Beckert, Walter (2005) Estimation of heterogeneous preferences, with an application to demand for internet services. *Review of Economics and Statistics* **87** (3): 495-502

ESTIMATION OF HETEROGENEOUS PREFERENCES, WITH AN APPLICATION TO DEMAND FOR INTERNET SERVICES

Walter Beckert*

Abstract—This paper presents a structural econometric framework for discrete and continuous consumer choices in which unobserved intrapersonal and interpersonal preference heterogeneity is modeled explicitly. It outlines a simulation-assisted estimation methodology applicable in this framework. This methodology is illustrated in an application to analyze data from the U.C. Berkeley Internet Demand Experiment.

I. Introduction

Situations in which consumers first make a discrete choice, such as a particular tariff, and then a continuous choice over service demanded are now very common for telephone, utilities, pay television, and many other services. Discrete-continuous consumer choice data are commonly available to service providers and, less often, to academics. This paper provides a structural-utility-based econometric model for the analysis of discrete and continuous consumer choices, which explicitly incorporates unobserved intra- and interpersonal preference heterogeneity. It demonstrates how this model can be estimated using a simulation-assisted estimation methodology. This econometric approach to demand estimation in the presence of unobserved preference heterogeneity is illustrated in a small-scale application to analyze demand for Internet access, using data from the U.C. Berkeley Internet Demand Experiment (INDEX).

An econometric methodology for modeling and estimating unobserved preference heterogeneity is of interest for a number of reason. Preference heterogeneity plays a central role in the management of capacity-constrained resources, such as network services. When consumers use services provided under capacity constraints, a consumer's service consumption and induced capacity utilization may impose a negative consumption externality on all contemporaneous users, degrading the effective quality of service. To manage such resources efficiently and effectively, it is therefore of interest to assess the entire distribution of service valuations and utilization among competing users. Internet access via local area networks (LANs) and via wireless networks are prime examples.

Quality of service can be assured through efficient capacity allocation. Nonlinear prices, at least in theory, enhance the efficiency of capacity allocations. The theoretical motivation for the welfare-enhancing effects of optimal nonlinear prices, and a primitive in their construction, is preference heterogeneity (Wilson, 1993). Therefore, modeling and measuring unobserved preference heterogeneity is necessary to implement optimal nonlinear prices.

Offering different qualities of service at different prices is akin to differentiated products and services. Estimating unobserved preference heterogeneity allows one to assess users' valuations of a given set of differentiated products and services, and to determine an optimal degree of product and service differentiation.

Finally, modeling unobserved preference heterogeneity reconciles the potential discordance between typical microeconometric choice rationality assumptions and revealed-preference violations of demand data. With precise measurements, such failures cannot be attributed to measurement error. Microeconometric demand analysis typically stipulates some notion of choice rationality, such as utility maximization. An econometric methodology acknowledging unobserved preference heterogeneity permits enough flexibility to reconcile the patterns in demand data with the maintained hypothesis of choice rationality.

The econometric methodology advanced in this paper builds on a random utility model for jointly endogenous discrete and continuous choices. Discrete-continuous consumer choice data available to service providers and academics are typically in the form of a nonequispaced, unbalanced panel. The econometric methodology in this paper exploits such data in order to empirically identify intrapersonal and interpersonal preference heterogeneity. Related work by Dubin and McFadden (1984) and Dubin (1985), using cross-section data on pairs of discrete electrical appliance and continuous consumption choices, calibrates interpersonal preference heterogeneity only, whereas Rust (1987, 1994) examines the dynamics of jointly endogenous sequential discrete equipment investment choices, suppressing jointly endogenous equipment utilization choices.

The paper proceeds as follows. The main part outlines an econometric methodology for the analysis of discrete-continuous choice data in the presence of unobserved preference heterogeneity. Section II describes the econometric model, and section III provides a suitable estimation methodology. Section IV presents a small-scale illustrative application of this methodology to data from INDEX. Section V concludes.

II. The Econometric Model

A. *Inter- and Intrapersonal Preference Heterogeneity*

Consider a generic quality-differentiated service. Users are presented with a menu of prices per unit time connected to the service provider, during which they are set up to

Received for publication February 14, 2002. Revision accepted for publication September 28, 2004.

* Birkbeck College, University of London.

I thank Chunrong Ai, Richard Blundell, David Brillinger, Daniel McFadden, Paul Ruud, David Sappington, Ron Smith, Kenneth Train, Pravin Varaiya, Hal Varian, the INDEX team, various seminar participants, the editor, and two referees for helpful comments and suggestions. All errors are mine. This work was funded in part by the National Science Foundation, grant ANI-9714559, and in part by the Cal@Silicon-Valley Fellowship of U.C. Berkeley.

utilize the service; and of prices per unit volume for service usage. These prices differ by quality of service, which can be thought of broadly as maximum volume per time unit. As an example, in calling-party-pays mobile telephony, users have a choice between pay-as-you-go service and monthly and annual service contracts providing various blocks of free minutes and/or text messages per month, possibly limited to the carrier, and beyond that incremental per-minute prices. Similar price structures are currently available for broadband Internet access. In landline telephony, a monthly connection charge combines with a per-minute calling rate, typically differentiated by the extent of low-priced or free calling periods.

In usage data for such a service one observes the user's discrete service quality choice $b \in B$, where B denotes the set of available service qualities; and one observes the time T the user is connected to the chosen quality level, as well as service usage in terms of volume v. Generally, the user's specific higher-level applications which the service feeds into are not observed. Applications typically differ in terms of their service quality requirements,[1] and such differences induce heterogeneity in choices—not only interpersonal preference heterogeneity between users, but also intrapersonal between a user's choices over time, as the user's higher-level applications may change.

A typical feature of such services is that the quality of service is chosen on the basis of anticipated usage, whereas the outcome of usage, in terms of its utility for higher-level applications, is ex ante uncertain; the resolution of this uncertainty, in the process of usage, may induce a discrepancy between anticipated and actual usage. This suggests a distinction between preferences giving rise to initial discrete service quality choices, referred to as ex ante service valuations, and preferences giving rise to subsequent usage choices, referred to as online service valuations. A way to succinctly parameterize it is that the (marginal) utility of service usage, v, is ex ante unknown, but becomes revealed in the usage process. This implies that actual service usage, in addition to being subject to service quality-specific prices per unit time and volume, may be subject to intrapersonal preference shifts, in response to the actual experience of consuming the service.

This distinction between ex ante and online valuations rationalizes a number of empirical regularities in choice data. During some of the time a user is connected, it may be that no usage occurs, but by staying connected the user retains the convenience or option to use the chosen service quality without reconnecting. This time, in excess of the time actually used, will be referred to as the convenience time t, where $t \leq T$. In service plans other than pay-as-you-go, it typically carries a pecuniary cost for the user. Users may be willing to pay for it, because, beyond its option value, it may be valuable in that it can be used to react to an ex ante uncertain utility of the service usage. The distinction between ex ante and online valuations thus rationalizes demand for convenience time. This demand, and its price sensitivity, can be empirically identified, as will be illustrated in section IV. In line with the application, in section IV, to demand for Internet access, a particular discrete-continuous choice instance, consisting of (b, v, t), will henceforth be referred to as a *connection.*

A choice bundle (v, t) can be priced either at the prices associated with the chosen service quality, or at prices of other service qualities in the choice menu. Often, the observed service quality ex post is not cost-minimizing for the observed bundle. This suggest, beyond interpersonal and intrapersonal heterogeneity between connections, a third relevant sort of preference heterogeneity: intrapersonal heterogeneity within connections. A user chooses a cost-minimizing bundle on the basis of ex ante utilization, but once the uncertainty about the utility of usage is resolved and the user learns his or her online service valuations, the user may make a usage decision which ex post, on the basis of the initial service quality choice, appears not cost-minimizing. The distinction between ex ante and online valuations decouples the initial discrete service quality and associated tariff choice from the subsequent continuous usage and convenience-time choices. It thus allows one to rationalize seemingly suboptimal discrete-continuous choices.

B. *Model Specification*

The User's Perspective: This section provides a functional-form specification of a service user's utility pertaining to a connection which parameterizes inter- and intrapersonal preference heterogeneity. The utility specification is *semistructural,* in the sense that utility is not defined over higher-level service applications, which are typically unobserved, but over observed service consumption attributes, which act as inputs to such applications. The service consumption attributes associated with a connection are the service quality $b \in B$, volume v, and convenience time t.

Consider a random parameter Cobb-Douglas utility model for a connection

$$U(v, t, x, b; \theta, \epsilon, \zeta_b) = e^{\epsilon_1+\theta} \ln(v) + \sigma_\theta^2 \ln(t) + e^{\epsilon_2} \ln(x) + \zeta_b$$

where $\epsilon = (\epsilon_1, \epsilon_2)'$ and x is an outside good.[2] The parameters ϵ and $\{\zeta_b, b \in B\}$ are assumed known to the user. The parameter θ influences the (marginal) utility of service usage in terms of volume v. It is random from the user's ex ante perspective, before the user makes any choices, and

[1] For example, music and video downloads require higher usage allowances than gaming and no-frills Internet for Web browsing and mail only. Mobile telephony, if used in parallel with a landline phone for social activities, may require higher usage allowance on weekends and evenings than if used as the user's single phone.

[2] The outside good is not directly observed by the analyst, but it may be possible to indirectly infer expenditure on it from the budget constraint.

parameterizes intrapersonal preference shifts within a discrete-continuous choice instance, that is, within a connection. It distinguishes the user's ex ante valuations, formed in ignorance about θ, from the user's online valuations, when θ is revealed in the process of service utilization.[3] The user is assumed to know the distribution of θ, with its variance σ_θ^2 measuring the ex ante uncertainty about θ. Ex ante uncertainty about the marginal utility of the volume υ, in this model, is governed by σ_θ^2 and induces demand for convenience time t. The marginal utility of convenience time increases with ex ante uncertainty about the marginal utility of volume. The service-quality-specific parameter ζ_b allows for the possibility that the user's valuations vary across $b \in B$.

In this model, the user makes choices for individual connections. The time line for a connection is as follows.[4] The user chooses a service quality $b \in B$ in ignorance about θ, on the basis of ex ante valuation $E_\theta[U(v, t, x, b; \theta, \epsilon, \zeta_b)]$ and unit prices $q(b)$ and $p(b)$ for volume and time of service quality b. Once a service quality b_c, say, is chosen, the user starts consuming the service, in the context of applications unobserved to the analyst. In the process of service consumption, the user learns the marginal utility of service usage v for the purpose of these applications, that is, a realization of the random variable θ is revealed to the user and induces the user's online valuations $U(v, t, x, b_c; \theta, \epsilon, \zeta)$, given the service quality choice b_c. Conditional on this realization of θ and b_c, the user chooses the volume v, the convenience time t, and the level of the outside good x. Let $v(\mathrm{p}, m, b; \theta, \epsilon)$ and $\tau\,(p, m, b; \theta, \epsilon)$ denote the functions for continuous demands, where m is total expenditure and $\mathrm{p} = ((q(b), p(b))_{b \in B})'$. For the Cobb-Douglas specification, they take the well-known form of isoelastic demands with random coefficients and depend nonlinearly on $(\epsilon', \theta)'$. Nonlinearity of stochastic continuous demand functions in random terms arising from unobserved preference heterogeneity is a well-known consequence of utility maximization (Brown and Walker, 1989; Lewbel, 2001) and has consequences for estimation which are discussed below.

The Econometrician's Perspective: The econometrician does not observe ϵ, θ and its variance σ_θ^2, and the service quality specific utility components ζ_b.[5] The random parameter vector ϵ can be interpreted as capturing utility associated with higher-level applications unobserved by the analyst. With the exception of σ_θ^2, these parameters are allowed to vary across connections for a given user, and, including σ_θ^2, they are allowed to vary across users. This introduces intrapersonal (within and across connections) and interpersonal preference heterogeneity from the analyst's perspective.

To assess statistical properties of the model and estimate it, assumptions on the joint distribution of these random parameters need to be made. Because the model will ultimately be estimated using simulation methods, in principle any economically justifiable distributional assumptions convenient for simulation can be maintained, provided they are not rejected by the data.

Convenient, though restrictive, distributional assumptions, maintained for estimation in the Internet demand application, are as follows. The parameter ϵ_1 is allowed to be correlated with θ, whereas ϵ_2 is assumed independent. The trivariate vector $(\epsilon', \theta)'$ is assumed trivariate normal, with mean 0 and a variance-covariance matrix Σ which corresponds to these correlation assumptions. The collection of parameters $\{\zeta_b, b \in B\}$ is assumed i.i.d. extreme-value with parameter 1, independent of each other and of $(\epsilon', \theta)'$; the extreme-value assumption induces the well-known multinomial logit choice probability for the discrete-choice part of the model (McFadden, 1974). Finally, independence across connections is assumed.

These distributional assumptions impose some nontrivial restrictions on the model. The justifiability of these assumptions is partly an empirical matter and can be assessed jointly with the fit of the model. The remainder of this section provides some discussion of the main potential limitations induced by these assumptions and how they could be overcome in an extended analysis.

Independence of ϵ_2 complements additive separability of x in the utility function and allows the model to be estimated conditional on observed expenditure in a connection. It amounts to assuming strict exogeneity of expenditure. Viewed through the lens of two-stage budgeting, in the context of deterministic utility, weak separability is enough for allocations within groups to be a function of within-group relative prices and group expenditure (see, for example, Blundell, 1988). In the generalization to random utility, even under the stronger assumption of additive separability, independence of random group-specific utility coefficients is necessary to preserve orthogonality conditions necessary for estimation. Independence of ϵ_2 implies, furthermore, that the discrete choice probabilities functionally depend on ϵ_1 alone. This assumption is testable, as discussed in section IV.

The assumption that ζ_b, $b \in B$, is statistically independent of ϵ_1 may lack plausibility. The reason is that, if ϵ_1 captures characteristics of applications the user intends to pursue online, then the value of any service quality is presumably correlated with these. The independence assumption is maintained for convenience. Nonetheless, discrete choice

[3] An alternative interpretation of this specification is that the utility $\bar{U}(V, t, x, b; \epsilon, \zeta_b) = e^{\epsilon_1} \ln(V) + \sigma_\theta^2 \ln(t) + e^{\epsilon_2} \ln(x) + \zeta_b$ is defined over a usage product V which is produced from volume v as $V = \mathrm{V}(v, \theta) = v^{e^\theta}$, with ex ante unknown productivity parameter $e^\theta \geq 0$. Then $U(v, t, x, b; \theta, \epsilon, \zeta_b) = \bar{U}(V(v, \theta), t, x, b; \epsilon, \zeta_b)$.

[4] Formal details are given in an appendix.

[5] As an aside, the econometrician may not observe x, p_x, and m either. This means that the model may have to be estimated conditional on observed expenditure for a connection. The given formulation of the model is useful to test for endogeneity of expenditures which, if not rejected, would bias estimation results. This test is described later, in section IV.

probabilities do depend functionally on ϵ_1. Hence, the statistical independence assumption turns out to be mild.[6]

Serial dependence, instead of the assumption of independence of connections, might arise from learning about θ or dependence in applications. It could be simulated, at additional computational cost, using Markov chain Monte Carlo techniques.[7]

III. Estimation Methodology

The model needs to be estimated from the reduced form. The nonlinearity of the reduced-form model $(v, t)' = h(p, m, b; \theta, \epsilon) = (v(p, m, b; \theta, \epsilon), \tau(p, m, b; \theta, \epsilon))'$ in $(\epsilon', \theta)'$ has two important consequences. First, the model cannot be estimated without simulation. Second, because the discrete-choice probabilities depend on ϵ_1, conditional on the discrete choice b the system h depends nonlinearly on θ and ϵ_2. The inversion of this system with respect to $(\theta, \epsilon_2)'$ is analytically intractable. Hence, evaluation of the likelihood function is not feasible.[8] Therefore, the model must be estimated from its moments. The method of choice is therefore the method of simulated moments (McFadden, 1989; Pakes and Pollard, 1989).[9]

The parameters to be estimated are the variance of ϵ_1, $\sigma^2_{\epsilon 1}$, which captures intrapersonal preference heterogeneity across connections or variation in ex ante valuations; the variance of θ, σ^2_θ, which captures intrapersonal preference heterogeneity within connections or variation in online valuations; the correlation $\rho_{\theta_c 1}$ of ϵ_1 and θ, which captures the extent to which nonpersistent preference shifts are correlated with latent applications; and random utility parameters ξ on exogenous covariates z related to a connection, capturing observed preference heterogeneity; these are included linear as $z'\xi$ and additive to ϵ_1 and appear in the reduced-form model. To simplify notation, collect the variance-covariance parameters in the matrix Σ. Denote the true parameters by ξ_0 and Σ_0.

One may start from conditional moments, given prices and expenditures, derived from $k = 2 + \text{card}(B)$ conditional moment functions

$$M(v, t, b_c, p, m; \xi, \Sigma) = \begin{bmatrix} (v, t)' - E[h(p, m, b_c; \xi, \theta, \epsilon)|b_c] \\ \{1_{\{b=b_c\}} - P(b; \xi, \Sigma);\ b \in B\} \end{bmatrix} = \begin{bmatrix} (v, t)' - E_\epsilon\left[E_{\theta|\epsilon, b_c}[h(p, m, b_c; \xi, \theta, \epsilon)]\ \frac{P(b_c|\epsilon_1; \xi, \Sigma)}{P(b_c; \xi, \Sigma)}\right] \\ \{1_{\{b=b_c\}} - E_{\epsilon_1}[P(b|\xi, \Sigma, \epsilon_1)];\ b \in B\} \end{bmatrix},$$

satisfying $E[M(v, t, b_c, p, m; \xi_0, \Sigma_0)|p, m] = 0$. These expectations are not analytically tractable, due to the inherent nonlinearity of the reduced-form expressions in ϵ and θ, but they can be simulated, drawing from the respective conditional and marginal distributions. One is not limited to conditional moments alone. Using an $r \times k$ array of suitable instruments $Z = Z(p, m; \xi, \Sigma)$, which can be any function of p, m, provided they are exogenous, and, optionally, the parameters of interest, one can form further moment functions

$$D(\xi, \Sigma) = Z(p, m; \xi, \Sigma)\, M(v, t, b_c, p, m; \xi, \Sigma),$$

which yield unconditional moments $E[D(\xi_0, \Sigma_0)] = 0$.[10] Index a user's connections by $s = 1, \ldots, S$, with observed choices $\{(v_s, t_s, b_s), s = 1, \ldots, S\}$. For each s, draw ϵ_1^* from the conditional distribution, given b_s, noting that the choice b_s in connection s contains information on the unobserved ϵ_1;[11] and draw θ^* from its conditional distribution, given ϵ_1^* and b_s. Averaging across T simulation draws for each s, form simulated counterparts to $D_s(\xi, \Sigma)$, denoted $D_s^*(\xi, \Sigma)$. For a positive definite, symmetric weighting matrix Q_S, form the quadratic

$$\psi_S(\xi, \Sigma) = E_S[D_s^*(\xi, \Sigma)]'\, Q_S E_S[D_S^*(\xi, \Sigma)],$$

where $E_S[\cdot]$ denotes a sample average. Then minimize $\psi(\xi, \Sigma)$ with respect to ξ_s and vec(Σ). This yields initial, consistent, yet inefficient estimators of ξ and Σ. To enhance the asymptotic efficiency of this MSM estimator, replace Q_S in a second estimation step by an estimate of the asymptotically optimal weighting matrix

$$Q^* = \{E[D(\xi_0, \Sigma_0)\, D(\xi_0, \Sigma_0)']\}^{-1}.$$

[6] If one wanted to relax it, from a computational point of view it would be preferable to replace the extreme-value assumption with joint normality of the $\{\zeta_b, b \in B\}$, as conditional normals are easier to simulate and, in the process of estimating a multinomial probit model (instead of the multinomial logit), automatically emerge in the simulation algorithm. See, for example, Hajivassiliou and Ruud (1994).

[7] See, for example, Chib and Greenberg (1996) and references provided therein.

[8] It is worth noting that this model is an instance where the maximum likelihood estimator exists in theory (that is, classical regularity conditions for its existence are satisfied) but is computationally infeasible.

[9] The method of simulated moments (MSM) is a variant of the generalized method of moments (GMM), replacing analytically intractable moments by simulated analogs. Like GMM, in estimation it minimizes moments in a metric defined by a positive definite weighting matrix. Moments are best thought of as orthogonality conditions between instruments and residuals, derived from an econometric model. The method is generally less efficient than GMM, as a consequence of simulation noise, unless the number of simulation draws increases sufficiently fast with sample size. GMM, in turn, is less efficient than maximum likelihood (ML) estimation, because it does not use distributional assumptions, only moment conditions. If the moments are given by the expected ML score, then the instruments are optimal by construction and GMM is fully efficient.

[10] It is a straightforward consequence of the expression for the probability law of the endogenous choice variables that the ideal instruments depend on the parameters to be estimated. Because the likelihood function cannot be evaluated [due to the analytical intractability of the inverse of h with respect to (θ, ϵ_2)], the ideal instruments, as part of the orthogonality conditions summarized in the score equations, are analytically intractable as well.

[11] An easy, but computationally expensive method to do this is the accept-reject algorithm; see McFadden and Ruud (1994).

A consistent estimator of Q* is $\hat{Q}_S = \{E_S[D_s(\bar{\xi}, \tilde{\Sigma})D_s(\tilde{\xi}, \tilde{\Sigma})']\}^{-1}$, where $\bar{\xi}$ and $\bar{\Sigma}$ are the initial, consistent estimators of ξ_0 and Σ_0 from the first step. For the first-round estimation, $Q_S = I_r$ can be chosen. Under regularity conditions,[12] this procedure yields consistent, asymptotically normal estimators. It will not be asymptotically efficient, for efficient moments are infeasible.

Intuitively, the parameter $\sigma^2_{\epsilon_1}$ is identified from variation in the discrete choice sequence which is induced by different applications a user runs. The parameters σ^2_θ and $\rho_{\theta_\epsilon 1}$ are identified through discrete-continuous choice pairs, which ex post appear suboptimal in a model that does not allow for intrapersonal preference shifts within connections. A formal identification argument for this class of models is provided in Beckert (2004).

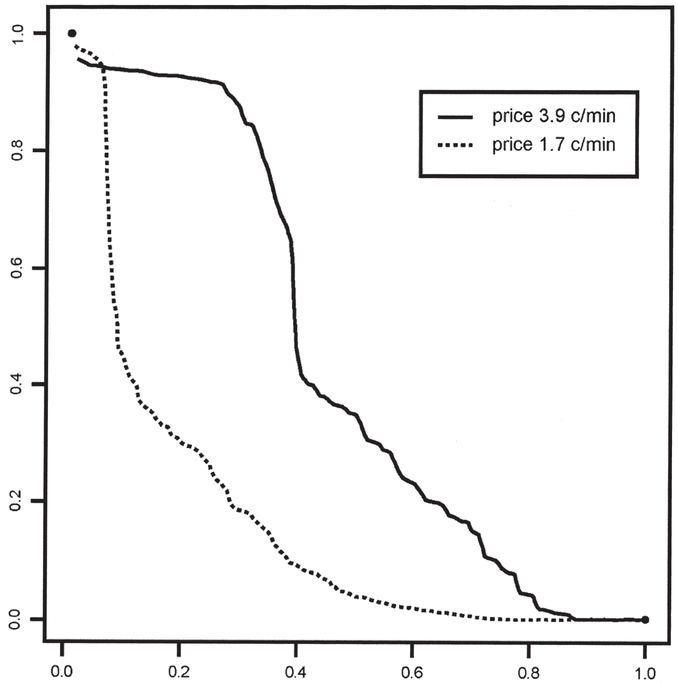
IV. Application: Demand for Internet Access

This section reports the results from an application of the model to a randomly chosen 10% subsample of approximately 70 participants total of the U.C. Berkeley Internet Demand Experiment (INDEX). The experiment is described in the appendix. Given the small scale of this investigation, the relatively select group of INDEX participants, and the lack of relevant covariates other than prices, the estimation results make no claim to be representative of the universe of Internet users, but are intended as illustration of the econometric methodology proposed in this paper.

In this application, b corresponds to the nominal transmission speed, measured in kilobits per second (Kbit/s), υ to the in- and outbound byte volume, measured in kilobits; and t to the convenience time, measured in seconds.[13] The outside good can be interpreted as all other Internet activity outside the INDEX environment.

Demand for convenience time can be readily identified in the INDEX data. Service utilization can be measured by the proportion of a minute in which the user is connected and transmits data. Utilization can be ranked, from 100% (the user is transmitting the entire minute) to 0% (the user does not transmit anything during that minute, but is connected entirely for convenience). Figure 1 plots ranked utilization against percentage of total connection time to 128 Kbit/s, for data from a user in two weeks with different prices for the 128-Kbit/s service; the upper curve plots ranked utilization when the cost was 3.9 cents per minute, the lower one when it was 1.7 cents per minute. The graph provides two important insights. First, it emerges that utilization is sensitive to per-minute prices. Second, the graphs suggests that the fraction of idle time (that is, time corresponding to zero

FIGURE 1.—CUMULATIVE UTILIZATION OF 128 KBIT/S AT TWO DIFFERENT PRICES, AGAINST FRACTION OF TIME.



utilization in which the user does not transmit any data) is price-sensitive as well. This confirms, in particular, that users value the convenience of staying connected, even if the connection is not actively used; equivalently, they value the mere option to transmit data, even when none are being transmitted. Furthermore it suggests that demand for this option responds to price.

Because the sequence $\{m_s\}$ of total expenditures is not observed, the model is estimated conditional on observed expenditures on Internet services in every connection. This has the consequence that $\sigma^2_{\epsilon_2}$ is not identifiable. Also, it is clear from the functions $\upsilon(\mathrm{p}, m, b; \theta, \epsilon)$, $\tau(\mathrm{p}, m, b; \theta, \epsilon)$, and $x(\mathrm{p}, m; \theta, \epsilon)$ that expenditures on volume υ and convenience time t in the general model specification may very well be correlated with the residuals in the stochastic demand equations. This is the case if the outside good x is not exogenous, conditional on expenditure. How to test for exogeneity is briefly discussed below. The estimable model parameters then are $\sigma^2_{\epsilon_1}$, σ^2_θ, $\sigma_{\theta_{\epsilon_1}}$, and the parameters ξ. Because at the time of this study no covariates other than the date of the observed connection and the start time are available, these are used to create proxies for whether the observed connection was work-related or not. Whether or not a connection is work-related may be reflected in a user's behavior. It may determine whether the user herself or her employer pays for the connection. It may also restrict the class of applications that make up the transmission activity of the connection. For lack of more accurate covariate data, the proxies used are two indicator variables, taking value 1 if the date of the connection corresponds to a regular working day (z_1) and a regular work hour, 7 A.M.–7 P.M. (z_2). These proxies are

[12] For consistency, these essentially amount to uniform convergence in probability of the objective function $\psi_S(\xi, \Sigma)$, compactness of the parameters space, and identification; asymptotic normality requires in addition that a central limit theorem apply to the gradient of the vector of moments, and a uniform law of large numbers to $\hat{Q}_S$.

[13] Convenience time is not always directly observed and needs to be estimated. An outline of the approach taken is available upon request.

TABLE 1.—DISTRIBUTION OF MSM ESTIMATES

Statistic	$\hat{\sigma}^2_{\epsilon_1}$	$\hat{\sigma}^2_\theta$	$\hat{\rho}_{\theta_\epsilon 1}$	$\hat{\xi}_1$	$\hat{\xi}_2$
Mean	2.51	4.01	−0.54	0.04	−0.38
Std. dev.	1.52	0.89	0.36	1.38	0.98
Min	0.48	2.42	−0.01	−1.73	−0.147
Max	4.60	4.91	−0.91	2.51	1.03

weak, for many of the INDEX subjects are UC Berkeley students and faculty members, whose schedules are likely to deviate from this notion of regularity.

MSM estimations are performed separately on the data for individual users. The number of observed connections, S, for a user then corresponds to the sample size.[14] The instruments used in estimation are the vector of prices and nonlinear transformations of these, such as their squares, their cubes, and their logarithms, leading to a total of $r = 20$ unconditional moments. By experimental design, prices were randomized and hence are valid instruments. For simulation, $T = 10$ simulation sample draws were used. Denoting the MSM estimators by $\hat{\xi}$, $\hat{\Sigma}$, approximate asymptotic standard errors are computed on the basis of the usual normal approximation, using $E_S[\nabla_{(\xi',\mathrm{vec}(\Sigma)')}D_s{}^*(\hat{\xi}, \hat{\Sigma})]$ as an estimate of $\mathrm{M}_0 = E[\nabla_{(\xi',\mathrm{vec}(\Sigma)')}D(\xi_0, \Sigma_0)]$ for corresponding expression in the asymptotic variance-covariance matrix $\frac{T+1}{T}(\mathrm{M}_0'\mathrm{Q}^{*-1}\mathrm{M}_0)^{-1}$.

Table 1 presents parameter estimates for the random subset of seven users. The estimation results point to a number of observations. Intrapersonal preference heterogeneity across connections, or variation in ex ante valuations, is estimated by $\hat{\sigma}^2_{\epsilon_1}$. This estimated measure varies considerably between users. This variation is attributable to different levels of diversity in unobserved higher-level Web applications across users.

Users also exhibit similarly strong, if not larger, estimated intrapersonal preference heterogeneity within connections, that is, variation in online valuations, estimated by $\hat{\sigma}^2_\theta$. Estimates $\hat{\sigma}^2_\theta$ dominate in size the estimates for the variation in ex ante valuations, $\hat{\sigma}^2_{\epsilon_1}$. This suggests that the consumption experience itself induces a discrepancy between ex ante and ex post valuations. The parameter σ^2_θ is identified through discrete-continuous choice combinations that appear suboptimal ex post. The fraction of ex post apparently suboptimal choices among the users in this analysis ranges between 4% and 27%. The associated pecuniary cost is on the order of 5 to 12 dollars, for monthly bills of up to 60 dollars. In light of these costs, seemingly suboptimal choices are unlikely to result from mere carelessness.

The parameter $\rho_{\theta_\epsilon 1}$ characterizes the correlation between unobserved intrapersonal preference heterogeneity within and across connections, or ex ante and online valuations. Its estimates are negative, uniformly across users. A negative (positive) correlation can be interpreted as an ex ante low (high) anticipated marginal valuation of the byte volume being revised upward (downward) online once the quality of information embodied in the data is revealed. In this sense, the estimates of $\rho_{\theta_\epsilon 1}$ suggest furthermore that users deviate in their online service valuations from their ex ante valuations.

The estimates of the coefficients on the work proxies are to be interpreted with caution, for reasons already pointed out. There does not appear to be a regular pattern applicable to all users. For some users, on the premise of the validity of the work proxy, the result suggests a tendency for work-related activity to reduce convenience time, conditional on prices and expenditure.

Various model-free diagnostic tests and out-of-sample predictions provide some reassurance regarding these estimates.[15] A common concern, however, is that joint expenditures on volume and convenience time may be correlated with the residuals from the difference between observed continuous choices and their expectation, given expenditure. This correlation amounts essentially to a selection bias in the moment conditions. Endogeneity of expenditures has received increasing attention in the applied and methodological literature dealing with analysis of demand data.[16] Exogeneity of expenditures can easily be tested. One includes a coefficient α on the component $\ln(x)$ in the utility function and tests for exogeneity by examining the null hypothesis $H_0 : \alpha = 0$. A score test is a convenient test procedure in this context, in that it obviates estimation of the alternative model. Under the null hypothesis, the score test statistic has a χ^2_1 distribution. The maximal score test statistic for the random sample of users is 3.35, which lies well below the 95% critical level for rejection, 3.85. Therefore, the null hypothesis of exogenous expenditures cannot be rejected at the 95% significance level.

V. Conclusion

This paper proposes a structural econometric framework for the analysis of discrete-continuous choice data in the presence of unobserved inter- and intrapersonal preference heterogeneity. Such data are becoming increasingly available, certainly in telecommunication services, whether wired or wireless, as well as in the

[14] For each subject, the number of observations was on the order of several hundred connections. By any standard, this is enough to justify using the asymptotic properties ascribed to the MSM estimators. In any event, network services that require establishing temporary connections and connection data for billing can reasonably be expected to deliver long enough data series for individual users to rely on asymptotic results and to render repetition of the data-generating process a plausible scenario.

[15] Details are omitted due to space limitations, but available on request.

[16] Endogeneity is usually dealt with by the instrumental variables approach, as in well-known parametric approaches; see Newey and Powell (2003), Darolles, Florens, and Renault (2002), and Hall and Horowitz (2003) for the purely nonparametric approach; Ai and Chen (2003) for the semiparametric approach; and Blundell, Chen, and Kristensen (2003) for the semi-nonparametric approach. An alternative is the control function approach, taken for example by Blundell and Smith (1994), Newey, Powell, and Vella (1999), and Das, Newey, and Vella (2003).

television, electricity, gas, and water industries; recent policy initiatives in Germany and the U.K. also envision usage-based charges for haulage vehicles on the national road network. In the telecommunication and electricity services especially, suboptimal capacity management leading to network congestion and ensuing quality-of-service degradation have been painfully felt in many highly developed economic areas. This is a concern to both corporate analysts and regulatory authorities. Both have access to such data and may find the approach taken in this paper a useful building block in the design of superior price-induced capacity allocation mechanisms and to assess the competitive implications of changes in market structures in these network industries.

APPENDIX A

Formal Sequential Choice Algorithm

The user is assumed to solve the sequential discrete-continuous choice problem according to the following algorithm. The user first forms expectations about θ and computes the expected utility function. Given expenditure m and prices for unit volume $q(b)$ and unit time $p(b)$, differentiated by service quality $b \in B$, the user maximizes the expected utility function over the budget constraint and derives the anticipated volume and time choices $\hat{c}$ and $\hat{\tau}$, for each $b \in B$. On the basis of these, the indirect utility of each $b \in B$ can be computed as the expected utility function evaluated at $\hat{v}$ and $\hat{\tau}$; due to the separability of x, the outside good is immaterial at this point. The user then chooses the service quality which is associated with the highest indirect utility, say b_c. Finally, being committed to b_c, with a realization of θ revealed in the process of service usage, the user chooses the actual levels of volume υ and convenience time t, and the outside good x.

Denote the expectation operator with respect to the random variable θ by $E_\theta[\cdot]$. Then, formally, anticipated choices $\hat{v}(\mathrm{p}, m, b; \epsilon)$, $\hat{T}(\mathrm{p}, m, b; \epsilon)$, and $\hat{x}(\mathrm{p}, m; \epsilon)$ maximize $E_\theta\,[U\,(v, t, x, b; \theta, \epsilon, \zeta_b)]$ subject to ${}_x x + p(b)t + [q(b) + p(b)/b]v = m$, for each $b \in B$; note here that the service quality b, defined as the maximum volume per unit time, implies that v/b is the required time to process volume υ, payable at price $p(b)$. The indirect utility of $b \in B$ is $V\,(b, \mathrm{p}, m; \epsilon) = E_\theta[U(\hat{v}(\mathrm{p}, m, b; \epsilon)$, $\hat{T}(\mathrm{p}, m, b; \epsilon)$, $\hat{x}(\mathrm{p}, m; \epsilon), b; \theta, \epsilon, \zeta_b)]$. The first-step discrete choice then is $b_{\mathrm{c}} = \arg\max_{b\in B} V(\mathrm{p}, m, b; \epsilon)$, and the second-step continuous choices, given b_{c} and θ, are $\upsilon^* = v(\mathrm{p}, m, b_{\mathrm{c}}; \theta, \epsilon)$, $t^* = \tau(\mathrm{p}, m, b_{\mathrm{c}}; \theta, \epsilon)$, and $x^* = x(\mathrm{p}, m; \theta, \epsilon)$, maximizing $U(v, t, x, b_{\mathrm{c}}; \theta, \epsilon, \zeta_{b*})$ subject to $p_x x + p(b_c)t + [q(b_c) + p(b_c)/b_c]v = m$.

With the specification of sections II B and III (that is, including ξ and Σ), we have

$$\hat{v}(p, m, b; \xi, \epsilon) = \frac{\exp\left(\epsilon_1 + \frac{1}{2}\sigma_\theta^2\right) + \xi' z}{\exp\left(\epsilon_1 + \frac{1}{2}\sigma_\theta^2 + \xi' z\right) + \exp(\epsilon_2) + \sigma_\theta^2} \frac{m}{q(b)},$$

whereas

$$v(p, m, b; \xi, \theta, \epsilon) = \frac{\exp(\epsilon_1 + \theta + \xi' z)}{\exp(\epsilon_1 + \theta + \xi' z) + \exp(\epsilon_2) + \sigma_\theta^2} \frac{m}{q(b)},$$

and analogously for $\hat{v}$, $\hat{x}$, τ, and x. Then, in abbreviated notation, the indirect ex ante utility of service quality b is $V(p, m, b; \xi, \Sigma, \epsilon) = \exp(\epsilon_1 + \frac{1}{2}\sigma_\theta^2 + \xi' \mathrm{z})\ \ln\,(\hat{v}) + \sigma_\theta^2 + \ln(\bar{\tau}) + \mathrm{e}^{\epsilon 2}\ \ln(\hat{x})$. Due to the separability of x in U, this implies that the discrete choice probabilities functionally only depend on ϵ_1 and, as a consequence of the extreme value assumption on $\{\zeta_b, b \in B\}$, are given by

$$P(b; \xi, \Sigma) = E_{\epsilon_1}\left[\frac{\exp[V(p, m, b; \epsilon)]}{\Sigma_{\beta\in B}\exp[V(p, m, \beta; \epsilon)]}\right].$$

APPENDIX B

The Experiment

The INDEX trial provides a group of 70+ subjects[17] with home Internet access via ISDN lines at various price-quality combinations. Service quality is defined in terms of nominal transmission speed or bandwidth, measured in kilobits per second (Kbit/s). ISDN lines are dedicated lines. This means that there is no capacity sharing between a user's home and the overprovisioned U.C. Berkeley campus network; there is capacity sharing for all Internet traffic that goes beyond the U.C. Berkeley network. In addition, there is however always potential sharing of destination server capacity. Hence, for many applications, the chosen nominal and actually delivered bandwidth are the same. This allows us to estimate a user's valuation of the chosen service, rather than of the mere promise, which is the primary objective of INDEX.

INDEX users pay for the service they use, according to their usage and the prevailing price structure. Customers participate in a sequence of subexperiments, or service plans. Each subexperiment runs over six to ten weeks and involves a different nonlinear price structure. The variable symmetric bandwidth experiment, for instance, levies prices per unit time symmetrically for in- and outbound traffic, differentiated by quality of service, that is, bandwidth. The volume experiment charges per unit byte volume, also differentiated by quality of service. Other experiments impose convex combinations of these price structures or tariffs involving flat-rate portions. The menu of available bandwidths is 8, 16, 32, 64, 96, and 128 Kbit/s. The first week in each experiment is a free trial period in which users can monitor the billing consequences of their choices under the prevailing price structure, without having to pay. In subsequent weeks, prices are drawn randomly for each user at the beginning of the week and remain in place throughout the week.

The data for this analysis come from the variable symmetric bandwidth experiment, in which users are constrained to select the same bandwidth for in- and outbound traffic and pay according to the time they spend in a chosen bandwidth. In the experiments, the users, in a first step, make a discrete quality-of-service choice, choosing a bandwidth out of the available menu. They thereby implicitly choose the price for subsequent utilization of this bandwidth, to the extent that unit prices are differentiated by bandwidth. In a second step, through the unobserved Internet-based applications they run, they make implicit byte volume and connection-time choices, given the initial bandwidth choice and the associated prices. Users can then switch bandwidths or disconnect entirely at any instant, virtually instantaneously and at no cost beyond the click of a mouse button.

The experiment records individual-level data whose finest resolution is (a fraction of) a one-minute time interval. For most of the analysis in this paper, the data for an individual user are aggregated to connections. For each connection, the chosen bandwidth, the in- and outbound volumes (in bytes), the duration (in seconds), the calendar day on which it was established, and the set of prevailing prices are recorded.

INDEX users are all affiliated with U.C. Berkeley. They belong to the faculty, the student population, or the staff. They are a relatively computer-literate sample. All of them have outside options to accomplish any Internet-based application from various campus locations or from home via the university dial-up system. The former option requires presence on campus. The latter does not afford the high transmission speeds at the upper end of the spectrum provided by INDEX. Usage of outside options cannot be monitored by INDEX and is a source of potential bias in estimation results.

Altmann, Rupp, and Varaiya (2001), Edell and Varaiya (1999), and Rupp et al. (1998) give supplementary information on INDEX, its

[17] Due to changes in residence, there has been some turnover in the subject pool.

technology, and its experimental design, as well as various model-free summary statistics. Varian (2000) provides a reduced-form analysis of INDEX data.

REFERENCES

Ai, C., and X. Chen, "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica* 71:6 (2003), 1795–1843.

Altmann, J., B. Rupp, and P. Varaiya, "Internet Demand under Different Pricing Schemes," Univ. of California, Berkeley, manuscript (2001).

Beckert, W., "On Specification and Identification of Stochastic Demand Models,' Birkbeck College, Univ. of London, manuscript (2004).

Blundell, R., "Consumer Behaviour: Theory and Empirical Evidence," *Economic Journal* 98 (1988), 16–65.

Blundell, R., X. Chen, and D. Kristensen, "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," Institute for Fiscal Studies, London, CEMMAP working paper (2003).

Blundell, R., and R. J. Smith, "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables," *Journal of Econometrics* 64 (1994), 355–373.

Brown, B. W., and M. B. Walker, "The Random Utility Hypothesis and Inference in Demand Systems," *Econometrica* 59 (1989), 925–951.

Chib, S., and E. Greenberg, "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory* 49:4 (1996), 327–335.

Darolles, S., J. P. Florens, and E. Renault, "Nonparametric Instrumental Variables Regression," GREMAQ, University of Toulouse, mimeograph (2002).

Das, M., W. K. Newey, and F. Vella, "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies* 70:1 (2003), 33–58.

Dubin, J. A., *Consumer Durable Choice and the Demand for Electricity* (Amsterdam: North-Holland, 1985).

Dubin, J. A., and D. L. McFadden, "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica* 52:2 (1984), 345–362.

Edell, R., and P. Varaiya, "Providing Internet Access: What We Learn from INDEX,' Univ. of California, Berkeley, manuscript (1999).

Hajivassiliou, V. A., and P. A. Ruud, "Classical Estimation Methods for LDV Models Using Simulation" (pp. 2382–2441), in R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics 4* (Amsterdam: North-Holland, 1994).

Hall, P., and J. Horowitz, "Nonparametric Methods for Inference in the Presence of Instrumental Variables," Department of Economics, Northwestern University, working paper (2003).

Lewbel, A., "Demand Systems With and Without Errors," *American Economic Review* 91:3 (2001), 611–618.

McFadden, D. L., "The Measurement of Urban Travel Demand," *Journal of Public Economics* 3:4 (1974), 303–328.

——— "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica* 57:5 (1989), 995–1026.

McFadden, D. L., and P. A. Ruud, "Estimation by Simulation," this *Review,* 76:4 (1994), 591–608.

Newey, W. K., and J. L. Powell, "Instrumental Variables Estimation for Nonparametric Models," *Econometrica* 71:5 (2003), 1565–1678.

Newey, W. K., J. L. Powell, and F. Vella, "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica* 67:3, (1999), 565–604.

Pakes, A., and D. Pollard, "Simulation and the Asymptotics of Optimization Estimators,' *Econometrica* 57:5 (1989), 1027–1057.

Rupp, B., R. Edell, H. Chand, and P. Varaiya, "INDEX: A Platform for Determining How People Value the Quality of Their Internet Access," in *Proceedings of the 6th IEEE/IFIP International Workshop on Quality of Service* (Piscataway, NJ: IEEE Press, 1998).

Rust, J., "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica* 55:5 (1987), 999–1033.

——— "Structural Estimation of Markov Decision Processes" (pp. 3081–3143), in R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics 4* (Amsterdam: North-Holland, 1994).

Varian, H. R., "Estimating the Demand for Bandwidth,' School for Information Management and Systems, Univ. of California, Berkeley, manuscript (2000).

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