Quant Mark Econ (2008) 6:299–338
DOI 10.1007/s11129-008-9044-x

Evaluating wireless carrier consolidation using semiparametric demand estimation

Patrick Bajari · Jeremy T. Fox · Stephen P. Ryan

Received: 29 March 2007 / Accepted: 26 August 2008 / Published online: 19 September 2008

Abstract The US mobile phone service industry has dramatically consolidated over the last two decades. One justification for consolidation is that merged firms can provide consumers with larger coverage areas at lower costs. We estimate the willingness to pay for national coverage to evaluate this justification for past consolidation. As market level quantity data are not publicly available, we devise an econometric procedure that allows us to estimate the willingness to pay using market share ranks collected from the popular online retailer Amazon. Our semiparametric maximum score estimator controls for consumers' heterogeneous preferences for carriers, handsets and minutes of calling time. We find that national coverage is strongly valued by consumers, providing an efficiency justification for across-market mergers. The methods we propose can estimate demand for other products using data from online retailers where product ranks, but not quantities, are observed.

P. Bajari
University of Minnesota, Minneapolis, MN, USA
e-mail: bajari@econ.umn.edu

P. Bajari · S. P. Ryan
National Bureau of Economic Research, Cambridge, MA, USA

J. T. Fox (✉)
University of Chicago, Chicago, IL, USA
e-mail: fox@uchicago.edu

S. P. Ryan
Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: sryan@mit.edu

Keywords Market share ranks · Semiparametric · Demand estimation · Amazon · Mergers · Antitrust · Telecommunications · Mobile phones · Online · Discrete choice

JEL Classification L4 · L63 · C35 · C13 · C14

1 Introduction

Currently the US mobile phone service industry is dominated by four large, national carriers: Cingular, Sprint, T-Mobile and Verizon. This market structure is a relatively recent phenomenon resulting from a series of mergers over the last 15 years. By comparison, in 1988 the company serving the largest number of the top 20 markets was US West, which served only four markets. As federal regulators can oppose mergers if consumer welfare is lowered through increased market power, cellular providers typically argue that mergers enhance consumer welfare through the introduction of better products. One major measure of product quality for a wireless carrier is the extent of its geographic coverage, which can be increased through cross-market mergers. Larger geographic coverage is desirable to consumers as otherwise they may face roaming fees and certain features on their phones may not work as well outside of their home calling area. The first plan to offer a comprehensive national network was the AT&T OneRate plan in 1998, which was offered after AT&T Wireless achieved a near-national scale through an extensive series of mergers. In this paper we provide an estimate of consumer willingness to pay for national coverage, which is useful for assessing whether the benefits from increased coverage counterbalance the reduced competition resulting from consolidation in this industry.

Academic work on wireless mergers has been hampered by a lack of quantity data. In the mobile phone industry, the relevant unit of analysis for demand estimation is a geographic market, such as a metropolitan area. Mobile phone carriers only release data on national customer counts, and not on the numbers of customers per geographic market. However, we exploit a novel source of market level data: product popularity rankings as reported by the online retailer Amazon. In addition to reporting the characteristics and features of cellular plans from a cross section of large U.S. markets, Amazon also ranks the popularity of each plan over a recent time interval. Several previous studies have used online rank data. Brynjolfsson et al. (2003), Chevalier and Goolsbee (2003), and Ghose and Sundararajan (2006) use insider knowledge and online experiments to empirically verify that book sales on Amazon follow a power law. The power law distribution is then used to construct estimates of the market shares to be used in traditional discrete choice estimation. In our setting, a consumer can only choose between 70–80 mobile phone subscription plans, many fewer than the millions of books on Amazon. It is unlikely that the same power law that holds for books can be applied to wireless calling plans.

We therefore devise a consistent estimator for discrete choice models when only rank data are available, building on Manski (1975)'s maximum score estimator.[1] The intuition of the estimator is straightforward: cellular plan A will be more popular than cellular plan B if and only if the mean utility from A exceeds the mean utility from B. Our estimator enumerates all possible comparisons of non-identical product pairs and maximizes the number of times that the predicted and actual ranking between two products are the same. Our analysis extends Manski's maximum score estimator in four ways. First, we show how to estimate utility parameters using aggregate data instead of individual level data. Second, we extend maximum score to the case where the dependent variable is a market share rank instead of an individual level choice. Third, we consider the case where the utility parameters are set identified instead of point identified. This is important in our application because one of Manski's requirements for point identification fails due to a lack of variation in calling plans across markets. For inference, we implement a version of the subsampling confidence regions from Romano and Shaikh (2008). Finally, our estimator allows for omitted product attributes and for heterogeneity in consumer willingness to pay for these attributes. Omitted product attributes are important in our application. For example, the econometrician will not observe the quality of cellular service from a given company in a particular city. However, this variable is likely to be important for consumer choice and may be positively correlated with price. Ignoring this form of unobserved heterogeneity will bias willingness to pay estimates.

The methods that we propose are useful for analyzing demand in other markets where product ranks are observable while market shares are unobservable. Many online retailers, including market leaders Amazon and Wal-Mart, allow the user to sort alternative products from the most to the least popular. Therefore, our methods can be used for data on a wide variety of product categories. All else held constant, it is of course preferable to use data with actual quantities because weaker econometric assumptions can be used. Despite this disadvantage, we believe in many cases data from online retailers have some strengths compared to alternative data sources. First, in an online market, the economist is able to observe the exact information about a product presented by the retailer to the consumer. In many empirical studies of differentiated product markets, there is a large gap between the economist's and the consumer's information about a product. Commonly, the economist only observes an average (or quantity weighted) price and a fairly incomplete list of product attributes. In online markets, the exact price is observed and online retailers often generate web pages in a manner that allows the consumers to compare a large number of characteristics across products. In our Amazon data, for example, it is possible to construct a matrix of 13 product characteristics for 70–80 plans across 22 markets. Second, measurement error

[1] The maximum score approach has previously been used in industrial organization and marketing by Briesch et al. (2002).

may be less problematic in data collected from online retailers. The web pages we downloaded from Amazon are contracts that describe the product and terms offered by the retailer. Online retailers have strong incentives to make sure that such information is reported accurately. By comparison, measurement error is a common (and ignored) problem in many studies of differentiated product markets. Third, online data can be collected freely for a diverse set of product offerings. Leading online retailers such as Amazon and Wal-Mart have product offerings in a wide array of categories. In many of these categories, high quality data are not publicly available to researchers from other sources.

2 Wireless carrier consolidation

Fox (2005) presents an overview of the history of mobile phone consolidation. The Federal Communications Commission (FCC) is the primary regulator of the mobile phone service industry. However, two carriers must win approval from the FCC and the Department of Justice to merge. Starting in 1996, the FCC began a gradual process of loosening and finally eliminating the spectrum cap, which is the fraction of the public radio waves allocated for mobile phone use that an individual carrier can control in a given geographic market. More across- and within-market mergers have occurred as the spectrum cap has been loosened and governmental objections to mergers have declined.

It is difficult to assess carrier market power at the level of a particular city using publicly available data sources.[2] Nevertheless, the degree of concentration in the mobile phone service industry has raised concerns about market power. FCC Commissioner Michael J. Copps, who has access to restricted-access market share estimates, writes in his statement of approval for the Sprint / Nextel merger the warning, "The average US market's HHI (Herfindahl-Hirschman Index) score has grown from 2,900 (before the Cingular/AT&T merger) to 3,100 (after the Cingular/AT&T merger) to 3,300 (after the Sprint/Nextel merger)." An HHI of 3,300 implies that there are now an equivalent of three equal sized competitors in most markets. The Department of Justice's horizontal merger guidelines suggest that any industry with an HHI above 1,800 is "highly concentrated."

The FCC writes in its approval of the Sprint/Nextel merger that a merger must "serve the public interest, convenience, and necessity." Carriers requesting regulatory approval of their mergers often write of the welfare-enhancing benefits of such mergers. One potential benefit of mergers is nationwide calling plans. AT&T Wireless initiated the first national calling plan in 1998. AT&T

[2]The FCC uses data on the number of telephone numbers assigned to carriers to approximate market shares, rather than using data on actual customers. The FCC writes in response to a Department of Justice request to access its data on market shares, "The Commission has recognized that disaggregated, carrier-specific forecast and utilization data should be treated as confidential and should be exempt from public disclosure under 5 U.S.C. Section 552(b)(4)."

introduced this service after it completed a series of takeovers and mergers that allowed it to achieve a nationwide calling scale. More carriers have begun offering national calling plans as more companies have merged.

Mergers could increase national calling plans for both technological and incentive reasons. On the technology side, the integration of features such as voice mail notifications, internet access and push-to-talk across different wireless networks is a complex task. A merger can standardize the implementation of advanced features across the entire carrier's coverage area.

On the incentive side, the theory of the firm suggests incentive problems may prevent the introduction of national calling plans. Data on the early cellular industry show that the per-minute customer charge for roaming was around 55% higher than the charge for placing calls in a customer's native coverage area (Fox 2005). Today, many subscription plans make roaming charges invisible to customers. However, published industry comments suggest that carriers transfer high per-minute fees between each other when one carrier's customer travels to a roaming market and places a call.[3] A related issue is pricing. Standard models of double marginalization predict that the sum of the profits of the roaming and home carriers will be less than the profits of a merged firm.

Whatever the reason, across-market mergers expand native calling areas and thus reduce roaming charges. If cost savings are passed on to consumers, lower roaming charges reduce the price premium national calling plans charge over regional plans. We present an econometric methodology that can be used to measure consumers' valuation for national calling. Due to data limitations, we do not estimate a total welfare analysis of mergers. However, our analysis is the first paper to provide estimates of the benefits of a key welfare parameter from cellular mergers.

3 Model and estimator

3.1 Consumer utility

In our data, we observe m geographically separated markets and a set of plans, along with their characteristics, in each market. Let J_m denote the set of plans offered in market m. We will let x_{jm} be a $d \times 1$ vector of plan characteristics, which in our application will includes features such as national coverage. Let

[3] Some carriers have decided to take advantage of the network aspect of their products by offering free in-network calling. If two Verizon or two Cingular customers talk, the length of the call is not deducted from either customer's bucket of included minutes from their subscription plans. Merging carriers create larger networks so that customers can better exploit free in-network calling. Unfortunately, major carriers either include unlimited in-network calling as part of all plans, or offer it as an add-on option. There is no variation within a carrier in whether in-network calling is included in a subscription plan, so we cannot estimate its value without using across-carrier variation in market share ranks.

$H_m \subset J_m$ denote a group of subscription plans. Similar to the nested logit, the set of all H_m will form a partition of J_m. As we will describe below, we form the nests H_m by grouping products with similar characteristics. In a duplication of notation, we sometimes use J_m to refer to the number of elements in a set, in addition to the set itself.

The utility u_{ijhm} of customer i for subscription plan j from nest h in geographic market m is

$$u_{ijhm} = x_{jm}^{'}\beta - p_{jm} + \nu_{im}^{h} + \xi_{jm} + \varepsilon_{ijm}. \tag{1}$$

We measure price in terms of cost per anytime minute, or

$$p_{jm} = \frac{\text{monthly price of plan } j \text{ in market } m}{\text{\# of anytime calling minutes of plan } j \text{ in market } m}. \tag{2}$$

We use prices per minute because they will simplify our interpretation of the value of national coverage.[4] In our model, national coverage is a dummy variable product characteristic in x_{jm} that shifts around the value of an anytime minute. Eliminating a surcharge is useful on every call. The parameter β^{Nat} is then a typical customer's per-minute willingness to pay for having no-surcharge national coverage instead of having to pay a surcharge when traveling. A motivation for approving a merger is that many customers have high valuations for national service and the merger could improve geographic coverage. If so, β^{Nat} should be large and statistically significant.

Assuming that demand is downward sloping, we normalize the coefficient on price to -1 without further loss of generality, as utility is only defined up to scale normalization. The $d \times 1$ vector β measures the willingness of the consumer to pay, relative to anytime calling minutes, for the product attributes x_{jm}. In (1), ν_{ihm} is a customer- and nest-specific unobserved preference (fixed effect),[5] ξ_{jm} is a market and subscription plan specific error, and ε_{ijm} is a customer and subscription plan specific error. Suppose that the nest h corresponds to family plans from Verizon offered in market m. The term ν_{ihm} would then represent the utility to household i of choosing a Verizon family plan. Our model has horizontal product differentiation that allows this utility to vary across households. Economic intuition suggests that such variation might be important. Young, single households may put little value on family plans compared to a household with several teenage children. The error term ξ_{jm} is a plan j and market m fixed effect. This term reflects vertical product differentiation in j within a single market. For example, a particular plan

[4] We did not collect data on overage charges. An overage charge is the per-minute cost for calls that exceed the calling minutes for plan j. While we have no data on the usage of plan minutes, our measure does not account for the entirety of a plan's price. Our assumption (relaxed a little in Section 5.4) is that a consumer uses all the minutes in his or her plan, and no more, so there are no overage charges.

[5] The nests do not represent a dynamic choice problem. Rather, each nest represents the set of products that have the same fixed effect for consumer i, ν_{ihm}.

with regional coverage may be popular in Atlanta.[6] The error term ε_{ijm} is a standard random shock to preferences; we describe our specific distributional assumptions on ε_{ijm} below.

Customer i chooses plan j when j maximizes his or her utility:

$$u_{ijhm} > u_{iklm} \forall k \in J_m, k \neq j, \tag{3}$$

where l is the nest of product k, and ties occur with probability zero.

3.2 Data generating process

Let X_m be the $J_m \times d$ matrix of the observable, non-price characteristics of all plans in market m. Let $\vec{p}_m$ be the vector of J_m prices. Let $\vec{\xi}_m$ be the vector of J_m market and product shocks. For a consumer i in market m, $\vec{\nu}_{im}$ is the vector of the H_m nest fixed effects and $\vec{\varepsilon}_{im}$ is the vector of the J_m product errors. Let I_m be the unobserved number of consumers in market m buying plans on Amazon. With some duplication of notation, let $h_m(j)$ be a function returning the nest of choice j. The following assumption summarizes the statistical assumptions on the exogenous variables.

Assumption 1 Across markets, $\left\{J_m, H_m, X_m, \vec{p}_m, I_m, \vec{\xi}_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}, \{\vec{\varepsilon}_{im}\}_{i\in I_m}\right\}$ is an independent and identically distributed random collection whose joint distribution and pattern of observability in the data satisfy the following properties:

1. J_m, H_m, X_m and $\vec{p}_m$ are observed by the econometrician.
2. I_m , $\vec{\xi}_m$, $\{\vec{\nu}_{im}\}_{i\in I_{m,}}$, and $\{\vec{\varepsilon}_{im}\}_{i\in I_m}$ are not observed by the econometrician.
3. I_m is a positive integer and, importantly, can be small.
4. There is some joint distribution $D\left(J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\right)$, which may be degenerate in some or all components, for the listed arguments.
5. Each plan and market error term ξ_{jm} for a plan in nest h is independent and identically distributed with the absolutely continuous density $g_{hm}\left(\xi_{jm} \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\right)$ and has full support on the real line. Each ξ_{jm} is conditionally independent from ξ_{jm}'s in other nests.
6. For a given consumer i plan j and nest h, ε_{ijm} is independent and identically distributed with with the absolutely continuous density $f_{ihm}\big(\varepsilon_{ijm} \mid J_m, H_m, X_m, \vec{p}_m, I_m, \vec{\xi}_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\big)$ with full support on the real line. Each ε_{ijm} is conditionally independent from ε_{ijm}'s in other nests.

Assumption 1 makes several key restrictions. Part 5 says that, conditioning on a nest of products in the same market, all product shocks ξ_{jm} have the same distribution. Therefore, while the vector of prices $\vec{p}_m$ can affect the shape of

[6]Consumer-specific fixed effects at the carrier level capture carrier-specific features such as the coverage near a consumer's house. In a model without fixed effects, the market and subscription plan specific errors ξ_{jm} would account for omitted variables such as coverage quality.

the marginal distribution, within a nest the price of product j in market m is independent of ξ_{jm}. Second, the within-nest, conditional iid assumption on the ξ_{jm}'s and the ε_{ijm}'s rules out some forms of consumer-level heterogeneity. We return to these concerns later, in Section 3.7.2.

On the other hand, Assumption 1 is flexible as to the shape of the marginal distributions of the errors. For example, two consumers in the same market may have completely different f_{ihm}'s, and these densities may even be a function of the realization of product levels shocks ξ_{jm}. This is in contrast to parametric discrete choice estimators, which almost always impose some restriction such as consumer errors are draws from the same distribution, such as the logit.

Finally, we are extremely flexible about the values of the nest and consumer specific ν_{im}^{h}'s. Each ν_{im}^{h} is a fixed effect: it can enter all conditioning arguments for consumer i and be correlated with all observables. For example, the ν_{im}^{h} can affect the shape of the marginal distribution of errors. Further, the ν_{im}^{h}'s of different consumers i can be correlated.

3.3 Consumer level choice probabilities

Manski (1975) introduced a pioneering semiparametric estimator known as maximum score.[7] In this paper, we seek to extend the original work in Manski on unordered choice with three or more choices to the case of aggregate data on market share ranks. Manski's maximum score estimator for individual-level data requires a property that choices with higher deterministic payoffs are chosen more often. This section reproduces Manski's rank order property at the individual consumer level, under our very similar assumptions.

Let $\Pr_{im}\left(j \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\right)$ be the probability that consumer i in market m chooses calling plan j, conditional on the number of plans, the number and division of plans into nests, plan observables including price, the unobserved number of consumers, and the nest fixed effects of consumer i and other consumers. $\Pr_{im}\left(j \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\right)$ is thus an integral over two sets of random variables: the J_m market and product shocks ξ_{jm} and the J_m consumer and product shocks ε_{ijm}. The subscript im emphasizes that this function is different for different consumers because the error densities can vary across consumers.

Lemma 1 *Let $h_m(j) = h_m(k)$, and let consumer i be given. Under Assumption 1,*

$$x_{jm}^{'}\beta - p_{jm} > x_{km}^{'}\beta - p_{km}$$

[7]Manski initially studied the properties of this estimator both for the two choice case and the three or more choices case. Much of the attention in the subsequent maximum score literature focuses on the two choice case, because the two choice case allows for relatively weak median independence assumptions about the relationship between errors and observables (Manski 1985; Horowitz 1992). Also, others have extended estimators for the two-choice case to more general ordered choice problems (Han 1987; Abrevaya 2000).

if and only if

$$Pr_{im}\left(j \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{v}_{im}\}_{i\in I_{m,}}\right)$$
$$> Pr_{im}\left(k \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{v}_{im}\}_{i\in I_{m,}}\right).$$

Proofs are in the Appendix. The innovation over the proof in Manski (1975) is minor: we work with two errors, ε_{ijm} and ξ_{jm}, and the distribution of ε_{ijm} is parameterized by ξ_{jm}.[8]

Lemma 1 compares only two plans in the same nest h of plans. Let h be the nest of all plans offered by the carrier Verizon. Let customer i have an unobserved willingness to pay of $\nu_{im}^{\text{Verizon}}$ for Verizon plans. Let plan j and k both be from Verizon. Customer i prefers j to k if

$$x_{jm}^{'}\beta - p_{jm} + \nu_{im}^{\text{Verizon}} + \xi_{jm} + \varepsilon_{ijm} > x_{km}^{'}\beta - p_{km} + \nu_{im}^{\text{Verizon}} + \xi_{km} + \varepsilon_{ikm}. \quad (4)$$

The common preference $\upsilon_{im}^{\text{Verizon}}$ differences out, and the choice inequality reduces to the condition

$$x_{jm}^{'}\beta - p_{jm} + \xi_{jm} + \varepsilon_{ijm} > x_{km}^{'}\beta - p_{km} + \xi_{km} + \varepsilon_{ikm}. \quad (5)$$

By comparing two plans from Verizon, we do not need to make a functional form assumption about how preferences for Verizon differ across the population, and about whether the preferences for Verizon are correlated with the preferences for plans from other carriers or with the observed characteristics of the plans (x_{jm}) from different carriers.

Nest h and agent i fixed effects ν_{im}^{h} satisfy four roles in our application:

1. Each consumer may have a certain need to talk on the phone. If a nest is restricted to plans with similar numbers of minutes of airtime, then we control for a consumer's heterogeneous tastes for mobile communication.
2. Each carrier has its own network of cellular base stations, so one carrier may have better coverage near a consumer's home or workplace than another carrier. The fixed effects capture a consumer's views on the quality of coverage for each carrier.
3. Consumers simultaneously buy calling plans and phones. Our data cannot pair individual phone and plan purchases. The fixed effects capture a consumer's views on each carrier's phones, as we define a nest narrowly enough so that all plans can be paired with the same set of phones.
4. Fixed effects can be correlated with observable plan characteristics, including price. This would be the case, for example, if Verizon optimally chose its phone lineup in conjunction with its menu of plans.

[8] We can weaken the assumption of i.i.d. errors across choices in the same nest to be an exchangeable joint density. See Fox (2007).

3.4 Market share ranks

On Amazon, we are only able to observe market share ranks, not data on individual purchases. To extend the maximum score estimation strategy to this case, we first want to show that plans with greater payoffs will have a higher expected market share rank. Recall that Assumption 1 does not in any way assume that the error densities of two consumers in the same market are the same. Also recall that the same ξ_{jm} appears in the choice problem of all consumers. Under these weak restrictions, the additive separability of a market share allows us to sum the choice probabilities of consumers despite the non-i.i.d. nature of choices across consumers. Also, we do not assume the researcher has data on I_m. Define the market share of product j in a market m with I_m consumers to be

$$s_{jm} = \frac{1}{I_m} \sum_{i=1}^{I_m} 1\left[i \text{ buys } j\right],$$

where $1\left[i \text{ buys } j\right]$ is an indicator equal to 1 when consumer i buys subscription plan j. Then the following lemma is true.

Lemma 2 *Let $h_m(j) = h_m(k)$. Under Assumption 1,*

$$E\left[s_{jm} \mid J_m, H_m, X_m, \vec{p}_m\right] > E\left[s_{km} \mid J_m, H_m, X_m, \vec{p}_m\right]$$

if and only if

$$x'_{jm}\beta - p_{jm} > x'_{km}\beta - p_{km}.$$

For a given market, we see whether one plan has a higher market share than another plan. Let r_{jm} be the rank of plan j in market m, with higher ranks corresponding to plans with higher shares. If there are 70 plans, the plan with the highest market share has a market share rank of 70, not 1. We want to prove the property that a plan with a higher mean payoff will have higher market share rank more often than not. Given two random variables a and b, it is possible that a has a higher mean than b even though the random variable $1\left[a > b\right]$ has a mean less than 1/2. Fortunately, our decision model is well behaved and the expected market share ranks of products in the same nest are rank ordered by their mean payoffs. Let $\Pr_m\left(r_{jm} > r_{km} \mid J_m, H_m, X_m, \vec{p}_m\right)$ be the conditional probability that product j is chosen more often than product k in market m. The subscript m emphasizes that this market share equation can vary across markets because different consumers may have different error densities, and because the densities of product level shocks may vary across markets as well.

Lemma 3 *Let $h_m(j) = h_m(k)$. Under Assumption 1,*

$$Pr_m\left(r_{jm} > r_{km} \mid J_m, H_m, X_m, \vec{p}_m\right) > Pr_m\left(r_{jm} < r_{km} \mid J_m, H_m, X_m, \vec{p}_m\right)$$

if and only if

$$x'_{jm}\beta - p_{jm} > x'_{km}\beta - p_{km}.$$

The proof of the lemma deals with both the integer nature of a market share rank (r_{jm} is an integer while s_{jm} is a real number) and the fact that each consumer has different fixed effects and a different distribution for the error terms ε_{ijm}. Details are contained in the Appendix.

3.5 Statistical objective function

We could estimate β by finding the set of parameters that maximize the objective function

$$Q_M(\beta) = \frac{1}{M}\sum_{m=1}^{M}\sum_{h=1}^{H_m}\sum_{j,k\in J_{hm},k\neq j} 1\left[r_{jm} > r_{km}, x'_{jm}\beta - p_{jm} > x'_{km}\beta - p_{km}\right]. \quad (6)$$

The objective function uses data on plan characteristics and market share ranks for M markets. For each market m, the objective function sums over the H_m pre-specified nests of plans. Each nest h has J_{hm} plans and the estimator compares all pairs of plans j and k. In the data and ignoring ties, one of the two products j or k has a greater market share rank r_{jm}. If in the data $r_{jm} > r_{km}$, for a trial guess of β the estimator asks if indeed j has a higher mean payoff $x'_{jm}\beta - p_{jm}$. If j does, the prediction from Lemma 3 is satisfied, and the objective function increases by 1.

As the goal is to maximize the count, or score, of correct predictions of Lemma 3, (6) is a maximum score objective function. As there are a finite number of inequalities, typically there might be a set of parameter vectors β that maximize the objective function. The set estimate $\hat{B}_M$ of the willingnesses to pay is the set of parameter vectors that maximize the objective function: $\hat{B}_M \equiv \arg\max Q_M(\beta)$. Note that maximum score is a partial identification estimator: the model has other components, such as the distribution of the error terms, but only β is estimated.

In our Amazon data, the set of plans is identical across markets ($p_{jm} = p_{jn}$, $x_{jm} = x_{jn}$), with one notable exception. It is likely that each carrier designs a default national set of plans to offer in all markets, and then makes small adjustments based upon regional conditions. Our data do not have the property that sampling new markets gives much variation in the set of plans in a market. We are far from the requirements in the semiparametric discrete choice literature that at least one plan characteristic has continuous support on the real line (Manski 1985). In the mobile phone calling plan industry, plans are chosen nationally. Collecting data on different geographic markets at the same point in time is not representative of the asymptotic argument needed for point identification. Our parameters are only set identified.

This section suggests that our estimator is consistent for the identified set B^0 as the number of markets M goes to infinity. For simplicity, we assume that all markets have the same set of J plans with characteristics X, and the only differences in the market shares of plans across geographic markets arise from product market specific shocks ξ_{jm} and the small number of consumers, I_m , that purchase plans over the period Amazon uses to calculate market share ranks.

With sampling error, many inequalities will not be satisfied, even when evaluated at the true parameter vector β^0. Here we prove that the probability limit of our objective function with sampling error is uniquely maximized by the parameters in the identified set. In other words, in the limit the introduction of sampling error does not alter the identified set. That the set of maximizers of the probability limit of the objective function equals the identified set is a property of our objective function's functional form; it may have been the case that sampling error changes the set of maximizers.[9]

Lemma 4 *Let the exogenous characteristics of products be the same in each market. Under Assumption* 1*, the set of parameters that maximizes the probability limit, as $M \to \infty$, of $Q_M(\beta)$ is B^0, the identified set defined as*

$$B^0 \equiv \Big\{\beta : x_j^{'}\beta - p_j > x_k^{'}\beta - p_k \text{ whenever } x_j^{'}\beta^0 - p_j > x_k^{'}\beta^0 - p_k,\ \forall\, j, k \in J_h, h \in H,\ j \neq k\Big\},$$

where β^0 is the true parameter in the data generating process.

Typically a lemma exploring the probability limit of the objective function is an input into a consistency proof along the lines of the general theorems in Newey and McFadden (1994). However, if the model is only set identified, one already has to report a set as the estimate. Typically one reports a 95% confidence set for the parameters in the identified set, rather than both an estimate of the set and a confidence set based on that set. We return to the technical details of set inference in Section 3.8.

The proof of identification under sampling error is in an Appendix. Note that because of sampling error, the theoretical maximum number of inequalities will not be satisfied, even in the limit. The identification proof works by adding more inequalities from new markets, rather than eliminating sampling error for a fixed number of markets.

Han (1987) presents a similar objective function to (6), and calls the objective function a maximum rank correlation estimator. Sherman (1993) derived the asymptotic distribution of this estimator. The main distinction between maximum score and maximum rank correlation, in this case, is the asymptotic

[9]For example, if we have added a penalty term $\left(x_j^{'}\beta - p_j > x_k^{'}\beta - p_k\right)^2$ in the degree of an inequality violation and minimized the resulting penalties, then there is no guarantee that the true parameter in the data generating process would be in the identified set.

argument to eliminate sampling error. In maximum rank correlation, each of the terms $J_{hm} \to \infty$. As new products are added to a nest, the number of inequalities increases at the rate J_{hm}^2, because of the double summation for each nest. In our empirical application, J_{hm} is typically 2 or 3 while M is 22. Although both samples are small, we believe the large M asymptotics may be more believable than the large J_{hm} asymptotics.[10]

3.6 Identification despite Amazon not offering all plans

We can identify the willingness to pay parameters β even though Amazon does not offer all plans. Many parametric demand models, such as the multinomial probit and random coefficients logit, are not consistent under similar conditions. Fox (2007) introduces and formally proves this property for the individual data maximum score estimator. Consistency is preserved because maximum score, and by extension our Assumption 1, involves only comparisons between pairs of products in the same nest of choices. Relative choice probabilities and hence market share ranks for products in the same nest are preserved by conditioning on the event that an agent purchased one of the pair of products. By contrast, the multinomial probit and random coefficients logit impose functional form assumptions for the distribution of heterogeneity that do not survive conditioning on an endogenous outcome. Conditioning on an endogenous outcome induces correlation between observables and unobservables, resulting in inconsistency due to selection.

3.7 Comparison to the BLP assumptions

Berry, Levinsohn, and Pakes (Berry et al. 1995), or BLP, introduce a set of econometric assumptions and estimation procedures that are now accepted as conventional when working with aggregate data on product characteristics and market share levels. For applications of the BLP framework, see Nevo (2001) and Petrin (2002). Below we describe how our model and estimator compare to BLP.

3.7.1 Correlation of price with unobserved product attributes

BLP use instruments to control for the price endogeneity from the correlation of p_{jm} with the unobserved ξ_{jm}. If instruments are available, one can control

[10] See Fox (2008) for another case where an objective function like (6) has both Manski (1975) and Han (1987) asymptotics. Note that Han (1987) motivates his estimator with ordered choice problems. His estimator involves combining different observations in a double summation. We study an unordered choice problem. Under a much stronger version of Assumption 1, we could use Han's estimator to interact observations across markets if all markets had exactly the same set of plans. While most of the plans on offer are the same across markets, there is some small degree of variation in the offered plans, so we do not pursue this further.

for price endogeneity using a maximum score estimator as well. Fox (2007) introduces a maximum score instrumental variables estimator. See also Hong and Tamer (2003) for the two choice case.

In many applications, instruments may not be available. Even when instruments are available, their validity may be a source of disagreement among researchers. We do not have instruments for price (such as carrier-specific cost shifters) in our mobile phone data from Amazon. Our approach relies instead on first differencing in order to deal with the endogeneity of price. Equations (4) and (5) shows that we can eliminate ν_{im}^{h} from our objective function. This intuition behind our identification strategy is straightforward. We will group plans into nests of products where a priori we believe that ν_{im}^{h} is identical, e.g. Verizon Family plans. We then use the within-nest variation in product characteristics in order to identify the preference parameters β.

Note that our framework allows for horizontal and vertical product differentiation in the omitted attribute. The BLP framework assumes that the omitted product attribute is purely vertical. This is natural in our application because, for example, Verizon family plans may not be highly valued by young single consumers compared to households with large families. Our framework allows for this type of heterogeneity while BLP does not. However, our framework imposes the restriction that ν_{im}^{h} does not vary within products in the same nest. Product specific demand shocks ξ_{jm} are assumed to be independent.

3.7.2 Heterogeneity over tastes for product characteristics

In many applications, consumers have heterogeneous tastes over the observable characteristics of products. For mobile phone calling plans, consumers are likely to have heterogeneous willingnesses to pay to talk on the phone. A salesperson with a lot of clients may be willing to buy an expensive plan that offers 6000 minutes of daytime calling a month, while a person who uses his or her phone only in emergencies may prefer the plan with the least amount of minutes.

BLP would capture heterogeneity in the willingness to pay for anytime minutes using a random coefficient specification. Typically, these random coefficients are assumed to be independent and normal. Also, most researchers allow for a small number of random coefficients. This is because of the computational burden of estimating more flexible models.

No scholar has generalized results for the two-choice model from Manski (1975) about the semiparametric identification of the mean willingness to pay under random coefficients to the case with more than two choices, which was also studied by Manski (1975).[11] Given this, researchers working with

[11]Fox (2007) discusses this point in more detail for the maximum score estimator with individual data.

our estimator and market share rank data cannot use a random coefficients specification.[12]

In our framework, we assume that taste heterogeneity across households can be captured by ν_{im}^{h} and ε_{ijm}. Therefore, heterogeneity in tastes is allowed to vary freely within products within the same nest. If product nests are defined quite narrowly, our framework can allow for considerable heterogeneity in a flexible framework. However, the tastes for the observed product characteristics are assumed not to vary across households and are captured by a fixed vector of parameters, β.

3.8 Set inference

Our objective function is set identified as M goes to infinity. There are several recently-developed methods for performing inference on set-identified estimators, for example Andrews et al. (2005), Chernozhukov et al. (2007), Beresteanu and Molinari (2008), Galichon and Henry (2006), Imbens and Manski (2005), Pakes et al. (2006), Romano and Shaikh (2008), and Rosen (2006).

Let β lie in the parameter space B, and let B_0 be the identified set that minimizes the probability limit of (10). For a given level α, we follow the approach of Romano and Shaikh (2008) to construct a confidence region for identifiable parameters C_M that satisfies

$$\liminf_{M\to\infty} \Pr\{\beta \in C_M\} \geq 1-\alpha \;\forall \beta \in B_0. \tag{7}$$

Our confidence region is

$$C_M = \left\{\beta \in B : a_M \left(Q_M(\beta) - \sup_{\beta'} Q_M\left(\beta'\right)\right) \leq \hat{d}_M(\beta, 1-\alpha)\right\}, \tag{8}$$

where a_M is a normalizing constant and the critical value $\hat{d}_M(\beta, 1-\alpha)$ for the parameter β is determined by subsampling. Let the subsample size $b_M < M$ be a sequence of positive integers satisfying $b_M \to \infty$ and $b_M/M \to 0$. Let there be N_M subsamples of size b_M drawn from the original data. The $1-\alpha$ critical value is the $1-\alpha$ quantile of the subsampled distribution of the size b_M objective function, or

$$\hat{d}_M(\beta, 1-\alpha) = \inf\left\{q : \frac{1}{N_M}\sum_{b=1}^{N_M} 1\left[a_{b_M}\left(Q_{b_M}(\beta) - \sup_{\beta'} Q_{b_M}\left(\beta'\right)\right) \leq q\right] \geq 1-\alpha\right\}.$$

The main challenge in using any of the existing set inference procedures for maximum score estimation is theoretically verifying the technical conditions.

[12]Bajari et al. (2007) prove the nonparametric identification of the distribution of random coefficients in the random coefficients logit model, with market share levels. Our results rely on continuous product characteristic variation across markets, the type of variation that we do not have in the mobile phone market. Therefore, we are skeptical about identifying the distribution of random coefficients with this type of data.

In Romano and Shaikh (2008), the main technical condition is that the limiting distribution of the normalized objective function $a_M Q_M(\beta)$ exists and (more importantly) has a continuous $1-\alpha$ quantile, for each β in the parameter space B. Unfortunately, it is not a valid technical argument to use the limiting distribution from the point identified case in Kim and Pollard (1990) to argue separately for each $\beta \in B$ that the limiting distribution has a continuous $1-\alpha$ quantile.

However, in another paper, Romano and Shaikh (2006) prove a distinct validity result for subsampling that requires only that the limiting distribution exists, with no requirement about the limiting distribution being continuous. This construction requires inflating the confidence regions using any constant $\delta > 0$. In particular, the modified δ-inflated confidence set is

$$C_M^{\delta} = \left\{ \beta \in B : a_M \left(Q_M(\beta) - \sup_{\beta'} Q_M\left(\beta'\right) \right) \leq \hat{d}_M(\beta, 1-\alpha) + \delta \right\}.$$

The only requirement on δ is that it is strictly positive, so it can be small like $\delta = 0.000001$. Note that the δ correction is on the objective function values, not the parameter space. The following lemma restates Theorem 5 in Romano and Shaikh (2006) for the special case of our application.

Lemma 5 *For some sequence a_M, let the distribution of $a_M\Big(Q_M(\beta) - \sup_{\beta'} Q_M\left(\beta'\right)\Big)$ exist. Then the confidence region C_M^{δ} for any scalar $\delta > 0$ is asymptotically valid, in the sense of satisfying (7).*

We use this Romano and Shaikh (2006) theorem to avoid having to derive the limiting distribution and show that it has a continuous $1-\alpha$ quantile. In our dataset, our subsampled confidence regions using the Romano and Shaikh (2008) procedure are exactly equal to the set of parameters that maximize the objective function, with or without the δ inflation. This is an unusual empirical result, and arises because the amount of variation in the ordering of any two plans' market share ranks is very small across markets.[13] We explain this empirical result in more detail in the results section. Consequently, the small-δ inflation does not change our empirical results, and, more surprisingly, there is no distinction between the set estimates and the 95% confidence sets.

[13] A referee points out that for some set-identified estimators, the 95% confidence sets and set estimates will be the same with probability 1. Maximum score is not such an estimator. If the dependent variable (market share ranks) vary a lot conditional on covariates, the estimates using some subsamples will not be the same as with the full sample.

Note that we are discussing variation in the dependent variable. If the independent variables vary a lot across markets, then the identified set will be smaller than a case with no or a small amount of characteristic variation. If at least one independent variable per product has continuous support, then Manski (1985) and others show that the model is point identified and that the maximum score estimator is consistent. The typical large support assumption for the continuous characteristic can be relaxed while still maintaining point identification, as Horowitz (1998) discusses.

We pick the normalizing constant $a_M = M^{2/3}$ as an informed choice given the rate of convergence derived by Kim and Pollard (1990) for the point identified case.

3.9 How to program the estimator

The objective function for our estimator is easy to program. First, for a given dataset the market share ranks are data are known. So in (6), only the inequalities corresponding to $r_{jm} > r_{km}$ are relevant and need to be hard coded into the objective function that is programmed. For our dataset, the objective function is the sum of the forthcoming equations (11), (12) and (14). These indicator functions are very simple to program. To find a global maximizer, one should use a global search algorithm such as differential evolution, simulated annealing or another stochastic algorithm. Set inference using subsampling is slightly harder to program. Santiago and Fox (2007) provide computer code that implements the Romano and Shaikh (2008) set inference subsampling procedure for maximum score objective functions.

4 Amazon data

4.1 Markets and plans

Data availability has been an impediment to studying antitrust issues in the wireless phone industry. Carriers release annual, nationwide data on total subscribers, not market-level subscribers. Also, carriers do not report on the popularity of their individual subscription plans. The FCC has access to the sets of phone numbers given to carriers, and the FCC uses the information to approximate market shares. However, the FCC does not release its confidential data to researchers. We asked.[14]

The online retailer Amazon sells mobile phones and attached subscription plans. An online retailer faces a disadvantage because consumers cannot physically examine a phone as they can in a brick-and-mortar store. Amazon's competitive advantage is that it offers deep mail-in rebates on phones, so its prices for buying new phones (when attached to a new one or two year subscription) are often lower than the prices on a carrier's own website. Amazon does not discount the monthly fees of subscription plans. Monthly fees are billed directly by the carrier providing service.

We collect detailed plan characteristic and market share rank data from Amazon's site. For plan j in market m, we observe the plan's monthly fee, p_{jm}, and a vector of d other plan features, x_{jm}. Many features, such as the

[14]Some companies collect phone bills from consumers. Bill harvest data are not entirely appropriate, as at any given point in time the stock of mobile phone users has plans purchased from the menus of plans available in many different time periods. The time of plan purchase may not be observable.

one-time activation fee, are constant across the comparisons that we include in our maximum score objective function. Amazon rank orders the top selling plans for different geographic markets, so we observe the rank order r_{jm} of the market share of each Amazon plan in each geographic market.[15]

Practically speaking, a carrier must own a FCC license in a geographic market to enroll subscribers there. The competitors in each market may differ, as a fixed number of licenses are issued per market. Therefore, the competition for customers is primarily local. The set of plans offered by carriers is relatively constant across markets; the choice of the menu of plans appears to primarily operate on the national level. We collect data on different markets to increase statistical precision.

We collected data for 22 of the largest metropolitan areas in the United States.[16] Amazon chooses the boundaries of markets; we do not know them. However, the boundaries appear to correspond to the popular notion of a metropolitan area. For example, the same plans and market share ranks appear for the nearby cities of Los Angeles and Riverside, CA, but the plans and market share ranks differ for San Diego, CA, which is typically considered a separate metropolitan area. We use only markets that both Amazon and we agree are separate metropolitan areas. We have verified that the market share ranks are not volatile over a week, particularly for the plans with the highest sales ranks. Ghose and Sundararajan (2006) mention that Amazon uses a rolling window of sales to compute market share ranks. In the summer of 2004, Amazon moved to a system that calculates market share ranks using exponential decays that give more weight to recent purchases.

Amazon offers plans from all five national carriers: Cingular, Nextel, Sprint, T-Mobile and Verizon. The smaller carrier EarthLink Wireless offers plans on Amazon in 15 of the 22 markets.[17] We use data on only plans from T-Mobile and Verizon. Only carriers that offer both regional and national plans provide variation that can identify the willingness to pay for national coverage. Table 1 lists the number and characteristics of the plans that we use

[15] If a researcher has another dataset with continuously measured shares s_{jm}, it is easy to convert those shares into ranks.

[16] We have used the Amazon site extensively and wish to explain a little of how the site worked in late 2005, when the data were collected. When you go to the site to shop for mobile phone plans, you are prompted to enter your zip code. Amazon uses the zip code to look up your geographic market. We collect market share rank data by choosing a zip code corresponding to each city. There are various pages. The page that presents the plans rank ordered by sales is reached by using the toolbar to search "Wireless Plans" for a blank string. All plans will appear, and you can sort them by sales rank. The resulting plans are presented in a matrix, with the top three plans in the first row, plans three to six in the second row, etc. A plan's rank comes from its position in the matrix, not from a text label, as Amazon includes for books. You can verify the ordering of plans in the matrix by consulting another page, which lists the top five plans in numeric order, with the rankings listed explicitly. We clicked on each plan and manually copied its characteristics.

[17] Amazon sells prepaid service, where a customer does not pay a set monthly fee. We do not consider the data on prepaid service, because plan characteristics such as price and anytime minutes are not comparable to the monthly values for subscription plans. A customer that uses all of his or her monthly minutes will find it cheaper to subscribe to a monthly plan.

Table 1 Plan characteristics do not vary across markets: Plans used in estimation

Carrier Nest	# of Plans	# of Markets used	Minutes and Prices
T-Mobile National	2	22	{2500, $99.99},{5000, $129.99}
T-Mobile Regional	1	22	{3000, $49.99}
Verizon National	4	7	{450, $39.99},{900, $59.99}, {1350, $79.99},{2000, $99.99}
Verizon Regional	3	7	{600, $49.99}, {1200, $69.99}, {1800, $89.99}
Verizon Family National	3	7	{700, $60},{1400, $80},{2100, $100}
Verizon Family Regional	3	7	{600, $49.99},{1200, $69.99}, {1800, $89.99}

The Verizon national plans are in all 22 markets; we only use markets that also have regional plans. The table does not list all the plans offered by these carriers. T-Mobile offers lower minute national plans and Verizon and Verizon Family offer higher minute national plans. These plans do not directly compare in minutes to regional plans, and so do not enter the maximum score objective function. Figure 1 will show plan popularity is inversely correlated with the number of minutes in the plan. The T-Mobile lower minute plans are popular; the Verizon high-minute plans are less popular.

in estimation. Table 1 shows that, for a typical market, we use data on three T-Mobile plans, seven Verizon plans, and six Verizon family plans. Note that the characteristics of individual plans do not vary across markets. In relation to the semiparametric discrete choice literature (Manski 1985), the lack of variation in the characteristics of individual plans across markets is why we argue our estimator is set and not point identified in the limit. A key characteristic of a plan is its number of anytime minutes: minutes of airtime that a consumer can use to make phone calls without paying more than the monthly price listed in Table 1. If a user exceeds the bucket of anytime minutes, the user faces a high incremental charge.

Table 1 also describes the number of plans that have only regional coverage. A plan can allow a traveler to make calls from across the United States with no surcharges (national coverage), or a carrier may levy such surcharges on travelers (regional coverage). Most plans in our data offer national coverage. However, there are important regional plans. T-Mobile offers one regional plan with 3000 minutes. T-Mobile charges a regional subscriber an extra 49 cents a minute when using his or her phone outside his or her home region. Verizon offers three regional individual plans and three regional family plans. Both family and individual plans charge 69 cents a minute for roaming. The Verizon regional plans are offered only in markets in the West: Denver, Los Angeles, Phoenix, Portland, San Diego, San Francisco and Seattle. The Verizon regional plans are only present in the West offline as well; Verizon does not offer regional plans in the remainder of the country. Note again that the characteristics of individual plans do not vary across markets: only the presence of Verizon regional plans at all varies.

All regional plans levy long distance surcharges when a customer phones a customer in another market. T-Mobile and Verizon both charge 20 cents a minute for long distance. Verizon allows regional subscribers to upgrade to free long distance for \$5 a month.[18]

We will estimate the willingness to pay for national coverage, the opposite of regional coverage. We only compare plans to plans in the same carrier nest and in the same market. We also only compare a plan to plans with slightly more or fewer anytime minutes. Therefore, our estimator uses only within-market variation. The fact that Verizon regional plans are not found in all markets is not a source of variation that our maximum score estimator exploits.

4.2 Popularity of national coverage

Our goal is to estimate the value of national coverage. Our identification strategy will compare the market share ranks of regional and national plans. This section describes our data on plan market share ranks, r_{jm}. We change market share rank from an integer to a percentile, by the formula

$$\tilde{r}_{jm} = \frac{r_{jm}}{\max_{j \in J_m} \left(r_{jm} \right)}. \tag{9}$$

This normalizes the ranks, so that the most popular plan in a market has a percentile rank of 1, and the least popular plan has a percentile rank of close to 0. We then compute the mean of each plan's percentile rank across the markets where the plan is offered. Most plans are offered in all markets.

Table 1 shows T-Mobile, Verizon and Verizon Family offer plans with regional coverage. For those three carrier nests, Fig. 1 plots the mean percentile ranks of each plan by its monthly bucket of anytime calling minutes, along with a fitted quadratic and its confidence interval. The plan labels are monthly prices in cents divided by monthly anytime minutes, our measure of price from (2). Figure 1 compares plans across the three carrier nests, which we will not do in our willingness to pay estimates. Figure 1 also lists some T-Mobile and Verizon plans that are not in our estimation sample; we include these to document the popularity of plans that are dropped because they are not in the same nests as regional plans, so they provide no information about the willingness to pay for national coverage.

Plans with more monthly minutes are less popular. Plans with many minutes cost less per minute: the 6000 minute Verizon individual plan costs 3.3 cents per minute, while the 450 minute plan costs 8.9 cents per minute. High minute plans are unpopular not because they are a bad deal, but because many consumers do not have such a strong need to talk on the phone. To address this unobserved heterogeneity in demand for anytime minutes, our

[18] While not in the table and our estimation sample, Nextel offers four plans that do not include free long distance. Nextel uses a proprietary phone technology that prohibits its customers from operating on networks owned by almost all other carriers. Consequently, Nextel does not levy charges to travelers, in part because its phones are incapable of operating off its network. We do not consider Nextel. Also, Cingular dropped its regional plans from Amazon just before data collection began.

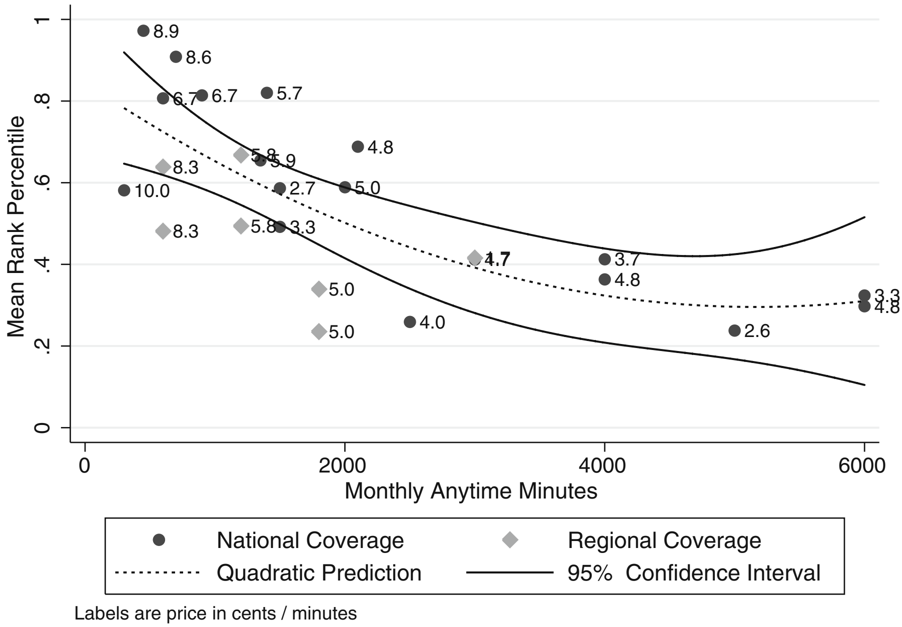

Fig. 1 Across-market mean market share ranks of national and regional plans. An observation is a calling plan. The rank percentile is (9). The number on the vertical axis is the mean rank percentile across the 22 geographic markets in our data. The number of anytime minutes of a plan are described in Table 1. The labels of the observations are the monthly prices of plans divided by the number of anytime minutes, our measure of price from (2). The labels are written in cents rather than dollars

structural estimator uses only comparisons between plans with similar numbers of minutes.

Figure 1 shows the relative popularity of various plans. Figure 1 shows that plans with regional coverage are less popular than plans with national coverage. Four of the seven regional plans are below the fitted quadratic's 95% confidence interval, and three plans are within the 95% confidence interval. No regional plans lie above the confidence interval, although many national plans do.

Verizon individual regional plans are always less popular than similar national plans. For example, the 600 minute Verizon regional plan has a mean percentile rank of 0.64, which is much lower than the ranks of 0.97 for the 450 minute national plan and 0.81 for the 900 minute national plan. On a price per minute basis, the Verizon regional plans are bad deals. The 450 national plan charges 8.9 cents per minute, while the 600 minute regional plan charges 8.3 cents per minute.

T-Mobile discounts regional plans more. The 3000 minute T-Mobile regional plan charges 1.7 cents per minute, while the 2500 and 5000 minute plans with national coverage charge 4.0 and 2.6 cents per minute, respectively. The 3000 minute regional plan is more popular, with a mean percentile rank of

0.42 compared to the ranks of 0.26 and 0.24 respectively for the 2500 and 5000 minute national plans. Nonetheless, the low price per minute of the T-Mobile regional plan suggests that consumers may have a positive willingness to pay for national coverage.

4.3 Representativeness of Amazon users

Customers who use Amazon differ in preferences from non-users. Section 3.6 shows that our estimator is consistent for the population willingness to pay even if people have different probabilities of knowing about Amazon's cell phone offerings, and if plans from some carriers are not offered on Amazon.

Another possibility is that Amazon users have a higher willingness to pay for national coverage. To address this, we turn to an auxiliary dataset on internet use. The market-research firm Forrester surveyed 68,664 Americans in its 2005 Technographics Benchmark survey. The Forrester data oversample heads of household, as only 2.3% of reported mobile phone users in the survey are under 25. As Amazon has no salespeople to field questions from new users, we suspect most mobile phone customers on Amazon are upgrading to a new phone, not buying a phone for the first time. We use the roughly 70% of respondents that report owning a mobile telephone as our base sample. Of mobile phone users, 40% have purchased an item online at least once in the past twelve months, and 13% have "shopped" at Amazon itself in the last 30 days. Unfortunately, we cannot isolate the presumably small sample of people who purchased phones on Amazon.

Among all mobile phone users, Amazon shoppers are younger and wealthier. The Forrester data suggest that 31% of Amazon phone users are under 40, while only 21% of non-Amazon users are under 40. Using a midpoint approximation to a survey question, Amazon households earn $23,000 more in a year than non-Amazon households. Forty-eight percent of Amazon shoppers are male, compared with 46% of non-Amazon mobile phone users. Forty-one percent of Amazon users have children under 18 at home, compared with 36% of non-users. National coverage will obviously be more valued by valued by frequent travelers. Amazon users are more likely to travel than non-users. Thirty-three percent of Amazon users report going on one or more business trips in the past twelve months, compared to 18% of non-users. Similarly, 47% of Amazon users have recently gone on a pleasure trip, compared to 33% of non-users. It is likely that our estimate of the valuation of national coverage is an upper bound, given the characteristics of Amazon users.

4.4 Applicability of discrete choice models to wireless plan choice

In the wireless industry, consumers first purchase a plan and then, usually for two or more years, decide how many minutes to consume each month. Recently, several papers model both plan choice and usage of minutes (Huang 2008; Narayanan et al. 2007). These more detailed models explain the interaction between usage and plan choice.

We lack data on usage. Our approach of allowing for fixed effects that are common to plans with similar number of minutes tries, as best as we can, to control for unobserved heterogeneity in the demand for minutes of calling time. Our focus in this paper is on a parameter, the willingness to pay for national coverage, with relevance to evaluating possible efficiency gains from wireless mergers. This type of question could never be asked with data on usage obtained from a wireless carrier; wireless carriers lack a financial incentive to support investigations into their mergers. While we make no claim that a consumer's planned usage does not interact with the decision to purchase a plan with regional coverage, we do not believe that this interaction between usage and regional coverage is a first order concern. We use publicly available data, from Amazon, on plan choice to estimate a parameter that is necessary to evaluate wireless mergers.

5 Willingness to pay

We now turn to our structural estimates. We use our nest fixed effects to control for the heterogeneous willingness to pay for anytime minutes. For each carrier, we compare a regional plan only to the national plans with slightly more and slightly fewer minutes. For example, Verizon offers national plans with 450, 900, 1350, 2000, 4000 and 6000 anytime minutes a month. Verizon also offers regional plans with 600, 1200 and 1800 minutes. We use the following comparisons: 450 to 600, 600 to 900, 900 to 1200, 1200 to 1350, 1350 to 1800, and 1800 to 2000. We order these seven plans, so that $J_{hm} = 7$ for Verizon. Our maximum score objective function is

$$Q_M(\beta) = \frac{1}{M}\sum_{m=1}^{M}\sum_{h=1}^{H_m}\sum_{j\in J_{hm}\setminus\{l\}}\Big(1\left[r_{jm} > r_{k(j),m}, x_j'\beta - p_j > x_{k(j)}'\beta - p_{k(j)}\right] + 1\left[r_{jm} < r_{k(j),m}, x_j'\beta - p_j < x_{k(j)}'\beta - p_{k(j)}\right]\Big), \quad (10)$$

where $k(j)$ is the next largest (in terms of minutes) plan to plan j in the nest h and l is some largest plan in the nest that appears in only one inequality. We difference out heterogeneity in the demand for anytime minutes to the greatest extent possible.

5.1 National coverage without other controls

Figure 2 shows the plot of the maximum score objective function, (10), when we include only one non-price characteristic: national coverage. The objective function attains its maximum of 103 out of 120 inequalities at a willingness to pay for national coverage of between 0.926 and 0.933 cents per minute. The mean price per minute of the 25 plans from T-Mobile, Verizon and Verizon

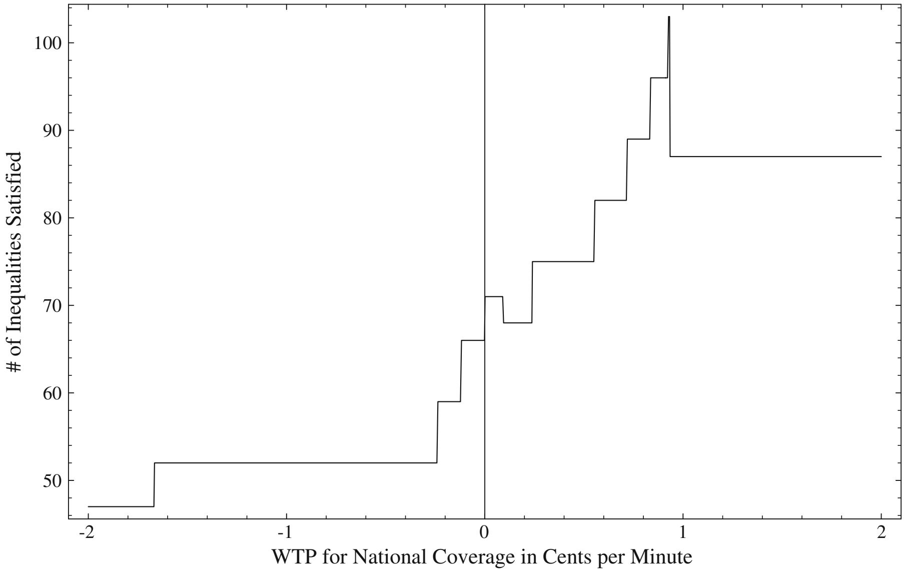

Fig. 2 Maximum score objective function for national coverage only

Family is 5.44 cents per minute. So the willingness to pay for national coverage is 17% of the monthly price per minute of subscription plans.

Table 1 shows that there are seven subscription plans that offer regional coverage. T-Mobile offers one plan and Verizon and Verizon Family each offer three. As our estimator compares only plans from one carrier to plans with the next highest and lowest numbers of minutes, the identification of the parameter for national coverage is generated by variation across plans within the carrier nests. We now describe the contribution of each carrier nest to our maximum score estimator.

T-Mobile offers one regional plan in all 22 markets. The regional plan offers 3000 minutes for $49.99, which comes to 1.67 cents a minute. Our estimator compares this 3000 minute regional plan to a 2500 minute national plan that costs 4.00 cents a minute, and a 5000 minute national plan that costs 2.60 cents per minute. The across-market popularity in decreasing order is the 3000, the 2500 and finally the 5000 minute plan. Given the realized within-market ranks, and ignoring division by the number of markets, the component of our objective function generated by T-Mobile plans is

$$19 \cdot 1\left[\beta^{\text{Nat}} < 0.933\right] + 21 \cdot 1\left[\beta^{\text{Nat}} < 2.333\right] + 1\left[\beta^{\text{Nat}} > 2.333\right] \\ +3 \cdot 1\left[\beta^{\text{Nat}} > 0.933\right]. \tag{11}$$

The number 0.933 is the difference in price per minute between the 3000 and 5000 minute plans, and the number 2.333 is the difference in price per minute between the 2500 and 3000 minute plans. In 19 out of 22 markets, the 3000 minute plan is more popular than the 5000 minute national plan. In 21 markets, the 3000 minute plan is more popular than the 2500 minute plan. Clearly, any value of national coverage less than 0.933 maximizes the T-Mobile component of the objective function by satisfying 40 out of 44 inequalities. One objection might be that 5000 minutes is a lot more than 3000 minutes. A T-Mobile subscriber with a strong need for around 5000 minutes might never consider the 3000 minute plan. In this case, the only valid comparison is the 3000 minute regional plan with the 2500 minute national plan. The regional plan is more popular in 21 out of 22 markets. If attention is restricted to these comparisons, any value of less than 2.33 cents maximizes the objective function.

Verizon offers three non-family regional plans and six non-family national plans, although we use data on only the four national plans comparable minutes to the regional plans. The regional plans are offered only in seven Western cities. The objective function for Verizon is

$$\begin{aligned} &1\left[\beta^{\text{Nat}} < -1.67\right] + 6 \cdot 1\left[\beta^{\text{Nat}} > -1.67\right] + 1\left[\beta^{\text{Nat}} < 0.0\right] + 6 \cdot 1\left[\beta^{\text{Nat}} > 0.0\right] \\ &\quad + 5 \cdot 1\left[\beta^{\text{Nat}} < 0.0927\right] + 2 \cdot 1\left[\beta^{\text{Nat}} > 0.0927\right] + 7 \cdot 1\left[\beta^{\text{Nat}} > 0.926\right] \\ &\quad + 7 \cdot 1\left[\beta^{\text{Nat}} > 0.833\right] + 7 \cdot 1\left[\beta^{\text{Nat}} > 0.555\right]. \end{aligned} \tag{12}$$

The Verizon objective function is maximized by any β^{Nat} greater than 0.926 cents per minute. The number 0.926 is the difference in the cents per minute of the 1350 minute national plan (5.925) and the 1800 minute regional plan (4.999). The 1800 minute regional plan is less popular in six out of seven markets than the 1350 minute national plan, and less expensive per minute, so we can form only a lower bound on β^{Nat}. However, a reasonable person might suspect that the reason the 1800 minute regional plan is not very popular is a 2000 minute national plan with the same price per minute: 4.999. Verizon regional plans tend to have one very near neighbor national plan in terms of anytime minutes. If we only compare each Verizon regional plan to its single nearest neighboring national plan, the objective function for the three regional plan comparisons becomes

$$\begin{aligned} &7 \cdot 1\left[\beta^{\text{Nat}} > 0.555\right] + 5 \cdot 1\left[\beta^{\text{Nat}} < 0.0927\right] + 2 \cdot 1\left[\beta^{\text{Nat}} > 0.0927\right] \\ &\quad + 1\left[\beta^{\text{Nat}} < 0.0\right] + 6 \cdot 1\left[\beta^{\text{Nat}} > 0.0\right]. \end{aligned} \tag{13}$$

By narrowing our set of comparison national plans, our lower bound for β^{Nat} becomes 0.555, which is the price per minute difference between Verizon's very popular 450 minute national plan (8.887 cents) and the somewhat less popular 600 minute regional plan (8.332 cents). In this case, using a more conservative choice of inequalities makes our bounds wider.

Verizon also offers three regional and six national family plans, although we use only the three national plans that compare to regional plans and inform the estimates of β^{Nat}. The objective function is

$$\begin{aligned} &7 \cdot 1\left[\beta^{\text{Nat}} > 2.739\right] + 7 \cdot 1\left[\beta^{\text{Nat}} > 0.715\right] + 7 \cdot 1\left[\beta^{\text{Nat}} > 0.240\right] \\ &\quad + 7 \cdot 1\left[\beta^{\text{Nat}} > -0.118\right] + 7 \cdot 1\left[\beta^{\text{Nat}} > -0.238\right]. \end{aligned} \tag{14}$$

In all cases the regional family plan is less popular than the national family plan. The lower bound for β^{Nat} of 2.739 comes from the largest price difference between two adjacent plans: a 700 minute national plan (8.571 cents) and a 1200 minute regional plan (5.832 cents). A lower bound of 2.739 is inconsistent with the upper bounds of, depending on the comparisons used, of 0.933 and 2.33 from the T-Mobile data. In the pooled sample of all carriers in Fig. 2, the T-Mobile plans dominate the Verizon Family plans as the T-Mobile regional plan is sold in all 22 markets, versus 7 for the Verizon Family regional plans.

However, we might want to restrict comparing the 1200 minute regional plan to its nearer 1400 minute national plan neighbor. By making only the closest possible plan comparisons, our objective function becomes

$$7 \cdot 1\left[\beta^{\text{Nat}} > 0.240\right] + 7 \cdot 1\left[\beta^{\text{Nat}} > -0.118\right] + 7 \cdot 1\left[\beta^{\text{Nat}} > -0.238\right].$$

By using only the most reasonable comparisons, our lower bound for β^{Nat} becomes 0.240, which is the price difference between the 600 minute regional plan (8.332 cents) and the 700 minute national plan (8.571 cents). The number 0.240 is not the tightest bound, as the lower bound from the Verizon individual plans is 0.555.

To conclude, we observe a very popular T-Mobile regional plan. This places an upper bound on the willingness to pay per minute for national coverage. We also observe unpopular Verizon regional plans, which place lower bounds on the willingness to pay for national coverage. We first consider a specification where we compare each regional plan to the national plan with the next fewer anytime minutes and the national plan with the next most anytime minutes. Using the fact that the T-Mobile plans appear in all 22 markets to weight them more than the Verizon Family plans, our bounds are very tight: the willingness to pay for national coverage is between 0.926 and 0.933 cents per minute. If we include only the more conservative comparison of each regional plan to its single most similar national plan neighbor, the willingness to pay for national coverage is bounded between 0.555 and 2.333 cents per minute.

5.2 Controls for other regional plan features

There are characteristics omitted in the above regional plans. First, the 3000 minute T-Mobile regional plan does not offer free calling during nights and weekends, a popular feature for non-business users. The 2500 and 5000 minute T-Mobile national plans that we compare the regional plan to do have unlimited nights and weekends. Second, both the T-Mobile and Verizon regional plans charge customers for making long distance calls. National coverage involves surcharges for placing calls when a user travels; long distance charges are incurred when a user in any region places a call to a number in a distant region. It is not clear why mergers would affect the ability of carriers to offer free long distance. This section controls for these other characteristics.

T-Mobile offers a 1500 minute national plan for \$39.99 that does not offer unlimited calling during nights and weekends. This compares closely to a 1500 minute national plan for \$49.99 that does offer unlimited nights and weekends.

In 13 of our of 22 markets (and 12 out of 13 markets where these plans were very popular), the \$39.99 plan is more popular, meaning an estimate for the willingness to pay for unlimited nights and weekends is any value less than \$10 a month. Unlimited nights and weekends adds extra minutes and so is a substitute rather than a complement for anytime minutes. Therefore, the willingness to pay for unlimited nights and weekends is not a structural constant in terms of cents per minute.

If we apply the $\beta^{\text{NightWeek}} \leq \$10$ willingness to pay to T-Mobile's 3000 minute regional plan, the value per minute for that plan is $0 \leq \beta^{\text{NightWeek}}/30 < 0.333$ cents per minute, where 30 is 3000 minutes per month divided by 100 cents per dollar. Considering that the T-Mobile national plans we compared the regional plan to had unlimited nights and weekends, our previous upper bound becomes $\beta^{\text{Nat}} + \beta^{\text{NightWeek}}/30 < 0.933$. The upper bound on $\beta^{\text{NightWeek}}/30$ does not lower the marginal upper bound on β^{Nat}, which is still 0.933. The theoretical lower bound of 0 for $\beta^{\text{NightWeek}}/30$ does not lower the upper bound on β^{Nat}, either. After examining unlimited nights and weekends, our upper bound is still $\beta^{\text{Nat}} < 0.933$. Likewise, our conservative upper bound remains 2.33. Note that our estimate of $\beta^{\text{NightWeek}} < 0.333 \cdot 30 = \$10$ a month is an univariate analysis from the two 1500 minute plans. Our full multivariate analysis below tightens this upper bound for $\beta^{\text{NightWeek}}$.

T-Mobile and Verizon's regional plans charge for long distance, in addition to deducting the call length from the standard allotment of minutes. However, Verizon does allow customers to add free domestic long distance to any regional plan for an extra \$5. Theoretically, free long distance should raise the value of an anytime minute by β^{FreeLong}. For whatever reason, Verizon charges a fixed fee. To more accurately find the lower bound on national coverage, we will assume that all Verizon users pay the extra \$5 a month fee. For example, a user who buys the 1200 minute regional plan faces an extra charge of $500/1200 = 0.417$ cents per minute.

Verizon's maximum score objective function when we compare each regional plan to the national plans with the next fewer and next most number of minutes, Eq. (12), is

$$\begin{aligned}
&1\left[\beta^{\text{Nat}} < -2.50\right] + 6 \cdot 1\left[\beta^{\text{Nat}} > -2.50\right] + 1\left[\beta^{\text{Nat}} < -0.278\right] \\
&\quad + 6 \cdot 1\left[\beta^{\text{Nat}} > 0.278\right] + 5 \cdot 1\left[\beta^{\text{Nat}} < -0.324\right] + 2 \cdot 1\left[\beta^{\text{Nat}} > -0.324\right] \\
&\quad + 7 \cdot 1\left[\beta^{\text{Nat}} > 0.648\right] + 7 \cdot 1\left[\beta^{\text{Nat}} > 0.416\right] + 7 \cdot 1\left[\beta^{\text{Nat}} > -0.278\right].
\end{aligned}$$

This objective function is maximized by $\beta^{\text{Nat}} > 0.648$. Verizon's conservative maximum score objective function in (13) becomes

$$\begin{aligned}
&7 \cdot 1\left[\beta^{\text{Nat}} > -0.2783\right] + 5 \cdot 1\left[\beta^{\text{Nat}} < -0.324\right] + 2 \cdot 1\left[\beta^{\text{Nat}} > -0.324\right] \\
&\quad + 1\left[\beta^{\text{Nat}} < -0.277\right] + 6 \cdot 1\left[\beta^{\text{Nat}} > -0.277\right].
\end{aligned}$$

Unfortunately, maximizing this objective function says that $\beta^{\text{Nat}} > -0.277$. If we are conservative in our choice of inequalities, our plan characteristics are not rich enough to distinguish between free long distance and national coverage. National coverage is a positive characteristic, so we set β^{Nat}'s lower bound to 0. The conservative lower bound is uninformative because Verizon's regional plans are an even worse financial deal if one considers the extra \$5 fee for free long distance.

T-Mobile does not allow regional plan users to upgrade to free long distance. For lack of a better solution, we assign Verizon's \$5 fee to T-Mobile 3000 minute regional plan. The per-minute cost is 0.167 cents per minute. Therefore, our tight upper bound for β^{Nat} decreases from 0.933 to 0.766 and our conservative upper bound decreases from 2.333 to 2.166. Applying the \$5 fee to the T-Mobile regional plan is a conservative approach: if a consumer with the 3000 minute regional plan valued free long distance more than \$5 a month, the upper bounds would decrease by more and become tighter. In conclusion, our tight bound for β^{Nat} is 0.648 to 0.766 cents per minute. Our conservative bound for β^{Nat} is 0 to 2.166 cents per minute.

Figure 3 shows the final maximum score objective function for pairwise comparisons, evaluated at the \$5 Verizon fee and the theoretical lower bound $\beta^{\text{NightWeek}} = 0$. Our tight bound for β^{Nat} is 0.648 to 0.766 cents per minute is clearly visible in the picture.

The comparison between the 1500 minute T-Mobile plans is not informative for the lower bound for the willingness to pay for unlimited off peak calling, so we set the lower bound's value to the theoretical constraint of 0. Previously, our univariate analysis of the two 1500 minute plans yielded an upper bound for the willingness to pay of unlimited off peak calling of $\beta^{\text{FreeNight}} = \$10$ a

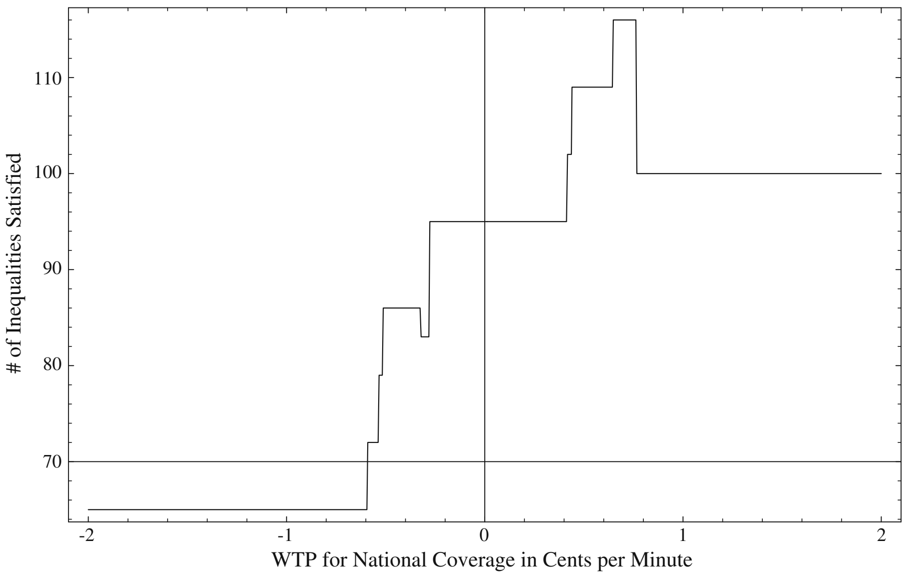

Fig. 3 Maximum score objective function for national coverage evaluated at \$5 long distance fee and $\beta^{\text{NightWeek}} = 0$

month, the difference in price between the two plans. Our specification with both the WTPs for national coverage and unlimited nights and weekends is multivariate. The value of unlimited nights and weekends enters the objective function for the T-Mobile 3000 minute regional plan as well. The upper bound on the value of unlimited nights and weekends comes from evaluating the maximum score objective function at the lower bound for the WTP for national coverage of $\beta^{\text{Nat}} = 0.648$. The objective function is

$$\begin{aligned} 19 \cdot 1 \left[\beta^{\text{FreeNight}}/30 < 0.119\right] + 21 \cdot 1 \left[\beta^{\text{FreeNight}}/30 < 1.52\right] \\ + 1 \left[\beta^{\text{FreeNight}}/30 > 1.52\right] + 3 \cdot 1 \left[\beta^{\text{FreeNight}}/30 > 0.119\right] \\ + 9 \cdot 1 \left[\beta^{\text{FreeNight}}/15 > 0.667\right] + 13 \cdot 1 \left[\beta^{\text{FreeNight}}/15 < 0.667\right], \end{aligned}$$

where all the plans are from T-Mobile, and the comparisons on the first row are between the regional and the two nearest national plans and the second row is the comparison of the two 1500 minute plans. The numbers 15 and 30 are 1500 and 3000 minutes per month divided by 100 cents per dollar, which converts the monthly value into per minute values for each plan. The objective function is maximized at 53 inequalities at $\beta^{\text{FreeNight}}/30 < 0.119$, or $\beta^{\text{FreeNight}} < \$3.57$ a month.

5.3 Subsampled confidence intervals

Section 3.8 discusses how we use the Romano and Shaikh (2008) procedure to estimate 95% confidence regions for each parameter. We now explain how we implement this procedure in our dataset. Like how a researcher must choose a bandwidth parameter in density estimation, we must choose a subsample size. As with bandwidth choices, there is not a developed theory for choosing the subsample size. In our past empirical experience, we have found that using a quarter of the sample produces intuitively plausible results. Therefore, we subsample by drawing fake datasets composed of subsets of 5 of our 22 markets. The other choice that must be made is the number of subsampled datasets to use. Unlike the number of markets per dataset, more datasets gives a better approximation to the limiting distribution than fewer datasets. We use 100 datasets.[19] Choosing the number of markets per subsample and the number of subsamples are the main qualitative judgements that must be made. In our application, subsampling allows for market-specific taste shocks ξ_{jm} and allows for sampling from a small number I_m of customers purchasing calling plans on Amazon. To the extent that market share ranks are similar across markets, our confidence sets will approximate the estimated bounds above.

We incorporate Section 5.1's comparisons of each regional plan to two national plans and Section 5.2's comparison of the 1500 minute T-Mobile plan without unlimited night and weekends to the 1500 minute plan with unlimited

[19]The software toolkit Santiago and Fox (2007) implements subsampling and is available on the internet.

Table 2 Set estimates and confidence regions for WTPs for univariate and multivariate specifications

Variable	Lower	Upper	Length
Set estimates			
WTP for national (cents/minute)	0.648	0.767	0.119
WTP for national (cents/minute)	0.648	0.767	0.119
WTP for unlimited off peak ($/month)	0	3.57	3.57
95% confidence region			
WTP for national (cents/minute)	0.648	0.767	0.119
WTP for national (cents/minute)	0.648	0.767	0.119
WTP for unlimited off peak ($/month)	0	3.57	3.57

nights and weekends. As before, β^{Nat} enters the value of an anytime minute, while $\beta^{\text{NightWeek}}$ is a monthly value, as unlimited nights and weekends are a substitute for anytime minutes. We apply the $5 Verizon fee for free long distance calling to all regional plans.

Table 2 reports set estimates as well as 95% confidence regions from the subsampling approach discussed in Section 3.8.[20] For each method, we report two specifications: one with only national coverage and one controlling for unlimited off peak calling. We see that the 95% confidence region for the willingness to pay for national coverage has the same tight bounds as reported before: 0.648 to 0.767 cents per minute.

The 95% confidence region is the same as the set estimate because of the strong similarity in market share ranks across our sample of 22 markets: the estimates using a subset of the markets are typically the same as for the entire sample. For example, look at (11), the objective function for T-Mobile that gives national coverage an upper bound, before adjusting for unlimited off peak calling, of 0.933 cents per minute. Notice how the regional plan is more popular in 19 out of 22 markets. As our subsample size is 5, the only way a different lower bound could occur is if all 3 of the markets where the regional plan is less popular appear in the subsample. This way, the 3 markets where the regional plan is less popular would outrank the 2 included markets where the regional plan is more popular. There are $\binom{22}{5} = 26,334$ possible subsamples but only $\binom{22-3}{2} = 171$ of them contain all 3 of the markets where the regional plan is less popular. The 171 subsamples are $171/26,334 = 0.6\%$ of the possible subsamples, and so do not affect the 0.025 and 0.975 quantiles of the subsampled distribution of the objective function.[21]

[20] A referee reports out that the natural generalization of a confidence set for a point estimate to set estimation is a confidence collection (a set of sets). It is hard to visually describe a set of sets. Given that the natural measure is the (hard to compute) set of sets, the marginal benefit of reporting the set estimates in addition to the confidence sets may be small. We report both in Table 2 only to drive home their similarity.

[21] A similar argument can be used to show that the confidence sets will be the same as the set estimates for other subsample sizes.

5.4 Sensitivity of WTP to assumptions about usage

A common impression of many mobile phone users is that they may contract for more minutes than they use each month. Huang (2008) uses aggregate data on carrier (but not individual plan) market shares and total carrier minute usage to estimate a joint model of plan choice and minute usage. It is less clear where identification of usage comes from without market share data or minute usage for individual plans. Our price measure, (2), assumes that consumers use all of the minutes in their plan.[22] Our assumption on usage is stronger, but identification of the WTP for plan features such as β^{Nat} in our semiparametric model is clearer given our dependent variable is plan purchase and not minute usage.

We can show how our estimates change under alternative assumptions about usage. For example, the T-Mobile upper bound on β^{Nat} in (11) comes from comparing a 1.67 cents a minute ($49.99), 3000 minute regional plan to a 2.60 cents per minute ($129.99), 5000 minute national plan. Say instead we assume that a consumer will at most use 3000 minutes of talk time. Then the effective price per minute of the national plan is $\$129.99/3000 = 4.3$ cents. $4.3 - 1.67 = 2.63$, which is actually greater than the 2.33 cents for the comparison of the 3000 minute regional plan with the 2500 minute national plan ($99.99) in (11). Therefore, the bound from (11) with the adjustment is $\beta^{\text{Nat}} < 2.33$ (2.16 after the $5 long distance charge), equal to our earlier conservative upper bound. For the 2500 to 3000 comparison, if we (in addition to our earlier adjustment) assume that the user consumes only 2000 minutes, then we get $\beta^{\text{Nat}} < 0.05 - 0.025 = 2.5$ cents, or 2.3 cents after the $5 long distance charge. As before, reasonable changes to our usage assumptions yield upper bounds nearly equivalent to our earlier conservative upper bound.

Verizon provides our lower bounds on β^{Nat}. Our binding comparison in (12) compares a 1350 minute national plan at 5.93 cents per minute to a 1800 minute regional plan at 5.00 cents per minute. If we assumed instead that a consumer was only going to talk for say 1200 minutes, then $\beta^{\text{Nat}} > 0.067 - 0.075 = -1$ cent by this comparison alone. The full maximum score estimate from (12) after this change would be $\beta^{\text{Nat}} > 0.833$, which is driven by a comparison of two other plans: a 1200 minute, $69.99 regional plan and a 900 minute, $59.99 national plan. Assuming, say, that a user only wants to speak 800 minutes will again make the lower bound, for this comparison only, drop below 0 (the regional plan becomes more expensive in price per minute). By decreasing the minutes consumed for all the pairwise comparisons in (12), the lower bound on β^{Nat} can continue to be dropped. Similar results hold for Verizon Family plans.

The lower bound on β^{Nat} in a sense is our main result. It is sensitive to the assumption that those purchasing Verizon's regional plans actually use most of

[22] We did not collect data on the overage charge, the per-minute fee for making calls that exceed the monthly bucket of minutes. Therefore, we consider only one aspect of the price.

the relatively large number of minutes that come with those plans. Otherwise, regional plans have no price advantage over national plans, and so there is no reason (other than demand shocks) for any consumer to buy them. If regional plans are dominated on every dimension, consumers are not facing any meaningful trade-off, and quantity data is not informative about the value of β^{Nat}. The fact that some consumers buy these plans suggests some consumers use the extra minutes. Therefore, we believe our earlier assumption of complete usage of the monthly minutes is better than an alternative assumption that makes the regional plans dominated in all dimensions.

5.5 Across-market heterogeneity in WTP for national coverage

Section 3.7.2 discusses why the multinomial maximum score estimator, and hence our approach to dealing with market share rank data, is incompatible with identification of a distribution random coefficients for consumers in the same market. Indeed, we have studied identification of the distribution of random coefficients in other work (Bajari et al. 2007), and our proof requires across-market variation in product characteristics. As we have argued, mobile phone carriers choose plans on a national level and across-market variation in characteristics is not available. Therefore, any attempt to identify unobserved heterogeneity using aggregate data in this industry seems ambitious, at best.

We can discuss how if we used data on each market as a separate dataset, our set estimate of β_m^{Nat} would vary across markets m. The answer is not much.[23] For T-Mobile, (11) shows only in 4 out of 40 market / plan comparisons is the regional plan less popular than a comparison national plan. These four exceptions are: the 3000 regional plan has a slightly lower rank than the two national plans in Atlanta and the 3000 minute regional plan has a slightly lower rank than the 5000 minute national plan in Pittsburgh and Washington, DC. Therefore, there is no upper bound for β_m^{Nat} in Pittsburgh and Washington and in Atlanta the T-Mobile data would provide a lower bound of 2.33 cents per minute for β_m^{Nat}, suggesting a huge value for national coverage in Atlanta, as even the attractively priced T-Mobile regional plan is not popular.

For Verizon in (12), the 4 out of 42 cases where the regional plan is more popular than a similar national plan are: the 600 minute regional plan is more popular than the 900 minute national plan in Denver, and the 1200 minute regional plan is slightly more popular than the 1350 national plan in Los Angeles, Phoenix and Portland. Neither of these comparisons drive the lower bound for β_m^{Nat} in (12), so the lower bound estimate for β_m^{Nat} is the same across all markets. For the Verizon Family comparisons in (14), in all 35 market / plan comparisons the regional plan is less popular.

[23] We cannot compute the presumably huge standard errors for an estimate using data on only market.

In terms of our econometric theory, we can interact observed market heterogeneity with the WTP for national coverage. However, there is so little variation in relative market share ranks of regional and national plans across markets that any attempt to relate WTP for national coverage will produce a near-zero coefficient on the interaction term. Interestingly, the popularity of individual carriers varies quite a lot across markets, perhaps reflecting path dependence in shares or variations in the quality of service. This emphasizes the need to only compare plans from the same carrier with each other, which allows for carrier and market specific fixed effects.

5.6 Interpreting the WTP for national coverage

As our conservative lower bound is uninformative, we will focus on our tight bound, which comes from comparing each regional plan to two national plans with similar numbers of minutes from the same carrier. An industry trade group, the CTIA–The Wireless Association, estimates there were more than 1.4 trillion wireless minutes used in the United States in 2005. Multiplying 1.4 trillion minutes by the tight lower bound for the willingness to pay for national coverage of 0.648 cents per minute yields a national willingness to pay of $9.1 billion. The CTIA reports the 2005 subscriber revenue of US carriers from their 208 million subscribers is $113.5 billion. $9.1 billion is therefore 8% of annual subscriber revenue. The $9.1 billion is a lower bound; our tight upper bound of 0.766 cents per minute comes to $10.7 billion annually, or 9.4% of the $113.5 in annual industry revenue.

On an individual basis, the most popular plans in our data are two 450 minute plans that each have a monthly fee of $40. For this type of plan, at our lower bound the monthly willingness to pay for national coverage is $0.648/100 \cdot 450$, or $2.92. Our lower bound for the willingness to pay for national coverage is equivalent to 7.2% of the price of a plan.

We interpret our willingness to pay estimates in terms of the actual plan details. Verizon regional plans charge travelers 69 cents a minute. At the lower bound, a consumer will be indifferent to paying a roaming fee of 69 cents a minute and buying a national plan that costs an extra 0.648 cents a minute if 0.9% of the consumer's minutes are for calls made while traveling. This low percentage is consistent with the overall pattern that national plans are much more popular than regional plans.

Our estimate of an annual willingness to pay of $9.1–10.7 billion may overstate the benefits of national coverage. Section 4.3 examined data from a survey of consumers and found that Amazon users take more trips than non-Amazon mobile phone users. Amazon users may value national coverage more. Second, our data contain customers who purchased only postpaid / monthly contracts. We do not observe customers using prepaid plans, some of which have national coverage, and some of which do not. As many prepaid customers have poor credit or use their phones infrequently, they may value national coverage less than those in our sample.

6 Conclusions

There has been a tremendous amount of consolidation in the wireless service industry. In 1988, the company serving the largest number of the top 20 markets was US West, which served only four markets. In 2007, four national carriers dominate the market. One motivation for consolidation is to offer customers seamless national coverage areas, where features such as data and voicemail will work without interruptions. The first national calling plan was initiated by AT&T Wireless, and it came only after that carrier achieved somewhat of a national calling scale.

This paper proposed semiparametric demand estimators for market share rank data to estimate the willingness to pay for national coverage. Three plan nests, T-Mobile, Verizon and Verizon family, offer both national and regional plans in the same geographic markets. We formally estimated the willingness to pay for national coverage, and found estimates between 0.648 and 0.766 cents per minute. The lower bound corresponds to 7.2% of the monthly bill of a customer with a popular 450 minute plan.

We extrapolated our bounds from the Amazon sample to the entire US population of 208 million mobile phone customers. We found that the annual consumer value from national calling plans is between $9.1–10.7 billion. These bounds are 8–9.4% of the industry's annual revenue of $113.5 billion.

We interpret our results as showing that national coverage is highly valued by consumers. To the extent that across-market mergers are necessary to provide national calling areas, the evidence from customer behavior on Amazon supports the view that across-market mergers are efficiency enhancing. Inter-firm roaming agreements are another alternative to mergers. As we discussed in Section 2, evidence from past industry data and current industry news reports suggests roaming agreements typically involve high per-minute transfers between the home carrier of the traveler and the carrier providing the coverage.

Our high estimates of the willingness to pay for national coverage are consistent with the behavior of firms in our industry. Just before our data collection began, Cingular discontinued its regional plans. As we write, large national carriers offer mainly national plans and do not price discriminate against travelers within their native calling areas.

Our econometric contribution is to extend semiparametric demand estimation to the use of market-level data. In particular, we show how to estimate a willingness to pay using only market share ranks. We hope our estimator will encourage others to use online retailers such as Amazon as an easily accessible source of data for demand estimation. We acknowledge that the lack of quantity data requires potentially stronger assumptions for identification. However, data from online retailers have potential advantages. In online data, the economist observes the exact information about the product presented to the consumer by retailers. Online data are also freely available for product categories, such as calling plans, where other high quality data sources may be lacking.

Acknowledgements Bajari thanks the National Science Foundation for generous research support. Fox thanks the National Science Foundation, the NET Institute, the Olin Foundation, and the Stigler Center for generous funding. Thanks to helpful comments from seminar participants at the International Industrial Organization Conference (IIOC), the Northwestern and University of Chicago IO-Marketing conference, and the STIET seminar at the University of Michigan. We thank Austan Goolsbee, Ali Hortacsu and Chad Syverson for providing access to the Forrester Data, and for helpful comments. We thank Dennis Carlton, Victor Chernozhukov, Anindya Ghose, Aviv Nevo, Amil Petrin, Bill Rogerson and Azeem Shaikh for detailed conversations. We thank anonymous referees for comments.

A Proofs

A.1 Lemma 1

In what follows, drop the indices i and m, and use the shorthand notation a_j for $x'_{jm}\beta - p_{jm} + v^h_{im}$. Also replace the conditioning arguments J_m, H_m, X_m, $\vec{p}_m$, I_m, $\{\vec{v}_{im}\}_{i\in I_{m,}}$ of the densities of ε_j and ξ_j with $\vec{a}$, the vector of the J a_j's.

Both ε_j and ξ_j are not in the data. The joint density of all J ε_J's and ξ_j's is

$$c\left(\vec{\varepsilon},\vec{\xi}\mid\vec{a}\right) = \prod_{h\in H}\prod_{k\in H} f_h\left(\varepsilon_j\mid\vec{a},\vec{\xi}\right)\cdot\prod_{h\in H}\prod_{k\in H} g_h\left(\xi_k\mid\vec{a}\right).$$

First we will integrate out ε. Let $\Pr\left(j\mid\vec{a},\vec{\xi}\right)$ be the probability of picking j conditional on the realization of the ξ_j's. The decision rule in Eq. (3) becomes

$$\varepsilon_l < a_j - a_l + \varepsilon_j + \xi_j - \xi_l$$

for all choices l.

First prove the "only if" direction: If $a_j+\xi_j > a_k+\xi_k$, then $\Pr\left(j\mid\vec{a},\vec{\xi}\right) > \Pr\left(k\mid\vec{a},\vec{\xi}\right)$. By the definition of a choice probability,

$$\Pr\left(j\mid\vec{a},\vec{\xi}\right) = \int_{-\infty}^{\infty}\left(\left\{\int_{-\infty}^{a_j-a_k+\varepsilon_j+\xi_j-\xi_k} f_{h(k)}\left(\varepsilon_k\mid\vec{a},\vec{\xi}\right)d\varepsilon_k\right\}\right.$$
$$\left.\cdot\prod_{l=1,l\neq j,k}^{J}\left\{\int_{-\infty}^{a_j-a_l+\varepsilon_j+\xi_j-\xi_l} f_{h(l)}\left(\varepsilon_l\mid\vec{a},\vec{\xi}\right)d\varepsilon_l\right\}\right)$$
$$\cdot f_{h(j)}\left(\varepsilon_j\mid\vec{a},\vec{\xi}\right)d\varepsilon_j,$$

where $h(l)$ is a convenience function returning the nest of choice l. If $h(j) = h(k)$, as in the statement of the lemma, $\Pr\left(j\mid\vec{a},\vec{\xi}\right)$ is the same function as $\Pr\left(k\mid\vec{a},\vec{\xi}\right)$, except that $a_k+\xi_k$ replaces $a_j+\xi_j$ in the upper limits, and $a_j+\xi_j$ replaces the one term where $a_k+\xi_k$ enters in $\Pr\left(j\mid\vec{a},\vec{\xi}\right)$. Let $W\left(a_j+\xi_j, a_k+\xi_k\right)$ be $\Pr\left(j\mid\vec{a},\vec{\xi}\right)$ as a function of $a_j+\xi_j$ and $a_k+\xi_k$.

As $a_j+\xi_j$ enters only upper limits of integrals in $W\left(a_j+\xi_j, a_k+\xi_k\right)$, $W\left(a_j+\xi_j, a_k+\xi_k\right)$ is increasing in $a_j+\xi_j$. Also, $a_k+\xi_k$ enters negatively in only one upper limit in $W\left(a_j+\xi_j, a_k+\xi_k\right)$. Because $f_{h(j)}\left(\varepsilon_j \mid \vec{a}, \vec{\xi}\right)$ has full support by Assumption 1, $W\left(a_j+\xi_j, a_k+\xi_k\right)$ is strictly increasing in $a_j+\xi_k$ and strictly decreasing in $a_k+\xi_k$. Then if $a_j+\xi_k > a_k+\xi_k$, as in the statement of the lemma, $W\left(a_j+\xi_j, a_k+\xi_k\right) > W\left(a_k+\xi_k, a_j+\xi_j\right)$. Likewise, the "if" direction is proved as the only way $W\left(a_j+\xi_j, a_k+\xi_k\right) > W\left(a_k+\xi_k, a_j+\xi_j\right)$ is if $a_j+\xi_j > a_k+\xi_k$.

The above argument conditioned on $\vec{\xi}$. We need to prove statements about $\Pr(j \mid \vec{a})$ and $\Pr(k \mid \vec{a})$. Again by the definition of a choice probability,

$$\Pr\left(j \mid \vec{a}, \vec{\xi}\right) = \int_{-\infty}^{\infty} \Pr\left(j \mid \vec{a}, \vec{\xi}\right) \prod_{h\in H} \prod_{k\in H} g_h\left(\xi_k \mid \vec{a}\right) d\vec{\xi}.$$

Because Assumption 1 states each $f_{h(j)}\left(\varepsilon_j \mid \vec{a}, \vec{\xi}\right)$ is exchangeable in the arguments ξ_j and ξ_k when $h(j) = h(k)$ and the ξ_{jm}'s are i.i.d. within a nest, then $\Pr(j \mid \vec{a})$ is the same function as $\Pr(k \mid \vec{a})$, except where a_j and a_k enter the upper limits in $\Pr\left(j \mid \vec{a}, \vec{\xi}\right)$. By a similar argument as with the W function above, the lemma is proved.

A.2 Lemma 2

Because expectation and integration are linear operators, conditioning on the number of consumers I_m results in:

$$\begin{aligned}
&E\left[s_{jm} \mid I_m, J_m, H_m, X_m, \vec{p}_m\right] \\
&\quad = E\left[\frac{1}{I_m}\sum_{i=1}^{I_m} 1\left[i\,\text{buys}\, j\right] \mid I_m, J_m, H_m, X_m, \vec{p}_m\right] \\
&\quad = \frac{1}{I_m}\sum_{i=1}^{I_m} E\left[1\left[i\,\text{buys}\, j\right] \mid I_m, J_m, H_m, X_m, \vec{p}_m\right] \\
&\quad = \frac{1}{I_m}\sum_{i=1}^{I_m} E\left[E_{\varepsilon,\xi}\left[1\left[i\,\text{buys}\, j\right] \mid \{\vec{\nu}_{im}\}_{i\in I_{m,}}, I_m, J_m, H_m, X_m, \vec{p}_m\right]\right. \\
&\qquad\qquad \left. \mid I_m, J_m, H_m, X_m, \vec{p}_m\right] \\
&\quad = \frac{1}{I_m}\sum_{i=1}^{I_m} E\left[\Pr_{im}\left(j \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\right) \mid I_m, J_m, H_m, X_m, \vec{p}_m\right],
\end{aligned}$$

where the second-to-last equality is from the law of iterated expectations and the last equality uses the definition of a choice probability that integrates out consumer product specific error terms of the form ε_{ijm} and product specific error terms of the form ξ_{jm}.

First consider the "if" direction. Consider two products j and k in the same nest h, and let $x'_{jm}\beta - p_{jm} > x'_{km}\beta - p_{km}$. Under Assumption 1, Lemma 1 states that

$$\begin{aligned}&\Pr_{im}\left(j \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\right)\\&\quad > \Pr_{im}\left(k \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\right)\end{aligned}$$

for all consumers. By the above market share algebra,

$$\begin{aligned}&E\left[s_{jm} \mid I_m, J_m, H_m, X_m, \vec{p}_m\right]\\&\quad = \frac{1}{I_m}\sum_{i=1}^{I_m} E\left[\Pr_{im}\left(j \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\right) \mid I_m, J_m, H_m, X_m, \vec{p}_m\right]\\&\quad > \frac{1}{I_m}\sum_{i=1}^{I_m} E\left[\Pr_{im}\left(k \mid J_m, H_m, X_m, \vec{p}_m, I_m, \{\vec{\nu}_{im}\}_{i\in I_{m,}}\right) \mid I_m, J_m, H_m, X_m, \vec{p}_m\right]\\&\quad = E\left[s_{km} \mid I_m, J_m, H_m, X_m, \vec{p}_m\right],\end{aligned}$$

as by Lemma 1 each consumer chooses j more often than k. As $E\left[s_{jm} \mid I_m, J_m, H_m, X_m, \vec{p}_m\right] > E\left[s_{km} \mid I_m, J_m, H_m, X_m, \vec{p}_m\right]$ for any number of consumers I_m, $E\left[s_{jm} \mid J_m, H_m, X_m, \vec{p}_m\right] > E\left[s_{km} \mid J_m, H_m, X_m, \vec{p}_m\right]$ unconditional on the unobserved (in our data) number of consumers I_m.

The "only if" direction just reverses these documents, as the only way the sum of choice j's probabilities can be greater than choice i's under Lemma 1 is when $x'_{jm}\beta - p_{jm} > x'_{km}\beta - p_{km}$.

A.3 Lemma 3

Drop the m index for simplicity. We are comparing products j and k, which are in the same nest. The rank of product j with a finite sample of I customers is $\hat{r}_j$. The rank orders the (unobserved) market shares s_j. The condition that $\hat{r}_j > \hat{r}_k$ can be rewritten as $\hat{s}_j > \hat{s}_k$. Dividing by a positive number $\hat{s}_j + \hat{s}_k$, the inequality becomes

$$\frac{\hat{s}_j}{\hat{s}_j + \hat{s}_k} > \frac{\hat{s}_k}{\hat{s}_j + \hat{s}_k}.$$

Define $\tilde{s}$ to be $\frac{\hat{s}_j}{\hat{s}_j+\hat{s}_k}$, leaving $1 - \tilde{s}$ to be $\frac{\hat{s}_k}{\hat{s}_j+\hat{s}_k}$. We want to show that $P\left(\tilde{s} > \frac{1}{2}\right) > P\left(\tilde{s} < \frac{1}{2}\right)$.

First consider the case without product market error terms ξ_j. We also condition on I_{jk}, the number of people who buy either j or k. An individual i prefers j over k with probability q_i. By Lemma 1, $q_i > \frac{1}{2}$. However, because the density of errors and the realization of fixed effects vary across consumers, q_i will be different for each consumer. We work with the number (rather then the fraction) of consumers who buy j out of the group who buy either j or k. Call this number r_j. The random variable $r_k = I_{jk} - r_j$ is the number of consumers who pick k over j. Let $r^\star = \frac{I_{jk}}{2}$. We want to show $\Pr\left(r_j > r^\star\right) > \Pr\left(r_k > r^\star\right)$.

Now, if $q_i = \frac{1}{2}$ for all i, $\Pr\left(r_j > r^\star\right) = \Pr\left(r_k > r^\star\right)$ by the properties of the binomial distribution. By a monotonicity arguments, increasing even one q_i will raise $\Pr\left(r_j > r^\star\right)$ and consequently weakly lower $\Pr\left(r_k > r^\star\right)$, as for odd I_{jk} $\Pr\left(r_j > r^\star\right) + \Pr\left(r_k > r^\star\right) = 1$ and for even I_{jk} $\Pr\left(r_j > r^\star\right) + \Pr\left(r_k > r^\star\right) = 1 - \Pr\left(r_k = r_j\right)$. So $\Pr\left(r_j > r^\star\right) > \Pr\left(r_k > r^\star\right)$, and the "only if" direction of the lemma is proved. The "if" direction just reverses the above steps, as the only way one product is ranked higher than another more frequently is when the first product has a higher payoff.

The above argument conditioned on a value of I_{jk}. As the lemma holds for any value of I_{jk}, it holds unconditionally as well.

Now consider the case with both ξ_j and ε_{ij} errors. For each realization of ξ_j, each consumer has a q_i that involves the remaining uncertainty over the ε_{ij} terms. Even if j has a higher mean payoff than k, it could be that $q_i < \frac{1}{2}$ because of the realization of ξ_j and ξ_k. The probability q_i is the probability of picking j over k given that the payoff of j is $x_j'\beta - p_j + \xi_j + \nu_{ih}$ and the payoff of k is $x_k'\beta - p_k + \xi_k + \nu_{ih}$. By Lemma 1, $q_i > \frac{1}{2}$ when $x_j'\beta - p_j + \xi_j + \nu_{ih} > x_k'\beta - p_k + \xi_k + \nu_{ih}$. In other words, either the realization of product and market specific shocks is such that $q_i > \frac{1}{2}$ for everyone or $q_i \le \frac{1}{2}$ for everyone. So the lemma being proved holds if $x_j'\beta - p_j + \xi_j > x_k'\beta - p_k + \xi_k$ more than half of the time when $x_j'\beta - p_j > x_k'\beta - p_k$, which it does because Assumption 1 states that the ξ's are i.i.d.

A.4 Lemma 4

Our identification under sampling error argument will show that the probability limit of $Q_M(\beta)$ is uniquely maximized by the parameter vectors in the identified set B^0, which is defined in the statement of the lemma. Use the notation r_j for the underlying random variable for market share ranks. If $M \to \infty$, the maximum score objective function converges to the population objective function

$$Q_\infty(\beta) = \sum_{h=1}^{H} \sum_{j,k \in J_h, k \neq j} 1\left[x_j'\beta - p_j > x_k'\beta - p_k\right] \cdot E\left\{1\left[r_j > r_k\right]\right\}, \tag{15}$$

where we have factored the fixed-across-markets product characteristics out of the expectation over the preferences of customers and the number of such customers in each market m. The probability limit can be rewritten to focus on unique pairs of products as

$$Q_\infty(\beta) = \sum_{h=1}^{H} \sum_{j,k \in J_h, k > j} \left\{1\left[x_j'\beta - p_j > x_k'\beta - p_k\right] \cdot E\left\{1\left[r_j > r_k\right]\right\} \right.$$
$$\left. +1\left[x_k'\beta - p_k > x_j'\beta - p_j\right] \cdot E\left\{1\left[r_k > r_j\right]\right\}\right\}.$$

For each pair of products, the objective function is the sum of two probabilities times mutually exclusive inequalities. $Q_\infty(\beta)$ is maximized if the inequality multiplying the (weakly) greater of $E\{1[r_j > r_k]\}$ and $E\{1[r_k > r_j]\}$ is set to 1. Lemma 3 shows that $E\{1[r_j > r_k]\}$ is larger than $E\{1[r_k > r_j]\}$ precisely when $x_j'\beta^0 - p_j$ is greater than $x_k'\beta^0 - p_k$. Therefore, $Q_\infty(\beta)$ is maximized for any $\beta \in B^0$, the identified set. Clearly $\beta^0 \in B^0$, the identified set.

For identification under sampling error, we also need to show that parameter vectors that are not part of the identified set do not maximize the objective function. Equivalently, we need to prove that if some β maximizes $Q_\infty(\beta)$ then $\beta \in B^0$. If β maximizes $Q_\infty(\beta)$, the larger of $E\{1[r_j > r_k]\}$ and $E\{1[r_k > r_j]\}$ enters the objective function for each pair of choices. That term multiplies one of the mutually exclusive indicator functions in β, so this β must maximize the non-sampling error objective function, (6). So by the definition of B^0 and Lemma 3, $\beta \in B^0$ and the identified set comprises the maximizers of the probability limit of the objective function under sampling error.

References

Abrevaya, J. (2000). Rank estimation of a generalized fixed-effects model. *Journal of Econometrics, 95*, 1–23.

Andrews, D. W. K., Berry, S., & Jia, P. (2005). Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location. Yale University working paper.

Bajari, P., Fox, J. T., Kim, K. il., & Ryan, S. (2007). A simple nonparametric estimator for the distribution of random coefficients in discrete choice models. Working paper, October.

Beresteanu, A., & Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica, 76*(4), July.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica, 63*(4), 841–90.

Briesch, R. A., Chintagunta, P. C., & Matzkin, R. L. (2002). Semiparametric estimation of brand choice behavior. *Journal of the American Statistical Association, 97*(460), 973–982, December.

Brynjolfsson, E., Hu, Y., & Smith, M. D. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science, 49*(11), 1580–1596, November.

Chernozhukov, V., Hong, H., & Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica, 75*(5), 1243–1284, September.

Chevalier, J., & Goolsbee, A. (2003). Price competition online: Amazon versus barnes and noble. *Quantitative Marketing and Economics, 1*(2), 203–222, June.

Fox, J. T. (2005). Consolidation in the wireless phone industry. NET Institute working paper 05–13.

Fox, J. T. (2007). Semiparametric estimation of multinomial discrete choice models using a subset of choices. *RAND Journal of Economics, 38*(4), 1002–1019.

Fox, J. T. (2008). Estimating matching games with transfers. Working paper.

Galichon, A., & Henry, M. (2006). Inference in incomplete models. Working paper.

Ghose, A., & Sundararajan, A. (2006). Software versioning and quality degradation? An exploratory study of the evidence. Working paper.

Han, A. K. (1987). Nonparametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics, 35*, 303–316.

Hong, H., & Tamer, E. (2003). Endogenous binary choice model with median restrictions. *Economics Letters, 80*, 219–225.

Horowitz, J. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica, 60*(3), 505–31.

Horowitz, J. L. (1998). *Semiparametric Methods in Econometrics*. *Lecture Notes in Statistics* (Vol. 131). Springer.

Huang, C.-I. (2008). Estimating demand for cellular phone services under nonlinear pricing. *Quantitative Marketing and Economics*. doi:10.1007/s11129-008-9040-1.

Imbens, G., & Manski, C. F. (2005). Confidence intervals for partially identified parameters. *Econometrica, 72*, 1845–1857.

Kim, J., & Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics, 18*, 191–219.

Manski, C. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics, 3*(3), 205–228.

Manski, C. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics, 27*, 313–333.

Narayanan, S., Chintagunta, P. K., & Miravete, E. J. (2007). The role of self selection, usage uncertainty and learning in the demand for local telephone service. *Quantitative Marketing and Economics, 5*(1), 1–34, March.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica, 69*(2), 307–342, March.

Newey, W., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics* (Vol. 4, pp. 2111–2245). Elsevier.

Pakes, A., Porter, J., Ho, K., & Ishii, J. (2006). Moment inequalities and their application. Working paper, November.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *The Journal of Political Economy, 110*(4), 705–729.

Romano, J. P., & Shaikh, A. M. (2006). Inference for the identified set in partially identified econometric models. Working paper, September.

Romano, J. P., & Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified economic models. *Journal of Statistical Planning and Inference, 138*(9), 2786–2807, 1 September.

Rosen, A. M. (2006). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. Working paper, June.

Santiago, D., & Fox, J. T. (2007). A toolkit for matching maximum score estimation and point and set identified subsampling inference. Working paper, January.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica, 61*(1), 123–137, January.