ECONOMETRIC MODELING OF BUSINESS TELECOMMUNICATIONS DEMAND USING RETINA AND FINITE MIXTURES

MASSIMILIANO MARINUCCI AND TEODOSIO PÉREZ-AMARAL*

Departamento de Economía Cuantitativa, Universidad Complutense Madrid,

Campus de Somosaguas, 28223 Madrid.

This version: May 16th 2005
Comments are welcome

Abstract

In this paper we estimate the business telecommunications demands for local, intra-LATA and inter-LATA services, using US data from a Bill Harvesting® survey carried out during 1997. We model heterogeneity, which is present among firms due to a variety of different business telecommunication needs, by estimating normal heteroskedastic mixture regressions. The results show that a three-component mixture model fits the demand for local services well, while a two-component structure is used to model intra-LATA and inter-LATA demand. We characterize the groups in terms of their differences among the coefficients, and then use RETINA to perform automatic model selection over an expanded candidate regressor set which includes heterogeneity parameters as well as transformations of the original variables.

Our models improve substantially the in-sample fit as well the out-of-sample predictive ability over alternative candidate models. RETINA suggests that the final demand specification should include telephone equipment variables as relevant regressors. On the other hand, the output of the firm, as well as its physical extension, have second order, yet significant effects on the demand for telecommunication services. Estimated elasticities are different for the three demands but always positive for access form (single-line or private network).

JEL-Codes: C21, C51, C52.

Keywords: Telecommunication Demand Models, Local calls, inter-LATA calls, intra-LATA calls, RETINA, Flexible Functional Forms, Heterogeneity, Finite Mixtures.

*Correspondence to: Massimiliano Marinucci, Departamento de Economía Cuantitativa, Universidad Complutense Madrid, Somosaguas, 28223 Madrid. E-mail: `mmarinuc@ccee.ucm.es` or Teodosio Pérez-Amaral, E-mail: `teodosio@ccee.ucm.es`.

1 Introduction

The literature on econometric modeling of Telecommunications demand is very extensive[1]. The theoretical framework for the modeling is well known and goes back to Artle and Averous [1], Von Rabenau and Stahl [19] and Rohlfs [14] among others. Empirical studies on business demand are not so abundant. A relevant contribution in this field is the pioneering work of Ben-Akiva and Gershenfeld [2] which focuses on the demand for different types of access lines[2]. Another example is the preliminary study of Pérez-Amaral and Marinucci [12] where they estimated the demand for local telecommunications services by using Bill Harvesting® data[3].

A theoretical framework on business telecommunication demand is provided by Taylor [15]. He proposes a valuable approach, as he recognized that the standard approach, which considers telecommunications as an input of a production function along with capital and labor, is often too inflexible to describe the variety of different telecommunications needs existing among firms. He points out that determinants of demand may vary widely depending on the size, the activity sector and the localization of the business. Since there is an immense diversity of communications needs, the presence of unaccounted heterogeneity in the data can pose serious problems in the modeling process, and becomes an important empirical question.

In this study we are interested in estimating a model for predicting the demand for short distance (local), medium distance[4] (intra-LATA) and long distance (inter-LATA) business telecommunication services. We use a cross-section of US telecommunications data from the Bill Harvesting® survey, carried out in 1997.

Modeling business telecommunications demand is not a trivial matter. We face two difficulties: 1) prices charged for the services are not available 2) heterogeneity is present among firms. In this situation, we do not have a precise, a priori hypothesis about the functional form to adopt for the demand equation.

We estimate the parameters of the unknown relationship which relates demand with available information by using 1) the model search features of RETINA[5] to select useful variables for prediction and 2) the clustering flexibility of the finite

[1] Many residential demand studies use Bill Harvesting® or other customer databases. For example Kridel, Rappoport and Taylor [4] presented a study of carrier choice, usage demand and price elasticities for the residential intra-LATA toll market using Bill Harvesting® data. Taylor [16] estimated competitive cross-price elasticities for the residential intra-LATA toll market with a two stage approach and using Bill Harvesting® data. Levy [6] estimated a semi-parametric generalized additive Tobit model of residential Intra-LATA Telephone demand on a cross-section of residential telephone consumers across 28 states using bills of GTE customers.

[2] They consider a discrete choice framework to estimate price elasticities with respect to the choice of different telephone systems (PBX, Centrex).

[3] Bill Harvesting® is a proprietary methodology of PNR & Associates (now TNS Telecoms).

[4] Medium and long distance calls are defined on the basis of communications between LATA's. LATA stands for "Local Access and Transport Area". A toll call from one point within a LATA to another point within the same LATA is an intra-LATA or medium distance toll call. Similarly a toll call from a point within a LATA to a different LATA is an inter-LATA call or long distance toll call.

[5] RETINA: RElevant Transformations of the Inputs Network Approach.

mixture modeling framework, in order to identify sources of heterogeneity in the data.

Retina (Pérez-Amaral, Gallo and White[11]) is an automated model selection tool which allows flexibility in specification, suggesting which variables or transformations can be used to approximate the unknown functional relationship. Its goal is to provide a model with improved predictive ability over alternative specifications.

Recent developments in finite mixture modeling (McLachlan and Peel [9]) have provided techniques to partition the data into homogeneous subgroups and perform regression analyses within each group simultaneously. Thus, by using Retina in combination with finite mixture models, we overcome the difficulties generated in 1) demand model specification and 2) accounting for heterogeneity among firms.

The rest of the paper is organized as follows: Section 2 discusses Taylor's approach in modeling Business Toll Telecommunications Demand, Section 3 describes the data, Section 4 discusses the empirical model and estimation techniques, Section 5 presents the results and Section 6 contains the conclusions. The appendixes follow.

2 Business Toll Demand Modeling

A formal discussion of business telecommunications demand, considered as a production input within a classical microeconomic framework, may be found in textbooks. Demand, considered as a function of the price and other production factors, is derived from the firm's cost minimization conditions. Nonetheless, due to limitations in the available information, this approach is not always useful in empirical research, as in our case.

Rather than formally discuss a microeconomic model of business telecommunications demand, we will follow an informal approach which recognizes the differences among firms in satisfying their communications needs and uses. For this purpose we will summarize the framework used by Taylor [16], which is the starting point of our empirical analysis. Taylor divides firms into four generic types, where each type is referred to as a stage. Firms need telecommunications services not only for external communications but also for internal use, and this need increases nonlinearly with the size, the location and the activity sector of the firm.

Stage I firms are assumed to operate from a single location and are supposed to have mostly external communications needs. Moreover, they are supposed to access the public network with few single-line telephone systems. These are usually small-sized businesses.

Stage II firms have multiple locations in the same locality. As the number of employees increases, the internal use of telecommunications grows. Increased usage can be accommodated by increasing the number of lines until the purchase (or rental) of a small private network is considered. Nonetheless, purchasing toll services in bulk (WATS, 800 service)[6] is frequent as a valid alternative to such a decision and small

[6] WATS: Wide Area Telephone Service is a flat rate or a special rate pay-by-the-minute (measured) billing for a specified calling area. It is usually offered by companies that buy transmission

Table 1: Summary of an a priori segmentation scheme proposed by Taylor†.

	Stage I	**Stage II to IV**
Locations	Single location	Multiple locations, same locality, or in multiple localities
Type of Usage	External	Internal +External
Type of access	Multiple Single Lines (Business Lines, Hunting Lines)	Multiple Single Lines Private Network (PBX, Centrex, WATS, 800 service)
Sociodemographic characteristics	The number of Employees may be low with respect to firms that are not in stage I	Number of employees larger than in stage I

†Based on Taylor [15].

businesses usually still work well with multiple single-lines.

Stage III firms in general tend to be larger than stage I and II firms. But the main difference is that they have multiple locations in different localities. Stage III firms may switch from multiple single lines to private networks if there is a sufficient volume of communications between fixed points. In this stage access to the public network is still required for external needs, while internal needs are largely satisfied by the private network. Nonetheless, frequently so-called smart switches are used to select the lowest cost for external or internal calls. This is done by routing a call over the private network and then into the corresponding local destination area.

Finally stage IV firms include multinational corporations located in multiple countries. The main difference with respect to previous stages is their bigger size and the fact that their workers are spread across different states and countries. Thus International Toll services are required for business activity.

3 The Data

Our Bill Harvesting® database consists of a cross-section of 13766 firms from south-eastern United States. The data has been provided by PNR & Associates (Philadelphia, PA) which today forms part of the TNS group.

Since the AT&T divestiture (January 1^{st}, 1984) local telecommunication services in this area are provided in a quasi-monopoly regime by Bell South. In fact 78% of the firms were served by this company and the rest by other independent carriers.

Prior to the model specification a large preprocessing stage was undertaken. Some details are reported in the appendix. Only 4463 observations had complete

capacity in bulk from other network operators in order to re-offer it to customers at lower prices.

data and our effective sample size varies along with the type of demand. Local services are used by all the firms, while intra-LATA and inter-LATA services have been used by only 29% and 27% of the businesses, respectively. Descriptive analyses and estimations were carried out twice, first just by using complete records, and second by using the total data set of 13743 observations where missing information is imputed with a method suggested by Troyanskaya et al. [18]. In general, results over the imputed data set differed slightly with respect to the results obtained over the reduced record set, and the results are not reported here.

After this preprocessing stage and prior to any estimation, outliers were detected by using an automated procedure proposed by Peña and Yohai [10]. The procedure is implemented in Retina Winpack[7] and may be run optionally by the user prior to model selection. This reduced the effective sample size of local demand to 4391 firms, while 1261 were kept for intra-LATA demand and 1176 for inter-LATA demand. Also, prior to estimation, data have been rescaled to avoid the potential negative effect of different orders of magnitude.

Since visual inspection of the histograms and empirical densities of the original variables shows highly skewed distributions, *log* transforms have always been considered. Logarithmic transformations tend to normalize the data, stabilize the variances and limit the potential negative effect of the most extreme observations. Variables with zero values, such as the number of lines, have all been augmented by a unit constant prior to transformations. We also consider $log - ratio$ transforms, by using the *log* of the ratio between the original variables *(BUS, HUN, PBX, CTX, SAL, EMT, SQFT)* and the number of workers employed locally *EMH*. Worker per capita forms, obtained by dividing the variables by the number of employees working locally *EMH*, have been chosen since they are common in the literature and reduce heteroskedaticity. A description of the original variables is reported in Table 2, while descriptive statistics of their transformations over the complete data sample are given in Table 3 and Figure 1.

The data include four types of variables:

Access form variables: There are four different types of lines, which may be grouped into two categories. The first includes single-line access equipment: business lines (BUS) and hunting lines (HUN). The second group represents private network access forms and includes PBX trunks (PBX) and Centrex lines (CTX).

Socio-demographic variables: These are the population habitat size (POP) and the States (*AL, GA, KY, LA, MS, NC, SC, TN*).

Business size and dispersion related variables, such as the number of employees in the whole business (EMT), the number of workers employed locally (EMH) and the physical extension of the firm ($SQFT$).

Output variable: the sales of the firm (SAL).

[7] See Appendix B and [8] for more details about Retina Winpack.

Table 2: 1997 Bill Harvesting Data: Variable definitions†.

Variable	Description
LOCAL	Total expenditures for local calls in dollars
INTRA	Total duration of intra-LATA calls in minutes
INTER	Total duration of inter-LATA calls in minutes
BUS^a	Number of Business Lines +1
HUN^b	Number of Hunting Lines +1
PBX^c	Number of PBX Trunks +1
CTX^d	Number of Centrex Lines +1
SAL	Sales expressed in dollars
EMT	Total number of employees
EMH	Number of employees working locally
SQFT	Square footage of the firm
POP	Population habitat size
IMILLS	Inverse of the Mills ratio (see Appendix B)
STAGE I	Binary variable= 1 if Firm is at stage I
BSOUTH	Binary variable= 1 if Service is provided by Bell South
AL	Binary variable= 1 if Alabama
GA	Binary variable= 1 if Georgia
KY	Binary variable= 1 if Kentucky
LA	Binary variable= 1 if Lousiana
MS	Binary variable= 1 if Missouri
NC	Binary variable= 1 if North Connecticut
SC	Binary variable= 1 if South Connecticut
TN	Binary variable= 1 if Tennessee
FL	Binary variable= 1 if Florida (omitted to avoid perfect colinearity)

†Source: PNR & Associates, Philadelphia, PA, now TNS.

a. *BUS*: Business Lines. A service that handles all the routine business telecommunications applications. Data transmissions for fax, email, and Internet access are usually charged at the same price as voice calls.

b. *HUN*: Hunting Lines. A service that bundles all the telephone lines (2 lines up) in the same location to be easily accessible with a single number (pilot number).

c. *PBX*: PBX Trunks. Connections between an organization's PBX (Private Branch eXchange) and the outside telephone network. Telephone users within the customer's company share these connections for making and receiving calls outside the company's network.

d. *CTX*: Centrex Lines. (Central office exchange service) is a service which is functionally equivalent to the PBX and consists of up-to-date phone facilities offered by the telephone company to business users so they do not need to purchase the equipment. The Centrex service effectively partitions part of its own centralized capabilities among its business customers. The customer is spared the expense of having to keep up with fast-moving technology changes and the phone company has a new set of services to sell. In many cases, Centrex has now replaced the private branch exchange. The central office has effectively become a huge branch exchange for all of its local customers. In most cases, the Centrex service provides customers with as much if not more control over the services they have than PBX did.

Notice that Business and Hunting Lines can be considered as single line access forms while PBX and Centrex services are network access forms.

Table 3: Univariate statistics of the *log* of each variable per worker.

	Mean	Std. Dev.	Median	Kurtosis	Skewness	n
$\ln(LOCAL/EMH)$	2.556	1.049	2.613	.580	$-.135$	4391
$\ln(INTRA/EMH)$	1.296	1.767	1.428	.001	$-.416$	1261
$\ln(INTER/EMH)$	2.538	1.573	2.693	$-.108$	$-.322$	1176
$\ln(BUS/EMH)$	-1.614	1.864	-1.061	.147	$-.898$	4391
$\ln(HUN/EMH)$	-1.919	1.447	-1.609	.820	$-.691$	4391
$\ln(PBX/EMH)$	-2.490	1.354	-2.398	.207	$-.323$	4391
$\ln(CTX/EMH)$	-2.259	1.699	-2.197	.294	$-.295$	4391
$\ln(SAL/EMH)$	1.249	3.499	.182	.280	1.176	4391
$\ln(EMT/EMH)$	.249	.706	.000	21.277	4.217	4391
$\ln(SQFT/EMH)$	5.928	1.235	5.968	1.399	$-.273$	4391
$\ln(POP)$	1.193	2.273	9.770	-1.461	.210	4391

Figure 1: Descriptive statistics, by the stage of the firm.

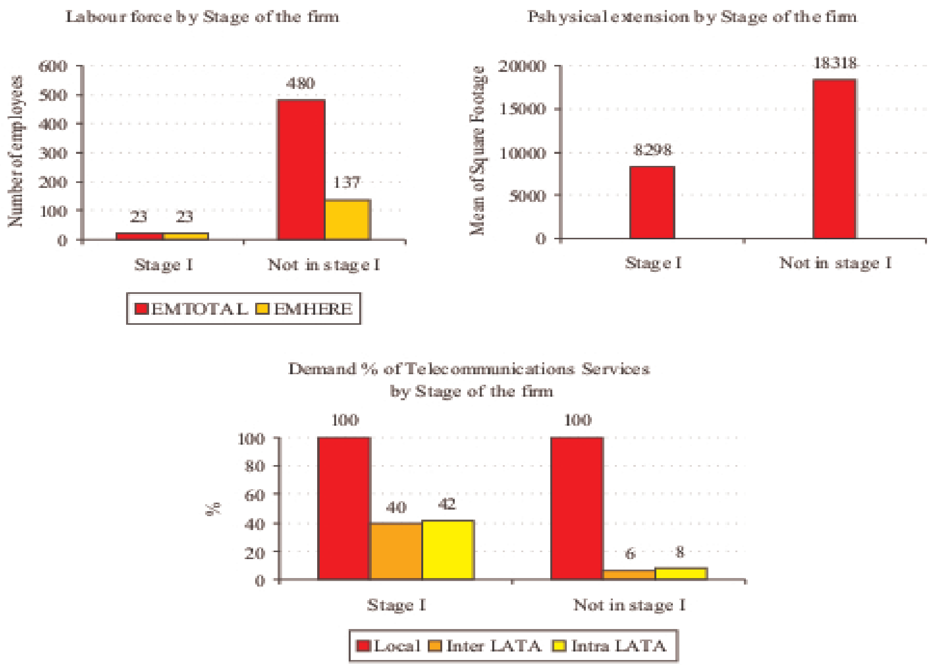


Table 4: Demand by type of access: Firms that demand intra-LATA and inter-LATA calls do not own private networks (Vertical %).

Type of access	**Firm demands only Local calls** ($n = 2921$)	**Firm demands intra-LATA or inter-LATA calls** ($n = 1542$)
Firm owns Multi-Single Lines (Business or Hunting lines)	80.2%	99.6%
Firm owns Private Networks (PBX or Centrex)	39.3%	.7%

Bivariate plots of the transformed variables, reported in Figures 2, 3 and 4, announce that the modeling problem is difficult especially because of non-linearities and heterogeneity among businesses with respect to telecommunications services.

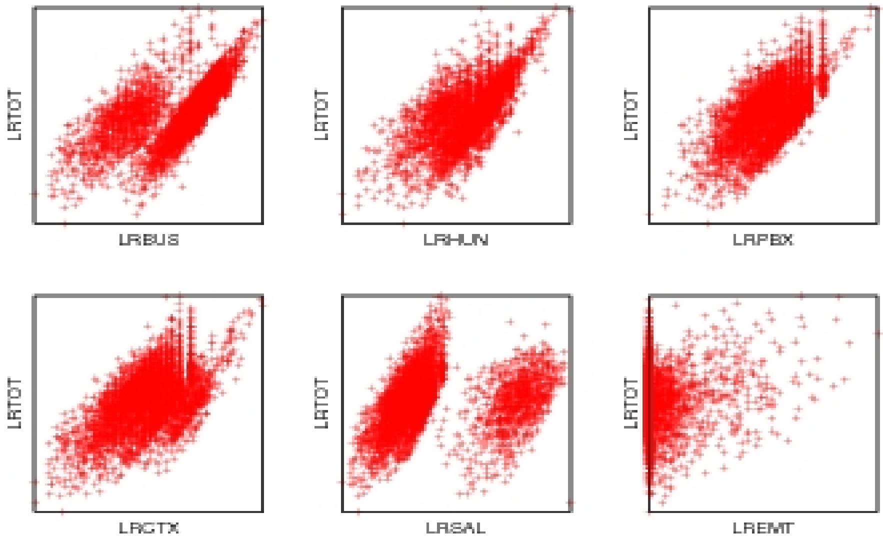
From these plots we find initial evidence of heterogeneity. In some cases as for local demand (Figure 2), two moderately separated clusters may be visually identified. Clusters appear upward sloping and as elliptic shaped clouds, suggesting that they may have different mean and covariance structures. Demand for medium and long distance services (Figures 3 and 4) also accounts for evident heterogeneity especially with respect to the firm output proxied by sales (SAL). Nonetheless for the remaining variables, heterogeneity is visually much less evident and statistical methods are necessary to assess its existence.

A new variable called Stage I was also added to the analysis. This is a dummy variable which proxies Taylor's definition of Stage I firms, i.e., single location businesses with only a single-line access form used for external communication purposes[8]. This variable is used to show some other interesting facts as reported in Figure 1. For example note that stage I firms are on average smaller, in terms of number of employees and their physical extension, than firms at higher stages, although to some extent this also depends on the nature of the markets sold in. Moreover, Figure 1 shows that intra-LATA and inter-LATA services are almost exclusively demanded by stage I firms. On the contrary, bigger or multiple-location firms that are not at stage I make a more intensive use of local services. Yet this seems plausible only if such firms use some "smart" switches which route non-internal calls over the private network and then into the appropriate local area through a local call.

Finally, from Table 4 we also learn that firms using intra-LATA or inter-LATA services use almost exclusively single-line equipment access forms (99.6%). In other words the private network dimension will not play a relevant role in the explanation of medium and long distance calls and may be dropped without losing relevant information during the modeling process. We conclude this by bearing in mind that

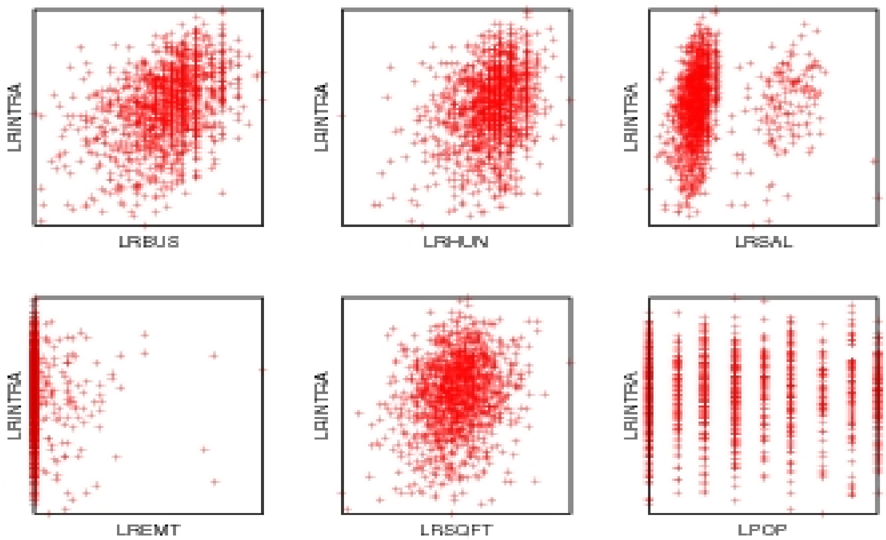
[8] Location conditions have been inferred from the difference between the number of employees working locally (EMH) and the total number of employees of the company (EMT). If the difference $EMT - EMH = 0$ then the firm is assumed to be single location.

Figure 2: Bivariate plots of Local demand vs. explanatory variables. (The LR-prefix stands for the *log* transformation of the original variables divided by *EMH*).



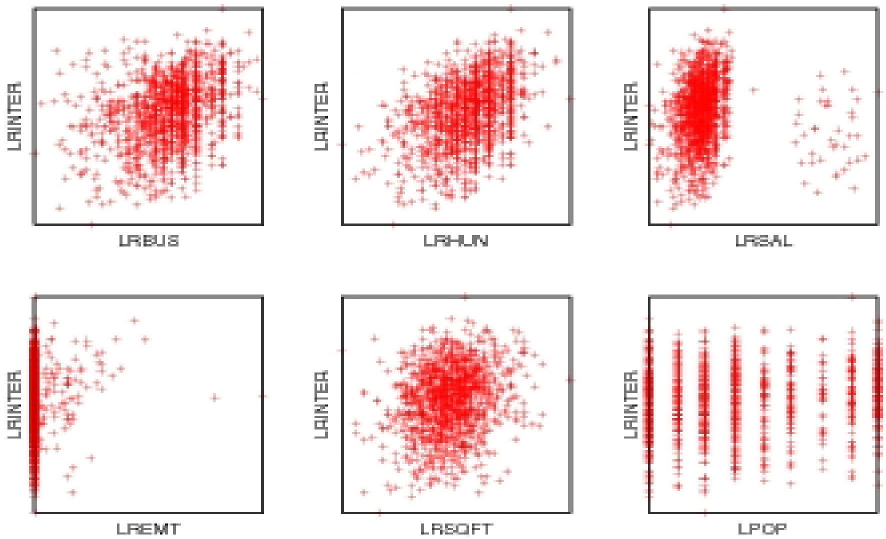
Two or more groups are visible. Heterogeneity patterns with respect to the demand for local services are visible for the number of Business Lines (LRBUS) and sales (LRSAL).

Figure 3: Bivariate plots of intra-LATA demand vs. explanatory variables. (The LR-prefix stands for the *log* transformation of the original variables divided by *EMH*).



Intra-LATA services show a possible two-cluster structure especially with respect to sales (LRSAL).

Figure 4: Bivariate plots of inter-LATA demand vs. explanatory variables. (The LR-prefix stands for the *log* transformation of the original variables divided by *EMH*).



Heterogeneity in inter-LATA services is not so evident. Non-linearities emerge from LREMT, the log of the ratio between the total number of employees (EMT) and the number of workers employed locally (EMH).

telecommunication services demanded by the firms are related with the dimension of the firm and its location.

4 Empirical Model and Estimation

4.1 Demand Specification

We start with the following telecommunication demand function.

$$D = f(\text{prices, employees, locations, markets sold in, output, equipment}) + error$$

However lack of data imposes some constraints on this specification. No prices are available, nor is there a direct measure of the number of locations of the business or the markets sold in. These limitations call for a reformulation of the demand. Thus we assume the function to be:

$$\begin{aligned} D &= f(\text{employees, physical extension, output,} \\ &\quad \text{equipment, socio-demographic variables}) + error \end{aligned}$$

Relevant information is available on the number of employees, which can give an idea about the dimension of internal communication needs. Output is proxied by sales although its relevance is unclear a priori, since phone calls are made by people and sales may influence the volume of calls only if the business involves a heavy tele-marketing activity. Sociodemographic variables such as the population habitat size and the geographic region are included in the general specification as well, but their effects are uncertain. The signs of the coefficients are expected to be positive in the case of the number of different types of lines and the relative size of the firm. Demand variables are defined as follows:

ln(LOCAL/EMH): $log - ratio$ of the expenditures in local calls in dollars per worker.

ln(INTRA/EMH): $log - ratio$ of the duration of intra-LATA calls in minutes per worker.

ln(INTER/EMH): $log - ratio$ of the duration of inter-LATA calls in minutes per worker.

Our analysis begins with the specification of a Benchmark Linear Model (BLM). The BLM is defined as the specified model without using any particular selection strategy. This is the natural starting point because the BLM represents a lower bound for evaluating the performance of our methodology. Recall that our final objective is specification of a Useful Representative Model or URM, which represents the function finally adopted for a representative firm. Its functional form is unknown a priori, however we would like 1) an improved predictive ability over the corresponding BLM, and 2) a parsimonious representation of the phenomena of interest. As starting point, we adopt a double logarithmic specification for the BLM, which expresses telecommunications demand in per worker terms as a function of the candidate predictors (or any transformation of them):

$$\begin{aligned}\ln\left(\frac{Y_{ij}}{EMH_i}\right) &= \beta_0 + \beta_1 \ \ln\left(\frac{BUS_i}{EMH_i}\right) + \beta_2 \ln\left(\frac{HUN_i}{EMH_i}\right) + \beta_3 \ln\left(\frac{PBX_i}{EMH_i}\right) \\ +\beta_4 \ \ln\left(\frac{CTX_i}{EMH_i}\right) &+ \beta_5 \ \ln\left(\frac{SAL_i}{EMH_i}\right) + \beta_6 \ \ln\left(\frac{EMT_i}{EMH_i}\right) + \beta_7 \ \ln\left(\frac{SQFT_i}{EMH_i}\right) \\ &+\beta_8 \ \ln\left(POP_i\right) + \varphi_1 \ IMILLS_i + \delta_1 \ STAGEI_i + \delta_2 \ BSOUTH_i + \\ &+ \sum_h \gamma_h \ POP_{ih} + u_i \qquad (1)\end{aligned}$$

Where $i = 1, \ldots, n \quad h = 1, \ldots, 9$ and u is the error term with $u \sim i.i.d. \ N(0, \sigma)$, $IMILLS$ is the inverse of the Mills ratio (which is explained in Appendix B) and Y_{ij} represents alternatively the total local bill ($j = 1$), the intra-LATA minutes ($j = 2$) or the inter-LATA minutes ($j = 3$). In the following sections we discuss how to obtain possible URM models, from which we can choose a final URM*.

4.2 Model Building and Selection using Retina

Equation (1) may be expressed in a more compact form as follows:

$$\ln\left(\frac{Y_j}{EMH}\right) = X'\beta + F'\delta + u \tag{2}$$

Where:

X: includes all the continuous candidate regressors which in the case of eq.(1) are: $\ln(BUS/EMH)$, $\ln(HUN/EMH)$, $\ln(PBX/EMH)$, $\ln(CTX/EMH)$, $\ln(SAL/EMH)$, $\ln(EMT/EMH)$, $\ln(SQFT/EMH)$, $\ln(POP)$.

F: includes the dummies, and the inverse Mills ratio.

In general, X represents a matrix of variables which may be transformed expanding the original set of regressors, while F represents a matrix of other auxiliary regressors that will not be transformed.

Now, if we allow the original regressor set X to be expanded by considering squares, cross-products, ratios and inverses of the original variables, we may generalize the demand as follows:

$$\ln\left(\frac{Y_j}{EMH}\right) = W'\beta + F'\delta + u \tag{3}$$

$W = x_i^{\alpha} x_j^{\beta}$ with $i, j = (1, \ldots, k)$ where k is the total number of untransformed continuous inputs.

$\alpha, \beta = -1, 0, 1$

The main difference with respect to the BLM specified in (1) is that here we allow transformations of the original regressors, while the BLM exclusively considers logs of ratios of variables per worker. We use (3) because we want Retina to generate the W transforms and identify which ones may help to predict better than the BLM.

We can further generalize equation (3) by using the dummy variables included in F to model group-specific slopes and allowing interactions between such dummy variables and the continuous regressors. Formally, assume H_g to be a subset matrix of F with $g-1$ columns, which represents some specific grouping which accounts for heterogeneity in the data set. This leads to:

$$\ln\left(\frac{Y_j}{EMH}\right) = W'\beta + [H_g \times W]'\beta_h + F'\delta + u \tag{4}$$

with: $H_g \subset F$.

This specification is akin to an analysis of covariance formulation where the parameters of W may vary across the categories by using dummy indicators included in F to model group-specific constants, or in H_g to model group-specific slopes.

4.3 Modeling Heterogeneity using Finite Mixtures

Since the presence of heterogeneity can in our case be visually detected from bivariate scatterplots as seen in Section 3, the problem of modeling heterogeneity may well be addressed by using available information at hand (geographic indicators, stage of the firm and so on). Nonetheless, this a priori information may not account for all the heterogeneity in the data set. Finite mixtures may then be used to detect or represent any additional group structure, if present, in the data.

The only assumption in this case is that the distribution of our dependent variable may be approximated as a weighted sum of normal distributions, each of which has an expected mean expressed as a function of the explanatory variables, without loss of generality if we define:

$$\ln\left(\frac{Y_j}{EMH}\right) = W'\beta_g + F'\delta_g + \sigma_g\, u$$

where $u \sim N(0,1)$ and $\beta_g, \delta_g, \sigma_g$ may assume different G values with probabilities $(\pi_1, \ldots, \pi_G)$: the conditional distribution of the dependent variable with respect to the candidate regressors may then be expressed formally as a mixture of G components as:

$$\ln(Y_j/EMH) \mid W, F \ \sim \sum_{g=1}^{G} \pi_g\, N(\, W'\beta_g + F'\delta_g, \sigma_g^2)$$

Using this formulation[9], the Expectation Maximization (EM) algorithm[10] is then used to estimate the maximum likelihood parameters of the regression equations of each group $\hat{\beta}_g, \hat{\delta}_g, \hat{\sigma}_g$, and the posterior probabilities $\hat{\pi}_g$ for each firm. Cluster membership (the H_g matrix) is then determined by assigning each observation to the group for which posterior probability is highest.

What is relevant for us is that this methodology allows us to obtain a consistent inference about H_g better suited to our objectives than any other traditional non-parametric clustering method, eg. K-means (MacQueen [7]), Ward (Ward [20]). Traditional clustering methods are concerned with grouping objects, in our case the firms, by minimizing some distance measure among them. The distance measures are defined on the basis of a specific metric which typically is chosen by the researcher on an a priori basis (Euclidean distance is usually considered). Thus traditional clustering methods do not involve the estimation of any a priori parametric model structure on the variables. With Finite Mixtures, on the contrary, distributional assumptions and conditional heterogeneity among the variables, rather than unconditional heterogeneity, are explicitly taken into account and a parametric (or semi-parametric) inference about a specific partition model is possible, as will be shown later.

[9] Note that we are assuming normal heteroskedastic components. See Appendix B for more details.

[10] See McLachlan and Peel [9] for a discussion on finite mixture modeling.

4.4 Combining RETINA and Finite Mixture Framework

The above discussion configures a two-step approach which is used to obtain a set of possible URM models from which we can choose a final URM*. The first step is to model heterogeneity, fitting finite mixtures to each demand. The second is to perform a variable selection on an expanded regressors set which, besides the original variables, also includes transformations of the form $x_i^{\alpha}\ x_j^{\beta}$ as well as heterogeneity parameters. More specifically:

1. First fit a mixture of regressions for each demand and for each proposed initial specification by estimating the maximum likelihood mixture parameters via the EM algorithm.

2. Decide the number G of clusters to be retained for subsequent analyses in each case (we use AIC and BIC).

3. Obtain the corresponding H_g matrixes (if any) by assigning each observation to the cluster for which the posterior cluster membership probability is highest.

4. Once a partition has been chosen, consider a general specification as in equation (4) but this time including the cluster membership matrix H_g:

   - into F in order to model group-specific constants
   - into H_g in order to model group-specific slopes

5. Then use Retina to automatically select only the most relevant predictors among W, F, and the $H_g \times W$ interactions between predictors and clusters. Obtain a candidate URM*.

This approach works well in practice. One can get different candidate URM's by running the above steps for different specifications of the inputs, namely X, F and H. All of them represent a *candidate model set* on which Multi-Model Inference, MMI [3] may be carried out by comparing the models on the basis of AIC and BIC criterion.

5 Empirical Results

In this section we present the main results of this study. Details about estimations are reported in the Appendix. We first present the results for the BLM models corresponding to each demand. We then comment the heterogeneity parameters we found via finite mixtures, and finally we discuss model selection, using Retina, as the final URM.

5.1 BLM Demand Models

In Table 9 in the Appendix, we report the Benchmark Linear Models for local, intra-LATA and inter-LATA demand. The estimations show that:

1. Demands appear to be sensitive to equipment variables (*BUS, HUN, PBX, CTX*.
2. Constant terms for intra-LATA and inter-LATA are not significant.
3. The sales (*SAL*) variables have wrong signs. This may be due to heterogeneity (see Figures 2, 3 and 4).
4. The Stage I indicator is negative for local calls, confirming that firms at stages higher than the first make a more intensive use of local services by routing long distance calls over their private network (*PBX*, *CTX*).
5. Dimension of the firm appears to be relevant for local services demand, again reflecting the fact that larger-sized firms demand *ceteris paribus* use more local services than firms at stage I.

The sample fit for local calls is quite satisfactory, ($\bar{R}^2 = .682$), but this is not the case of intra-LATA ($\bar{R}^2 = .191$) and inter-LATA demands ($\bar{R}^2 = .243$). These results suggest that alternative specifications should be taken into account.

5.2 Mixtures of Linear Demand Models

After applying the EM algorithm, the number of groups was selected by examining both the AIC and BIC criteria over three different specifications, say S_1, S_2 and S_3 where $S_1 \equiv$ BLM, S_2 the relevant regressors of the BLM selected by Retina, and S_3 the BLM excluding all dummy variables. The AIC and BIC statistics of the fitted mixture of linear demand models using S_1, S_2 and S_3 as initial specification are reported in Table 10[11].

Strong evidence for a two group solution was found for intra-LATA demand using the S_2 specification suggested by Retina, while for local calls we adopted both a two cluster and a three cluster solution using the S_1 BLM specification. For inter-LATA demand there is apparently weaker evidence of heterogeneity although finally a bootstrap likelihood ratio test was finally used to choose a two group structure using the S_2 specification proposed by Retina.

The estimated models are reported in Tables 11, 12, 13 and 14 in the Appendix. Interestingly the results show that most differences among clusters can be captured by differences in constants. For example, while in the intra-LATA or inter-LATA BLM's the constant term was not statistically significant (see Table 9), homogeneous clusters found by using mixture modeling show significant variations across the

[11] See also Appendix B for more details about the justification of using AIC and BIC as selection criterion for mixture models.

constant terms of two groups (see Table 13 for intra-LATA and Table 14 for inter-LATA demand). Basically, this means that firms belonging to clusters with higher constants may be "heavy users", while components with lower constants may be "light users" of the service. Other differences among groups are associated with component variances and slope parameter estimates. Interestingly we find a close relationship between these results and the descriptive statistics shown in Table 1.

For example, consider the coefficient of $\ln(EMT/EMH)$ for inter-LATA demand (Table 14). This parameter gives an indication of the effect of the relative size of the local subsidiary with respect to the whole business. It gives an approximation of the dimension of the firm's internal communication needs. As we can see from Table 14 inter-LATA "heavy users" (cluster 2) are not sensitive to the $\ln(EMT/EMH)$ ratio since the corresponding coefficient is not significantly different from zero. This reflects the fact that "heavy users" of inter-LATA service are mostly stage I firms, which are smaller and single location firms. In fact, since the proportion of single location firms is higher in this cluster, EMT tends to EMH and this causes the $\ln(EMT/EMH)$ ratio to tend towards zero.

More evidence of heterogeneity is reported in Table 12. Here, local services demand is decomposed into three components: cluster 1 with a constant term of 4.112 (virtually equal to the whole sample estimate), cluster 2 with a constant term of 2.497 and cluster 3 with a constant term of 3.125. For the sake of convenience we will call cluster 1 "heavy users" cluster 2 "light users", and cluster 3 "medium users". We observe that estimated demand elasticities of single-line accesses such as business (BUS) and hunting (HUN) lines have positive signs as expected and are significant. Nonetheless, for network systems such as PBX trunks (PBX) and Centrex lines (CTX), the signs of the elasticities vary across clusters: PBX trunk elasticities are negative (-1.259) for "light users", and Centrex line elasticities are also negative (-1.229) for "medium users" - with very high $t - values$.

A final comment is due for Sales, which is the variable that proxies the firm output. Heterogeneity of demand with respect to sales (SAL) is evident from Figure 2, where the upward sloping cloud may suggest a positive relationship between local demand and the firm's sales. Nonetheless the estimated parameter has negative signs across clusters (Table 12). This suggests that the heterogeneity attributed to sales has no correlation with heterogeneity due to different access equipment in the firm, which in turn is represented by four variables (BUS, HUN, PBX, CTX) and accounts for a greater proportion of explained variance.

5.3 Model Selection using RETINA

5.3.1 Local Demand URM

Summary statistics for a set of local demand URM's are reported in Table 5. The final selected model is URM_6 which has been chosen among six possible URM's suggested by Retina by varying the inputs as detailed in Table 15.

We start by defining a new specification, say URM_1, and adding to the BLM the heterogeneity parameters of the optimal three-cluster solution. In Table 5 we see that URM_1 slightly improves predictive ability with respect to the BLM and $\bar{R}^2$ increases from .682 (BLM) to .708 (URM_1).

However substantial improvement in prediction is achieved with the use of W transformations generated by RETINA. This is the case of URM_2, which includes W transforms of worker per capita log-ratios. With 27 parameters URM_2 has an $\bar{R}^2 = .883$, thus explaining an increased variance of about 20% with respect to the BLM and about 18% with respect to URM_1. Out of sample predictive ability, measured by the Robust Cross Mean Square Prediction Error (RCMSPE)[12] increases substantially (about 60% of the BLM) as do the information statistics (AIC, BIC).

Perhaps the most interesting results are obtained for URM_3 and URM_4 in which we exclude all mixture heterogeneity parameters and just use per capita log-ratios together with W transforms of the logs of the original variables. Both models slightly outperform URM_2, in terms of predictive ability without using mixture heterogeneity parameters. URM_3 is a very appealing specification suggested by RETINA because it has just 20 parameters, almost as many as the number of parameters of the BLM (19), while URM_4 has 27 parameters and shows a modest forecasting improvement with respect to URM_3. We can say more about URM_3 by looking at its specification:

$$\begin{aligned}
\ln\left(\widehat{\frac{LOCAL}{EMH}}\right) = {} & \underset{(102.99)}{2.630} + \underset{(79.07)}{2.722}\ln(EMH) + \underset{(82.60)}{1.098}\ln\left(\frac{BUS}{EMH}\right) + \underset{(52.15)}{.694}\ln\left(\frac{HUN}{EMH}\right) \\
& + \underset{(89.10)}{1.219}\ln\left(\frac{PBX}{EMH}\right) + \underset{(101.88)}{.774}\ln\left(\frac{CTX}{EMH}\right) + \underset{(6.84)}{.184}\ln\left(\frac{EMH}{SAL}\right) \\
& - \underset{(-25.99)}{.194}\ln(BUS)\ln(HUN) - \underset{(-6.34)}{.214}\ln(BUS)\ln(PBX) \\
& - \underset{(-36.98)}{.204}\ln(HUN)\ln(PBX) - \underset{(-35.95)}{.143}\ln(HUN)\ln(CTX) \\
& + \underset{(8.19)}{.012}\ln(EMH)\ln(EMT) + \underset{(12.70)}{.179}\,BSOUTH + \underset{(14.58)}{.262}\,AL + \underset{(6.34)}{.107}\,GA \\
& + \underset{(6.97)}{.155}\,KY + \underset{(15.97)}{.371}\,LA + \underset{(19.78)}{.445}\,MS + \underset{(11.09)}{.262}\,SC + \underset{(1.49)}{.249}\,TN
\end{aligned} \tag{5}$$

$$n = 4391 \quad \bar{R}^2 = .891 \quad \hat{\sigma} = .346 \quad \text{RCMSPE (1000)} = .349$$
$$\sum \hat{\epsilon}^2 = 522.485 \quad \text{AIC} = -9305 \quad \text{BIC} = -9171$$
$$(t - statistics \text{ are reported in parentheses})$$

Note that RETINA suggests that interaction effects are not negligible for the final specification. Selected W transformations mainly involve interactions between different types of lines: $\ln(BUS)\ln(HUN)$, $\ln(BUS)\ln(PBX)$, $\ln(HUN)\ln(PBX)$ and $\ln(HUN)\ln(CTX)$. All of them have negative signs indicating a negative impact on demand.

[12] See [8] for details on RCMSPE.

Table 5: Local Demand: Comparison of selected statistics of candidate URM models with respect to the BLM†.

Specification	**BLM**	URM_1	URM_2	URM_3	URM_4	URM_5	URM_6
No. of Parameters	19	21	27	20	26	37	41
No. of Clusters	1	3	3	1	1	2	3
RETINA Selection	No	Yes	Yes	Yes	Yes	Yes	Yes
W Transforms	No	No	Yes	Yes	Yes	Yes	Yes
Specific Constants	No	Yes	Yes	No	No	Yes	Yes
Specific Slopes	No	No	No	No	No	Yes	Yes
$\bar{R}^2$	.682	.708	.883	.891	.896	.912	.930
$\hat{\sigma}$	.592	.567	.359	.346	.339	.312	.278
RCMSPE(1000)[a]	.595	.571	.365	.349	.343	.317	.286
$\sum \hat{\epsilon}^2$	1530	1403	562	522	500	423	336
AIC[b]	−4590	−4965	−8973	−9305	−9484	−10199	−11207
BIC[c]	−4462	−4824	−8794	−9171	−9312	−9956	−10939

†Different URM models have been selected by RETINA using different initial specifications for X,F and H. Details are reported in Table 15.

- URM_1: Obtained starting with BLM + three specific constants corresponding to the optimal S_1, G_3 three cluster solution.
- URM_2: As in URM_1 + W transforms.
- URM_3: Here heterogeneity mixture parameters are excluded. Auxiliary log-transforms of original variables (*BUS, HUN, PBX, CTX, EMT, EMH, SQFT*) are used to generate W transforms and original log-ratios are included in F, the untransformed inputs.
- URM_4: A different specification proposed by RETINA using the same specification as in URM_3.
- URM_5: As in URM_3, but this time allowing heterogeneity parameters corresponding to the S_1, G_2 two-cluster solution.
- URM_6: As in URM_3 and including heterogeneity parameters of the S_1, G_3 optimal three-cluster solution.

a. Robust Cross Mean Square Prediction Error is an approximation of the out of sample $\hat{\sigma}^2$ using 1000 bootstrap random selection of three disjointed sub-samples. See [8] for details.

b. Here AIC is specified as $n \ln(\hat{\sigma_\epsilon}^2) + 2k$, where n is the sample size and k is the number of parameters.

c. Here BIC is specified as $n \ln(\hat{\sigma_\epsilon}^2) + \ln(n)k$, where n is the sample size and k is the number of parameters.

Ramsey's RESET test [13] was computed for URM_3 to test departure from the null hypothesis of correct model specification. With an $F(2, 4369) = 105.24$ the null hypothesis of correct specification is rejected, suggesting that there is room to improve the results. Just as URM_3, URM_4 is still not well specified, RESET F (2, 4363) = 69.74, and thus we reject the null hypothesis of correct specification. A natural way to re-specify URM_3 is to add heterogeneity parameters suggested by finite mixtures. Both URM_5 and URM_6 incorporate two-cluster and three-cluster mixture parameters, respectively. Estimates of URM_6 are shown in Table 16. Inclusion of heterogeneity parameters improves prediction ability at the expense of having a larger number of parameters (37 and 41 for URM_5 and URM_6, respectively). But this gain in prediction ability is larger than the loss in precision of the estimates since AIC and BIC statistics both show evidence in favor of URM_5 and URM_6 over previous models. URM_6 has an $\bar{R}^2$ of .930 and RCMSPE which is about half that of the BLM. Both models include line-equipment interactions as in URM_3 (see Table 16), but additional demand variation is modeled by cluster-specific slopes, namely regressors that are selected by Retina from the $[H_g \times W]$ term of equation (4). Going back to Table 5, the out of sample prediction ability (RCMSPE) is only 48% of the BLM for URM_6 (.286/.595) and 53% for URM_5 (.317/.595) while it is 59% for URM_3 (.349/.595).

Almost all of the variables already used in the earlier specification of the BLM are included in the URM. These are: Business Lines (BUS), Hunting Lines (HUN), PBX trunks (PBX) and Centrex lines (CTX). RESET test for URM_6 gave $F(2, 4349) = 1.10$ which does not reject the null hypothesis of correct specification.

In URM_6, the specification suggested by Retina includes untransformed variables as well as interactions and cross-ratios between them. Equipment variables (such as type and number of lines) have non-linear effects on demand. Non-linearities may arise due to a variety of reasons including the unavailability of other relevant variables such as the nature of the business activity or whether usage is primarily internal or external. In order to capture the above mentioned non-linearities, the proposed URM_6 for local services includes a variety of transformations that go beyond the a priori specification of the BLM.

Final WLS estimations that incorporate heteroskedasticity correction of URM_6 are shown in Table 17. $F - tests$ for variable exclusion were also carried out, since some of the initial 41 variables were no longer significant, finally reducing the number of parameters of URM_6 from 41 to 37.

5.3.2 Intra-LATA and inter-LATA URM

The intra-LATA and inter-LATA demand results are quite different. As seen from Table 9, both BLM's show relatively poor fits and high standard errors of the estimation over the whole data set. The estimations suggest that both demands are sensitive to the number of single-line accesses in the business. Moreover we observe that the constant term in both BLM's is not significant. The negative sign of the (EMT/EMH) coefficients is due to the fact that both medium and long distance services are mostly demanded by single-location and small sized firms. Since these results are somewhat unsatisfactory from the prediction point of view, we apply here a selection strategy similar to the one used for local demand. For the sake of brevity, for intra-LATA and inter-LATA demand we report just the final selected URM models.

5.3.2.1 Intra-LATA URM

The final selected Useful Representative Model (URM) for intra-LATA minutes is reported in equation 6: Retina selects a very simple formulation as URM of intra-LATA demand. This model has only 5 parameters and, with the inclusion of just one specific constant for cluster 1 (H_1), we take into account heterogeneity in the data set. Recall Cluster 1 may be defined as "light users" which explains the negative sign of the specific H_1 constant. Moreover the Bell South effect is negative, reflecting the fact that intra-LATA services tend to be provided by alternative companies.

$$\ln\left(\frac{\widehat{INTRA}}{EMH}\right) = \underset{(51.90)}{3.015} + \underset{(23.88)}{.662}\ \ln\left(\frac{BUS}{EMH}\right) - \underset{(-45.75)}{2.637}\, H_1 - \underset{(-3.36)}{.203}\, BSOUTH - \underset{(-14.80)}{1.205}\, LA \tag{6}$$

$$\text{Weighted Statistics}$$
$$n = 1261 \quad \bar{R}^2 = .701 \quad \hat{\sigma} = 1.766$$

$$\text{Non Weighted Statistics}$$
$$n = 1261 \quad \bar{R}^2 = .711 \quad \hat{\sigma} = .950 \quad \text{RCMSPE (1000)} = .955$$
$$\sum \hat{\epsilon}^2 = 1134.094 \quad \text{AIC} = -121.755 \quad \text{BIC} = -90.918$$
$$(t - statistics \text{ in parentheses})$$

But perhaps the most interesting characteristic of the inter-LATA demand model concerns the $\ln(BUS/EMH)$ ratio, which represents the effect of basic single line access demand. In other words, intra-LATA demand is found to be especially sensitive to the number of business lines, while the effect of the other variables negligible.

The model passes the RESET specification test; with F(2,1254) = .958 we do not reject the null hypothesis of correct specification. Then we applied weighted OLS to correct for heteroskedisticity. The $\bar{R}^2$ of the intra-LATA URM increases from .191 to .711 (.701 for weighted estimation), while the RCMSPE is about a fifth of the BLM corresponding value. Also the standard error of estimate is about 60% with respect to the corresponding BLM value. This model shows very appealing features because its specification includes only five variables in modeling the demand of intra-LATA calls. With respect to the corresponding BLM, we gained in terms of predictive ability and also in terms of a more parsimonious representation.

5.3.2.2 Inter-LATA URM

For inter-LATA demand, we also obtain a quite parsimonious representation with just 9 parameters, after considering a set of potential URM candidates suggested by Retina. The selected URM for inter-LATA minutes is:

$$
\begin{aligned}
\ln\left(\frac{\widehat{INTER}}{EMH}\right) &= \underset{(44.27)}{3.858} + \underset{(7.49)}{.481}\ln\left(\frac{BUS}{EMH}\right) + \underset{(3.85)}{.234}\ln\left(\frac{HUN}{EMH}\right) - \underset{(-24.03)}{2.051}\,H_1 \\
&+ \underset{(9.20)}{.626}\ln(HUN)^2 + \underset{(2.19)}{.001}\ln(POP)^2 - \underset{(-8.01)}{.397}\ln(BUS)\ln(HUN) \\
&- \underset{(-3.29)}{.787}\,H_1\frac{\ln(SAL)}{\ln(POP)} + \underset{(5.05)}{.840}H_1\ \ln\left(\frac{EMT}{EMH}\right)
\end{aligned}
\tag{7}
$$

Weighted Statistics

$n = 1176 \quad \bar{R}^2 = .733 \quad \hat{\sigma} = 1.774$

Non Weighted Statistics

$n = 1176 \qquad \bar{R}^2 = .730 \quad \hat{\sigma} = .818 \quad \text{RCMSPE } (1000) = .827$

$\sum \hat{\epsilon}^2 = 780 \quad \text{AIC} = -463 \quad \text{BIC} = -412$

($t - statistics$ in parentheses)

The model has been estimated by WLS for heteroskedasticity correction. Here, significant effects are provided by the number of lines per capita, namely the number of business (BUS) and hunting (HUN) lines. Also their interaction is relevant, as well as the square of $\ln(HUN)$. Again, these interactions have negative signs. $\bar{R}^2$ is .730 versus .243 of the corresponding BLM, and RCMSPE (.827) is only 59% of the corresponding BLM (1.392) value. The model suggested by Retina is a very significant improvement over the BLM.

5.3.3 Elasticities

We are interested in evaluating the leading elasticities both for local and inter-LATA final URM models. In the case of the intra-LATA URM, since the demand specification is very simple, we do not need to make further calculations to evaluate the elasticities because the corresponding coefficients may be interpreted directly. Elasticity of intra-LATA demand with respect to the number of business lines is .66 (see eq. 6). On the other hand, evaluation of the local and inter-LATA elasticities is more tedious because the respective URM's often embed nonlinear transformations of the inputs.

As a consequence, expressions for the elasticities of the local and the inter-LATA URM also embed heterogeneity parameters and other non linearities represented by further transformations of the inputs, as shown in Table 18 and Table 19. Note that the reported expressions in most cases depend on the values assumed by other variables. We evaluate the elasticities at the average values of the influencing variables. The results are shown in Table 6.

Table 6: Selected Elasticities based on URM estimates†.

	BUS	*HUN*	*PBX*	CTX	SAL	EMT	SQFT	POP
Local ($)	1.02	.21	.87	.74	-.04	.02	.00	.00
intra-LATA (min.)	.66	-	-	-	-	-	-	-
inter-LATA (min.)	.29	.44	-	-	.03	.36	-	.03

† See Tables 18 and 19 for elasticity expressions of local and inter-LATA demand, respectively.

For local demand, we found larger positive elasticities for telephone equipment of the firm. Elasticities with respect to the number of basic accesses, namely the number of business lines, is close to one. Elasticities with respect to network access forms, PBX trunks and Centrex lines, are .87 and .74, respectively. Demand elasticities are quite irrelevant for the other explanatory variables, including the number of workers in the firm (*EMH, EMT*), sales (*SAL*), and physical extension (*SQFT*) and population habitat size (*POP*).

Single line access forms were also positively related to demand for inter-LATA services. Elasticity is .29 for business lines and .44 for hunting lines. On the other hand, the elasticity with respect to the total number of employees is .36.

5.3.4 Discussion

For each of the three demands, we may summarize the results obtained so far as follows:

- In Table 7 we report summary statistics of the Benchmark Linear Models in comparison with the final URM models suggested by Retina. They show that modeling heterogeneity and non-linearities substantially increases the overall fit and predictive ability of the estimated models with respect to the correspondent BLM's. The $\bar{R}^2$ increases for all the proposed models which is a significant improvement in within-sample-fit. Also the RCMSPE drops to between one half and one third of the benchmark model, which is a marked improvement in the out-of-sample forecast ability.

- Retina often suggests the inclusion of access equipment variables in the demand models. Relevant first order effects for medium distance (intra-LATA) and for long distance calls (inter-LATA) are single line access forms[13], whereas local demand additionally includes network access equipment variables[14] in the final specification[15]. As expected, the signs of these effects are positive.

- The specification of the three telecommunication demands never includes the physical extension of the firm[16].

[13] $\ln(BUS/EMH)$ and $\ln(HUN/EMH)$.
[14] $\ln(PBX/EMH)$ and $\ln(CTX/EMH)$.
[15] URM_6: Table 17.
[16] $\ln(SQFT/EMH)$.

Table 7: Comparison of Benchmark Linear Models (BLM) and Useful Representative Models (URM)†.

	Local (n=4391)		Intra-LATA (n=1261)		Inter-LATA (n=1176)	
	BLM	URM_6	BLM	URM	BLM	URM
Parameters	19	41	18	5	16	10
R^2	.682	.930	.191	.711	.243	.730
Std.Err. Estimate	.592	.278	1.589	.950	1.369	.818
Robust CMSPE	.595	.286	1.619	.955	1.392	.827
AIC	-4590	-11207	1188	-122	757	-463
BIC	-4462	-10945	1285	-91	843	-412

† Here we use non-weighted models for direct comparison between BLM and URM. The overall fit of the estimated URM models improves with respect to the corresponding BLM's.

- First order effects never include the output of the firm (SAL) in the final specification. However this variable appears in second order terms.
- There are significant pairwise interactions between access equipment variables[17] for local demand and between single access systems[18] for inter-LATA demand. The signs are always negative.
- Heterogeneity parameters estimated via finite mixtures are always included in the demand functions, in the form of specific constants or slopes.
- These heterogeneity parameters also influence elasticity of the demands with respect to the relevant predictors. We observe that access form variables, namely single access lines (Business and Hunting lines) and network accesses (PBX trunks and Centrex), produce larger relative variations in demand than the remaining explanatory variables.

The above results suggest that:

1. Access equipment variables are good predictors of telecommunication demand.
2. Interactions between different telephone access equipments, are not negligible.
3. The sales account only for a small proportion of explained variance for the proposed models, since their effects are second order.
4. Heterogeneity needs to be taken into account to represent the data and evaluate leading elasticities with respect to the relevant inputs.

[17] $\ln(BUS)$, $\ln(HUN)$, $\ln(PBX)$, $\ln(CTX)$.
[18] $\ln(BUS)\ln(HUN)$.

6 Conclusions

In this paper we estimate business telecommunications demands for local, intra-LATA and inter-LATA services using US Telecommunications data. Graphical bivariate analysis and Benchmark Linear Model estimation show strong evidence of heterogeneity which must be modeled in order to achieve a useful representation of the data. We achieve this goal by first using finite mixtures of normal heteroskedastic components to partition the data into homogeneous subgroups. For local demand we fit three components, while two components were fitted both for intra-LATA and inter-LATA demand. We describe the clusters, observing significant differences between constant terms and regression coefficients of each component. We then perform an automatic model search using the RETINA algorithm to obtain a flexible model useful for out of sample prediction. RETINA generates an expanded regressor set using the firm group membership as a heterogeneity parameter to estimate specific constants and specific slopes. In addition RETINA includes interactions and nonlinear transformations of the original variables as candidate regressors. We find that telephone equipment variables are almost always selected as relevant first order effects. Moreover, the corresponding coefficients are always positive. Also heterogeneity parameters and negative interactions between different forms of access are significant and play an important role in demand prediction. As a result, the demand elasticities, evaluated for the relevant variables at the average values, show that:

- Local calls demand is most sensitive to a relative variation of the number of business lines (1.02) and network access equipment (PBX Trunks (.87) and Centrex (.74)), while a change in the remaining explanatory variables is not significantly linked to relative variations of demand.
- Intra-LATA demand was sensitive only to single line access equipment represented by the number of business lines (elasticity is .66), while the effect of most of the remaining explanatory variables was negligible.
- Inter-LATA demand elasticity is positive with respect to business lines (.29) and hunting lines (.44) but also shows a positive relationship with respect to the total number of workers of the whole business (.36).

With these results we are tempted to claim that modeling of business telecom demand for this data set is adequate for its intended primary use of out of sample forecasting. These results are very encouraging in that the proposed methodology could also work well in other contexts. It is likely that the use of finite mixtures and automatic modeling procedures such as RETINA will become more widespread due to the availability of richer data sets, new software and enhanced computing power.

References

[1] Artle and Averous. "The Telephone System as a Public Good". *Bell Journal of Economics and Management Science*, 4:89–100, 1973.

[2] M. Ben-Akiva and S. Gershenfeld. *"Analysis of Business Establishment Choice of Telephone Systems"*. Cambridge Systematics, Cambridge , MA, May 1989.

[3] K.P. Burnham and D.R. Anderson. *"Model Selection and Multi-Model Inference. A Practical Information-Theoretic Approach"*. Springer-Verlag, New York, 2nd edition, 2002.

[4] P.N. Kridel, D.J. Rappoport and L.D. Taylor. "IntraLATA Long-Distance Demand: Carrier Choice, Usage Demand and Price Elasticities". *International Communications Forecasting Conference*, June 25-27 1997.

[5] F. Leisch. "FlexMix: A general framework for finite mixture models and latent class regression in R". Technical Report 86, Adaptive Information Systems and Modelling in Economics and Management Science, 2003. SFB.

[6] A. Levy. "Semi-parametric Estimation of Demand for Intra-LATA Telecommunications". Master's thesis, Department of Economics, North Carolina State University, Raleigh, NC 27695-8110, 1998.

[7] J.B. MacQueen. "Some Methods for Classification and Analysis of Multivariate Observations". Proc. Symp. Math. Statist. And Probability, 5th, pages 281–297, Berkeley, CA, 1967. Univ. of California Press. AD 669871.

[8] M. Marinucci. *"RETINA Winpack for Real Data: A quick Guide for Automatic Model Selection"*. Universidad Complutense de Madrid, Somosaguas, Madrid, Spain, December 2004.

[9] G.J. McLachlan and D. Peel. *"Finite Mixture Models"*. Wiley Series in Probability & Statistics. John Wiley & Sons, New York, 2000.

[10] D. Peña and V. Yohai. "A Fast Procedure for Outlier Diagnostic in Large Regression Problems". *Journal of the American Statistical Association*, 94:434–445, 1999.

[11] T. Pérez-Amaral, G. Gallo, and H. White. "A Flexible Tool for Model Building: the Relevant Transformation of the Inputs Network Approach (RETINA)". *Oxford Bulletin of Economics and Statistics*, 65:821–838, 2003.

[12] T. Pérez-Amaral and M. Marinucci. "Econometric Modelling of Business Telephone Toll Demand for Individual Firms using a new model selection approach, RETINA". 13th Regional Conference of the International Telecommunications Society-Madrid, September 2002.

[13] J.B. Ramsey. "Tests for Specification Errors in Classical Linear Least Square Regression Analysis". *Journal of the Royal Statistical Society*, B(31):350–371, 1969.

[14] J. Rohlfs. "A Theory of Interdependent Demand for a Communications Service". *Bell Journal of Economics and Management Science*, 5:16–37, 1974.

[15] L.D. Taylor. *"Telecommunications Demand Modelling: Theory and Applications"*. Kluwer, 1994.

[16] L.D. Taylor. "Competitive Own- and Cross-Price Elasticities in the Intralata Toll Market: Estimates from the Bill Harvesting ® II Database". Technical report, Department of Economics, University of Arizona, Tucson AZ, September 1996.

[17] R Development Core Team. *"R: A Language and Environment for Statistical Computing"*. R Foundation for Statistical Computing, Vienna, Austria, 2004. http://www.R-project.org.

[18] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, Botstein D., and R. B Altman. "Missing Value Estimation Methods for DNA Microarrays". *Bioinformatics*, 17(6):520–525, 2001.

[19] B. Von Rabenau and K. Stahl. "Dynamic A of Public Goods: a further Analysis of the Telephone System". *Bell Journal of Economics and Management Science*, 5(2):651–669, 1974.

[20] J.H. Ward Jr. "Hierarchical Grouping to Optimize an Objective Function". *Journal of the American Statistical Association*, 58(301):236–244, 1963.

[21] H. White. A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, 48:817–838, 1980.

Table 8: Probit models for intra-LATA and inter-LATA demand ($t - statistics$ are reported in parenthesis)

Dependent Variable	intra-LATA (Yes=1)	inter-LATA (Yes=1)
Observations	4463	4463
log-Lik	−1806.962	−1343.420
restricted log-Lik	−2675.425	−2594.076
Chi-sq (dgf)	1736.927 (18)	2501.312 (16)
sig.	.0000	.0000
$Constant$	−.815 (−3.381)	−.954 (−3.565)
$\ln(BUS/EMH)$	.389 (11.439)	.376 (9.095)
$\ln(HUN/EMH)$	−.238 (−6.288)	−.153 (−3.590)
$\ln(PBX/EMH)$	.205 (5.929)	.313 (8.446)
$\ln(CTX/EMH)$	−.015 (−.578)	.065 (2.316)
$\ln(SAL/EMH)$	−.119 (−1.793)	−.255 (−16.439)
$\ln(EMT/EMH)$	−.102 (−1.987)	−.198 (−3.063)
$\ln(SQFT/EMH)$	−.020 (−.806)	−.026 (−.925)
$\ln(POP)$	−.006 (−.459)	.026 (1.870)
$STAGEI$	.774 (9.534)	1.053 (1.785)
$BSOUTH$	−.083 (−1.314)	.010 (.143)
AL	.503 (5.390)	–
GA	−.043 (−.514)	.103 (1.247)
KY	1.599 (15.670)	1.904 (18.568)
LA	.493 (4.597)	1.327 (12.299)
MS	.911 (9.186)	1.442 (14.330)
NC	1.347 (12.342)	2.240 (17.634)
SC	1.575 (15.117)	–
TN	1.004 (1.055)	1.720 (15.638)

Predicted counts for intra-LATA and inter-LATA probit models

intra-LATA				inter-LATA			
	Predicted				*Predicted*		
Actual	0	1	*Total*	*Actual*	0	1	*Total*
0	2890	292	3182	0	3111	156	3267
1	503	778	1281	1	364	832	1196
Total	3393	1070	4463	*Total*	3475	988	4463

Table 9: Benchmark Linear Models for Local, intra-LATA and inter-LATA traffic ($t - statistics$ are reported in parenthesis).

Dependent Variable	$\ln(LOCAL/EMH)$	$\ln(INTRA/EMH)$	$\ln(INTER/EMH)$
Observations	4391	1261	1176
Adjusted R^2	.682	.191	.243
Std.err.est.	.592	1.589	1.369
Robust CMSPE	.595	1.619	1.392
AIC	−4590	1188	757
BIC	−4462	1285	843
$Constant$	4.112 (45.115)	1.374 (1.297)	.941 (1.349)
$\ln(BUS/EMH)$	.290 (29.320)	.694 (3.515)	.713 (5.330)
$\ln(HUN/EMH)$	.168 (15.181)	.111 (1.228)	.509 (6.632)
$\ln(PBX/EMH)$	.157 (15.371)	–	–
$\ln(CTX/EMH)$	.112 (16.950)	–	–
$\ln(SAL/EMH)$	−.024 (−5.796)	.009 (.165)	−.343 (−5.985)
$\ln(EMT/EMH)$	.138 (9.520)	−.324 (−1.615)	−.188 (−.945)
$\ln(SQFT/EMH)$	.001 (.087)	.017 (.379)	.036 (.878)
$\ln(POP)$	.004 (.804)	.009 (.378)	.110 (5.577)
$IMILLS$	–	.256 (.420)	1.159 (3.591)
$STAGEI$	−.490 (−17.197)	−.064 (−.176)	.125 (.448)
$BSOUTH$	.221 (9.142)	−.355 (−2.956)	.051 (.544)
AL	.092 (2.884)	.859 (2.516)	–
GA	.047 (1.614)	.356 (1.726)	−.374 (−2.172)
KY	−.275 (−7.100)	.881 (1.383)	1.242 (3.171)
LA	.089 (2.199)	−.517 (−1.551)	.712 (2.130)
MS	.004 (.099)	.646 (1.417)	1.119 (3.221)
NC	−.598 (−13.489)	.763 (1.281)	2.156 (4.520)
SC	−.156 (−3.808)	.819 (1.289)	–
TN	−.101 (−2.450)	.435 (.920)	.923 (2.472)

Table 10: Model selection of BLMM (Benchmark Linear Mixture Models)†.

AIC Statistic									
	Local			*intra* − *LATA*			*inter* − *LATA*		
Groups	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
1	7871	8161	8628	4766	4774	4796	4094	4170	4139
2	4155	4960	5415	4721	4730*	4766	4067	4139*	4116
3	3403	4405	4823	4674	4734	4769	4020	4149	4110
4	3300	4231	4592	4614	4739	4749	3984	4154	4101
5	3209*	4146*	4561*	4564*	4747	4744*	3928*	4160	4092*
BIC Statistic									
	Local			*intra* − *LATA*			*inter* − *LATA*		
Groups	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
1	7999	8238	8692	4864*	4795	4842*	4180*	4190	4185*
2	4417	5145	5550	4922	4776*	4864	4244	4184*	4212
3	3799*	4648	5027	4978	4806	4918	4288	4220	4257
4	3831	4556*	4866*	5020	4837	4949	4344	4250	4299
5	3873	4614	4906	5073	4870	4996	4379	4260	4340

†In Table 10 we show up to five groups solution for each demand. Solutions were obtained using both $k - means$ starting values and 100 random starting values for each partition. Values marked with an asterisk represent the lowest values of AIC and BIC along specifications S_1, S_2 and S_3. More in detail:

- S_1 : An initial specification as in equation (1). This is adopted as a natural starting point, since it is the BLM specification.
- S_2 : An initial specification suggested by performing a variables selection on eq. (1). Here we choose a more parsimonious specification than the BLM, where selected regressors were suggested by RETINA.
- S_3 : An initial specification using a specification as in (1) but excluding all the dummy variables. This is just an additional specification allowing only continuous regressors.

Solutions proposed by both criteria do not generally coincide. BIC criterion is in general the preferred statistic since AIC has been observed to over-estimate the number of components [9]. In fact AIC tends to suggest a higher dimensional solution excluding some special cases. The lowest AIC value for local demand models corresponds to a five group solution ($\text{AIC}_{LOCAL,G5,S1} = 3209$) while $\text{BIC}_{LOCAL,G5,S1} = 3799$ suggests a three group solution using specification S_1. Note that these are the lowest values with respect to alternative group solutions, but also with respect to specifications S_2 and S_3. Nonetheless, since a two cluster solution is visually expected we also take into consideration a two group solution for subsequent steps. A similar reasoning is applied to intra-LATA demand. We find evidence for a two groups solution since the lowest BIC statistic across alternative specifications corresponds to S_2 for which [19] $\text{BIC}_{INTRA,G5,S1} = 4776$. The choice of the number of groups is more difficult in the inter-LATA case. Heterogeneity is not strongly supported as in the case of local and intra-LATA demand, since lowest information statistics provide opposite results: we find the lowest $\text{AIC}_{INTER,G5,S1} = 3928$ suggests 5 groups using specification S_1, but $\text{BIC}_{INTER,G1,S1} = 4180$ suggests evidence in favor of absence of heterogeneity in the data proposing a one-cluster solution. Nonetheless we observe that the second best solution is for the two groups S_2 specification, for which $\text{BIC}_{INTER,G2,S2} = 4184$. But the differences $\text{BIC}_{INTER,G2,S2} - \text{BIC}_{INTER,G1,S1} = 4184 - 4180 = 4$ and $\text{BIC}_{INTER,G1,S2} - \text{BIC}_{INTER,G2,S2} = 4190 - 4184 = 6$ indicate only a weak evidence in favor of the S_1 and S_2 *absence of heterogeneity* model. To verify this hypothesis at least on the S_2 specification we run a bootstrapped likelihood ratio test where null hypothesis is the one group solution and the alternative is a two group solution. Departure from the null hypothesis was significant using $n = 100$ replications at $\alpha = .0001$ level thus we finally decided on a two cluster solution.

Table 11: Local Demand: Two-cluster solution Benchmark Linear Mixture Models (BLMM)($t - statistics$ in parenthesis).

$\ln(LOCAL/EMH)$	Total Sample	cluster 1	cluster 2
Observations	4391	3287	1104
AdjustedR^2	.682	.962	.764
Std.err.est.	.592	.191	.566
$Constant$	4.112 (45.115)	2.482 (7.516)	4.742 (26.067)
$\ln(BUS/EMH)$	.290 (29.320)	1.178 (166.347)	.171 (7.734)
$\ln(HUN/EMH)$	.168 (15.181)	.249 (47.777)	.032 (1.796)
$\ln(PBX/EMH)$	.157 (15.371)	−1.165 (−123.813)	.341 (2.932)
$\ln(CTX/EMH)$	.112 (16.950)	.761 (137.977)	.219 (19.808)
$\ln(SAL/EMH)$	−.024 (−5.796)	−.017 (−11.296)	−.052 (−4.125)
$\ln(EMT/EMH)$	.138 (9.520)	.008 (1.171)	.169 (8.813)
$\ln(SQFT/EMH)$	.001 (.087)	.000 (−.044)	−.018 (−1.018)
$\ln(POP)$	.004 (.804)	.005 (2.676)	−.005 (−.598)
$STAGEI$	−.490 (−17.197)	.029 (2.512)	−.272 (−4.372)
$BSOUTH$	.221 (9.142)	.237 (26.234)	.033 (.696)
AL	.092 (2.884)	.406 (32.997)	−.143 (−2.274)
GA	.047 (1.614)	.243 (22.739)	−.345 (−5.098)
KY	−.275 (−7.100)	.347 (23.727)	−.324 (−4.040)
LA	.089 (2.199)	.316 (19.833)	.271 (3.559)
MS	.004 (.099)	.451 (29.652)	−.059 (−.753)
NC	−.598 (−13.489)	.034 (1.817)	−.800 (−1.211)
SC	−.156 (−3.808)	.415 (28.702)	−.623 (−5.004)
TN	−.101 (−2.450)	.355 (24.162)	−.132 (−1.216)

Table 12: Local Demand: Three-cluster Benchmark Linear Mixture Models (BLMM) for Local Calls Billing ($t - statistics$ in parenthesis)†.

$\ln(LOCAL/EMH)$	Total Sample	cluster 1	cluster 2	cluster 3
Observations	4391	585	2622	1184
AdjustedR^2	.682	.759	.972	.984
Std.err.est.	.592	.586	.163	.134
$Constant$	4.112 (45.115)	4.113 (15.347)	2.497 (73.969)	3.125 (74.760)
$\ln(BUS/EMH)$	.290 (29.320)	.146 (4.684)	1.248 (179.245)	1.046 (139.117)
$\ln(HUN/EMH)$	.168 (15.181)	.072 (3.051)	.203 (41.950)	.104 (2.523)
$\ln(PBX/EMH)$	.157 (15.371)	.247 (1.484)	−1.259 (−138.700)	1.101 (136.011)
$\ln(CTX/EMH)$	.112 (16.950)	.215 (15.284)	.841 (165.416)	−1.229 (−106.006)
$\ln(SAL/EMH)$	−.024 (−5.796)	−.054 (−2.96)	−.026 (−18.348)	−.043 (−22.712)
$\ln(EMT/EMH)$	.138 (9.520)	.256 (9.212)	−.002 (−.368)	−.017 (−2.661)
$\ln(SQFT/EMH)$	.001 (.087)	.013 (.493)	−.001 (−.156)	−.007 (−1.593)
$\ln(POP)$	.004 (.804)	.003 (.283)	.004 (2.204)	−.001 (−.527)
$STAGEI$	−.490 (−17.197)	−.040 (−.465)	.037 (3.387)	−.042 (−3.225)
$BSOUTH$	.221 (9.142)	−.165 (−2.349)	.158 (18.509)	.331 (29.423)
AL	.092 (2.884)	−.046 (−.502)	.425 (37.658)	.211 (13.002)
GA	.047 (1.614)	−.430 (−4.037)	.068 (6.296)	.391 (31.415)
KY	−.275 (−7.100)	−.382 (−2.898)	.218 (14.668)	.330 (21.140)
LA	.089 (2.199)	.481 (4.542)	.311 (2.804)	.220 (11.350)
MS	.004 (.099)	.175 (1.584)	.357 (24.757)	.451 (24.196)
NC	−.598 (−13.489)	−.584 (−5.266)	−.002 (−.116)	.242 (11.329)
SC	−.156 (−3.808)	−.444 (−2.617)	.392 (28.314)	.358 (18.019)
TN	−.101 (−2.450)	−.030 (−.213)	.299 (2.840)	.362 (18.870)

†These parameter estimates correspond to the optimal three-cluster solution of specification S_1. See Table 10.

Table 13: Selected Benchmark Linear Mixture Models (BLMM) for intra-LATA minutes ($t - statistics$ in parenthesis).

$\ln(INTRA/EMH)$	Total Sample	cluster 1	cluster 2
Observations	1261	472	788
AdjustedR^2	.177	.361	.349
Std.err.est.	1.603	1.030	.902
$Constant$	1.923 (32.597)	.281 (4.451)	2.849 (68.752)
$\ln(BUS/EMH)$	.738 (15.691)	.776 (15.490)	.626 (18.713)
LA	−1.037 (−5.361)	−.973 (−4.507)	−1.244 (−9.348)

Table 14: Selected Benchmark Linear Mixture Models (BLMM) for inter-LATA minutes ($t - statistics$ in parenthesis).

$\ln(INTER/EMH)$	Total Sample	cluster 1	cluster 2
Observations	1176	505	665
AdjustedR^2	.184	.389	.396
Std.err.est.	1.422	.928	.763
$Constant$	3.404 (48.208)	2.133 (29.648)	4.354 (88.110)
$\ln(HUN/EMH)$	.693 (15.891)	.743 (16.666)	.637 (2.937)
$\ln(EMT/EMH)$	.442 (2.982)	1.153 (6.073)	−.102 (−.110)

Table 15: Local Demand: Specification of X,F and H inputs of RETINA†.

	URM_1	URM_2	URM_3	URM_4	URM_5	URM_6
$\ln(BUS/EMH)$	X	W	F	F	F	F
$\ln(HUN/EMH)$	X	W	F	F	F	F
$\ln(PBX/EMH)$	X	W	F	F	F	F
$\ln(CTX/EMH)$	X	W	F	F	F	F
$\ln(SAL/EMH)$	X	W	F	F	F	F
$\ln(EMT/EMH)$	X	W	F	F	F	F
$\ln(SQFT/EMH)$	X	W	F	F	F	F
$\ln(BUS)$	–	–	W	W	W	W
$\ln(HUN)$	–	–	W	W	W	W
$\ln(PBX)$	–	–	W	W	W	W
$\ln(CTX)$	–	–	W	W	W	W
$\ln(SAL)$	–	–	W	W	W	W
$\ln(EMT)$	–	–	W	W	W	W
$\ln(EMH)$	–	–	W	W	W	W
$\ln(SQFT)$	–	–	W	W	W	W
$\ln(POP)$	X	W	W	W	W	W
$STAGEI$	F	F	F	F	F	F
$BSOUTH$	F	F	F	F	F	F
AL	F	F	F	F	F	F
GA	F	F	F	F	F	F
KY	F	F	F	F	F	F
LA	F	F	F	F	F	F
MS	F	F	F	F	F	F
NC	F	F	F	F	F	F
SC	F	F	F	F	F	F
TN	F	F	F	F	F	F
H_1	F/H	F/H	–	–	F/H	F/H
H_2	F/H	F/H	–	–	–	F/H

† Each letter of the table is referred to a specification as in model 4.

Table 16: Local Calls: URM_6 parameter estimates.

Observations	4391
AdjustedR^2	.930
Std.err.est.	.278
Robust CMSPE	.286
AIC	-11207
BIC	-10939

	Variable	coefficient	t-statistic
	$Constant$	2.644	8.783
A priori transforms	$\ln(EMH)$	2.427	61.851
	$\ln(BUS/EMH)$	1.095	76.492
	$\ln(HUN/EMH)$	.341	18.568
	$\ln(PBX/EMH)$	1.201	77.146
	$\ln(CTX/EMH)$	.767	55.685
Interaction Terms	$\ln(BUS)\ \ln(HUN)$	$-.077$	-1.642
	$\ln(BUS)\ \ln(PBX)$	$-.145$	-5.040
	$\ln(HUN)\ \ln(PBX)$	$-.101$	-15.432
	$\ln(HUN)\ \ln(CTX)$	$-.055$	-13.331
	$\ln(EMH)/\ln(SAL)$	.119	4.100
	$[\ln(SAL)\ \ln(SQFT)]^{-1}$	4.511	9.162
Specific constants	$STAGEI$	.028	2.031
	$BSOUTH$	.163	14.159
	AL	.307	2.694
	GA	.099	7.064
	KY	.180	9.861
	LA	.357	18.921
	MS	.389	21.035
	SC	.317	16.586
	TN	.265	13.747
Specific slopes of Cluster 1	$H_1\ \ln(EMH)^2$	.058	19.144
	$H_1\ \ln(BUS)\ \ln(EMH)$	$-.172$	-22.787
	$H_1\ \ln(HUN)\ \ln(EMH)$	$-.027$	-4.916
	$H_1\ \ln(PBX)\ \ln(EMH)$	$-.139$	-25.710
	$H_1\ \ln(CTX)\ \ln(EMH)$	$-.064$	-16.913
	$H_1\ \ln(HUN)\ \ln(PBX)$	.082	8.469
	$H_1\ [\ln(SAL)\ \ln(SQFT)]^{-1}$	-3.185	-4.723
	$H_1\ \ln(BUS)/\ln(POP)$	2.195	6.373
	$H_1\ \ln(HUN)/\ln(POP)$	.450	1.986
	$H_1\ \ln(EMH)/\ln(POP)$	-1.381	-7.654
	$H_1\ \ln(EMT)/\ln(POP)$	1.630	14.357
Specific slopes of Cluster 2	$H_2\ \ln(BUS)^2$	.022	4.174
	$H_2\ \ln(SAL)^{-2}$	-1.151	-8.161
	$H_2\ \ln(SQFT)^{-2}$	-9.551	-8.977
	$H_2\ \ln(CTX)\ \ln(EMH)$	.015	5.185
	$H_2\ \ln(EMH)\ \ln(POP)$	$-.003$	-4.505
	$H_2\ \ln(BUS)/\ln(SAL)$	.161	2.756
	$H_2\ \ln(HUN)/\ln(SQFT)$	.506	3.472
	$H_2\ \ln(EMH)/\ln(POP)$	$-.446$	-6.724

Table 17: WLS estimation of URM_6 with heteroskedasticity correction†.

Observations	4391		
AdjustedR^2	.973		
Std.err.est.	.278		
	Variable	coefficient	t-statistic
	$Constant$	2.534	166.664
A priori transforms	$\ln(EMH)$	2.592	11.937
	$\ln(BUS/EMH)$	1.150	155.441
	$\ln(HUN/EMH)$	.370	29.58
	$\ln(PBX/EMH)$	1.239	119.72
	$\ln(CTX/EMH)$	.804	6.735
Specific constants	$STAGEI$	.178	26.892
	AL	.380	38.466
	GA	.119	13.372
	KY	.233	21.59
	LA	.309	23.085
	MS	.405	35.803
	SC	.377	46.607
	TN	.307	33.193
Interaction Terms	$\ln(BUS)\ \ln(HUN)$	$-.100$	-19.099
	$\ln(BUS)\ \ln(PBX)$	$-.200$	-6.318
	$\ln(HUN)\ \ln(PBX)$	$-.111$	-21.988
	$\ln(HUN)\ \ln(CTX)$	$-.054$	-15.52
	$\ln(EMH)/\ln(SAL)$	.057	3.346
	$[\ln(SAL)\ \ln(SQFT)]^{-1}$	5.336	2.662
Specific slopes of Cluster 1	$H_1\ \ln(EMH)^2$	.058	12.556
	$H_1\ \ln(CTX)\ \ln(EMH)$	$-.076$	-15.211
	$H_1\ \ln(HUN)\ \ln(EMH)$	$-.027$	-5.391
	$H_1\ \ln(HUN)\ \ln(PBX)$	.100	7.721
	$H_1\ \ln(PBX)\ \ln(EMH)$	$-.151$	-19.665
	$H_1\ [\ln(SAL)\ \ln(SQFT)]^{-1}$	$-.149$	-13.382
	$H_1\ \ln(EMH)/\ln(POP)$	-1.112	-3.587
	$H_1\ \ln(EMT)/\ln(POP)$	1.863	9.001
	$H_1\ \ln(BUS)/\ln(POP)$	1.625	4.180
	$H_1\ \ln(SAL)/\ln(POP)$	$-.531$	-3.153
Specific slopes of Cluster 2	$H_2\ \ln(SAL)^{-2}$	-1.320	-16.472
	$H_2\ \ln(SQFT)^{-2}$	-9.782	-17.047
	$H_2\ \ln(EMH)\ \ln(CTX)$	.008	2.835
	$H_2\ \ln(EMH)\ \ln(POP)$	$-.002$	-4.989
	$H_2\ \ln(EMH)/\ln(POP)$	$-.389$	-1.103
	$H_2\ \ln(BUS)/\ln(SAL)$	.332	9.438
	$H_2\ \ln(HUN)/\ln(SQFT)$	.369	4.034

† Several transformations have been dropped since they were no longer significant after WLS estimations.

Table 18: Selected Elasticities from local calls URM$_6$ weighted model (Table 17)

$$\begin{aligned}
\frac{\partial \ln(LOCAL)}{\partial \ln(BUS)} &= 1.150 - .149\,H_1 \ln(EMH) - .100 \ln(HUN) - .200 \ln(PBX) + \\
&+ \frac{1.626\,H_1}{\ln(POP)} + \frac{.332\,H_2}{\ln(SAL)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln(LOCAL)}{\partial \ln(HUN)} &= .029 - .149\,H_1 \ln(BUS) - .076\,H_1 \ln(CTX) + .008\,H_2 \ln(CTX) + \\
&- .027\,H_1 \ln(HUN) - .151\,H_1 \ln(PBX) + .116\,H_1 \ln(EMH) + \\
&- .002\,H_2 \ln(POP) - \frac{1.112 H_1}{\ln(POP)} - \frac{.389\,H_2}{\ln(POP)} + \frac{.057}{\ln(SAL)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln(LOCAL)}{\partial \ln(PBX)} &= 1.239 - .200 \ln(BUS) - .151\,H_1 \ln(EMH) - .111 \ln(HUN) + \\
&+ .100\,H_1 \ln(HUN)
\end{aligned}$$

$$\frac{\partial \ln(LOCAL)}{\partial \ln(CTX)} = .804 - .076\,H_1 \ln(EMH) + .008\,H_2 \ln(EMH) - .054 \ln(HUN)$$

$$\begin{aligned}
\frac{\partial \ln(LOCAL)}{\partial \ln(SAL)} &= -\frac{.531\,H_1}{\ln(POP)} + \frac{2.637\,H_2}{\ln(SAL)^3} - \frac{.332\,H_2 \ln(BUS)}{\ln(SAL)^2} \\
&- \frac{.057 \ln(EMH)}{\ln(SAL)^2} - \frac{5.336}{\ln(SAL)^2 \ln(SQFT)}
\end{aligned}$$

$$\frac{\partial \ln(LOCAL)}{\partial \ln(EMT)} = \frac{1.863\,H_1}{\ln(POP)}$$

$$\begin{aligned}
\frac{\partial \ln(LOCAL)}{\partial \ln(EMH)} &= .029 - .149\,H_1 \ln(BUS) - .076\,H_1 \ln(CTX) + .008\,H_2 \ln(CTX) + \\
&+ .116\,H_1 \ln(EMH) - .027\,H_1 \ln(HUN) - .151\,H_1 \ln(PBX) - \frac{1.112 H_1}{\ln(POP)} + \\
&- \frac{.389\,H_2}{\ln(POP)} - .002\,H_2 \ln(POP) + \frac{.057}{\ln(SAL)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln(LOCAL)}{\partial \ln(POP)} &= -.002\,H_2 \ln(EMH) - \frac{1.625\,H_1 \ln(BUS)}{\ln(POP)^2} + \frac{1.112\,H_1 \ln(EMH)}{\ln(POP)^2} + \\
&\frac{.389\,H_2 \ln(EMH)}{\ln(POP)^2} - \frac{1.863\,H_1 \ln(EMT)}{\ln(POP)^2} + \frac{.531\,H_1 \ln(SAL)}{\ln(POP)^2}
\end{aligned}$$

Table 19: Selected Elasticities from inter-LATA URM weighted model (eq.7).

$$\frac{\partial \ln(INTER)}{\partial \ln(BUS)} = \ln(HUN)$$

$$\frac{\partial \ln(INTER)}{\partial \ln(HUN)} = .234 - .397 \ln(BUS) + 1.251 \ln(HUN)$$

$$\frac{\partial \ln(INTER)}{\partial \ln(EMT)} = .840\, H_1$$

$$\frac{\partial \ln(INTER)}{\partial \ln(EMH)} = .285 - .840\, H_1$$

$$\frac{\partial \ln(INTER)}{\partial \ln(POP)} = .002\ \ln(POP) + .787\, H_1\ \ln(SAL)\ \ln(POP)^{-2}$$

$$\frac{\partial \ln(INTER)}{\partial \ln(SAL)} = \frac{.787\, H_1}{\ln(POP)}$$

A Data Preprocessing

Prior to any analysis, the data have been inspected for errors and checked for consistency. From the whole sample, 23 observations have been definitively deleted because telephone usage was absent. From the remaining 13743 observations, 4463 had complete data[20].

In modeling intra-LATA and inter-LATA demand, PBX trunks (PBX) and Centrex lines (CTX) have been excluded a priori from the respective BLM's due to the lack of variation across the considered sample. Also, for the same reason, the Alabama and South Connecticut dummy indicators have been excluded from the specification of inter-LATA traffic. Furthermore, in order to avoid perfect collinearity between the State dummy indicators, Florida is taken as the reference level and is always removed from the analysis.

Finally all the equipment variables such as business lines (BUS), hunting lines (HUN), PBX trunks (PBX) and Centrex lines (CTX) have been augmented by one because these variables present zero values, and therefore *log* transformations would be undefined.

B Estimation

Estimation of the BLM is straightforward for local traffic but not for intra-LATA and inter-LATA demand since, as seen in Section 3, not all firms use public carriers for this type of service. Direct estimation of the BLM by OLS, using only the sample with nonzero demand, would be inconsistent since the mean of the error estimates could be biased by sample selection effects. In these cases we first use a probit model to explain the probability of the firm having a non-zero demand. Probit analysis provides us with a new variable called the inverse of the Mills ratio ($IMILLS$). After this step we can go ahead with the OLS estimation of the BLM considering only those firms with some toll calling activity. But this time, among the regressors, we include the inverse of the Mills ratio as an explanatory variable because it adjusts the mean of the error term which is not necessarily equal to zero. Probit estimations for intra-LATA and inter-LATA demand are provided in Table 8.

We fitted a number of Gaussian mixture models to capture additional sources of variation for each demand. We specify the dependent variable to be distributed as a mixture of normal distributions with heteroscedastic components allowing different variances for each component. Indeed, there are many different initial specifications that may be used for clustering our data via finite mixtures. Moreover, within each specification, the number of groups of the resulting partition must be assessed after estimations. Interested readers may refer to Table 10 for details.

When heteroscedastic components are specified, the likelihood function is unbounded for the component covariances, which in turn implies that a global maxi-

[20] The variables which reported missing values were (number of missing values reported in parenthesis): $SQFT$ (9270), EMH (416), EMT (2458), SAL (2735).

mizer does not exist, (see [9]). This means that great care must be taken in order to ensure that the provided estimations do not correspond to a spurious local solution on the edge of the parameter space for σ_g, $g = (1 \ldots G)$, which should be discarded. For this reason, we compared a wide range of solutions by using different strategies to select the starting parameter values. For their definition, we used both $k-means$ clustering [7] and 100 random initial partitions of the original data set. Using this strategy we fitted up to 5 groups for each initial model specification relative to each demand.

As regards the number of groups to be retained for subsequent analysis, since regularity conditions do not hold for the log-likelihood function, usual likelihood ratio tests cannot be applied here. Thus the decision on the number of partitions to retain is based on information criteria (both AIC and BIC in our case) as well as on an a priori hypothesis about a two-cluster structure especially for local and intra-LATA demand. As discussed in the foot note of Table 10, only in the case of inter-LATA demand was there a need to assess a two cluster structure using a bootstrap likelihood ratio test. Computations were carried out using the *Flexmix* [5] and the *Mixreg* packages designed for the R software [17].

For the model selection step, we used a modified version of the Retina algorithm, called Retina Winpack [8]. The Winpack allows to perform model selection and estimate a variety of econometric models including those corresponding to equations (2), (3) and (4). The Winpack version of Retina incorporates part of the original algorithm proposed by Pérez-Amaral, Gallo and White [11] . Some modifications especially designed to analyze real data sets have been introduced. More details about Retina Winpack may be found in [8] and is available at:

`http://personales.ya.com/max_mar/retina_v0b_p.exe`.

Interested readers may refer to [8] and to [11] for details. Using Retina, estimations were performed following a bottom up strategy, beginning from the simplest specification of equation (2) and then expanding the regressors set with available information on heterogeneity sources in the data, such as the geographic region or the stage of the firm. Different URM candidates were estimated for each demand by considering original variables, their possible transformations, and specific heterogeneity parameters available in the data[21] or obtained via finite mixtures. Final checks for heteroscedasticity on residuals have been performed using White's test [21]. When necessary, weighted least squares are applied to correct heteroscedasticity of the errors.

[21] These are the dummy variables available in the data set: the geographic region, the Bell South and Stage I dummy variable.