# RISE | RICE INITIATIVE *for the* STUDY *of* ECONOMICS

RISE Working Paper 14-003

# "Structural Analysis of Nonlinear Pricing"

by

Yao Luo, Isabelle Perrigne, Quang Vuong

RICE

Department of Economics
Baker Hall, MS22
6100 Main Street, Houston, Texas 77005
https://economics.rice.edu

# Structural Analysis of Nonlinear Pricing

Yao Luo

University of Toronto

Isabelle Perrigne

Rice University

Quang Vuong

New York University

July 2014

* This paper has been circulated under different titles. We thank Dan Ackerberg, Don Andrews, Gaurab Aryal, Steve Berry, Philippe Fevrier, Jin Hahn, Phil Haile, Ken Hendricks, Guy Laroque, Pascal Lavergne, Vadim Marmer, Rosa Matzkin, Ariel Pakes, Daniel Quint, John Riley, Jean-Marc Robin and Marc Rysman for constructive comments. We also benefited from participants' comments at the Stanford Institute for Theoretical Economics 2008, Cowles Summer Conference 2008, Econometric Society European Meeting 2008, Latin American Meeting of the Econometric Society 2008, IDEI Workshop on the Econometrics of Industrial Organization 2008, North American Summer Meeting of the Econometric Society 2012 and seminars at California Institute of Technology, Carnegie Mellon University, Massachusetts Institute of Technology, New York University, Paris School of Economics, Simon Fraser University, University of Arizona, University of California at Davis, University of California at Los Angeles, University College London, University of Fortaleza, University of Maryland, University of Montreal, University of Southern California and University of Wisconsin at Madison. The first author thanks the NET Institute (www.NETinst.org) for financial support while the last two authors gratefully acknowledge financial support from the National Science Foundation through grant SES 1148149.

Correspondence to: I. Perrigne, Department of Economics, Rice University, Houston, TX 77251-1892, e-mail address : iperrigne@gmail.com

**Abstract**

This paper proposes a methodology for analyzing nonlinear pricing data with an illustration on cellular phone. The model incorporates consumers exclusion. Assuming a known tariff, we establish identification of the model primitives using the first-order conditions of both the firm and the consumer up to a cost parameterization. Next, we propose a new one-step quantile-based nonparametric method to estimate the consumers inverse demand and their type distribution. We show that our nonparametric estimator is $\sqrt{N}$-consistent. We then introduce unobserved product heterogeneity with an unknown tariff. We show how our identification and estimation results extend. Our analysis of cellular phone consumption data assesses the performance of alternative pricing strategies relative to nonlinear pricing.

Keywords: Nonlinear Pricing, Nonparametric Identification, Empirical Processes, Quantile, Transformation Model, Unobserved Heterogeneity, Telecommunication.

# Structural Analysis of Nonlinear Pricing

Yao Luo, Isabelle Perrigne & Quang Vuong

## 1 Introduction

When facing heterogeneous consumers, a firm can discriminate consumers by offering different prices across purchase sizes and/or qualities. This practice is referred as nonlinear pricing or second degree price discrimination. Such a practice is common in electricity, cellular phone and advertising among others. See Wilson (1993) for examples. Seminal papers by Spence (1977), Mussa and Rosen (1978) and Maskin and Riley (1984) provide nonlinear pricing models within an imperfect information framework. The basic idea is to consider the consumer's taste/type as a parameter of adverse selection. The firm then designs an incentive compatible tariff discriminating consumers while endogeneizing the offered quantity/quality. This is achieved by giving up some rents to consumers. The resulting optimal price schedule is concave in quantity/quality implying discounts.[1]

The economic importance of price discrimination has led to an important empirical literature. Early empirical studies by Lott and Roberts (1991) and Shepard (1991) to name a few focus on exhibiting evidence of nonlinear pricing.[2] More recently, empirical studies evaluate the impact of nonlinear pricing on profits, consumer surplus and economic efficiency. Starting with Leslie (2004), a random utility discrete choice model for consumers' preferences is used to recover the consumers' taste distribution treating the price schedule as exogenous. See also McManus (2007), Cohen (2008) and Economides, Seim and Viard

[1]Extensions to oligopoly competition or several products include Oren, Smith and Wilson (1983), Ivaldi and Martimort (1994), Stole (1995), Armstrong (1996), Rochet and Chone (1998), Armstrong and Vickers (2001), Rochet and Stole (2003) and Stole (2007). Because of multidimensional screening, the optimal price schedule becomes less tractable though closed-form solutions might be obtained for some specifications.

[2]Other studies by Borenstein (1991), Borenstein and Rose (1994) and Busse and Rysman (2005) document the impact of competition on patterns of nonlinear pricing.

(2008). A third trend endogeneizes the optimal price and quantity schedules to estimate the demand and cost structure. See Ivaldi and Martimort (1994), Miravete (2002), Miravete and Roller (2004) and Crawford and Shum (2007).

In this paper, we propose a methodology for the structural analysis of nonlinear pricing data. We consider the Maskin and Riley (1984) model with consumer exclusion. This model contains all the relevant features to develop a structural setting for nonlinear pricing data. In this respect, the conclusion discusses how our results extend to more advanced pricing models with bundling and differentiated products. In the spirit of the recent literature in empirical industrial organization, we investigate the nonparametric identification of the model primitives from observables, which are the individual purchases and payments. See Laffont and Vuong (1996) and Athey and Haile (2007) for surveys on the nonparametric identification of auction models.

To simplify, we first consider the benchmark case in which the firm's tariff is known. We remark that our identification problem is reminiscent of Ekeland, Heckman and Neishem (2004) and Heckman, Matzkin and Neishem (2010) who study the nonparametric identification of hedonic price models. Instead of relying on instruments, we use the first-order conditions of both the consumer and the firm to identify the model primitives and exploit the one-to-one mapping between the unobserved consumer's type and his observed purchase. Given a parameterization of the cost, we also exploit the consumer exclusion condition to identify the cost parameters. Next, we propose a computationally convenient nonparametric procedure for estimating the marginal payoff and the type distribution that relies on empirical processes and quantile estimators. In contrast to previous papers on the estimation of incomplete information models, our estimator is one-step only. We establish its uniform consistency and show that it is $\sqrt{N}$-consistent. As a result, our nonparametric type density estimator converges at the parametric rate which is much faster rate than that of Guerre, Perrigne and Vuong (2000) in the context of auctions.

We then relax the assumption of a known tariff while introducing product unobserved heterogeneity. Such heterogeneity is important as observed payments and quantities do not necessarily satisfy a deterministic relationship. We discuss several options with a focus on their respective assumptions. The option that we retain leads to a transformation model. See Horowitz (1996). More precisely, we use the semiparametric model of Linton, Sperlich and Van Keilegom (2008). We then show how our identification results extend by establishing identification of the tariff function from observables. Regarding estimation, we study how our estimator extends while retaining its $\sqrt{N}$ consistency. In particular, we show how the estimation of the tariff parameters affects the asymptotic distributions of our estimators.

An illustration on cellular service data shows the importance of product unobserved heterogeneity and supports the nonlinear pricing model. Counterfactuals assess the performance of alternative pricing strategies such as two-part tariffs and quantity forcing relative to nonlinear pricing. A menu of two-part tariffs performs well at the cost of excluding more consumers. The conclusion briefly discusses some extensions of our methodology opening several avenues for future research. For instance, multiproduct firm, bundling and product differentiation can be entertained from our setting. Beyond nonlinear pricing, analysts can also use and extend our methodology to analyze contract data in retailing and labor to name a few where incomplete information plays a key role.

The paper is organized as follows. Section 2 introduces the model and establishes identification with a known tariff. Section 3 develops a one-step quantile-based nonparametric estimation procedure. Section 4 introduces product unobserved heterogeneity with an unknown tariff. It shows how the previous identification and estimation results extend. Section 5 presents an empirical application to cellular service data. Section 6 concludes with future lines of research. Three appendices collect proofs of statements in Sections 2, 3 and 4.

# 2 Model and Identification

## 2.1 The Model

Assumptions and Model Primitives

We consider the canonical model of nonlinear pricing by Maskin and Riley (1984). This model contains all the main components to analyze nonlinear pricing data. In the conclusion we will discuss further how we can incorporate bundling, multiple products and product differentiation. Consumers or agents are characterized by a scalar taste parameter $\theta$ distributed as $F(\cdot)$ with a continuous density $f(\cdot) > 0$ on $[\underline{\theta}, \overline{\theta}]$ with $0 \leq \underline{\theta} < \overline{\theta} < \infty$. This taste parameter is private information, i.e. it is unknown to the firm or principal. In general, not all agents consume as the firm excludes consumers with low tastes for the product. Specifically, the firm chooses optimally a threshold level $\theta^*$ above which consumers will buy its product.

Each consumer has a utility $U(Q; \theta) = \theta U_0(Q)$ and faces a tariff $T(Q)$ where $Q$ is the purchased quantity/quality.[3] Thus, the consumer's payoff is

$$\theta U_0(Q) - T(Q) \tag{1}$$

The firm incurs a cost $C(Q)$ for each consumer.[4] We make standard assumptions on the model primitives $[U_0(\cdot), F(\cdot), C(\cdot)]$. Hereafter, we use a variable as a subscript to indicate the derivative of a function with respect to this variable.

**Assumption A1:**
*(i) The base utility function $U_0(\cdot)$ is continuously differentiable on $[0, +\infty)$, and $\forall Q \geq 0$ $U_0(Q) \geq 0$, $U_{0Q}(Q) > 0$ and $U_{0QQ}(\cdot) < 0$,*
*(ii) The function $\theta - [(1 - F(\theta))/f(\theta)]$ is strictly increasing in $\theta \in [\underline{\theta}, \overline{\theta}]$,*
*(iii) The cost function $C(\cdot)$ is continuously differentiable on $[0, +\infty)$ with marginal cost satisfying $C_Q(Q) > 0 \quad \forall Q \geq 0$,*

THE OPTIMIZATION PROBLEM AND FIRST-ORDER CONDITIONS

The consumer chooses a quantity/quality $Q(\cdot)$ as a function of his type $\theta$. This quantity maximizes (1) and leads to

$$T_Q(Q(\theta)) = \theta U_{0Q}(Q(\theta)). \tag{2}$$

In words, his marginal utility equals the marginal tariff. In addition, consuming should provide a larger utility than not consuming, namely

$$\theta U_0(Q(\theta)) - T(Q(\theta)) \geq \theta U_0(0) \tag{3}$$

Despite the RHS depending on $\theta$, there are no countervailing incentives here because (3) is equivalent to $\theta[U_0(Q) - U(0)] - T(Q) \geq 0$, which is strictly increasing in $\theta$ by A1-(i) for any given $Q$.[5] Equation (2) and (3) are the so-called incentive compatibility (IC) and individual rationality (IR) constraints.

[3] The multiplicative separability in $\theta$ is common in the theoretical literature. It is also an identifying assumption as a general functional form $U(Q; \theta)$ is not identified. See also Section 2.2. An equivalent specification of utility is $\int_0^Q \theta v_0(x)dx$, where $v_0(\cdot)$ expresses the consumer willingness-to-pay for the $Q$th unit of product also called the inverse demand.

[4] Previous versions of the paper consider a general cost function $C(\cdot)$ for the total amount produced. See Riley (2012) and Section 2.2.

[5] See Lewis and Sappington (1989) and Maggi and Rodriguez-Clare (1995) for studies on countervailing incentives.

The firm's profit is

$$\int_{\theta^*}^{\overline{\theta}} \left[T(Q(\theta)) - C(Q(\theta))\right] f(\theta) d\theta.$$

The firm chooses optimally $\theta^*$, $Q(\cdot)$ and $T(\cdot)$ to maximize its profit subject to the agents' IR and IC constraints.[6] This gives

$$\max_{\theta^*, Q(\cdot), T(\cdot)} \int_{\theta^*}^{\overline{\theta}} \left[T(Q(\theta)) - C(Q(\theta))\right] f(\theta) d\theta, \tag{4}$$

subject to the IC and IR constraints. The next proposition establishes the necessary conditions for the solution $[\theta^*, Q(\cdot), T(\cdot)]$.

**Proposition 1:** *Under A1 and $Q_\theta(\cdot) > 0$, there exists a threshold value $\theta^* \in [\underline{\theta}, \overline{\theta}]$ below which consumers are not served. In addition, for $\theta \in [\theta^*, \overline{\theta}]$, the functions $Q(\cdot)$ and $T(\cdot)$ that solve the firm's optimization problem (4) satisfy*

$$\theta U_{0Q}(Q(\theta)) = C_Q(Q(\theta)) + \frac{1 - F(\theta)}{f(\theta)} U_{0Q}(Q(\theta)) \tag{5}$$

$$T_Q(Q(\theta)) = \theta U_{0Q}(Q(\theta)). \tag{6}$$

*If $\theta^* \in (\underline{\theta}, \overline{\theta})$, then $\theta^*$ solves the optimal exclusion condition*

$$\theta^* U_0(Q(\theta^*)) - C(Q(\theta^*)) - \frac{1 - F(\theta^*)}{f(\theta^*)} U_0(Q(\theta^*)) = 0. \tag{7}$$

*Moreover, the $\theta^*$-consumer gets zero net utility, i.e. the boundary condition $\theta^* U_0(Q(\theta^*)) = T(Q(\theta^*))$ holds.*

Conditions (5) and (6) characterize the optimal quantity schedule $Q(\cdot)$ and tariff $T(\cdot)$, respectively. Equation (5) says that the marginal payoff for each type equals the marginal cost plus a nonnegative distortion term due to incomplete information. Hence, all consumers buy less than the efficient (first-best) quantity/quality except for the $\overline{\theta}$ consumer for whom there is no distortion. Once $Q(\cdot)$ is determined, (6) characterizes the optimal price schedule $T(\cdot)$ using the boundary condition. Equation (7) expresses the trade-off between expanding the customer base and lowering the tariff, while the latter gives no rent to the $\theta^*$-agent.[7]

[6] If the firm can discriminate consumers based on some observed characteristics as in third degree price discrimination, such characteristics will show up in $F(\cdot)$ as conditioning variables and/or in $U_0(Q)$ as additional variables.

[7] We assume that $\theta^* \in (\underline{\theta}, \overline{\theta})$. If the LHS of (7) is always strictly positive, then there is no exclusion, in which case the boudary condition becomes $\underline{\theta} U_0(Q(\underline{\theta})) = T(Q(\underline{\theta}))$.

Under additional assumptions, we can show that $Q(\cdot)$ is strictly increasing and continuously dfferentiable on $[\theta^*, \overline{\theta}]$, while $T(\cdot)$ is strictly increasing and twice continuously differentiable on $[\underline{Q}, \overline{Q}] \equiv [Q(\theta^*), Q(\overline{\theta})]$. Regarding the verification of the second-order conditions, see Maskin and Riley (1984). Tirole (1988) indicates that the second-order derivative $T_{QQ}(\cdot) < 0$, i.e. the price schedule is strictly concave in $Q$. For formal proofs of Proposition 1, see Maskin and Riley (1984), Tirole (1988) or Riley (2012) among others.

## 2.2 Identification

We address the identification of the model when the tariff $T(\cdot)$ is known. In Section 4 we relax this assumption with the introduction of product unobserved heterogeneity.

General Setting and Discussion

Following Section 2.1, the model primitives are $[U_0(\cdot), F(\cdot), C(\cdot)]$, which are the consumer's base utility, his type distribution and the firm's cost function. We consider a situation where the analyst has information on the price schedule as well as data on consumers' purchased quantities. Thus, identification investigates whether the primitives can be uniquely recovered from the observables $[T(\cdot), G^{Q*}(\cdot)]$.[8]

A first-order condition similar to (6) arises in hedonic models whose nonparametric identification is studied by Ekeland, Heckman and Neishem (2004) and Heckman, Matzkin and Neishem (2010). Both papers show that the marginal utility is nonidentified without further restrictions. Ekeland, Heckman and Neishem (2004) establish identification of the consumer utility and the distribution of unobserved heterogeneity up to location and scale by exploiting variations in some consumers' continuous exogenous variables that are independent of the term of unobserved heterogeneity. As emphasized there, this result is obtained without the need to consider the firm's optimization problem. Heckman, Matzkin and Nesheim (2010)

[8]We use $G(\cdot)$ for the distribution of observables and use a superscript to indicate the random variable of interest. The superscript $*$ refers to a truncation as only consumers with a type $\theta \geq \theta^*$ consume the good. As is frequently the case, the analyst does not have information on consumers who choose the outside option. Otherwise, we could identify the proportion of such consumers $F(\theta^*)$ and hence $F(\cdot)$ instead of $F^*(\cdot)$ on $[\theta^*, \overline{\theta}]$ in view of Proposition 2 below. The data may also provide some exogenous agent's characteristics $Z$. Because second-degree price discrimination imposes the same tariff across agents, we can view $\theta$ as a scalar aggregation of the consumer's observed and unobserved heterogeneity. Also, we consider data from a single market. See Luo, Perrigne and Vuong (2013) for the introduction of market and consumers heterogeneity in a more general setting.

also consider exogenous variables for single and multimarket data.[9] Taking the logarithm of (6) gives $\log T_Q(Q) = \log U_{0Q}(Q) + \log\theta$. In a standard setup, $U_{0Q}(\cdot)$ would be identified by assuming (say) $\mathrm{E}[\log\theta|Q] = 0$. But from (5), $Q = Q(\theta)$ creating an endogeneity problem traditionally solved with instruments. It seems impossible to find instruments that are correlated with $Q$ but independent of $\theta$. Consequently, the consumer's first-order condition (6) is not sufficient to identify the utility function and the type distribution.

Our problem is also reminiscent of identification in models with incomplete information that lead to an equilibrium relationship between an observable and the agent's private information. For instance, in auctions Guerre, Perrigne and Vuong (2000) exploit this one-to-one mapping to recover the bidders' private value distribution. When the model contains more primitives to identify, several strategies can be entertained. When bidders are risk averse, Guerre, Perrigne and Vuong (2009) exploit exogenous variations in the number of bidders leading to some exclusion restrictions to identify the bidders' utility function. In the case of contract models, recent papers use exogenous variations and exclusion restrictions to identify the model primitives. Considering labor contracts, D'Haultfoeuille and Février (2011) exploit some exogenous variations in the offered contracts that do not affect the agents' primitives to identify the latter. Since contract models provide optimality conditions for both the principal and the agent, an alternative strategy is to exploit the entire set of first-order conditions to identify the model primitives as in Perrigne and Vuong (2011) for a procurement model with adverse selection and moral hazard. We follow this strategy and exploit both first-order conditions (6) and (7). In particular, our identification results do not require the existence of exogenous variables characterizing the agent and/or principal.

Identification of $[U_0(\cdot), F(\cdot), C(\cdot)]$

We first show that a scale normalization is necessary as both the type $\theta$ and the base utility $U_0(\cdot)$ are unknown. The next lemma formalizes this result. Let $\mathcal{S}$ be the set of structures $[U_0(\cdot), F(\cdot), C(\cdot)]$ satisfying A1.

**Lemma 1:** *Consider a structure $S = [U_0(\cdot), F(\cdot), C(\cdot)] \in \mathcal{S}$. Define another structure $\tilde{S} = [\tilde{U}_0(\cdot), \tilde{F}(\cdot), C(\cdot)]$, where $\tilde{U}_0(\cdot) = \frac{1}{\alpha}U_0(\cdot)$ and $\tilde{F}(\cdot) = F(\cdot/\alpha)$ for some $\alpha > 0$. Thus, $\tilde{S} \in \mathcal{S}$ and the two structures $S$ and $\tilde{S}$ lead to the same set of observables $[T(\cdot), G^{Q*}(\cdot)]$, i.e. the two structures are observationally equivalent.*

[9] Both papers consider a general utility function $U(Q, \theta)$ and show that it is not identified. The former considers separable additivity in $\theta$, while the latter considers alternative functional forms.

Several scale normalizations can be entertained. Three natural choices are to fix $\underline{\theta}$, $\theta^*$ or $\overline{\theta}$. We propose a convenient normalization after establishing the identification of $[U_0(\cdot), F(\cdot)]$.

Considering a model of adverse selection, D'Haultfoeuille and Février (2007) show that at least one of their three primitives, namely the surplus, the type distribution or the cost function, needs to be known to identify the model. We parameterize the cost function as formalized below. To simplify the expressions, we choose a constant marginal cost specification though alternative functional forms such as $C(Q) = \kappa(1+Q)^\gamma$ can be entertained.[10]

**Assumption B1:** *The cost function is of the form $C(Q) = \kappa + \gamma Q$ for $Q \geq 0$.*

The model primitives become $[U_0(\cdot), F(\cdot), \kappa, \gamma]$. Following B1, $C_Q(Q(\theta))$ becomes $\gamma$ in (5). Thus evaluating (5) and (6) at $\overline{\theta}$ and noting that $\overline{Q} = Q(\overline{\theta})$, $\gamma$ is identified by $\gamma = T_Q(\overline{Q})$.

We now turn to the identification of the marginal utility $U_{0Q}(\cdot)$ and the unobserved type distribution $F(\cdot)$. Our argument is based on quantiles. Let $\theta(\alpha)$ and $Q(\alpha)$ denote the $\alpha$-quantiles of the unobserved truncated type distribution $F^*(\cdot)$ and the observed truncated consumption distribution $G^{Q*}(\cdot)$. Hereafter, we define $\theta(0) = \theta^*$ and $Q(0) = \underline{Q}$. Thus $\theta(\cdot)$ and $Q(\cdot)$ are defined on $[0,1]$ because $\overline{\theta}$ and $\overline{Q}$ are finite. We first rewrite the first-order conditions (5) and (6) in terms of quantiles. This gives

$$\begin{aligned} \theta(\alpha)U_{0Q}(Q(\alpha)) &= \gamma + \frac{1-\alpha}{f^*(\theta(\alpha))}U_{0Q}(Q(\alpha)) \\ T_Q(Q(\alpha)) &= \theta(\alpha)U_{0Q}(Q(\alpha)) \end{aligned}$$

for $\alpha \in [0,1]$. Using the relationship between the density and its quantile function, i.e. $f^*(\theta(\alpha)) = 1/\theta_\alpha(\alpha)$ and $U_{0Q}(Q(\alpha)) = T_Q(Q(\alpha))/\theta(\alpha)$ in the first equation give

$$\frac{\theta_\alpha(\alpha)}{\theta(\alpha)} = \frac{T_Q(Q(\alpha)) - \gamma}{(1-\alpha)T_Q(Q(\alpha))}. \tag{8}$$

Integrating (8) from 0 to $\alpha$ gives

$$\log\frac{\theta(\alpha)}{\theta^*} = \int_0^\alpha \frac{1}{1-u}\left[1 - \frac{\gamma}{T_Q(Q(u))}\right]du \tag{9}$$

since $\theta(0) = \theta^*$. Because the RHS of (9) is known from the observables, it follows that $\theta(\cdot)$ is identified on $[0,1]$ up to $\theta^*$. This suggests the following natural normalization.

[10] Alternatively, we could consider the cost of the total amount produced, i.e. $C\left[\int_{\theta^*}^{\overline{\theta}} Q(\theta)f(\theta)d\theta\right]$. In this case, the marginal cost for the total amount produced is identified by combining (5) and (6) evaluated at the upper boundary $\overline{\theta}$. For the nonparametric identification of the cost function, see Luo, Perrigne and Vuong (2013) using multiple market data.

**Assumption B2:** *We normalize* $\theta^* = 1$.

Because $F^*(\cdot) = \theta^{-1}(\cdot)$, the type distribution is identified. Moreover, the marginal base utility is identified from (6) evaluated at $\theta = \theta(\alpha)$. Integrating the marginal base utility from $\underline{Q}$ to $Q$ identifies $U_0(\cdot)$ using the boundary condition $U_0(\underline{Q}) = T(\underline{Q})$. It remains to identify the cost parameter $\kappa$. We use the exclusion condition (7) for this purpose. The next proposition formalizes the identification of $[U_0(\cdot), F^*(\cdot), \kappa, \gamma]$.

**Proposition 2:** *Under assumptions A1, B1 and B2, the cost parameters are identified by*

$$\gamma = T_Q(\overline{Q}), \quad \kappa = \gamma \left( \frac{T(\underline{Q})}{T_Q(\underline{Q})} - \underline{Q} \right).$$

*The base utility* $U_0(\cdot)$ *is identified on* $[\underline{Q}, \overline{Q}]$ *as* $U_0(Q) = T(\underline{Q}) + \int_{\underline{Q}}^{Q} U_{0Q}(x)dx$. *The truncated consumers' type distribution* $F^*(\cdot)$ *is identified on* $[\theta^*, \overline{\theta}] = [1, \overline{\theta}]$.

We remark that $U_0(\cdot)$ and $F^*(\cdot)$ are not identified on $[0, \underline{Q})$ and $[\underline{\theta}, \theta^*)$, respectively. Intuitively, the purchase and price data do not provide any variation to identify these functions on those ranges as the minimum observed quantity is $\underline{Q}$.

# 3 Estimation

Our estimation method follows identification. Specifically it relies on (8) and (9). We remark that (8) and (9) involve the derivative of the tariff $T_Q(\cdot)$. As in Section 2.2, we consider that the tariff $T(\cdot)$ is known. Section 4 addresses the identification and estimation of $T(\cdot)$ with unobserved product heterogeneity. In contrast to the previous literature on the estimation of incomplete information models, e.g. Guerre, Perrigne and Vuong (2000), we develop one-step estimators for both the utility function $U_0(\cdot)$ and the type density $f^*(\cdot)$. In addition, our estimators achieve the parametric rate thereby circumventing the large data requirements associated with nonparametric estimators.

A One-Step Nonparametric Procedure

Our new estimator is based on quantiles.[11] Intuitively, (9) suggests that we can estimate $\theta(\cdot)$ at the parametric rate since the estimator of the quantile $Q(\cdot)$ is $\sqrt{N}$-consistent. Using (8), it follows that $f^*(\theta(\cdot)) = 1/\theta_\alpha(\cdot)$ and $U_{0Q}(Q(\cdot)) = T_Q(Q(\cdot))/\theta(\cdot)$ can also be estimated at the parametric rate. We now develop formally our estimators.

[11] Marmer and Shneyerov (2012) develop a two-step quantile-based estimator for auction models.

Let $N$ denote the number of consumers purchasing a quantity/quality $Q_i, i = 1, 2, ..., N$. Since (9) involves the cost parameter $\gamma$, we first study its estimation. We use the identifying relationship $\gamma = T_Q(\overline{Q})$. We propose a maximum estimator for $\overline{Q}$ leading to $Q_{\max} = \max_i Q_i$. This estimator has the main advantage to converge at a fast rate. Regarding the fixed cost parameter $\kappa$, we use the identifying equation of Proposition 2, which involves $\underline{Q}$. We use a minimum estimator leading to $Q_{\min} = \min_i Q_i$. This gives the following estimators

$$\hat{\gamma} = T_Q(Q_{\max}), \quad \hat{\kappa} = \hat{\gamma}\left(\frac{T(Q_{\min})}{T_Q(Q_{\min})} - Q_{\min}\right). \tag{10}$$

Next in view of (8) and (9), our estimators for $\theta(\cdot)$ and its derivative $\theta_\alpha(\cdot)$ are

$$\hat{\theta}(\alpha) = \exp\left\{\int_0^\alpha \frac{1}{1-u}\left[1 - \frac{\hat{\gamma}}{T_Q[\hat{Q}(u)]}\right]du\right\}, \quad \hat{\theta}_\alpha(\alpha) = \frac{\hat{\theta}(\alpha)}{1-\alpha}\frac{T_Q[\hat{Q}(\alpha)] - \hat{\gamma}}{T_Q[\hat{Q}(\alpha)]}, \tag{11}$$

for $\alpha \in [0,1]$, where $\hat{Q}(\cdot)$ is an estimator of the quantile function from the observed quantities. The relationships $f^*[\theta(\alpha)] \equiv [f^* \circ \theta](\alpha) = 1/\theta_\alpha(\alpha)$ and $U_{0Q}[Q(\alpha)] \equiv [U_{0Q} \circ Q](\alpha) = T_Q[Q(\alpha)]/\theta(\alpha)$ then provide estimators for the type density and the marginal base utility. This leads to the estimators of $f^*(\cdot)$ and $U_{0Q}(\cdot)$ at their $\theta$ and $Q$ quantiles, respectively

$$\widehat{f^*\circ}\theta(\alpha) = \frac{1}{\hat{\theta}_\alpha(\alpha)}, \quad \widehat{U_{0Q}\circ}Q(\alpha) = \frac{T_Q[\hat{Q}(\alpha)]}{\hat{\theta}(\alpha)}. \tag{12}$$

Hence, estimators at any value $\theta \in [1, \hat{\overline{\theta}}] = [1, \hat{\theta}(1)]$ and $Q \in [Q_{\min}, Q_{\max}]$ are

$$\hat{f}^*(\theta) = \widehat{f^*\circ}\theta[\hat{\theta}^{-1}(\theta)], \quad \hat{U}_{0Q}(Q) = \widehat{U_{0Q}\circ}\, Q[\hat{G}^{Q*}(Q)], \tag{13}$$

where $\hat{\theta}^{-1}(\cdot)$ is the estimated inverse of $\hat{\theta}(\cdot)$ and $\hat{G}^{Q*}(\cdot)$ is the empirical c.d.f. of $\{Q_i; i = 1, \ldots, N\}$. We note that $\theta^{-1}(\cdot) = F(\cdot)$, so that $\hat{\theta}^{-1}(\cdot)$ is our estimator $\hat{F}(\cdot)$ of $F(\cdot)$.

A standard estimator of the quantile function $Q(\cdot)$ is the inverse of the empirical c.d.f $\hat{G}_N^{Q*-1}(\cdot)$. In particular, $\hat{Q}(\cdot)$ is a left-continuous step function on $(0,1]$ with steps at $1/N < 2/N < \ldots < (N-1)/N$ with values equal to the ordered statistics $Q^1 \equiv Q_{\min} < Q^2 < \ldots < Q^N \equiv Q_{\max}$. At $\alpha = 0$, we define $\hat{Q}(0) = Q_{\min}$. This implies that the integral in (11) can be replaced by a finite sum of integrals leading to the computationally simple expression

$$\begin{aligned}
\log\hat{\theta}(\alpha) &= \sum_{j=1}^{J-1}\int_{(j-1)/N}^{j/N}\frac{1}{1-u}\left[1 - \frac{\hat{\gamma}}{T_Q(Q^j)}\right]du + \int_{(J-1)/N}^{\alpha}\frac{1}{1-u}\left[1 - \frac{\hat{\gamma}}{T_Q(Q^J)}\right]du \\
&= \sum_{j=1}^{J-1}\left[1 - \frac{\hat{\gamma}}{T_Q(Q^j)}\right]\log\left(\frac{N-j+1}{N-j}\right) + \left[1 - \frac{\hat{\gamma}}{T_Q(Q^J)}\right]\log\left(\frac{N-J+1}{N(1-\alpha)}\right) \quad (14)
\end{aligned}$$

for $(J-1)/N \leq \alpha \leq J/N$, where $J = 1, 2, \ldots, N$. We remark that $\hat{\theta}(\alpha)$ is continuous and increasing in $\alpha \in [0, (N-1)/N]$ since $T_Q[\hat{Q}(u)] < T_Q[Q_{\max}] = \hat{\gamma}$ by concavity of $T(\cdot)$ so that the integrand in (11) is strictly positive for $u \in [0, (N-1)/N]$.[12] Thus, the inverse $\hat{F}(\cdot) = \hat{\theta}^{-1}(\cdot)$ can be readily computed from (14) for any $\theta \in [1, \hat{\theta}[(N-1)/N])$.

Asymptotic Properties

We make the following assumption on the data generating process.

**Assumption C1:** *The unobserved types $\theta_i$, $i = 1, \ldots, N$ are i.i.d. distributed as $F^*(\cdot)$.*

Since $Q_i = Q(\theta_i)$, for $i = \ldots, N$, the observed quantities are also i.i.d.

The next lemma establishes the strong consistency of $\hat{\kappa}$ and $\hat{\gamma}$ with rates of convergence faster than $\sqrt{N}$. It also provides their asymptotic distributions. These results follow from the delta method combined with known properties of extreme order statistics from e.g. Galambos (1978). Let $\mathcal{E}(\lambda)$ denote the exponential distribution with parameter $\lambda$.

**Lemma 2:** *Under A1, B1 and C1, as $N \to \infty$, we have*

*(i) $\hat{\gamma} = \gamma + O_{a.s.}[(\log\log N)/N]$ and $\hat{\kappa} = \kappa + O_{a.s.}[(\log\log N)/N]$,*

*(ii) $N(\hat{\gamma} - \gamma) \xrightarrow{D} \mathcal{E}(\lambda)$ where $\lambda = -g^{Q*}(\overline{Q})/T_{QQ}(\overline{Q}) > 0$ and $N(\hat{\kappa} - \kappa) \xrightarrow{D} \mathcal{E}(\lambda_1) + \mathcal{E}(\lambda_2)$ where $\mathcal{E}(\lambda_1)$ and $\mathcal{E}(\lambda_2)$ are mutually independent with*

$$\lambda_1 = -\frac{g^{Q*}(\underline{Q}) T_Q^2(\underline{Q})}{T_Q(\overline{Q}) T(\underline{Q}) T_{QQ}(\underline{Q})} > 0, \quad \lambda_2 = -\frac{g^{Q*}(\overline{Q}) T_Q(\underline{Q})}{T_{QQ}(\overline{Q})(T(\underline{Q}) - \underline{Q} T_Q(\underline{Q}))} > 0.$$

We remark that $\lambda_2 = \lambda\gamma/\kappa$. We note that the sum of two independent exponentially distributed variables has marginal density $f(t) = [\lambda_1\lambda_2/(\lambda_1 - \lambda_2)][\exp(-\lambda_2 t) - \exp(-\lambda_1 t)]$. Following Campo, Guerre, Perrigne and Vuong (2011), we can estimate consistently $g^{Q*}(\overline{Q})$ and $g^{Q*}(\underline{Q})$ by one-sided kernel density estimators $\hat{g}^{Q*}(\cdot)$ evaluated at $Q_{\max}$ and $Q_{\min}$, respectively. As we can replace $\underline{Q}$ and $\overline{Q}$ by their estimates $Q_{\min}$ and $Q_{\max}$, we can use the asymptotic distributions in (ii) to construct confidence intervals for $\gamma$ and $\kappa$.

In view of (12) and (13), we need to study the properties of $\hat{\theta}(\cdot)$ and $\hat{\theta}_\alpha(\cdot)$. Following the empirical process literature introduced in econometrics by Andrews (1994), we view $\hat{\theta}(\cdot)$ and $\hat{\theta}_\alpha(\cdot)$ as random elements in the space $\ell^\infty[0, \alpha_\dagger]$ of bounded functions on $[0, \alpha_\dagger]$

[12] For $(N-1)/N < \alpha \leq 1$, the last term of (14) is equal to zero since $T_Q(Q^N) = T_Q(Q_{\max}) = \hat{\gamma}$. Thus, $\hat{\theta}(\alpha) = \hat{\theta}[(N-1)/N]$ for $\alpha \in [(N-1)/N, 1]$.

for any $\alpha_\dagger \in (0,1)$.[13] As usual, we equip $\ell^\infty[0,\alpha_\dagger]$ with its uniform metric $\|\psi_1 - \psi_2\|_\dagger = \sup_{\alpha\in[0,\alpha_\dagger]} |\psi_1(\alpha) - \psi_2(\alpha)|$. Weak convergence on the space $\ell^\infty[0,\alpha_\dagger]$ is denoted by "$\Rightarrow$."[14] The next lemma establishes the asymptotic properties of $\hat\theta(\cdot)$ and $\hat\theta_\alpha(\cdot)$.

**Lemma 3:** *Under A1, B1, B2 and C1, for any $\alpha_\dagger \in (0,1)$, as $N \to \infty$, we have*

*(i) $\|\hat\theta(\cdot) - \theta(\cdot)\|_\dagger \xrightarrow{a.s.} 0$ and $\|\hat\theta_\alpha(\cdot) - \theta_\alpha(\cdot)\|_\dagger \xrightarrow{a.s.} 0$,*

*(ii) as random functions in $\ell^\infty[0,\alpha_\dagger]$,*

$$\sqrt{N}[\hat\theta(\cdot) - \theta(\cdot)] \Rightarrow \gamma\theta(\cdot)\mathcal{Z}(\cdot),$$
$$\sqrt{N}[\hat\theta_\alpha(\cdot) - \theta_\alpha(\cdot)] \Rightarrow \gamma\theta_\alpha(\cdot)\left[\mathcal{Z}(\cdot) - \frac{T_{QQ}[Q(\cdot)]}{T_Q[Q(\cdot)](T_Q[Q(\cdot)] - \gamma)}\frac{\mathcal{B}_{G^{Q*}}[Q(\cdot)]}{g^{Q*}[Q(\cdot)]}\right],$$

*where $\mathcal{Z}(\cdot)$ is a tight Gaussian process defined on $[0,\alpha_\dagger]$ by*

$$\mathcal{Z}(\cdot) = -\int_{\underline{Q}}^{Q(\cdot)} \frac{T_{QQ}(q)}{T_Q^2(q)}\frac{\mathcal{B}_{G^{Q*}}(q)}{1 - G^{Q*}(q)}dq, \tag{15}$$

*with $\mathcal{B}_{G^{Q*}}(\cdot)$ denoting the $G^{Q*}$-Brownian bridge on $[\underline{Q},\overline{Q}]$.*[15]

In particular, $\sqrt{N}[\hat\theta(0) - \theta(0)] \xrightarrow{D} 0$ and $\sqrt{N}[\hat\theta_\alpha(0) - \theta_\alpha(0)] \xrightarrow{D} 0$ because $\mathcal{Z}(0) = 0$ and $\mathcal{B}_{G^{Q*}}[Q(0)] = 0$. This is expected since $\hat\theta(0) - \theta(0) = 1 - 1 = 0$ while $\hat\theta_\alpha(0) - \theta_\alpha(0) = \{[T_Q(Q_{\min}) - \hat\gamma]/T_Q(Q_{\min})\} - \{[T_Q(\underline{Q}) - \gamma]/T_Q(\underline{Q})\}$, which is $N$-consistent by Lemma 2-(ii).

Lemma 3 provides the uniform consistency and asymptotic distributions of the estimators (12) of $f^*(\cdot)$ and $U_{0Q}(\cdot)$ at their quantiles $\theta(\cdot)$ and $Q(\cdot)$, respectively. Namely, using $f^*[\theta(\cdot)] = 1/\theta_\alpha(\cdot)$ and $U_{0Q}[Q(\cdot)] = T_Q[Q(\cdot)]/\theta(\cdot)$ we obtain

$$\sqrt{N}[\widehat{f^*\circ\theta}(\cdot) - f^*\circ\theta(\cdot)] \Rightarrow -\gamma f^*[\theta(\cdot)]\left[\mathcal{Z}(\cdot) - \frac{T_{QQ}[Q(\cdot)]}{T_Q[Q(\cdot)](T_Q[Q(\cdot)] - \gamma)}\frac{\mathcal{B}_{G^{Q*}}[Q(\cdot)]}{g^{Q*}[Q(\cdot)]}\right] \tag{16}$$

$$\sqrt{N}[\widehat{U_{0Q}\circ Q}(\cdot) - U_{0Q}\circ Q(\cdot)] \Rightarrow -U_{0Q}[Q(\cdot)]\left[\gamma\mathcal{Z}(\cdot) + \frac{T_{QQ}[Q(\cdot)]}{T_Q[Q(\cdot)]}\frac{\mathcal{B}_{G^{Q*}}[Q(\cdot)]}{g^{Q*}[Q(\cdot)]}\right] \tag{17}$$

[13] As noted in footnote 12, the ratio $[T_Q(\hat Q(\alpha)) - \hat\gamma]/(1-\alpha)$ is zero for $\alpha \in [(N-1)/N, 1)$. In contrast, (8) and Proposition 2 show that, as $\alpha \to 1$, $[T_Q(Q(\alpha)) - \gamma]/(1-\alpha)$ converges to $\gamma/[\overline{\theta}f^*(\overline{\theta})]$, which is finite and positive. Thus we could improve our estimator around the upper boundary by imposing this restriction on $\hat Q(\cdot)$. To keep it simple, we choose instead to derive the asymptotic properties on $[0,\alpha_\dagger)$, where $\alpha_\dagger \in (0,1)$.

[14] Measurability issues are ignored hereafter. This can be addressed by considering either the projection $\sigma$-field on $\ell^\infty[0,\alpha_\dagger]$ as in Pollard (1984) or outer probabilities as in van der Vaart (1998). Alternatively, we may use another metric such as the Skorohod metric as in Billingsley (1968).

[15] The $G^{Q*}$-Brownian bridge on $[\underline{Q},\overline{Q}]$ is the limit of the empirical process $(1/\sqrt{N})\sum_i\{\mathbb{1}(Q_i \le \cdot) - G^{Q*}(\cdot)\}$ indexed by $[\underline{Q},\overline{Q}]$. See (say) van der Vaart (1998, p.266). It is a tight Gaussian process with mean 0 and covariance $G^{Q*}(Q)[1 - G^{Q*}(Q')]$, where $\underline{Q} \le Q \le Q' \le \overline{Q}$.

on $[0, \alpha_\dagger]$ using Property (P4) in Appendix B. In particular, both estimators are $\sqrt{N}$-consistent. The same remark as above applies to $\widehat{f^* \circ \theta}(0) = 1/\hat{\theta}_\alpha(0)$ and $\widehat{U_{0Q} \circ Q}(0) = T_Q(Q_{\min})$, which converge to $f^*(\theta^*) = 1/\theta_\alpha(0)$ and $U_{0Q}(\underline{Q}) = T_Q(\underline{Q})$ at rate $N$, respectively.

The next proposition gives the asymptotic properties of $\hat{f}^*(\cdot)$ and $\hat{U}_{0Q}(\cdot)$ on $[\theta^*, \theta_\dagger]$ and $[\underline{Q}, Q_\dagger]$, respectively, where $\theta_\dagger \in (\theta^*, \overline{\theta})$ and $Q_\dagger \in (\underline{Q}, \overline{Q})$. For instance, $\theta_\dagger = \theta(\alpha_\dagger)$ and $Q_\dagger = Q(\alpha_\dagger)$ with $0 < \alpha_\dagger < 1$. Let $\ell^\infty[\theta^*, \theta_\dagger]$ and $\ell^\infty[\underline{Q}, Q_\dagger]$ denote the space of bounded functions on $[\theta^*, \theta_\dagger]$ and $[\underline{Q}, Q_\dagger]$ equipped with their uniform metric $\|\cdot\|_\dagger$, respectively.

**Proposition 3:** *Under A1, B1, B2 and C1, for any $\theta_\dagger \in (\theta^*, \overline{\theta})$ and $Q_\dagger \in (\underline{Q}, \overline{Q})$, as $N \to \infty$, we have*

*(i) $\|\hat{f}^*(\cdot) - f^*(\cdot)\|_\dagger \xrightarrow{a.s} 0$ and $\|\hat{U}_{0Q}(\cdot) - U_{0Q}(\cdot)\|_\dagger \xrightarrow{a.s} 0$,*

*(ii) as random functions in $\ell^\infty[\theta^*, \theta_\dagger]$ and $\ell^\infty[\underline{Q}, Q_\dagger]$,*

$$\sqrt{N}[\hat{f}^*(\theta) - f^*(\theta)] \Rightarrow -\gamma[f^*(\theta) + \theta f^*_\theta(\theta)]\mathcal{Z}[F^*(\theta)] + \frac{H_\theta(\theta)}{H(\theta)}\mathcal{B}_{F^*}(\theta)$$

$$\sqrt{N}[\hat{U}_{0Q}(Q) - U_{0Q}(Q)] \Rightarrow -U_{0Q}(Q)\left[\gamma\mathcal{Z}[G^{Q*}(Q)] + \frac{T_Q(Q) - \gamma}{T_Q(Q)}\frac{\mathcal{B}_{G^{Q*}}(Q)}{1 - G^{Q*}(Q)}\right],$$

*where $H(\theta) = [1 - F^*(\theta)]/[\theta f^*(\theta)]$, $\mathcal{B}_{F^*}(\theta) = \mathcal{B}_{G^{Q*}}[Q(\theta)]$ and $\mathcal{Z}(\cdot)$ is defined in (15).*

The first part establishes the uniform almost sure convergence of $\hat{f}^*(\cdot)$ and $\hat{U}_{0Q}(\cdot)$ on any subsets $[\theta^*, \theta_\dagger] \subset [\theta^*, \overline{\theta})$ and $[\underline{Q}, Q_\dagger] \subset [\underline{Q}, \overline{Q})$, respectively. The second part gives the asymptotic distributions of these estimators. It is worthnoting that their rates of convergence are the parametric rate $\sqrt{N}$.[16] To our knowledge, this contrasts with the previous literature on estimation of incomplete information models such as auctions in which the optimal convergence rates for the model primitives are slower. See e.g. Guerre, Perrigne and Vuong (2000) for the bidders' private value density. Achieving the parametric rate is useful in practice as it allows to analyze medium size data sets while avoiding the curse of dimensionality typically associated with nonparametric estimators.

We remark that $\mathcal{Z}[G^{Q*}(Q)]$ is given by the right-hand side of (15) with $Q(\cdot)$ replaced by $Q$ since $Q[G^{Q*}(Q)] = Q$, while $\mathcal{Z}[F^*(\theta)]$ is given by

$$\mathcal{Z}[F^*(\theta)] = -\frac{1}{\gamma}\int_{\theta^*}^{\theta} H_\theta(x)\frac{\mathcal{B}_{F^*}(x)}{1 - F^*(x)}dx. \tag{18}$$

[16] As for Lemma 3, at the lower boundary, $\hat{f}^*(\theta^*)$ and $\hat{U}_{0Q}(\underline{Q})$ converge at a faster rate, namely at rate $N$, to their limits $f^*(\theta^*)$ and $U_{0Q}(\underline{Q}) = T_Q(\underline{Q})$, respectively.

The proof of the latter is given in Appendix B. Using Proposition 3-(ii), we can derive the asymptotic distribution of $\hat{U}_0(Q) = T(Q_{\min}) + \int_{Q_{\min}}^{Q} \hat{U}_{0Q}(q)dq$. Because $Q_{\min}$ is $N$-consistent, we obtain

$$\sqrt{N}[\hat{U}_0(Q) - U_0(Q)] \Rightarrow -\int_{\underline{Q}}^{Q} U_{0Q}(q)\left[\gamma\mathcal{Z}[G^{Q*}(q)] + \frac{T_Q(q)-\gamma}{T_Q(q)}\frac{\mathcal{B}_{G^{Q*}}(q)}{1-G^{Q*}(q)}\right]dq$$

uniformly in $Q \in [\underline{Q}, Q_\dagger]$.

The previous limits are tight Gaussian processes with zero means and finite covariance functions. In practice, we can use such asymptotic distributions to conduct large sample hypotheses tests and construct pointwise or uniform confidence intervals provided the asymptotic variances can be estimated consistently and uniformly. Since we are mainly interested in the two primitives $f^*(\cdot)$ and $U_{0Q}(\cdot)$, we focus on them. Specifically, using Proposition 3-(ii), we have

$$\sqrt{N}[\hat{f}^*(\theta) - f^*(\theta)] \xrightarrow{D} \mathcal{N}\left(0, V_{f^*}(\theta)\right), \quad \sqrt{N}[\hat{U}_{0Q}(Q) - U_{0Q}(Q)] \xrightarrow{D} \mathcal{N}\left(0, V_{U_{0Q}}(Q)\right)$$

for $\theta \in [\theta^*, \theta_\dagger]$ and $Q \in [\underline{Q}, Q_\dagger]$, respectively. Appendix B provides detailed computations of $V_{f^*}(\theta)$ and $V_{U_{0Q}}(Q)$. It also discusses their consistent estimation.

Up to now, we have considered the situation where the analyst knows the payment schedule and has data on purchased quantities. Equivalently, we can consider the case where the analyst knows the payment schedule and has data on payments. In particular, our estimators can be written in terms of the payments $\{t_i, i = 1, \ldots, N\}$. Since $Q = T^{-1}(t)$ and $T_Q(Q) = 1/T_t^{-1}(t)$, where $T^{-1}(\cdot)$ is strictly increasing, it is easy to see that (10)–(13) become

$$\hat{\gamma} = \frac{1}{T_t^{-1}(t_{\max})}, \qquad \hat{\kappa} = \hat{\gamma}\left(t_{\min}T_t^{-1}(t_{\min}) - T^{-1}(t_{\min})\right) \quad (19)$$

$$\hat{\theta}(\alpha) = \exp\left\{\int_0^\alpha \frac{1}{1-u}\left[1 - \hat{\gamma}T_t^{-1}(\hat{t}(u))\right]\right\}du, \qquad \hat{\theta}_\alpha(\alpha) = \frac{\hat{\theta}(\alpha)}{1-\alpha}\left[1 - \hat{\gamma}T_t^{-1}(\hat{t}(\alpha))\right] \quad (20)$$

$$\widehat{f^*\circ\theta}(\alpha) = \frac{1}{\hat{\theta}_\alpha(\alpha)}, \qquad \widehat{U_{0Q}\circ Q}(\alpha) = \frac{1}{\hat{\theta}(\alpha)T_t^{-1}[\hat{t}(\alpha)]} \quad (21)$$

$$\hat{f}^*(\theta) = \widehat{f^*\circ\theta}[\hat{\theta}^{-1}(\theta)], \qquad \hat{U}_{0Q}(Q) = \widehat{U_{0Q}\circ Q}[\hat{G}^{t*}(T(\cdot))], \quad (22)$$

where $\hat{t}(\cdot) = \hat{G}^{t*-1}(\cdot)$ is the standard estimator of the quantile function $t(\cdot)$ of $G^{t*}(\cdot)$ with $\hat{t}(0) \equiv \underline{t}$, and $\hat{G}^{t*}(\cdot)$ is the empirical cdf of payments. Lemmas 2 and 3 as well as Proposition 3 still hold, while the asymptotic variances can be writen directly in terms of $G^{t*}(\cdot)$. For instance, because $t = T(Q)$, the Gaussian process $\mathcal{Z}(\cdot)$ on $[0, \alpha_\dagger]$ can be written as

$$\mathcal{Z}(\cdot) = \int_{\underline{t}}^{t(\cdot)} T_{tt}^{-1}(t)\frac{\mathcal{B}_{G^{t*}}(t)}{1-G^{t*}(t)}dt \quad (23)$$

from the change-of-variable $t = T(q)$ in (15) so that $T_Q(q) = 1/T_t^{-1}(t)$, $T_{QQ}(q) = -T_{tt}^{-1}(t)/T_t^{-1}(t)^3$, and $G^{Q*}(q) = G^{t*}(t)$.

# 4 Unobserved Heterogeneity

In this section, we consider the case when the analyst does not know the payment schedule $T(\cdot)$. Instead, he/she observes the pairs of payments and quantities $(t_i, q_i), i = 1, \ldots, N$. The nonlinear pricing model of Section 2 implies that the observed payments and quantities lie on the curve $t = T(q)$ as the payment $t_i$ and quantity $q_i$ depend on the consumer type or unobserved heterogeneity $\theta_i$, which is the only unobserved random term in the econometric model of Section 3. In practice, the observed prices and quantities may not lie on a curve thereby calling for an additional source of randomness. Several reasons can be invoked to rationalize this addition. For instance, the analyst may not observe perfectly the payment and/or consumption. In addition, the product may be horizontally and/or vertically differentiated in more than a single dimension. See Luo, Perrigne and Vuong (2012, 2013) when the attributes are observed. This second random term is denoted $\epsilon$.

Discussion

We discuss three options for introducing $\epsilon$ in the econometric model. A first option is to consider a measurement error on the payment leading to $t_i = T(Q_i) + \epsilon_i$ with $q_i = Q_i$. In Perrigne and Vuong (2011), $\epsilon_i$ is interpreted as a deviation from optimal payments in a procurement model where $\epsilon_i$ could incorporate corruption, side payments or political capture. With the normalization $E[\epsilon_i|Q_i] = 0$, the tariff function $T(\cdot)$ is nonparametrically identified.

Alternatively, $\epsilon$ can be viewed as representing product unobserved heterogeneity. In the empirical application of Section 5, data from a cellular phone company include the quantity $q_i$ of phone calls measured in minutes. As a matter of fact, the bill contains additional services such as roaming, voice mail services, phone rings, etc. More generally, data may not always contain detailed information on all the components or product attributes that make up the payment. We then assume that the analyst observes the quantity $q_i$ while the nonlinear pricing mechanism is based on $Q_i$ which is a function of $q_i$ and $\epsilon_i$, i.e. $Q_i = m(q_i, \epsilon_i)$. Hereafter, $Q_i$ is called the contracted quantity. The tariff equation then becomes $t_i = T(Q_i) = T[m(q_i, \epsilon_i)]$, where $T(\cdot)$ is strictly increasing and concave. This equation can also been written using the inverse of the tariff leading to $T^{-1}(t_i) = m(q_i, \epsilon_i)$. This general model is not identified. We discuss below some special cases.

In the second option, we view $\epsilon_i$ as a measurement error on the contracted quantity $Q_i$, i.e. $q_i = Q_i \epsilon_i$ with $\epsilon_i$ independent of $Q_i$. Thus, $m(q_i, \epsilon_i) = q_i/\epsilon_i$. The model of Section 2 remains with $Q_i = q_i/\epsilon_i$. Intuitively, product unobserved heterogeneity acts as a quantity multiplier. More precisely, by taking the logarithm, we have $\log q_i = \log Q_i + \log \epsilon_i$. The independence of $Q_i$ and $\epsilon_i$ is equivalent to assuming $\theta_i$ independent of $\epsilon_i$ since $Q_i = Q(\theta_i)$ from the model of Section 2. The equation identifying the tariff is $\log q_i = \log T^{-1}(t_i) + \log \epsilon_i$, where $t_i = T(Q_i)$ is independent of $\epsilon_i$. The tariff is then identified under the normalization $\mathrm{E}[\log \epsilon] = 0$.

A third option assumes independence of the two components of $Q_i$, namely $q_i$ and $\epsilon_i$ leading to $T^{-1}(t_i) = \tilde{m}(q_i)\epsilon_i$, where $m(q_i, \epsilon_i) \equiv \tilde{m}(q_i)\epsilon_i$ since the general specification $m(q_i, \epsilon_i)$ is not identified. Taking the logarithm gives the transformation model $\log T^{-1}(t_i) = \log \tilde{m}(q_i) + \log \epsilon_i$. See Carroll and Ruppert (1988) for parametric transformation models. Under some location and scale normalizations, Horowitz (1996) establishes the nonparametric identification of $T(\cdot)$ with $\tilde{m}(\cdot)$ parametric. Relying on this result, Ekeland, Heckman and Neishem (2004) show the nonparametric identification of $T(\cdot)$ and $\tilde{m}(\cdot)$ while Chiappori, Komunjer and Kristensen (2013) obtain a similar result under a weaker independence assumption. The model of Section 2 remains with $Q_i = \tilde{m}(q_i)\epsilon_i$. Because $Q_i = Q(\theta_i)$, the third option allows for dependence between $\theta_i$ and $\epsilon_i$, while the first two options imply at the minimum that $\theta_i$ and $\epsilon_i$ are uncorrelated.

Regarding estimation, we can use any nonparametric regression estimator for $T(\cdot)$ in the first option. In particular, we can choose sieve estimators to impose shape restrictions such as monotonicity and concavity of the tariff. As is well known, the resulting estimator converges at a rate slower than $\sqrt{N}$. This also applies to the estimation of $T^{-1}(\cdot)$ in the second option. In contrast, the third option has the advantage to lead to a $\sqrt{N}$-consistent estimator of $T^{-1}(\cdot)$ and hence of $T(\cdot)$ as shown by Chiappori, Komunjer and Kristensen (2013), while the estimator of $\tilde{m}(\cdot)$ inherits the usual nonparametric rate. Following Horowitz (1996), estimation of $T^{-1}(\cdot)$ involves integrating kernel estimators. Alternatively, estimation of $T^{-1}(\cdot)$ can be based on Chen (2002) rank estimator.

In Section 3, however, estimation of the model primitives involves the derivative $T_Q(\cdot)$ of the tariff. The $\sqrt{N}$-consistency property is lost with any of the above nonparametric estimators since $T_Q(\cdot)$ is estimated at a slower rate.[17] As we want to maintain the parametric rate for estimating the model primitives, we consider a parameterization of the tariff function, more precisely of its inverse $T^{-1}(\cdot) = T^{-1}(\cdot;\beta)$ for $\beta \in I\!R^{\dim\beta}$. This leads to the semiparametric transformation model studied by Linton, Sperlich and Van Keilegom (2008), i.e. $\log T^{-1}(t_i;\beta) = \log \tilde{m}(q_i) + \log \epsilon_i$. This specification has at least three appealing features. First, it provides a $\sqrt{N}$-consistent estimator of the derivative $T_Q(\cdot)$ and hence $\sqrt{N}$-consistent estimators of the model primitives. In addition, this allows us to assess the effect of estimating $T(\cdot)$ through $\beta$ on the asymptotic distributions of the latter. Second, we can readily impose monotonicity and convexity on $T^{-1}(\cdot)$, while maintaining flexibility on how the observed quantity $q_i$ affects the contracted quantity $Q_i$ as $Q_i = \tilde{m}(q_i)\epsilon_i$. Third, we can test the chosen parametric specification by comparing the nonparametric and parametric estimates of $T^{-1}(\cdot)$. Though we do not pursue this issue here, an appealing feature is that a Cramer-von Mises-type test will be $\sqrt{N}$-consistent. In view of this discussion, hereafter we retain the semiparametric transformation model to introduce product unobserved heterogeneity.

Estimation

Let $F^\epsilon(\cdot)$ be the distribution of $\log\epsilon$ with support $[\log\underline{\epsilon}, \log\overline{\epsilon}] \subseteq I\!R$ and density $f^\epsilon(\cdot) > 0$. Our model for product unobserved heterogeneity is

$$\log T^{-1}(t;\beta) = \log \tilde{m}(q) + \log \epsilon, \tag{24}$$

where $T^{-1}(\cdot)$ is strictly increasing in $t$, $\beta \in I\!R^{\dim\beta}$, and $\epsilon$ is independent of $q$ with $\mathrm{E}(\log\epsilon) = 0$.[18] Identification of $[\beta, \tilde{m}(\cdot), F^\epsilon(\cdot)]$ is ensured by the nonparametric identification of $[T^{-1}(\cdot), \tilde{m}(\cdot), F^\epsilon(\cdot)]$. Hereafter, we assume that the observations $(t_i, q_i)$, $i = 1, \ldots, N$ are i.i.d. Thus, $(q_i, \epsilon_i), i = 1, \ldots, N$ are also i.i.d. thereby implying that $Q_i = \tilde{m}(q_i)\epsilon_i, i = 1, \ldots, N$ are i.i.d. which is consistent with C1.[19]

[17] In addition to loosing $\sqrt{N}$-consistency, we foresee additional problems in the implementation of Horowitz (1996) and Chen (2002) estimators. Identification of $T^{-1}(\cdot)$ requires a location normalization. In our case, a natural normalization is $\underline{\epsilon} = 1$. With the transformation model, such a normalization leads to estimate the lower envelope of the scatter plot of the observations $(t_i, q_i)$. As such, the estimates of the tariff function would be sensitive to outliers.

[18] Possible extensions include heteroscedasticity in which the variance of $\epsilon$ depends on $q$ . See e.g. Zhou, Lin and Johnson (2008) and Khan, Shin and Tamer (2011) in a semiparametric settting and Chiappori, Komunjer and Kristensen (2013) in a nonparametric one.

[19] The DGP may be subject to some restrictions as the distributions of the unobserved $\epsilon$ and the observed quantities $q$ as well as the function $\tilde{m}(\cdot)$ need to lead to a distribution $G^{Q*}(\cdot)$ of contracted quantity $Q$ that is

Let $\hat{\beta}$ be any $\sqrt{N}$-asymptotically normal estimator of $\beta$, i.e. $\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{D} \mathcal{N} \sim \mathcal{N}(0,\Omega)$. The empirical study of Section 5 uses Linton, Sperlich and Van Keilegom (2008) minimum distance estimator from independence. Once $T^{-1}(\cdot)$ has been estimated by $\hat{T}^{-1}(\cdot) \equiv T^{-1}(\cdot;\hat{\beta})$, we can use (19)–(21), which use directly the payments $\{t_i; i=1,\ldots,N\}$ instead of using $\hat{Q}_i \equiv \hat{T}^{-1}(t_i)$ in (10)–(13). This minimizes computation errors. Let $\tilde{\gamma}$, $\tilde{\kappa}$, $\tilde{\theta}(\cdot)$, $\tilde{\theta}_\alpha(\cdot)$, $\widetilde{f^*\circ\theta}(\cdot)$, $\widetilde{U_{0Q}\circ Q}(\cdot)$, $\tilde{f}^*(\cdot)$ and $\tilde{U}_{0Q}(\cdot)$ denote our estimators. We have

$$\tilde{\gamma} = \frac{1}{\hat{T}_t^{-1}(t_{\max})}, \qquad \tilde{\kappa} = \tilde{\gamma}\left(t_{\min}\hat{T}_t^{-1}(t_{\min}) - \hat{T}^{-1}(t_{\min})\right) \quad (25)$$

$$\tilde{\theta}(\alpha) = \exp\left\{\int_0^\alpha \frac{1}{1-u}\left[1-\tilde{\gamma}\hat{T}_t^{-1}(\hat{t}(u))\right]\right\}du, \qquad \tilde{\theta}_\alpha(\alpha) = \frac{\tilde{\theta}(\alpha)}{1-\alpha}\left[1-\hat{\gamma}\hat{T}_t^{-1}(\hat{t}(\alpha))\right] \quad (26)$$

$$\widetilde{f^*\circ\theta}(\alpha) = \frac{1}{\tilde{\theta}_\alpha(\alpha)}, \qquad \widetilde{U_{0Q}\circ Q}(\alpha) = \frac{1}{\tilde{\theta}(\alpha)\hat{T}_t^{-1}[\hat{t}(\alpha)]} \quad (27)$$

$$\tilde{f}^*(\theta) = \widetilde{f^*\circ\theta}[\tilde{\theta}^{-1}(\theta)], \qquad \tilde{U}_{0Q}(Q) = \widetilde{U_{0Q}\circ Q}[\hat{G}^{t*}(\hat{T}(\cdot))], \quad (28)$$

where $\hat{T}_t^{-1}(\cdot) \equiv T_t^{-1}(\cdot;\hat{\beta})$, $\hat{G}^{t*}(\cdot)$ is the empirical cdf of payments, and $\hat{t}(\cdot)$ is the usual quantile estimator for payments with $\hat{t}(0) \equiv \underline{t}$. It is worthnoting that our estimators are straightforward to compute given $\hat{\beta}$. In particular, letting $t^1 \equiv t_{\min} < t^2 < \ldots < t^N \equiv t_{\max}$ be the payment ordered statistics, we have

$$\log\tilde{\theta}(\alpha) = \sum_{j=1}^{J-1}\left[1-\hat{\gamma}\hat{T}_t^{-1}(t^j)\right]\log\left(\frac{N-j+1}{N-j}\right) + \left[1-\hat{\gamma}\hat{T}_t^{-1}(t^J)\right]\log\left(\frac{N-J+1}{N(1-\alpha)}\right)$$

for $(J-1)/N \le \alpha \le J/N$, where $J=1,2,\ldots,N$. Moreover, $\tilde{\theta}(\cdot)$ is continuous and increasing in $\alpha \in [0,(N-1)/N]$ since $\hat{T}_t^{-1}[t(u)] < \hat{T}_t^{-1}[t_{\max}] = \hat{\gamma}$ by convexity of $\hat{T}^{-1}(\cdot)$. On the other hand, $\tilde{\theta}(\alpha) = \tilde{\theta}[(N-1)/N]$ for $\alpha \in [(N-1)/N, 1]$.

Asymptotic Properties

We note that $g^{T*}(\cdot) > 0$ on its support $[\underline{t},\overline{t}]$ with $0 < \underline{t} < \overline{t} < \infty$ since $t = T[Q(\theta)]$, $T_Q(\cdot) > 0$ on $[\underline{Q},\overline{Q}]$, $Q_\theta(\cdot) > 0$ on $[\theta^*,\theta]$, and $f^*(\cdot) > 0$ on its support $[\theta^*,\overline{\theta}]$. Thus, because $T^{-1}(\cdot;\hat{\beta})$ and $T_t^{-1}(\cdot;\hat{\beta})$ converge uniformly on $[\underline{t},\overline{t}]$ to $T^{-1}(\cdot) = T^{-1}(\cdot;\beta)$ and $T_t^{-1}(\cdot) = T_t^{-1}(\cdot;\beta)$, respectively, consistency or uniform consistency of our estimators (25)–(28) is easily established. Hereafter, we focus on their asymptotic distributions. In particular,

rationalized by the economic model $[F(\cdot), U_0(\cdot), \gamma, \kappa]$. Specifically, given a type distribution $F(\cdot)$, we obtain $G^{Q*}(\cdot)$ through the equilibrium mapping $Q_i = Q(\theta_i)$ for $\theta_i \ge \theta^*$. Thus, in view of the independence of $q$ and $\epsilon$, the distribution of $\log\tilde{m}(q_i)$ is obtained by Fourier-inverting the ratio of the characteristic functions of $\log Q$ and $\log\epsilon$, provided such a ratio is a proper characteristic function.

our derivation underscores the effects of estimating $T(\cdot)$, more precisely of estimating the parameters $\beta$ of $T^{-1}(\cdot;\beta)$.

Following Section 3, we begin with the estimators of the cost parameters $\gamma$ and $\kappa$. The next lemma is useful for deriving their asymptotic distributions.

**Lemma 4:** *Under A1, B1 and C1, as $N \to \infty$ we have*
*(i)* $\sqrt{N}(\tilde{\gamma}-\gamma) = -\gamma^2 T_{t\beta}^{-1}(\bar{t};\beta)\sqrt{N}(\hat{\beta}-\beta) + o_P(1)$,
*(ii)* $\sqrt{N}(\tilde{\kappa}-\kappa) = \gamma\left(\underline{t}T_{t\beta}^{-1}(\underline{t};\beta) - T_{\beta}^{-1}(\underline{t};\beta) - \kappa T_{t\beta}^{-1}(\bar{t};\beta)\right)\sqrt{N}(\hat{\beta}-\beta) + o_P(1)$.

Since $\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{D} \mathcal{N} \sim \mathcal{N}(0,\Omega)$, Lemma 4 implies that $\tilde{\gamma}$ and $\tilde{\kappa}$ are asymptotically normal. In contrast to Lemma 2, their convergence rate is now $\sqrt{N}$ instead of $N$. This is because $\hat{\beta}$ converges at a slower rate than the infeasible estimators $\hat{\gamma}$ and $\hat{\kappa}$. As usual, consistent estimation of the asymptotic variances of $\sqrt{N}(\tilde{\gamma}-\gamma)$ and $\sqrt{N}(\tilde{\kappa}-\kappa)$ are obtained by replacing $\gamma$, $\underline{t}$, $\bar{t}$, and $\beta$ by their consistent estimators $\tilde{\gamma}$, $t_{\min}$, $t_{\max}$, and $\hat{\beta}$ upon consistent estimation of the asymptotic variance $\Omega$ of $\sqrt{N}(\hat{\beta}-\beta)$.

Next, we turn to $\tilde{f}^*(\cdot)$ and $\tilde{U}_{0Q}(\cdot)$. Following Lemma 3, we first study the asymptotic properties of the estimators of the quantile $\theta(\cdot)$ and its derivative $\theta_\alpha(\cdot)$.

**Lemma 5:** *Under A1, B1, B2 and C1, for any $\alpha_\dagger \in (0,1)$, as $N \to \infty$ we have*

$$\begin{aligned}\sqrt{N}[\tilde{\theta}(\cdot)-\theta(\cdot)] &= \sqrt{N}[\hat{\theta}(\cdot)-\theta(\cdot)] + \gamma\theta(\cdot)I(\cdot)\sqrt{N}(\hat{\beta}-\beta) + o_P(1),\\ \sqrt{N}[\tilde{\theta}_\alpha(\cdot)-\theta_\alpha(\cdot)] &= \sqrt{N}[\hat{\theta}_\alpha(\cdot)-\theta_\alpha(\cdot)]\\ &\quad +\gamma\theta_\alpha(\cdot)\left(I(\cdot)-\frac{1}{H[\theta(\cdot)]}a(\cdot)\right)\sqrt{N}(\hat{\beta}-\beta)+o_P(1),\end{aligned}$$

*uniformly on $[0,\alpha_\dagger]$, where*

$$I(\cdot) = -\int_0^{\cdot}\frac{1}{1-u}a(u)du, \quad a(\cdot) = T_{t\beta}^{-1}[t(\cdot);\beta] - \gamma T_t^{-1}[t(\cdot);\beta]T_{t\beta}^{-1}(\bar{t};\beta) \tag{29}$$

*are nonstochastic $(1\times \dim\beta)$ vector functions both defined on $[0,\alpha_\dagger]$.*

Lemma 5 shows the effects of estimating $\beta$ relative to the infeasible estimators $\hat{\theta}(\cdot)$ and $\hat{\theta}_\alpha(\cdot)$.

We are now in a position to derive the asymptotic distributions of our estimators $\tilde{f}^*(\cdot)$ and $\tilde{U}_{0Q}(\cdot)$ on $[\theta^*,\theta_\dagger]$ and $[\underline{Q},Q_\dagger]$, where $\theta_\dagger \in (\theta^*,\overline{\theta})$ and $Q_\dagger \in (\underline{Q},\overline{Q})$.

**Proposition 4:** *Under A1, B1, B2 and C1, for any $\theta_\dagger \in (\theta^*,\overline{\theta})$ and $Q_\dagger \in (\underline{Q},\overline{Q})$, as $N\to\infty$ we have*

$$\sqrt{N}[\tilde{f}^*(\theta)-f^*(\theta)] \Rightarrow -\gamma[f^*(\theta)+\theta f_\theta^*(\theta)]\mathcal{Z}[F^*(\theta)] + \frac{H_\theta(\theta)}{H(\theta)}\mathcal{B}_{F^*}(\theta) - \gamma b(\theta)\mathcal{N},$$

$$\sqrt{N}[\tilde{U}_{0Q}(Q)-U_{0Q}(Q)] \Rightarrow -U_{0Q}(Q)\left[\gamma\mathcal{Z}[G^{Q*}(Q)]+\frac{T_Q(Q;\beta)-\gamma}{T_Q(Q;\beta)}\frac{\mathcal{B}_{G^{Q*}}(Q)}{1-G^{Q*}(Q)}+c(Q)\mathcal{N}\right],$$

*as random functions in* $\ell^\infty[\theta^*,\theta_\dagger]$ *and* $\ell^\infty[\underline{Q},Q_\dagger]$*, respectively, where*

$$\begin{aligned} b(\theta) &= [f^*(\theta)+\theta f^*_\theta(\theta)]I[F^*(\theta)]-\frac{f^*(\theta)}{H(\theta)}a[F^*(\theta)], \\ c(Q) &= \left[\gamma I[G^{Q*}(Q)]+\frac{T^{-1}_{t\beta}(T;\beta)}{T^{-1}_t(T;\beta)}+\frac{U_{0QQ}(Q)}{U_{0Q}(Q)}T^{-1}_\beta(T;\beta)\right] \end{aligned}$$

*are nonstochastic* $(1\times\dim\beta)$ *vectors and* $T=T(Q;\beta)$*.*

For each statement, the first two terms are identical to the limiting process of the corresponding infeasible estimator in Proposition 3, while the third term arises from estimating $\beta$ as $\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{D}\mathcal{N}\sim\mathcal{N}(0,\Omega)$.

The limiting processes in Proposition 4 are tight Gaussian processes with zero means and finite covariance functions. A difficulty in determining the latter is that the processes $\mathcal{Z}(\cdot)$ and $\mathcal{B}(\cdot)$ are not independent from the random vector $\mathcal{N}$ as Lemma C.1 shows. This is because $\mathcal{Z}[G^{Q*}(\cdot)]$ and $\mathcal{B}_{G^{Q*}}(\cdot)$ arises from $\{Q_i=T^{-1}(t_i);i=1,\ldots,N\}$, while $\mathcal{N}$ arises from $\hat{\beta}$, which depends on $\{(t_i,q_i);i=1,\ldots,N\}$. Using Proposition 4, we obtain

$$\sqrt{N}[\tilde{f}^*(\theta)-f^*(\theta)]\xrightarrow{D}\mathcal{N}\left(0,\omega^2_{f^*}(\theta)\right),\quad \sqrt{N}[\tilde{U}_{0Q}(Q)-U_{0Q}(Q)]\xrightarrow{D}\mathcal{N}\left(0,\omega^2_{U_{0Q}}(Q)\right)$$

for $\theta\in[\theta^*,\theta_\dagger]$ and $Q\in[\underline{Q},Q_\dagger]$, respectively. Given a linear representation of $\sqrt{N}(\hat{\beta}-\beta)$, Appendix C provides detailed computations of $\omega^2_{f^*}(\theta)$ and $\omega^2_{U_{0Q}}(Q)$. It also discusses their consistent estimation.

# 5 Application to Mobile Service

This section applies the methods of Section 4 to mobile service data. Because we do not have information on all the components that make up the payment, we consider product unobserved heterogeneity. Some counterfactuals assess the benefits of nonlinear pricing over alternative pricing strategies.

Data

We obtained data from a major asian mobile phone company. A random sample of 4,000 consumers covers the billing period of May 2009 and consists of subscribers who were under the May 2009 tariff.[20] For each consumer, we observe the amount paid and the total number of minutes consumed. The first two lines of Table 1 provide summary statistics on the number of minutes and the payment expressed in U.S. dollars. Figure 1 displays the scatter plot of observations. A striking feature is that there are as many pairs $(t_i, q_i)$ as observations. The data also show a large variability in prices at a given quantity and in quantities at a given price. This arises from additional features that subscribers consume but that we do not observe. This includes roaming, phone rings, call forwarding, directory assistance, voice mail, etc. These extra features are captured in the product unobserved heterogeneity term $\epsilon$. Regressing payments on quantities and their squares gives an $R^2$ of about 60% suggesting again an important unobserved heterogeneity while the mean tariff is strictly concave.

Implementation

We use the transformation model (24) with a flexible parametric specification for $T^{-1}(\cdot)$. Specifically, we use the spline-based functional form

$$T^{-1}(t;\beta) = \beta_0 + \beta_1 t + \sum_{k=1}^{K} \delta_k \psi_k(t) \tag{30}$$

where $\beta = (\beta_0, \beta_1, \delta_1, \ldots, \delta_K)$ with $\dim \beta = K + 2$ and $K$ is the number of interior knots. Following Dole (1999) the $\psi_k(\cdot)$s are the basis functions

$$\psi_k(t) = \begin{cases} 0 & \text{if } t \in [-\infty, \tau_{k-1}] \\ (t-\tau_{k-1})^3/[6(\tau_k - \tau_{k-1})] & \text{if } t \in [\tau_{k-1}, \tau_k] \\ ((t-\tau_{k+1})^3/[6(\tau_k - \tau_{k+1})]) + a_1 t + a_0 & \text{if } t \in [\tau_k, \tau_{k+1}] \\ a_1 t + a_0 & \text{if } t \in [\tau_{k+1}, +\infty] \end{cases}$$

where $a_1 = (\tau_{k+1} - \tau_{k-1})/2$ and $a_0 = [(\tau_k - \tau_{k-1})^2 - (\tau_k - \tau_{k+1})^2 - 3\tau_k(\tau_{k+1} - \tau_{k-1})]/6$. In practice, we partition the range $[\underline{t}, \overline{t}]$ into equally spaced $K + 1 = 5$ bins of the form $[\tau_{k-1}, \tau_k)$ for $k = 1, \ldots, 5$ with $\tau_0 = t_{\min}$ and $\tau_5 = t_{\max}$. We estimate the inverse tariff imposing shape constraints, i.e. the tariff is increasing and concave. With the Dole basis functions, monotonicity and convexity of $T^{-1}(\cdot)$ is achieved with $\beta_1 > 0$, $\delta_k > 0$, $k = 1, \ldots, 4$.

As is well known, identification of the transformation model (24) is achieved under a location and scale normalization. Several options are possible. We choose a normalization

[20]The company changes their tariff every year. Consumers can switch to the new tariff upon request at no extra cost. In addition, consumers pay for what they consume avoiding the typical problem associated with usage uncertainty. See Miravete (2002) and Grubb and Osborne (2014).

at the lowest payment $t_{\min} = 13.97$. Specifically, given (30), we remark

$$T^{-1}(t_{\min}) = \beta_0 + \beta_1 t_{\min}, \quad T_t^{-1}(t_{\min}) = \beta_1. \tag{31}$$

We propose to estimate $T^{-1}(t_{\min})$ and $T_t^{-1}(t_{\min})$ by a second degree local polynomial regression of $q$ on $t$ using a Gaussian kernel and a standard bandwidth. We obtain $\hat{T}^{-1}(t_{\min}) = 195.2$ and $\hat{T}_t^{-1}(t_{\min}) = 46.11$ thereby providing $\hat{\beta}_0 = -448.96$ and $\hat{\beta}_1 = 46.11$ by solving (31).

Hereafter, $\beta = (\hat{\beta}_0, \hat{\beta}_1, \delta_1, \delta_2, \delta_3, \delta_4)$. To estimate $(\delta_1, \delta_2, \delta_3, \delta_4)$, we use Linton, Sperlich and Van Keilegom (2008) minimum distance estimator from independence. For any fixed $\beta$, let $\log \hat{\epsilon}_i(\beta)$ be the $i$th residual of a nonparametric regression of $\log T^{-1}(t;\beta)$ on $q$, i.e. $\log \hat{\epsilon}_i(\beta) = \log T^{-1}(t_i;\beta) - \log \hat{\tilde{m}}(q_i;\beta)$, where $\log \hat{\tilde{m}}(\cdot;\beta)$ is a kernel estimator of the regression $\mathrm{E}[\log T^{-1}(t;\ \beta)|q = \cdot]$. For this regression, we use a triweight kernel $K(u) = (35/32)(1-u^2)^3 I\!I(|u| \leq 1)$ and a standard bandwidth of the form $h = 1.06\hat{\sigma}_q N^{-1/5}$, where $\hat{\sigma}_q$ is the empirical standard deviation of the minute consumptions. For any fixed $\beta$, consider the following empirical distribution functions

$$\hat{G}^q(q) = \frac{1}{N}\sum_{i=1}^{N} I\!I(q_i \leq q), \quad \hat{F}^{\epsilon(\beta)}(e) = \frac{1}{N}\sum_{i=1}^{N} I\!I(\hat{\epsilon}_i(\beta) \leq e),$$
$$\hat{F}^{q,\epsilon(\beta)}(q,e) = \frac{1}{N}\sum_{i=1}^{N} I\!I(q_i \leq q) I\!I(\hat{\epsilon}_i(\beta) \leq e).$$

The estimator $\hat{\beta}$ is the minimizer of the criterion function

$$Q_N(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{F}^{q,\epsilon(\beta)}(q_i, \hat{\epsilon}_i(\beta)) - \hat{G}^q(q_i)\hat{F}^{\epsilon(\beta)}(\hat{\epsilon}_i(\beta))\right]^2.$$

Once we have the estimated $\hat{\beta}$, an estimate of $\tilde{m}(\cdot)$ is obtained by a kernel regression of $\log T^{-1}(t_i;\hat{\beta})$ on $q_i$ giving $\log \hat{\tilde{m}}(\cdot) = \hat{\mathrm{E}}[\log T^{-1}(t;\hat{\beta})|q = \cdot]$. The density of $\log \epsilon$ is then estimated by a kernel density estimator from the estimated residuals $\log \hat{\epsilon}_i(\hat{\beta})$ using the above kernel and bandwidth based on $\hat{\sigma}_{\hat{\epsilon}}$.

### Estimation Results

The estimated tariff as a function of $Q = \hat{\tilde{m}}(q)\hat{\epsilon}$ is displayed in Figure 2. Though not shown, the estimated $\tilde{m}(\cdot)$ function is increasing in $q$ and slightly convex. This agrees with the economic intuition that the unobserved contracted quantity $Q$ is increasing with the consumption $q$. The tariff shows an important curvature as suggested by the preliminary data analysis. Figure 3 displays a skewed estimated density for $\epsilon$. Most of its values range in the $[0, 2]$ interval. From Table 1, the variability in unobserved heterogeneity is more important than in consumption of minutes as measured by their respective coefficients of variation, 0.81 and 0.61.

With the estimated tariff, we proceed to the estimation of the cost parameters using (25). These require estimates of $\underline{Q}$ and $\overline{Q}$. We use the minimum and maximum estimators, respectively, i.e. $Q_{\min} = \hat{T}^{-1}(t_{\min}) = 195.2$ (by the location normalization) and $Q_{\max} = \max_i \hat{Q}_i = 9,171.03$ from Table 1. We obtain $\hat{\gamma} = 0.0065$. This leads to a variable cost of $0.0065 \times 1,087 = 7.07$, which can be compared to a payment of 30.37 for the median consumer. Moreover, the ratio of the variable cost over the payment varies and tends to increase for large consumers. The estimated fixed cost is $\hat{\kappa} = 2.92$. This means that the fixed cost to serve the median consumer is roughly 9.6% of his bill.

Implementing the quantile estimator of Section 4 provides the estimated base utility and type density. Figure 4 displays $\hat{\theta}(\alpha)$, which is strictly increasing. Figure 5 displays $\hat{U}_{0Q}(\cdot)$, which satisfies A1-(i), i.e. the estimated base utility is concave. Figure 6 displays the estimated truncated type density $\hat{f}^*(\cdot)$. We remark that the figure displays a truncation at one since we normalize $\theta^*$ at one and we cannot identify the type density below $\theta^*$. The shape suggests an exponential distribution with a thin tail though the variance is much smaller than the square of the mean. The estimated density verifies A1-(ii) as shown in Figure 7 which displays the estimated $\theta - [1 - \hat{F}(\theta)]/\hat{f}(\theta)$, which is strictly increasing. Lastly, Table 1 provides summary statistics on the estimated type and the informational rents left to consumers. Using (1), the individual informational rent is estimated by $\hat{\theta}_i \hat{U}_0(\hat{Q}_i) - \hat{T}(\hat{Q}_i)$, where $\hat{U}_0(\cdot) = \hat{T}(Q_{\min}) + \int_{Q_{\min}}^{\cdot} \hat{U}_{0Q}(q)dq$ and $\hat{\theta}_i = \hat{\theta}(\hat{\alpha}_i)$ with $\hat{\alpha}_i$ being the quantile of the payment $t_i$. The rent ratio (rent divided by payment) is on average 47% with an important variability following the important heterogeneity of consumers. Because of skewness, it is more informative to report its median, which is 41%.

Counterfactuals

We simulate the outcomes of three alternative pricing strategies. For each pricing strategy, we assume that the company can choose four price parameters. We perform a grid search for their values that maximize the firm's profit. We then measure the loss in profit as well as the potential gain/loss for consumers. Since we cannot identify the type distribution and the utility function on their full supports, a natural question is how this influences the simulation results. Wong (2012) shows that it is generally not profitable for the firm to exclude more consumers when implementing alternative pricing strategies. This result relates to the dominance of nonlinear pricing in terms of firm's profit subject to some property of the type density. It is worthnoting that a large class of type densities satisfies such a property. In view of this result, the nonidentification of $f(\theta)$ for values below one does not constitute an obstacle to perform counterfactuals. As a matter of fact, we show that alternative pricing strategies tend to exclude more consumers. To measure the net consumer surplus or informational rents, we assume that the utility function is linear for values below $Q_{\min}$ with continuity at $Q_{\min}$ and $U_0(0) = 0$.

The results of our counterfactuals are summarized in Table 2. We measure the total informational rent or consumer surplus, the firm's profit, the total welfare (as the sum of the two), the total amount of contracted quantity assuming that the level of unobserved product heterogeneity remains the same and the total payment made by consumers. The nonlinear pricing (NLP) column reports the actual values while the other columns report the corresponding values for the three alternative pricing strategies in proportions relative to the NLP values. Table 3 assesses the winners and losers upon dividing the sample into four equal subsamples ranking the consumers from the lowest to the highest type.

A first counterfactual consists in a menu of two-part tariffs of the form $t = c_j + p_j Q$, $j = 1, 2$, where $c_j$ and $p_j$ are the fixed fee and the marginal price. The values maximizing the firm's profit are $c_1 = 12.51$, $p_1 = 0.0167$, $c_2 = 34.76$ and $p_2 = 0.0094$. The results confirm that a simple two-part tariff does almost as well as nonlinear pricing as the firm's profit is smaller by less than 1% despite excluding about 5% of our sample of 4,000 consumers. As a result, the low-type group is the most hurt by the two-part tariff because of the exclusion of consumers though those who buy tend to consume more. Overall, this policy tends to benefit the two medium groups as the marginal price tends to be more advantageous for them. The high-type group is about indifferent though they tend to consume less because they do not benefit as much from price discounts. To summarize, simple two-part tariffs can perform efficiently at the social cost of excluding more consumers though this effect can be reduced by increasing the number of two-part tariffs beyond two.

A second counterfactual consists in a menu of linear tariffs with minimum purchase of the form $t = p_j Q$ for $m_j \leq Q < m_{j+1}$, $j = 1, 2$, where $m_j$ and $p_j$ are the minimum quantity and the marginal price, respectively. The values maximizing the firm's profit are $m_1 = 920.64$, $p_1 = 0.0327$, $m_2 = 3,415.6$ and $p_2 = 0.0166$. This pricing excludes even more consumers, namely 20%, as the minimum price paid becomes 30.10 while the minimum purchased quantity increases from 195 as observed in the data to 921. Not surprisingly, this pricing strategy greatly hurts the low-type group as their surplus is only 8% of what it was under nonlinear pricing. All the other groups are hurt as well though not as much with the high-type group loosing the least. Overall, all the indicators of interest decrease by 4% to 12% hurting the consumers the most.

A third counterfactual consists of a menu of quantity forcing or plans with a fixed maximum number of minutes, i.e. the tariff is of the form $t = T_j$ with $Q_j$ for $j = 1, 2$, where $T_j$ and $Q_j$ are the monthly fee and the number of minutes, respectively. The values maximizing the firm's profit are $T_1 = 20.49$, $Q_1 = 522.34$, $T_2 = 53.50$ and $Q_2 = 3,305.5$. This pricing strategy tends to correspond to what we observe in several countries though the degree of customization remains important including in the US. Additional consumer exclusion is minimal (15 out of 4,000 consumers). This tariff benefits the low-type group because it offers a larger possible consumption for a small increase in payment. The other groups lose at different levels. It seems that the medium-high types are hurt the most as they are forced to pay more than under nonlinear pricing. The loss in consumer surplus for the high-type group is only 2%. These consumers would be willing to consume larger quantities. Overall, this pricing does not provide as much profit and consumer surplus with losses of 6% and 4%, respectively.

# 6 Conclusion

This paper studies the identification and estimation of the nonlinear pricing model. Identification is achieved by exploiting the first-order conditions of both the firm and the consumer under a parameterization of the cost function. As in the previous literature on the identification of incomplete information models, the one-to-one mapping between the unknown consumer's type and his observed consumption plays a crucial role. We propose a new quantile-based nonparametric estimator for the model primitives. A striking property of our estimator is its $\sqrt{N}$-consistency, a rate that has not been attained so far in the estimation of incomplete information models. In addition, we introduce product unobserved heterogeneity and show how our results extend to an unknown tariff. An illustration with cellular phone data assesses the performance of alternative pricing strategies relative to nonlinear pricing.

We find that a two-part tariff would be the least hurtful to consumers and the firm.

Our paper proposes a general methodology for second-degree and third-degree price discrimination as third-degree price discrimination can be entertained by introducing some observed consumers' characteristics that affect the tariff. This paper represents a step toward the identification and estimation of incomplete information models for bundling and differentiated products while endogeneizing the product attributes. For instance, Luo (2012) develops a model with nonlinear pricing and bundling and shows how our results extend to his case. Relying on Armstrong (1996) model, Luo, Perrigne and Vuong (2012) extend our results to a multiproduct firm or equivalently to a product with mutiple continuous attributes. Luo, Perrigne and Vuong (2013) consider a general framework for differentiated products with several endogenous continuous and discrete attributes. This paper also shows how to exploit multimarket data to identify the cost function nonparametrically. Despite the difficulties associated to multidimensional screening in these extensions, the basic ideas of our results remain though some adjustments need to be made. Lastly, the model contains all the key ingredients to analyze contract data under incomplete information such as in labor or retailing. Thus, our methodology can be used for this purpose thereby opening several avenues for future research.

.

## Appendix A

This appendix collects the proofs of Lemma 1 and Proposition 2 in Section 2.2.

**Proof of Lemma 1:** Let $\tilde{\theta} = \alpha\theta$, which is distributed as $\tilde{F}(\cdot)$ on $[\underline{\tilde{\theta}}, \overline{\tilde{\theta}}] = [\alpha\underline{\theta}, \alpha\overline{\theta}]$. Let $\tilde{T}(\cdot) \equiv T(\cdot)$, $\tilde{Q}(\cdot) \equiv Q(\cdot/\alpha)$, $\tilde{\theta}^* = \alpha\theta^*$. First we show that $\tilde{T}(\cdot)$, $\tilde{Q}(\cdot)$ and $\tilde{\theta}^*$ satisfy the necessary conditions (5), (6) and (7). We then show that $G^{\tilde{Q}*}(\cdot) = G^{Q*}(\cdot)$, where $G^{\tilde{Q}*}(\cdot)$ is the truncated distribution of $\tilde{Q}$. Hence, the observables $\left[\tilde{T}(\cdot), G^{\tilde{Q}*}(\cdot)\right]$ generated by the structure $\tilde{S}$ are the same observables $\left[T(\cdot), G^{Q*}(\cdot)\right]$ generated by the structure $S$. Lastly, we show $\tilde{S} \in \mathcal{S}$.

To show $\tilde{T}_Q(\tilde{Q}(\tilde{\theta})) = \tilde{\theta}\tilde{U}_{0Q}(\tilde{Q}(\tilde{\theta}))$ for all $\tilde{\theta} \in (\tilde{\theta}^*, \overline{\tilde{\theta}}]$, we rewrite this equation using the definition of $\tilde{T}(\cdot)$, $\tilde{U}_0(\cdot)$ and $\tilde{Q}(\cdot)$. This gives $T_Q(Q(\tilde{\theta}/\alpha)) = (\tilde{\theta}/\alpha)U_{0Q}(Q(\tilde{\theta}/\alpha))$ for all $\tilde{\theta} \in (\tilde{\theta}^*, \overline{\tilde{\theta}}]$, which is true because of (6) with $\theta = (\tilde{\theta}/\alpha) \in [\theta^*, \overline{\theta}]$. To show $\tilde{\theta}\tilde{U}_{0Q}(\tilde{Q}(\tilde{\theta})) = C_Q(\tilde{Q}(\tilde{\theta})) + [(1 - \tilde{F}(\tilde{\theta})/\tilde{f}(\tilde{\theta})]\tilde{U}_{0Q}(\tilde{Q}(\tilde{\theta}))$ for all $\tilde{\theta} \in (\tilde{\theta}^*, \overline{\tilde{\theta}}]$, we rewrite this equation using the definition of $\tilde{U}_0(\cdot)$, $\tilde{Q}(\cdot)$ and $\tilde{F}(\cdot)$:

$$\frac{\tilde{\theta}}{\alpha}U_{0Q}(Q(\tilde{\theta}/\alpha)) = C_Q(Q(\tilde{\theta}/\alpha)) + \frac{1 - F(\tilde{\theta}/\alpha)}{f(\tilde{\theta}/\alpha)}U_{0Q}(Q(\tilde{\theta}/\alpha))$$

for all $\tilde{\theta} \in (\tilde{\theta}^*, \overline{\tilde{\theta}}]$. The above equation holds for all $\theta = \tilde{\theta}/\alpha \in (\theta^*, \overline{\theta}]$ in view of (5). Regarding (7), we follow the same steps and obtain the equivalent of (7) with $\tilde{\theta}^*/\alpha$ for the argument of $Q(\cdot)$. This equation holds for $\tilde{\theta}^*/\alpha = \theta^*$ in view of (7).

Next, we show that the observables coincide. Since $\tilde{T}(\cdot) = T(\cdot)$, it suffices to show $G^{\tilde{Q}*}(\cdot) = G^{Q*}(\cdot)$. Namely,

$$\begin{aligned} G^{\tilde{Q}*}(y) = \Pr[\tilde{Q}(\tilde{\theta}) \leq y | \tilde{Q}(\tilde{\theta}) > \tilde{Q}(\tilde{\theta}^*)] &= \Pr[\tilde{\theta} \leq \tilde{Q}^{-1}(y) | \tilde{\theta} > \tilde{Q}^{-1}(Q(\tilde{\theta}^*/\alpha))] \\ &= \Pr[\alpha\theta \leq \alpha Q^{-1}(y) | \alpha\theta > \alpha Q^{-1}(Q(\theta^*))] \\ &= \Pr[\theta \leq Q^{-1}(y) | \theta > Q^{-1}(Q(\theta^*))] \\ &= \Pr[Q(\theta) \leq y | Q(\theta) > Q(\theta^*)] = G^{Q*}(y), \end{aligned}$$

using the monotonicity of $\tilde{Q}(\cdot)$ and $Q(\cdot)$.

Lastly, we verify that the structure $\tilde{S}$ belongs to $\mathcal{S}$. Assumption A1-(i,iii) is trivially satisfied. Regarding A1-(ii), we have

$$\tilde{\theta} - \frac{1 - \tilde{F}(\tilde{\theta})}{\tilde{f}(\tilde{\theta})} = \tilde{\theta} - \frac{1 - F(\tilde{\theta}/\alpha)}{(1/\alpha)f(\tilde{\theta}/\alpha)} = \alpha\left[\frac{\tilde{\theta}}{\alpha} - \frac{1 - F(\tilde{\theta}/\alpha)}{f(\tilde{\theta}/\alpha)}\right],$$

which is strictly increasing in $\tilde{\theta}/\alpha$ and hence in $\tilde{\theta}$. □

**Proof of Proposition 2:** In view of the discussion in the text, it suffices to establish the identification of $\kappa$. From (7), we obtain

$$\begin{aligned}\kappa &= \theta^* U_0(Q(\theta^*)) - \gamma Q(\theta^*) - \frac{1-F(\theta^*)}{f(\theta^*)} U_0(Q(\theta^*)) \\ &= U_0(\underline{Q})\left(\theta^* - \frac{1-F(\theta^*)}{f(\theta^*)}\right) - \gamma \underline{Q} \\ &= \gamma \frac{U_0(\underline{Q})}{U_{0Q}(\underline{Q})} - \gamma \underline{Q} = \gamma\left(\frac{T(\underline{Q})}{T_Q(\underline{Q})} - \underline{Q}\right),\end{aligned}$$

where the third inequality is obtained from (5) evaluated at $\theta^*$ and the fourth equality exploits the boundary condition $\theta^* U_0(\underline{Q}) = T(\underline{Q})$ and (6) evaluated at $\theta^* = 1$ by B2. $\square$

## Appendix B

This appendix collects the proofs of Lemmas 2 and 3, Proposition 3 in Section 3 as well as the derivation and estimation of asymptotic variances.

**Proof of Lemma 2:** We note that $g^{Q*}(\cdot)$ is continuous and bounded away from zero on $[\underline{Q}, \overline{Q}]$. Thus, from Galambos (1978), we have (i) $Q_{\max} = \overline{Q} + O_{a.s.}[(\log\log N)/N]$ and $Q_{\min} = \underline{Q} + O_{a.s.}[(\log\log N)/N]$, and (ii) $N(Q_{\max} - \overline{Q}) \xrightarrow{D} -\mathcal{E}[g^{Q*}(\overline{Q})]$ and $N(Q_{\min} - \underline{Q}) \xrightarrow{D} \mathcal{E}[g^{Q*}(\underline{Q})]$ as $N \to \infty$. Specifically, (i) follows from Galambos (1978) Theorem 4.3.1 and Example 4.3.2 by letting $u_N = \overline{Q} - \delta(\log\log N)/N$ for any $\delta > 1$ so that $\sum_{N=2}^{\infty}[1 - G^{Q*}(u_N)]\exp\{-N[1 - G^{Q*}(u_N)]\} \equiv \sum_{N=2}^{\infty} v_N < \infty$ since $v_N \sim \tilde{v}_N \equiv \delta g^{Q*}(\overline{Q})[\log\log N]/[N(\log N)^{\delta g^{Q*}(\overline{Q})}]$ as $N \to \infty$ with $\sum_{N=2}^{\infty} \tilde{v}_N < \infty$. Thus, $\Pr[\overline{Q} - Q_{\max} \geq \delta(\log\log N)/N \text{ i.o.}] = 0$, i.e. $\Pr[0 \leq (N/\log\log N)(\overline{Q} - Q_{\max}) \leq \delta \text{ for } N \text{ sufficiently large}] = 1$. Similarly, (ii) follows from Galambos (1978) Theorem 2.1.2 and Section 2.3.1 with $a_N = \overline{Q}$ and $b_N = \overline{Q} - G^{Q*-1}(1 - 1/N)$. Specifically, since $\lim_{t\to\infty}[1 - G^{Q*}(\overline{Q} - 1/(tx))]/[1 - G^{Q*}(\overline{Q} - 1/t)] = 1/x$ for $x > 0$, we obtain $(Q_{\max} - \overline{Q})/b_N \xrightarrow{D} -\mathcal{E}(1)$, i.e. (ii) as $b_N \sim 1/[g^{Q*}(\overline{Q})N]$ as $N \to \infty$. A similar argument applies to $\underline{Q}$ with appropriate adjustments. Moreover, from Galambos (1978, p. 118), $N(Q_{\max} - \overline{Q})$ and $N(Q_{\min} - \underline{Q})$ are asymptotically independent.

The lemma then follows from the standard delta method. Namely, using a Taylor expansion and the continuous differentiability of $T_Q(\cdot)$, we have after some algebra

$$\begin{aligned}\hat{\gamma} - \gamma &= T_Q(Q_{\max}) - T_Q(\overline{Q}) = T_{QQ}(\tilde{Q})(Q_{\max} - \overline{Q}), \\ \hat{\kappa} - \kappa &= -T_Q(\overline{Q})\frac{T(\tilde{\tilde{Q}})T_{QQ}(\tilde{\tilde{Q}})}{T^2(\tilde{\tilde{Q}})}(Q_{\min} - \underline{Q}) + T_{QQ}(\tilde{Q})\left(\frac{T(\underline{Q})}{T_Q(\underline{Q})} - \underline{Q}\right)(Q_{\max} - \overline{Q}) \\ &\quad - T_{QQ}(\tilde{Q})\frac{T(\tilde{\tilde{Q}})T_{QQ}(\tilde{\tilde{Q}})}{T^2(\tilde{\tilde{Q}})}(Q_{\max} - \overline{Q})(Q_{\min} - \underline{Q})\end{aligned}$$

where $Q_{\max} < \tilde{Q} < \overline{Q}$ and $\underline{Q} < \tilde{\tilde{Q}} < Q_{\min}$. Statements (i) and (ii) of the lemma follow. □

To prove Lemma 3, we first need some properties of the empirical quantile process.

**Lemma B.1:** *Under B3, the empirical quantile process* $\hat{Q}(\cdot) = \hat{G}^{Q*-1}(\cdot)$ *satisfies*

$$(\text{P1}) : |\hat{Q}(\cdot) - Q(\cdot)| \xrightarrow{a.s.} 0, \quad (\text{P2}) : \sqrt{N}[\hat{Q}(\cdot) - Q(\cdot)] \Rightarrow -\frac{\mathcal{B}(\cdot)}{g^{Q*}[Q(\cdot)]},$$

*uniformly on* $[0,1]$ *as* $N \to \infty$, *where* $\mathcal{B}$ *is the standard Brownian bridge on* $[0,1]$.

**Proof of Lemma B.1:** Under C1, the quantities $Q_i = Q(\theta_i)$ are i.i.d. as $G^{Q*}(\cdot)$. Properties (P1)-(P2) on $(0,1)$ follow from the Hadamard differentiability of the inverse map $\phi(\cdot) : \mathcal{D}[\underline{Q}, \overline{Q}] \mapsto \ell^\infty(0,1)$ at $G^{Q*}$ tangentially to $\mathcal{C}[\underline{Q}, \overline{Q}]$. See Lemma 3.9.23-(ii) and Example 3.9.24 in van der Vaart and Wellner (1996) as $g^{Q*}(\cdot)$ is strictly positive with compact support $[\underline{Q}, \overline{Q}]$. Because $\hat{Q}(0) = Q_{\min}$ and $\overline{Q}$ finite, defining $Q(0) = \underline{Q}$ allows us to obtain (P1) and (P2) on $[0,1]$ instead of $(0,1)$ using Lemma 2-(i). In particular, $N[\hat{Q}(0) - Q(0)] \xrightarrow{D} 0$ and $N[\hat{Q}(1) - Q(1)] \xrightarrow{D} 0$.□

**Proof of Lemma 3:** From (P1)-(P2) and the continuous differentiability of $T_Q(\cdot)$ on $[\underline{Q}, \overline{Q}]$, we have the following properties

$$(\text{P3}) : |T_Q[\hat{Q}(\cdot)] - T_Q[Q(\cdot)]| \xrightarrow{a.s.} 0, \quad (\text{P4}) \sqrt{N}\left\{T_Q[\hat{Q}(\cdot)] - T_Q[Q(\cdot)]\right\} \Rightarrow -T_{QQ}[Q(\cdot)]\frac{\mathcal{B}(\cdot)}{g^{Q*}[Q(\cdot)]}$$

uniformly on $[0,1]$ as $N \to \infty$. Property (P3) follows from the Continuous Mapping Theorem, while (P4) follows from the Functional Delta Method. See Theorems 18.11 and 20.8 in van der Vaart (1998), respectively. The Hadamard derivative of $T_Q[\cdot]$ as a map $\mathcal{D}_{T_Q} \mapsto \ell^\infty[0,1]$ at every $\psi(\cdot) \in \mathcal{D}_{T_Q}$ is $T_{QQ}[\psi(\cdot)]$, where $\mathcal{D}_{T_Q} = \{\psi(\cdot) \in \ell^\infty[0,1] : \underline{Q} \le \psi(\cdot) \le \overline{Q}\}$ by Lemma 3.9.25 in van der Vaart and Wellner (1996).

From (8), (9) and (11), $\theta(\alpha) > 0$ and $\theta_\alpha(\alpha) > 0$ as well as $\hat{\theta}(\alpha) > 0$ and $\hat{\theta}_\alpha(\alpha) > 0$ for any $\alpha \in [0, \alpha_\dagger]$ with $\alpha_\dagger \in (0,1)$ since $T_Q(Q(\cdot))$ is decreasing and $\alpha_\dagger < (N-1)/N$ for $N$ sufficiently large. Thus,

$$\begin{aligned}
\log\frac{\hat{\theta}(\alpha)}{\theta(\alpha)} &= \int_0^\alpha \frac{1}{1-u}\left[\frac{\gamma}{T_Q[Q(u)]} - \frac{\hat{\gamma}}{T_Q[\hat{Q}(u)]}\right]du \\
&= \int_0^\alpha \frac{1}{1-u}\left[\frac{\gamma(T_Q[\hat{Q}(u)] - T_Q[Q(u)])}{T_Q[Q(u)]T_Q[\hat{Q}(u)]} - (\hat{\gamma}-\gamma)\frac{1}{T_Q[\hat{Q}(u)]}\right]du, \qquad \text{(B.1)} \\
\log\frac{\hat{\theta}_\alpha(\alpha)}{\theta_\alpha(\alpha)} &= \log\frac{\hat{\theta}(\alpha)}{\theta(\alpha)} + \log\frac{T_Q[\hat{Q}(\alpha)] - \hat{\gamma}}{T_Q[Q(\alpha)] - \gamma} - \log\frac{T_Q[\hat{Q}(\alpha)]}{T_Q[Q(\alpha)]}, \qquad \text{(B.2)}
\end{aligned}$$

for $\alpha \in [0, \alpha_\dagger]$. Note also that $T_Q[Q(\cdot)] \ge \gamma > 0$ on $[0,1]$ so that $T_Q[Q(\cdot)]$ is bounded away from zero on $[0,1]$.

Proof of (i): From (B.1), Lemma 2-(i) and Property (P3), we have $\|\log(\hat{\theta}/\theta)\|_\dagger \xrightarrow{a.s.} 0$. By the Continuous Mapping Theorem, we obtain $\|\hat{\theta} - \theta\|_\dagger \xrightarrow{a.s.} 0$ as $\|\theta\|_\dagger < \infty$. From the convergence of $\hat{\theta}(\cdot)$ as shown above, Lemma 2-(i) and Property (P3), we have $\|\log(\hat{\theta}_\alpha/\theta_\alpha)\|_\dagger \xrightarrow{a.s.} 0$. By the Continuous Mapping Theorem, we obtain $\|\hat{\theta}_\alpha - \theta_\alpha\|_\dagger \xrightarrow{a.s.} 0$ as $\|\theta_\alpha\|_\dagger < \infty$.

Proof of (ii): Using Lemma 2-(ii) and Property (P3), it follows from (B.1)

$$\begin{aligned}\sqrt{N}\log\frac{\hat{\theta}(\alpha)}{\theta(\alpha)} &= \int_0^\alpha \frac{1}{1-u}\left[\sqrt{N}\frac{\gamma(T_Q[\hat{Q}(u)] - T_Q[Q(u)])}{T_Q[Q(u)]T_Q[\hat{Q}(u)]} - \frac{\log\log N}{\sqrt{N}}\frac{N}{\log\log N}(\hat{\gamma}-\gamma)\frac{1}{T_Q[\hat{Q}(u)]}\right]du \\ &= \int_0^\alpha \frac{1}{1-u}\left[\sqrt{N}\frac{\gamma(T_Q[\hat{Q}(u)] - T_Q[Q(u)])}{T_Q[Q(u)]T_Q[\hat{Q}(u)]}\right]du + o_{a.s.}(1),\end{aligned}$$

uniformly in $\alpha \in [0, \alpha_\dagger]$. Hence, it follows from Properties (P3)-(P4) and the Continuous Mapping Theorem

$$\sqrt{N}\log\frac{\hat{\theta}(\cdot)}{\theta(\cdot)} \Rightarrow \int_0^\cdot \frac{-\gamma}{1-u}\frac{T_{QQ}[Q(u)]}{T_Q^2[Q(u)]}\frac{\mathcal{B}(u)}{g^{Q*}(Q(u))}du$$

on $[0, \alpha_\dagger]$. With the change of variable $u = G^{Q*}(q)$, we obtain

$$\sqrt{N}\log\frac{\hat{\theta}(\cdot)}{\theta(\cdot)} \Rightarrow \gamma\int_{\underline{Q}}^{Q(\cdot)} \frac{-1}{1-G^{Q*}(q)}\frac{T_{QQ}(q)}{T_Q^2(q)}\mathcal{B}_{G^{Q*}}(q)dq \equiv \gamma\mathcal{Z}(\cdot) \tag{B.3}$$

on $[0, \alpha_\dagger]$, where $\mathcal{B}_{G^{Q*}}(q) = \mathcal{B}\circ G^{Q*}(q)$ is the $G^{Q*}$-Brownian bridge. Thus, from the Functional Delta Method, it follows that $\sqrt{N}[\hat{\theta}(\alpha) - \theta(\alpha)] = \sqrt{N}\left\{\exp[\log\hat{\theta}(\alpha)] - \exp[\log\theta(\alpha)]\right\} \Rightarrow \gamma\theta(\cdot)\mathcal{Z}(\cdot)$ since the Hadamard derivative of $\exp\psi$ at $\log\theta(\cdot)$ is $\theta(\cdot)$. Moreover, the process $\mathcal{Z}(\cdot)$ is Gaussian as it is a continuous linear functional of $\mathcal{B}_{G^{Q*}}(\cdot)$ by Lemma 3.9.8 in van der Vaart and Wellner (1996). It is also tight by Theorem 1.4 in Billingsley (1968) since all its sample paths are continuous, while $\mathcal{C}[0, \alpha_\dagger]$ is complete and separable with respect to the uniform norm.

Using (B.2), (B.3), Property (P4) and the Functional Delta Method, we obtain

$$\begin{aligned}\sqrt{N}\log\frac{\hat{\theta}_\alpha(\alpha)}{\theta_\alpha(\alpha)} &\Rightarrow \gamma\mathcal{Z}(\alpha) - \frac{T_{QQ}[Q(\alpha)]}{T_Q[Q(\alpha)]-\gamma}\frac{\mathcal{B}(\alpha)}{g^{Q*}(Q(\alpha))} + \frac{T_{QQ}[Q(\alpha)]}{T_Q[Q(\alpha)]}\frac{\mathcal{B}(\alpha)}{g^{Q*}(Q(\alpha))} \\ &= \gamma\mathcal{Z}(\alpha) - \gamma\frac{\mathcal{B}_{G^{Q*}}(Q(\alpha))}{g^{Q*}(Q(\alpha))}\frac{T_{QQ}[Q(\alpha)]}{T_Q[Q(\alpha)](T_Q[Q(\alpha)]-\gamma)}\end{aligned}$$

uniformly in $\alpha \in [0, \alpha_\dagger]$, where the first equality follows from $\sqrt{N}\{T_Q[\hat{Q}(\alpha)] - \hat{\gamma}) - (T_Q[Q(\alpha)] - \gamma)\} = \sqrt{N}(T_Q[\hat{Q}(\alpha)] - T_Q[Q(\alpha)]) + o_{a.s.}(1)$ by Lemma 2-(i), while the Hadamard derivative of $\log\psi$ is $1/\psi$. Using a similar argument as above, we obtain

$$\sqrt{N}[\hat{\theta}_\alpha(\cdot) - \theta_\alpha(\cdot)] \Rightarrow \gamma\theta_\alpha(\cdot)\left[\mathcal{Z}(\cdot) - \frac{T_{QQ}[Q(\cdot)]}{T_Q[Q(\cdot)](T_Q[Q(\cdot)]-\gamma)}\frac{\mathcal{B}_{G^{Q*}}(Q(\cdot))}{g^{Q*}(Q(\cdot))}\right]$$

on $[0, \alpha_\dagger]$. □

**Proof of Proposition 3:** We use the following properties of the empirical c.d.f.

$$(\text{P5}):\ |\hat{G}^{Q*}(\cdot) - G^{Q*}(\cdot)| \xrightarrow{a.s.} 0, \qquad (\text{P6}):\ \sqrt{N}[\hat{G}^{Q*}(\cdot) - G^{Q*}(\cdot)] \Rightarrow \mathcal{B}_{G^{Q*}}(\cdot)$$

uniformly on $[\underline{Q},\overline{Q}]$ as $N\to\infty$. Given C1, Properties (P5) and (P6) are well-known properties of the empirical c.d.f. for i.i.d. observations. In particular, (P5) follows from the Glivenko-Cantelli Theorem, while (P6) follows from the Functional Central Limit Theorem. See van der Vaart (1998, p.266).

Proof of (i): Lemma 3-(i) shows the uniform consistency of the quantile estimator $\hat{\theta}(\cdot)$ on $[0,\alpha_\dagger]$. Thus, $\hat{F}^*(\cdot)=\hat{\theta}^{-1}(\cdot)$ is uniformly consistent on $[\theta^*,\theta_\dagger]$ by the Continuous Mapping Theorem. Noting that $\hat{f}^*(\cdot)-f^*(\cdot)=\widehat{f^*\circ\theta}[\hat{\theta}^{-1}(\cdot)]-f^*\circ\theta[\theta^{-1}(\cdot)]=\widehat{f^*\circ\theta}[\hat{\theta}^{-1}(\cdot)]-f^*\circ\theta[\hat{\theta}^{-1}(\cdot)]+f^*\circ\theta[\hat{\theta}^{-1}(\cdot)]-f^*\circ\theta[\theta^{-1}(\cdot)]$, the uniform consistency of $\hat{f}^*(\cdot)$ to $f^*(\cdot)$ on $[\theta^*,\theta_\dagger]$ follows from the uniform consistency of $\widehat{f^*\circ\theta}(\cdot)$ on $[0,\alpha_\dagger]$ as discussed in the text after Lemma 3 and the uniform consistency of $\hat{\theta}^{-1}(\cdot)=\hat{F}^*(\cdot)$ on $[\theta^*,\theta_\dagger]$ as noted above combined with the continuity of $f^*\circ\theta(\cdot)$ on $[0,\alpha_\dagger]$. A similar argument establishes the uniform consistency of $\hat{U}_{0Q}(\cdot)$ to $U_{0Q}(\cdot)$ on $[\underline{Q},Q_\dagger]$.

Proof of (ii): We first derive the asymptotic distribution of $\hat{F}^*(\cdot)=\hat{\theta}^{-1}(\cdot)$. We use the Functional Delta Method and the Hadamard derivative of the inverse mapping. In particular, from Lemma 3-(ii) and van der Vaart and Wellner (1996) Lemma 3.9.23, we have

$$\sqrt{N}[\hat{\theta}^{-1}(\cdot)-\theta^{-1}(\cdot)]\Rightarrow-\left(\frac{\gamma\theta\mathcal{Z}}{\theta_\alpha}\right)\circ\theta^{-1}(\cdot)=-\gamma\theta[\theta^{-1}(\cdot)]f^*(\cdot)\mathcal{Z}[\theta^{-1}(\cdot)],$$

uniformly on $[\theta^*,\theta_\dagger]$ noting that $\theta^{-1}(\cdot)=F^*(\cdot)$ and $\theta_\alpha[F^*(\cdot)]=1/f^*(\cdot)$.

Turning to the asymptotic distribution of $\hat{f}^*(\theta)=\widehat{f^*\circ\theta}[\hat{\theta}^{-1}(\cdot)]$, we apply van der Vaart and Wellner (1996) Lemma 3.9.27 to the composition map $\phi:\ell^\infty([\theta^*,\theta_\dagger],[0,\alpha_\dagger])\times\ell^\infty([0,\alpha_\dagger],I\!\!R)\mapsto\ell^\infty([\theta^*,\theta_\dagger],I\!\!R)$ with $\phi[\theta^{-1},f^*\circ\theta](\cdot)=(f^*\circ\theta)[\theta^{-1}(\cdot)]$. This gives the Hadamard derivative

$$\phi'_{[\theta^{-1},f^*\circ\theta]}(h_1,h_2)(\cdot)=h_2[\theta^{-1}(\cdot)]+f^*_\theta(\cdot)\theta_\alpha(\theta^{-1}(\cdot))h_1(\cdot)$$

since $(f^*\circ\theta)'_{\theta^{-1}(\cdot)}[h_1(\cdot)]=f^*_\theta[\theta(\theta^{-1}(\cdot))]\theta_\alpha[\theta^{-1}(\cdot)]h_1(\cdot)$, where $[h_1(\cdot),h_2(\cdot)]\in\ell^\infty([\theta^*,\theta_\dagger],[0,\alpha_\dagger])\times\mathrm{UC}([0,\alpha_\dagger],I\!\!R)$. Thus from the Functional Delta Method, the asymptotic distribution of $\widehat{f^*\circ\theta}(\cdot)$ in (16) and the asymptotic distribution of $\hat{\theta}^{-1}(\cdot)$ given above, we obtain

$$\begin{aligned}&\sqrt{N}\left[\widehat{f^*\circ\theta}[\hat{\theta}^{-1}(\theta)]-f^*\circ\theta[\theta^{-1}(\theta)]\right]\\ \Rightarrow\ &-\gamma f^*(\theta)\left[\mathcal{Z}[\theta^{-1}(\theta)]-\frac{T_{QQ}[Q(\theta)]}{T_Q[Q(\theta)](T_Q[Q(\theta)]-\gamma)}\frac{\mathcal{B}_{GQ*}[Q(\theta)]}{g^{Q*}[Q(\theta)]}\right]\\ &-f^*_\theta(\theta)\theta_\alpha[\theta^{-1}(\theta)]\gamma\theta[\theta^{-1}(\theta)]f^*(\theta)\mathcal{Z}[\theta^{-1}(\theta)]\end{aligned}$$

uniformly in $\theta\in[\theta^*,\overline{\theta}]$. Collecting the terms in $\mathcal{Z}(\cdot)$ gives the desired result using $\theta_\alpha[\theta^{-1}(\theta)]=1/f^*(\theta)$, $\theta^{-1}(\theta)=F^*(\cdot)$ and

$$\frac{H_\theta(\theta)}{H(\theta)}=\frac{\gamma f^*(\theta)}{g^{Q*}[Q(\theta)]}\frac{T_{QQ}[Q(\theta)]}{T_Q[Q(\theta)](T_Q[Q(\theta)]-\gamma)}. \tag{B.4}$$

Equation (B.4) follows from differentiating $[1-F^*(\theta)]/[\theta f^*(\theta)] = [T_Q[Q(\theta)]-\gamma]/T_Q[Q(\theta)]$ and using $g^{Q*}[Q(\theta)]Q_\theta(\theta) = f^*(\theta)$.

Similarly, applying Lemma 3.9.27 in van der Vaart and Wellner (1996) and the Functional Delta method to the composition map $\phi : \ell^\infty([\underline{Q}, Q_\dagger], [0, \alpha_\dagger]) \times \ell^\infty([0, \alpha_\dagger], I\!R) \mapsto \ell^\infty([\underline{Q}, Q_\dagger], I\!R)$ with $\phi[G^{Q*}, U_{0Q}\circ Q](\cdot) = (U_{0Q}\circ Q)[G^{Q*}(\cdot)]$, we obtain using Property (P6) and (17)

$$\begin{aligned}&\sqrt{N}\left[\widehat{U_{0Q}\circ}Q[\hat{G}^{Q*}(Q)] - U_{0Q}\circ Q[G^{Q*}(Q)]\right]\\ &\Rightarrow -U_{0Q}(Q)\left[\gamma\mathcal{Z}[G^{Q*}(Q)] + \frac{T_{QQ}(Q)}{T_Q(Q)}\frac{\mathcal{B}_{G^{Q*}}(Q)}{g^{Q*}(Q)}\right] + \frac{U_{0QQ}(Q)}{g^{Q*}(Q)}\mathcal{B}_{G^{Q*}}(Q),\end{aligned}$$

uniformly in $Q \in [\underline{Q}, Q_\dagger]$ since $(U_{0Q}\circ Q)'_{G^{Q*}(\cdot)}[h_1(\cdot)] = U_{0QQ}(\cdot)Q_\alpha[G^{Q*}(\cdot)]h_1(\cdot)$ and $Q_\alpha[G^{Q*}(\cdot)] = 1/g^{Q*}(\cdot)$. Collecting the terms in $\mathcal{B}_{G^{Q*}}(\cdot)$ gives the desired result using

$$\frac{T_{QQ}(Q)}{T_Q(Q)} - \frac{U_{0QQ}(Q)}{U_{0Q}(Q)} = \frac{T_Q(Q)-\gamma}{T_Q(Q)}\frac{g^{Q*}(Q)}{1-G^{Q*}(Q)}. \tag{B.5}$$

Equation (B.5) follows from differentiating the logarithm of $T_Q(Q)/U_{0Q}(Q) = \theta[G^{Q*}(Q)]$ and using (8) expressed at $\alpha = G^{Q*}(Q)$.□

**Proof of (18)**: From (15) expressed at $F^*(\theta)$, we make the change of variable $q = Q(t)$. This gives

$$\mathcal{Z}[F^*(\theta)] = -\int_{\theta^*}^{\theta}\frac{T_{QQ}[Q(t)]}{T_Q^2[Q(t)]}\frac{\mathcal{B}_{F^*}(t)}{1-F^*(t)}Q_\theta(t)dt \tag{B.6}$$

using $Q[F(\theta)] = Q(\theta)$, $\mathcal{B}_{G^{Q*}}[Q(t)] = \mathcal{B}_{F^*}(t)$ and $G^{Q*}[Q(t)] = F^*(t)$. We remark that the derivative of $1-\gamma/T_Q[Q(\theta)]$ with respect to $\theta$ is equal to $\gamma T_{QQ}[Q(\theta)]Q_\theta[Q(\theta)]/T^2[Q(\theta)]$. But $1-\gamma/T_Q[Q(\theta)] = [1-F^*(\theta)]/[\theta f^*(\theta)] \equiv H(\theta)$ from the FOC (5) and (6).□

**Computation of $V_{f^*}(\theta)$:** We show that

$$\begin{aligned}V_{f^*}(\theta) \;&=\; 2[f^*(\theta)+\theta f^*_\theta(\theta)]f^*(\theta)\int_{\theta^*}^{\theta}\frac{H(x)f^*(x)}{[1-F^*(x)]^2}dx\\ &\quad+[f^*(\theta)+\theta f^*_\theta(\theta)]^2\int_{\theta^*}^{\theta}\frac{H^2(x)f^*(x)}{[1-F^*(x)]^2}dx + f^{*2}(\theta)\frac{F^*(\theta)}{1-F^*(\theta)}\end{aligned} \tag{B.7}$$

We begin with the covariance of the process $\mathcal{Z}[F^*(\cdot)]$. From the covariance of the Brownian bridge (see e.g. van der Vaart (1998, p.266)), we remark that for $\theta^* \le \theta \le \theta' < \overline{\theta}$

$$\mathrm{E}\left[\frac{\mathcal{B}_{F^*}(\theta)}{1-F^*(\theta)}\frac{\mathcal{B}_{F^*}(\theta')}{1-F^*(\theta')}\right] = \frac{F^*(\theta)}{1-F^*(\theta)},$$

which is independent of $q'$. From the definition of $\mathcal{Z}[F^*(\cdot)]$, we have

$$\mathrm{E}\left\{\mathcal{Z}[F^*(\theta)]\mathcal{Z}[F^*(\theta')]\right\} \;=\; \frac{1}{\gamma^2}\int_{\theta^*}^{\theta}\left[\int_{\theta^*}^{\theta'}H_\theta(t)H_\theta(t')\mathrm{E}\left\{\frac{\mathcal{B}_{F^*}(t)}{1-F^*(t)}\frac{\mathcal{B}_{F^*}(t')}{1-F^*(t')}\right\}dt'\right]dt$$

$$
\begin{aligned}
&= \frac{1}{\gamma^2}\int_{\theta^*}^{\theta}\left[H_\theta(t)\int_{\theta^*}^{t}H_\theta(t')\frac{F^*(t')}{1-F^*(t')}dt' + H_\theta(t)\frac{F^*(t)}{1-F^*(t)}\int_t^{\theta'}H_\theta(t')dt'\right]dt \\
&= \frac{1}{\gamma^2}\frac{1-F^*(\theta)}{\theta f^*(\theta)}\int_{\theta^*}^{\theta}H_\theta(t')\frac{F^*(t')}{1-F^*(t')}dt' - \frac{1}{\gamma^2}\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{tf^*(t)}dt \\
&\quad +\frac{1}{\gamma^2}\frac{1-F^*(\theta')}{\theta' f^*(\theta')}\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{1-F^*(t)}dt - \frac{1}{\gamma^2}\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{tf^*(t)}dt \\
&= \frac{1}{\gamma^2}[H(\theta)+H(\theta')]\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{1-F^*(t)}dt - \frac{2}{\gamma^2}\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{tf^*(t)}dt, \qquad \text{(B.8)}
\end{aligned}
$$

where the third equality follows from an integration by parts for the first term. Moreover,

$$
\begin{aligned}
\mathrm{E}[\mathcal{Z}[F^*(\theta)]\mathcal{B}_{F^*}(\theta')] &= -\frac{1}{\gamma}\int_{\theta^*}^{\theta}H_\theta(t)\mathrm{E}\left[\frac{\mathcal{B}_{F^*}(t)}{1-F(t)}\mathcal{B}_{F^*}(\theta')\right]dt \\
&= -\frac{1-F^*(\theta')}{\gamma}\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{1-F^*(t)}dt. \qquad \text{(B.9)}
\end{aligned}
$$

Now using the limit process in Proposition 3-(ii), (B.8) and (B.9) and setting $\theta = \theta'$ give

$$
\begin{aligned}
V_{f^*}(\theta) &= 2[f^*(\theta)+\theta f^*_\theta(\theta)]^2\left[H(\theta)\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{1-F^*(t)}dt - \int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{tf^*(t)}dt\right] \\
&\quad +2[f^*(\theta)+\theta f^*_\theta(\theta)]\frac{H_\theta(\theta)}{H(\theta)}[1-F^*(\theta)]\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{1-F^*(t)}dt + \frac{H^2_\theta(\theta)}{H^2(\theta)}F^*(\theta)[1-F^*(\theta)] \\
&= -2[f^*(\theta)+\theta f^*_\theta(\theta)]f^*(\theta)\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{1-F^*(t)}dt - 2[f^*(\theta)+\theta f^*_\theta(\theta)]^2\int_{\theta^*}^{\theta}H_\theta(t)H(t)\frac{F^*(t)}{1-F^*(t)}dt \\
&\quad +\frac{H^2_\theta(\theta)}{H^2(\theta)}F^*(\theta)[1-F^*(\theta)]
\end{aligned}
$$

upon collecting terms. Integration by parts of the two integrals followed by some algebra gives (B.7).□

**Computation of $V_{U_{0Q}}(Q)$:** We show that

$$
V_{U_{0Q}}(Q) = U^2_{0Q}(Q)\int_{\underline{Q}}^{Q}\left(\frac{T_Q(q)-\gamma}{T_Q(q)}\right)^2\frac{g^{Q*}(q)}{[1-G^{Q^*}(q)]^2}dq. \qquad \text{(B.10)}
$$

Using its limit process from Proposition 3-(ii), (B.8) and (B.9) as well as noting $\mathcal{Z}[G^{Q*}(Q)] = \mathcal{Z}[F^*(\theta)]$, $\mathcal{B}_{G^{Q*}}(Q) = \mathcal{B}_{F^*}(\theta)$ with $\theta = \theta(Q)$, we have

$$
\begin{aligned}
V_{U_{OQ}} &= U^2_{0Q}(Q)\left[2H(\theta)\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{1-F^*(t)}dt - 2\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{tf^*(t)}dt\right. \\
&\quad \left. -2\frac{T_Q(Q)-\gamma}{T_Q(Q)}\int_{\theta^*}^{\theta}H_\theta(t)\frac{F^*(t)}{1-F^*(t)}dt + \left(\frac{T_Q(Q)-\gamma}{T_Q(Q)}\right)^2\frac{F^*(\theta)}{1-F^*(\theta)}\right] \\
&= U^2_{0Q}(Q)\left[\left(\frac{T_Q(Q)-\gamma}{T_Q(Q)}\right)^2\frac{G^{Q*}(Q)}{1-G^{Q*}(Q)} - 2\int_{\theta^*}^{\theta}H_\theta(t)H(t)\frac{F^*(t)}{1-F^*(t)}dt\right] \\
&= U^2_{0Q}(Q)\left[\left(\frac{T_Q(Q)-\gamma}{T_Q(Q)}\right)^2\frac{G^{Q*}(Q)}{1-G^{Q*}(Q)} - 2\gamma\int_{\underline{Q}}^{Q}\frac{T_{QQ}(q)}{T^2_Q(q)}\frac{T_Q(q)-\gamma}{T_Q(q)}\frac{G^{Q*}(q)}{1-G^{Q*}(q)}dq\right],
\end{aligned}
$$

where the second equality uses $H(\theta) = [1 - F^*(\theta)]/[\theta f^*(\theta)] = [T_Q[Q(\theta)] - \gamma]/T_Q[Q(\theta)]$, while the third equality uses the change of variable $q = Q(t)$. Integrating by parts gives (B.10).□

**Estimation of $V_{f^*}(\theta)$ and $V_{U_{0Q}}(Q)$:** Natural estimators are obtained by replacing the unknown quantities in (B.7) and (B.10) by their estimators. Specifically, regarding $V_{U_{0Q}}(\cdot)$, we remark that it is an expectation with respect to the distribution of $Q$ leading to

$$\hat{V}_{U_{0Q}}(Q) = \hat{U}_{0Q}^2(Q) \frac{1}{N} \sum_{i=1}^{N} 1\!\!1(Q_i \leq Q) \left(1 - \frac{\hat{\gamma}}{T_Q(Q_i)}\right)^2 \frac{1}{[1 - \hat{G}^*(Q_i)]^2}.$$

We can use a similar idea to estimate the two integrals in $V_{f^*}(\cdot)$. Using $H(x) = [T_Q[Q(x)] - \gamma]/T_Q[Q(x)]$ and making the change of variable $q = Q(x)$, we can estimate these two integrals by

$$\frac{1}{N} \sum_{i=1}^{N} 1\!\!1[Q_i \leq \hat{Q}(\theta)] \left(1 - \frac{\hat{\gamma}}{T_Q(Q_i)}\right) \frac{1}{[1 - \hat{G}^{Q*}(Q_i)]^2},$$
$$\frac{1}{N} \sum_{i=1}^{N} 1\!\!1[Q_i \leq \hat{Q}(\theta)] \left(1 - \frac{\hat{\gamma}}{T_Q(Q_i)}\right)^2 \frac{1}{[1 - \hat{G}^{Q*}(Q_i)]^2},$$

respectively, where $\hat{Q}(\cdot) = \hat{G}^{Q*-1}[\hat{\theta}^{-1}(\cdot)]$. It remains the problem of estimating the derivative $f^*_\theta(\cdot)$ of the density in (B.7). We remark that $f^*(\theta) + \theta f^*_\theta(\theta)$ is the derivative of $\theta f^*(\theta) = \{T_Q[Q(\theta)][1 - G^{Q*}(Q(\theta))]\}/[T_Q[Q(\theta)] - \gamma]$. Thus, the term $f^*(\theta) + \theta f^*_\theta(\theta)$ can be estimated by

$$\left(-\hat{\gamma} \frac{T_{QQ}[\hat{Q}(\theta)]}{(T_Q[\hat{Q}(\theta)] - \hat{\gamma})^2} \frac{1 - \hat{G}^{Q*}[\hat{Q}(\theta)]}{\hat{g}^{Q*}[\hat{Q}(\theta)]} - \frac{T_Q[\hat{Q}(\theta)]}{T_Q[\hat{Q}(\theta)] - \hat{\gamma}}\right) \hat{f}^*(\theta),$$

where $\hat{g}^{Q*}(\cdot)$ is (say) the standard kernel density estimator.

## Appendix C

This appendix collects the proofs of Lemmas 4 and 5, Proposition 4 in Section 4 as well as the derivation of estimation of asymptotic variances.

**Proof of Lemma 4:** We have $\tilde{\gamma} - \gamma = \tilde{\gamma} - \hat{\gamma} + \hat{\gamma} - \gamma$. Thus, $\sqrt{N}(\tilde{\gamma} - \gamma) = \sqrt{N}(\tilde{\gamma} - \hat{\gamma}) + o_P(1)$ since $\sqrt{N}(\hat{\gamma} - \gamma) = o_P(1)$ by Lemma 2-(i). Now, from (19) and (25) we have

$$\begin{aligned} \sqrt{N}(\tilde{\gamma} - \hat{\gamma}) &= \sqrt{N} \left( \frac{1}{T_t^{-1}(t_{\max}; \hat{\beta})} - \frac{1}{T_t^{-1}(t_{\max}; \beta)} \right) \\ &= -\gamma^2 T_{t\beta}^{-1}(\bar{t}; \beta) \sqrt{N}(\hat{\beta} - \beta) + o_P(1) \end{aligned} \tag{C.1}$$

using $t_{\max} \xrightarrow{a.s.} \bar{t}$, a Taylor expansion around $\beta$, and $\gamma = 1/T_t^{-1}(\bar{t}; \beta)$. Part (i) follows.

Similarly, we have $\sqrt{N}(\tilde{\kappa}-\kappa)=\sqrt{N}(\tilde{\kappa}-\hat{\kappa})+o_P(1)$ since $\sqrt{N}(\hat{\kappa}-\kappa)=o_P(1)$ by Lemma 2-(i). Moreover, from (19) and (25) we have

$$\begin{aligned}
\sqrt{N}(\tilde{\kappa}-\hat{\kappa}) &= \sqrt{N}\left[\tilde{\gamma}\left(t_{\min}T_t^{-1}(t_{\min};\hat{\beta})-T^{-1}(t_{\min};\hat{\beta})\right)-\hat{\gamma}\left(t_{\min}T_t^{-1}(t_{\min};\beta)-T^{-1}(t_{\min};\beta)\right)\right] \\
&= \sqrt{N}(\tilde{\gamma}-\hat{\gamma})\left[t_{\min}T_t^{-1}(t_{\min};\hat{\beta})-T^{-1}(t_{\min};\hat{\beta})\right] \\
&\quad +\hat{\gamma}t_{\min}\sqrt{N}\left[T_t^{-1}(t_{\min};\hat{\beta})-T_t^{-1}(t_{\min};\beta)\right]-\hat{\gamma}\sqrt{N}\left[T^{-1}(t_{\min};\hat{\beta})-T^{-1}(t_{\min};\beta)\right] \\
&= \frac{\kappa}{\gamma}\sqrt{N}(\tilde{\gamma}-\hat{\gamma})+\gamma\left[\underline{t}T_{t\beta}^{-1}(\underline{t};\beta)-T_{\beta}^{-1}(\underline{t};\beta)\right]\sqrt{N}(\hat{\beta}-\beta)+o_P(1)
\end{aligned}$$

where the last equality uses $\hat{\beta}\xrightarrow{P}\beta$, $\hat{\gamma}\xrightarrow{P}\gamma$, $t_{\min}\xrightarrow{P}\underline{t}$, and a Taylor expansion around $\beta$. Part (ii) then follows from (C.1). □

**Proof of Lemma 5:** From (20) and (26) we obtain

$$\begin{aligned}
\sqrt{N}\log\left(\frac{\tilde{\theta}(\alpha)}{\hat{\theta}(\alpha)}\right) &= \sqrt{N}\int_0^\alpha \frac{-1}{1-u}\left(\tilde{\gamma}T_t^{-1}[\hat{t}(u);\hat{\beta}]-\hat{\gamma}T_t^{-1}[\hat{t}(u);\beta]\right)du \\
&= \sqrt{N}\int_0^\alpha \frac{-\tilde{\gamma}}{1-u}\left(T_t^{-1}[\hat{t}(u);\hat{\beta}]-T_t^{-1}[\hat{t}(u);\beta]\right)du \\
&\quad -\int_0^\alpha \frac{1}{1-u}T_t^{-1}[\hat{t}(u);\beta]du\sqrt{N}\left(\tilde{\gamma}-\hat{\gamma}\right) \\
&= \int_0^\alpha \frac{-\gamma}{1-u}T_{t\beta}^{-1}[t(u);\beta]du\sqrt{N}(\hat{\beta}-\beta) \\
&\quad -\int_0^\alpha \frac{1}{1-u}T_t^{-1}[t(u);\beta]du\sqrt{N}\left(\tilde{\gamma}-\hat{\gamma}\right)+o_P(1) \\
&= \gamma I(\alpha)\sqrt{N}(\hat{\beta}-\beta)+o_P(1)
\end{aligned}$$

uniformly in $\alpha\in[0,\alpha_\dagger]$, where the third equality uses $\|\hat{t}(\cdot)-t(\cdot)\|_\dagger\xrightarrow{a.s.}0$ and a Taylor expansion around $\beta$, while the last equality uses (C.1) and (29). Hence,

$$\sqrt{N}[\tilde{\theta}(\cdot)-\hat{\theta}(\cdot)]=\gamma\theta(\cdot)I(\cdot)\sqrt{N}(\hat{\beta}-\beta)+o_P(1) \tag{C.2}$$

uniformly on $[0,\alpha_\dagger]$. The desired result follows.

Turning to $\tilde{\theta}_\alpha(\cdot)$, we have from (20) and (26)

$$\begin{aligned}
\sqrt{N}[\tilde{\theta}_\alpha(\alpha)-\hat{\theta}_\alpha(\alpha)] &= \sqrt{N}\left[\frac{\tilde{\theta}(\alpha)}{1-\alpha}\left(1-\tilde{\gamma}T_t^{-1}[\hat{t}(\alpha);\hat{\beta}]\right)-\frac{\hat{\theta}(\alpha)}{1-\alpha}\left(1-\hat{\gamma}T_t^{-1}[\hat{t}(\alpha);\beta]\right)\right] \\
&= \sqrt{N}\,\frac{\tilde{\theta}(\alpha)-\hat{\theta}(\alpha)}{1-\alpha}\left(1-\tilde{\gamma}T_t^{-1}[\hat{t}(\alpha);\hat{\beta}]\right) \\
&\quad -\frac{\hat{\theta}(\alpha)}{1-\alpha}\left[\sqrt{N}(\tilde{\gamma}-\hat{\gamma})T_t^{-1}[\hat{t}(\alpha);\hat{\beta}]+\hat{\gamma}\sqrt{N}\left(T_t^{-1}[\hat{t}(\alpha);\hat{\beta}]-T_t^{-1}[\hat{t}(\alpha);\beta]\right)\right] \\
&= \frac{1-\gamma T_t^{-1}[t(\alpha);\beta]}{1-\alpha}\sqrt{N}[\tilde{\theta}(\alpha)-\hat{\theta}(\alpha)] \\
&\quad -\frac{\theta(\alpha)}{1-\alpha}\left[T_t^{-1}[t(\alpha);\beta]\sqrt{N}(\tilde{\gamma}-\hat{\gamma})+\gamma T_{t\beta}^{-1}[t(\alpha);\beta]\sqrt{N}(\hat{\beta}-\beta)\right]+o_P(1)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\gamma\theta(\alpha)}{1-\alpha}\Big[\Big(1-\gamma T_t^{-1}[t(\alpha);\beta]\Big) I(\alpha)+\gamma T_t^{-1}[t(\alpha);\beta]T_{t\beta}^{-1}(\bar{t};\beta) \\
&\qquad -T_{t\beta}^{-1}[t(\alpha);\beta]\Big]\sqrt{N}(\hat{\beta}-\beta)+o_P(1)
\end{aligned}
$$

uniformly in $\alpha \in [0,\alpha_\dagger]$, where the third equality uses $\|\hat{t}(\cdot)-t(\cdot)\|_\dagger \xrightarrow{a.s.} 0$ and a Taylor expansion around $\beta$, while the last equality uses (C.1)-(C.2). Since $\theta_\alpha(\alpha)=\theta(\alpha)\left(1-\gamma T_t^{-1}[t(\alpha);\beta]\right)/(1-\alpha)$, $H(\theta)=1-\gamma/T_Q[Q(\theta)]=1-\gamma T_t^{-1}[T(Q(\theta);\beta);\beta]$, and $t(\alpha)=T[Q(\alpha);\beta]$, the desired result follows. □

**Proof of Proposition 4:** The proof follows that of Proposition 3. We begin with $\tilde{f}(\cdot)$. From van der Vaart and Wellner (1996) Lemma 2.9.23, the Hadamard derivative of the inverse map at $\theta(\cdot)$ is the map $h \mapsto -h[\theta^{-1}(\cdot)]/\theta_\alpha[\theta^{-1}(\cdot)]$. Thus, from the Functional Delta Method, we have

$$
\sqrt{N}[\tilde{\theta}^{-1}(\cdot)-\theta^{-1}(\cdot)] = -f^*(\cdot)\sqrt{N}[\tilde{\theta}(\theta^{-1}(\cdot))-\theta(\theta^{-1}(\cdot))]+o_P(1) \tag{C.3}
$$

uniformly on $[\theta^*,\theta_\dagger]$ since $f^*(\cdot)=1/\theta_\alpha[\theta^{-1}(\cdot)]$. Moreover, using (27) and a Taylor expansion, we obtain

$$
\sqrt{N}[\widetilde{f^*\circ}\theta(\cdot)-f^*\circ\theta(\cdot)] = -f^{*2}[\theta(\cdot)]\sqrt{N}[\tilde{\theta}_\alpha(\cdot)-\theta_\alpha(\cdot)]+o_P(1) \tag{C.4}
$$

uniformly on $[0,\alpha_\dagger]$ since $f^*[\theta(\cdot)]=1/\theta_\alpha(\cdot)$. Thus, using (28), the Hadamard derivative of the composition map $\phi[\theta^{-1},f^*\circ\theta]=(f^*\circ\theta)\circ\theta^{-1}$ in the proof of Proposition 3, and (C.3)-(C.4) we obtain from the Functional Delta Method

$$
\begin{aligned}
\sqrt{N}[\tilde{f}^*(\cdot)-f^*(\cdot)] &= -f^{*2}(\cdot)\sqrt{N}[\tilde{\theta}_\alpha(\theta^{-1}(\cdot))-\theta_\alpha(\theta^{-1}(\cdot))] \\
&\quad -f^*_\theta(\cdot)\sqrt{N}[\tilde{\theta}(\theta^{-1}(\cdot))-\theta(\theta^{-1}(\cdot))]+o_P(1)
\end{aligned}
$$

uniformly on $[\theta^*,\theta_\dagger]$ since $f^*(\cdot)=1/\theta_\alpha[\theta^{-1}(\cdot)]$. Note that the same equation holds with $\tilde{\theta}$ replaced by $\hat{\theta}$. Thus, taking the difference, using Lemma 5 and noting that $\theta^{-1}(\cdot)=F^*(\cdot)$ give

$$
\begin{aligned}
\sqrt{N}[\tilde{f}^*(\cdot)-f^*(\cdot)] &= \sqrt{N}[\hat{f}^*(\cdot)-f^*(\cdot)] \\
&\quad -\gamma f^*(\cdot)\left(I[F^*(\cdot)]-\frac{1}{H(\cdot)}a[F^*(\cdot)]\right)\sqrt{N}(\hat{\beta}-\beta) \\
&\quad -\gamma f^*_\theta(\cdot)\theta[F^*(\cdot)]I[F^*(\cdot)]\sqrt{N}(\hat{\beta}-\beta)+o_P(1).
\end{aligned}
$$

Collecting terms, the desired result follows from Proposition 3-(ii) and $\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{D} \mathcal{N}$.

Next, we turn to $\hat{U}_{0Q}(\cdot)=\widetilde{U_{0Q}\circ}Q[\hat{G}^{t*}(T(\cdot;\hat{\beta}))]$ by (28). From (21) and (27) we have

$$
\begin{aligned}
\sqrt{N}[\widetilde{U_{0Q}\circ}Q(\alpha)-\widehat{U_{0Q}\circ}Q(\alpha)] &= \sqrt{N}\left(\frac{1}{\tilde{\theta}(\alpha)T_t^{-1}[\hat{t}(\alpha);\hat{\beta}]}-\frac{1}{\hat{\theta}(\alpha)T_t^{-1}[\hat{t}(\alpha);\beta]}\right) \\
&= \frac{-1}{\theta(\alpha)T_t^{-1}[t(\alpha);\beta]}\left(\frac{\sqrt{N}[\tilde{\theta}(\alpha)-\hat{\theta}(\alpha)]}{\theta(\alpha)}\right.
\end{aligned}
$$

$$+\frac{\sqrt{N}\left(T_t^{-1}[\hat{t}(\alpha);\hat{\beta}]-T_t^{-1}[\hat{t}(\alpha);\beta]\right)}{T_t^{-1}[t(\alpha);\beta]}\Bigg)+o_P(1)$$

$$= \quad -U_{0Q}[Q(\alpha)]\left(\gamma I(\alpha)+\frac{T_{t\beta}^{-1}[t(\alpha);\beta]}{T_t^{-1}[t(\alpha);\beta]}\right)\sqrt{N}(\hat{\beta}-\beta)+o_P(1),$$

uniformly in $\alpha\in[0,\alpha_\dagger]$, where we have used the uniform consistency of $\hat{\beta}$, $\hat{t}(\cdot)$, $\hat{\theta}(\cdot)$ and $\tilde{\theta}(\cdot)$ for $\beta$, $t(\cdot)$ and $\theta(\cdot)$ on $[0,\alpha_\dagger]$ together with Lemma 5, $\sqrt{N}(\hat{\beta}-\beta)=O_P(1)$ and $\sqrt{N}[\tilde{\theta}(\cdot)-\hat{\theta}(\cdot)]=O_P(1)$. Thus, using (17) and $\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{D}\mathcal{N}$ we obtain

$$\begin{aligned}\sqrt{N}[\widetilde{U_{0Q}\circ}Q(\cdot)-U_{0Q}\circ Q(\cdot)] \quad &\Rightarrow \quad -U_{0Q}[Q(\cdot)]\Bigg[\gamma\mathcal{Z}(\cdot)+\frac{T_{QQ}[Q(\cdot)]}{T_Q[Q(\cdot)]}\frac{\mathcal{B}_{G^{Q*}}[Q(\cdot)]}{g^{Q*}[Q(\cdot)]}\\ &\qquad +\left(\gamma I(\cdot)+\frac{T_{t\beta}^{-1}[t(\cdot);\beta]}{T_t^{-1}[t(\cdot);\beta]}\right)\mathcal{N}\Bigg] \qquad \text{(C.5)}\end{aligned}$$

on $[0,\alpha_\dagger]$. Let $G^{t*}(\cdot)$ be the empirical cdf of $t_i, i=1,\ldots,N$. We have

$$\begin{aligned}\sqrt{N}\left(\hat{G}^{t*}[T(\cdot;\hat{\beta})]-G^{t*}[T(\cdot;\beta)]\right) \quad &= \quad \sqrt{N}\left(\hat{G}^{t*}[T(\cdot;\hat{\beta})]-G^{t*}[T(\cdot;\hat{\beta})]\right)\\ &\qquad +\sqrt{N}\left(G^{t*}[T(\cdot;\hat{\beta})]-G^{t*}[T(\cdot;\beta)]\right)\\ &= \quad \sqrt{N}\left(\hat{G}^{t*}[T(\cdot;\beta)]-G^{t*}[T(\cdot;\beta)]\right)\\ &\qquad +g^{t*}[T(\cdot;\beta)]T_\beta(\cdot;\beta)\sqrt{N}(\hat{\beta}-\beta)+o_P(1)\\ &\Rightarrow \quad \mathcal{B}_{G^{t*}}[T(\cdot;\beta)]+g^{t*}[T(\cdot;\beta)]T_\beta(\cdot;\beta)\,\mathcal{N} \qquad \text{(C.6)}\end{aligned}$$

on $[\underline{Q},Q_\dagger]=[Q(0),Q(\alpha_\dagger)]$, where the second equality follows from asymptotic equicontinuity of $\sqrt{N}[\hat{G}^{t*}(\cdot)-G^{t*}(\cdot)]$ by Theorem 18.14-(ii) in van der Vaart (1998). We now consider the composition map $\phi\{G^{t*}[T(\cdot;\beta)],U_{0Q}[Q(\cdot)]\}=U_{0Q}[Q(\cdot)]\circ G^{t*}[T(\cdot;\beta)]$ with its Hadamard derivative

$$\phi'_{G^{t*}[T(\cdot;\beta)],U_{0Q}[Q(\cdot)]}(h_1,h_2)(\cdot)=h_2\left(G^{Q*}(\cdot)\right)+\frac{U_{0QQ}(\cdot)}{g^{Q*}(\cdot)}h_1(\cdot)$$

by Lemma 3.9.27 in van der Vaart and Wellner (1996) since $t=T(Q;\beta)$ so that $G^{t*}[T(\cdot;\beta)]=G^{Q*}(\cdot)$ and $U_{0Q}[Q(\cdot)]'_{G^{Q*}(\cdot)}[h_1(\cdot)]=U_{0QQ}(\cdot)Q_\alpha[G^{Q*}(\cdot)]h_1(\cdot)$ with $Q_\alpha[G^{Q*}(\cdot)]=1/g^{Q*}(\cdot)$. Thus, from (C.5)-(C.6) and the Functional Delta Method we obtain

$$\begin{aligned}\sqrt{N}[\tilde{U}_{0Q}(\cdot)-U_{0Q}(\cdot)] \quad &\Rightarrow \quad -U_{0Q}(\cdot)\Bigg[\gamma\mathcal{Z}[G^{Q*}(\cdot)]+\frac{T_{QQ}(\cdot)}{T_Q(\cdot)}\frac{\mathcal{B}_{G^{Q*}}(\cdot)}{g^{Q*}(\cdot)}\\ &\qquad +\left(\gamma I[G^{Q*}(\cdot)]+\frac{T_{t\beta}^{-1}[T(\cdot;\beta);\beta]}{T_t^{-1}[T(\cdot;\beta);\beta]}\right)\mathcal{N}\Bigg]\\ &\qquad +\frac{U_{0QQ}(\cdot)}{g^{Q*}(\cdot)}\left(\mathcal{B}_{G^{Q*}}(\cdot)+g^{t*}[T(\cdot;\beta)]T_\beta(\cdot;\beta)\,\mathcal{N}\right)\end{aligned}$$

since $G^{Q*}(\cdot)=G^{t*}[T(\cdot;\beta)]$. Collecting terms and using (B.5), $g^{Q*}(\cdot)=g^{t*}[T(\cdot;\beta)]/T_t^{-1}[T(\cdot;\beta);\beta]$, and $T_\beta(\cdot;\beta)=-T_\beta^{-1}[T(\cdot;\beta);\beta]/T_t^{-1}[T(\cdot;\beta);\beta]$, give the desired result.□

Herefater we assume that $\sqrt{N}(\hat{\beta}-\beta)$ satisfies the linear representation

$$\sqrt{N}(\hat{\beta}-\beta) = -\overline{\Gamma}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} V(q_i,\epsilon_i) + o_p(1), \tag{C.7}$$

for some matrix $\overline{\Gamma}$ and $(\dim\beta \times 1)$ vector function $V(\cdot,\cdot)$. See e.g. Linton, Sperlich and Van Keilegom (2008). The next lemma derives the correlation between $\mathcal{B}_{G^{t*}}(\cdot)$ and $\mathcal{Z}(\cdot)$ with $\mathcal{N}$.

**Lemma C.1:** *Let $G^{t*}(\cdot)$ be the cdf of $t$. We have*

$$\begin{aligned}
\mathrm{E}\left[\mathcal{B}_{G^{t*}}(\cdot)\mathcal{N}\right] &= -\overline{\Gamma}^{-1}\mathrm{E}\left[1\!\!I(t\le\cdot)V(q,\epsilon)\right],\\
\mathrm{E}\left[\mathcal{Z}(\cdot)\mathcal{N}\right] &= -\overline{\Gamma}^{-1}\int_{\underline{t}}^{t(\cdot)} T_{tt}^{-1}(\tau;\beta)\frac{\mathrm{E}\left[1\!\!I(t\le\tau)V(q,\epsilon)\right]}{1-G^{t*}(\tau)}d\tau,
\end{aligned}$$

*where $\mathcal{B}_{G^{t*}}(\cdot)$ is the $G^{t*}$-Brownian Bridge on $[\underline{t},\overline{t}]$ associated with $\{t_i; i=1,\dots,N\}$.*

**Proof of Lemma C.1:** Recall that $\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{D}\mathcal{N}\sim\mathcal{N}(0,\Omega)$. Moreover, by the Functional Central Limit Theorem, we have

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}1\!\!I(t_i\le\cdot)-G^{t*}(\cdot)\right)\Rightarrow\mathcal{B}_{G^{t*}}(\cdot) \tag{C.8}$$

on $[\underline{t},\overline{t}]$. Thus, the first statement follows from (C.7) and $\mathrm{E}[V(q,\epsilon)]=0$.

Turning to the second statement, the change-of-variable $\tau=T(q;\beta)$ in (15) gives

$$\mathcal{Z}(\cdot)=\int_{\underline{t}}^{t(\cdot)} T_{tt}^{-1}(\tau;\beta)\frac{\mathcal{B}_{G^{t*}}(\tau)}{1-G^{t*}(\tau)}d\tau \tag{C.9}$$

since $dq=T_t^{-1}(\tau;\beta)d\tau$, $T_Q(q)=1/T_t^{-1}(\tau;\beta)$, $T_{QQ}(q)=-T_{tt}^{-1}(\tau;\beta)/T_t^{-1}(\tau;\beta)^3$, $G^{Q*}(q)=G^{t*}(\tau)$ and $g^{Q*}(q)=g^{t*}(\tau)/T^{-1}(\tau;\beta)$. The desired result follows from (C.7), (C.8), and $\mathrm{E}[V(q,\epsilon)]=0.\Box$

**Computation of $\omega^2_{f^*}(\theta)$ and $\omega^2_{U_{0Q}}(Q)$:** We have $F^*(\theta)=G^{t*}[T(Q(\theta);\beta)]$ since $t=T(Q(\theta);\beta)$. In particular, $\mathcal{B}_{F^*}(\theta)=\mathcal{B}_{G^{t*}}[T(Q(\theta);\beta)]$. Thus, from $\mathcal{N}\sim\mathcal{N}(0,\Omega)$, Propositions 3 and 4 and Lemma C.1 we obtain

$$\begin{aligned}
\omega^2_{f^*}(\theta) &= V_{f*}(\theta)+2\gamma^2[f^*(\theta)+\theta f^*_\theta(\theta)]b(\theta)\mathrm{E}\left\{\mathcal{Z}\Big[G^{t*}[T(Q(\theta);\beta)]\Big]\mathcal{N}\right\}\\
&\quad -2\gamma\frac{H_\theta(\theta)}{H(\theta)}b(\theta)\mathrm{E}\left\{\mathcal{B}_{G^{t*}}[T(Q(\theta);\beta)]\mathcal{N}\right\}+\gamma^2 b(\theta)\Omega b(\theta)^T\\
&= V_{f*}(\theta)-2\gamma^2[f^*(\theta)+\theta f^*_\theta(\theta)]b(\theta)\overline{\Gamma}^{-1}\int_{\underline{t}}^{T[Q(\theta);\beta]} T_{tt}^{-1}(\tau;\beta)\frac{E\left[1\!\!I(t\le\tau)V(q,\epsilon)\right]}{1-G^{t*}(\tau)}d\tau\\
&\quad -2\gamma\frac{H_\theta(\theta)}{H(\theta)}b(\theta)\overline{\Gamma}^{-1}\mathrm{E}\Big\{1\!\!I\Big[t\le T[Q(\theta);\beta]\Big]V(q,\epsilon)\Big\}+\gamma^2 b(\theta)\Omega b(\theta)^T,
\end{aligned} \tag{C.10}$$

where $V_{f^*}(\theta)$ is given in (B.7).

Similarly, we have $G^{Q*}(Q) = G^{t*}[T(Q;\beta)]$ since $t = T(Q;\beta)$. In particular, $\mathcal{B}_{G^*}(Q) = \mathcal{B}_{G^{t*}}[T(Q;\beta)]$. Thus, from $\mathcal{N} \sim \mathcal{N}(0,\Omega)$, Propositions 3 and 4 and Lemma C.1, we obtain

$$\begin{aligned}
\omega^2_{U_{0Q}}(Q) &= V_{U_{0Q}}(Q) + U_{0Q}(Q)^2\Big[2\gamma c(Q)\mathrm{E}\Big\{\mathcal{Z}\Big[G^{t*}[T(Q;\beta)]\Big]\mathcal{N}\Big\} \\
&\qquad +2\frac{T_Q(Q;\beta)-\gamma}{T_Q(Q;\beta)}\mathrm{E}\left\{\frac{\mathcal{B}_{G^{t*}}[T(Q;\beta)]\mathcal{N}}{1-G^{t*}[T(Q;\beta)]}\right\} + c(Q)\Omega c(Q)^T\Bigg] \\
&= V_{U_{0Q}}(Q) + U_{0Q}(Q)^2\left[-2\gamma c(Q)\overline{\Gamma}^{-1}\int_{\underline{t}}^{T(Q;\beta)} T_{tt}^{-1}(\tau;\beta)\frac{\mathrm{E}\left[1\!\!1(t\le\tau)V(q,\epsilon)\right]}{1-G^{t*}(\tau)}d\tau\right. \\
&\qquad \left. -2\frac{T_Q(Q;\beta)-\gamma}{T_Q(Q;\beta)}\overline{\Gamma}^{-1}\mathrm{E}\left\{\frac{1\!\!1[t\le T(Q;\beta)]V(q,\epsilon)}{1-G^{t*}[T(Q;\beta)]}\right\} + c(Q)\Omega c(Q)^T\right], \quad \text{(C.11)}
\end{aligned}$$

where $V_{U_{0Q}}(Q)$ is given in (B.10).□

**Estimation of $\omega^2_{f^*}(\theta)$ and $\omega^2_{U_{0Q}}(Q)$:** Given that the observables are $(t_i, q_i)$, estimation of $V_{f^*}(\theta)$ and $V_{U_{0Q}}(Q)$ need to be adjusted accordingly. In particular, the argument in Appendix B applies by noting that $T_Q(Q) = 1/T_t^{-1}(t)$, $T_{QQ}(Q) = -T_{tt}^{-1}(t)/T_t^{-1}(t)^2$, $G^{Q*}(Q) = G^{t*}(t)$ and $g^{Q*}(Q) = g^{t*}(t)/T^{-1}(t)$. In addition, the indicators $1\!\!1[Q_i \le Q]$ and $1\!\!1[Q_i \le \hat{Q}(\theta)]$ are now replaced by $1\!\!1[\hat{T}^{-1}(t_i) \le Q]$ and $1\!\!1[t_i \le \hat{T}(\theta)]$ with $\hat{T}(\cdot) = G^{t*-1}[\hat{\theta}^{-1}(\cdot)]$, respectively.

We assume that we have consistent estimators $\hat{V}(\cdot,\cdot)$, $\hat{\overline{\Gamma}}$ and $\hat{\Omega}$. See e.g. Linton, Sperlich and Van Keilegom (2008). Regarding estimation of $b(\cdot)$ and $c(\cdot)$, the only term that requires attention is $I(\cdot)$. We note that $I(\alpha)$ for $\alpha \in [0, \alpha_\dagger]$ can be written as an expectation upon making the change of variable $t = t(u)$ thereby leading to the estimator

$$\hat{I}(\alpha) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1\!\!1[t_i \le \hat{t}(\alpha)]}{1-\hat{G}^{t*}(t_i)}\left\{T_{t\beta}^{-1}(t_i;\hat{\beta}) - \hat{\gamma}T_t^{-1}(t_i;\hat{\beta})T_{t\beta}^{-1}(t_{\max};\hat{\beta})\right\}\right].$$

In $c(\cdot)$, the term $U_{0QQ}(Q)/U_{0Q}(Q)$ can be rewritten using (B.5) and expressed in term of $T^{-1}(\cdot;\beta)$ and its derivatives as above. We remark that the expectation $\mathrm{E}[1\!\!1(t \le \tau)V(q,\epsilon)]$ in (C.10) and (C.11) can be estimated by $(1/N)\sum_{i=1}^{N} 1\!\!1[t_i \le \tau]\hat{V}(q_i,\hat{\epsilon}_i)$ where $\hat{\epsilon}_i = T^{-1}(t_i;\hat{\beta})/\hat{\hat{m}}(q_i)$. Lastly, $T[Q(\cdot);\beta]$ can be replaced by $T(\cdot) = G^{t*-1}[\theta^{-1}(\cdot)]$ and estimated accordingly.

## References


**Andrews, D.** (1994): "Empirical Process Methods in Econometrics," in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics, Volume IV*, North Holland.

**Armstrong, M.** (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51-76.

**Armstrong, M.** and **J. Vickers** (2001): "Competitive Price Discrimination," *Rand Journal of Economics*, 32, 579-605.

**Athey, S.** and **P. Haile** (2007): "Nonparametric Approaches to Auctions," in J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Volume VI*, North Holland.

**Billingsley, D.** (1968): *Convergence of Probability Measures*, John Wiley and Sons.

**Borenstein, S.** (1991): "Selling Costs and Switching Costs: Explaining Retail Gasoline Margins," *Rand Journal of Economics*, 22, 354-369.

**Borenstein, S.** and **N. Rose** (1994): "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, 102, 653-683.

**Busse, M.** and **M. Rysman** (2005): "Competition and Price Discrimination in Yellow Pages Advertising," *Rand Journal of Economics*, 36, 378-390.

**Campo, S.**, **E. Guerre**, **I. Perrigne** and **Q. Vuong** (2011): "Semiparametric Estimation of First-Price Auctions with Risk Averse Bidders," *Review of Economic Studies*, 78, 112-147.

**Carroll, R.J.** and **D. Ruppert** (1988): *Transformation and Weighting in Regression*, Chapman and Hall.

**Chen, S.** (2002): "Rank Estimation of Transformation Models," *Econometrica*, 70, 1683-1697.

**Chiappori, P.A.**, **I. Komunjer** and **D. Kristensen** (2013): "Identification and Estimation of Nonparametric Transformation Models," Working Paper, University of California San Diego.

**Cohen, A.** (2008): "Package Size and Price Discrimination in Paper Towels," *International Journal of Industrial Organization*, 26, 502-516.

**Crawford, G.S.** and **M. Shum** (2007): "Monopoly Quality Choice in Cable Television," *Journal of Law and Economics*, 50, 181-209.

**D'Haultfoeuille, X.** and **P. Février** (2007): "Identification and Estimation of Incentive Problems: Adverse Selection," Working Paper, CREST-INSEE.

**D'Haultfoeuille, X.** and **P. Février** (2011): "The Provision of Wage Incentives: A Structural Estimation Using Contract Variation," Working Paper, CREST-INSEE.

**Economides, N.**, **K. Seim** and **V.B. Viard** (2008): "Quantifying the Benefits of Entry into Local Phone Service," *Rand Journal of Economics*, 39, 699-730.

**Ekeland, I.**, **J. Heckman** and **L. Neishem** (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112, 60-109.

**Galambos, J.** (1978): *The Asymptotic Theory of Extreme Order Statistics,* Wiley.

**Grubb, M.** and **M. Osborne** (2014): "Cellular Service Demand: Biased Beliefs, Learning and Bill Shock," *American Economic Review*, forthcoming.

**Guerre, E.**, **I. Perrigne** and **Q. Vuong** (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525-574.

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

**Heckman, J.**, **R. Matzkin** and **L. Neishem** (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78, 1569-1591.

**Horowitz, J.** (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," *Econometrica*, 64, 103-137.

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

**Khan, S.**, **Y. Shin** and **E. Tamer** (2011): "Heteroskedastic Transformation Models with Covariate Dependent Censoring," *Journal of Business Economics and Statistics*, 29, 40-48.

**Laffont, J.J.** and **Q. Vuong** (1996): "Structural Analysis of Auction Data," *American Economic Review, Papers and Proceedings*, 86, 414-420.

**Leslie, P.** (2004): "Price Discrimination in Broadway Theatre," *Rand Journal of Economics*, 35, 520-541.

**Lewis, T.** and **D. Sappington** (1989): "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, 49, 294-313.

**Linton, O.**, **S. Sperlich** and **I. Van Keilegon** (2008): "Estimation of a Semiparametric Transformation Model," *The Annals of Statistics*, 36, 686-718.

**Lott, J.** and **R. Roberts** (1991): "A Guide to the Pitfalls of Identifying Price Discrimination," *Economic Inquiry*, 29, 14-23.

**Luo, Y.** (2012): "Bundling and Nonlinear Pricing in Telecommunications," Working Paper, Pennsylvania State University.

**Luo, Y.**, **I. Perrigne** and **Q. Vuong** (2012): "Multiproduct Nonlinear Pricing: Mobile Voice Service and SMS," Working Paper, Pennsylvania State University.

**Luo, Y.**, **I. Perrigne** and **Q. Vuong** (2013): "A General Framework for Nonlinear Pricing Data," Working Paper, New York University.

**Marmer, V.** and **A. Shneyerov** (2012): "Quantile-Based Nonparametric Inference for First-Price Auctions," *Journal of Econometrics*, 167, 345-357.

**Maskin, E.** and **J. Riley** (1984): "Monopoly with Incomplete Information," *The Rand Journal of Economics*, 15, 171-196.

**Maggi, G.** and **A. Rodriguez-Clare** (1995): "On Countervailing Incentives," *Journal of Economic Theory*, 66, 238-263.

**McManus, B.** (2007): "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee," *Rand Journal of Economics*, 38, 512-532.

**Miravete, E.** (2002): "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans," *Review of Economic Studies*, 69, 943-971.

**Miravete, E.** and **L.H. Roller** (2004): "Competitive Nonlinear Pricing in Duopoly Equilibrium: The Early U.S. Cellular Telephone Industry," Working Paper , University of Pennsylvania.

**Mussa, M.** and **S. Rosen** (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301-317.

**Oren, S.**, **S. Smith** and **R. Wilson** (1983): "Competitive Nonlinear Tariffs," *Journal of Economic Theory*, 29, 49-71.

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

**Pollard, D.** (1984): *Convergence of Stochastic Processes*, Springer Verlag.

**Riley, J.** (2012): *Essential Microeconomics*, Cambridge University Press.

**Rochet, J.C.** and **P. Chone** (1998): "Ironing, Sweeping and Multidimensional Screening," *Econometrica*, 66, 783-826.

**Rochet, J.C.** and **L.A. Stole** (2003): "The Economics of Multidimensional Screening," in Dewatripont, M., L.P. Hansen, and S.J. Turnovsky, eds, *Advances in Economic Theory: Eighth World Congress*, Cambridge University Press.

**Shepard, A.** (1991): "Price Discrimination and Retail Configuration," *Journal of Political Economy*, 99, 30-53.

**Spence, A.M.** (1977): "Nonlinear Prices and Welfare," *Journal of Public Economics*, 1, 1-18.

**Stole, L.A.** (1995): "Nonlinear Pricing and Oligopoly," *Journal of Economic Management and Strategy*, 4, 529-562.

**Stole, L.A** (2007): "Price Discrimination in Competitive Environment," in M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization, Volume III*, North Holland.

**Tirole, J.** (1988): *The Theory of Industrial Organization*, MIT Press.

**van der Vaart, A.** (1998): *Asymptotic Statistics*, Cambridge University Press.

**van der Vaart, A.** and **J. Wellner** (1996): *Weak Convergence and Empirical Processes*, Springer.

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

**Wong, A.** (2012): " Menu Implementability and Practical Pricing Schemes," Working Paper, Shangai University of Finance and Economics.

**Zhou, X.H.**, **H. Lin** and **E. Johnson** (2008): "Nonparametric Heteroscedastic Transformation Regression Models for Skewed Data with an Application to Health Care Costs," *Journal of the Royal Statistical Society: Series B*, 70, 1029-1047.

Table 1: Summary statistics

| | Observations | Mean | Median | Min | Max | STD |
|---|---|---|---|---|---|---|
| $t$ | 4,000 | 34.63 | 30.37 | 13.97 | 99.78 | 16.52 |
| $q$ | 4,000 | 704.32 | 620.00 | 10 | 3,498.00 | 427.88 |
| $Q$ | 4,000 | 1,650.15 | 1,087.00 | 195.20 | 9,171.03 | 1,561.51 |
| $\hat{\theta}$ | 4,000 | 1.78 | 1.56 | 1.00 | 5.03 | 0.74 |
| $\hat{\epsilon}$ | 4,000 | 1.19 | 0.94 | 0.13 | 15.74 | 0.96 |
| Rent | 4,000 | 21.92 | 12.40 | 0 | 160.16 | 26.12 |
| Rent Ratio | 4,000 | 0.47 | 0.41 | 0 | 1.63 | 0.35 |

Table 2: Comparisons with Alternative Pricing Strategies

| Indicators | NLP | Two-part Tariffs | Minimal Quantities | Plans |
|---|---|---|---|---|
| Consumer Surplus | 87,688 | 0.9973 | 0.8824 | 0.9653 |
| Profit | 83,907 | 0.9926 | 0.9575 | 0.9433 |
| Welfare | 171,590 | 0.9950 | 0.9191 | 0.9546 |
| Total $Q$ | 6,600,600 | 0.9875 | 0.8975 | 0.9564 |
| Total Payment | 138,540 | 0.9875 | 0.9258 | 0.9519 |
| Consumers | 4,000 | 0.9505 | 0.8023 | 0.9963 |

Table 3: Winners and Losers

| Group | Variable | NLP | Two-part Tariffs | Minimal Quantities | Plans |
|---|---|---|---|---|---|
| Low Types | CS | 1,742 | 0.8601 | 0.0810 | 1.1506 |
| | Total Payment | 18,686 | 0.9934 | 0.3364 | 1.0803 |
| | Total $Q$ | 418,360 | 1.2248 | 0.4599 | 1.2295 |
| Medium low types | CS | 7,748 | 1.0097 | 0.7005 | 0.9778 |
| | Total Payment | 25,968 | 1.0005 | 1.1583 | 0.7892 |
| | Total $Q$ | 807,440 | 1.0024 | 1.1401 | 0.6468 |
| Medium High Types | CS | 19,763 | 0.9944 | 0.8452 | 0.9022 |
| | Total Payment | 35,746 | 1.0224 | 0.9800 | 1.0544 |
| | Total $Q$ | 1,524,900 | 1.0732 | 0.9076 | 1.2930 |
| High Types | CS | 58,435 | 1.008 | 0.9430 | 0.9794 |
| | Total Payment | 58,139 | 0.9583 | 0.9780 | 0.9202 |
| | Total $Q$ | 3,849,800 | 0.9248 | 0.8902 | 0.8583 |

Figure 1: Scatter plot $(q_i, t_i)$

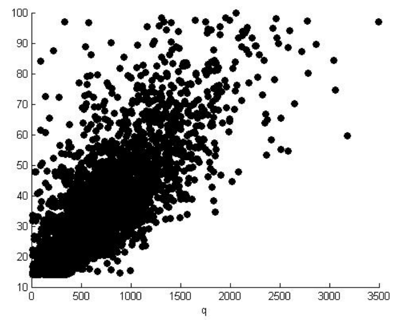


Figure 2: Tariff $\hat{T}(\cdot)$

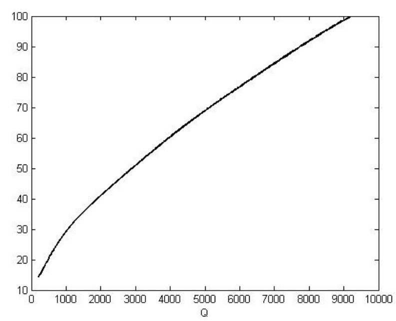


Figure 3: Density of Unobserved Heterogeneity $\hat{f}_\epsilon(\cdot)$

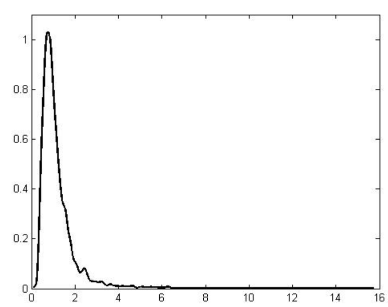

Figure 4: Quantile Type $\hat{\theta}(\cdot)$

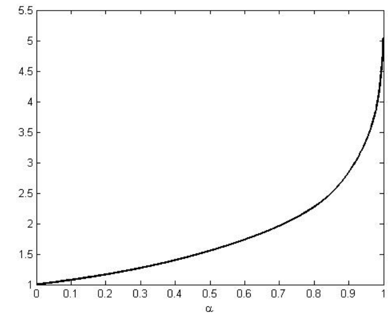


Figure 5: Marginal Base Utility $\hat{U}_{0Q}(\cdot)$

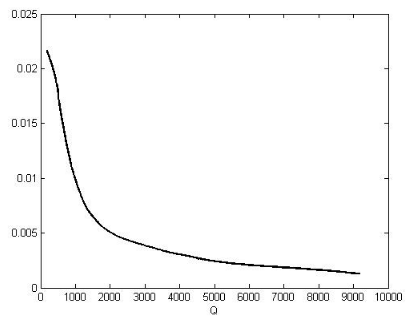


Figure 6: Type Density $\hat{f}^{*}(\cdot)$

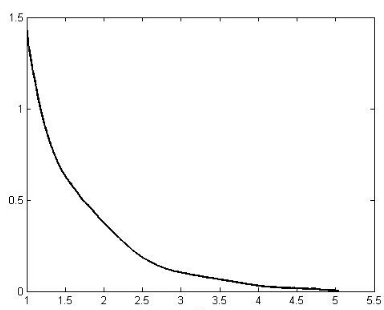

Figure 7: $\theta - [1 - \hat{F}(\theta)]/\hat{f}(\theta)$