ANNALS OF ECONOMICS AND FINANCE **1**, 165–185 (2001)

An Econometric Study of the Residential Demand for Non-Listed, Non-Published, and Special Non-Published Services

Cheng Hsiao

Department of Economics, University of Southern California
Los Angeles, CA 90089-0253

J. S. Chen

Business Impact/Advisory Services Global Telecoms Taylor Nelson Sofres Plc.
16601 N. Blanco Rd. Suite 219 San Antonia, Texas 78232

Li Gan

Department of Economics, University of Texas at Austin
Austin, Texas 78712-1173

and

R. B. Williamson

Business Impact/Advisory Services Global Telecoms Taylor Nelson Sofres Plc.
16601 N. Blanco Rd. Suite 219 San Antonia, Texas 78232

We use a stratified customer survey data to study the patterns of residential telephone subscriber listings. A conditional logit model to capture the unobserved individual-choice-specific effects is suggested to provide revenue impact analysis under different price scenarios. © 2001 Peking University Press

Key Words: Telephone listing services; Disorete choice; Demand analysis.
JEL Classification Numbers: C35, C53, D12.

1. INTRODUCTION

In this study we provide an econometric analysis of the patterns of residential telephone subscriber listings using customer survey data provided by a marketing research firm. We also provide revenue impact analysis under different price scenarios.

1529-7373/2001
Copyright © 2001 by Peking University Press
All rights of reproduction in any form reserved.

The services available today to residential customers in the U.S. consist of the standard white page (WP) residential customer listing, the Non-listed listings (NL) and the Non-published listings (NP). Standard white page listings have been in existence since the earliest customer directories were published in the latter part of the 19*th* century. Indeed, there were two reasons for making the publication of a customer's listing the normal practice. First, the earliest subscribers were businesses and later, private citizens with self-interest in being accessible to the calling public. A telephone listing demonstrated both the modernity and the means to have a telephone installed in their offices or homes. Thus, the subscriber's directory listing was a statement as well as a commercially valuable tool. Second, and just as importantly, the telephone companies in the last decade of the 19*th* century already knew about positive network externalities in the provisioning and pricing of telephone access. That is, the more subscribers they could list, the more valuable network access and usage became to a potential subscriber. This period proved to be a fertile time for experiments in access and usage pricing, network databases of addressable locations (i.e., listings), and the concept of circuit switched networking that permeates voice telecommunications to this very day.

Non-published listing services came into existence by the early part of the 20*th* century to accommodate *residential* subscribers with telephones that they wished to keep private. It was then and now seldom used by business customers. The Non-listed and Non-published listing services were never advertised or promoted, because telephone companies believed the practice diminished the economic value of their subscriber lists and of their networks, even if subscribers, by their willingness to pay to be de-listed, showed that there was market for the practice. Despite the lack of publicity and the requirement that a subscriber had to specifically request the delisting, the Non-published and Non-listed services became very highly penetrated by the time of this research. Market penetrations of NP and NL could reach a combined 30% to 40% of households in major metropolitan areas, with reportedly higher local penetrations on the eastern and western coasts of the United States.

At the time of the market research survey, a local telephone company in the United States sought to determine if there was a market among residential customers for a new service called Special Non-published listings (SNP). The differences between these subscriber listing services are consequential with respect to privacy and accessibility. The standard white page directory listing has no charge, and the residential customers are given a printed listing in the local telephone company's white pages directory. The Non-listed service does not list a customer in the telephone company's printed directory, but the customer's telephone number will be provided to anyone calling a Directory Assistance operator and specifically asking for the non-listed customer's name. The Non-published listings are more exclusive in that not only is the customers name and telephone number not printed in the white page directories, but also, the Directory Assistance operator will not give out the listing to anyone calling Directory Assistance. However, a Directory Assistance supervisor can reach a Non-published customer in an emergency situation. The new concept of Special Non-published (SNP) listing that was tested in this research would be the most exclusive listing service available to residential telephone customers. Not only is a customer's name and address not printed in the published Directory and not available through Directory Assistance, but also the customer's existence is not acknowledged. Neither is the Directory Assistance supervisor is able to reach a Special Non-published customer in an emergency situation.

The data for this analysis consists of 2,824 completed surveys from a stratified random sample of residential customers. There are several unique features of this data set. First, in the states covered by this market research survey, the NL and NP listing services were available to residential customers by annual subscription, and paid for in uniform monthly fees of $0.90 and $1.10, respectively. The SNP listing was not available as an option yet. The penetration levels at those prices were approximately 25% of residential customers in a five state area who subscribed to either NL or NP already at the time of the study. Second, the sample is not a random sample. Instead, the sample consists of 707 completed surveys from customers of White Page listings, 705 from the NL listings and 1,412 from the NP listings. Third, even though the telephone survey is cross sectional in nature, the way the data were generated by matching current customer records against survey responses provides longitudinal information ($T = 2$). The longitudinal information provides the possibility of controlling for unobserved individual specific effects e.g. (Hsiao (1986, 92)), and hence provides the possibility of obtaining a more precise estimation of the coefficients of the included explanatory variables and predictions. Fourth, while the data provides information on individual behavior, the interest is in deriving the aggregate price elasticities and making aggregate forecasts. Therefore, the development of econometric models will have to take account of the goal of the study and the unique features of this data set to provide the best combination of accuracy and cost for a specific forecast situation.

In section 2 we set up the basic framework. Section 3 discusses the nature of the data and Section 4 provides the empirical specification and estimation methods. Results are in section 5. Section 6 uses the basic individual choice model to generate aggregate prediction under various price scenarios. Conclusions are in Section 7.

2. THE BASIC FRAMEWORK

When a decision-maker is confronted with a set of mutually exclusive and collectively exhaustive alternatives, we assume that the decision-maker selects the alternative with the highest utility among those available at the time a choice is made. Let $C = \{0, 1, \ldots, M\}$ be the choice set and U_{jnt} be the utility associated with the j-th alternative for the n-th individual at time t. We assume that U_{jnt} can be decomposed into two components,

$$U_{jnt} = \mu_{jnt} + \epsilon_{jnt}, \quad j = 0, 1, \ldots, M, \tag{2.1}$$

where μ_{jnt} is a nonstochastic function of personal attributes $\tilde{x}_{jnt}$ and mode specific attributes $\tilde{z}_j$ and ϵ_{jnt} represents the effects of all other factors, assumed to possess the properties of a random variable. Let the observed variable y_{jnt} be defined as

$$y_{jnt} = \begin{cases} 1 \text{ if } & U_{jnt} = \max(U_{ont}, \ldots, U_{Mnt}), \\ 0 \text{ otherwise.} & \end{cases} \tag{2.2}$$

Then, the probability that the n-th individual will choose the j-th alternative is given by

$$P(y_{jnt} = 1) = \text{Prob } (\epsilon_{lnt} < \epsilon_{jnt} + \mu_{jnt} - \mu_{\ell nt}, \text{ for all } \ell \neq j) \tag{2.3}$$

Under the assumption that ϵ_{jnt} is independently identically distributed across j and n with density function $f(\cdot)$, (2.3) can be written as

$$P(y_{jnt} = 1) = \int_{-\infty}^{\infty} \Pi_{\ell \neq j} F(\epsilon_{jnt} + \mu_{jnt} - \mu_{\ell nt}) \cdot f(\epsilon_{jnt}) d\epsilon_{jnt}, \tag{2.4}$$

where $F(\cdot)$ denotes the cumulative distribution function of $f(\cdot)$. If the density function of ϵ_{jnt} is the Type I extreme-value distribution,

$$f(\epsilon_{jnt}) = \exp\left[-\epsilon_{jnt} - \exp\left(-\epsilon_{jnt}\right)\right], \tag{2.5}$$

then the probability that the nth individual will choose the jth alternative is given by the conditional logit model

$$P(y_{jnt} = 1) = \frac{e^{\mu_{jnt}}}{\sum_{\ell=0}^{M} e^{\mu_{\ell nt}}}, \tag{2.6}$$

(e.g. McFadden (1974)). Dividing both the numerator and denominator of (2.6) by $e^{\mu_{ont}}$, we have

$$P(y_{jnt}=1)=\frac{e^{\mu_{jn_t}-\mu_{ont}}}{1+\sum_{\ell=1}^{M}e^{\mu_{\ell nt}-\mu_{ont}}}. \tag{2.7}$$

Since only differences in utility appears as relevant factors affecting the choice probability (2.7), the conditional logit model allows a substantial simplification for the specification of choice probabilities. For instance, suppose

$$\mu_{jn}=\tilde{\alpha}'\tilde{w}_n+\tilde{\beta}_j'\tilde{x}_{jn}+\tilde{\gamma}_j'\tilde{z}_j, \quad j=0,1,\ldots,M, \tag{2.8}$$

then

$$P(y_{jnt}=1)=\frac{\exp\,(\tilde{\beta}_j'\tilde{x}_{jnt}-\tilde{\beta}_0'\tilde{x}_{ont}+\tilde{\gamma}_j'\tilde{z}_j-\tilde{\gamma}_0'\tilde{z}_0)}{1+\sum_{\ell=1}^{M}\exp\,(\tilde{\beta}_j'\tilde{x}_{jnt}-\tilde{\beta}_0'\tilde{x}_{ont}+\tilde{\gamma}_j'\tilde{z}_j-\tilde{\gamma}_0'\tilde{z}_0)}. \tag{2.9}$$

That is, the factors $\tilde{w}_n$ that have a common effect on all $M+1$ choices can be eliminated from consideration. Only choice specific factors need to be included in the specification of choice probabilities.

In the event that the stochastic component ϵ_{jnt} contains an individual-choice specific component η_{jn}, possibly representing the nth individual's time invariant preference for the jth choice, we can write ϵ_{jnt} as the sum of two components,

$$\epsilon_{jnt}=\eta_{jn}+\nu_{jnt}, \tag{2.10}$$

where ν_{jnt} is assumed to be independently, identically distributed over j,n, and t. Then if η_{jn} were known, a more accurate specification of the probability of $y_{jnt}=1$ is given by

$$P(y_{jnt}=1)=\frac{e^{\mu_{jnt}+\eta_{jn}}}{\sum_{\ell=0}^{M}e^{\mu_{\ell nt}+\eta_{\ell n}}}. \tag{2.11}$$

Unfortunately, in general, η_{jn} is unknown. To approximate η_{jn}, the following specification may be used.

$$\eta_{jn}=\tilde{\alpha}_j'\tilde{w}_{jn}+l_{jn}, \tag{2.12}$$

where $\tilde{w}_{jn}$ are the observable variables representing part of the nth individual's preference for the jth alternative and l_{jn} are the unobserved residuals assumed to be independently, identically distributed over j and n. Then, the utility associated with the jth alternative by the nth individual at time t may be written as in the form,

$$U_{jnt} = \mu_{jnt} + \tilde{\alpha}'_j \tilde{w}_{jn} + \epsilon^*_{jnt}, \tag{2.13}$$

where

$$\epsilon^*_{jnt} = l_{jn} + \nu_{jnt}, \tag{2.14}$$

and the marginal probability of $y_{jnt} = 1$ given μ_{jnt} and $\tilde{\omega}_{jn}$ can be written as

$$P(y_{jnt} = 1 \mid \mu_{jnt}, \tilde{w}_{jn}) = \frac{e^{\mu_{jnt} + \tilde{\alpha}'_j \tilde{w}_{jn}}}{\sum_{\ell=0}^{M} e^{\mu_{\ell nt} + \tilde{\alpha}'_l \tilde{w}_{ln}}}. \tag{2.15}$$

However, since

$$E\epsilon^*_{jnt}\epsilon^*_{jnt'} = El^2_{jn} \neq 0, \text{ for } t \neq t', \tag{2.16}$$

then

$$E(\epsilon^*_{jnt} \mid \epsilon^*_{jn,t-1}) \neq E\epsilon^*_{jnt}. \tag{2.17}$$

In other words, the marginal probability of $y_{jnt} = 1$ given $\tilde{x}_{jnt}, \tilde{w}_{jn}, \tilde{z}_j$ will not be equal to the conditional probability of $y_{jnt} = 1$ given $\tilde{x}_{jnt}, \tilde{w}_{jn}, \tilde{z}_j$ and $y_{jn,t-1}$,

$$P(y_{jnt} \mid \tilde{x}_{jnt}, \tilde{w}_{jn}, \tilde{z}_j) \neq P(y_{jnt} \mid \tilde{x}_{jnt}, \tilde{w}_{jn}, \tilde{z}_j, y_{jn,t-1}) \tag{2.18}$$

Hence, if information on $y_{jn,t-1}$ are available, a more accurate prediction of the probability of $y_{jnt} = 1$ may be obtained by considering the model

$$P(y_{jnt} \mid \tilde{y}_{n,t-1}) = \frac{e^{\mu_{jnt} + \delta_j y_{jn,t-1} + \tilde{\alpha}'_j \tilde{w}_{jn}}}{\sum_{\ell=0}^{M} e^{\mu_{\ell nt} + \delta_\ell y_{\ell n,t-1} + \tilde{\alpha}'_\ell \tilde{w}_\ell}}, \tag{2.19}$$

where $\tilde{y}_{n,t-1}$ is the $M \times 1$ vector denoting the nth individual's status at time $t - 1$.

3. THE DATA

The data were based on 2,824 completed telephone surveys of a stratified random sample from the existing service listings of residential customers. The existing services at the time of the survey were White Page Directories

(WP) listings, Non-listed (NL) listings and Non-published (NP) listings. The observed population shares of these three services were .75, .005, and .245, respectively. The sample fractions for them were predetermined at $H_0 = 1/4, H_1 = 1/4$, and $H_2 = 1/2$, respectively and 2,824 random samples were drawn from the three groups (H_g, $g = 0, 1$ and 2). Each individual in the sample was given four choices — WP, NL, NP and Special Non-published (SNP) listings. WP listing was free of charge. There was a monthly fee for NL, NP and SNP listings. The monthly fee for NL listings was set at either one of \$.90, \$1.40 or \$1.90; for NP listings it was set at either one of \$1.10, \$2,10, or \$3.10; and for SNP listings was at either one of \$1.00, \$2.25 or \$3.50. In total, there were 27 different price scenarios for NL, NP and SNP listings. It was equally likely for a respondent to face any one of these 27 price sets and the respondent's stated service preference was recorded. In addition, customer lifestyle, economic and demographic profiles such as income, occupation, number of adults and children in the household, and customer preferences for the WP, NL, NP and SNP listings when the services were available at no charge (customer interest variable), etc. were also collected.

4. EMPIRICAL SPECIFICATION AND ESTIMATION

Tables 1, 2 and 3 provide the 2×2 contingency tables of the sample customers' current status, choices and interests in the telephone listing categories if there were no cost.[1] In view of this summary information, we provide a maintained hypothesis of the multinominal choice model for telephone listings as summarized in Table 4. In this table, the White Page Directory (WP) listing is treated as the default alternative ($\mu_{ont} = 1$).

The maximum likelihood estimator for the multinominal logit model based on random sampling and exogenous sampling is not applicable here. As stated in Section 3, the sample was generated from stratified random sampling in which the population was divided into 3 mutually exclusive and collectively exhaustive groups — White Page Directory (WP) listings, designated as the alternative 0, Non-listed (NL) listings, designated as the alternative 1, and Non-published (NP) listings, designated as the alternative 2. The total sample size was 2,824 and the fractions were $H_0 = 1/4$, $H_1 = 1/4$ and $H_2 = 1/2$.

[1]In the telephone interview, the following specific question was asked: "Now thinking about your main residential telephone line, the regular white page listing service is currently available to you at no charge. Please assume the non-listed, the non-published, and the special non-published services are also available to you at no charge and you had to select one of these four services for your telephone lines, which one would you choose? Would you choose the (a) regular white page listing, (b) non-listed, (c) non-published, or (d) special non-published?"

TABLE 1.

The Distribution of Choices by Current Status

Choices Current Status	WP	NL	NP	SNP	Total
WP	575	56	55	21	707
	(81.33%)	(7.92%)	(7.78%)	(2.97%)	(100%)
NL	135	413	123	34	705
	(19.15%)	(58.58%)	(17.45%)	(4.82%)	(100%)
NP	180	193	710	329	1,412
	(12.75%)	(13.67%)	(50.28%)	(23.30%)	(100%)
Total	890	662	888	384	2,824

TABLE 2.

The Distribution of Choices by Current Interest Category if No Price is Charged for the Services

Choices Interest	WP	NL	NP	SNP	Total
WP	645	23	20	11	699
	(92.28%)	(3.29%)	(2.86%)	(1.57%)	(100%)
NL	111	514	53	22	700
	(15.86%)	(73.43%)	(7.57%)	(3.12%)	(100%)
NP	104	101	740	73	1,018
	(10.22%)	(9.92%)	(72.69%)	(7.17%)	(100%)
SNP	30	24	75	278	407
	(7.37%)	(5.90%)	(18.43%)	(68.31%)	(100%)
Total	890	662	888	384	2,824

TABLE 3.

The Distribution of Current Status by Interest Category if No Price is Charged for the Services

Current Status Interest	WP	NL	NP	Total
WP	525 (75.11%)	67 (9.59%)	107 (15.31%)	699 (100%)
NL	80 (11.43%)	440 (62.86%)	180 (25.71%)	700 (100%)
NP	83 (8.15%)	165 (16.21%)	770 (75.64%)	1,018 (100%)
SNP	19 (4.67%)	33 (8.11%)	355 (87.22%)	407 (100%)
Total	707	705	1,412	2,824

TABLE 4.

Maintained Hypothesis of the Multinominal Choice Model for Telephone Listing

Coefficients	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
Non -listed (NL) Utility	1	0	0	NL Price	0	0	Interest in NL listing dummy	0	0
Non Published (NP) Utility	0	1	0	0	NP Price	0	0	Interest in NP listing dummy	0
Special Non-Published (SNP) Utility	0	0	1	0	0	SNP Price	0	0	Interest in SNP listing dummy

TABLE 4—*Continued*

Coefficients	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}
NL Utility	Interest in White Page (WP) listing dummy	Current State in NL listing dummy	0	0	Current State in WP listing dummy
NP Utility	Interest in WP listing dummy	0	Current State in NP listing dummy	0	Current State in WP listing dummy
SP Utility	Interest in WP listing dummy	0	0	Current State in SNP listing dummy	Current State in WP listing dummy

Let d_{gn} be the current status dummy with

$$d_{gn} = \begin{cases} 1 & \text{if the observation came from the g-th stratum,} \\ 0 & \text{otherwise,} \end{cases} \tag{4.1}$$

and y_{jn} denote the response of the nth sample customer with

$$y_{jn} = \begin{cases} 1 & \text{if the jth alternative was chosen by the nth individual,} \\ 0 & \text{otherwise,} \end{cases} \tag{4.2}$$

$j = 0, 1, 2$ and 3. Denote the set of individuals that came from the g-th group by D_g, where $D_g = \{n \mid d_{gn} = 1\}, g = 0, 1,$ and 2. For each individual n from the g-th stratum, $n \in D_g$, it was equally likely that one of the 27 sets of prices for NL, NP and SNP listing was given, and the customer's choice and attributes were observed. Let the population proportion of each group be $W_g, g = 0, 1,$ and 2, then the likelihood function of the sample is equal to

$$L = \Pi_{g=0}^{2} \Pi_{n \in Dg} \Pi_{j=0}^{3} \left[\frac{f(j, \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn} \mid d_{gn} = 1) H_g}{W_g} \right]^{y_{jn}}, \tag{4.3}$$

where $f(j, \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn} \mid d_{gn} = 1)$ denotes the joint density of $y_{jn}, \tilde{x}_{jn}$, $\tilde{z}_j, \tilde{w}_{jn}$ given $d_{gn} = 1$.

Taking the logarithmic transformation, we have

$$\begin{aligned}\log\ L\ &= \sum_{g=0}^{2}\sum_{n\in D_g}\sum_{j=0}^{3} y_{jn}\ \log\ P(j \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{gn}=1)\\ &+\sum_{g=0}^{2}\sum_{n\in D_g}\sum_{j=0}^{3} f(\tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn} \mid d_{gn}=1)\\ &+\sum_{g=0}^{2} N_g\ \log\ H_g - \sum_{g=0}^{2} N_g\ \log\ W_g.\end{aligned} \tag{4.4}$$

where N_g denotes the total number of observations for each group. Provided W_g can be specified, maximizing (4.4) with respect to the unknown parameters yields consistent estimators. However, even though the specification of W_g is possible, because it involves the integration over the gth group, the computation of the MLE can be unwieldy. Therefore, instead of maximizing (4.4), we maximize the following weighted log-likelihood function

$$\log\ L^* = \sum_{g=0}^{2} \frac{W_g}{H_g} \sum_{n\in D_g}\sum_{j=0}^{3} y_{jn}\ \log\ P(j \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{gn}=1). \tag{4.5}$$

Generalizing the Manski and Lerman (1977) consistency proof of the weighted maximum likelihood estimator for the choice based sampling, we can show that maximizing (4.5) yields a consistent estimator of $\tilde{\beta}$, with the asymptotic covariance matrix equal to

$$\text{Asy Cov }(\tilde{\beta}) = A^{-1}\Lambda A^{-1}, \tag{4.6}$$

where $A = E\frac{\partial^2 \log L}{\partial\tilde{\beta}\partial\tilde{\beta}'}$ and $\Lambda = E\frac{\partial \log L}{\partial\tilde{\beta}}\frac{\partial \log L}{\partial\tilde{\beta}'}$.

5. RESULTS

In Table 5 we report the estimation results for the model listed in Table 4 and the two variants, one without dummy variables for the current state ($\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = 0$) and one without the dummy variables indicating the interest ($\beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$). Both the WMLE and the (false) MLE based on the random or exogenous sampling assumption are provided. In implementing the WMLE for the model, we let the population proportion be $W_0 = .75$, $W_1 = .005$ and $W_2 = .245$. However, instead of using the complete sample we only use 2,692 observations. This is because the first line of Table 1 indicates that there were 132 residential customers

TABLE 5.

Estimation Results for Multinomial Choice Models

	Model I		Model II	
Variable Name	WMLE	MLE (False)	WMLE	MLE (False)
NL Constant	−2.722	−0.0945	−1.426	−0.0054
(β_1)	(−4.561)	(−0.572)	(−2.617)	(−0.0224)
NL Constant	−1.424	0.40041	−0.269	0.7498
(β_2)	(−5.435)	(3.43827)	(−0.863)	(4.134)
SNP Constant	−2.209	−3.3925	−1.086	0.3317
(β_3)	(−6.912)	(−2.951)	(−3.344)	(1.759)
NL Price	−0.2497	−0.2085	−0.177	−0.291
(β_4)	(−0.595)	(−1.8384)	(−0.482)	(−1.855)
NP Price	−0.200	−0.22418	−0.438	−0.366
(β_5)	(−1.669)	(−4.3444)	(−3.734)	(−5.146)
SNP Price	−0.185	−0.23256	−0.423	−0.567
(β_6)	(−1.347)	(−4.12319)	(−3.337)	(−6.929)
Interested			1.122	2.328
(β_7)			(3.724)	(16.558)
Interested In NP			1.922	1.736
(β_8)			(7.908)	(12.86)
Interested In SNP			3.102	2.785
(β_9)			(11.107)	(16.005)
Interested In WP			−4.560	−3.643
(β_{10})			(−8.341)	(−18.959)
Current State In NL				
(β_{11})				
Current State In NP				
(β_{12})				
Current State In NP				
(β_{13})				
Current State In WP				
(β_{14})				
Log-likelihood	−1973.4	−3573.3	−943.67	−1932.8
Number of Observations	2692	2692	2692	2692
Percent Correctly Predicided	33.061	34.25	63.856	77.526

Asymptotic t-values are in parentheses

TABLE 5—*Continued*

	Model III		Model IV	
Variable Name	WMLE	MLE (False)	WMLE	MLE (False)
NL Constant	0.466	0.46057	0.25428	0.36094
(β_1)	(1.720)	(2.19)	(0.75230)	(1.352)
NL Constant	0.418	0.43462	0.24522	0.22283
(β_2)	(2.312)	(2.552)	(1.0162)	(1.003)
SNP Constant	−0.949	−0.847	−0.38209	−0.27653
(β_3)	(−4.086)	(−3.744)	(−1.389)	(−1.04059)
NL Price	−0.287	−0.28256	−0.27902	−0.354356
(β_4)	(−1.563)	(−2.109)	(−1.26319)	(−2.110)
NP Price	−0.2497	−0.2579	−0.31823	−0.35453
(β_5)	(−3.832)	(−4.435)	(−3.837)	(−4.756)
SNP Price	−0.20	−0.24918	−0.52059	−0.58478
(β_6)	(−3.224)	(−4.215)	(−5.923)	(−7.006)
Interested			2.18068	2.10326
(β_7)			(8.72531)	(13.623)
Interested In NP			1.45734	1.61347
(β_8)			(8.725)	(11.221)
Interested In SNP			2.55053	2.56852
(β_9)			(12.84511)	(14.01208)
Interested In WP			−2.60357	−2.51009
(β_{10})			(−10.23164)	(−11.76105)
Current State In NL	1.057	1.0547	0.21773	0.24888
(β_{11})	(7.341)	(7.344)	(1.16418)	(1.3772)
Current State In NP	1.479	1.4797	1.28939	1.27169
(β_{12})	(9.782)	(9.833)	(7.16418)	(7.133)
Current State In NP	1.991	1.994	1.41836	1.41327
(β_{13})	(9.27)	(9.336)	(5.601)	(5.639)
Current State In WP	−16.22	−15.25	−14.80933	−13.90392
(β_{14})	(−0.175)	(−0.267)	(−0.2009)	(−0.2933)
Log-likelihood	−812.58	−2462.9	−550.36	−1666.8
Number of Observations	2692	2692	2692	2692
Percent Correctly Predicided	63.076	63.076	80.052	80.089

who moved from the current WP listing to one of the three subscription based listings. Because none of the sample faced a price scheme for subscription based listings which was less than the current price of $0.90 for the NL listing and $1.10 for the NP listing, their transitions to the choice of NL, NP, or SNP listings in the hypothetical market exercise could be consistent with utility maximization behavior if (1) they represented those individuals that did not know the existence of these unlisted services, or (2), the telephone respondent was not the member of the household who initially accepted the standard WP listing when applying for telephone service. In either case, although this is useful marketing information, their behavior did not reflect rational consumer response to price change under complete information. Therefore, for the analysis here, they were excluded from consideration and the actual fraction for each stratum is modified to $H_0 = \frac{575}{2{,}692}, H_1 = \frac{705}{2{,}692}$, and $H_2 = \frac{1{,}412}{2{,}692}$.

We call attention to the following results:

1. A multinominal logit model with alternative-specific constants, prices, interest in the specific type of services, and current status variables, explains surprisingly well the demand for specific types of services. These fourteen variables correctly predict 80% of the choices of the 2,692 sample residential customers. The inclusion of additional income and sociodemographic variables, although improving the closeness of fit of the sample, failed to yield statistically significant results, and many of them also have the wrong signs, probably due to severe multicollinearity. Since the purpose of this study is not to find a model that most closely fits the data, but to generate reliable forecasts for the aggregate price elasticities and aggregate proportions of demand for specific type of services under different price scenarios, it appears that the model stated in Table 4 and results reported in Table 5, under the heading of Model IV, is the best compromise between the criteria of closeness of the fit to the observed sample and the accuracy of the forecast.

2. The price coefficients all have the correct signs and the t-values are reasonable.

3. The weighted MLE and unweighted (false) MLE yield very different predictions for the proportions of the aggregate demand for different types of services. For instance, consider the Model I of Table 5. Under the price scenario of $0.90 for the Non-listed, $1.10 for the Non-published and $2.25 for the Special Non-published services, the aggregate proportion of the demand for these three types of services based on weighted MLE is .24131. However, if we base our prediction on the unweighted (false) MLE, the aggregate proportion is 0.69885.

4. Models not using the current status dummy variables (Model I and II of Table 5) can predict at most about 60% of the outcomes correctly. When current status dummies are included, the explanatory power of the

basic model is increased substantially. Model IV of Table 5 can correctly predict 80% of the outcomes.

5. In view of the stability of the price and other coefficients, the current status dummy variables can be viewed as capturing the unobserved individual-choice-specific effects rather than describing some sort of state-dependency effects (e.g. Chamberlain (1978), Heckman (1978, 81a,b) and Hsiao (1986, 1995)).

6. The results are based on sample survey responses. It is well known that the survey responses could be subject to a number of biases, such as strategic choice behavior in order to influence the policy. Also, the telephone survey by itself can be viewed as an education process. Many customers in the beginning might not know the existence or the difference between the different types of services, or might have a wrong view about their price differentials. The interview process served to inform them of the availability and the difference between the different types of services. Therefore, even if there were no biases associated with survey responses, the results reported here should be viewed as applicable to a mature market where consumers are well informed about the availability of the services.

6. REVENUE SIMULATION

The choice models estimated in Section 5 predict the probabilities with which any particular individual will take various actions. However, predictions for a specific individual are generally of little use to planners and decision makers. The real-world decisions are usually based on the forecast of some aggregate demand. To derive the aggregate proportions we need to find the linkage between the disaggregate models and the aggregate level of forecasts. Since the disaggregate model is based on the conditional probability of choosing the jth alternative given (say) $\tilde{x}^*$, then the aggregate proportion of selecting the jth alternative is given by

$$\overline{P}(j) = \int P(y_j \mid \tilde{x}^*) f(\tilde{x}^*) d\tilde{x}^*, \tag{6.1}$$

where $f(\tilde{x}^*)$ denotes the population density of $\tilde{x}^*$.

Unfortunately, $f(\tilde{x}^*)$ is unknown. If the sample were randomly drawn, $\overline{P}(j)$ can be approximated by

$$\hat{\overline{P}}(j) = \frac{1}{N} \sum_{n=1}^{N} P(y_{jn} \mid \tilde{x}_n^*). \tag{6.2}$$

If N is large, $\hat{\overline{P}}(j)$ will be close to $\overline{P}(j)$. However, our sample is a stratified random sample. Therefore, to estimate $\hat{\overline{P}}(j)$, we use the formula

$$\begin{aligned}\hat{\overline{P}}(j) &= 0.75 \ \times \tfrac{1}{N_0} \sum_{n\in D_0} P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{0n}=1) \\ &+0.005 \ \times \tfrac{1}{N_1} \sum_{n\in D_1} P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{1n}=1) \\ &+0.245 \ \times \tfrac{1}{N_2} \sum_{n\in D_2} P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{2n}=1).\end{aligned} \tag{6.3}$$

In addition to the interest in the population proportions of the specific types of services, the direct and cross-elasticities are often of interest. Suppose the value of some variable x_{jkn} for each individual is altered by some increment so that

$$\frac{dx_{jkn}}{x_{jkn}} = \frac{dx_{jkn'}}{x_{jkn'}} = \frac{dx_{jk}}{x_{jk}}, \text{ for all } n \text{ and } n', \tag{6.4}$$

where $x_{jk} = \frac{1}{N}\sum_{n=1}^{N} x_{jnk}$, then the direct and cross elasticities of the logit model can be approximated by

$$e_{jk} = \frac{\partial \log \overline{P}(j)}{\partial \log x_{jk}} \simeq \frac{\beta_{jk} x_{jk}}{\overline{P}(j)} \ \cdot \ \frac{1}{N}\sum_{n=1}^{N} P(y_{jn} \mid \tilde{x}_n^*)\left[1 - P(y_{jn} \mid \tilde{x}_n^*)\right], \tag{6.5}$$

and

$$e_{\ell k} = \frac{\partial \log \overline{P}(j)}{\partial \log x_{jk}} \simeq -\frac{\beta_{\ell k} x_{\ell k}}{\overline{P}(j)} \ \cdot \ \frac{1}{N}\sum_{n=1}^{N} P(y_{jn} \mid \tilde{x}_n^*) P(y_{\ell n} \mid \tilde{x}_n^*). \tag{6.6}$$

Since we have a stratified random sample, we approximate the direct and cross-elasticities by the formula:

$$\begin{aligned}\hat{e}_{jk} \ &= \ -\tfrac{\beta_{jk} x_{jk}}{\hat{\overline{P}}(j)} \ \cdot \ \Big\{ 0.75 \ \cdot \ \tfrac{1}{N_0} \sum_{n\in D_0} P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{0n}=1) \\ &\cdot \ [1 - P(y_{jn} \mid \tilde{w}_{jn}, \tilde{z}_j, \tilde{x}_{jn}, d_{0n}=1)] \\ &+ \ 0.005 \ \times \tfrac{1}{N_1} \sum_{n\in D_1} P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{1n}=1) \\ &\cdot \ [1 - P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{1n}=1)] \\ &+ \ 0.245 \ \times \tfrac{1}{N_2} \sum_{n\in D_2} P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{2n}=1) \\ &\cdot \ [1 - P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{2n}=1)] \Big\},\end{aligned} \tag{6.7}$$

TABLE 6.
The Take Rate for White Page (WP), Non-Listed (NL) Non-Published (NP) and Special Non-Published (SNP) Services Under Different Price Scenarios

		NL	NP	SNP	SUM
1	NL=\$0.90, NP=\$1.10, SNP=∞	4.79	17.56	0	22.34
2	NL=\$0.90, NP=\$1.10, SNP=\$1.00	3.43	13.01	7.04	23.48
3	NL=\$0.90, NP=\$1.10, SNP=\$1.10	3.46	13.11	6.90	23.46
4	NL=\$0.90, NP=\$1.10, SNP=\$2.25	3.76	14.14	5.36	23.25
5	NL=\$0.90, NP=\$1.10, SNP=\$3.50	4.04	15.10	3.87	23.02
6	NL=\$0.90, NP=\$2.10, SNP=\$2.25	4.21	12.86	5.86	22.93
7	NL=\$0.90, NP=\$2.10, SNP=\$3.50	4.55	13.79	4.30	22.65
8	NL=\$0.90, NP=\$3.10, SNP=\$3.50	5.08	12.44	4.71	22.23
9	NL=\$1.40, NP=\$2.10, SNP=\$2.25	3.89	13.05	3.94	22.88
10	NL=\$1.40, NP=\$2.10, SNP=\$3.50	4.21	14.01	4.37	22.59
11	NL=\$1.40, NP=\$3.10, SNP=\$3.50	4.70	12.67	4.79	22.16
12	NL=\$1.90, NP=\$2.10, SNP=\$2.25	3.59	13.24	6.01	22.83
13	NL=\$1.90, NP=\$2.10, SNP=\$3.50	3.88	14.22	4.43	22.53
14	NL=\$1.90, NP=\$3.10, SNP=\$3.50	4.34	12.88	4.86	22.08

$$\begin{aligned} \hat{e}_{\ell k} \; = \; & -\tfrac{\beta_{\ell k} x_{\ell k}}{\hat{\bar{P}}(j)} \cdot \Big\{ 0.75 \; \times \tfrac{1}{N_0} \sum_{n \in D_0} P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{0n} = 1) \\ & \cdot \; P(y_{\ell n} \mid \tilde{x}_{\ell n}, \tilde{z}_\ell, \tilde{w}_{\ell n}, d_{0n} = 1) \\ & + \; 0.005 \; \times \tfrac{1}{N_1} \sum_{n \in D_1} P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{1n} = 1) \\ & \cdot \; P(y_{\ell n} \mid \tilde{x}_{\ell n}, \tilde{z}_\ell, \tilde{w}_{\ell n}, d_{1n} = 1) \\ & + \; 0.245 \; \times \tfrac{1}{N_2} \sum_{n \in D_2} P(y_{jn} \mid \tilde{x}_{jn}, \tilde{z}_j, \tilde{w}_{jn}, d_{2n} = 1) \\ & \cdot \; P(y_{\ell n} \mid \tilde{x}_{\ell n}, \tilde{z}_\ell, \tilde{w}_{\ell n}, d_{2n} = 1) \Big\} . \end{aligned} \tag{6.8}$$

In Tables 6 and 7 we provide the estimated Take Rates and direct and cross-price elasticities for different services under different price scenarios. We note that first, the direct and cross price elasticities all have the correct signs. The absolute values of the direct and cross price elasticities are in general small, however, and they increase with the price of the services. For instance, the direct price elasticities for the NL service vary from -0.146 when the NL price is \$.90 to -0.309 when the price is \$1.90. The direct price elasticities for the NP service vary from -0.080 when the NP price is \$1.10 to -0.340 when the price is \$3.10. The direct price elasticities for the SNP service vary from -0.207 when the SNP price is \$1.00 to -

TABLE 7.

Aggregate Own and Cross Price Elasticities

		NL	NP	SNP	WP
1	NL = $0.90	−0.146	0.032	0	0.002
	NP = $1.10	0.164	−0.080	0	0.008
	SNP = ∞				
2	NL = $0.90	−0.146	0.023	0.019	0.001
	NP = $1.10	0.125	−0.105	0.094	0.004
	SNP = $1.00	0.083	0.076	−0.207	0.002
3	NL = $0.90	−0.146	0.023	0.020	0.001
	NP = $1.10	0.123	−0.105	0.095	0.004
	SNP = $1.10	0.089	0.082	−0.229	0.003
4	NL = $0.90	−0.144	0.025	0.022	0.001
	NP = $1.10	0.130	−0.098	0.104	0.004
	SNP = $2.25	0.147	0.132	−0.052	0.005
5	NL = $0.90	−0.144	0.026	0.027	0.001
	NP = $1.10	0.136	−0.094	0.125	0.005
	SNP = $3.50	0.187	0.166	−1.008	0.008
6	NL = $0.90	−0.141	0.028	0.024	0.001
	NP = $2.10	0.228	−0.211	0.177	0.009
	SNP = $2.25	0.142	0.157	−0.509	0.007
7	NL = $0.90	−0.141	0.030	0.029	0.001
	NP = $2.10	0.239	−0.203	0.204	0.011
	SNP = $3.50	0.202	0.134	−0.953	0.010
8	NL = $0.90	−0.139	0.034	0.032	0.002
	NP = $3.10	0.326	−0.340	0.261	0.017
	SNP = $3.50	0.215	0.182	−0.913	0.012
9	NL = $1.40	−0.223	0.040	0.034	0.002
	NP = $2.10	0.231	−0.207	0.180	0.010
	SNP = $2.25	0.158	0.143	−0.506	0.007
10	NL = $1.40	−0.223	0.041	0.042	0.002
	NP = $2.10	0.242	−0.198	0.207	0.011
	SNP = $3.50	0.201	0.176	−0.944	0.010
11	NL = $1.40	−0.220	0.048	0.045	0.003
	NP = $3.10	0.265	−0.333	0.329	0.018
	SNP = $3.50	0.214	0.185	−0.902	0.013
12	NL = $1.90	−0.309	0.050	0.043	0.002
	NP = $2.10	0.234	−0.204	0.182	0.010
	SNP = $2.25	0.159	0.145	−0.503	0.007
13	NL = $1.90	−0.308	0.051	0.053	0.003
	NP = $2.10	0.245	−0.195	0.210	0.011
	SNP = $3.50	0.200	0.178	−0.936	0.011
14	NL = $1.90	−0.303	0.060	0.055	0.003
	NP = $3.10	0.332	−0.326	0.269	0.019
	SNP = $3.50	0.214	0.187	−0.893	0.010

1.008 when the price is \$3.50. Second, the effect of price increases leads more to the substitution between the NL, NP and SNP services rather than the substitution between them and the WP service. In other words, the effect of the concern for some sort of privacy appears to dominate the price consideration. The sum of the cross price elasticities from the NL, NP and SNP to WP service is a tiny .007 when the NL price is \$0.90, the NP is \$1.10 and the SNP is \$1.00. Even at the NL price of \$1.90, NP of \$3.10 and the SNP of \$3.50, the sum is still only 0.032. Therefore, the total take rates for NL, NP and SNP services are remarkably stable. The total take rates vary between 22.08% when the price for the NL is \$1.90, the NP is \$3.10, the SNP is \$3.50; and 23.462% when the price for the NL is \$0.90, the NP is \$1.10 and the SNP is \$1.10. When the price for the NL is \$1.40, the NP is \$2.10 and the SNP is \$2.25, the total is 22.877%. Third, the take rate for the NL is considerably higher and the take rate for the NP is considerably lower than the current take rates of 0.5% and 24.5%, respectively. When the price for the NL is \$1.40, the NP is \$2.10, and the SNP is \$2.25, the take rate for the NL is 3.886% and the take rate for the NP is 13.055%. Fourth, there is a market for the SNP service. The take rate for the SNP is 5.935% at the above price scenario and varies between 4.7% when the price for the NL is \$0.90, the NP is \$3.10, the SNP is \$3.50, and 6.9% when the price for the NL is \$0.90, the NP is \$1.10 and the SNP is \$1.10.

The above analyses are all based on the assumption that the sample customers' behaviors are similar to those in the population, and customers in the sample and in the population possess the same kind of information with regard to the availability and specifics of various kinds of services. If a certain proportion of residential customers are not aware of the availability of certain kinds of services or have misconceptions about the prices of these services, then customers in the sample and in the population do not necessarily possess the same information set. The telephone interview process itself serves to provide the respondents with the specifics of different kinds of services.[2] Therefore, the customers in the sample can be viewed as possessing the complete information but not all the customers in the population can be perceived so. In other words, the projected take rates should be viewed as what may happen in a mature market or after a full scale promotional or advertising campaign so that customers are aware of the services.

In addition, the survey responses could be subject to a number of biases (e.g. Hsiao and Sun (1999)). For instance, respondents wishing to influence company policies may opt for certain choices which they may not take otherwise. Furthermore, respondents may be indicating that they no longer need the service, or again, that they may not be the same individuals in

[2]In particular, the services were explained and prices were given during the interview.

the households who initially ordered either NL or NP, and therefore see less value in these services for the prices quoted. In fact, there is evidence that such a bias does exist. When we fit the same basic model using the current status dummy as the dependent variable we obtain projected take rates of 0.047 for the NL service and 0.245 for the NP service, which are very close to the actual take rates. On the other hand, when the survey response is used as the dependent variable, the model (Table 5, Model IV) yields the aggregate take rates for NL, NP and SNP services of 1.5% to 2.5% below the actuals when we use a price scheme of $.90 for NL, $1.10 for NP and either infinity (indicating non-availability) or $1.10 (indicating identical to NP) for SNP. However, given the stability of the aggregate take rates for the NL, NP and SNP services (Table 6), it may be safe to conclude that even if estimated price elasticities may be higher than actual, the bias is probably small.

7. CONCLUSION

In this study we suggest a conditional logit model that exploits the information contained in the stratified random sample to capture the unobserved individual-choice-specific effects. The model fits remarkably well to the residential telephone customer data. The estimated coefficients for the choice probabilities for the Non-listed, Non-published and Special Non-published services all have the correct signs and reasonable t-values. Moreover, over 80% of the responses of the 2,692 residential customers can be correctly predicted using only the price, customer interest, and current status variables as explanatory variables.

In this study we also suggest a simple procedure to derive the aggregate take rates and aggregate direct and cross elasticities from the individual choice model when the population distributions of individual attributes are unknown. The estimated aggregate price elasticities are increasing functions of prices. However, they remain small when the monthly fee for the services is below $8.00. There are some substitution possibilities between different kinds of services, but the predominant substitution is between the Non-listed, Non-published and Special Non-published services. There are very small substitution possibilities between any one of these services and the white page standard listing. This is a clear indication that the implied value upon a level of privacy afforded by these services dominates the price consideration. The predicted sum of the aggregate take rates for the Non-listed, Non-published and Special Non-published services is remarkably stable. It varies between 23.48% when the NL price $0.90, the NP is $1.10, the SNP is $1.00, and 22.08% when the NL price is $1.90, NP is $3.10, and the SNP is $3.50. Because the consumers' demand for the NL, NP and SNP services are largely determined by a preference and willingness to pay for a level of privacy, there is substantial room for the telephone company to raise revenue at very little change in costs by such measures as raising prices for non-publisehd listing services and promoting or publicizing the purpose and availability of unlisted subscriber listings.

REFERENCES

Chamberlain, G., 1978, On the use of panel data, paper presented at the Social Science Research Council Conference on Life Cycle Aspects of Employment and the Labor Market, Mt. Kisco, New York.

Heckman, J. J., 1978, Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. *Annales De L'insee* No. **30-31**, 227-269.

— 1981a, Statistical models for discrete panel data. In: *Structural Analysis of Discrete Data with Econometric Applications*, ed. by Manski C. F. and D. McFadden, 114-178, Cambridge: MIT Press.

— 1981b, Heterogeneity and state dependence. In: *Studies in Labor Markets*, ed. by Rosen S., Chicago: Chicago University Press, 91-139.

Hsiao, C., 1986, *Analysis of Panel Data*. Econometric Society monographs No.**11**, New York: Cambridge University Press.

— 1995, Panel analysis for metric data. In: *Handbook of Statistical Modelling in the Social and Behavioral Sciences*, ed. by G. Arminger, C. C. Clogg, and M. E. Sobel, Plenum, 361-400.

Hsiao, C. and B. H. Sun, 1999, Modeling survey response bias—With an application to the demand for an advanced electronic device. *Journal of Econometrics* **89**, 15-39.

Manski, C. and S. Lerman, 1977, The estimation of choice probabilities from choice-based samples. *Econometrica* **45**, 1977-1988.

McFadden, D., 1974, Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Econometrics*, ed. by Zarembka P., New York: Academic Press.