisfies Assumption 2-(iv-a) (Model 1), Assumption 2-(iv-b) (Model 2), or Assumption 2-(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 $(Z, X_i, \{q_i, p_i\})$ 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 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 $D_1 = (q, p), D_2 = (q, p, X_1), D_3 = (q, p, X, Z)$ distributed, respectively, as $\Psi_{D_\ell}(\cdot), \ell = 1, 2, 3,$ and let

\[
\begin{align*}
\mathcal{M}_1 &= \{(F_\theta(\cdot), C(\cdot)) \in \mathcal{F} \times \mathcal{C} : \text{satisfy Assumption 2-} - (i) - (iv - a), (v)\} \\
\mathcal{M}_2 &= \{(F_\theta(\cdot), C(\cdot)) \in \mathcal{F} \times \mathcal{C} : \text{satisfy Assumption 2-} - (i) - (iv - b), (v)\} \\
\mathcal{M}_3 &= \{(F_\theta(\cdot), C(\cdot), Z) \in \mathcal{F} \times \mathcal{C}_Z : \text{satisfy Assumptions 2-} - (i) - (iv - c), (v), 4 \text{ and } 5\}
\end{align*}
\]

Define the following conditions:

\begin{enumerate}
\item $\Psi_{D_\ell}(\cdot) = \delta[p = P(q)] \times M(q), \text{ with density } m(q) > 0 \text{ for all } q \in Q^1 \cup Q^2.$
\item There is a subset $Q^1 \subsetneq Q$ which is a $J - 1$ dimensional flat (hyperplane) in $\mathbb{R}^J_+$. \\
\item $p = P(q)$ has non-vanishing gradient and Hessian for all $q \in Q^2$. \\
\item Let $\{W\} := \{\nabla P(q) : q \in Q^2\}$. Then $F_W(w) = \Pr(W \leq w) = M^*(q)$ and let $m^*(\cdot) > 0$ be the density of $M^*(\cdot)$.
\end{enumerate}
C5. Let $C(\cdot)$ be the solution of the differential equation
\[ \text{div} \left\{ \frac{m^*(q)}{|\text{det}(D\nabla P(q))|} (\nabla P(q) - \nabla C(q)) \right\} = -\frac{m^*(q)}{|\text{det}(D\nabla P(q))|}, \] (18)
with boundary conditions
\[ \frac{m^*(q)}{|\text{det}(D\nabla P(q))|} (\nabla C(q) - \nabla P(q)) \cdot \hat{n} (\nabla P(q)) = 0. \]

4.1. Linear Utility. For every consumer we observe $D_1$, and our objective is to determine the necessary and sufficient conditions on the joint distribution $\Psi_{D_1}(\cdot, \cdot)$ to be rationalized by model $M_1$.

Lemma 4.1. If $M_1$ 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 C1 – C5 then there is a model $M_1$ that generates $D_1$.

Proof. The if part. Since $F_\theta(\cdot)$ is such that the density $f_\theta(\cdot) > 0$ everywhere on $S_\theta$ and the equilibrium allocation rule $\rho : S_\theta \to Q$ is onto and continuous, the CDF $M(q)$ is well defined and the density $m(q) > 0$. Moreover, since the equilibrium allocation rule is deterministic, for every $q$ there is only one price $P(q)$, hence the Dirac measure, which completes C1. Rochet-Choné showed that in equilibrium the bunching set $Q^1$ is nonempty, and hence $m(q) > 0, q \in Q^1$.

Moreover, the allocation rule $\rho : S^1_\theta \to Q^1$ is not bijective, and as a result, $Q^1$ as a subset of $\mathbb{R}^J_+$ is flat, which completes C2. The optimality condition for the types that are perfectly screened is $\theta = \nabla P(q) := \hat{\theta}(q)$, and incentive compatibility implies that the indirect utility function is convex and $P(q)$ has non-vanishing gradient and Hessian, which completes C3. Then, $M^*(q) = \Pr(q \leq q) = \Pr(\nabla P(q) \leq \nabla P(q)) = \Pr(W \leq w) = F_W(w)$, hence C4. Finally, using (6) to replace $f_\theta(\cdot)$ in $\alpha(\theta) = 0, \forall \theta \in S^2_\theta$, with the boundary condition $\beta(\theta) = 0, \forall \theta \in \partial S^2_\theta \cap \partial S_\theta$, we get C5.

The only if part. Now, we show that if $\Psi_{D_1}(\cdot)$ satisfies all C1 – C5 conditions listed above then we can determine a model $M_1$ that rationalizes $\Psi_{D_1}(\cdot)$. Let $C(\cdot)$ satisfy C5, then we can determine the cost function $C(\cdot)$. Moreover, it is real analytic so it can be extended uniquely to all $Q$. From C4. we can determine the vector $W$, which is also the type $\theta$, and it satisfies the first-order optimality condition. Thus, the indirect utility of the type $\theta$ that corresponds to the choices $q \in Q^2$ is convex, and therefore satisfies the incentive compatibility constraint. Moreover, since $m^*(q) > 0$ the density $f_\theta(\cdot|2) > 0$ and
\[ F_\theta(\cdot|2) = \int_{\theta \in [W]} \nabla P(q, q \in Q^2)} f_\theta(\theta|2) d\theta. \] As far as \( F_\theta(\cdot|1) \) is concerned, we can simply ignore bunching and define \( F_\theta(\theta) = M(q|q \in Q^1) \) where \( \theta = \nabla P(q) \).

### 4.2. Bi-Linear Utility

Now consider the case of bi-linear utility function. Since \( X_2 \) is redundant information, we ignore it. The only difference between this and the previous model is that now there is \( X \) and the notations we can still use the same conditions \( C1. - C5. \) as long as they are understood with respect to \( D_2 \). For instance, \( C1. \) becomes \( \Psi_{D_2}(\cdot) = \delta[p = P(q; X_1)] \times M(q) \times \Psi_{X_1}. \)

**Lemma 4.2.** If \( M_2 \) rationalizes \( \Psi_{D_2}(\cdot) \) then \( \Psi_{D_2}(\cdot) \) satisfies conditions \( C1. - C5. \) Conversely, if \( F_\theta(\cdot|0) \) is known, \( \text{dim}(X_1) = \text{dim}(q) = J, \) and \( \Psi_{D_2}(\cdot) \) satisfies \( C1. - C5 \) then there is a model \( M_2 \) that generates \( D_2 \).

The proof of this lemma is very similar to that of Lemma 4.1, except here the menu (allocation and prices) depends on \( X_1 \) but the cost function and the conditional density \( f_\theta(\cdot|1) \) can be determined from the data. Since the argument is same as above, the proof is omitted.

### 4.3. Nonlinear Utility

Finally, consider the case of nonlinear utility and introduce two more conditions:

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

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

\[
\max_{\ell(q, \theta) \in \mathcal{L}(Q^2, S_\delta)} \mathbb{E}(\theta \cdot q|X),
\]

where \( \mathcal{L}(Q^2, S_\delta) \) is as defined in (17) is given by a mapping \( \theta = \nabla q \Gamma(q, X) \) for some convex function \( \Gamma(q, X) \) 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 \( M_3 \) rationalizes \( \Psi_{D_3}(\cdot) \) then \( \Psi_{D_3}(\cdot) \) satisfies \( C1. - C3., C4'. - C6. \) Conversely, if \( F_\theta(\cdot|0) \), and a quantile \( \tilde{\theta}(q) \) are known, \( \text{dim}(X_1) = \text{dim}(q) = J, Q^2_{X,z_k} = Q^2_{X,z_\ell}, \) (common support) and \( \Psi_{D_3}(\cdot) \) satisfies \( C1. - C3., C4'. - C6., \) then there exists a model \( M_3 \) that rationalizes \( \Psi_{D_3}(\cdot) \).

**Proof. The if part.** The CDF is \( F_\theta(\cdot) \) with density \( f_\theta(\cdot) > 0 \) everywhere on the support \( S_\theta. \) Moreover, the equilibrium allocation rule \( \rho : S_\theta \times X \times Z \rightarrow Q \) is
onto and continuous for given \((X, Z)\). Therefore the CDF \(F_{q;X,Z}(\cdot|\cdot,\cdot)\) is a push forward of \(F_\theta(\cdot)\) given \((X, Z)\). Since \(Q = Q^2_{X,Z} \cup Q^1_{X,Z} \cup \{q_0\}\) the (truncated) density \(m_{q;X,Z}(q|\cdot,\cdot) > 0\) for all \(q \in Q^2_{X,Z} \cup Q^1_{X,Z}\). In equilibrium, for a given \((q, X, Z)\) the pricing function is deterministic, therefore the distribution is degenerate at \(p = P(q; X, Z)\), i.e., a Dirac measure. This completes C1. For C2, note that the allocation rule is not bijective, and as a result \(\rho(S^1_\theta; X, Z) = Q^1 \subseteq \mathbb{R}_+^d\) is a hyperplane. For the high-types, optimality requires that the marginal utility \(\theta \cdot v(q; X_2)\) is equal to the marginal price \(P(q; X, Z)\), and since \(v(\cdot; X_2)\) has non-vanishing Hessian, \(P(\cdot; X, Z)\) also has non-vanishing gradient \(\nabla P(\cdot; X, Z)\) and Hessian, which completes C3. Since \(Z \parallel \theta\), using Equation (13) gives \(F_\theta(\xi|2) = M_{q;X,Z}(\rho_1(\xi)|X, z_1) = M_{q;X,Z}(\rho_2(\xi)|X, z_2)\), as desired for \(C4'\). The condition \(C5\), follows once we replace \(m^*(\cdot)\) and \(P(q)\) in (18) with \(m^*_{q;X,Z}(\cdot|\cdot,\cdot)\) and \(P(q; X, Z)\), respectively, and observe that for any pair \((X, Z)\) the equilibrium for high-type is given by \(\alpha(\theta) = 0\). Since \(F_\theta(\cdot|2)\) is known and \(M^*_{q;X,Z}(\cdot)\) is determined, condition \(C6\) follows from Lemma 2.

**The only if part.** We want to show that if \(\Psi_{D_3}(\cdot)\) satisfies all conditions in the statement, then we can construct a model \(M_3\) that rationalizes \(\Psi_{D_3}(\cdot)\). At \(Z = z_k\), using condition \(C6\), we can determine two cost functions \(C(\cdot, z_1)\) and \(C(\cdot, z_2)\). Since (18) is applicable only to \(Q^2_{X,Z}\), we need to extend the domain of the cost function. Of many ways to extend the domain, the simplest is to assume that the cost is quadratic, i.e., \(C(q; X, Z) = 1/2 \sum_{j=1}^J q_j^2\) for all \(q \in Q^1_{X,Z} \cup \{q_0\}\). Using the exclusion restriction and (16) for all \(q \in Q^2_{X,Z}\), we determine the function \(\hat{\theta}(q_r; Z = z_k)\) along a set \(\hat{Q}^2_{X,Z} \subseteq Q^2_{X,Z}\) for \(k = 1, 2\). If the set \(\hat{Q}^2_{X,Z}\) is a dense subset then there is a unique extension of \(\hat{\theta}(\cdot; \cdot)\) over all \(Q^2_{X,Z}\). If not, then, let us linearly extend the function to the entire domain of \(Q^2_{X,Z}\). Then, define \(v(q; X_2) = \nabla P(q; X, Z) \circ (\hat{\theta}(q))^{-1}\). Finally, to extend the function to \(Q\), we assume that each function \(v_j(q_j; X) = q_j^{1/2}, j = 1, \ldots, J\) for all \(q \in Q^1_{X,Z} \cup \{q_0\}\). As far as \(F_\theta(\cdot|1)\) is concerned, we ignore bunching and define \(F_\theta(\theta) = M(q|q \in Q^1)\) where \(q \in Q^1\) is such that \(\theta = \nabla P(q; X, Z) \circ (v(q; X_2))^{-1}\). Since the probability of \(q = \{q_0\}, q \in Q^1_{X,Z}\) and \(q \in Q^2_{X,Z}\) is equal to the probability of \(\theta \in S^1_\theta, \theta \in S^1_\theta\) and \(\theta \in S^2_\theta\), respectively, we can determine \(F_\theta(\cdot)\).

It is easy to verify that these three terms thus constructed belong to \(M_3\). □

**5. Extensions and Discussion**

5.1. **Unobserved Taste Shifter.** So far we have assumed that \(\theta\) is the exhaustive list of consumers’ taste profile. But it is possible to imagine a situation
where some market level characteristics, denoted as $Y \in \mathbb{R}^{++}$, affects $\theta$, but it is unobserved by the researcher. Suppose further that $Y$ enters the consumer heterogeneity multiplicatively, and is common to all consumers who observe it.

**Assumption 8.**

1. Let the consumer heterogeneity $(\theta, Y) \sim F_{\theta, Y}(\cdot, \cdot)$ with strictly positive density $f_{\theta, Y}(\cdot, \cdot)$ over the support $\mathcal{S}_\theta \times \mathbb{R}^{++}$ and $\mathbb{E}(\log Y) = 0$.
2. Let $\theta^* := Y \times \theta$ be such that $\theta^* \sim F_{\theta^* \mid Y}(\cdot \mid y) = F_{\theta^*}(\cdot)$.

These assumptions, including the location normalization $\mathbb{E}(\log Y) = 0$ are widely used in the literature on measurement error models; see, for example, Carroll, Ruppert, Stefanski, and Crainiceanu (2006). Let $\mathcal{S}_{\theta^* \mid Y}$ denote the types that are perfectly screened. Then, under Assumption 8, optimality of these types means $\theta^*_i = \nabla P(q_i)$, and since $\theta^*_i = \theta_i y, i \in [N_2]$, we want to identify $F_{\theta}(\cdot)$ and $F_Y(\cdot)$ from above. Dividing $[N_2]$ into two parts and reindexing $\{1, \ldots, N_{21}\}$ and $\{1, \ldots, N_{22}\}$ and taking the log of the above we get

$$\log \theta^*_i = \log \theta_i + \log Y, \quad i = 1, \ldots N_{2j}, j = 1, 2.$$

Let $Ch(\cdot, \cdot)$ be the joint characteristic function of $(\log \theta_1, \log \theta_2)$ 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 shorthand for $\theta_{i j}, i_j \in [N_{2j}]$. Then from Kotlarski (1966) we know that

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

which together with $Ch_{\log \theta_1}(\xi) = \frac{Ch(\xi, 0)}{Ch_{\log Y}(\xi)}$, identifies $Ch_{\log \theta_1}(\xi)$, which in turn identifies $F_{\theta}(\cdot)$ as a Fourier inverse of $Ch_{\log \theta_1}(\xi)$. This result is summarized in the following Lemma.

**Lemma 5.1.** Under Assumption 8, the model $[F_{\theta}(\cdot), F_Y(\cdot), C(\cdot), v(\cdot)]$ with unobserved heterogeneity is identified.

5.2. **Measurement Errors.** So far, we have assumed that prices and the product characteristics are measured without error. It is possible that store-level price might be different from the out-of-pocket price. Similarly, there might be errors in how product characteristics are recorded in the data. In this subsection we explore the robustness of the identification strategy with respect to measurement errors in prices and product characteristics.
5.2.1. **Measurement Error in Prices.** If only the prices are measured with additive error, and if the error is independent of the true price, then the model is still identified. That is because with additive error, the price function is

\[ P^\varepsilon(q) = P(q) + \varepsilon, \quad P(q) \perp \varepsilon, \]

so the observed marginal prices are \( \nabla P^\varepsilon(\cdot) = \nabla P(\cdot) \). Thus the previous identification argument is still valid.

**Lemma 5.2.** If \( \{ P^\varepsilon = P + \varepsilon \} \) is observed, where \( P \) is the price and \( P \perp \varepsilon \) is the measurement error, then the model parameters \([F_\theta(\cdot), C(\cdot)]\) are identified.

Now let us suppose that the data contain information about the listed (optimal price) but the consumers at the retail level paid different prices, owing to, say, discount coupons (directed advertisements). Then the price faced by the consumer would be the optimal price minus the (unobserved) discount,

\[ P^\varepsilon(q) = P(q) - \delta. \]

If there is targeted advertisement, then \( \delta \) will be correlated with \( P(q) \) and the marginal price faced by the consumer is \( \nabla P^\varepsilon(q) = \nabla P(q) - \nabla \delta \) and without \( \nabla \delta \neq 0 \) or without other form of exclusion restriction our identification strategy will not work.

5.2.2. **Measurement Error in Choices.** Now let consider the situation when the choices of \( q \) are observed with error. To simplify the problem let’s suppose \( \eta \in \mathbb{R}_+ \) and we measure \( q^\eta = q + \eta \cdot 1 \), where \( 1 \) is \( J \)-dimensional vector of ones. Let us also assume that \( \eta \perp q \) and \( \eta \sim F_\eta(\cdot) \). The data consist of \( \{ P, q^\eta \} \) pair for every consumer with type \( \theta \in \mathcal{S}_2^J \). Then \( P = P(q) = P(q^\eta - \eta \cdot 1) \) implies that \( \nabla P(q) \neq \nabla P(q^\eta) \), which means that without correcting for \( \eta \) the taste parameter \( \theta \) cannot be identified. Following the same logic as in Lemma 5.1, we can identify \( F_\eta(\cdot) \), but that still does not mean we can identify \( \theta \), because we have \( J + 1 \) unknowns and only \( J \) equations for each consumer. This result is summarized in the following Lemma.

**Lemma 5.3.** If \( \{ q^\eta = q + \eta \cdot 1 \} \) is observed, where \( \eta \perp q \) is the measurement error then the model \([F_\theta(\cdot), C(\cdot)]\) cannot be identified.

5.3. **Unobserved Product Characteristic.** Next we consider a situation where one of the \( J \) characteristics is unobserved by the econometrician. It is possible that not all product characteristics (or products, in a multiproduct interpretation) are observed. Suppose we only observe the first \( (J-1) \) attributes
\( q_{-J} \equiv (q_1, \ldots, q_{J-1}) \) but not \( q_J \), then it is possible that \( q_{-J} >> q'_{-J} \) yet their respective prices \( P \) and \( P' \) are such that \( P' > P \), as in, say, Bajari and Benkard (2005). This is because price depends on both observed \( q_{-J} \) and unobserved \( q_J \) characteristics, i.e., \( P = P(q_{-J}, q_J) \). We can write the econometrics model as

\[
P = P(q_{-J}, q_J)
\]

\[
\left( \begin{array}{c}
q_{-J} \\
q_J
\end{array} \right) = \rho(\theta_1, \ldots, \theta_{J-1}, \theta_J, F_\theta, C),
\]

(19)

with unknown (\( q_J, P(\cdot, \cdot), F_\theta(\cdot, \cdot), C(\cdot) \)). Since \( q_J \) is endogenous, the observed product characteristics are correlated with the unobserved characteristic \( q_J \), the model cannot be identified without additional structure.

To that end suppose we follow Bajari and Benkard (2005) and normalize \( \theta_J \equiv 1 \) (i.e., there is no heterogeneity in how consumers value \( J^{th} \) characteristic) and assume that the unobserved product characteristic is missing at random, i.e., \( q_{-J} \| q_J \). Then we can see that \( q_J \) can be treated as an unobserved “noise,” which means the identification strategy that works when we observe all \( J \) characteristics extends to this case as long as we can identify the nonadditive random price function \( P(\cdot, \cdot) \). Then a sufficient condition for identification of \( P(\cdot, \cdot) \) is that there be a special bundle \( q^*_J \) such that \( P(q^*_J, q_J) = q_J \) for all \( q_J \).

We refer interested reader to Matzkin (2003) for a formal argument. Once the price function is identified, we can reapply the previous identification strategy.

5.4. Discussion. Research on “empirical mechanism design” as a way to study market outcomes is at its early stage and still a lot remains to be done. We conclude this section by discussing two limitations of this paper that requires further research. The first is the assumption that we observe continuous choices and the second is that there is a single seller. It appears that, in both cases, empirical research (in multidimensional adverse selection) is stymied by the lack of proper theoretical insights, and that it is likely that more theoretical progress will aid future empirical work in this area.

Discreteness of Product Characteristics. One of the fundamental assumptions is that we observe continuous and optimal product line as the theory predicts. Yet there are many instances where a seller offers only finite options even when it is suboptimal. This gap between the theory and the data are coherently conceptualized by (Wilson, 1993) and more recently Chu, Leslie, and Sorensen (2011) for nonlinear pricing. One obvious, but unsatisfactory, solution

\[\text{I want to thank Ali Hortaçsu and Ariel Pakes for discussion on this topic.}\]
would be to assume that consumer type \( \Theta \) is also discrete. In addition, very little is known about possible supply side constraints that can rationalize this behavior in our environment with multidimensional private information (unobserved heterogeneity). And without the supply side, we lose the identification of the cost function too.

A similar feature is also observed in multi-unit auctions (Kastl, 2011) where bidders submit substantially fewer steps than what would be profitable. To address this, Kastl (2011), and more recently Cassola, Hortaçsu, and Kastl (2013), posit that there is a fixed cost of submitting one more step that is unobserved to the researcher. Another avenue of research is to look for optimal nonlinear pricing that are also “robust” with respect to the distribution of the consumer heterogeneity. The idea is simple: if the seller cares about profits, but at the same time also allows for the possibility that she does not know the correct distribution of consumers type, then a robust nonlinear pricing might induce bunching everywhere. If these regions of bunching are indeed discrete then that would provide a better fit to the data. This consideration for robustness might even result in partial identification, but that would at least be within the framework of optimal supply side. Even though these are important extensions, they are beyond the scope of this paper and are left for future research.

**Multiple Sellers.** One of the key advantages of using mechanism design is that it allows for a flexible and multidimensional consumer heterogeneity. However, one disadvantage of this paper is that we focus on only a single seller environment and as such cannot address the problem of competing sellers. One of the major difficulties is that with competition among sellers the theoretical model becomes substantially difficult, even when there is only one dimensional consumer heterogeneity; see, for example, Epstein and Peters (1999); Martimort and Stole (2002) and Rochet and Stole (2002); and Bonatti (2011). This could partially explain why there are very few empirical papers (Ivaldi and Martimort, 1994; Aryal, 2016) that allow multiple sellers and use supply side optimality conditions. Any progress in identification and estimation will depend on the progress made in the theoretical literature. Perhaps a good starting point would be to allow \( \theta_j \) to be independent of \( \theta_{j'} \) for all pairs \( j, j' = 1, 2, \ldots, J \). We leave the identification of an oligopoly model with multidimensional unobserved preferences for future research.
In this paper we study the identification of the joint density of consumer’s multidimensional preferences using disaggregated choice data from a market with price discrimination. In particular the market is served by a single seller who sells products with multiple characteristics, or equivalently a multiple differentiated products (multiproduct monopoly). We show that if the utility is linear or bi-linear, then we can use the optimality of both the supply side and the demand side to nonparametrically identify the multidimensional unobserved consumer taste distribution and the cost function of the seller without resorting to strong functional form assumptions.

The key to identification is to exploit the equilibrium bijection between unobserved types and observed choices and the fact that in equilibrium, the consumer chooses a bundle that equates marginal utility to marginal prices. With multidimensional preference, however, the allocation rule need not be bijective for all types. For those consumer types who are bunched, the model can be identified under the assumption that there are sufficient consumer covariates that are orthogonal to unobserved preferences but affect the utility by interacting with the product characteristics.

When product characteristics enter utility in a nonlinear and unknown way, we need data from at least two periods that are separated by a binary and exogenous cost shifter for identification. The identification strategy is based on comparing multidimensional quantiles of choices across two cost regimes. If on top, researchers have information about consumer covariates then the model is over identified. We also characterize all testable restrictions of the model on the data. As far as we are aware, this is the first paper that provides tools to test optimality of principal-agent model with multidimensional private information. And finally we also extend the identification to consider measurement error and unobserved heterogeneity.

In summary, it is worth emphasizing that the next step for this line of research, before we can fully understand the nature of multidimensional unobserved heterogeneity, i.e., adverse selection, would be to study the identification with discrete menu and competitive sellers.
Appendix A. Multivariate Quantiles

In this section we introduce the definition of multivariate quantiles of (Koltchinskii (1997)). The objective is to give a short introduction to multivariate quantiles, and interested readers are referred to Koltchinskii (1997).\textsuperscript{20}

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 $L-$ integrable function 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

$$g_{L,t}(\mathbf{q}) := g_L(\mathbf{q}) - \langle \mathbf{q}, t \rangle,$$

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 $\partial g_L$ 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, $\partial g_L^{-1}$ is a single-valued map, and hence we get a unique quantile.\textsuperscript{21}

One can choose any kernel function $g$ as long as it satisfies the conditions mentioned above to define a multivariate quantile. Then, from Proposition 2.6 and Corollary 2.9 in Koltchinskii (1997), we know that $\partial g_L$ 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

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

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

\textsuperscript{20} Also see Chernozhukov, Galichon, Hallin, and Henry (2015) for an alternative approach.
\textsuperscript{21} For example, in a 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


