Nonparametric analysis of cost complementarities in the telecommunications industry

Arie Beresteanu*

Using data on Local Exchange Carriers (LECs), I estimate the total cost function of LECs operating between 1988 and 1995. First, I show that cost complementarities can be nonparametrically identified in many situations where economies of scope and subadditivity - the focus of previous empirical research - cannot be identified. Next, I implement a feasible nonparametric estimation model restricting the cost function to satisfy properties required by economic theory. The results support the assumption of cost complementarities in the production of local and toll calls. The degree of complementarity is computed and shown to be larger for small companies.

1 Introduction

The deregulation of the telecommunications industry, in the 1980s, sought to separate markets in which natural monopolies exist from markets that can benefit from competition. To that end, the disvestiture of American Telephone and Telegraph company (AT&T) in 1984 separated the market for local calls and the market for long distance and international calls, and prohibited local carriers from operating in the long distance market. The Telecommunications Act of 1996 was intended to break these regulatory barriers (see Brock (2002) for an historical overview). According to the act, local carriers would be allowed into the long-distance market if they convinced the Federal Communication Commission (FCC) that they had opened their local networks to competitors. In recent years local telephone companies have wished to expand the scope and the scale of their operations. They seek to enter new markets where they were not allowed to operate and to increase the scale of their operations through mergers with similar firms.[1] This paper tries to shed some light on the question of whether increases in the scope and

[0]* Duke University; arie@econ.duke.edu.
This paper is based on a chapter from my thesis. I wish to thank Chuck Manski, Rosa Matskin, Rob Porter and Tom Severini for many helpful comments and discussions. Comments from Gregory Besharov, Tim Bollerslev, Uli Doraszelski, Paul Ellickson, Jean Piere Dube, Hiroshi Ohashi, Holger Sieg, George Tauchen, Huseyin Yildrim and the participants of seminars at Duke and The Hebrew University are appreciated. I thank Ariel Pakes and two referees for many helpful comments that greatly improved the paper. All remaining errors are mine.

[1]For example, the merger between GTE and Bell-Atlantic and the merger between US West and Qwest were two mergers between local telephone carriers. Additional mergers happened in the growing sectors of wireless communication and internet services.

scale of operations are efficient through an analysis of cost functions of Local Exchange Carriers (LECs).

Most of the empirical analysis that emerged after the divestiture of AT&T employs the technology-based approach developed by Baumol (1977) and Baumol, Panzar and Willig (1982). This approach emphasizes the role of the firms' cost function in the study of market structure. Although in agreement about the methodology, the various analyses disagree on the conclusions.[2] The first attempt to address the question of natural monopolies in the telecommunications industry was by Evans and Heckman (1983, 1984). They use a trans-log function and annual time series from the period prior to the divestiture to estimate a cost function in which the two outputs are local and toll calls. They reject the hypothesis that the Bell System's cost function is subadditive and conclude that AT&T was not a natural monopoly. Roller (1990) uses a Generalized-CES-Quadratic approximation function[3] with the same data and reaches the opposite conclusion of Evans and Heckman. Shin and Ying (1992) use disaggregated panel data on 58 LECs[4] operating in the time before AT&T was forced to give up its local services. They find that the LECs' cost function is not subadditive and conclude that it is implausible that AT&T was a natural monopoly. Wilson and Zhou (2001) use data on LECs in the period after the divestiture of AT&T. The data set contains an unbalanced panel of 66 LECs operating between 1988 and 1995. The variables they use resemble those used in Shin and Ying (1992). They conclude that the cost function of the LECs is subadditive, suggesting that local telephone markets are a natural monopoly. This stands in contrast with the result in Shin and Ying (1992).

The conflicting empirical results with the use of different parametric methods raise the concern that properties like economies of scope and subadditivity are not nonparametrically identified. Section 2 introduces the notion of nonparametric identification to the literature of cost function estimation. I show that identification of subadditivity and economies of scope, which were the focus of previous empirical research, requires assumptions about the underlying distribution of the covariates that rarely hold in practice. These requirements include being able to observe firms that produce product mixes that we do not observe in the data. As a result,

[2] See Fuss and Waverman (2002) for a survey of previous literature on cost function estimation.

[3] For an exact formulation and discussion of the properties of the Generalized-CES-Quadratic approximation see Roller (1990).

[4] The local exchange carriers were AT&T's daughter firms providing the telecommunication services in the area in which they operated. All local and toll calls were made through the local exchange carriers.

subadditivity and economies of scope are not likely to be identified, and additional functional form assumptions are needed in order to make inference. I consider a different property of the cost function: cost complementarities. Cost complementarities exist if the marginal cost from an increase in production of one output decreases when the amount produced of the other product increases. Cost complementarities can be identified under relatively mild conditions. In addition, even if economies of scope and subadditivity can be identified in the population, finite samples may restrict our ability to use these properties in estimation. I show that cost complementarities are less vulnerable to finite sample effects.

The empirical focus of this paper is the market for local and toll calls in the period after the divestiture of AT&T but before the Telecommunications Act of 1996. The data at my disposal is based on reports submitted by the LECs and are published annually by the FCC in the *Statistics of Communications Common Carriers*.[5] The total cost as a function of three groups of variables - outputs, input prices and firm characteristics - is estimated. Starting with a general model, I list the basic properties that a cost function needs to satisfy according to economic theory. Assuming separability of the cost function with respect to the three groups of variables, the model can be written as an additive regression model. This assumption is restrictive but allows a nonparametric estimation while maintaining an estimation model as general as possible. The estimation method uses the backfitting algorithm developed by Hastie and Tibshirani (1987). This algorithm allows one to impose the desired properties of the cost function on the estimator.

Two models are estimated. In the first, the unrestricted model, only the basic properties of the cost function are imposed. In the second, the restricted model, I impose cost complementarities with respect to the outputs in addition to the basic restrictions. Nonparametric estimation under the condition of complementarities is developed in Beresteanu (2001, 2004) and is adopted here to estimate the restricted model. Using a bootstrap technique, I compare the restricted and the unrestricted estimators. The null hypothesis that the cost function exhibits cost complementarities with respect to outputs cannot be rejected. Therefore, I conclude that there are cost complementarities in the production of local and toll calls. A measure of the degree of complementarities is suggested and I show that cost complementarities are larger for small companies than they are for big companies. The results are used to address a pertinent policy question - the "Internet Freedom and Broadband Deployment" Act.[6]

[5] The data set was provided to me by Wilson and Zhou and was used by them in Wilson and Zhou (2001).

[6] 107-H.R.1542, passed by the House of Representatives in 2002. See more discussion in Section 4.

The rest of the paper is organized as follows. Section 2 defines economies of scope, subadditivity and cost complementarities and discusses their identification and additional issues specific to finite samples. Section 3 describes the estimation model and the data. Section 4 presents the empirical results, and section 5 concludes. The appendix contains the technical details of the estimation employed in Section 4.

2 Economies of scope, subadditivity and cost complementarities

Three properties of cost functions are discussed here: cost economies of scope, subadditivity and cost complementarities. These properties are commonly used in the literature to determine the market structure in the industry in which a multi-product firm operates (see Panzar (1989) for a comprehensive discussion). Moreover, the first two are extensively used in the empirical analysis of telephone companies. This section discusses economies of scope, subadditivity and cost complementarities in estimation. Both identification and finite sample issues are addressed. To simplify notation and discussion, attention, in this section, is restricted to the cost of production and to outputs produced, ignoring other explanatory variables.

2.1 Definitions

Assume that in some industry firms produce a vector of products chosen from a set $\mathcal{Y}$ of all feasible production plans. Let I be the population of firms in the industry. I assume here that all firms in the population have the same cost function up to a firm specific scalar. The total cost function of firm $i \in I$ is therefore,

$$C_i = \Xi_i \cdot C(Y_i) \tag{1}$$

where C_i is the total cost, Y_i is a vector of outputs in the set $\mathcal{Y}$ and Ξ_i is firm specific. In this formulation Ξ_i can be viewed as the firm's relative efficiency factor. $\Xi_i > E(\Xi_i)$ means that firm i's cost of producing Y is relatively higher than the industry average and thus firm i is less efficient than the average. Throughout the following discussion, I assume for simplicity that $\mathcal{Y} \subset \Re^2_+$ and $C_i \in \Re_+$ for all i. The set $\mathcal{Y}$ represents the technology available to the firms in this industry. Finally, the projections of vectors in $\Re^2$ on the first and second dimensions are denoted by Π_1 and Π_2 respectively.

One of the key questions in the discussion about breaking up AT&T in the 1980s was what is market structure in the telephone industry.[7] Arguments in favor of the divestiture included the claim that producing a vector of outputs, Y, costs less when each element of the output vector Y is produced separately by a different single-product firm than when Y is produced together by one multi-product firm. In other words, splitting AT&T into firms which produce distinct output vectors is based on the idea that the cost function of AT&T exhibits diseconomies of scope. The following definition formalizes this concept.

Definition 1 *The firm's cost function exhibits **cost economies of scope** over a region $\mathcal{A} \subset \mathcal{Y}$ if*

$$C(Y) \leq C(\Pi_1 Y) + C(\Pi_2 Y) \tag{2}$$

for all vectors of outputs $Y \in \mathcal{A}$, such that $\Pi_1 Y, \Pi_2 Y \in \mathcal{A}$.

The two companies, whose cost functions appears on the right-hand side of inequality (2), produce different outputs. As Theorem 1 below shows, the ability to test (2) is very limited. This is one of the reasons why researchers focused on a different property of the cost function named subadditivity. The second reason is that subadditive cost function implies that natural monopolies exist in this industry (see Baumol (1977) and Baumol, Panzar and Willig (1982, chapter 7)). Subadditivity modifies the above definition by relaxing the specialization requirement.

Definition 2 *The firm's cost function is said to be **subadditive** over a region $\mathcal{A} \subset \mathcal{Y}$ if*

$$C(Y) \leq C(Y_1) + C(Y_2) \tag{3}$$

for all $Y_1, Y_2, Y \in \mathcal{A}$ such that $Y_1 + Y_2 = Y$.

Milgrom and Roberts (1990) reintroduced the concept of complementarities in economics in the context of production. Modern theoretical analysis of complementarity has benefited substantially from the mathematical theory of supermodular functions, which makes explicit the necessary and sufficient conditions for complementarity without imposing auxiliary functional form assumptions. Milgrom and Shannon (1994) and Topkis (1998) give a comprehensive description of economic questions that can benefit from this type of analysis. This paper adopts

[7] Apart from the local carriers, AT&T also included Western Electric, its manufacturing arm, which supplied most of the equipment used by the Bell System, and Bell Laboratories, which performed most of the research and development for the Bell System. Two other important issues were the potential impact of the divestiture of AT&T on research and development and on the market for telecommunications equipment. For additional discussion on the role of economics in the AT&T divestiture case see Lovell and Sickles (1999).

the analysis of complementarities to cost functions. Cost complementarities exist if the marginal cost from an increase in production of one output decreases when the amount produced of the other product increases. The next definition relates the concept of cost complementarities to submodular functions.

Definition 3 *The firm's cost function, C, exhibits* ***cost complementarities*** *on $\mathcal{A} \subset \mathcal{Y}$ if C is submodular*[8] *in $\mathcal{A}$.*

$$\left[C(Y+Y') - C(Y)\right] \leq \left[C(Y+\Pi_1 Y') - C(Y)\right] + \left[C(Y+\Pi_2 Y') - C(Y)\right] \tag{4}$$

for any $Y \in \mathcal{A}$ and $Y' > 0$ such that $Y+Y'$,$Y+\Pi_1 Y'$, $Y+\Pi_2 Y' \in \mathcal{A}$.
The firm's cost function, C, exhibits ***log cost complementarities*** *on $\mathcal{A} \subset \mathcal{Y}$ if C is log-submodular in $\mathcal{A}$.*

$$\frac{C(Y+Y')}{C(Y)} \leq \frac{C(Y+\Pi_1 Y')}{C(Y)} \cdot \frac{C(Y+\Pi_2 Y')}{C(Y)} \tag{5}$$

for any $Y \in \mathcal{A}$ and $Y' > 0$ such that $Y+Y'$,$Y+\Pi_1 Y'$, $Y+\Pi_2 Y' \in \mathcal{A}$.

The right-hand side of inequality (4) represents the sum of two independent increases in cost when we increase production from Y to $Y+\Pi_1 Y'$ and from Y to $Y+\Pi_2 Y'$. The left-hand side represents the increase in cost when output is increased jointly by $Y' = \Pi_1 Y' + \Pi_2 Y'$. Inequality (5) describes the same idea but in ratios. A set, $\mathcal{A}$, for which any $Y \in \mathcal{A}$ and $Y' > 0$, $Y+Y' \in \mathcal{A}$ implies $Y+\Pi_1 Y'$, $Y+\Pi_2 Y' \in \mathcal{A}$ is called a sub-lattice (for an alternative definition see Topkis (1998)). If the set $\mathcal{A}$ includes the origin and the axes, then subadditivity implies economies of scope. If in addition the set $\mathcal{A}$ is a sub-lattice and $C(0) = 0$, then cost complementarities imply economies of scope.[9] Cost complementarities can be viewed as local economies of scope where the point of reference is not the origin but any point in the set $\mathcal{A}$.

2.2 Identification

Studies of identification examine what properties of the model can be learned from the combination of the joint distribution of the observed variables and the assumptions of the model. I follow the methodology described in Roehrig (1988) and Matzkin (1994) and give a brief description of the concept of nonparametric identification analysis. An econometric model, M^*, consists of the

[8] f is submodular if $(-f)$ is supermodular.

[9] Sharkey (1982) and Topkis (1998) refer to the property in Definition 3 as 'weak cost complementarities' and define a stronger property which implies also subadditivity of the cost function given that $\mathcal{A}$ includes the origin and axes and is a sub-lattice. I thank a referee for pointing this out.

set of observed and unobserved variables, the set of known functional relationships among the variables and the set of restrictions on the unknown functions and distributions in the model. If we can find another model $M \neq M^*$, which is identical to M^* in terms of what is assumed and that implies the same joint distribution of the observed variables, we can not distinguish between the two models. In this case we say that M^* is not identified. In this paper, I show that what drives identification are conditions on the support of the marginal distribution of the covariates. I also show that the conditions on the support of the marginal distribution of the outputs, Y, are weaker in the case of cost complementarities than in the cases of economies of scope or subadditivity. Finite sample issues are discussed in the next section.

The observed variables in the model are C, the total cost and Y, the vector of outputs. Ξ, firm's efficiency, is the unobserved variable. The data generating process of (C, Y, Ξ) satisfies $C = \Xi \cdot f(Y)$ for some unknown continuous function f and $E(\Xi|Y) = k$ a.s. for some constant $k > 0$. The data are an i.i.d sample from this data generating process and reveals only C and Y. Denote the support[10] of the marginal distribution of Y (the outputs) by S. The set S is the set of all production plans that firms in the population actually choose to produce and thus S can be a proper subset of $\mathcal{Y}$, the technology set. Our ability to verify the conditions on the cost functions in definitions 1, 2 and 3 depends on the support of the marginal distribution of the outputs. If the last does not include points which are necessary for testing the conditions in these definitions, then these properties are not verifiable. Theorem 1 makes explicit the conditions on the support that allows identification of economies of scope and subadditivity.

Theorem 1 ***Identification of economies of scope and subadditivity****. Let C, Y, Ξ be three random variables which satisfy $C = \Xi \cdot f(Y)$ and $E(\Xi|Y) = k$ a.s. for some constant $k > 0$. Then, with no additional information*

1) The biggest set in which economies of scope can be identified is

$$\tilde{S}_{scope} = \{Y \in S : \Pi_1 Y \in S \text{ and } \Pi_2 Y \in S\}.$$

2) For $i = 1, 2$ let $s_i = \inf\{\Pi_i S\}$. The biggest set in which subadditivity can be identified is

$$\tilde{S}_{sub} = \{Y \in S : \Pi_1 Y \geq 2s_1 \text{ and } \Pi_2 Y \geq 2s_2\}.$$

[10] The support of a distribution is the set of all points which every neighborhood around them will be visited infinitely many times in an infinite sample.

Proof.

1) To identify inequality (2), it should be possible to observe situations in which companies specialize in producing disjoint subsets of the product vector. Therefore, inequality (2) can be identified in the population only for points $Y \in S$ *such that also* $\Pi_1 Y \in S$ *and* $\Pi_2 Y \in S$*.*

2) Take $y \in S \backslash \tilde{S}_{sub}$ *(where* $S \backslash \tilde{S}_{sub} = \left\{ x \in S | x \notin \tilde{S}_{sub} \right\}$*) and let* y' *and* y'' *be such that* $y' + y'' = y$*. Then either* $y_1 < 2s_1$ *or* $y_2 < 2s_2$*. If the first is true then it must be that either* $y'_1 < s_1$ *or* $y''_1 < s_1$*. If the second is true then either* $y'_2 < s_2$ *or* $y''_2 < s_2$*. This means that for such* y *we cannot find* y' *and* y'' *such that* $y' + y'' = y$ *and both* y' *and* y'' *are in* S*. Therefore, for* $y \in S \backslash \tilde{S}_{sub}$ *subadditivity is not identified.* ■

Theorem 1 highlights the difference between the technology set $\mathcal{Y}$ and the support of the outputs S. For example, the set $\mathcal{Y}$ satisfies the free disposal property and thus for each $Y \in \mathcal{Y}$, $\{\Pi_1 Y, \Pi_2 Y\} \subset \mathcal{Y}$. This is merely a property of the technology available for the firms. The population of firms may never actually choose to produce on the axes and thus $\Pi_1 Y, \Pi_2 Y \notin S$ for all $Y \in S$. In this case, economies of scope are nowhere identified. The requirement that S contains the axes is a strong requirement which is rarely supported by the data. This is one of the reasons for the interest in subadditivity. The following corollary shows that lack of identification can occur even with subadditivity.

Corollary 1 *Given the conditions in Theorem 1,*

1) If either $s_1 > 0$ *or* $s_2 > 0$ *then* $\exists Y \in S$ *such that subadditivity is not identified for* Y*.*

2) If either $2s_1 \geq \sup \{\Pi_1 S\}$ *or* $2s_2 \geq \sup \{\Pi_2 S\}$ *then subadditivity is nowhere identified.*

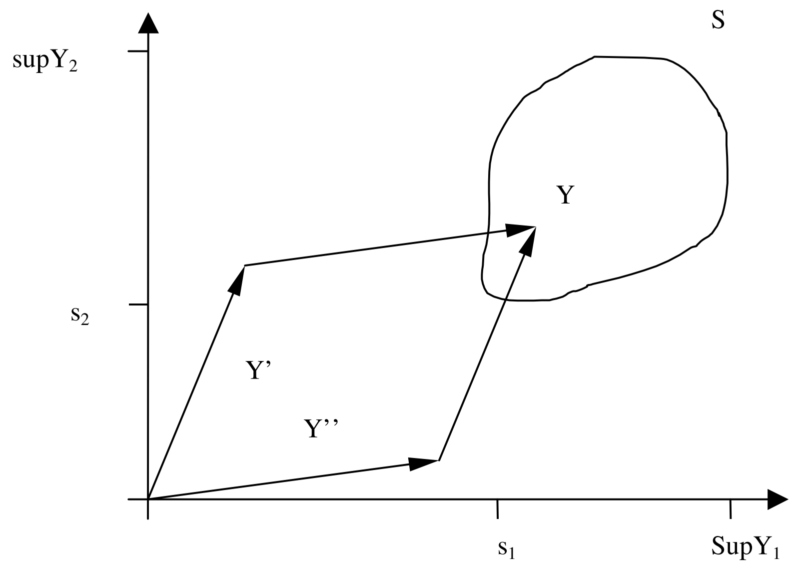
Proof.

(1) Let $s_1 = \inf \{\Pi_1 y : y \in S\}$ *,* $s_2 = \inf \{\Pi_2 y : y \in S\} > 0$ *and* $\tilde{S} = \{y \in S : (s_1, s_2) \leq y \leq (2s_1, 2s_2)\}$*. Then for* $y \in \tilde{S}$ *inequality (3) cannot be tested. However if both* $s_1 = 0$ *and* $s_2 = 0$ *then* $\tilde{S} = \{0\}$ *and (3) is trivially satisfied. In the case described in diagram 1* $2s_1 \geq \sup \{\Pi_1 y : y \in S\}$ *and* $2s_2 \geq \sup \{\Pi_2 y : y \in S\}$ *which mean that* $\tilde{S} \supset S$*.*

(2) Using the argument in (1) $S \subset \tilde{S}$ *and therefore subadditivity is nowhere identified.* ■

Identification of subadditivity at a point Y depends on its proximity to the origin. This fact may cause complete lack of identification as Figure 1 demonstrates or at least partial lack of identification as Corollary 1 clarifies. For any point Y in the set S in Figure 1, any break-up combination $Y' + Y'' = Y$ will have at least one of the points Y' or Y'' outside the set S. The reason is that in Figure 1 the infimum of the points in S is too big and twice this infimum exceeds the supremum. With cost complementarities, the location of the support with respect to the

Figure 1: An example of S for which subadditivity is nowhere identified



origin or the axes is irrelevant. For example, monotone transformations (including shifts) of the explanatory variables do not affect the property of cost complementarities. This is due to the fact that submodularity is an ordinal property of the function.[11] Cost complementarities (and log-cost complementarities) can be identified on any sub-lattice, $\tilde{S}$, contained in the support of the covariates, S. The axes need not be included in S and identification does not depend on the proximity of the points to the axes. In many cases where subadditivity and economies of scope cannot be identified we can identify cost complementarities. For example, in Figure 1, S is a sub-lattice and cost complementarities can be identified on S even though subadditivity is nowhere identified.

2.3 Finite sample issues

Identification, or lack of identification, of properties like economies of scope, subadditivity and cost complementarities depends on the support of the marginal distribution of the covariates.

[11] Suppose that f is a supermodular function defined on a sub-lattice S with a partial order $\preceq$. If S' is another set with a partial order $\preceq'$ and $T : S' \to S$ is such that $s \preceq t$ in S' iff $T(s) \preceq' T(t)$ in S then $f \circ T$ is supermodular in S'.

Even if the population support includes the axes or is such that subadditivity is identified, small samples may restrict our ability to take into account these properties in estimation. Low density near the axes is sufficient for identification in the population but will probably result in no or few observations in this region. In this case no information about economies of scope can be obtained from the sample. Restricting our attention to some properties is thus a result of both what can be identified in the population and what can be learned from finite samples. In this section I discuss the difference between the three properties of cost functions described above from a finite sample point of view. Since lack of nonparametric identification may suggest the use of parametric methods, they are the starting point of the discussion here.

Lack of identification implies that we can find two different functional forms that produce comparable fits on the support and that one is inconsistent with inequality (2) and the other confirms it. In addition to the question of which functional form to use, there is also a question of how far away from the observed sample should we extrapolate the estimators. The extrapolation area depends on our belief about the support of the marginal distribution of the outputs. This belief about the support is often called an admissible set and is denoted here as $\hat{S}$. Identification of economies of scope, as Theorem 1 shows, requires that $\hat{S}$ includes the axes. This is usually not the case with subadditivity. As Definition 2 reveals, as long as for a point Y in $\hat{S}$ we can find at least one pair of points Y' and Y'' also in $\hat{S}$ such that $Y' + Y'' = Y$, subadditivity can be tested for Y. The choice of $\hat{S}$ determines for which points subadditivity can be tested and also the number of possible break-up combinations for each point in $\hat{S}$.[12] A bigger admissible set means that for each point Y there are more possible break-up combinations included in $\hat{S}$. Therefore, the choice of $\hat{S}$ influences the likelihood of any test to accept subadditivity.

The last point can be better understood using an example from Evans and Heckman (1984). Evans and Heckman indicate that the nature of the observations requires extrapolation of the cost function. They define an admissible set on which the cost function is extrapolated. The admissible set chosen by them is assumed to satisfy two requirements. First, the "mother" firm cannot be divided into firms that produce zero in all elements but one. Instead they suggest requiring that the hypothetical "baby" firms produce at least the sample minimum in each product. More specifically, if we observe $\{y_t\}_{t=1}^{T}$ output vectors, define $y_m = \wedge_{i=1}^{T} y_i$.[13] For an

[12] A break-up combination for a point Y is any pair Y' and Y'' such that $Y' + Y'' = Y$.

[13] $\wedge$ is the minimum taken coordinate wise on a set of vectors. y_m is, therefore, a vector whose j^{th}coordinate equals the minimum over the j^{th}coordinates of all the vectors y_i.

observation $\tilde{y} \in \{y_t\}_{t=1}^T$ a possible break-up, y' and y'', satisfies $y', y'' \geq y_m$. As a result only observations after 1958 (i.e. only 20 out of the available 31 observations) are included in the admissible set. The second requirement further limits the area of extrapolation. They suggest that any "baby" firm does not specialize in one product (local or toll calls) in a rate bigger or smaller than the maximal or minimal rates observed. More specifically, define $R_L = \min_t \frac{y_{1t}}{y_{2t}}$ and $R_U = \max_t \frac{y_{1t}}{y_{2t}}$. The admissible region on which subadditivity is tested is therefore $\hat{S} = \left\{ y \in \Re^2 : y \geq 2y_m, R_L \leq \frac{y_1}{y_2} \leq R_U \right\}$.[14] They claim that this region is conservative enough and that extrapolation on that region is reasonable.[15] Evans and Heckman (1983, 1984) use versions of the trans-log approximation to estimate the cost function of AT&T. Using this technique they reject the hypothesis that the Bell System's cost function is subadditive and conclude that AT&T was not a natural monopoly. Roller (1990) suggests using a Generalized-CES-Quadratic approximation function. He tests also for economies of scope and therefore extrapolates his estimator to the axes. Using a different cost function and different admissible set but the same data, Roller (1990) reaches the opposite conclusion from Evans and Heckman (1983, 1984).[16] These conflicting results further illustrate the claim that the results can dramatically change with the choice of functional form since subadditivity is not nonparametrically identified.

Testing for cost complementarities does not require a substantial increase of the support S. The estimator employed in this paper is based on a spline smoother constructed on a grid of points built around the observed sample. The extrapolation needed to have a sub-lattice structure is still within the scope of the nonparametric methods used here. Figure 2 in Section 4 illustrates a choice of a grid that leads to a small amount of extrapolation. Furthermore, since lack of identification is a property of the population, bigger samples do not change the need for extrapolation in the case of testing for subadditivity. In the case of cost complementarities, a bigger sample enables a choice of a finer estimation grid which decreases the amount of extrapolation needed.

[14] See figure 2 in Evans and Heckman (1984).

[15] Evans and Heckman (1984) compare the cost for each observed production configuration with all possible combinations of break-ups that fall in the admissible region. Since the smallest production configuration that can be tested is two times the minimum (i.e. production in 1947), only points from 1958 could actually be tested. The test over all admissible break-up combinations is done on a grid and is described in their paper.

[16] For another critique of Evans and Heckman's method, see Diewert and Wales (1991).

3 Shape restricted estimation of cost functions of LECs

This section describes estimation of the cost function of telephone companies based on a sample of Local Exchange Companies (LECs). I start from a general model and describe the properties that a cost function should satisfy based on economic theory. Then I describe the data used in this paper and the way it was constructed. The next stage is to describe a set of assumptions that make the model estimateable with existing techniques.

3.1 A general model for the cost function

The total cost function, C, is a function of outputs, Y, factor prices, W, observed firm characteristics, X, and unobserved firms characteristics, Ξ.

$$C = C(Y, W, X, \Xi) \tag{6}$$

Using economic theory, the function $C(\cdot)$ has the following properties:
$(M1)$ For any value of (W, X, Ξ), the function $C(\cdot)$ is monotone in Y.
$(M2)$ For any value of (Y, X, Ξ), the function $C(\cdot)$ is monotone in W.
(H) For any value of (Y, X, Ξ), the function $C(\cdot)$ is homogeneous of degree one in W.
(C) For any value of (Y, X, Ξ), the function $C(\cdot)$ is concave in W.
The cost function may also satisfy the following property:
(LSM) For any value of (W, X, Ξ), the function $C(\cdot)$ is log-submodular in Y.
The null hypothesis is that the function $C(\cdot)$ satisfies conditions $(M1), (M2), (H)$, (C) and (LSM) and the alternative is that $C(\cdot)$ satisfies conditions $(M1), (M2), (H)$ and (C) only. An estimator that imposes $(M1), (M2), (H)$ and (C) on the function $C(\cdot)$ is referred to as the "unconstrained estimator". The estimator that imposes (LSM) as well is referred to as the "constrained estimator". Testing for cost complementarities in the cost function is done by comparing the constrained and the unconstrained estimators.

3.2 The data

This section provides a short description of the variables used to estimate the cost functions. For a detailed explanation on how the variables were constructed and what assumptions have been made, see Wilson and Zhou (2001) and Shin and Ying (1992).[17] The variables are based

[17] The data construction by Wilson and Zhou (2001) mirrors that in Shin and Ying (1992). There are, however, two differences. Shin and Ying (1992) use data from 1976 to 1983 and Wilson and Zhou (2001) use data from 1988 to 1995. Second, the regulatory environment changed through time, leading to some differences in reporting

on reports submitted by the LECs and are published annually by the FCC in the *Statistics of Communications Common Carriers*. This publication contains annual data on major LECs. The data set provided to me by Wilson and Zhou contains 66 companies operating between 1988 and 1995.[18] The LECs included in the sample account for more than 90% of the local telephone lines served in the US.[19] Most of the LECs that cover the remaining 10% of the market are rather small. Therefore, the estimators presented in this paper and other papers that use the same data can be treated as estimators suitable for drawing conclusions on relatively big telephone companies.

I divide the set of explanatory variables into three groups: firm characteristics, factor prices and outputs. Firm characteristics include the number of access lines, percent of electronic switching equipment, number of central offices and average loop length. These variables include information on the size of the area covered by the firm and the technology used by the firm.

Factor prices are computed by dividing the expense related to that factor by the number of units used. For example, Labor price P_L is the ratio of total employment compensation to the number of employees. The construction of capital expenses requires a fair amount of assumptions and data on capital prices.[20] Material prices were computed using the residual (non-labor and non-capital) expenses divided by the number of access lines.

Outputs are minutes of local calls and toll calls as reported by the companies. Total cost includes all operating expenses but does not include depreciation and amortization expenses.

Table 1 lists the variables in the data. LC, TC and AL can be argued to be outputs of the firm. Wilson and Zhou (2001) argue that LC and TC are colinear and suggest using the ratio $r = \frac{TC}{TC+LC}$ and AL as the two outputs; however, colinearity is a result of the linear model assumed in Wilson and Zhou (2001). In this paper I regard the two usage outputs LC and TC as the two outputs in the model. The variable AL is treated as one of the firm's characteristics.

requirements. The FCC addopted a new accounting system that became effective at the begining of 1998.

[18] Observations with missing or suspicious values were removed. The raw data set contains 420 observations and the clean one 401 observations. Apart from missing data, observations were removed as suspect if the computed factor shares were negative. See Wilson and Zhou (2001) for details.

[19] Based on Table 2.3 in the Statistics of Communications Common Carriers reports.

[20] The construction of this variable was severely criticized by Gabel and Kennet (1994).

Table 1: The variables in the data set

Group	Variable	Description	Average	Std.	Min	Max
	L_p	Labor price ($\times 10^3$ \$)	40.37	6.70	27.26	66.81
Prices	K_p	Capital price ($\times 10^3$ \$)	.221	.0499	.129	.635
	M_p	Material price ($\times 10^3$ \$)	.204	.060	.062	.450
	LC	Local calls ($\times 10^6$minutes)	8,500	13,784	148	90,294
Outputs	TC	Toll calls ($\times 10^6$minutes)	1,477	2,372	36	16,822
	AL	Access lines ($\times 10^6$)	2.79	4.12	.076	22.6
Firm characteristics	TK	Percent of electronic switching equipment (%)	95.2	6.8	68.9	100
	CO	Central offices	371	406.4	23	1,967
	ALL	Average loop length (miles)	.043	.032	.0035	.170
Cost	C	Total cost ($\times 10^6$ \$)	1,507	2,150	63.9	11,902

Therefore, the following variables are used:

$$Y = (LC, TC)$$
$$W = (L_p, K_p, M_p)$$
$$X = (AL, TK, CO, ALL)$$

3.3 A feasible nonparametric estimation model for the cost function

In this section I describe a set of assumptions that make it feasible to estimate model (6) with the existing estimation methods. The set of assumptions that I employ here are chosen based on two considerations: first, the ability to take into account the restrictions on the model implied by economic theory; and second, the desire to be as general as possible.

The first assumption is separability of the unobserved characteristics. The firm's cost function is written as follows:

$$C = C(Y, W, X)F(\Xi). \tag{7}$$

Let lowercase letters denote quantities in logs and let $\xi = \log F(\Xi)$; then

$$c = c(Y, W, X) + \xi. \tag{8}$$

I assume that ξ is mean independent of (Y, W, X). In other words, we can regard the function $c(Y, W, X)$ as the regression function of log cost conditioned on the outputs Y, factor prices W and other observed characteristics of the firm X. ξ corresponds to the efficiency term in (1) and can be regarded as the unobserved relative efficiency of the firm.

A note on the mean independence assumption is in order. Olley and Pakes (1996) investigate the production function of companies in the telecommunications equipment industry. They raise two concerns about the mean independence assumption that are relevant in our case as well. First, productivity shocks that appear in the error term can be correlated with the other covariates. The main concern here is a possible correlation between the error term and the outputs. This correlation happens if prices respond to productivity shocks and demand (and therefore outputs) responds to prices. Since the telephone companies in our sample are utility companies and since regulations do not allow them to change prices freely, this correlation may be weaker here than in other applications. Nevertheless, this possible correlation is being ignored here.[21]

The second concern raised by Olley and Pakes (1996) involves selection problems introduced by entry and exit of firms in the sample. The data set is an unbalanced panel. Table 2 summarizes the reasons for entry and exit in the sample. Leaving name changes, joint reporting, and missing observations aside, there are 12 entry episodes and 8 exit episodes (see footnotes to table 2). The small number of firms in the sample limits the analysis of entry and exit. Therefore, I choose to maintain the assumption of mean independence of the error term in (8).

The following theorem shows how the assumptions made for the general model (6) transfer to the log version of the model (8).

Theorem 2 *Let the function $C(\cdot)$ be as in model (6). Then the following claims hold:*
1) The following condition is necessary and sufficient for $(M1)$
$(m1)$ For any value of (W, X), the function $c(\cdot)$ is monotone in Y.
2) The following condition is necessary and sufficient for $(M2)$
$(m2)$ For any value of (Y, X), the function $c(\cdot)$ is monotone in W.
3) Let $W = (W_1, ..., W_n)$ be the vector of all factor prices. Define the following polar transformation of W: $r = \left(\sum_{j=1}^{n} W_j^2\right)^{\frac{1}{2}}, \theta_j = \arctan(\frac{W_{j+1}}{W_j})$ for $j = 1, ..., n-1$. Then for $C(\cdot)$ differentiable in W the following is a necessary and sufficient condition for (H)
(h) For any value of (Y, X), the function $c(\cdot)$ can be decomposed to $c(Y, W, X) = \tilde{c}(Y, \theta_1, ..., \theta_{n-1}, X) + r$.
4) For any value of (Y, X) the following is a sufficient condition for (C).
(c) For any value of (Y, X), the function $c(\cdot)$ is concave in W.

[21] The firms in our sample are assumed to be price takers in the labor, capital and material markets (See Olley and Pakes (1996) for a discussion on increased competition in the material input market). The observed firm's characteristics represent mostly features of the markets in which the firm operates. Therefore, The error term is assumed to be mean independent of input prices and firm's characteristics as well.

Table 2: Entry and exit from the sample

Category	Number of firms	Reasons for entry	Reasons for exit
Appears from 1988 to 1995	32	N/A	N/A
Appears in 1988 and exits before 1995	11	N/A	Merged into another firm - 5 Name change - 6
Enters after 1988 and stays until 1995	17	Passed the reporting threshold - 8 Name change[a] - 5 Missing observations - 4	N/A
Enters after 1988 and exits before 1995	6	Passed the reporting threshold - 4 Name change - 1 Missing observations - 1	Merged into another firm - 2 Name change - 2 Passed the reporting threshold[b]- 1 Missing observation[c] - 1
Total	66	Passed the reporting threshold - 12 Name change - 6 Missing observations - 5	Merged into another firm - 7 Name change - 8 Passed the reporting threshold[b]- 1 Missing observation[c] - 1

[a]South Central Bell and Southern Bell filed separately until 1991 and jointly from 1992 under the name Bellsouth, The Mountain State Telephone and Telegraph Company, Northwestern Bell and Pacific Northwest Telephone Company reported separately until 1991 and filed jointly under the name US West from 1992. Other firms simply changed their name and therefore are counted as two firms and generate entry and exit.

[b]Contel of West fell below the reporting threshold after selling part of its holding to other carriers.

[c] Citizens Utilities acquired Contel of New York in 1994 and the FCC waived its reporting obligation for 1995.

5) For any value of (W, X)*, and if* $(M1)$ *holds then the following condition is necessary and sufficient for* (LSM)

(sm) *For any value of* (W, X)*, the function* $c(\cdot)$ *is submodular in* Y*.*

Proof. *The first two parts follow immediately from the monotonicity of the* log *and* exp *functions. Part 3 follows from the analysis described in detail in Beresteanu (2001, Chapter 3). Part* 4 *means that log-concavity implies concavity and is a known result from real-analysis. Part* 5 *is based on the definition of log-submodularity (see Topkis (1998, section 2.6)).* ■

Theorem 2 deals with taking logs of the left-hand side variable - the total cost. The following corollary justifies using logs of the outputs as covariates instead of their levels. Conditions $(m1)$ and (sm) can be written in terms of $y = \log(Y)$ instead of Y. Conditioning on $\log(Y)$ rather than on Y has no effect on the conditional expectation. However, since y is close to being uniformly distributed it is more convenient to use it in the nonparametric estimation described in the next section.

Corollary 2 *conditions* $(m1)$ *and* (sm) *can be written using* y *(log outputs). That is, for any value of* (W, X)*, the function* $c(\cdot)$ *is monotone and submodular in* y*.*

Proof. *The* log *function is a monotone increasing function. This implies that* $\log(Y \wedge Y') = \log Y \wedge \log Y'$ *and* $\log(Y \vee Y') = \log Y \vee \log Y'$ *where* log *is taken coordinate wise and* $\vee$ *and* $\wedge$ *are the maximum and minimum coordinate wise taken on two vectors, respectively. Therefore, submodularity is unchanged when the variables are transformed by a monotone increasing function. It is easy to see that the same applies for monotonicity.* ■

At this stage, apart from assuming separability of the unobserved characteristics of the firm, no further assumptions have been made on the structure of the total-cost function $c(y, W, X)$. With no parametric assumptions on the regression function and with continuous explanatory variables (y, W, X), it is practically impossible to impose the restrictions discussed above on the regression function using existing nonparametric estimation procedures. I choose to address the trade-off between feasibility and flexibility by assuming separability of the various components of the regression function $c(\cdot)$. This specification is obviously a less general one but feasible and for which consistency has been established by Hastie and Tibshirani (1986). In order to take into account all the assumptions mentioned in Theorem 2, I assume the following.

$$C = G_1(y)G_2(W)G_3(X)F(\Xi)$$

or in log terms

$$c = g_1(y) + g_2(W) + g_3(X) + \xi, \tag{9}$$

where lowercase letters denote logs of variables or functions denoted in uppercase letters and $\xi = \log F(\Xi)$.

The variables are divided into three groups. The formulation in (9) does not allow for interactions between variables from different groups.[22] However, interactions between variables *within* each group are left to be arbitrary.[23] The additivity assumption and the nonparametric estimation procedure (see the appendix) balance between flexibility and feasibility.

To take into account the basic properties that a cost function needs to satisfy, I make the following assumptions. I assume that $g_1(y)$ is monotone in y, $g_2(W)$ is monotone and homogenous of degree one in W. Since concavity of $g_2(\cdot)$ is only sufficient for concavity of $G_2(\cdot)$ I do not impose this condition on $g_2(\cdot)$. The constrained estimator restricts $g_1(\cdot)$ to be submodular. There are no assumptions made on the variables X. For simplicity, I assume that $g_3(X) = \alpha + X'\beta$.

4 Empirical results

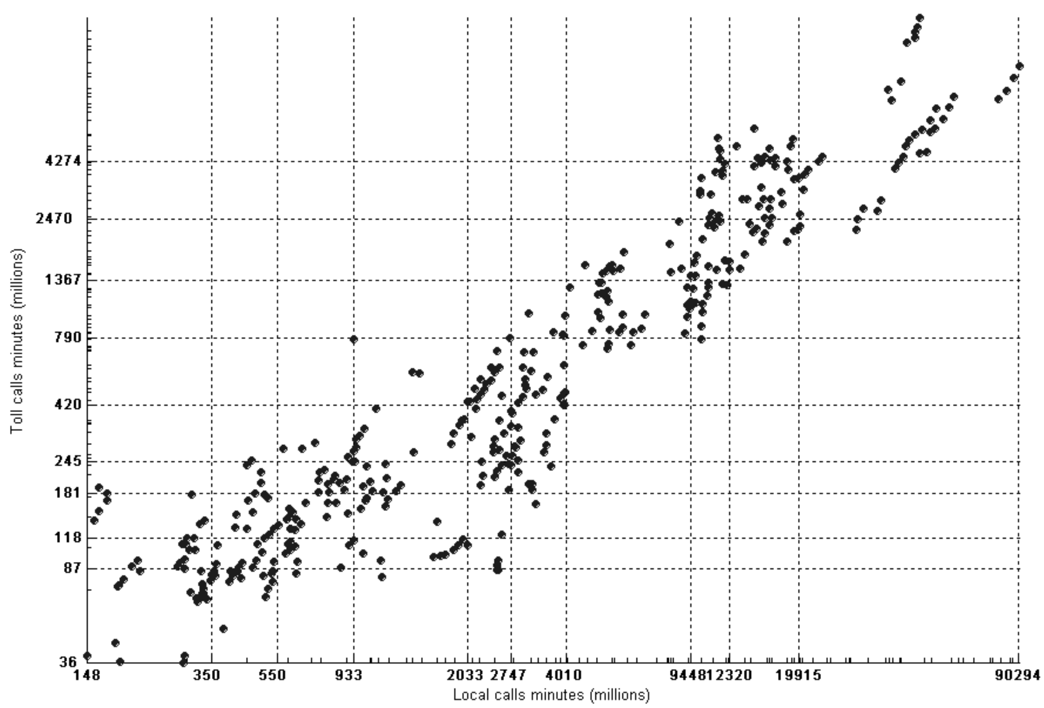
The cost function in (9) is estimated using the backfitting algorithm developed by Hastie and Tibshirani (1987) as described in the appendix. A scatter plot of the output variables is shown in Figure 2. First, assumption (*sm*) is tested in the following way. The restricted model, which represents the null hypothesis, is estimated and the residuals from this model are computed. Since no asymptotic distribution has been developed for this estimator, a bootstrap method is used. The specific bootstrap procedure I am using is wild bootstrap.[24] 500 bootstrap samples

[22] The interactions are not allowed only in the log equation. Obviously, in levels this implies multiplicative interactions between the variables.

[23] The trans-log approximation allows for interactions between all variables but restricts interactions between variables both within and between groups to be linear. Furthermore, the expansions used in previous papers include a large number of parameters to be estimated. For example, Wilson and Zhou (2001) have about 70 parameters (depending on the specification) for the variables and the interactions between them and an additional 55 parameters for firm dummies while having only 401 observations. Even though feasible, a high ratio of parameters to observations causes the regression to be very sensitive to multicolinearity and the estimated parameters to vary across different specifications of the model. Gallant (1982, pages 320-322) argues that conceptually the trans-log approximation hinges on a Taylor expansion argument and thus desired statistical properties of this estimator require one to believe that the trans-log function is the true data generating process.

[24] See appendix for additional notes on the wild bootstrap.

Figure 2: Local and toll calls (millions of minutes)



are generated and for each bootstrap sample both the restricted and the unrestricted estimators are computed. The following statistic is computed for the original sample and for the bootstrap samples

$$\Delta = \frac{\frac{1}{N}\sum_{i=1}^{N}(\tilde{c}_i - \tilde{c}_{0,i})^2 - \frac{1}{N}\sum_{i=1}^{N}(\tilde{c}_i - \tilde{c}_{1,i})^2}{\frac{1}{N}\sum_{i=1}^{N}(\tilde{c}_i - \tilde{c}_{1,i})^2} \tag{10}$$

where $\tilde{c}$ are the bootstrapped costs, $\tilde{c}_0$ is the estimator for the cost under the null and $\tilde{c}_1$ is the estimator under the alternative - both computed using the bootstrap sample. Under the null we expect Δ to be equal to zero since the additional restriction, namely (sm), is non-binding. Under the alternative, however, we expect Δ to diverge to a number different than zero.[25] Since the (sm) property signs the cross partial derivative of the log-cost function, a one-sided test is in order. A 95% critical value for the one-sided test is constructed from the 500 test statistics computed for each bootstrap sample. The statistic in (10) computed from the original sample is $\Delta = 5.96 \times 10^{-3}$. The 95% critical value is 8.64×10^{-3} (and the 90% critical value is 7.24×10^{-3}). Therefore, I conclude that the null hypothesis that the log-cost function is submodular cannot be rejected and that cost complementarities are consistent with the data. The estimators of g_1, g_2 and g_3 under the null are presented in the appendix.[26]

The next stage is to get an idea about the size of cost complementarities in this industry. I compute the following indicator:

$$Comp(y_1, y_2) = \left[\frac{[g_1(y_1+\delta_1, y_2) - g_1(y_1, y_2)] + [g_1(y_1, y_2+\delta_2) - g_1(y_1, y_2)]}{[g_1(y_1+\delta_1, y_2+\delta_2) - g_1(y_1, y_2)]} - 1\right] \cdot 100 \tag{11}$$

for some positive numbers δ_1 and δ_2. The statistic in (11) is positive in the presence of cost complementarities (see inequality (4)) and its size indicates the amount of complementarities. Table 3 reports the results computed from the estimator (under the null).[27] The amount of complementarities is evaluated at the quantiles of the observed outputs. The empty cells belong to areas with no observations. I refer to carriers producing less than the median local calls

[25] The distance between the true cost function, c, and the best predictor to the cost function when forcing it to satisfy the wrong constraint (sm), c_0, is positive and thus the numerator in (10) converges to a positive number while the denominator converges to the variance of the error term.

[26] g_1 and g_2 are nonparametric functions with two and three explanatory variables respectively and therefore can not be easily graphed or tabulated.

[27] To improve the computational properties of (11) at a point (y_1, y_2) a two sided smoothing is used: $Comp(y_1, y_2) = \left[\frac{[g_1(y_1+\delta_1,y_2)-g_1(y_1-\delta_1,y_2)]+[g_1(y_1,y_2+\delta_2)-g_1(y_1,y_2-\delta_2)]}{[g_1(y_1+\delta_1,y_2+\delta_2)-g_1(y_1-\delta_1,y_2-\delta_2)]} - 1\right] \cdot 100$. The δ's were chosen such that $y_i - \delta_i$ and $y_i + \delta_i$ create a $\pm 2\%$ interval around y_i.

(2.747×10^9 minutes) as small firms and to carries producing above median local calls as large firms. Large cost complementarities exist mostly for small carriers. The amounts of complementarities that can be achieved peaks around the median point and then drops above the median. This result suggests that mergers of small firms will allow them to take advantage of the cost complementarities. For large firms, however, cost complementarities are exhausted although still positive. Therefore, a merger between firms operating in the low cost complementarities area can be less efficient in terms of benefits from complementarities.

To better understand the difference between small and large carriers we look at how cost complementarities vary with the product mix. Let r be the proportion of toll calls in the total of local and toll calls, $r = \frac{TC}{TC+LC}$. At the two median points $r \approx 13\%$ and its average in the sample is a little over 16%. Wilson and Zhou (2001) use the variable r and r^2 in estimating the cost function.[28] They conclude that high levels of r increase the cost and do so in an accelerating way since the coefficient of r^2 is also positive. This result is interpreted as an evidence for the existence of economies of scope. As was explained before, we focus here on cost complementarities which are nonparametrically identified. Since cost complementarities are local economies of scope,[29] their result concerning economies of scope is consistent with the existence of cost complementarities. A closer look at Table 3 reveals that cost complementarities do not behave in a uniform way. For small carriers large cost complementarities are associated mostly with product ratios which are higher than the ratio at the median point (usually 20% to 40%).[30] By using the local network to provide both toll and local calls they are able to increase production without incurring much additional cost. Being able to utilize a bigger local network by merging with another firm can be efficient in terms of benefits from cost complementarities. Thus, we should not be surprised to see mergers of this type of firms. In fact, most of the carriers operating in this region of the production mix belonged to the Contel family of local carriers, which gradually merged into GTE in the years following 1995.[31]

Since the data used here is from telephone companies operating in the post AT&T monopoly era, our ability to draw sharp conclusions concerning the AT&T divestiture case is limited. It is

[28] See specifications 3 and 4 in Table 3 in their paper (r is labeled there as PM).

[29] See the discussion at the end of Section 2.1.

[30] For large carriers the picture is less clear as the amount of complementarities are relatively small anywhere.

[31] See also footnotes to Table 2.

Table 3: Cost complementarities in production of local and toll calls

		local calls quantiles (levels $\times 10^6$minutes)								
		10 (350)	20 (550)	30 (933)	40 (2,033)	50 (2,747)	60 (4,010)	70 (9,448)	80 (12,320)	90 (19,915)
toll calls quantiles (levels $\times 10^6$min.)	10 (87)	.006	1.12	.85						
	20 (118)	.006	1.20	.89	.004	.87	1.33			
	30 (181)	.006	1.32	.96	.004	.93	1.40			
	40 (245)		3.01	3.57	4.28	.99	0.93			
	50 (420)		4.45	5.79	7.93	1.11	1.01			
	60 (790)				.016	1.05	1.08	1.13	1.20	
	70 $(1,367)$					.97	1.22	1.05	1.11	1.24
	80 $(2,470)$					1.10	1.14	1.19	1.27	1.45
	90 $(4,274)$							.005	.005	.005

plausible, however, to conclude that if the results in Table 3 can be extrapolated to the output bundle produced by AT&T, then cost complementarities could have been zero or even negative. This extrapolation of the results supports the claim that AT&T grew beyond its optimal size and merited a divestiture.

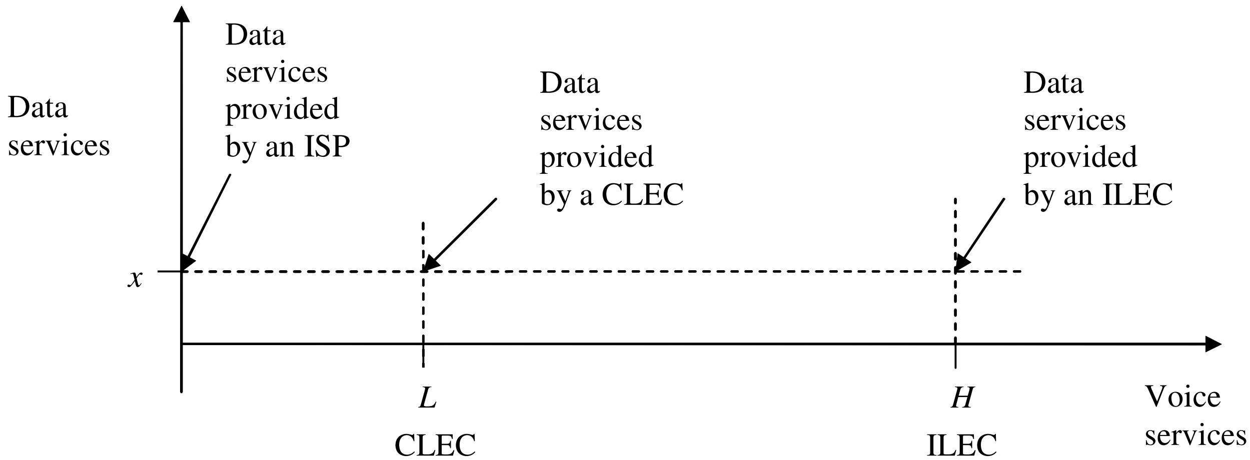
The methods developed in this paper can be applied to a wide range of questions and markets. I choose to demonstrate this claim by analyzing another relevant policy question in the telecommunications industry.

The Internet Freedom and Broadband Deployment Act

As a result of the divestiture of AT&T, the US local telephone market was divided into nearly 200 local access and transport areas (LATAs). The baby-bell companies who were the incumbent LECs (ILECs) were only allowed to provide services within each LATA they served (intraLATA). The 1996 Telecommunication Act lifted this constraint, allowing the ILECs to provide services between LATAs (interLATA) as long as they fulfill the 14 requirements detailed in section 271 of the Act. These requirements mostly involve opening the local network to other competitive LECs (CLECs). The act applies to both voice and data services. During its 107th session, the House of Representatives passed the "Internet Freedom and Broadband Deployment Act" (H.R. 1542, also known as the Tauzin-Dingell Bill). If enacted, the bill will allow the ILECs to provide interLATA data services without having to fulfill the requirements of section 271. The ILECs, who are the main proponents of this bill, claim that since cable companies (who also provide data services) are not subjected to section 271 requirements, they enjoy an unfair advantage. On the other hand, the opponents of this bill argue that lifting these requirements will substantially diminish the incentives that ILECs have for opening their local networks and therefore will stifle the competition in the local telephone markets.

A key element in the debate about the Tauzin-Dingell Bill involves the extent to which the local telephone network can be utilized to provide additional services beyond local and toll calls and the degree of cost complementarities between the existing services and the new services. To demonstrate the need to consider cost complementarities when considering mergers between providers of data services and phone companies, I use the estimates reported in Table 3. The patterns of cost complementarities depicted there are assumed to apply to the new product. In other words, cost complementarities between phone calls and internet services are assumed to exhibit the same pattern of complementarities as local and toll calls. For simplicity, I treat both local and toll calls as one product (voice services) and internet services as the second product (data services). Given this assumption, we analyze the following case. Assume that quantity

Figure 3: Potential complementarities between voice and data services



x of data services is provided in a certain market by an internet service provider (ISP). Two options are considered: a) a merger between the ISP and a CLEC who will provide both voice and data services - i.e. producing a bundle (L, x); b) a merger of the ISP with an ILEC who will provide both voice and data services - i.e. producing a bundle (H, x). $L < H$ are the amounts of voice services currently provided in this market by the CLEC and the ILEC respectively. This situation is depicted in Figure 3.

Assuming that $C(0, 0) = 0$, we assume that

$$\begin{aligned}\frac{C(0,x)+C(L,0)}{C(L,x)} &= (1+\alpha_1)\\ \frac{C(0,x)+C(H,0)}{C(H,x)} &= (1+\alpha_2)\end{aligned}$$

where α_1 and α_2 are the cost complementarities with respect to the origin at the relevant production levels. If the ISP merges with the CLEC, the total cost in producing the bundle $(L + H, x)$ is

$$T_1 = C(L,x) + C(H,0) = \frac{C(0,x)+C(L,0)}{1+\alpha_1} + C(H,0).$$

If the ISP merges with the ILEC, the total cost in producing the bundle $(L + H, x)$ is

$$T_2 = C(L,0) + C(H,x) = \frac{C(0,x)+C(H,0)}{1+\alpha_2} + C(L,0).$$

To see whether $T_2 > T_1$ we look at

$$T_2 - T_1 = \left\{ C(0,x) \left[\frac{1}{1+\alpha_2} - \frac{1}{1+\alpha_1} \right] \right\} + \left\{ C(L,0) \frac{\alpha_1}{1+\alpha_1} - C(H,0) \frac{a_2}{1+\alpha_2} \right\}. \quad (12)$$

When $\alpha_1 > \alpha_2$, the first element on the right-hand side of (12) is positive. The second element on the right-hand side of (12) can be evaluated using the estimates from Table 3 and the observed costs in the data. Consider the following numerical example. Both Ohio Bell and United Telephone of Ohio operated in the same market in 1995. Assume that the complementarities between data services and voice services are at the same levels as in Table 3.[32] Table 4 computes the second element on the right-hand side of (12).[33]

Table 4: The effect of complementarities on production of voice and data services

	Cost complementarities $\alpha =$	Total cost in producing voice services only (10^6\$) $C(y,0) =$	Effect on Cost savings (10^6\$) $\frac{\alpha}{1+\alpha} C(y,0) =$
United Tel. of OH, $y = L$	$\alpha_1 = 7.5\%$	348	24.3
Ohio Bell, $y = H$	$\alpha_2 = 1.2\%$	1807	21.4
Difference			2.9

This example shows that due to higher complementarities the bundle $(L+H, x)$ can be produced at a lower total cost if the ISP merges with the CLEC and not with the ILEC. It would be interesting to compare this analysis to results based on data from companies who provide both voice and data services.

5 Conclusions

This paper examines the structure of the US market for local and toll calls between 1988 and 1995 using data on Local Exchange Companies (LECs). I estimate the total cost function as a function of outputs, factor prices and additional characteristics of the firm. The contribution of this paper is twofold. First, it offers a methodological discussion on the identification of properties of cost functions. I show that identification of economies of scope and subadditivity depends on the distance of the support of the covariates from the origin. Partial or complete lack of identification is likely to occur. However, I am able to show that in many cases cost

[32] The results hold as long as the ratio between a_1 and a_2 is the same as in Table 3.

[33] Table 4 shows rounded figures reflecting the two firms' costs in 1995 .

complementarities can be identified with nonparametric methods. The second contribution concerns the estimation methodology. I describe a semiparametric estimation procedure that takes into account conditions coming from economic theory involving mild separability assumptions but without having to impose a specific functional form on the regression.

The empirical results establish the existence of cost complementarities in the production of local telephone services. Moreover, cost complementarities tend to be higher for small and medium companies than for larger companies like the Baby-Bells. These results suggest that the degree of cost complementarities in production of similar outputs should be taken into account when mergers of telecommunications companies are considered. The importance of the patterns of cost complementarities in analyzing telecommunications markets calls for further research.

A Technical details of the estimator

This appendix discusses the technical aspects of the estimator of the regression in (9). The estimation technique employs the backfitting algorithm developed in Hastie and Tibshirani (1987). This method is based on the following three equations:

$$E(c - g_2(W) - g_3(X)|Y) = g_1(Y) \tag{A-1}$$

$$E(c - g_1(Y) - g_3(X)|W) = g_2(W) \tag{A-2}$$

$$E(c - g_1(Y) - g_2(W)|X) = g_3(X). \tag{A-3}$$

If $\hat{g}_2(W)$ and $\hat{g}_3(X)$ are good estimators of $g_2(W)$ and $g_3(X)$ then $g_1(Y)$ can be estimated by regressing $c - \hat{g}_2(W) - \hat{g}_3(X)$ on Y. The same logic applies to $g_2(W)$ and $g_3(X)$ using (A-2) and (A-3). The steps of the backfitting algorithm are summarized in Table 5.

Table 5: The backfitting algorithm

Step 0 :		Select initial estimates $\hat{g}_1^0 \equiv 0, \hat{g}_2^0 \equiv 0, \hat{g}_3^0 \equiv \frac{1}{N}\sum c_i$
Step i :	i_1	Obtain $\hat{g}_1^i$ by nonparametrically regressing $c - \hat{g}_2^{i-1} - \hat{g}_3^{i-1}$ on Y imposing monotonicity and possibly submodularity
	i_2	Obtain $\hat{g}_2^i$ by nonparametrically regressing $c - \hat{g}_1^{i-1} - \hat{g}_3^{i-1}$ on W imposing monotonicity and homogeneity of degree 1
	i_3	Obtain $\hat{g}_3^i$ by a linear regression of $c - \hat{g}_1^{i-1} - \hat{g}_2^{i-1}$ on X
Convergence		Continue iteration until there is little change in individual estimates

The advantage of this algorithm is in its ability to take into account restrictions coming from economic theory on the elements of the model. I start by describing the steps in Table 5.

Step i_1:

$g_1(Y)$ is estimated using a constrained piecewise-linear spline developed in Beresteanu (2001, Chapter 2). I bring here a short description of the estimator. The function g_1 is written in terms of log outputs rather than the levels for a reasons explained before Corollary 2. The vector of log outputs is scaled to be in the box $[0,1]^2$. This box is then divided into equally spaced grid. The estimator for g_1 is a piecewise linear spline whose knots are the points of the equally spaced grid. In Beresteanu (2001) I show that it is enough to impose monotonicity and submodularity on the values that the estimator takes on the grid points and that linear interpolation preserves these properties over the whole domain. Thus, the values that $g_1(y)$ takes on the grid are chosen to minimize a least squares criteria subject to linear inequality constraints assuring that it satisfies the desired properties. The exact form of these linear inequality restrictions are described in Beresteanu (2001, Chapter 2). Next, the grid parameter (m_1, m_2) is chosen to minimize the mean square error of the estimator. The choice of the optimal grid parameter is based on a variant of the cross-validation procedure. Cross-validation procedures can be computationally intensive. With 401 observations, one has to run the backfitting algorithm 401 times for each possible value of (m_1, m_2). The cross-validation function is

$$CV(m_1, m_2) = \frac{1}{N} \sum_{i=1}^{N} (c_i - \hat{g}_{1,-i}(y_i) - \hat{g}_{2,-i}(w_i) - \hat{g}_{3,-i}(x_i))^2 \tag{A-4}$$

where c_i is the observed cost and $\hat{g}_{1,-i}, \hat{g}_{2,-i}$ and $\hat{g}_{3,-i}$ are the estimators for $\hat{g}_1, \hat{g}_2$ and $\hat{g}_3$ respectively based on the sample excluding observation i. I modify (A-4) in the following way. It is more important to have better predicting power in areas where data are dense. These parts of the support have greater influence on the value of $CV(m_1, m_2)$. Therefore, I randomly choose K observations $(K << N)$, $i_1, ..., i_K$, and compute the following modified cross validation function

$$\widetilde{CV}(m_1, m_2) = \frac{1}{K} \sum_{k=1}^{K} (c_{i_k} - \hat{g}_{1,-i_k}(y_{i_k}) - \hat{g}_{2,-i_k}(w_{i_k}) - \hat{g}_{3,-i_k}(x_{i_k}))^2 . \tag{A-5}$$

Random sampling imitates the density of the sample. K should be big but not too big to maintain a feasible computation time. I used $K = 40$.

The grid that minimizes the modified cross validation criteria in (A-5) is $(m_1 = 7, m_2 = 7)$. The grid contains 49 boxes out of which only 20 contain at least one observation. This yields 35 grid points which consist of the boxes corners. The number of linear inequality restrictions imposed on the regression to achieve monotonicity is 54 and additional 20 are required for submodularity.

Step i_2:

$g_2(W)$ is estimated using the estimator described in Beresteanu (2001, Chapter 3). This estimator is based on transforming the data to its polar representation. The data contains three prices L_p, K_p and M_p. The polar representation used here is

$$\begin{aligned} r &= \sqrt{L_p^2 + K_p^2 + M_p^2} \\ \theta_1 &= \arctan\left(\frac{K_p}{L_p}\right) \\ \theta_2 &= \arctan\left(\frac{M_p}{K_p}\right). \end{aligned}$$

As explained in Beresteanu (2001, Chapter 3), g_2 can be decomposed to $g_2(L_p, K_p, M_p) = h(\theta_1, \theta_2) + \log(r)$ where $h(\cdot, \cdot)$ is estimated nonparametrically using kernel smoothing. The result is a homogeneous of degree 1 and monotone estimator.

Step i_3:

$g_3(X)$ is assumed to be log-linear. This is done for simplicity due to the high dimension of X and the fact that no assumptions are made on the influence of these variables. More specifically, I use the four variables described in Table 1, time dummies to take into account possible changes in technology and another dummy variable to indicate if the firm is a "baby bell".

Convergence

Convergence of the backfitting algorithm is determined using a measure of change in the estimators from step $i-1$ to step i:

$$R_j = \frac{1}{N} \sum_{k=1}^{N} \left(\hat{g}_j^i(Y_k) - \hat{g}_j^{i-1}(Y_k) \right)^2$$

for $j = 1, 2, 3$.

Finally, The functions g_1, g_2 and g_3 in model (9) are identified up to an additive constant. I normalize by associating the constant with the linear part of the model, g_3. The backfitting algorithm showed convergence after 20 iterations (all R_j were of order 10^{-4} or lower).

Bootstrap

As Hastie and Tibshirani (1987) show, the above procedure converges and is consistent. However, the asymptotic distribution is not clear especially in light of the shape restrictions imposed on the estimator. To circumvent the need for asymptotic distribution a bootstrap procedure for building confidence intervals is used in this paper. The specific variant employed here is the *wild bootstrap* which is shortly described below. The model is estimated under the null once using the optimal mesh $m_1 = m_2 = 7$ and once using a smaller mesh $m_1 = m_2 = 5$ (over-smoothing). The residuals are computed from the model estimated with the optimal mesh. The residual in each observation point is then used to construct a discrete distribution with mean zero and second and third moments that imitate the second and third moments implied by the residual at this point. Hardle (1990, page 106) gives the exact formula of this distribution. Using this distribution a new residual is drawn for each observation. The bootstrap samples are constructed by keeping the covariates unchanged and computing the left-hand side variable by adding the bootstrap residuals to the fitted values computed from the over-smoothed estimator. Each bootstrap sample is then used to compute the test statistic. The 500 test statistics are ordered and the $0.95 \cdot 500^{th}$ value is the 95% critical value. This procedure preserves the possible heteroscedastic nature of the residuals and maintains the properties of the conditional model under the null. A comprehensive discussion of the merits, properties as well as the limitations of this bootstrap method appears in Horowitz (2001).

Estimators for g_1, g_2 and g_3

The function g_1 is a function of the two outputs - local and toll calls. Table 6 reports the estimator for g_1 computed at the quantiles of the outputs. This estimator was used to compute the complementarities reported in Table 3. The function g_2 is a function of three prices and can not be tabulated or graphed. Instead I computed the average derivatives of $\hat{g}_2$ with respect to the three outputs. They are reported in Table 7. The average derivatives are the expected value of the partial derivatives with respect to each log price. The method for estimating average derivatives is described in Hardle and Stocker (1989).

Table 6: Nonparametric piecewise linear spline estimator for g_1

$10^{-2}\times$		log local calls quantiles								
		10	20	30	40	50	60	70	80	90
	10	.274	1.537	3.199	4.375	4.580	4.582			
	20	.493	1.663	3.200	4.376	4.580	4.582			
	30	.805	1.842	3.202	4.377	4.581	4.583			
log toll	40		2.320	3.385	4.405	4.582	4.584			
calls	50		3.740	4.011	4.498	4.583	4.584			
quantiles	60			4.226	4.532	4.585	4.586	4.589	4.590	
	70					4.587	4.588	4.590	4.591	4.592
	80					4.590	4.590	4.591	4.592	4.593
	90							5.060	5.060	5.061

Table 7: Average derivatives of the log-cost function with respect to the log-prices in g_2

	labor	material	capital
average derivative	.25	.24	.42

The function g_3 is reported in Table 8. The following conclusions can be drawn from the coefficients of g_3. The effect of technological changes on cost is measured using time dummies. The coefficients of these dummies are negative, with some of them being insignificantly different from zero, and they become more negative with time.[34] Negative coefficients imply decreasing

[34] The 95% confidence interval for the coefficients is constructed using the 500 bootstrap samples that

costs with time. We can also see that the coefficient of tk is negative. The use of electronic switching equipment rather than analogue equipment reduces costs. The amount of access lines served definitely increases costs. The coefficient of *baby Bell* is negative as well indicating that the baby Bell companies are on average more cost efficient. Other variables are not significantly different than zero.

Table 8: Coefficient estimators for g_3

Estimator of g_3			
	point estimator	confidence interval	
variables	coefficients	2.5% lower bound	97.5% upper bound
intercept	13.364	13.356	13.398
baby Bell	-.0682	-.0936	-.0619
1989 dummy	.0129	-.00293	.0249
1990 dummy	-.0107	-.0299	.00430
1991 dummy	-.0230	-.0523	-.0155
1992 dummy	-.0318	-.0730	-.0306
1993 dummy	-.0576	-.109	-.0570
1994 dummy	-.0896	-.131	-.0881
1995 dummy	-.109	-.166	-.110
average loop length	-.0133	-.0173	.0062
central offices	.00490	-.00188	.0213
switching equipment	-.0130	-.0183	-.0073
access lines	1.392	1.337	1.400

are used to test the (sm) assumption.

References

Baumol, W. J. (1977). On the proper cost tests for natural monopoly in a multiproduct industry, *American Economic Review* **67**(5): 809–822.

Baumol, W. J., Panzar, J. C. and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace Jovanovich, Publishers, New York.

Beresteanu, A. (2001). *Nonparametric Estimation of Supermodular Functions with Application to the Telecommunication Industry*, PhD thesis, Northwestern University.

Beresteanu, A. (2004). Nonparametric estimation of regression functions under restrictions on partial derivatives, Working Paper 06-04, Duke University.

Brock, G. W. (2002). Historical overview, *in* S. K. M. Martin E. Cave and I. Vogelsang (eds), *Handbook of Telecommunications Economics*, Vol. 1, North-holland, chapter 2, pp. 44–74.

Diewert, W. E. and Wales, T. J. (1991). Multiproduct cost functions and subadditivity tests: A critique of the evans and heckman research on the u.s. bell systems, The University of British Columbia, Department of Economics, Discussion Paper No. 91-21.

Evans, D. and Heckman, J. (1983). Multiproduct cost function estimates and natural monopoly tests for the bell system, *in* D. Evans (ed.), *Breaking Up Bell: Essays on Industrial Organization and Regulation*, North-Holland, chapter 10, pp. 253–282.

Evans, D. and Heckman, J. (1984). A test for subadditivity of the cost function with an application to the bell system, *The American Economic Review* **74**(4): 615–623.

Fuss, M. A. and Waverman, L. (2002). Econometric cost functions, *Handbook of Telecommunications Economics*, North-Holland, chapter 5, pp. 143–177.

Gabel, D. and Kennet, D. (1994). Economies of scope in the local telephone exchange market, *Journal of Regulatory Economics* **6**: 381–398.

Gallant, R. (1982). Unbiased determination of production technologies, *Journal of Econometrics* **20**: 285–323.

Hardle, W. (1990). *Applied Nonparametric Regression*, Econometric Society Monographs, Cambridge University Press.

Hardle, W. and Stocker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives, *Journal of the Americal Statistical Association* **84**: 986–995.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models, *Statistical Science* **1**(3): 297–318.

Hastie, T. and Tibshirani, R. (1987). Generalized additive models: Some applications, *Journal of the American Statistical Association* **82**(398): 371–386.

Horowitz, J. L. (2001). The bootstrap, *in* J. J. Heckman and E. Leamer (eds), *Handbook of Econometrics*, Vol. 5, Elsevier Science B.V., chapter 52, pp. 3159–3228.

Lovell, C. K. and Sickles, R. C. (1999). Causes and consequences of expert disagreement: Methodological lessons from the u.s v. at&t debate, *in* D. J. Slottje (ed.), *The Role of the Academic Economist in Litigation Support*, Amsterdam: North-holland, chapter 12, pp. 189–205.

Matzkin, R. (1994). Restrictions of economic theory in nonparametric methods, *Handbook of Econometrics IV*, Vol. 4, North-Holland, chapter 42, pp. 2524–2558.

Milgrom, P. and Roberts, J. (1990). The economics of modern manufacturing: Technology, strategy, and organization, *American Economic Review* **80**(3): 511–528.

Milgrom, P. and Shannon, C. (1994). Monotone comparative statics, *Econometrica* **62**(1): 157–180.

Olley, S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry, *Econometrica* **64**(6): 1263–1297.

Panzar, J. (1989). Determinants of firm and industry structure, *Handbook of Industrial Organization*, North-Holland, chapter 1, pp. 3–59.

Roehrig, C. S. (1988). Conditions for identification in nonparametric and parametric models, *Econometrica* **56**(2): 433–447.

Roller, L. (1990). Proper quadratic cost functions with an application to the bell system, *The Review of Economics and Statistics* **72**: 202–210.

Sharkey, W. W. (1982). *The Theory of Natural Monopoly*, Cambridge University Press.

Shin, R. and Ying, J. (1992). Unnatural monopolies in local telephone, *RAND Journal of Economics* **23**(2): 171–183.

Topkis, D. (1998). *Supermodularity and Complementarity*, Princeton University Press, Princeton NJ.

Wilson, W. and Zhou, Y. (2001). Telecommunications deregulation and subadditive costs: Are local telephone monopolies unnatural?, *International Journal of Industrial Organization* **19**: 909–930.