DISCUSSION PAPER SERIES

No. 2699

QUANTITY DISCOUNTS FOR TIME-VARYING CONSUMERS

Eugenio J Miravete

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

ISSN 0265-8003

QUANTITY DISCOUNTS FOR TIME-VARYING CONSUMERS

Eugenio J Miravete, University of Pennsylvania and CEPR

Discussion Paper No. 2699
February 2001

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 2699

February 2001

ABSTRACT

Quantity Discounts for Time-Varying Consumers*

When a monopolist asks consumers to choose a particular non-linear tariff option, consumers do not completely know their type. Their valuations of the good and/or optimal quantity purchases are only fully realized after the optional tariff has been subscribed. In order to characterize the menu of optimal non-linear tariffs when consumers' demands are stochastic, I show that the increasing hazard rate property of distributions is preserved under convolution. This result, together with very weak assumptions on demand (common to standard non-linear pricing), ensures that the continuum of optional non-linear tariffs is characterized by quantity discounts. I test non-parametrically the theoretical prerequisites of the model using data directly linked to consumer types from the 1986 Kentucky Telephone Tariff experiment. I show that the distribution of actual calls second order stochastically dominates the distribution of expected calls, which fully supports the suggested type-varying theoretical model. Finally, I analyse possible welfare effects of the introduction of optional tariffs and their relative expected profitability using the empirical distribution of consumer types in two local exchanges with differentiated calling patterns. The evidence suggests that a menu of optional two-part tariffs dominates any other pricing strategy.

JEL Classification: D42, D82, L96
Keywords: convolution distributions, increasing hazard rate, optional non-linear pricing, quantity discounts, stochastic dominance

Eugenio J Miravete
Department of Economics
University of Pennsylvania
McNeil Building / 3718 Locust Walk
Philadelphia, PA 19104-62971
USA
Tel: (1 215) 898 1505
Fax: (1 215) 573 2057
Email: miravete@ssc.upenn.edu

*I thank seminar participants for their comments at Barcelona (Autónoma), Carlos III, New York, Pennsylvania, Pompeu Fabra, Princeton, Rutgers, Stern School, Wisconsin-Madison and Wissenschaftszentrum Berlin für Sozialforschung, as well as those attending the 1999 Workshop on Empirical Industrial Organization in Vigo, the 2000 Workshop on Applied Microeconomics and Econometrics at Johns Hopkins University, and the 2000 Winter Meeting of the Econometric Society in Boston. I am grateful to José Pernías for speeding up my simulation routines. Partial funding from the Ameritech Foundation through the Consortium for Research on Telecommunications Strategy and Policy is gratefully acknowledged.

Submitted 12 December 2000

NON-TECHNICAL SUMMARY

Price discrimination may take multiple expressions. A common business practice in many industries requires that consumers first sign up for a contract that specifies some rates, fixed monthly payments, and perhaps some 'free' consumption allowance per month. This type of contract is currently very common in the mobile phone industry.

Why are these tariffs so common? Why do firms generally find it profitable to offer different tariff options to consumers? Do they make additional profits because consumers cannot commit to future levels of consumption when they subscribe to a particular tariff option?

The standard alternative is to offer consumers a single tariff that will determine total payments depending on each customer's own consumption and the discounts included in the tariff. In the current practice in the mobile phone industry, this is the option that does not require any further contract obligation. Consumers are not required to pay a monthly fee, but per minute rates are by far the most expensive alternative. Price discrimination makes use of all these devices to extract the maximum possible amount of consumer surplus from individuals with different intensity of preferences. Consumers have different types depending on their willingness to pay for a good or service.

In this Paper I distinguish between *ex ante* and *ex post* type. The *ex post* type determines the consumption level. Since consumers are not fully aware of their future consumption when they sign up for a particular tariff option, I assume that they have a prior (*ex ante*) type. Between the choice of the tariff and the consumption decision any remaining individual uncertainty is resolved, and thus the *ex ante* and *ex post* type differs by some type shock.

When firms decide to offer a single tariff to discriminate among consumers they normally offer volume discounts on realized demand. This tariff takes into account the distribution of the *ex post* types. Here, the choice of consumption and the determination of payment are simultaneous. This tariff will be characterized by quantity discounts under fairly general assumptions, including that the distribution of *ex post* types should be increasing hazard rate (IHR).

If the firm decides to offer tariff options, it screens consumers only with respect to *ex ante* types. If these tariff options are two-part tariffs, consumers keep the informational rents related to type shocks. If tariff options are non-linear, however, firms introduce further incentives aimed at reducing these informational rents.

This Paper offers a general solution for the *ex post* and *ex ante* non-linear tariffs. It also shows that no additional assumption is needed for these tariffs to be characterized by quantity discounts since IHR is preserved under convolution. Finally, the Paper makes use of an exceptional data set that includes direct instruments related to the types of consumers to validate the theoretical prerequisites of the model, and to empirically evaluate the profit and welfare implications of the introduction of optional tariffs.

1 Introduction

Consumers have to choose frequently among sets of class of services. For instance, telephone customers have to choose among different long distance plans offered by competing firms, or among different subscription contracts to the local telephone monopolist. Internet access providers also allow choosing among different connection plans depending on the expected usage of the network. Cable companies offer a variety of channel options for monthly subscription at different rates and bundling discounts. Car rental rates depend on the duration of the lease, mileage, and/or fuel option chosen. Public transportation systems offer the possibility of advance purchase of passes of varied duration at different discount rates depending on the expected usage of the system. Banks ask their customers to select one among few checking and savings accounts depending on their average expected balance and number of monthly checks drawn. Finally, health clubs charge different monthly rates depending on registration fees related to the duration of the contract.

What do all these examples have in common? All these situations are characterized by a two–stage decision process: first consumers decide which class of service they sign up for, and later, once their demand needs are known with certainty, they decide how much to buy from the firm, contingent on the rates of the tariff plan previously chosen. Therefore, consumers are not signing a contingent contract, and they are not committing to any particular purchase level in the future while the choice among alternatives does not embody any attempt to minimize risk. Companies, either because of reputation, repeated interaction with consumers, or legal restrictions, are not allowed to switch customers from one class of service to a different one, neither to take advantage of customers consumption decisions, or to favor them. Thus, firms can only profit from the stochastic dimension of consumers' demand through the design of the offered options.

But furthermore, most business applications of nonlinear pricing are characterized by quantity discounts, *i.e.*, unit price decreases with volume purchased by consumers. This feature is particularly convenient for natural monopolists and firms with important fixed costs. Charging a higher price per unit for the first units sold allows covering fixed costs, while discounts increase efficiency as large customers are priced closer to marginal costs. Fortunately, very general demand specifications and consumer taste distributions lead to optimal nonlinear price schedules characterized by quantity discounts. At least since the work of Maskin and Riley (1984), it is well known that quantity discounts are present if the distribution of types is increasing hazard rate, if consumers utility function satisfies the single–crossing property, and if some (not very restrictive) third derivatives of the utility function have the appropriate sign. Thus, general concave pricing mechanisms lead to quantity discounts. In fact, when the tariff function is concave, it can be implemented using a menu or continuum of self–selecting two–part tariffs whose marginal charge is decreasing with the volume purchased by the consumer [Faulhaber and Panzar (1977)]. An important result of the paper is to show that no additional assumptions on demand need to be made, and that in fact it suffices to assume that each distribution of taste components are increasing hazard rate to ensure that the corresponding optional tariffs are characterized by quantity discounts.

Why cannot the above examples be addressed with the existing nonlinear pricing theory? The concept of self–selecting tariff has been incorrectly used as synonym of optional tariffs. This is particularly true in many works dealing with pricing of telecommunications services. A common mistake present in all the related empirical literature is to neglect the existence of two stages and assume that consumers make purchases and choose among class of services simultaneously.[1] If this were the case, the only relevant information for consumers to make that decision would be known at the time of consumption, and therefore the "choice" of the corresponding self–selecting tariff plan would be exactly dual to the usage decision. Obviously, within this framework, there is no possibility of an *ex–post* "mistake" in the choice of the tariff plan.

A correct approach should explicitly account for this two–stage nature of the problem: consumers first choose the tariff plan that better suit their needs given their expectation on future consumption level or expected valuation of the good or service. Later, once their needs or actual valuation are known, they decide how much to consume contingent on their previous tariff choice. The difficulty of this approach is that individual consumers' demands become stochastic since the expected consumption at the time of the tariff choice need not necessarily coincide with the purchase in the second stage of this game. Individual stochastic demands break the duality between consumption and choice of the corresponding self–selecting tariff. Consumers who chose different tariffs in stage 1 may end up paying and consuming the same at stage 2 if they receive demand shocks of opposite sign. Similarly, the same consumption level at stage 2 could be purchased at different rates depending on the disparate choice of tariffs of different consumers at stage 1. Thus, the optimal nonlinear tariff is no longer the lower envelope of a set of self–selecting two–part tariffs, and screening consumers becomes a more difficult task.

General treatments of nonlinear pricing when demand is stochastic are still not available.[2] This paper provides a characterization of the *ex–ante* nonlinear tariff (the one that considers two–stage decision problem) and relates it to the standard *ex–post* nonlinear tariff (where consumption and tariff choice are simultaneous). The key issue for these two pricing problems to be properly defined is that the hazard rate properties of the distributions of consumer types is preserved under convolution. This paper shows that this is the case under very general conditions, and furthermore that there might be some ordering of the hazard rates of the distributions used in each problem, *ex–ante vs. ex–post* that leads to unambiguous ranking of markups for every purchase level possible.

---

[1] See for instance Hobson and Spady (1988), Kling and van der Ploeg (1990), MacKie–Mason and Lawson (1993), and Mitchel and Vogelsang (1991, §8).

[2] Only recently this topic has attracted some attention, although few and incomplete attempts to model optional tariffs have been carried out. See for instance Clay, Sibley, and Srinagesh (1992), Courty and Li (2000), and Miravete (1996 and 2000b).

In order to deal with the stochastic nature of consumer demand, the suggested model assumes that consumers' types have two components: the *ex–ante* type θ_1, and the type shock θ_2. Together they define the *ex–post* type θ that drives purchase decisions. The *ex–ante* type is always known by consumers, and it determines the choice of the class of service. This type dimension is private information and defines something similar to the average consumption level for each consumer (or expected valuation of the product). Type shock θ_2 represents deviations from the average consumption due to unpredictable events (or unexpected changes in valuation due to any general or individual circumstances). The type shock is different for each individual and remains private information to each consumer. The monopolist will design each tariff option, and within each option the corresponding quantity discounts, to maximize his expected profits given the information set of consumers at each stage. The realization of θ_1 critically conditions the choice among tariffs, while the value of θ_2 together with the tariff plan chosen determines the actual level of usage in the second stage of the game. The paper studies first the design of optional two–part tariffs that screen only with respect to θ_1, and later the design of fully nonlinear options that screen all components of consumer types separately.

This type–varying setup rises many side issues that are intentionally neglected in the present paper to avoid unnecessary complexity and to limit the scope of the model in dealing with the effects of stochastic demands. I state two explicit simplifying assumptions here. First, the pricing game remains essentially static. Consumers first choose the optional tariff, and later decide how much to consume. I do not consider repeated versions of this game –as in Baron and Besanko (1984)– because it would require to model how informative is θ_1 with respect to θ, how are successive θ_2's correlated over time, and ultimately model the updating of future usage expectations, θ_1's. The closed–loop equilibrium tariff options of this richer model will be more difficult to characterize (if not impossible analytically), and this added complexity will however not help answering whether optional tariffs could be properly defined or not. Second, types remain single–dimensional. Thus, θ "moves around" θ_1 depending on the magnitude and sign of the type shock θ_2 and the single–dimensional definition of the *ex–post* type as a function of the *ex–ante* type and the shock. Additional dimensions should only be considered if they address different attributes in the definition of consumers' utility functions, so that the monopolist can screen consumers different taste dimensions *simultaneously*. Nevertheless, in this paper I assume that consumers have heterogeneous preferences defined on just one single dimension.[3] However, they have different knowledge about their preferences at each stage of the game. Consumers have a more or less intensive expected or actual valuation of the quantity consumed, but they do not take into account any other quality characteristic of the product. Thus, the stochastic

[3] The main result of the multidimensional screening literature is that type bunching is optimal due to a conflict between participation constraints and second order incentive compatibility conditions [Rochet and Choné (1998)]. My single–dimensional assumption helps focusing the analysis on the stochastic feature of demand since multidimensional nonlinear pricing can only be solved explicitly for utility functions that are radial symmetric in type dimensions [Armstrong (1996), Wilson (1993, §12–14)], where monotonicity of the optimal tariff does not hold in general [Wilson (1995)]. Thus, within a multidimensional framework, it would not be possible to isolate whether the lack of monotonicity is due to the violation of any sufficient condition or to the interactions among multiple type dimensions.

nature of the problem allows the monopolist to screen *sequentially* each component of the *ex–post* type.

An area where optional tariffs are prevalent is telecommunications. It is commonly reported in many telecommunications demand studies that telephone customers show a biased, even irrational, preference for flat tariff options.[4] I exploit the unique data of the 1986 Kentucky telephone tariff experiment to test the empirical implications and make policy evaluations using the suggested type–varying model. The interesting feature of this data set is that it includes direct observations of θ and θ_1. Thus, I compute Anderson's (1996) nonparametric test of stochastic dominance to provide with evidence in favor of the suggested type–varying model. The advantage of using instruments directly linked to consumers' tastes, is that the empirical analysis is not subject to the common identification and misspecification of structural models dealing with asymmetric information issues.[5]

The paper is organized as follows. Section 2 presents the solution of the standard nonlinear pricing problem when the monopolist offers a continuum of *ex–post* self–selecting two–part tariffs, and studies whether they will be characterized by quantity discounts. Section 3 introduces optional nonlinear pricing, first through a menu of optional two–part tariffs, and later by means of nonlinear options. Section 4 proves that under very general assumptions the increasing hazard rate property of the distribution of consumers' private information parameter is preserved under convolution, and thus shows that *ex–post* pricing is well defined and generally characterized by quantity discounts when type components are stochastically independent. Section 5 presents evidence in favor of the taste-varying model. Section 6 analyzes whether the monopolist and/or consumers prefer *ex–ante* to *ex–post* pricing, and empirically evaluates the welfare effects of the introduction of optional tariffs by using the kernel distribution of the observed θ, θ_1, and θ_2 in two local exchanges of Kentucky. Section 7 concludes.

2 Quantity Discounts in Nonlinear Pricing

This section briefly reviews the standard (*ex–post*) nonlinear pricing problem. I discuss the main assumptions of the pricing mechanism necessary to generate a separating Perfect–Bayesian Nash Equilibrium for the static game of incomplete information played by consumers and the monopolist. I also isolate sufficient constraints on demand and distribution of consumer's single–dimensional taste index so that screening of different

---

[4] See Hobson and Spady (1988), Kridel, Lehman, and Weisman (1993), and Srinagesh (1992). Train, Ben–Akiva, and Atherton (1989) use the same argument to explain the choice of tariff service to pay for domestic electricity consumption while Train, McFadden, and Ben–Akiva (1987) report that telephone customers switch options less frequently than expected from a pure cost minimization perspective. Using the same data set than the present paper, Miravete (2000a) shows that the choice behavior of customers of local telephone service learn very fast which tariff option is less expensive for their consumption profile.

[5] Empirical models as those of Ivaldi and Martimort (1994), Miravete (2000b), and Wolak (1996) identify the effects of asymmetric information through some structural restrictions and/or distribution assumptions. It is then difficult to acknowledge whether the estimates actually isolate the effect of asymmetry of information or those of the misspecification of the structural model.

types of consumers is achieved by means of quantity discounts. In order to provide a reference framework to compare the solution of the optional nonlinear pricing mechanism given in Section 3, I develop the mechanism using consumer's indirect instead of direct utility function, and marginal tariffs instead of quantities as the monopolist's control variable.

I assume an environment where consumers' preference heterogeneity is captured by a single–dimensional index, θ. This taste indicator is private information for consumers while the monopolist only knows the population distribution of such index, $F(\theta)$. Given this informational constraint, the monopolist designs a fully nonlinear tariff to maximize his expected profits given the distribution of θ, by extracting consumer surplus in a different proportion depending on consumers' purchase levels. Thus, consumers are given incentives to self–select their purchase levels according to their preference intensity, θ. I need to assume that $F(\theta)$ is increasing hazard rate (IHR) to ensure a separating equilibrium and avoid bunching of types at any given consumption or marginal tariff levels [Maskin and Riley (1984, §4)]. This property characterizes most common distributions used in economics, and the assumption should not be considered restrictive.

Definition 1: If a univariate random variable θ has density $f(\theta)$ and distribution function $F(\theta)$, then the hazard rate of either θ or $F(\theta)$ is the ratio: $r(\theta) = f(\theta)/[1 - F(\theta)]$ on $\{\theta \in \Theta : F(\theta) < 1\}$. A univariate random variable θ or its cumulative distribution function $F(\theta)$ are said to be increasing hazard rate if $r'(\theta) > 0$ on $\{\theta \in \Theta : F(\theta) < 1\}$.

Assumption 1: The *ex–post* preference index θ has a continuously differentiable probability density function $f(\theta) \geq 0$ on $\Theta = [\underline{\theta}, \overline{\theta}] \subseteq \Re$, such that the cumulative distribution function given by:

$$F(\theta) = \int_{\underline{\theta}}^{\theta} f(z)dz, \tag{1}$$

is absolutely continuous. Furthermore while θ remains private information for each consumer, $F(\theta)$ is common knowledge and IHR.

The monopolist sells a single product x at a marginal charge p. Consumers' income is taken as *numeraire*. In addition, and for simplicity, I assume that there are no income effects for consumers or capacity constraints for the monopolist. If consumers demand and $F(\theta)$ are properly behaved, the existence of quantity discounts is equivalent to the concavity of the tariff, which is just the lower envelope of the menu of self–selecting two–part tariffs. This equivalence results guarantees the duality between the choice of any quantity under the nonlinear tariff and the unique corresponding choice of a marginal tariff and a fixed fee characterizing a hypothetical self–selecting two–part tariff that leads to the same consumption level. Thus, for analytical convenience, I will assume that consumers choose the pair $\{A(x), p(x)\}$ instead of choosing x directly. The assumed indirect utility function net of fixed fee payment A is:

$$V(p, A, \theta) = v(p, \theta) - A = \int_{p}^{\infty} x(z, \theta) dz - A, \tag{2}$$

so that Roy's identity ensures that:

$$V_p(p, \theta, A) = v_p(p, \theta) = -x(p, \theta). \tag{3}$$

In order to characterize the optimal nonlinear schedule, some structure has to be imposed on the set of preferences as well as on the distribution of types. Focusing on demand, in order to ensure the existence of a separating equilibrium, it is necessary that consumers' demands do not cross so that consumers can be ranked by their preference intensity, θ. This is the well known *single–crossing property* (SCP).

ASSUMPTION 2: The indirect utility function is convex in price and increasing in the taste parameter. Thus $V_p(\cdot) = -x(\cdot) \leq 0$; $V_{pp}(\cdot) = -x_p(\cdot) > 0$; and $V_\theta(\cdot) > 0$; which implies $v_{p\theta}(\cdot) = -x_\theta(\cdot) < 0$ (SCP).

The incentive compatibility constraint (IC) ensures that each consumer type keeps enough informational rents to consume according to their true preferences. Thus, when choosing a particular two–part tariff each consumer chooses the one that maximizes her utility given the tariff schedule offered by the monopolist. After making use of the Envelope Theorem, the monopolist's IC constraint becomes:

$$V'(\theta) = v_\theta(p(\theta), \theta). \tag{4}$$

For simplicity, I will assume that all consumers are served by the monopolist. It therefore suffices, because of monotonicity of the optimal marginal charge $\hat{p}(\theta)$, to ensure that the lowest valuation consumer $\underline{\theta}$ participates in the market:[6]

$$V(\underline{\theta}) \geq 0. \tag{5}$$

Hence, given (4) – (5) a monopolist with marginal cost c maximizes the following profit function:[7]

$$\max_{p(\theta), V(\theta)} \int_{\Theta} \left[A(\theta) + (p(\theta) - c)\, x(p(\theta), \theta)\right] dF(\theta). \tag{6}$$

The solution of this problem is a pair of functions $\{\hat{A}(\theta), \hat{p}(\theta)\}$ that relates each optimal two–part tariff offered by the monopolist to each consumer type θ:

[6] The proof of this standard result is ommited. Because of SCP and IC, the chosen $\hat{p}(\theta)$ is nonincreasing in θ, and it suffices that the IR constraint holds only at the lower bound of Θ.

[7] Fixed cost is omitted to simplify the notation since it does not play any role in characterizing the optimal tariff, neither it can be identified with the available data at the simulation stage.

$$\hat{p}(\theta) = c - \frac{1}{r(\theta)}\left[\frac{v_{p\theta}(p(\theta),\theta)}{v_{pp}(p(\theta),\theta)}\right], \tag{7}$$

$$\hat{A}(\theta) = v(\hat{p}(\theta),\theta) - \int_{\underline{\theta}}^{\theta} v_\theta(\hat{p}(z),z)dz. \tag{8}$$

Equation (7) presents the classical result that only the highest consumer type is efficiently priced. The magnitude of the price distortion for each type θ therefore depends not only on the characteristics of demand, but also critically on the monopolist's knowledge of the population distribution of tastes. The spread of this distribution is related to the importance of the asymmetry of information between the monopolist and his customers regarding consumers' preferences. Therefore, conditional on the available information, the monopolist charges the optimal mark–up over marginal cost for each consumption level. The hazard rate of this distribution captures the economic effect of informational asymmetries and plays an important role in defining the magnitude of the price distortion (deviation from c) for each *ex–post* consumer type.

Provided that the single crossing property holds, and the hazard rate of the distribution of *ex–post* types is increasing, the monopolist can screen consumers by offering a continuum of self–selecting two–part tariffs that implement the optimal nonlinear pricing solution. Since each consumer type finds one and only one of these tariff plans to maximize her utility, each two–part tariff is the optimal solution for only one *ex–post* consumer type, and therefore the equilibrium is ensured to be fully separating. As mentioned before, a sufficient condition for this continuum of two–part tariffs to be self–selecting is that its lower envelope be concave in consumption, *i.e.*, a tariff with quantity discounts. This is equivalent to the marginal tariff being decreasing in θ:

$$\hat{p}'(\theta) = \frac{r'(\theta)}{r^2(\theta)}\frac{v_{p\theta}(p(\theta),\theta)}{v_{pp}(p(\theta),\theta)} - \frac{1}{r(\theta)}\frac{\partial}{\partial\theta}\left[\frac{v_{p\theta}(p(\theta),\theta)}{v_{pp}(p(\theta),\theta)}\right] \leq 0. \tag{9}$$

Since $F(\theta)$ is IHR, for the tariff to show quantity discounts it will suffice that the following condition, related to the second term on the right hand side of (9), and involving third derivatives of the indirect utility function, holds:[8]

$$\frac{\partial}{\partial\theta}\left[\frac{v_{p\theta}(p(\theta),\theta)}{v_{pp}(p(\theta),\theta)}\right] \geq 0. \tag{10}$$

Any discrete version of the model will therefore consist of a menu of two–part tariffs where lower marginal rates $\hat{p}_1 > \hat{p}_2 > \ldots > \hat{p}_n$, are associated to higher fixed fees $\hat{A}_1 < \hat{A}_2 < \ldots < \hat{A}_n$ –because of the IC equation–, and thus, these two–part tariffs will be self–selecting, and characterized by quantity discounts.

[8] For inequality (10) to hold it suffices that $x(p,\theta)$ be concave in θ and that the price elasticity of demand be non–increasing in θ. These regularity conditions are rather technical and have little economic content, but they are assumed to hold in order to ensure a concave nonlinear schedule.

3 Optional Nonlinear Tariffs

This section studies the design of optimal pricing mechanisms when consumer demand is stochastic. I analyze two cases: optional two–part tariffs and optional nonlinear tariffs. The sequential analysis of these two differentiated stages clearly points out the role of the distribution of each component of the type on the features of the tariff options. The single–dimensional *ex–ante* type, *ex–post* type, and type shock are defined as follows:

$$\theta = \theta_1 + \theta_2 \tag{11}$$

ASSUMPTION 3: The *ex–ante* preference index θ_1 and the type shock θ_2 have continuously differentiable probability density functions $f_i(\theta_i) \geq 0$, $i = 1, 2$, on $\Theta_1 = [\underline{\theta}_1, \overline{\theta}_1] \subseteq \Re$ and $\Theta_2 = [\underline{\theta}_2, \overline{\theta}_2] = [\underline{\theta} - \overline{\theta}_1, \overline{\theta} - \underline{\theta}_1] \subseteq \Re$ respectively, such that the cumulative distribution functions given by:

$$F_i(\theta_i) = \int_{\underline{\theta}_i}^{\theta_i} f_i(z)dz \qquad ; \qquad i = 1, 2 \tag{12}$$

are absolutely continuous. As in the standard case, θ_1 and θ_2 remain private information for each consumer while $F_1(\theta_1)$ and $F_2(\theta_2)$ are both common knowledge and IHR.

ASSUMPTION 4: θ_1 and θ_2 are independent random variables.

This last assumption is needed to solve the pricing problem explicitly, analyze how are the properties of the involved distributions related, characterize the properties of optional tariffs, and compare them to the standard nonlinear pricing solution. Section 4.3 explores the robustness of theoretical results when θ_1 and θ_2 are correlated.

3.1 Menu of Two–Part Tariffs

Consumers first choose an optional tariff characterized by a fixed payment A, and by a particular marginal tariff p. At the time of their choice, consumers are not fully aware of their preferences. They only know θ_1, and the distribution of θ_2. This means that consumers do not know how much will they consume when they choose the optional tariff plan. The choice of the tariff plan is final, and neither the monopolist can take advantage by switching consumers to a different plan, nor the consumer can request such a change in the interim between the tariff subscription and the consumption decision. If there is any tariff switching, it will only apply to future billing periods. Thus, given consumers' private information θ_1 and their expectation on type shocks, consumers choose the tariff plan that maximizes their expected net rent, which given equation (11) and Assumption 4 leads to the following *ex–ante* IC constraint that applies to the choice of tariff options:

$$\tilde{V}'(\theta_1) = E_2[v_\theta(\tilde{p}(\theta_1), \theta)]. \tag{13}$$

Once the tariff option has been chosen, consumers learn their *ex–post* type through the realization of an individual type shock. Then, their consumption level is decided, contingent on the previously chosen tariff plan $\{A(\theta_1), p(\theta_1)\}$. The optimal consumption decision maximizes the actual rent given the tariff option, which leads to the following *ex–post* IC constraint:

$$U'[x(\tilde{p}(\theta_1), \theta)]x_\theta(\tilde{p}(\theta_1), \theta) = \tilde{p}(\theta_1), \tag{14}$$

where $U[\cdot]$ denotes the direct utility function. The monopolist has to consider now two sets of participation constraints. To make the different solutions directly comparable, I assume that all households participate subscribing the service regardless of whether they later buy anything or not:

$$\tilde{V}(\underline{\theta}_1) \geq 0, \tag{15}$$

while *ex–post* actual rents should be non–negative in order to consume, and given the previous choice of service it is required that:

$$V(\theta_1, \underline{\theta}_2(\theta_1)) = v(\tilde{p}(\theta_1), \theta_1 + \underline{\theta}_2(\theta_1)) - \tilde{A}(\theta_1) \geq 0. \tag{16}$$

Thus, each *ex–ante* consumer type who chose the option $\{\tilde{A}(\theta_1), \tilde{p}(\theta_1)\}$ faces a different *ex–post* participation constraint. If the type shock is negative enough, $\theta_2 \leq \underline{\theta}_2(\theta_1)$, consumers with *ex–ante* type θ_1 do not to buy anything and thus the monopolist only gets the fixed fee $\tilde{A}(\theta_1)$ from them.[9] Then, given constraints (13) and $(15) - (16)$, the monopolist maximizes the following profit function:

$$\max_{\tilde{p}(\theta_1), \tilde{V}(\theta_1)} \int_{\Theta_1} \left[\tilde{A}(\theta_1) + (\tilde{p}(\theta_1) - c) \int_{\underline{\theta}_2(\theta_1)}^{\overline{\theta}_2} x(\tilde{p}(\theta_1), \theta) dF_2(\theta_2) \right] dF_1(\theta_1), \tag{17}$$

The solution of this problem is again a pair of functions $\{\tilde{A}(\theta_1), \tilde{p}(\theta_1)\}$ now associating an optional two–part tariff offered by the monopolist to each consumer with *ex–ante* type θ_1:

$$\tilde{p}(\theta_1) = c - \frac{1}{r_1(\theta_1)} \left[\frac{E_2[v_{p\theta}(\tilde{p}(\theta_1), \theta_1 + \theta_2) \,|\, \theta_2 \geq \underline{\theta}_2(\theta_1)]}{E_2[v_{pp}(\tilde{p}(\theta_1), \theta_1 + \theta_2) \,|\, \theta_2 \geq \underline{\theta}_2(\theta_1)]} \right], \tag{18}$$

$$\tilde{A}(\theta_1) = E_2 \left[v(\tilde{p}(\theta_1), \theta_1 + \theta_2) - \int_{\underline{\theta}_1}^{\theta_1} v_\theta(\tilde{p}(z), z + \theta_2) dz \;\middle|\; \theta_2 \geq \underline{\theta}_2(\theta_1) \right]. \tag{19}$$

This solution resembles that of the *ex–post* pricing very closely. With the exception of the *ex–post* participation constraint, the menu of optional two–part tariffs does not screen consumers with respect to their *ex–post* type, since θ_2 is integrated out in the monopolist's objective function. The monopolist thus just screens consumers with respect to θ_1 by

[9] The *ex–ante* type dependent cut–off shock $\underline{\theta}_2(\theta_1)$ is uniquely defined in (16) for each θ_1 due to continuity of all functions involved and monotonicity of $v_\theta(\cdot) > 0$.

offering them a menu of optional two–part tariffs that accounts for consumer differences before θ_2 is realized. The type shock only determines the amount that each consumer will purchase depending on the tariff option previously chosen. Denoting by $E_2^\star$ the conditional expectation with respect to the shock given that the ex–post participation constraints is fulfilled, and by differentiating (18) with respect to θ_1 we have:

$$\tilde{p}'(\theta_1) = \frac{r_1'(\theta_1)}{r_1^2(\theta_1)} \left[\frac{E_2^\star[v_{p\theta_1}(\tilde{p}(\theta_1), \theta_1 + \theta_2)]}{E_2^\star[v_{pp}(\tilde{p}(\theta_1), \theta_1 + \theta_2)]} \right] - \frac{1}{r_1(\theta_1)} \frac{\partial}{\partial \theta_1} \left[\frac{E_2^\star[v_{p\theta_1}(\tilde{p}(\theta_1), \theta_1 + \theta_2)]}{E_2^\star[v_{pp}(\tilde{p}(\theta_1), \theta_1 + \theta_2)]} \right] \leq 0. \quad (20)$$

Equation (20) shows that it is optimal to offer tariff options with lower marginal tariffs for future consumption if they are associated to higher actual fixed payments. Thus, the *ex–ante* tariff is characterized by quantity discounts, which means that there is a concave, lower envelope function underlying the optional tariffs. This concave function $\tilde{T}(\theta_1)$ is the mathematical lower envelope of the menu of two–part tariffs denoted by $\{\tilde{A}(\theta_1), \tilde{p}(\theta_1)\}$. But this function is not the tariff lower envelope in the traditional sense. For each *ex–ante* type θ_1 and tariff choice $\{\tilde{A}(\theta_1), \tilde{p}(\theta_1)\}$ there is a unique type shock $\theta_2 = \theta_2^*(\theta_1)$ so that total payments equal those of the lower envelope. We know that $\theta_2^*(\theta_1)$ is unique because the SCP requires that demand is increasing in the type, $x_\theta(\cdot) > 0$, and the marginal tariff $\tilde{p}(\theta_1)$ is given. Thus, if consumers receive any other shock different from $\theta_2^*(\theta_1)$ they will move along the tariff option chosen and will always pay more under the chosen tariff regime than if the had "correctly" anticipated their future consumption, in which case they had moved along the lower envelope by choosing a different two–part tariff option. If we now repeat the analysis for other *ex–ante* types who chose different tariff options, we could easily check that the shape of the actual *ex–post* tariff is state–dependent, and that the payment outlay function is not ensured to be concave unless we unrealistically restrict the behavior of θ_2. However, since the distribution of θ_1 is IHR, the mathematical lower envelope $\tilde{T}(\theta_1)$ is still concave, and thus the optimal two–part tariff options are such that they lead to quantity discounts by offering a lower marginal rate associated to higher fixed fees. Furthermore, quantity discounts do not require any additional assumption on demand relative to those made for the *ex–post* case. The reduction of the marginal tariffs for class of services designed for high volumes of consumption is ensured because the distribution of *ex–ante* types is IHR, which suffices to ensure IC of ex–ante plans.

A comparison with some related literature in Regulatory Economics may be useful at this point. Laffont and Tirole (1993, §1.4) show that optimal regulatory mechanisms can generally be implemented by a menu of linear contracts as long as the distribution of the adverse selection parameter of firms is IHR. This corresponds to the analysis carried out in Section 2 when the fully nonlinear *ex–post* tariff represented the lower envelope of a menu of self–selecting two–part tariffs. In both cases there is a one–to–one relationship between quantity purchased or reported cost level and the slope and intercept of the linear contract.

The analysis of this subsection has made clear that when the monopolist offers a menu of optional tariffs and consumers do not commit to purchase some particular level at the time of tariff choice, the consumption decision and tariff choice are no longer equivalent solutions of dual problems. In the Regulation literature, Caillaud, Guesnerie, and Rey (1992) and Laffont and Tirole (1986) among others prove that linear contracts in reported cost are robust to the existence of additive shocks in the cost functions of firms. In these models firms' objective functions are linear in the cost noise that might exist. Thus, substituting its expected value, firms' IC and participation constraints are unchanged. The realization of the cost shock still affects total payments (as in the present model), but the lower envelope and the linear contracts remain unchanged. This is not the case for the model presented here. Uncertainty enters nonlinearly in consumers' objective function, thus affecting the *ex–ante* IC and participation constraints (13) and (15) respectively. Neither the tariff's lower envelope or the two–part tariff options are immune to the existence of uncertainty, but the main result of this subsection is to prove that even if this is the case, the tariff can still be implemented by a menu of linear options.[10]

If the IHR assumption is not fulfilled and the optimal regulatory mechanism cannot be decentralized through linear contracts, Picard (1987) shows that it could still be implemented by a menu of appropriately chosen quadratic options. In the context of the present research this means that optimal tariffs involve quantity premium. My alternative approach is to assume that the distribution of θ_1 remains IHR through the analysis in order to make possible the existence of quantity discounts. Then, next subsection identifies the conditions for nonlinear options (not necessarily restricted to be quadratic) to be concave.

3.2 Menu of Nonlinear Tariffs

We now deal with the general problem of a menu of nonlinear tariffs that also induce self–selection of consumers with respect to their type shocks. I will characterize this tariff using a constructive approach starting from the solution of the previous section. Each tariff option in Section 3.1 was a two–part tariff and thus, each consumer selected one among $\{\tilde{A}(\theta_1), \tilde{p}(\theta_1)\}$ depending on her *ex–ante* type θ_1 only but there was no further incentive for consumers to *ex–post* self–select according to the realization of θ_2. Two consumers with the same expected consumption or *ex–ante* valuation of the product would choose the same two–part tariff option. But they will later consume different amounts depending on their respective type shocks. Since both of them consume at the level where their marginal utility equates their marginal tariffs chosen in advance, the monopolist extracts some rent based on the consumer *ex–ante* type while consumers keep all their *ex–post* informational rent exclusively due to the learned type shock. To avoid this insufficient screening, nonlinear tariff options also induce *ex–post* self–selection by means of quantity discounts and thus further reduce consumers' *ex–post* informational rents. Provided that

[10] Baron and Besanko (1999) rise the question of the equivalence of solutions when the type of an alliance θ comprises the types of the alliance members θ_1 and θ_2 when the type of the alliance is defined as in equation (11). This is not the case in the present model because of the sequential nature of the screening process, as well as for the fact that the unresolved uncertainty about θ_2 affects the IC and participation constraint of each consumer. A direct mechanism $\{\hat{A}(\theta_1{+}\theta_2), \hat{p}(\theta_1{+}\theta_2\}$ will not be equivalent to $\{\tilde{A}(\theta_1), \tilde{p}(\theta_1\}$ unless the distribution of θ_2 becomes degenerate.

each nonlinear tariff option is concave, they can also be implemented by a continuum of self–selecting two–part tariffs. The task of the monopolist is not to design the optimal menu of menus of linear tariff. At stage 1, when consumers only know θ_1 they choose a nonlinear tariff option $\tilde{\tilde{T}}(\theta_2 \mid \theta_1)$, or alternatively a particular continuum of *ex–post*, self–selecting, two–part tariffs $\{\tilde{\tilde{A}}(\theta_2 \,|\, \theta_1), \tilde{\tilde{p}}(\theta_2 \,|\, \theta_1)\}$. Given consumers' private information θ_1, their expectation on type shocks, and their knowledge of the "shapes" of tariff options for each θ_1, they choose the tariff plan that maximizes their expected net rent. Later, once θ_2 is realized, the mechanism determines consumption and payments conditional on the previous choice of tariff.

General characterizations of the menu of nonlinear tariffs are difficult and cumbersome. However, the fact that types components are statistically independent from each other proves to be very useful in obtaining the characterization of this menu of nonlinear tariff options. Since the shock is independent of the *ex–ante* type, the mathematical lower envelope, $\tilde{T}(\theta_1)$, still capture the optimal incentive mechanism to screen consumers with respect to their *ex–ante* type dimension regardless of whether tariff options are two–part tariffs or more general nonlinear functions. Thus, $\tilde{\tilde{T}}(\theta_2 \mid \theta_1)$ can be considered to be composed of two elements: one that screens consumers with respect to θ_1, represented by $(18) - (19)$, and another that induces self–selection of *ex–post* types given the optimal tariff choices of each *ex–ante* type θ_1. This second component increases the revenue of the monopolist by reducing consumers informational rents exclusively related to θ_2. Obviously, if θ_1 and θ_2 were not independent it would be impossible to separate the origin of the rent extraction as screening for θ_1 should also account for the related distribution of θ_2. In order to characterize the optimal menu of nonlinear options $\tilde{\tilde{T}}(\theta_2 \,|\, \theta_1)$, observe that since tariff options are nonlinear, the *ex–post* IC constraint of a consumer with *ex–ante* type θ_1 solves:

$$\theta_2 \in \arg\max_{\theta_2'} \left[v(\tilde{\tilde{p}}(\theta_2' \,|\, \theta_1), \theta_1 + \theta_2) - \tilde{\tilde{A}}(\theta_2' \,|\, \theta_1) \right]. \tag{21}$$

Observe also that since type components are independent each nonlinear option should be tangent only once to $\tilde{T}(\theta_1)$ as characterized by equations $(18) - (19)$, so that $\{\tilde{A}(\theta_1), \tilde{p}(\theta_1)\}$ remains the optimal mechanism to screen the *ex–ante* type dimension θ_1 as the effect of θ_2 has already been integrated out. Therefore, only one particular two–part tariff of each menu of *ex–post* nonlinear tariffs that characterizes each nonlinear option coincides with one of the optional two–part tariffs of the problem solved in the previous section. Thus, total and marginal payments will be the same, as well as consumption, when the realized shock equals $\theta_2^*(\theta_1)$.[11] Under independence, this *ex–ante*, type specific, critical shock is implicitly defined by:

$$E_2[\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1)] = \tilde{\tilde{p}}(\theta_2^*(\theta_1) \,|\, \theta_1), \tag{22}$$

The solution of the menu of nonlinear options builds upon the menu of optional two–part tariffs of Section 3.1. Taking the solution of the menu of optional two–part

[11] Uniqueness of $\theta_2^*(\theta_1)$ is ensured by monotonicity of equation (10) in θ.

tariffs (18) – (19) as the boundary condition at $\theta_2^*(\theta_1)$, the optimal screening process with respect to θ_2 results in deviations of the *ex–post* marginal tariff and fixed fee payment relative to those of $\{\tilde{A}(\theta_1), \tilde{p}(\theta_1)\}$. Thus, let define:

$$\Delta\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1) = \tilde{\tilde{p}}(\theta_2 \,|\, \theta_1) - \tilde{p}(\theta_2 \,|\, \theta_1), \tag{23}$$

$$\Delta\tilde{\tilde{A}}(\theta_2 \,|\, \theta_1) = \tilde{\tilde{A}}(\theta_1 + \theta_2) - \tilde{A}(\theta_1). \tag{24}$$

The IC constraint (21) can be rewritten as follows:

$$\tilde{\tilde{V}}'(\theta_2 \,|\, \theta_1) = v_\theta(\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1), \theta_1 + \theta_2). \tag{25}$$

In addition, there is another boundary constraint for this problem:

$$\Delta\tilde{\tilde{V}}(\theta_2^*(\theta_1) \,|\, \theta_1) = 0, \tag{26}$$

so that $\tilde{\tilde{p}}(\theta_2^*(\theta_1) \,|\, \theta_1) = \tilde{p}(\theta_1)$, and thus each nonlinear tariff option is ensured to be tangent to $\tilde{T}(\theta_1)$ only once if $\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1)$ is monotone.

Given all these constraints the monopolist's problem solves, for each possible nonlinear option, the change in marginal rate that will maximize the increase in revenues from the corresponding "boundary two–part tariff" option:

$$\max_{\Delta\tilde{\tilde{p}}, \Delta\tilde{\tilde{V}}} \int_{\Theta_2} \left[\tilde{\tilde{A}}(\theta_2 \,|\, \theta_1) + \tilde{\tilde{p}}(\theta_2 \,|\, \theta_1) x(\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1), \theta_1 + \theta_2) - \tilde{A}(\theta_1) - \tilde{p}(\theta_1) x(\tilde{p}(\theta_1), \theta_1 + \theta_2) \right] dF_2(\theta_2). \tag{27}$$

Solving this optimal control problem, the optimal changes of the marginal tariff and fixed fee relative to the optimal two–part tariff option chosen by an *ex–ante* type θ_1 are:

$$\Delta\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1) = -\frac{F_2(\theta_2^*(\theta_1)) - F_2(\theta_2)}{f_2(\theta_2)} \left[\frac{v_{p\theta}(\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1), \theta_1 + \theta_2)}{v_{pp}(\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1), \theta_1 + \theta_2)} \right], \tag{28}$$

$$\begin{aligned} \Delta\tilde{\tilde{A}}(\theta_2 \,|\, \theta_1) = {} & v(\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1), \theta_1 + \theta_2) - v(\tilde{p}(\theta_1), \theta_1 + \theta_2) \\ & - \int_{\theta_2^*(\theta_1)}^{\theta_2} [v_\theta(\tilde{\tilde{p}}(\theta_1 + z), \theta_1 + z) - v_\theta(\tilde{p}(\theta_1), \theta_1 + z)] dz. \end{aligned} \tag{29}$$

These two equations in conjunction with (18) – (19) characterize a menu of optional nonlinear tariffs $\{\tilde{\tilde{A}}(\theta_2 \mid \theta_1), \tilde{\tilde{p}}(\theta_2 \mid \theta_1)\}$ for each value of θ_1. Observe that equation (28) implies that consumers with *ex–ante* type θ_1 faces higher marginal charges than $\tilde{p}(\theta_1)$ if they receive a small shock $\theta_2 < \theta_2^*(\theta_1)$, but on the contrary, marginal tariffs will be smaller than $\tilde{p}(\theta_1)$ if $\theta_2 > \theta_2^*(\theta_1)$.

The final question that I have to address in this section is whether any further assumption is necessary to ensure that each nonlinear tariff $\tilde{\tilde{T}}(\theta_2 \,|\, \theta_1)$ is concave, so that screening consumers with respect to their type shocks could also be achieved through quantity discounts. As Section 3.1 proved that $\tilde{T}(\theta_1)$ is concave, there only remains to analyze whether marginal tariffs $\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1)$ are decreasing in θ_2. Thus, for each particular nonlinear option $\{\tilde{\tilde{A}}(\theta_2 \,|\, \theta_1), \tilde{\tilde{p}}(\theta_2 \,|\, \theta_1)\}$ to be concave it is required that:

$$\begin{aligned}\frac{\partial \Delta \tilde{\tilde{p}}(\theta_2 \,|\, \theta_1)}{\partial \theta_2} &= \left[\frac{r_2'(\theta_2)}{r_2(\theta_2)} - \frac{f_2'(\theta_2)}{f_2^2(\theta_2)} \frac{f_2(\theta_2^*(\theta_1))}{r_2(\theta_2^*(\theta_1))}\right] \left[\frac{v_{p\theta}(\tilde{\tilde{p}}(\theta), \theta)}{v_{pp}(\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1), \theta_1 + \theta_2)}\right] \\ &+ \left[\frac{1}{f_2(\theta_2)} \frac{f_2(\theta_2^*(\theta_1))}{r_2(\theta_2^*(\theta_1))} - \frac{1}{r_2(\theta_2)}\right] \frac{\partial}{\partial \theta_2} \left[\frac{v_{p\theta}(\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1), \theta_1 + \theta_2)}{v_{pp}(\tilde{\tilde{p}}(\theta_2 \,|\, \theta_1), \theta_1 + \theta_2)}\right] \leq 0. \quad (30)\end{aligned}$$

The ratio $v_{p\theta}/v_{pp}$ is negative because the indirect utility function is convex in price, and because of the SCP as stated by Assumption 1, while its derivative with respect to θ_2 is ensured to be positive by equations (10) and (11). Hence, the concavity of the nonlinear tariff option critically depends on the signs of the terms between brackets. The first term between brackets in equation (30) is ensured to be positive only if $r_2'(\theta_2) > f_2'(\theta_2)[1 - F_2(\theta_2^*(\theta_1))]/[1 - F_2(\theta_2)]^2$, while the second term between brackets in equation (30) is negative only as long as the shock θ_2 does not exceed $\theta_2^*(\theta_1)$.

Observe that even if $f_2'(\theta_2) \leq 0$ and the distribution of θ_2 is IHR, concavity of each nonlinear option in this more complicated problem also requires that $\theta_2 \leq \theta_2^*(\theta_1)$, *i.e.*, concavity will be ensured for lower consumption levels. If $f_2'(\theta_2) > 0$ or if $\theta_2 > \theta_2^*(\theta_1)$, the IHR assumption $r_2'(\theta_2) > 0$ does not suffice to ensure that each nonlinear option $\tilde{\tilde{T}}(\theta_2 \,|\, \theta_1)$ is concave. Therefore we need the impose the more restrictive assumption that the hazard rate of the distribution of the type shock is sufficiently increasing in order to compensate the effect of large shocks or increasing density functions.[12] The IHR property is still critical for the model to be well behaved, but it is not sufficient to ensure that each nonlinear tariff option leads to quantity discounts. If we just require that $r_2'(\theta_2) > 0$, we may find an asymmetric treatment of consumers with different θ_1: nonlinear tariff options chosen by high θ_1 are most likely concave, while on the contrary, low θ_1 choosing tariff options "designed" for low consumption levels would suffer important premia if they consume much more than what they expected. But still, the IHR property proves to be critical for the model to be well behaved.

Therefore, going from a menu of two–part tariffs to a menu of nonlinear tariffs only requires the additional assumption that the hazard rate of the distribution of the

[12] There are cases where these conditions are fulfilled. One of such cases is when $\theta_2 \in [0, 1]$ is distributed as a standard unit beta distribution of the first kind with parameters $p = 1$ and $q = \lambda_2$. This distribution is IHR as long as $\lambda_2 > 0$, and the density function is always decreasing when $\lambda_2 > 1$. The hazard rate of this distribution varies from λ_2 when $\theta_2 = 0$ to ∞ when $\theta_2 = 1$. Thus, it is always possible to find a large enough value of λ_2 to ensure that the nonlinear tariff option is concave, even when $\theta_2 > \theta_2^*(\theta_1)$.

shock is sufficiently increasing to ensure the existence of quantity discounts for every single nonlinear tariff option. The basic setup thus remains unchanged: the IHR property proves to be critical for the model to be well behaved. The following section studies how IHR properties are affected by the existence of stochastic components, so that the relative profitability of the *ex–ante* and *ex–post* pricing could be addressed.

4 Shocks, Convolutions, and Stochastic Dominance

The previous two sections have shown how to solve in isolation either the standard *ex–post* nonlinear pricing problem, or the more complex *ex–ante* optional nonlinear pricing problem. A most relevant question is whether these two solutions can be compared by the monopolist in order to choose the most profitable one in expectation. Thus, these two problems have to be consistently defined so that the *ex–post* pricing solution accounts for all statistical properties derived from the fact that θ is actually the result of the convolution of θ_1 and θ_2. In this section I focus on the relationship between properties of the distributions of the *ex–ante* type θ_1 and the type shock θ_2, and show how are they related to the features of the distribution of the *ex–post* type θ.

This section covers three related topics. First, it is shown that if the distribution of the components of the types is IHR, then the distribution of their convolution is also IHR. Second, I explore environments in which the comparison of these hazard rates can be ordered unambiguously. Third, I discuss how robust are these results to the definition of θ and the assumed stochastic independence of its components.

4.1 Preservation of IHR under Convolution

I should start this section by defining the distribution of the *ex–post* type in terms of the distribution of its components.

Definition 2: Let θ_1 and θ_2 be independent, univariate, random variables with cumulative distribution functions $F_i(\theta_i) : \Theta_i \to [0, 1],\ i = 1, 2$. The cumulative distribution function of $\theta = \theta_1 + \theta_2$ is given by the Fourier convolution:

$$F(\theta) = \int_{\Theta_2} F_1(\theta - \theta_2) f_2(\theta_2) d\theta_2. \tag{31}$$

Therefore, given any arbitrary, but well behaved, distribution function for the *ex–ante* type and the type shock, it is always possible to identify the distribution of *ex–post* types up to a linear transformation.

Showing that the IHR property of the distributions of the components of the type, $\{\theta_1, \theta_2\}$, is passed through to the distribution of the *ex–post* type, θ, is absolutely necessary to study the relationship between the features of the *ex–ante* optional and the *ex–post* standard nonlinear tariffs as well as to test the empirical implications of a model with type changes. The following Proposition summarizes a key result for the proper characterization of the *ex–ante* and *ex–post* tariffs. The proof is included in the Appendix.[13]

PROPOSITION 1: *If the distribution of the components of the type, $F_1(\theta_1)$ and $F_2(\theta_2)$ are both IHR, then the distribution of the convolution $F(\theta)$ is IHR.*

This result is important because both the *ex–ante* and *ex–post* pricing problems are consistent, well defined, and have separating equilibria involving quantity discounts as long as the distribution of the components of the *ex–post* type are IHR.

4.2 Implications

The definition of the *ex–post* type given in equation (11) together with the regularity conditions of the distributions of the type components discussed so far ensure that θ second order stochastically dominate θ_i, *i.e.*, $\theta_i \leq^{st} \theta$. This is a direct testable implication of the type–varying model. However, as we will see in Section 6.1, second order stochastic dominance does not generally suffice to compare the relative expected efficiency of *ex–ante* and *ex–post* nonlinear tariffs as higher or lower markups are inversely related to the magnitude of the hazard rate of the involved type distribution. Second order stochastic dominance alone allows for non–uniform orderings of the markups of the *ex–ante* tariff relative to those of the *ex–post* tariff, depending on particular consumption ranges.

A sufficient condition to compare the optimal solutions of the *ex–ante* and *ex–post* nonlinear pricing mechanisms is to require a particular hazard rate ordering of the involved distributions, such as in Laffont and Tirole (1993, §1.4). Since optimal nonlinear solutions critically depend on the value of the hazard rate of the corresponding distribution I have to establish how large is the hazard rate of the convolution distribution $F(\theta)$ relative to those of the components of the *ex–post* type, and thus determine whether a type–varying model may lead to a uniform ordering of hazard rates and markups of each pricing mechanism. Proposition 2, also proved in the Appendix, shows that for the present type–varying model, θ dominates in hazard rate to θ_i if the support of the distributions is restricted to $\Re_+$.

PROPOSITION 2: *Let $F_i(\theta_i)$ be* IHR*, i.e., $r_i'(\theta_i) > 0$ in θ_i on $\{\theta_i > 0 : F_i(\theta_i) < 1\}$, for $i = 1, 2$. Let $F(\theta)$ denote the cumulative convolution distribution of $\theta = \theta_1 + \theta_2$, with hazard rate $r(\theta)$. Then $r(\theta) \leq \min\{r_1(\theta), r_2(\theta)\}$ on $\{\theta > 0 : F(\theta) < 1, F_i(\theta) < 1; i = 1, 2\}$.*

Proposition 2 implies that the distribution $F(\cdot)$ always puts more weight on higher values of the type than the distribution $F_1(\cdot)$. Therefore given some value $\hat{\theta}$, the probability

[13] In a recent paper, Biais, Martimort, and Rochet (2000) claim that the convolution distribution is IHR if at least the distribution of one of its components is IHR and the other has a bounded support. Miravete (2001) discusses the accuracy of such statement. I still require that both distribution of the components is IHR. But relative to the assumptions of Biais, Martimort, and Rochet (2000), Proposition 1 is more general because I do not exclude the set of IHR distributions whose density functions are not log–concave, or whose support is not bounded.

that $\theta > \hat{\theta}$ always exceeds the probability that $\theta_1 > \hat{\theta}$. This intuitive result is formalized in the following corollary.

COROLLARY 1: *If* $r(\theta) \leq r_i(\theta)$ *on* $\{\theta > 0 : F(\theta) < 1, F_i(\theta) < 1; i = 1, 2\}$, *then* θ *first order stochastically dominates* θ_i.

PROOF: Since $r(\theta) = -d \log[1 - F(\theta)]/d\theta$ it follows that $\forall \theta > 0$:

$$1 - F(\theta) = \exp\left[-\int_0^\theta r(z)dz\right] \geq \exp\left[-\int_0^\theta r_i(z)dz\right] = 1 - F_i(\theta), \tag{32}$$

and therefore $F(\theta) \leq F_i(\theta)$ $\forall \theta > 0$, which is the definition of first order stochastic dominance of θ over θ_i, *i.e.*, $\theta_i \leq^{st} \theta$. ∎

Some pricing or agency problems, other than the telecommunications case studied in later sections, could easily define environments where the support of type components is constrained in a natural way. For instance, we could think of $\theta_1 \in \Re_+$ as general skills of workers before being hired (*e.g.*, acquired through education and/or working experience in other jobs). If hired, workers could develop some specific skills and abilities due to learning by doing, and therefore increase their productivity. It is not unreasonable within this framework to exclude the possibility of negative learning, and thus θ_2 could also be restricted to take only positive values. Consumption of electricity also provides a related example. While households consume according to their habits and location, *i.e.*, the base load $\theta_1 \in \Re_+$, changes in temperature (public information and common to all consumers) may induce additional seasonal demand: whenever it is too cold or too warm, consumers increase their demand for electricity by turning on the heating or the air conditioner. Thus, θ_2 could also be restricted to take only positive values, and the model will produce strong empirical implications.

According to Laffont and Tirole's interpretation (1993, §1.4–1.5), Proposition 2 means that the distribution of θ is more favorable than the distribution of θ_1. Corollary 1 shows that this result could be obtained within the type–varying framework because of the existence of an independent, but systematically positive type shock ensures that the actual purchase (or valuation) is always higher in stochastic sense than the expected purchase (or valuation). Similarly, Maskin and Riley (1984, §4) already considered the effect of changes in the distribution of consumer types on the shape of the nonlinear tariffs. As I show in Section 6.1, a nonlinear schedule based on $F(\theta)$ generally involves higher markups than the nonlinear tariff based on $F_1(\theta_1)$ for all consumption levels, which is a direct consequence of the hazard rate dominance of θ over θ_1 in a model with type varying consumers if the support of the distributions are restricted to $\Re_+$.

The testable implications of Proposition 2 and Corollary 1 are, strictly speaking, limited to situations where $\theta, \theta_i \in \Re_+$, which exclude the empirical application of this paper. There is no reason to expect that consumers always underestimate their future local telephone usage, and thus the type shock θ_2 is not restricted to take only positive values. However, a strong empirical evidence in favor of first stochastic dominance of θ over θ_1 will be consistent with the underlying hazard rate dominance of θ over θ_1. Thus, in such a case, an *ex–post* nonlinear tariff could lead to higher expected profits than an *ex–ante* nonlinear tariff because the markups of $\hat{T}(\theta)$ uniformly dominates those of $\tilde{T}(\theta_2 \mid \theta_2)$ for every consumption level.

4.3 General Distributions and Type Definitions

Equation (31) establishes that given the distribution of θ_1 and θ_2 it is always possible to identify the distribution of θ up to a linear transformation of the distributions of the type components. This identification issue rises the question of whether the results of the model are limited to a particular definition of the *ex–post* type in terms of the *ex–ante* type and the type shock. However, the distinction between convolution and composition distribution is mostly irrelevant because the same preferences can always be represented by any monotone transformation of a given utility function. Thus for instance, assume that the utility function is multiplicatively separable in θ and x. Assume also that $\theta = \theta_1\theta_2$, each of which is beta distributed with appropriate parameters so that the composition distribution is uniform. Observe that the same set of preferences could be represented by a logarithmic transformation of that utility function. But in that case, θ and x are additively separable, and $\theta = \theta_1 + \theta_2$ where θ is now distributed as a $\chi^2(2)$ while θ_1 and θ_2 are now two exponential generalized beta distributions of the second kind. Thus the functional form that relates the *ex–ante* type and the shock with the *ex–post* type is not independently identifiable from the assumed distributions and/or the utility function. Since the preference index is just a theoretical construction to describe situations of asymmetric information, we can always define a monotone transformation of the utility function that scale the index and its distribution appropriately to represent the same preferences, which ensures the generality of the results of the present model. In the end, types shocks, regardless of whether they are linearly related to the *ex–ante* type or not, are only identifiable as non–price shifts in consumer demand. Therefore, I can focus on the convolution case to analyze, without loss of generality, the implications of the existence of type shocks for the design of optimal pricing mechanisms.

Another important issue is the independence of θ_1 and θ_2. It is not difficult to envision situations where large consumers also make more or less mistakes than small consumers. For example, consider the reference case where θ_1 and θ_2 are independent and $F(\theta) \leq F_1(\theta)\ \forall \theta \in \Theta \subseteq \Re$. Assume now that the *ex–ante* type and the shock are negatively correlated, and denote by $F^\star(\cdot)$ the cumulative distribution of $\theta = \theta_1 + \theta_2$ under negative correlation. The distribution of θ is now less dispersed, with less mass of probability at the tails of the distribution than if θ_1 and θ_2 were independent. In some sense the monopolist is now "less uncertain" about the value that consumer types may take, because there is a larger mass of probability around the mean of θ. Thus, for low values of θ (below the mean), the probability of finding a type above a given θ is higher under negative correlation than under independence. This is just because the probability distribution function of $\theta_1+\theta_2$ is more concentrated around the mean. Thus, the hazard rate function is lower under negative correlation than under independence for low values of θ. Just the contrary holds for high values of θ, *i.e.*, the hazard rate of the distribution with negative correlation will exceed that of the distribution of independent type components. If $r^\star(\theta) \leq r(\theta)$ only for low values of θ, then for large customers *ex–post* nonlinear pricing markups will be lower under negative correlation of type components than under the assumption of independence as markups and hazard rate of the distribution of θ are inversely related. Consumer types are more concentrated around the mean under negative correlation than under independence, and thus it is necessary to introduce important distortions to distinguish among low consumers and preserve the IC property of the mechanism.[14] Thus, the results of this paper in general need to be qualified for particular cases where type components are allowed to be correlated.

5 Empirical Evidence

Results of Section 4 provide us with direct testable implications of the "taste–varying" approach. The goal of this section is to test these theoretical prerequisites in a particular case where actual data provides with a direct instrument for θ, θ_1, and θ_2. Contrary to many applied works, the source of asymmetric information in the application studied here is not identified through the specification of some distribution of unobserved characteristics, but rather using direct observations of consumers' taste parameters. The empirical analysis exploits the information available from the 1986 Kentucky Local Telephone Tariff Experiment.

In November of 1984, the Kentucky Public Service Commission (KPSC) established Administrative Case No. 285 to study the economic feasibility of providing local measured service telephone rates. Directly linked to Case No. 285, South Central Bell (SCB) carried out an extensive tariff experiment in the second half of 1986 in two cities of Kentucky to provide the commission with evidence in favor of introducing the optional local measured service. Prior to this tariff experiment, in spring of 1986, when all households in Kentucky were on mandatory flat rates, SCB collected demographic and economic information for about 5,000 households in the local exchanges of Bowling Green and Louisville. In the second half of 1986 tariffs where modified in these two cities. In Bowling Green all customers where placed on a mandatory measured service, *i.e.*, all of them paid according to their realized consumption. In Louisville, customers had the ability to remain in the previous flat tariff regime or switch to a measured service option. Both measured services, in Bowling Green and Louisville, included a monthly fixed fee, an allowance and

[14] These intuitive results are however difficult to prove except, maybe, for particular distributions. The reason is that under correlation, the probability density function of $\theta = \theta_1 + \theta_2$ cannot be factorized as the product of the components' probability density functions, and thus $F(\theta)$ is no longer the convolution distribution of θ_1 and θ_2. One of the cases where $F^\star(\theta)$ can be written explicitly is that of $\theta = \theta_1 + \theta_2$ where $(\theta_1, \theta_2) \sim BVN[\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho]$. In this case, $\theta \sim N[\mu = \mu_1 + \mu_2, \sigma^2 = \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2]$. To illustrate the argument of this paragraph, I computed the hazard rate functions of $\theta_1 + \theta_2$ under independence, $r(\theta)$, and under perfect negative correlation ($\rho = -1$), $r^\star(\theta)$. For the case where $\mu_1 = 0$, $\sigma_1^2 = 1$, $\mu_2 = 1$, $\sigma_2^2 = 0.5$, I found that $r^\star(\theta) > r(\theta)$ $\forall \theta > 0.12$.

a nonlinear multidimensional tariff that sometimes accounted for call distance within the local exchange, two or three periods of time for load pricing, as well as setup and duration of calls. The regulated monopolist also collected monthly information on usage (number and duration of calls classified by time of the day, day of the week, and distance), and payments during two periods of three months in spring and the fall of that year.

It is remarkable that in addition to demographic and economic variables, SCB also collected information on telephone customers' usage expectations. SCB explicitly requested customers' own estimates of their weekly average number of calls. These individual estimates are particularly useful because local calls were never priced before and consumers were not aware of the tariff experiment that was going to be held in the second half of the year. Thus, neither marginal tariffs or strategic considerations influence these estimates of customers' own satiation levels. This information, available for most households of the sample can be compared with the actual number of weekly phone calls for every month in the study. A direct test of the suggested taste–varying model will be constructed in the next subsection using the comparison between the expected and the actual weekly number of phone calls during the spring months of 1986.

Table 1 presents basic descriptive statistics of the sample. These two cities have quite different demographic structures. Residents in Bowling Green make a significantly higher income and households are larger, including the proportion of teenagers. Households with married couples and college graduates are also more common in Bowling Green than in Louisville. In this latter city, on the contrary, it is more common to find retired people, those who receive some kind of social benefits to support their income, and a smaller percentage of households that have moved in the last five years. Racial composition of these cities is also different. Only 6% of the population in Bowling Green, but about 12% of the population in Louisville, is black. There is also a significant difference between usage and expected usage of local telephone service across these two local exchanges. While the number of calls is higher in Louisville than in Bowling Green, the expected consumption is much more accurate in the latter exchange. On average, Bowling Green residents underestimate telephone usage by 2% and Louisville residents underestimate their usage by 29%. The difference in magnitude of the bias (type shock of the model) is remarkable. Perhaps it could be explained by positive network effects of the size of the local exchanges [Taylor (1994, §9)]. Bowling Green barely reached 50,000 inhabitants by the end of the 1980's but Louisville had a population that exceeded 250,000.

5.1 Are Data Consistent with the Type–Varying Model?

A common problem in estimating demand when consumers face nonlinear budgets is that the choices of consumption and the marginal tariff are simultaneous and therefore the relevant price is endogenous [*e.g.*, MacKie–Mason and Lawson (1993, §3.2)]. Regarding this point, observe that comparing the expected weekly number of calls with the actual number of calls during the spring months is qualitatively different from comparing those expectations with the actual number of calls during the fall months. In the second case the number of phone calls is a function of the tariff chosen (in Louisville) and the marginal charge per call, which varies with the time of the day and distance of the outgoing call (both in Bowling Green and Louisville's measured service option), as well as of customers' accumulated monthly usage of telephone services. However, this is not the case during the spring months because all local telephone customers were placed under a mandatory flat rate regime. Price was a relevant economic variable for the decision to subscribe the telephone service, but after that any additional call involves a zero marginal charge, and consequently local telephone customers should consume at their satiation levels.

Focusing on the spring months of 1986, the present data set provides us with an uncommonly available direct indicator for θ_1, the expected number of weekly calls, and also for θ, the actual number of weekly calls. The first column of Table 2 shows the average usage expectation bias, μ_2, which is positive for customers of these two local exchanges, but it is about seventeen times larger in Louisville than in Bowling Green. A more detailed analysis by demographic strata shows further differences between residents of these two exchanges. While in Louisville the bias is always positive and large, independently of the demographic characteristic considered, in Bowling Green it is more balanced and in several occasions it takes negative values. In both cities consumers tend to underestimate their future usage, but in Louisville they do it by more than an order of magnitude. The smaller average bias in Louisville (single and male household) is still more than seven times larger than the average bias in Bowling Green. Figures 1.7–1.8 show the empirical density function of type shocks. Although these expectation bias are quite disperse, small mistakes around the mean are the most frequent event. The "PAT" column presents further evidence in favor of the type varying model by computing Pearson's analog goodness of fit test for the equality of $F(\cdot)$ and $F_1(\cdot)$. That hypothesis is always strongly rejected and therefore we can conclude that the distribution of θ_2 is not degenerate and that the suggested type varying model is an accurate representation of consumers preferences.

But there is also significant heterogeneity by demographic strata for local telephone usage expectation bias. In both cities there is evidence (stronger in Louisville) in favor of a mean increasing spread of the distribution of θ relative to that of θ_1. However, a systematic ordering of the means of θ and θ_1 (through a positive μ_2) is not sufficient to ensure the stochastic dominance of θ over θ_1, since the whole distribution matters. Figures 1.1–1.2 present the empirical frequency distributions of actual and expected weekly number of local calls for the spring months of the experiment in the local exchanges of Bowling Green and Louisville respectively. It is evident that the distribution of expected weekly calls is characterized by the accumulation of frequencies on a few "focal points" of the usage range. More informative is the empirical cumulative distribution functions shown in Figures 1.3–1.4, which clearly indicates that in both cities telephone customers tend to underestimate their future local telephone usage, which leads to the relative ordering of the averages of θ and θ_1 discussed in Table 2.

If the mean of θ_2 is finite, –*i.e.*, $E[\theta_2] = \mu_2$–, and given that θ is additively separable in θ_1 and θ_2, –*i.e.*, equation (11)–, $F(\cdot)$ will be a stochastic spread of $F_1(\cdot)$, and therefore θ_1 should second order stochastically dominate θ. The stochastic spread will be mean preserving, mean increasing, or mean decreasing depending on whether μ_2 is zero, positive, or negative. Figure 1.4 appears to indicate that θ first order stochastically dominates θ_1 in Louisville, although Figure 1.3 fails to prove the same for Bowling Green. While first order stochastic dominance (FOSD) implies second order stochastic dominance (SOSD) and therefore supports the suggested type–varying model, FOSD is also much more restrictive than SOSD because it implies that consumers systematically underestimate their future consumption, not only independently of their demographic characteristics, but also independently of the magnitude of their local telephone usage. Thus, strong FOSD of θ over θ_1 will be consistent with a model where θ dominates in hazard rate to θ_1.

In order to test the hypotheses of FOSD and SOSD, I computed Anderson's (1996) nonparametric test of stochastic dominance. The test is based on comparing weighted differences of frequency functions of two variables within given mutually exclusive fractiles. For each demographic strata, stochastic dominance of any order is rejected if one ratio is significantly positive for any single fractile. Thus, Table 3 reports for each demographic strata the maximum of these ratios among 20 fractiles in which the range of phone calls is divided. Table 3 provides with strong evidence in favor of the suggested type–varying model, as SOSD of θ over θ_1 is only rejected for two demographic categories in Bowling Green for very large consumption ranges (exceeding 90 calls per week).[15] FOSD of θ over θ_1 is generally rejected in Bowling Green but never in Louisville. This FOSD result explains why the type shock has always a positive mean in Louisville, and its rejection in Bowling Green is consistent with the negative average bias found for some demographic strata.

6 Welfare Analysis

The monopolist has a choice between the *ex–post* and the *ex–ante tariff*. If we approach the problem from an *ex–post* perspective, the monopolist should always prefer the *ex–post* tariff to an *ex–ante* tariff made of two–part tariff options because in this latter case the monopolist is not screening consumers with respect to θ_2. The same comparison with a menu of nonlinear options is however not so straightforward because in this case both type components are used in the design of the tariff, but the monopolist screens them sequentially rather than simultaneously.

However, the monopolist, as well as the regulator, has to evaluate the choice among alternative ways to screen consumers *ex–ante*. As it was shown before, under sequential screening, consumers' expectations affect the IC and participation constraints, and thus integrating out the effect or θ_2 still affects the shape of the *ex–ante* tariff.

[15] I furthermore checked that SOSD was never rejected for neither of the two cities in any single month, using 10 and 15 fractiles.

In the end, the monopolist will prefer the tariff that introduces higher markups for every consumption level, thus reducing the informational rent. This may happen for certain environments as those discussed in Section 4.2 where distributions can be ordered according to their hazard rates. But most likely, the lower envelopes of the tariffs $\hat{T}(\theta)$ and $\tilde{T}(\theta_1)$ will cross each other, defining regions for which one leads to higher profits than the other. Thus, the evaluation of expected profit or welfare gains from the introduction of optional pricing becomes difficult to characterize in general.

6.1 Freedom of Choice vs. Mandatory Pricing

In this section I analyze the role of the distribution of asymmetric information parameters in the solution of the optimal *ex–post* nonlinear pricing problem. Suppose that consumer type θ could be distributed with respect to either $F(\theta)$ or $G(\theta)$. Which of these two distributions is more informative for the monopolist? The following Proposition summarizes a well known result in the mechanism design literature:

PROPOSITION 3: *Let $F(\theta)$ and $G(\theta)$ be* IHR, *i.e., such that $r'_F(\theta) > 0$ in θ on $\{\theta > 0 : F(\theta) < 1\}$, and $r'_G(\theta) > 0$ in θ on $\{\theta > 0 : G(\theta) < 1\}$. Assume also that $r_F(\theta) \leq r_G(\theta)$, $\forall\theta$. Then, the price mark–up and the marginal tariff will be uniformly higher under the $F(\theta)$ distribution than under the $G(\theta)$ distribution.*

PROOF: Differentiation of the price mark–up using solution (7) leads to:

$$\frac{\partial}{\partial r(\theta)}\left(\frac{\hat{p}(\theta)-c}{\hat{p}(\theta)}\right) = \frac{c \cdot \dfrac{v_{p\theta}(p(\theta),\theta)}{v_{pp}(p(\theta),\theta)}}{r^2(\theta)\left[c - r^{-1}(\theta)\cdot\dfrac{v_{p\theta}(p(\theta),\theta)}{v_{pp}(p(\theta),\theta)}\right]^2} < 0. \tag{33}$$

Thus, evaluating these expresions at $r_F(\theta)$ and $r_G(\theta)$ respectively, it follows that the price mark–up will be higher under $F(\theta)$ than under $G(\theta)$. ∎

The monopolist will charge a higher price mark–up under $F(\cdot)$ than with $G(\cdot)$ because $F(\cdot)$ dominates in hazard rate to $G(\cdot)$. Distribution $F(\cdot)$ puts more weight on consumers of high type.[16] The fact that $F(\cdot)$ is more favorable than $G(\cdot)$ implies that the optimal pricing has to create stronger incentives for inframarginal consumers to self–select according to their true type. Proposition 3 shows that maintenance of the IC constraint under more favorable distributions requires higher price distortions for inframarginal consumer types. Thus, reducing consumers' expected informational rents, the monopolist is able to screen among the many particular type values that a smaller proportion of his customers can reveal through their tariff choice or usage decisions.

Information structures that lead to the hazard rate ordering provide with a unique case where different nonlinear tariffs can be sorted. But more frequently, comparison

[16] This is true strictly speaking when the support of the distribution is restricted to $\Re_+$. Otherwise, hazard rate dominance is a regularity condition that has to be imposed to obtain this result.

among informational structures will not lead to situations in which one distribution is more favorable than the other over the whole support of the distribution of types. This is the case of the present case study. Figures 1.5–1.6 compare the hazard rate of the distributions of expected and actual number of calls in the two local exchanges. In both cases, there is an alternating dominance in hazard rate of $F(\cdot)$ over $F_1(\cdot)$ and *vice versa.* But in Louisville, consistent with the FOSD result, $F_1(\cdot)$ dominates in hazard rate to $F(\cdot)$ only in very small ranges. Thus, without strict hazard rate dominance, markups can be higher under one tariff only for a given range of consumption. The question that remains to be answered is whether, regardless of all these issues, something can be said about the desirability of *ex–ante vs. ex–post* tariffs for different agents.

Results for the monopolist are conclusive. Optimal tariff functions $T(\theta)$ are necessarily increasing, $T'(\theta) = p(\theta) > 0$. Furthermore, if the problem is well behaved, tariff functions will be concave (quantity discounts), $T''(\theta) = p'(\theta) < 0$. The monopolist generally expects an increase in profits by introducing optional pricing as the following proposition shows. These results are a direct consequence of the classical conditions of Hadar and Russell (1969) to order outcomes under uncertainty.

PROPOSITION 4: *Expected profits are higher under ex–ante pricing if any of the following conditions hold:*
(i) $T'(\theta) > 0$ *and* $F(\cdot)$ *FOSD* $F_1(\cdot)$,
(ii) $T'(\theta) > 0$, $T''(\cdot) < 0$, *and* $F(\cdot)$ *SOSD* $F_1(\cdot)$.

PROOF: Under circumstances of part *(i)*, the difference of expected profits between *ex–post* and *ex–ante* tariffs is (integrating by parts):

$$\int_{\Theta} T(x)[F(x) - F_1(x)]dx = -\int_{\Theta} T'(x)[F(x) - F_1(x)]dx \geq 0, \tag{34}$$

while for part *(ii)* the result is obtained integrating (34) by parts again:

$$\int_{\Theta} T''(x) \int_{\Theta} [F(y) - F_1(y)]dydx - T'(x) \int_{\Theta} [F(y) - F_1(y)]dy\Big|_{x=\underline{\theta}}^{x=\overline{\theta}} \geq 0, \tag{35}$$

which completes the proof. ■

Therefore, more favorable distributions (FOSD) increase expected profits even for cases where the pricing problem does not fulfill all required conditions to discriminate among consumers by means of quantity discounts. But if these quantity discounts are optimal, then less restrictive stochastic orderings (SOSD) also lead to the same conclusion. The commonly observed practice of using optional nonlinear tariffs is therefore profit maximizing under very general conditions, which should suffice to explain its widespread use.

Unfortunately, I cannot affirm the same about consumers. Assumption 2 only requires that the indirect utility function be increasing in θ. But the effect on the net rent $v(p(\theta), \theta) - T(\theta)$ will depend on many factors. If $v(\cdot)$ is more increasing than $T(\cdot)$, then part *(i)* of Proposition 4 could be applied, and consumers will prefer optional pricing to mandatory *ex–post* pricing. However, it is possible to observe that type shocks are so biased that $F(\cdot) \leq F_1(\cdot)$ and consumers still prefer the mandatory measured service. In this case preferences just fail to be increasing enough in θ. This appears to happen in the empirical analysis of the Louisville sample. A similar analysis could be made for the case of SOSD in order to apply part *(ii)* of Proposition 4. In addition to $v(\cdot)$ being more increasing than $T(\cdot)$, it would now require that $v(\cdot)$ is more concave than $T(\cdot)$. Thus, even more restrictive preferences are necessary to obtain a definite ordering of pricing strategies under increasingly less restrictive stochastic environments.

Obviously, this difficulty in ordering pricing schemes according to the expected consumer surplus is translated to the Regulator's welfare measure. The result is more unclear the more weight is given to consumers in the Regulator's objective function. But still something can be said, at least in limiting cases. For instance, if $\sigma_2^2 = 0$ then $F(\theta) = F_1(\theta_1)$. Obviously, if the variance of the shock is zero, all consumer differences are captured by the distribution of the *ex–ante type*. Therefore, there is no real distinction between tariff choice and usage decision. All consumers and also the monopolist would be indifferent between an *ex–ante* and *ex–post* tariffs. A model like this is actually equivalent to one where consumers are able to commit *ex–ante* to future consumption. Without being too precise I conclude that the smaller is the variance of the shock relative to the variance of the ex–ante type, the more likely is that *ex–ante* tariffs are welfare increasing and *vice versa*. To confirm this intuition, the other extreme case should also be analyzed. If $\sigma_1^2 = 0$ then $F(\theta) = F_2(\theta_2)$. In this case, consumers only differ *ex–post*. Since consumers are all alike *ex–ante*, the optimal *ex–ante* tariff will be a single two–part tariff. But that is not the welfare enhancing in expectation (neither profit maximizing) tariff because consumers will be considerably more diverse *ex–post* than *ex–ante*, and both welfare and profits are increasing in the number of self–selecting tariffs [Faulhaber and Panzar (1977, §4); Wilson (1993, §8)], which implies that the optimal strategy should be an *ex–post* based fully nonlinear tariff. However, the discussion falls short of determining the threshold levels of the ratio σ_1^2/σ_2^2 that make *ex–ante* pricing dominate *ex–post* tariffs or *vice versa*. Determinants of this threshold are application specific, and will critically depend on the specification of the utility function and the distributions used.

Thus, the stochastic nature of type shocks, whether they make the *ex–post* distribution more favorable (hazard rate effect) or whether consumers become more or less heterogeneous after learning θ_2 (variance effect) appear to drive welfare results in particular applications. These relationships are complex enough to make precise predictions almost impossible. Instead of solving the model for one of such *ad hoc* cases that avoids ambiguity, the following section explores the magnitude of the welfare effects associated to different pricing strategies using empirical distributions of θ, θ_1, and θ_2 as identified in the Kentucky tariff experiment, which is a more interesting analysis because direct observations of types are rarely available.

6.2 Simulations

In this subsection I evaluate the average expected consumer surplus, profits (revenues), and total welfare of screening local telephone customers through either a mandatory *ex–post* pricing, a continuum of optional two–part tariffs, or a continuum of fully nonlinear options. In order to run the Monte Carlo simulations, I assume that the indirect utility function is:

$$V(p, A, \theta) = \frac{\theta}{\alpha} \exp[-\alpha p] - A \quad ; \quad \alpha > 0, \tag{36}$$

which leads to the following demand equation:

$$x(p, \theta) = \theta \exp[-\alpha p]. \tag{37}$$

This specification has been used before in telecommunications demand analysis because it is bounded under the flat rate option. If $p = 0$, consumers purchase their satiation level $x(0, \theta) = \theta$. Similarly, when $p = 0$, the expected usage equals $E_2[\theta] = \theta_1 + \mu_2$. The solutions of $\hat{T}(\theta)$, $\tilde{T}(\theta_1)$, and $\tilde{\tilde{T}}(\theta_2 \mid \theta_1)$ for this particular demand function and general distributions, are shown in the Appendix. Nonlinear pricing solutions based on (37) are constructed under the assumption that the monopolist considers only the possibility of allowing for call discounts, instead of price discrimination based on duration of the call, time of the day, distance, or any other criteria.[17] Table 4 presents the results of evaluating these tariffs and their associated welfare effects for the two Kentucky local exchanges where the tariff experiment was conducted.[18]

Kernel estimates are shown in Figures 2–3 for the two local exchanges of Bowling Green and Kentucky. There are not many differences among the distributions of these exchanges. In both cases, the estimates identify important focal points around 50 and 100 expected calls per week respectively. The hazard rate of all distributions can be considered increasing. The first increasing portions of the different hazard rates of Figures 2 and 3 account for most of the corresponding mass of probability. As all increasing variances of the kernel estimates of the hazard rates show, most variations in $\hat{r}(\cdot)$ after the initial increasing section are most likely due to purely random effects than to genuine increases or decreases of the hazard rate for particular regions.

The values of θ and θ_1 are identified as the actual and expected number of calls during the spring months, when consumers faced a zero marginal charge. The existence of

[17] The available data does not identify any effect other than potential volume discounts based on the total number of calls. This is because it includes expectations for total number of weekly calls during the spring months when the effective marginal tariff is zero, but it does not include anything regarding expected duration of calls or average time/distance profile of these calls.

[18] I compute an adaptive Gaussian kernel with optimal bandwidth chosen to minimize the mean integrated square error of the estimation of the distributions of θ, θ_1, and θ_2 (actual or expected calls and estimation bias respectively) corresponding to each local exchange. The estimation procedure discretizes the ranges of θ, θ_1, and θ_2 around a 128 point grid to obtain the kernel estimation of each density by means of a fast Fourier transform. Estimation of $f(\cdot)$ and $F(\cdot)$ for intermediate values of θ, θ_1, or θ_2 is obtained by polynomial interpolation (with all 128 point estimates of the kernel) using Neville's algorithm.

a positive charge per call could lead to a selection effect in Louisville where the flat rate was still an option later in the fall, and/or a suppression effect in Bowling Green (mandatory measured) and Louisville (optional measured) due to the negative slope of demand. During the fall months in which these tariffs applied, customers in Bowling Green made 134.33 local calls on average every month. This number identifies the number of calls of the *ex–post* tariff in my base case for Bowling Green since it already includes the effect of a positive marginal tariff. In Louisville this number is significantly higher as it averages the number of calls of 10% of the customers on optional measured service, 86.69, and the 189.28 monthly calls of the remaining 90% of customers on optional flat rate service in that exchange. The value of 179.02 is therefore used in the base case to identify the volume of demand under the *ex–ante* pricing regime in Louisville.

The price elasticity of demand function (37) is given by $\varepsilon = -\alpha p$. Therefore, for any marginal rate, it is always possible to modify the value of α in order to fit a demand function with the desired value of the elasticity. The simulations are run for four values of price elasticity (evaluated at the average p) as reported in four independent empirical studies of local telephone demand: -0.1 [Park, Wetzel, and Mitchell (1983)], -0.17 [Kling and Van Der Ploeg (1990)], -0.45 [Train, McFadden, and Ben–Akiva (1987)], and -0.7 [Hobson and Spady (1988)]. Because of the richness of the data available, the estimate of Park, Wetzel, and Mitchell is probably the most accurate. I however decided to choose $\varepsilon = -0.17$ for the base case common to the two cities because this number allows for comparisons with other situations, both with higher and lower elasticities of demands. But also, and more importantly, because Park, Wetzel, and Mitchell (1983, §5) acknowledge that demand elasticities are price dependent and they indicate that elasticities could likely be higher for higher prices (as in my case). After comparing local tariffs and telephone usage patterns in the two local exchanges, I chose an average cost per call of 7 cents as representative for the base case of the simulations.

Table 4 evaluates each particular nonlinear pricing solution and its associated welfare magnitudes: consumer surplus V, profits π, and total welfare W. All average values of simulations in Table 4 are shown in 1986 dollars per month. Reported simulations are the average of 10,000 independent draws from the kernel estimation of the empirical distribution of types. I focus on the case where $\varepsilon = -0.17$. Thus, in Bowling Green, the optimal *ex–post* tariff involves an average marginal rate of \$0.07, and an average monthly fee of \$44.07. Given the empirical distributions of types in that local exchange, consumers enjoy an average expected money surplus of \$11.25, the local monopolist expects to make \$44.92 in profits per customer, and total expected welfare amounts to \$56.17 per person.

Average monthly fees are slightly higher under optional pricing than with the standard *ex–post* nonlinear tariffs, although almost no distinction is found between optional two–part tariffs and optional nonlinear tariffs. Marginal rates are 34% lower with optional two part tariffs than with *ex–post* pricing while under optional nonlinear tariffs they rise 18%. These are however average magnitudes. Thus, the higher consumption under optional nonlinear tariffs relative to optional two–part tariff could be explained by a likely reduction in the average marginal tariff under optional nonlinear tariff relative to optional two part tariffs as consumption increases for each chosen tariff. This increase in consumption explains the 14% increase in expected consumer surplus under optional nonlinear tariff due to a 5% expansion of demand relative to *ex–post* pricing, as compared to the 1% expansion induced by optional two–part tariffs.

Introduction of optional two–part tariffs enhances welfare by about 2%, mostly due to a 4% increase in profits, because consumer surplus is reduced by 4% (of an initial smaller amount). Optional nonlinear tariffs reduce welfare by 5%, but the distribution of its components is quite different from the two–part tariff options case. The effect of the reduction of marginal rates for large consumers under optional nonlinear tariffs dominates, and thus consumers benefit more from the introduction of nonlinear options than from the introduction of optional two–part tariffs, although the latter one is the welfare maximizing pricing policy in expectation among the three analyzed here.

Finally, all magnitudes considered (with the exception of consumption) are inversely related to the absolute value of the elasticity of demand. Thus, the more inelastic is the demand, the higher is the average fixed fee as well as the average marginal tariffs. But also the average expected consumer surplus, profits and total welfare. The welfare analysis carried out before for the reference scenario when $\varepsilon = -0.17$ is also valid for the others, so that the conclusion of optional two–part tariffs being the preferred pricing option appears to be robust to different values of the elasticity of demand.

For the case of Louisville, the reference case of two–part tariff options is characterized again with an average marginal rate of \$0.07, and the average monthly fee of \$63.59.[19] Individual expected consumer surplus is \$10.03, expected profits per customer are \$65.00, and total expected welfare amounts to \$75.03 per person. The welfare analysis of the results of Louisville is very similar to that one of Bowling Green. There are two sources of differences between these two exchanges that affect the results of simulations. First, consumption pattern may vary due to differences in demographics, socioeconomic variables, tariff options, and/or the size of the local network. The effect of all these variables have already been captured through the identification of exchange specific levels of telephone usage under different tariff regimes. The other source is the disparate behavior of type shocks in these two cities. Systematic underestimation of future consumption is the origin of the wider effects of welfare in Louisville relative to Bowling Green when comparing pricing alternatives. Thus, for instance, for the $\varepsilon = -0.17$ scenario, going from *ex–post* pricing to optional two–part tariffs reduces the expected consumer rents by 4% and increases expected profits by 4% in Bowling Green, while in Louisville the expected consumer surplus reduction is about 20% and the increase in expected profits reaches 7%. However, optional two–part tariffs are again the welfare maximizing among the pricing strategies considered here.

[19] Observe that average marginal rates are normalized to \$0.07 both for the optional two–part and fully nonlinear option cases. Since consumption (independent of ε) is also normalized across scenarios, the average marginal rate is always the same for these two alternative pricing strategies.

Observe that the simulation results regarding differences of expected profits are in accordance to the theoretical results of the previous sections, and of Proposition 4 in particular. Welfare increases in expectation when we implement optional two–part tariffs instead of *ex–post* nonlinear pricing. The SOSD of θ over θ_1 is the dominant factor driving this result. The FOSD of Louisville, with mean increasing effect on the usage level accounts for the stronger magnitude of the increase of expected profits (7% in Louisville *vs.* 4% in Bowling Green). Finally, the additional 4% increase in profits obtained when nonlinear tariff option are in use instead of optional two–part tariffs should be explained by the monopolist being able to discriminate consumers also with respect to θ_2 instead of just θ_1. Expected profits increase as the number of options increases and accounts for *ex–post* differences.

7 Conclusions

Optional nonlinear pricing has not attracted much attention among economists until very recently. Traditionally, economists have incorrectly extended the application of results of the standard nonlinear pricing theory to situations where consumption and tariff choice were not simultaneous. The early treatment of Clay, Sibley, and Srinagesh (1992) studied the design of optimal two–part tariffs, but restricted their attention to the discrete type case. They also limited drastically the range of variation of θ_2 to ensure that the same SCP held both *ex–ante* and *ex–post*, so that the ordering of individual consumer preferences remained unaltered after the realization of the shock. Miravete (1996) extended this model to the case of a continuum of two–part tariff options with a continuum of types, independently of whether the ordering of consumer tastes changed or not after the realization of the shock. Miravete (2000b) used a particular closed form solution of this model to analyze the estimation bias of not dealing with asymmetric information and self–selection issues in a cross–section framework. Finally, Courty and Li (2000) analyzed a general model of sequential screening with a continuum of types but limiting the analysis to consumers with unit demands and biased type shocks in the sense of FOSD.

Relative to all these works, the present paper contributes by characterizing a fully nonlinear tariff when consumers buy more than one unit, and by making explicit the role of the statistical assumptions on the existence of quantity discounts (IHR of the distribution of type components), and welfare effects (FOSD and SOSD of θ over θ_1). This paper also compares different optimal nonlinear tariffs depending on whether they are designed *ex–ante* or *ex–post*, through the preservation of the IHR property of the distribution of type components through convolution. Finally, the paper also contributes to this literature by providing very strong evidence in favor of the suggested type–varying model based on direct observation of consumer types. Furthermore, using simulations from the kernel distributions of these types, the paper reports results that favor optional two–part tariffs as the welfare maximizing strategy in two local exchanges of Kentucky.

References

Anderson, G. (1996): "Tests of Stochastic Dominance in Income Distributions." *Econometrica*, 64, 1183–1194.

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

Baron, D.P. and D. Besanko (1984): "Regulation and Information in a Continuing Relationship." *Information Economics and Policy*, 1, 267–302.

Baron, D.P. and D. Besanko (1999): "Informational Alliances." *Review of Economic Studies*, 66, 743–768.

Biais, B., D. Martimort, and J.C. Rochet (2000): "Competing Mechanisms in a Common Value Environment." *Econometrica*, 68, 799–837.

Caillaud, B., R. Guesnerie, and P. Rey (1992): "Noisy Observation in Adverse Selection Models." *Review of Economic Studies*, 59, 595–615.

Clay, K., D.S. Sibley, and P. Srinagesh (1992): "Ex Post *vs.* Ex Ante Pricing: Optional Calling Plans and Tapered Tariff." *Journal of Regulatory Economics*, 4, 115–138.

Courty, P. and H. Li (2000): "Sequential Screening." *Review of Economic Studies*, 67, 697–717.

Faulhaber, G.R. and J.C. Panzar (1977): "Optimal Two–Part Tariffs with Self–Selection." Bell Laboratories Economic Discussion Paper No. 74.

Hadar, J., and W.R. Russell (1969): "Rules for Ordering Uncertain Prospects." *American Economic Review*, 59 25–34.

Hobson, M. and R.H. Spady (1988): "The Demand for Local Telephone Service Under Optional Local Measured Service." Bellcore Economics Discussion Paper No. 50.

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

Karlin, S. (1968): *Total Positivity*, Vol. I. Stanford University Press.

Kling, J.P. and S.S. van der Ploeg (1990): "Estimating Local Call Elasticities with a Model of Stochastic Class of Service and Usage Choice," in A. de Fontenay, M.H. Shugard, and D.S. Sibley (eds.): *Telecommunications Demand Modelling*. North–Holland.

Kridel, D., D. Lehman, and D. Weisman (1993): "Option Value, Telecommunications Demand, and Policy." *Information Economics and Policy*, 17, 69–75.

Laffont, J.J. and J. Tirole (1986): "Using Cost Observations to Regulate Firms." *Journal of Political Economy*, 94, 614–641.

Laffont, J.J. and J. Tirole (1993): *A Theory of Incentives in Procurement and Regulation*. MIT Press.

MacKie–Mason, J.K. and D. Lawson (1993): "Local Telephone Calling Demand when Customers Face Optimal and Nonlinear Price Schedules." Working Paper. Department of Economics. University of Michigan.

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

Miravete, E.J. (1996): "Screening Consumers Through Alternative Pricing Mechanisms." *Journal of Regulatory Economics*, 9, 111–132.

Miravete, E.J. (2000a): "Choosing the Wrong Calling Plan? Ignorance, Learning, and Risk Aversion." CEPR Discussion Paper No. 2562.

Miravete, E.J. (2000b): "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans." CEPR Discussion Paper No. 2635.

Miravete, E.J. (2001): "On Preservation of Increasing Hazard Rate Under Convolution." CARESS Working Paper No. 01–05. Department of Economics, University of Pennsylvania.

Mitchell, B.M. and I. Vogelsang (1991): *Telecommunications Pricing. Theory and Practice*. Cambridge University Press.

Park, R.E., B.M. Wetzel, and B.M. Mitchell (1983): "Price Elasticities for Local Telephone Calls." *Econometrica*, 51, 1699–1730. Picard, P. (1987): "On the Design of Incentive Schemes under Moral Hazard and Adverse Selection." *Journal of Public Economics*, 33, 305–331.

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

Srinagesh, P. (1992): "A Dynamic Stochastic Model of Choice." Bellcore Economics Discussion Paper No. 78.

Stoline, M.R. and H.K. Ury (1979): "Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons Among Means." *Technometrics*, 21, 87–93.

Taylor, L.D. (1994): *Telecommunications Demand in Theory and Practice*, 2nd edition. Kluwer Academic Publishers.

Train, K.E., D.L. McFadden, and M. Ben–Akiva (1987): "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices." *Rand Journal of Economics*, 18, 109–123.

Train, K.E., M. Ben–Akiva, and T. Atherton (1989): "Consumption Patterns and Self–Selecting Tariffs." *The Review of Economics and Statistics*, 50, 62–73.

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

Wilson, R.B. (1995): "Nonlinear Pricing and Mechanism Design," in H. Amman, D. Kendrick, and J. Rust (eds.): *Handbook of Computational Economics*, Vol. I. North–Holland.

Wolak, F. (1996): "Estimating Regulated Firm Production Functions with Private Information: An Application to California Water Utilities," *Annales d'Economie et de Statistique*, 34, 13–69.

Appendix

• Proof of Proposition 1

First note that if any distribution function $F_i(\theta_i)$ is IHR, this is equivalent to the corresponding survival function $1 - F_i(\theta_i)$ being log concave:

$$\frac{\partial^2 \log[1 - F_i(\theta_i)]}{\partial \theta_i^2} = \frac{\partial}{\partial \theta_i}\left[\frac{-f_i(\theta_i)}{1 - F_i(\theta_i)}\right] \leq 0. \tag{A.1}$$

Second, note that by Assumption 3, the survival function is continuously differentiable. Therefore, it is a Pólya Frequency function of order 2 (PF_2). Survival function $1 - F_i(\theta_i)$ is PF_2 if $\forall x_1 < x_2 \in X \subseteq \Re$ and $\forall y_1 < y_2 \in Y \subseteq \Re$:

$$\begin{vmatrix} 1 - F_i(x_1 - y_1) & 1 - F_i(x_1 - y_2) \\ 1 - F_i(x_2 - y_1) & 1 - F_i(x_2 - y_2) \end{vmatrix} \geq 0. \tag{A.2}$$

To see the equivalence, assume without loss of generality that $x_1 < x_2$ and $0 = y_1 < y_2 = \Delta$. Then, from the definition of PF_2 and making use of common properties of determinants, the following equivalent inequalities hold:

$$\begin{vmatrix} 1 - F_i(x_1) & 1 - F_i(x_1 - \Delta) \\ 1 - F_i(x_2) & 1 - F_i(x_2 - \Delta) \end{vmatrix} \geq 0, \tag{A.3}$$

$$\Delta \cdot \begin{vmatrix} \dfrac{1 - F_i(x_1) - [1 - F_i(x_1 - \Delta)]}{\Delta} & 1 - F_i(x_1 - \Delta) \\ \dfrac{1 - F_i(x_2) - [1 - F_i(x_2 - \Delta)]}{\Delta} & 1 - F_i(x_2 - \Delta) \end{vmatrix} \geq 0. \tag{A.4}$$

Since $\Delta > 0$, we can take limits in the latter determinant to obtain:

$$\lim_{\Delta \to 0} \begin{vmatrix} \dfrac{1-F_i(x_1)-[1-F_i(x_1-\Delta)]}{\Delta} & 1-F_i(x_1-\Delta) \\ \dfrac{1-F_i(x_2)-[1-F_i(x_2-\Delta)]}{\Delta} & 1-F_i(x_2-\Delta) \end{vmatrix} = \begin{vmatrix} -f_i(x_1) & 1-F_i(x_1) \\ -f_i(x_2) & 1-F_i(x_2) \end{vmatrix} \geq 0, \tag{A.5}$$

which leads to:

$$\frac{f_i(x_1)}{1 - F_i(x_1)} \leq \frac{f_i(x_2)}{1 - F_i(x_2)}, \tag{A.6}$$

i.e., $F_i(\cdot)$ is IHR. Thus, I have to prove that the survival function of the convolution distribution is log–concave, *i.e.*, for $x_1 < x_2$ and $y_1 < y_2$:

$$D = \begin{vmatrix} 1 - F(x_1 - y_1) & 1 - F(x_1 - y_2) \\ 1 - F(x_2 - y_1) & 1 - F(x_2 - y_2) \end{vmatrix} \geq 0. \tag{A.7}$$

Applying Definition 2 of the Fourier convolution to the survival function we get:

$$D = \begin{vmatrix} \int[1-F_1(x_1-z)]f_2(z-y_1)dz & \int[1-F_1(x_1-z)]f_2(z-y_2)dz \\ \\ \int[1-F_1(x_2-z)]f_2(z-y_1)dz & \int[1-F_1(x_2-z)]f_2(z-y_2)dz \end{vmatrix} \geq 0. \qquad (A.8)$$

Using the commutative property of convolutions:

$$\int F_1(x-z)f_2(z-y)dz = \int f_1(x-z)F_2(z-y)dz, \qquad (A.9)$$

equation $(A.6)$ becomes:

$$D = \begin{vmatrix} \int[1-F_1(x_1-z)]f_2(z-y_1)dz & \int f_1(x_1-z)[1-F_2(z-y_2)]dz \\ \\ \int[1-F_1(x_2-z)]f_2(z-y_1)dz & \int f_1(x_2-z)[1-F_2(z-y_2)]dz \end{vmatrix} \geq 0. \qquad (A.10)$$

The final step involves the application of the *Basic Composition Formula* to convolutions [Karlin (1968, §1.2)]:

$$D = \int\limits_{z_1<}\int\limits_{z_2} \begin{vmatrix} 1-F_1(x_1-z_1) & f_1(x_1-z_2) \\ 1-F_1(x_2-z_1) & f_1(x_2-z_2) \end{vmatrix} \cdot \begin{vmatrix} f_2(z_1-y_1) & 1-F_2(z_2-y_1) \\ f_2(z_1-y_2) & 1-F_2(z_2-y_2) \end{vmatrix} dz_1 dz_2 \geq 0. \qquad (A.11)$$

Observe that for this last expression to be positive and thus ensure that the distribution $F(\cdot)$ is IHR, each determinant has to be positive. Assuming without loss of generality that $0 = z_1 < z_2 = \Delta$, the condition that the first determinant is positive requires that:

$$[1-F_1(x_1)]f_1(x_2-\Delta) - [1-F_1(x_2)]f_1(x_1-\Delta) \geq 0, \qquad (A.12)$$

which implies:

$$\frac{f_1(x_2-\Delta)}{1-F_1(x_2-\Delta)} \cdot \frac{1-F_1(x_2-\Delta)}{1-F_1(x_2)} \geq \frac{f_1(x_1-\Delta)}{1-F_1(x_1-\Delta)} \cdot \frac{1-F_1(x_1-\Delta)}{1-F_1(x_1)}. \qquad (A.13)$$

But since $\Delta > 0$ and $x_1 < x_2$:

$$\frac{f_1(x_2-\Delta)}{1-F_1(x_2-\Delta)} \geq \frac{f_1(x_1-\Delta)}{1-F_1(x_1-\Delta)}, \qquad (A.14)$$

which is just the hypothesis that $F_1(\cdot)$ is IHR. Similarly, comparing the other elements of inequality $(A.13)$:

$$\frac{1-F_1(x_2-\Delta)}{1-F_1(x_2)} \geq \frac{1-F_1(x_1-\Delta)}{1-F_1(x_1)}, \qquad (A.15)$$

which is equivalent to:

$$\begin{vmatrix} 1-F_1(x_1) & 1-F_1(x_1-\Delta) \\ \\ 1-F_1(x_2) & 1-F_1(x_2-\Delta) \end{vmatrix} \geq 0, \qquad (A.16)$$

that is the condition for the survival function $1-F_1(\cdot)$ to be log–concave, which we have proved to be equivalent to the assumption of $F_1(\cdot)$ being IHR. A similar argument proves that if $F_2(\cdot)$ is IHR, the second determinant in inequality $(A.11)$ is also positive. Thus, $F(\cdot)$ is IHR. ■

- **Proof of Proposition 2**

Suppose not, *i.e.*, for instance assume that for any common θ, $r_1(\theta) < r(\theta)$, that is:

$$\frac{f_1(\theta)}{1-F_1(\theta)} < \frac{f(\theta)}{1-F(\theta)}. \tag{A.17}$$

Using the definition of Fourier convolution, this inequality is equivalent to the following two inequalities:

$$f_1(\theta)[1-F(\theta)] - f(\theta)[1-F_1(\theta)] < 0, \tag{A.18}$$

$$\int_0^\infty \Big(f_1(\theta)[1-F_1(\theta-z)] - [1-F_1(\theta)]f_1(\theta-z)\Big) f_2(z)dz < 0. \tag{A.19}$$

Since $f_2(\theta) \geq 0$ on $0 \leq \theta < \infty$, it must be the case that the term between brackets is negative $\forall\theta \geq 0$. But observe that this condition then requires:

$$\frac{f_1(\theta)}{1-F_1(\theta)} \leq \frac{f_1(\theta-z)}{1-F_1(\theta-z)} \qquad \forall z \geq 0, \tag{A.20}$$

so that $F_1(\cdot)$ should be decreasing hazard rate. Similarly, $r_2(\theta) < r(\theta)$ violates $F_2(\cdot)$ being IHR. Contradiction. ▪

- **Tariff Solutions for Exponential Demand**

It is assumed that $c = 0$. This parameter only changes the scale of the marginal charge, but the comparisons of expected welfare under pricing regimes remain unaffected. Thus, the optimal *ex–post* tariff is characterized by:

$$\hat{p}(\theta) = \frac{1}{\alpha\theta r(\theta)}, \tag{A.21}$$

$$\hat{A}(\theta) = \frac{1}{\alpha}\left\{\theta \exp[-\{\theta r(\theta)\}^{-1}] - \int_{\underline{\theta}}^{\theta} \exp[-\{z r(z)\}^{-1}]dz\right\}. \tag{A.22}$$

Since the sample only includes active consumers, $F_2[\underline{\theta}_2(\theta_1)] = 0$ for all possible θ_1, and $E_2[\theta_2 \mid \theta_2 \geq \underline{\theta}_2(\theta_1)] = \mu_2$, which is straightforward to compute from the data. The menu of optional two–part tariffs is given by:

$$\tilde{p}(\theta_1) = \frac{1}{\alpha(\theta_1+\mu_2)r_1(\theta_1)}, \tag{A.23}$$

$$\tilde{A}(\theta_1) = \frac{\exp[-\{(\theta_1+\mu_2)r_1(\theta_1)\}^{-1}]}{\alpha}\left\{(\theta_1+\mu_2) - \int_{\underline{\theta}_1}^{\theta_1} \exp[-\{(z+\mu_2)r_1(z)\}^{-1}]dz\right\}. \tag{A.24}$$

For the utility function (36), condition (31) only requires that $\theta_2^*(\theta_1) = \mu_2$ for all possible *ex–ante* types θ_1. The optimal departure from the "boundary two–part tariff" is then:

$$\Delta\tilde{\tilde{p}}(\theta_2\,|\,\theta_1) = \frac{F_2(\mu_2) - F_2(\theta_2)}{\alpha(\theta_1+\theta_2)f_2(\theta_2)}, \tag{A.25}$$

$$\begin{aligned}\Delta\tilde{\tilde{A}}(\theta_2\,|\,\theta_1) = \frac{\exp\left[-\{(\theta_1+\mu_2)r_1(\theta_1)\}^{-1}\right]}{\alpha}\Bigg\{&(\theta_1+\theta_2)\left(\exp\left[\frac{F_2(\mu_2) - F_2(\theta_2)}{(\theta_1+\theta_2)f_2(\theta_2)}\right] - 1\right)\\ &-\int_{\mu_2}^{\theta_2}\left(\exp\left[\frac{F_2(\mu_2) - F_2(z)}{(\theta_1 + z)f_2(z)}\right] - 1\right)dz\Bigg\},\end{aligned} \tag{A.26}$$

which together with $(A.23)-(A.24)$ defines for each *ex–ante* type θ_1, the menu of nonlinear options $\{\tilde{\tilde{A}}(\theta_2\,|\,\theta_1), \tilde{\tilde{p}}(\theta_2\,|\,\theta_1)\}$.

Table 1. Descriptive Statistics

	Bowling Green	Louisville	TEST
CALLS, θ	32.0489	36.6112	-6.63
	(26.902)	(38.197)	
EXPCALLS, θ_1	31.4137	25.9329	8.02
	(36.123)	(30.827)	
BIAS, θ_2	0.6352	10.6783	-12.64
	(37.179)	(39.966)	
log(INCOME)	7.3097	7.0847	13.55
	(0.798)	(0.819)	
HHSIZE	2.7960	2.5381	9.02
	(1.266)	(1.493)	
TEENS	0.3711	0.2309	10.31
	(0.713)	(0.619)	
DINCOME	0.1328	0.1603	-3.78
	(0.339)	(0.370)	
AGE1	0.0614	0.0625	-0.22
	(0.240)	(0.242)	
AGE2	0.2524	0.2644	-1.34
	(0.434)	(0.441)	
AGE3	0.6861	0.6730	1.37
	(0.464)	(0.469)	
COLLEGE	0.2803	0.2244	6.31
	(0.449)	(0.417)	
MARRIED	0.6926	0.5059	18.85
	(0.462)	(0.500)	
RETIRED	0.1525	0.2550	-12.40
	(0.360)	(0.436)	
BLACK	0.0622	0.1168	-9.25
	(0.242)	(0.321)	
CHURCH	0.2082	0.1692	4.88
	(0.406)	(0.375)	
BENEFITS	0.2063	0.3152	-12.11
	(0.405)	(0.465)	
MOVED	0.4820	0.4074	7.34
	(0.500)	(0.491)	
ONLYMALE	0.0452	0.1053	-10.99
	(0.208)	(0.307)	
MARCH	0.3288	0.3325	-0.38
	(0.470)	(0.471)	
APRIL	0.3318	0.3318	0.00
	(0.471)	(0.471)	
MAY	0.3394	0.3357	0.38
	(0.474)	(0.472)	
Observations	5241	4349	

Mean and standard deviations (between parentheses) of demographics for the spring sample. The "TEST" column shows the test of differences of means for each variable in these two cities. Full description of the variables can be found in the Appendix of Miravete (2000a).

Table 2. Consumption Expectation Bias

Bowling Green					Louisville			
Obs.	PAT	Avg.Bias	Std.Dev.	Strata	Avg.Bias	Std.Dev.	PAT	Obs.
5241	2652.59	0.6352	(37.179)	ALL	10.6783	(39.966)	2353.89	4249
1723	879.39	0.9765	(37.076)	MARCH	11.6001	(43.581)	758.78	1446
1739	903.94	0.6571	(37.014)	APRIL	10.5580	(39.119)	791.41	1443
1779	879.94	0.2834	(37.457)	MAY	9.8842	(36.946)	819.24	1460
1967	1029.82	2.9062	(39.662)	LOW INCOME	15.9668	(50.592)	917.78	1645
3274	1662.00	-0.7291	(35.541)	HIGH INCOME	7.4610	(31.388)	1484.04	2704
714	293.15	0.0920	(18.198)	HHSIZE=1	6.2131	(34.470)	597.57	1095
1774	1016.19	-1.1249	(30.470)	HHSIZE=2	6.4538	(27.637)	874.67	1502
1290	704.12	2.9518	(33.353)	HHSIZE=3	13.8281	(38.995)	426.18	776
980	562.48	-0.0021	(47.312)	HHSIZE=4	14.3265	(43.909)	336.77	582
483	281.00	3.0087	(59.734)	HHSIZE $\geq$ 5	27.6001	(71.748)	277.91	394
3798	1941.58	-0.3655	(29.838)	TEENS=0	7.5578	(35.786)	2060.40	3653
1029	611.62	0.9405	(54.873)	TEENS=1	23.4185	(47.131)	252.33	460
414	225.09	9.0571	(42.156)	TEENS $\geq$ 2	34.1479	(65.503)	164.79	236
322	217.03	-4.7589	(26.910)	AGE1=1	8.4026	(32.578)	205.51	272
1323	869.76	-2.7377	(42.171)	AGE2=1	9.0469	(38.949)	723.88	1150
3596	1677.65	2.3592	(35.866)	AGE3=1	11.5307	(40.955)	1514.95	2927
1469	828.09	-3.4543	(37.277)	COLLEGE=1	4.6580	(28.899)	524.11	976
3772	1878.68	2.2279	(37.024)	COLLEGE=0	12.4203	(42.480)	1908.92	3373
3630	1851.96	0.5463	(36.427)	MARRIED=1	10.6344	(32.603)	1243.15	2200
1611	835.40	0.8355	(38.830)	MARRIED=0	10.7232	(46.315)	1166.71	2149
799	338.42	1.3146	(28.672)	RETIRED=1	9.6512	(35.496)	561.92	1109
4442	2361.63	0.5130	(38.512)	RETIRED=0	11.0299	(41.384)	1844.82	3240
326	237.93	11.6811	(71.411)	BLACK=1	29.3614	(66.110)	454.15	508
4915	2488.20	-0.0974	(33.587)	BLACK=0	8.2073	(34.340)	1957.76	3841
1091	600.92	-1.8867	(45.088)	CHURCH=1	7.8696	(52.922)	329.06	736
4150	2107.23	1.2982	(34.779)	CHURCH=0	11.2505	(36.754)	2056.26	3613
1081	493.97	2.2926	(35.188)	BENEFITS=1	13.8292	(42.011)	726.25	1371
4160	2201.68	0.2046	(37.671)	BENEFITS=0	9.2277	(38.910)	1661.81	2978
2526	1334.84	0.0820	(40.646)	MOVED=1	10.7220	(39.305)	1100.09	1772
2715	1381.03	1.1500	(33.634)	MOVED=0	10.6482	(40.422)	1303.97	2577
237	145.27	-3.5797	(23.912)	ONLYMALE=1	4.6319	(27.237)	265.54	458
5004	2541.78	0.8349	(37.682)	ONLYMALE=0	11.3900	(41.151)	2127.43	3891

“PAT” column reports Pearson analog goodness of fit test for equality of the distribution ”of the expected and actual number of calls. This test is distributed as a $\chi^2(19)$, with 0.05 and 0.01” critical values at 30.14 and 36.19 respectively. All statistics have p–values lower than 0.01.

Table 3. Test of Stochastic Dominance

	Bowling Green		Louisville	
Order:	1st	2nd	1st	2nd
ALL	2.72	0.51	-5.65	-8.44
MARCH	1.03	-0.26	-3.27	-4.37
APRIL	1.52	0.25	-3.56	-4.74
MAY	2.16	0.91	-2.94	-5.15
LOW INCOME	-0.08	-1.94	-6.15	-6.15
HIGH INCOME	3.91	2.06	-1.92	-4.77
HHSIZE=1	2.64	0.51	-0.65	-2.73
HHSIZE=2	5.66	4.09	-0.79	-3.37
HHSIZE=3	0.27	-0.93	-3.28	-3.70
HHSIZE=4	0.79	0.00	-1.63	-2.69
HHSIZE $\geq$ 5	0.00	0.00	-2.55	-2.55
TEENS=0	3.59	2.14	-1.75	-5.77
TEENS=1	1.27	-0.41	-2.12	-2.12
TEENS $\geq$ 2	-0.18	-0.45	-0.58	-0.58
AGE1=1	3.74	2.68	2.73	1.92
AGE2=1	3.64	2.70	-1.61	-1.83
AGE3=1	0.65	-1.44	-5.34	-8.58
COLLEGE=1	4.25	3.59	0.06	-1.63
COLLEGE=0	0.69	-1.48	-5.89	-8.60
MARRIED=1	2.46	0.59	-3.46	-4.90
MARRIED=0	2.16	0.03	-4.51	-6.69
RETIRED=1	1.73	0.57	-1.65	-4.28
RETIRED=0	2.99	0.38	-5.43	-5.94
BLACK=1	-2.16	-2.16	-3.72	-3.72
BLACK=0	4.27	2.31	-3.00	-6.49
CHURCH=1	2.01	1.23	0.09	-0.85
CHURCH=0	3.40	-0.08	-6.57	-7.41
BENEFITS=1	1.47	-0.01	-4.60	-6.68
BENEFITS=0	3.27	0.57	-3.81	-5.48
MOVED=1	3.61	-0.22	-2.59	-2.72
MOVED=0	4.10	1.86	-4.00	-6.94
ONLYMALE=1	3.10	2.39	0.66	-1.48
ONLYMALE=0	2.66	0.36	-5.28	-8.51

Maximum ratios by demographics of Anderson's (1996) test for a uniform 20–fractile division of the calling range. These ratios are distributed as a studientized maximum modulus distribution [Stoline and Ury (1979)]. With 20 multiple comparisons and infinite degrees of freedom the 5% and 1% one–tail critical values are 3.03 and 3.49 respectively.

Table 4. Simulation Results

BOWLING GREEN					
Tariff		$\varepsilon = -0.10$	$\varepsilon = -0.17$	$\varepsilon = -0.45$	$\varepsilon = -0.70$
Ex–Post	A	74.911	44.065	16.647	10.702
	p	0.119	0.070	0.026	0.017
	x	134.340	134.340	134.340	134.340
	V	19.127	11.251	4.250	2.732
	π	76.357	44.916	16.968	10.908
	W	95.484	56.167	21.219	13.641
Op. TPT	A	77.047	45.322	17.122	11.007
	p	0.079	0.046	0.018	0.011
	x	136.312	136.312	136.312	136.312
	V	18.371	10.807	4.083	2.625
	π	79.389	46.699	17.642	11.341
	W	97.760	57.506	21.724	13.966
Op. NLT	A	77.134	45.373	17.141	11.019
	p	0.141	0.083	0.031	0.020
	x	141.431	141.431	141.431	141.431
	V	21.868	12.863	4.859	3.124
	π	68.936	40.551	15.319	9.848
	W	90.803	53.414	20.179	12.972
LOUISVILLE					
Tariff		$\varepsilon = -0.10$	$\varepsilon = -0.17$	$\varepsilon = -0.45$	$\varepsilon = -0.70$
Ex–Post	A	100.530	59.135	22.340	14.361
	p	0.195	0.115	0.043	0.028
	x	174.076	174.076	174.076	174.076
	V	21.323	12.543	4.739	3.046
	π	103.002	60.590	22.889	14.715
	W	124.326	73.133	27.628	17.761
Op. TPT	A	108.266	63.686	24.059	15.467
	p	0.119	0.070	0.026	0.017
	x	179.017	179.017	179.017	179.017
	V	17.046	10.027	3.788	2.435
	π	110.498	64.999	24.555	15.785
	W	127.543	75.026	28.343	18.221
Op. NLT	A	115.539	67.964	25.675	16.506
	p	0.119	0.070	0.026	0.017
	x	179.017	179.017	179.017	179.017
	V	9.773	5.749	2.172	1.396
	π	114.385	67.286	25.419	16.341
	W	124.158	73.034	27.591	17.737

Average value of 10,000 random draws from Gaussian kernel estimates of the corresponding probability density functions.

Figure 1. Empirical Distributions

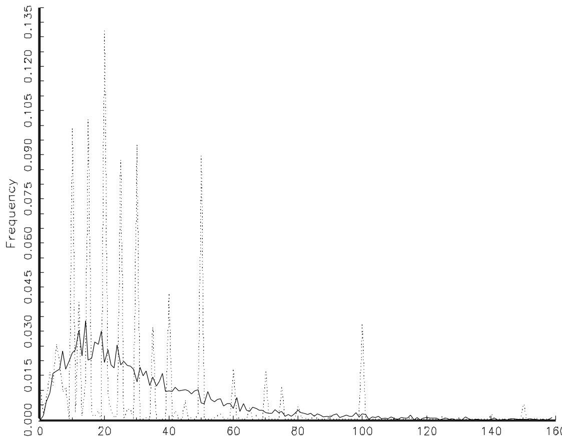

FIGURE 1.1: Actual and Expected (dotted) Calls in BG

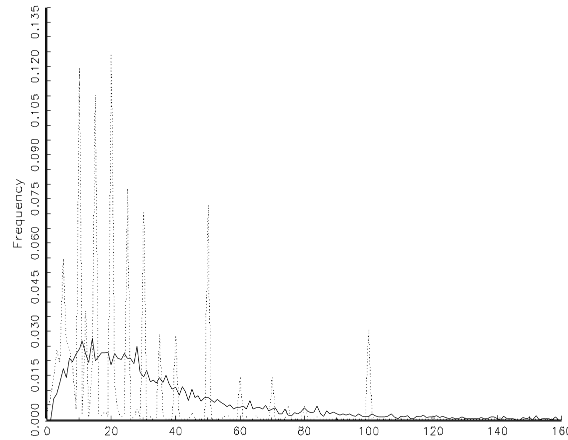

FIGURE 1.2: Actual and Expected (dotted) Calls in LV

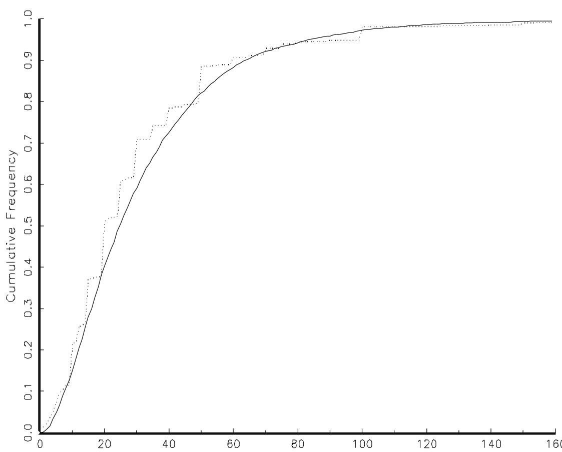

FIGURE 1.3: Actual and Expected (dotted) Calls in BG

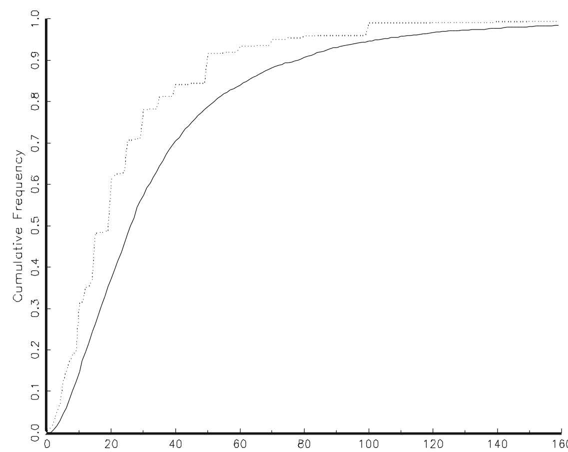

FIGURE 1.4: Actual and Expected (dotted) Calls in LV

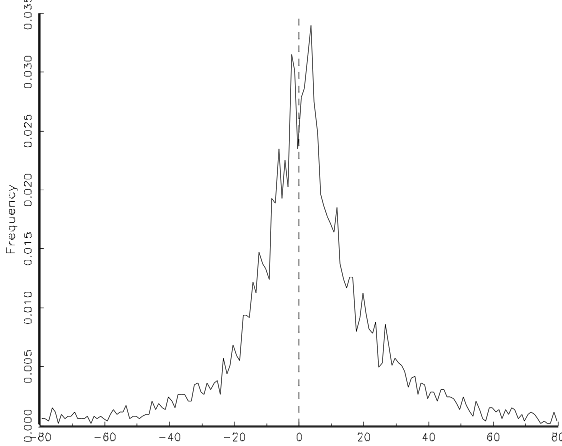

FIGURE 1.5: Call Expectation Bias in BG

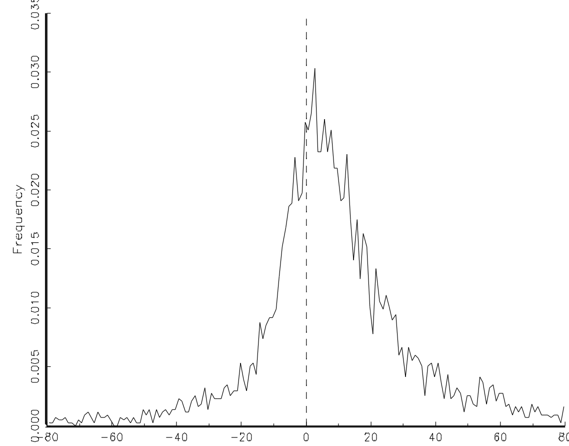

FIGURE 1.6: Call Expectation Bias in LV

Figure 2. Bowling Green: Kernel Estimates

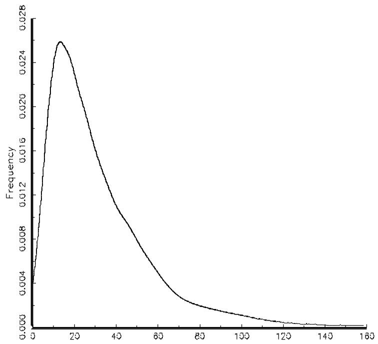

FIGURE 2.1: Actual Number of Calls

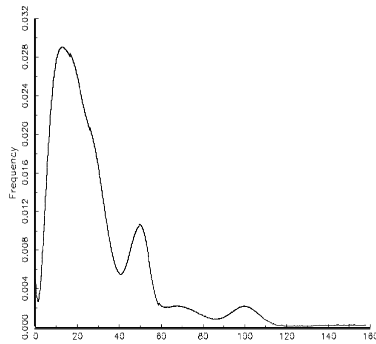

FIGURE 2.2: Expected Number of Calls

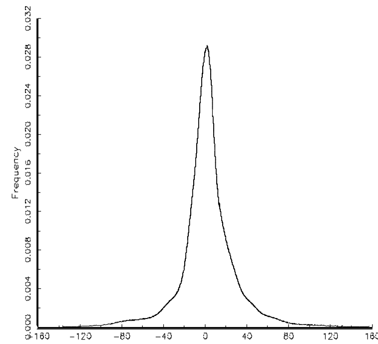

FIGURE 2.3: Call Expectation Bias

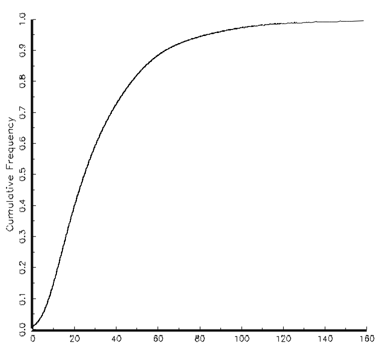

FIGURE 2.4: Actual Number of Calls

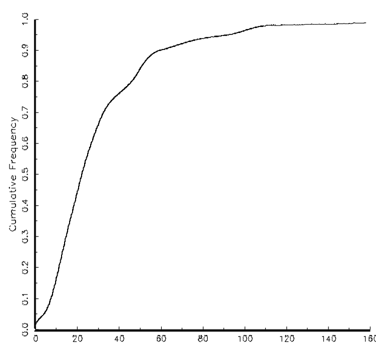

FIGURE 2.5: Expected Number of Calls

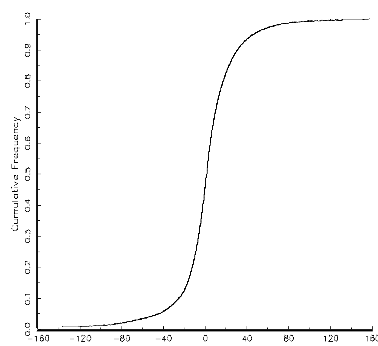

FIGURE 2.6: Call Expectation Bias

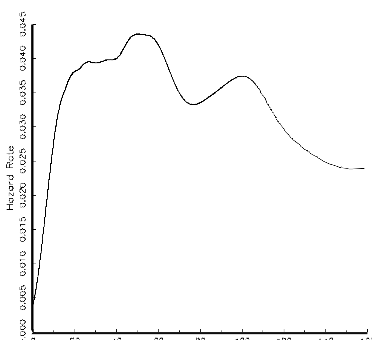

FIGURE 2.7: Actual Number of Calls

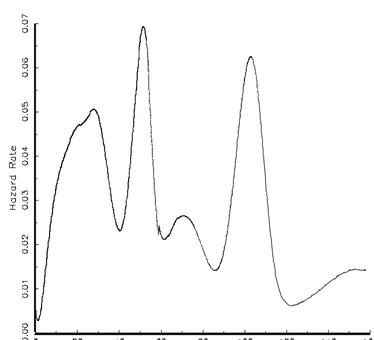

FIGURE 2.8: Expected Number of Calls

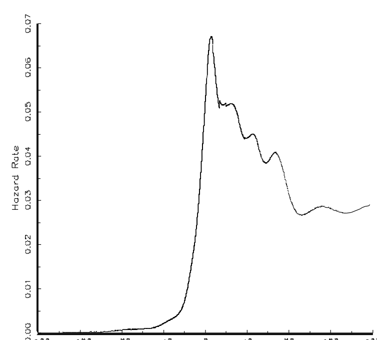

FIGURE 2.9: Call Expectation Bias

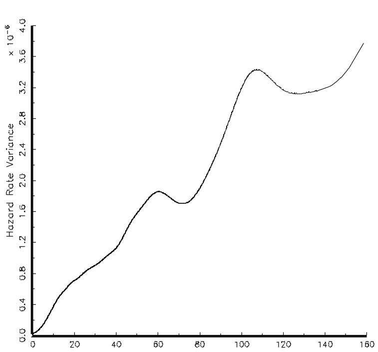

FIGURE 2.10: Actual Number of Calls

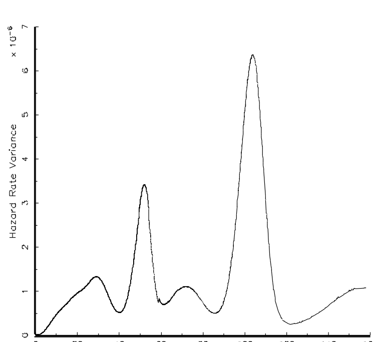

FIGURE 2.11: Expected Number of Calls

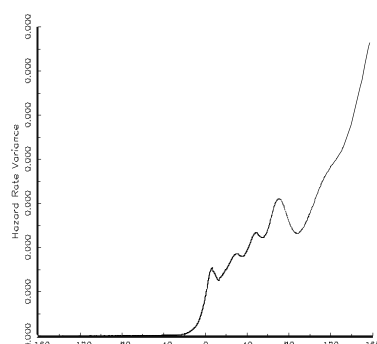

FIGURE 2.12: Call Expectation Bias

Figure 3. Louisville: Kernel Estimates