Working Paper #04-15

January 2005

Estimating Continuation Values in an FCC Spectrum Auction

Patrick Bajari (Duke University and NBER)

& Jeremy T. Fox (University of Chicago)

Estimating Continuation Values in an FCC Spectrum Auction

Patrick Bajari and Jeremy T. Fox*

Duke University and NBER

University of Chicago

January 28, 2005

Abstract

The US government auctions licenses for mobile phone spectrum using a simultaneous, ascending bids auction design. While this auction design has many novel features, previous theoretical and experimental work also suggests that it could generate equilibria that are inefficient, collusive or display other undesirable properties. We empirically examine bidding in the C Block of the U.S. spectrum auctions. We propose a simulation-based estimator that recovers estimates of bidders' continuation value functions at the end of the auction. The value function estimates allow us to test for properties of the equilibrium allocation. Our estimator contributes to the econometrics of semiparametric discrete choice models by allowing agents to choose from very large, but discrete sets of alternatives which is important in other applications including housing or matching. We also estimate complementarities between the licenses in a package, an important issue when deciding the optimal geographic area size of mobile phone licenses.

*Thanks to the NET Institute for financial support. Thanks to helpful comments from Lawrence Ausubel, Nicholas Economides, Philippe Fevrier, James Heckman, Paul Milgrom, Gregory Rosston, Andrew Shepherd, Andrew Sweeting and Edward Vytlacil. Thanks to Todd Schuble for help with GIS software, to Peter Cramton for sharing his data on license characteristics, and to Chad Syverson for sharing data on airline travel. Excellent research assistance has been provided by Luis Andres, Stephanie Houghton, Dionysios Kaltis and Wai-Ping Chim.
Our email addresses are bajari@econ.duke.edu and fox@uchicago.edu.

1 Introduction

The US Federal Communications Commission (FCC) auctions licenses of radio spectrum for mobile phone service. A license covers a specific block of radio spectrum and a specific geographic area. Based on the recommendations of academic economists, the FCC replaced traditional auction formats with a novel simultaneous ascending auction. Bidding lasts for multiple rounds, and there is simultaneous bidding on all licenses. The sale of all licenses closes when no more bids for any license are forthcoming. The auction rules mean bidders can assemble a packages of licenses exhibiting the potential for substantial complementarities (or synergies). For instance, a company that holds two geographically adjacent licenses can offer mobile phone users a greater geographically contiguous coverage area.

The economic importance of the spectrum auctions (revenues usually exceed $1 billion in these auctions) has stimulated a large and vibrant theoretical literature. However, with the notable exception of the papers listed below, there is relatively little empirical work on estimating the payoffs of bidders in spectrum auctions. One reason is formulating a tractable econometric model of bidding behavior in the spectrum auctions is difficult. These auctions are dynamic games of incomplete information. The repeated play nature of dynamic games means they can have a large number of equilibria, including highly inefficient and even collusive equilibria. To demonstrate the difficulty of finding an equilibrium in an auction with complementarities, consider the work of Rosenthal and Szentes (2003) and Szentes (2004). These papers only consider the case of two or three objects for sale with two bidders making one-time simultaneous sealed bids. Further highlighting the difficulty of the problem, the equilibria found are in mixed strategies and are sensitive to the rules for breaking ties.

In this paper, we propose and estimate a structural model of bidding using data from the 1995 auction of licenses for the C Block of the 1900 MHz spectrum band. We begin by providing some background information on the auctions and discuss previous evidence about complementarities. Empirical work Ausubel et al. (1997) and Moreton and Spiller (1998) documents

that some bidders purchased licenses that were geographically adjacent, and winning bids are higher in markets where the second-highest bidder won adjacent licenses.

Next, we we turn to the problem of specifying an empirical model of bidder behavior. The main difficulty, as discussed above, is that there is no solvable model to make predictions for bidding in spectrum auctions. In the spirit of Haile and Tamer (2003), we search for predictions that are robust across a wide range of theoretical models. One prediction that must hold in any equilibrium model (with pure strategies) is that at the end of the auction, the bidder's continuation value from the chosen action must be at least as large as the continuation value from actions that are not chosen.

We demonstrate that recovering bidders' continuation values at the end of the auction is equivalent to estimating a semiparametric multiple discrete choice model. The choice set is the set of packages that bidders are eligible to bid on given the rules of the auction. This multiple discrete choice problem is computationally intractable using standard methods. The set of alternative actions that bidders face are very large. In the C block auction, bidders have more choices than the number of atoms in the universe. We propose a two-stage estimator for semiparametric discrete choice models. The two-stage estimator is computationally simple and can be applied to problems with very large numbers of choices. As we shall discuss in the text, our analysis extends previous results on identification of maximum score and related estimators to the case where the estimator uses a subset of choices purely for computational reasons.

Using our estimates of bidder continuation values, we then test three hypotheses about equilibrium bidding. First, we ask if the observed equilibrium is inefficient. In more easily analyzed auction settings, Gul and Stacchetti (2000) and Hatfield and Milgrom (2003) show that auctions can achieve efficient outcomes when licenses are substitutes, but may not achieve efficient outcomes when they are complements. By interpreting our estimates of continuation values as final payoffs, we can calculate the level of complementarities consistent with bidder behavior in this auction. Second, we ask whether the observed bidding appears to be collusive. Previous papers have argued that bidders may choose to bid less aggressively in order to avoid a price

war see, for instance, Brusco and Lopomo (2002). We address this by examining whether the end-of-auction continuation values are functions of the characteristics of the actual winner of licenses that could have been bid on in some counterfactual situation. Third, we ask whether observed bidding was distorted by perverse incentives to default on licenses. Zhèng (2001) analyzes a first-price sealed-bid auction and suggests that bidders may strategically anticipate the possibility of default. In an equilibrium with default, bidders with less collateral bid more aggressively because they will have less at risk in a state of the world where default occurs. Default then leads to an adverse selection problem where bidders with fewer financial resources are more successful in winning licenses.

One policy question we are particularly interested is the optimal geographic size of spectrum licenses. Auctioning licenses for small geographic areas leads to the possibility that license ownership will be fragmented, and carriers will be unable to realize the potential complementarities we measure through bidder behavior in this paper. Auctioning licenses for larger regions ensures that many of these potential complementarities will be realized.[1] In Europe, where countries are smaller than the United States, governments usually auction nationwide licenses. On the other hand, the data we use come from a US spectrum auction with 493 distinct geographic markets. This policy issue is timely, as the FCC has announced future plans to auction spectrum recovered from television broadcasters for mobile phone use. The FCC's forecasted licensing scheme breaks the United States into only six geographic markets, so there is some indication the FCC itself believes geographic complementarities are important.

[1]Of course, firms could realize these complementarities through inter-firm contracts, such as roaming agreements. The theory of the firm provides various explanations for why firm boundaries matter. The empirical work in this paper can be interpreted as using the revealed preferences of bidders to test whether the bidders themselves felt contracts are sufficient to realize potential complementarities.

2 Background for the C Block Auction

2.1 FCC Spectrum Auctions for Mobile Phones

Wireless phones in the United States work on two major regions of radio spectrum. The FCC assigned 800 MHz licenses in the 1980s using comparative worth regulatory hearings, lotteries, and induced partnerships among applicants. In the 1990's, the FCC decided the mobile phone industry could support more competitors, and so allocated additional spectrum in the 1900 MHz block to mobile phone carriers. This time, the FCC assigned spectrum licenses using auctions.

There were three initial auctions of mobile phone spectrum between 1995 and 1997. The first auction (the AB blocks) sold 99 licenses for 30 MHz of spectrum for 51 large geographic regions. The auction raised \$7.0 billion for the US Treasury. The second auction (the C block) sold 493 30 MHz licenses in more narrowly defined geographic regions to smaller bidders that met certain eligibility criteria. The auction closed with bids totaling \$10.1 billion, although some bidders were unable to make payments, and their licenses were later re-auctioned. The third auction (the DEF blocks) sold licenses for 10 MHz in the same 493 markets as in the C block. The bids totaled \$2.5 billion in the DEF blocks.

This paper uses data from the C block. The number of observations in our estimation procedure is the number of bidders, and there are 255 bidders in the C block, compared to only 30 in the AB blocks 155 in the DEF blocks. Further, many of the bidders in the AB and DEF blocks were incumbent mobile phone carriers, and for antitrust reasons were ineligible to bid in geographic markets where they already held licenses.[2] This creates additional complexities in estimation. The C block, by comparison, featured almost only potential new entrants, and so all bidders could potentially bid on all licenses. An additional reason we do not consider the DEF block is that it suffered from collusion through signal-bidding (Cramton and Schwartz, 2002), which is outside the scope of our analysis.

[2] In particular, parties owning more than a 40% interest in an existing wireless license in an area could not bid on another license in that area. This policy may have lowered competition in the AB auction (Ausubel et al., 1997; Salant, 1997).

2.2 Auction Rules

Similar rules govern all FCC auctions for spectrum for mobile phone uses. Each auction operates in an ascending bid, simultaneous-close format. In other words, each auction lasts multiple rounds, where in a round all licenses were available for bidding. A round lasts one business day. Bidding on all licenses closes at the same day. This allows bidders to assemble a useful package of licenses from those available, without risk of a necessary license to complete a package being unavailable because of an early close.[3] These auction rules were explicitly designed to allow bidders to assemble packages exhibiting complementarities, while letting the bidders themselves and not the FCC determine where, if anywhere, true complementarities lie.

Bidders pay an upfront amount of money for eligibility. Eligibility is the total population (pops) that a bidder can bid on during an individual round. For example, a bidder who pays to be eligible for 100 million pops cannot bid on licenses that cover geographic areas that together contain more than 100 million people. Eligibility cannot be increased after the auction starts. From an empirical point of view, eligibility payments provide early evidence on a bidder's willingness to devote financial resources towards winning a large number of spectrum licenses.[4] This paper does not consider strategic motives (such as intimidating rivals) for choosing eligibility levels.[5]

[3]An additional potential problem is bidding high on two complementary licenses, winning only one and being forced to purchase it even though the bidder does not value it without its complement. The FCC allows each bidder a low number of withdrawal credits in order to mitigate this exposure problem.

[4]The eligibility payments were 2 cents for MHz-individual in a hypothetical license for the AB blocks, 1.5 cents for the C block, and 6 cents in the DEF blocks. Compared to the closing auction prices, these payments are trivial.

[5]The FCC is concerned with concluding the auction in a reasonable amount of time. Therefore, each bidder is required to make a certain number of bids, in terms of pops, or lose some of its eligibility. As information is revealed about the demands of other bidders, bidders drop their activity levels to an amount corresponding to the licenses they hope to win. By the close of the auction, a bidder's eligibility is generally only slightly higher than the licenses it wins.

2.3 The C Entrepreneurship Block

This paper studies bidder behavior in the C block auction. By issuing discounts to small businesses, the FCC effectively allowed only such small businesses to bid.[6] This policy fulfilled a Congressional mandate to encourage smaller companies to offer wireless phone service. Bidding for the C block was more aggressive than in the AB block, with bids (for only half the spectrum sold in the AB blocks) totaling $9.2 billion. The largest winner in the C block auction was NextWave, which spent a total of $4.2 billion for 56 licenses, including close to $1 billion for the New York City license.

Bidders were given an extended payment plan of 10 years. Many of the bidders planned to secure outside funding for both their license bids and other carrier startup costs after the auction. Securing licenses first and financing later was an extremely important part of the business plan of what was until the late 1990s the most successful American mobile phone carrier, McCaw Cellular.[7] McCaw grew from a regional cable provider to a multi-billion dollar mobile phone carrier by purchasing licenses and then using the licenses as collateral to secure loans. This strategy was based on McCaw's (correct) forecast of the revenue potential in mobile phones, which was higher than the forecasts of larger companies (James B. Murray, 2001). It is possible that many of the C block bidders were trying to recreate McCaw's strategy. Without a license, a C block bidder is not really a serious negotiating partner for financiers. With a scarce license, a small business bidder becomes a relevant player in the mobile phone industry, and can expect to hold serious discussions with financiers.

[6] Plans to give additional advantages to women and minorities were dropped because of litigation. Small business ownership requirements were not overly strict. In future revisions, we wish to examine the extent to which small businesses partnered with larger carriers. There are two ownerships structures that qualify bidders as small businesses. The first is a control group must hold 25% of the businesses' equity. Of that 25%, 15% (or 3/5) of the equity must be held by qualifying entrepreneurs. Of the remaining 75% of equity, no more than 25% can be controlled by any one entity. An alternative structure says the control group can be 50% of equity, with 30% being entrepreneurs. This allows the other 50% to be held by one outside entity, which in effect allows the company to partner with a major firm. The most famous case of partnering is Cook Inlet, an Alaskan native corporation that partnered with the incumbent carrier Western Wireless.

[7] McCaw was purchased by AT&T for $17.4 billion and renamed AT&T Wireless in 1993. AT&T Wireless was itself purchased by Cingular in 2004.

Compared to McCaw, the C block winners did not have an early-mover advantage. As it turns out, many C block winners were unable to meet their financial obligations to the FCC. These new carriers were unable to secure enough outside funding to both operate a mobile phone company and pay back the FCC. Many C block winners returned their licenses to the FCC, where they were re-auctioned. Others companies merged with with larger carriers (forming a large part of the licenses held by T-Mobile USA, for example), or were able to protect their licenses in bankruptcy court. NextWave is the most famous case of bankruptcy protection. NextWave was eventually able to settle with the FCC, and sell some of its licenses to other carriers for billions of dollars. Ex-post, the C block bidders, who were accused of bidding too aggressively at the time, considerably undervalued the eventual market value of the licenses. However, much of this value was to larger carriers, not small business entrants who could not secure the financing to operate as a mobile phone carrier. In 2004, only a few C block winners, such as MetroPCS, remain true independent carriers marketing service under their own brand.

The FCC limited any one bidder from winning more than 98 total licenses in the C and F entrepreneurs blocks. Only NextWave came close to meeting this limit.[8] Even though our estimator is based upon revealed preference, this FCC rule allows our estimates to explore the efficiency and revenue implications of a nationwide package (for example), without ruling out that allocation as inefficient because it did not occur.

Table 1 lists characteristics of the 85 winning and 170 non-winning bidders in the continental United States. Bidders have to pay an initial fee to be eligible to bid during an FCC spectrum auction. A bidder who pays more is eligible to bid on a set of licenses that together have a larger population. While it is possible that bidders invest in this fee strategically in order to signal rivals, we treat it as a measure of commitment to the auction. The average winning bidder paid fees to be eligible to bid on licenses covering 10 million people, while the average losing bidder was eligible to bid on licenses covering only 5 million people. Bidders also had to

[8] Ausubel et al. (1997) point out that because the limit was in total licenses rather than pops, NextWave had incentives to purchase licenses with the highest total population.

submit financial disclosure forms in order to qualify as entrepreneurs for the C block, which was limited to new entrants only. Here we see that the financial characteristics of winners and non-winners are similar, which leads me to believe that these financial forms did not represent the true resources of bidders. In the estimates of continuation values, we use only initial eligibility as a bidder characteristic.

2.4 Did the Auction Produce a Functioning Market?

This paper focuses on estimating the continuation values for agents at the final round of the C block spectrum auctions. As mentioned in the introduction, issues we are concerned with include the potential for collusion, aggressive bidding by bidders who can default if they cannot pay later, and the valuation of potential complementarities.

This section shows that despite all of these potential complications and the lack of solved theoretical models for this class of dynamic games, the C block auction created a market where the underlying characteristics of licenses were the primary source of variation in price across licenses.

The most important characteristic of a license is the number of people living in it. These people represent potential subscribers to mobile phone service. Figure 1 shows the winning bids by the population of the license, along with a fitted regression line. The slope of the regression line is \$52.7. For the most part, even the large population licenses, such as New York and Los Angeles, are only a little above or below the regression line. As population is so highly correlated with price, to some degree this paper is about the extent to which a license was sold for above or below the regression line.

Figure 1 is somewhat misleading because most of the markets have fewer than five million residents, and are clumped together at the left-hand side of the figure. Figure 2 plots the winning bid per resident for licenses with fewer than five million people. Here we see the mean winning bid per license is well below \$52.7, meaning that that number was driven by especially large

markets. Figure 2 shows that the final price per resident of more populated licenses is in fact greater. For example, there are no licenses with more than 1 million residents where the closing bid price is more than \$20. This pattern of higher prices for larger licenses could be driven by other license characteristics (such as demographics), but is also consistent with increasing returns to scale in mobile phone carrier operation, which may also create complementarities across licenses, as one way to increase scale is to win more licenses.

2.5 Previous Evidence on Complementarities

Many aspects of the design of the FCC spectrum auctions focus on the possibility that a package of licenses might be worth more than the sum of the values of the licenses if won by different bidders. Licenses with these properties exhibit complementarities (synergies).

Before looking at the auction results, one's prior might be that complementarities are not important in the spectrum auctions. The FCC chooses market boundaries to be in sparsely settled areas in order to minimize complementarities across markets. Furthermore, 1900 MHz wireless phone service is mainly deployed in urban areas and along major highways.[9] Salant (1997) makes the additional point that buildout costs were expected to be two or three times the cost of the licenses themselves. Finally, companies can coordinate with contracts (roaming agreements) if the same company does not own the adjacent licenses.[10]

Researchers examining the auction results have generally concluded that complementarities were important. Ausubel et al. (1997) present maps that show several large winners in the AB and C block auctions win licenses in markets adjacent to each other or adjacent to markets where the bidder is a mobile phone incumbent, or a landline telephone carrier. For example,

[9]To some extent, 1900 MHz licenses are primarily built out in urban areas because the FCC requires build-outs to cover a certain fraction of the population of the market, rather than a fraction of the market's land area. 800 MHz carriers tend to cover both urban and rural areas because the FCC requires coverage as a large fraction of the land area of those licenses.

[10]The Coase Theorem suggests that, in a frictionless world, such contracts will implement the efficient outcome. Our paper uses revealed preference to investigate whether bidders thought the Coase Theorem would be operative in the post-auction mobile phone service industry.

Pacific Bell, at the time a California telephone company, won AB block licenses in California. On the other hand, other bidders, such as the forerunners of Sprint PCS and AT&T Wireless, embarked on a strategy of winning licenses in as many markets as allowed. Still other bidders, such as NextWave in the C block, purchased clumps of adjacent licenses in different areas of the country. Finally, the majority of winning bidders won only a few licenses, meaning that complementarities were probably not critical in the common parts of valuations for those licenses. We calculate that only 20 out of 89 C block winning bidders won packages of licenses where the population in adjacent licenses within the package was more than 1 million.[11] From the maps alone, it appears some winning bidders cared more about complementarities than others.

Salant (1997), a consultant during the AB block auctions for GTE, provides an insider's take on bidder valuations. GTE did value complementarities, in that it wanted to acquire licenses in areas where it was a landline phone company, and in areas that would fill in holes in its existing wireless phone network. GTE was unwilling to bid on certain potentially lucrative licenses, such as Los Angeles, because GTE felt it would not be financially worth it to win such an expensive license.

Ausubel et al. (1997) and Moreton and Spiller (1998) examine whether adjacent licenses exhibited complementarities by regressing the log of winning bids on market and bidder characteristics. Ausubel et al. (1997) study the AB and C block auctions and find that the log of winning bids are positively related to measures of potential complementarities for the runner-up bidders, as one might expect in an ascending-bid auction. Moreton and Spiller (1998) have better measures of incumbency, and also find that winning bids are positively related to the runner-up bidder's measures of complementarities. The results are the most statistically significant for the C block auction.[12]

[11] This complementarity measure is calculated over pairs of licenses. If a license is adjacent to two others in a package, its population will be counted twice.

[12] Ausubel et al. (1997) and Moreton and Spiller (1998) do not claim their price regressions correspond to hedonic estimates of bidder valuations. Rather, they specify descriptive or in-sample prediction regressions designed to summarize facts about the closing bid prices.

The approach in this paper differs from regressing bids on market and bidder characteristics. The dependent variable in our estimation procedure is the package of licenses won by a bidder. An observation is an individual bidder. Implicitly, we relate the closing bid prices to the decisions of all bidders, not just the winning or runner-up bidders. Therefore, we discipline our results to apply equally to bidders who win no or only one license as to big winners such as NextWave. Intuitively, identification comes from the joint coincidence of licenses within the same package and the equilibrium matching between heterogeneous licenses and bidders, in addition to price variation.[13]

The previous authors are also interested in global complementarities or increasing returns the scale, the notion that a wireless network involves fixed costs that can be spread out among more customers in a larger carrier. Global complementarities can be represented as a valuation convex in package characteristics such as total population.

2.6 Properties of Winning Packages

Table 2 lists characteristics of winning packages. The non-price package characteristics represent proxies for operating scale and geographic scope that will form the basis for estimating continuation values for bidders in the auction. The average winning bidder agreed to pay $116 million and won a license covering 2.9 million people. The largest winner bid $4.2 billion for a package covering 94 million people. Counting all the pairwise combinations of licenses in a package, a the mean winning package had 175 million pairwise combinations of residents in different markets a normalized distance of 1000 km (620 miles) from each other.[14] Not all mobile phone subscribers travel between markets by land. Two measures that preview future work but have not been used in fully-converged estimates are a measure of (unweighted) total trips

[13] If there are L licenses for sale, the market share of each license is $1/L$. Thus, traditional demand estimation techniques that focus on relating market shares to product characteristics cannot be used.

[14] More formally, for a package $\mathcal{L}_i$, the geographic scope proxy is $\sum_{j\in\mathcal{L}_i}\sum_{k\in\mathcal{L}_i,k\neq j}\frac{\text{population}\,k}{\text{distance}(j,k)}$, with population in millions and distance in kilometers. Distances are measured from the population-weighted centroid (as calculated by GIS software) of each market.

between all combinations of two licenses in the American Travel Survey, and the total number of airline passengers traveling one-way between all pairs of licenses in a package.[15] The airline data show the mean package has 321,000 passengers traveling by air within it. Note that for all geographic and commercial scope proxies, some fraction of the winning packages have a value of 0. For example, 26 out of the 85 winning packages contain only one license in the continental United States. Therefore, looking at only the actions of a few large carriers may distort one's impression of how important scope economies are. The fact that singleton packages are observed means that other factors influence wireless industry structure.

3 An Empirical Model of Spectrum Auctions

In this section, we propose an empirical model of bidding for spectrum licenses. The FCC spectrum auctions are ascending-bid, multiple round auctions that can take more than a hundred days to complete. Then, formally speaking, a spectrum auction is a dynamic game, potentially with incomplete information. The auction can potentially go on forever, so it is, formally, a (countably) infinite horizon game.[16]

3.1 Basic Model

We index rounds, or business days in the actual auction, by t. There are $i = 1, ..., N$ bidders who compete to win licenses $l = 1, ..., L$. In the C block auction, N is 255 and L is 493. An FCC spectrum auction is a multiple unit auction and therefore bidders can submit bids on multiple licenses. The auctioneer, the FCC, keeps track of the highest bid on each license. Therefore, the simplest version of the state space at round t, s_t, is the highest bidder and identity of the highest bidder for each of the L licenses. The state space s_t evolves according to a deterministic rule given the bids submitted in a round. The state space at $t + 1$, s_{t+1}, is s_t with new highest

[15] The airline data are from the Airline Origin and Destination Survey.

[16] The FCC gave itself reserve powers to end the auction if the normal course of bidding failed to do so. As these powers were not used, we do not model them.

bids and bidders for licenses that had activity during round t. The FCC does not entertain bids below the current highest bid for a license.

The state space can be expanded to handle more realistic details of the auction. At each round t, each bidder has an eligibility level, measured in total population of licenses in a bidder's package. The sum of the population of licenses for a bidder's current highest bids and its new bids must be less than its eligibility. In order to speed the conclusion of the auction, the FCC reduces the eligibility of bidders that do not submit enough bids. The state space thus can be extended to include the vector of remaining eligibilities of all N bidders, $(a_{1t}, \ldots, a_{Nt})$. The notation a stands for activity, which is the way a bidder maintains eligibility. The initial eligibility level is purchased by a bidder before the beginning of the auction, and is labeled a_{i0} for bidder i. Also, certain Nash equilibria in a dynamic game may involve strategic interaction between players. In this case, the history of past actions may be relevant. We label the history of past actions known at time t as h_t.

While bidders submit bids on individual licenses, a bidder is concerned about the package of bids he wins. A package of licenses is an element of the power set of $\{1, ..., L\}$ which we denote as $\mathcal{P}(\{1, ..., L\})$. We let $\mathcal{L}_{it} \in \mathcal{P}(\{1, ..., L\})$ denote a package of licenses bid on by bidder i at round t and $p_{it}(\mathcal{L}_{it})$ will denote the vector of the bids submitted for this package. If a bidder is currently the highest bidder on a license, he is constrained to keep its bid the same on that license.[17] The vector $p_{it}(\mathcal{L}_{it})$ is our representation of an action of a bidder in the auction. Because of the eligibility rules, $\mathcal{L}_{it} \in A_{it}(s_t)$, the set of packages i can bid on, given its current winning bids and free eligibility. Importantly for our purposes, in the C and F block auctions $A_{it}(s_t)$ includes the constraint that no one bidder can win more than 98 licenses between the two auctions combined.

Auctions are different from other dynamic games because a bidder's payoff is based only upon the package it wins at the end of the auction, and the price paid for that package. Label a

[17]The auction rules give each bidder a small number of chances to withdraw a bid. Our model ignores this possibility.

generic terminal round to the game T. The payoff function at T is

$$u\left(\mathcal{L}_{it}, x_{\mathcal{L}_{iT}}, \xi_{\mathcal{L}_{iT}}, z_i, \varepsilon_i\right) - \sum_{j \in \mathcal{L}_{iT}} p_{jT}, \tag{1}$$

where p_{jT} is the component of $p_{it}\left(\mathcal{L}_{it}\right)$ corresponding to a license j in the package $\mathcal{L}_{iT}$. In equation (1), the term $u\left(\mathcal{L}_{it}, x_{\mathcal{L}_{iT}}, \xi_{\mathcal{L}_{iT}}, z_i, \eta_i, \varepsilon_i\right)$ denotes i's valuation for the licenses $\mathcal{L}_{iT}$. This valuation is a function of the characteristics of these licenses, which we denote as $x_{\mathcal{L}_{iT}}$; characteristics of the licenses that are observed by the bidders, but not the econometrician, $\xi_{\mathcal{L}_{iT}}$; characteristics of bidder i observed in the data, z_i; and ε_i, the private information of bidder i.[18] Note that $u\left(\mathcal{L}_{it}, x_{\mathcal{L}_{iT}}, \xi_{\mathcal{L}_{iT}}, z_i, \varepsilon_i\right)$ does not depend on the private information of other bidders. Therefore, we implicitly rule out common values in bidders' payoffs.[19] We rule out common values for analytic tractability. Also, at the end of the auction, the period at which we focus, it can plausibly be argued that players have already revealed a great deal about their private signals as a function of their participation and history of submitted bids.

In a pure-strategy, subgame perfect Bayes-Nash equilibrium of this extensive form game, bidders maximize expected discounted utility at every state in the game tree. Utility is computed by taking expectations about the probability of reaching the various terminal nodes of the game tree as a function of bidder i's own strategy and the strategies of other players. In particular, at any state s_t, bidder i has a continuation value, or Bellman equation,

$$V_{it}\left(s_t\right) = \max_{p_{it}(\mathcal{L}_{it}) \in A_{it}(s_t)} 0 + E\left[V_{it+1}\left(s_{t+1}\right) \mid p_{it}\left(\mathcal{L}_{it}\right)\right].$$

[18] The structural auction literature does not traditionally allow for information known to all agents but not the econometrician about individual bidders, although the focus on heterogeneous competitors is of central importance in the literatures on demand estimation and entry. Our estimator currently follows the literature on auctions by assuming the information sets of agents and the econometrician are similar.

[19] It is likely that the major source of common uncertainty is future demand for wireless services by consumers. This is not a market or bidder-specific source of uncertainty. We also rule out the idea that winning bidders are worried about post-auction competition with other bidders in the wireless phone service market. This is a small concern in the C block auction, where only one license for a given market was for sale. Competition with incumbents not bidding in the auction is an important part of bidder valuations we do not explicitly measure.

We write "$0+$"to emphasize that there is no current-period payoff in an auction.[20] The expectation in the continuation value is taken knowing the optimal decision rule of i at state s_t, $p_{it}(\mathcal{L}_{it})$, which is just the $p_{it}(\mathcal{L}_{it})$ that maximizes $E[V_{it+1}(s_{t+1}) \mid p_{it}(\mathcal{L}_{it})]$. The set of licenses where new bids are entered must satisfy the eligibility rules, $A_{it}(s_t)$.

The new state, s_{t+1}, evolves according to the submitted bids of all players. The uncertainty is over the unknown private information of the other bidders, which may cause uncertainty in their bidding.[21] At a Nash equilibrium, the strategies of all bidders as a function of the unknown private information are known. For an individual bidder, the strategies of rivals are subsumed into the expectation operator. At a terminal node,

$$V_{iT}(s_T) = u\left(\mathcal{L}_{it}, x_{\mathcal{L}_{iT}}, \xi_{\mathcal{L}_{iT}}, z_i, \varepsilon_i\right) - \sum_{j \in \mathcal{L}_{iT}} p_{jT}, \tag{2}$$

the payoff of a bidder. Note that in our notation, a terminal node T is the round after the last round where the FCC solicits bids. By definition, a bidder has no possible actions at that point.

In the data, the auction ended at some round we call T. If i knows the auction will end if it makes no new bids, its observed action to end the auction must have had a higher payoff than the expected payoff from making new bids.[22] If the auction ends, i earns the payoff in equation (1). If i enters a new bid, it enters down a new path in the game tree. So, formally, revealed preference implies that if i bids on $\mathcal{L}_{iT}^*$ at the end of the auction, then

$$V_{iT}(s_T) \geq V_{iT}(s_T, \mathcal{L}'_{iT}) \ \forall \ \mathcal{L}'_{i,T} \in A_{i,T}(s_T), \tag{3}$$

where $V_{iT}(s_T, \mathcal{L}'_{iT})$ is the continuation value that i receives from bidding on a package $\mathcal{L}'_{i,T}$ at the end of the auction. Revealed preference implies that the continuation value from the

[20] We ignore discounting, or impatience to end the auction.

[21] Alternatively, a rival bidder might be playing a mixing strategy, although we do not allow for mixed strategies in estimation.

[22] Of course, i can never know for sure if the auction will end, because all bidders must cease making new bids, but i can know the end is near with high probability.

package of licenses that i bid on was superior to packages of licenses $\mathcal{L}'_{i,T}$ that it could have bid on.

3.2 The Number of Outcomes

In markets with heterogeneous goods, the number of different final allocations is immense, as the identity of each good matters. From the point of view of an individual bidder in the C block auction, the total space of outcomes is all packages of the 493 licenses with 98 or fewer total licenses. License combinatorics yield $3.58\text{x}10^{105}$ such packages.[23] There are more packages than atoms in the universe.[24] Clearly, any theoretical model or estimation procedure that attempts to tackle a problem of such high dimension is going to require simplifying assumptions. Note that the large number of packages precludes certain auction designs, such as particular types of package bidding, because the evaluation of all bids to find the bids that maximize the seller's revenue is computationally infeasible.

4 Estimation of Continuation Values using Maximum Score

The revealed preference inequalities (3) can be used to formulate a discrete choice model. Bidder i can choose from a potentially large set $A_{i,T}(s_T)$ of actions. We will attempt to recover an estimate of $V_{iT}(s_T, \mathcal{L}_{iT})$, i's continuation value from bidding on a particular package $\mathcal{L}_{i,T} \in A_{i,T}(s_T)$.

In the introduction we noted that auction theorists have have not solved these types of dynamic auction games. Therefore, our estimator imposes only a limited structure. The continuation value $V_{iT}(s_T, \mathcal{L}_{iT})$ is an outcome of a game, not a structural primitive. However, we estimate it as if it is such a primitive, although we do not impose a parametric family for the

[23]The power set of all packages, ignoring the FCC's rule of no more than 98 licenses per package, has size $2^{493} = 2.557\text{x}10^{148}$.

[24]Physicists estimate that the total number of atoms in the universe ranges from 10^{79} to 10^{81}, clearly a good deal fewer than the number of packages in the C block auction.

pieces of private information that distinguish one identical bidder's behavior from another's. A second difficulty that we face in estimating $V_{iT}(s_T, \mathcal{L}_{iT})$ is that there is a very large number of possible choices, potentially more than the number of atoms in the universe, as we demonstrated in Section 3.2. Therefore, standard approaches to estimation of discrete choice models cannot be applied.

4.1 A General Discrete Choice Model

We propose a method for estimation of a general class of semiparametric discrete choice models. Since the framework we propose will be more general than the spectrum auction problem at hand, we introduce an alternative notation.

Let the true model be a random utility model where the agent makes a choice j out of J alternatives. Call the set of choices $\mathcal{J}$. An agent i picks choice j if

$$x'_{ij}\beta - p_j + \delta\xi_j + \epsilon_{ij} \geq x'_{ik}\beta - p_k + \delta\xi_k + \epsilon_{ik} \;\forall\; k \in \mathcal{J}, \tag{4}$$

where β and δ are model parameters, the x_{ij} represent observed product characteristics potentially interacted with consumer i's demographics, p_j is the observed price of good j, and ξ_j is an product attribute (such a product quality) that is not observed in the data. The unobserved attribute ξ_j introduces the econometric problem that price p_j is endogenous because it may be a function of ξ_j (higher-quality products command higher prices). Note that the utilities functions use the normalization that the parameter on price is -1, or that the payoff function is expressed in monetary units.

Our identification and consistency theorem below requires a least one observed covariate that has support equal to the real line. This is a standard condition in semiparametric identification (Horowitz, 1998).

Assumption 1. *For each agent i and choice j, at least one choice-specific component x_{i1j} of the*

covariates has a conditional density $g\left(x_{i1j} \mid x_i \backslash x_{i1,-j}\right)$ (conditioning on the other covariates) with positive support on the entire real line and no mass points. Further the parameter β_1 on x_i is nonzero.

This condition means there exist a continuum of moment restrictions (one for each x), and moment restrictions that are relevant for every potential value of the unknown parameter β. In our case, the number of choices J is large, but still finite. Itemizing over the entire choice set only provides a finite number of moments, while adding additional observations with new x variables (the exercise in identification) creates a continuum of restrictions. Thus, semiparametric identification and estimation take advantage of support conditions and do not require examination of the entire choice set.[25]

Assumption 2. *Let z be a vector of instruments distributed independently of ξ_j and ϵ_{ij}. Furthermore, let there be an auxiliary pricing equation, so that*

$$p_j = z_j'\gamma + \xi_j + \eta_j, \tag{5}$$

with a pricing error η_j, which is independent of z, and ξ_j in an omitted product characteristic that captures the correlation between prices and unobserved consumer tastes.

If we substitute the price equation into the utility equation, the total utility of choice j as a function of x and z is

$$x_{ij}'\beta - \left(z_j'\gamma + \xi_j + \eta_j\right) + \delta\xi_j + \epsilon_{ij} = x_{ij}'\beta - z_j'\gamma + (\delta - 1)\,\xi_j - \eta_j + \epsilon_{ij}$$

From now on we call the composite error term $\mu_{ij} = (\delta - 1)\,\xi_j - \eta_j + \epsilon_{ij}$.

Assumption 3. *Let the composite error μ_{ij} have an iid-across choices distribution $F\left(\mu_{ij} \mid x, z\right)$.*

[25] Horowitz (1998) provides a more in-depth discussion of similar support conditions needed for identification in simpler single-index models. If the covariate support condition is violated, then bounds can be placed upon β.

For a given set of covariates and instruments, the stochastic terms enter utility must be iid across choices. However, the parametric family of the distribution of errors can vary with the observed variables x and z. For example, if a consumer's sex is a covariate included in x, men may have mixed normal errors, while women may have Laplace errors.

The role of the iid error assumption in the multinomial maximum score estimator of Manski (1975) is found in the following lemma. The lemma explains that comparing linear indices leads to maximizing correct predictions, as choice probabilities are monotonic in deterministic payoffs.

Lemma 1. *Under Assumptions 1, 2 and 3,*

$$\mathrm{Prob}(j \mid \beta_0, \gamma_0, x, z) > \mathrm{Prob}(k \mid \beta_0, \gamma_0, x, z) \Leftrightarrow x_j'\beta_0 - z_j'\gamma_0 > x_k'\beta_0 - z_k'\gamma_0, \; \forall\, j, k \in \mathcal{J}, \tag{6}$$

where β_0 and γ_0 are the true parameter values of the data-generating process.

Proof. The proof is Case (b) of Step 2 on pages 212-213 of the consistency theorem in Manski (1975), and is omitted here for brevity. Assumption 1 is needed to make the inequality strict. □

4.2 Maximum Score is Identified and Consistent for a Subset of Choices

In the first stage, form the estimator $\hat{\gamma}$ from regressing p_j on z_j. The regression line forms predicted product prices $z_j'\hat{\gamma}$. The second stage estimator of β is then

$$\hat{\beta}_N^{\mathcal{K}} = \arg\max_{\beta} \sum_{i \in \mathcal{N}(\mathcal{K})} \sum_{k \in \mathcal{K} \setminus \{j_i\}} 1\left[x_{ij_i}'\beta - z_j'\hat{\gamma} > x_{ik}'\beta - z_k'\hat{\gamma}\right], \tag{7}$$

where j_i is the observed choice of agent i, and $1[\cdot]$ is the indicator function equal to 1 when the condition in brackets is true, and 0 otherwise. The term "maximum score" comes about because if the payoff from an observed choice j is greater than some arbitrary alternative k, then the model predicts that j is more likely to be chosen, and the objective function increases

its prediction score by 1. On the other hand, if the model predicts k is more likely, the score of correct predictions does not increase.

A feature of our definition of the maximum score estimator that is new, to our knowledge, is that the estimator is defined for a fixed number $K = |\mathcal{K}|$ of alternatives. We call the fixed subset of estimation choices $\mathcal{K} \subseteq \mathcal{J}$. Note that as a result, only the observations observed to pick a choice in $\mathcal{K}$, $\mathcal{N}(\mathcal{K})$, contribute to the objective function. We address this below. Our estimator is consistent for $K = 2$, even if the true number of products in a market, J, is on the order of thousands or millions. The computational expense for evaluating the maximum score objective is on the order of $|\mathcal{N}(\mathcal{K})| \cdot K^2$. If we set $K = J$ for large J, evaluating the objective function would be computationally infeasible. To our knowledge, the only previous consistent estimator that is computationally feasible for multinomial discrete choice models with large numbers of choices is the logit sampling estimator of McFadden (1978), which relies on the restrictive independence from irrelevant alternatives property of the logit model.

Appendix B contains more details on subset maximum score estimation. The appendix discusses using multiple estimation subsets to increase the number of observations making positive contributions to the objective function. We also present a Monte Carlo study where the true model is iid mixed normal. The Monte Carlo experiments shows the advantages of the semiparametric subset maximum score estimator over the only existing computationally feasible discrete choice estimator for the case of many choices, the logit sampling estimator of McFadden (1978).

The following theorem states that maximum score is a consistent estimator when a subset of choices are used in estimation. The same argument shows that multinomial discrete choice models are identified under these assumptions, although previous work has shown this as well (Manski, 1975; Thompson, 1989; Matzkin, 1993).

Theorem 1. *Under Assumptions 1, 2 and 3, the maximum score estimator $\hat{\beta}_N^{\mathcal{K}}$ is a consistent estimator for β_0, the true parameter in the data generating process.*

We rely on Theorem 9.6.1 in Amemiya (1985) to prove the consistency of maximum score under estimation using a subset of choices.[26] The proof in this note is inspired by a proof in Amemiya. We split the proof of the theorem into two lemmas, which correspond to the two major requirements of Theorem 9.6.1. Lemma 2 shows that the limit of the objective function has a unique maximum at the true value of the parameters of the linear index underlying the random utility model. Lemma 2 is also a constructive proof for the identification of multinomial discrete choice models under our assumptions. Lemma 3 proves that the objective function converges to its limit uniformly in probability.

Lemma 2. *The limit of the maximum score objective function has a unique global maximum at the true parameter value, β_0. Therefore, β_0 is identified.*

Proof. Note that the maximum score objective function can be rewritten as

$$\frac{1}{N} Q_N^{\mathcal{K}}(\beta) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j \in \mathcal{K}} y_{ij} \sum_{k \in \mathcal{K} \setminus \{j\}} \mathbb{1}\left[x'_{ij}\beta - z'_j \hat{\gamma} > x'_{ik}\beta - z'_k \hat{\gamma}\right],$$

where y_{ij} is an indicator variable equal to 1 when agent i is observed to pick choice j in the data, and 0 otherwise. This indicator variable y_{ij} suppresses the notation of $\mathcal{N}(\mathcal{K})$, the set of observations where the observed choice is in $\mathcal{K}$.

The limit of $N^{-1} Q_N^{\mathcal{K}}(\beta)$ as the number of observations goes to infinity is therefore

$$Q^{\mathcal{K}}(\beta) = E_{x,z}\left[\sum_{j \in \mathcal{K}} \mathrm{Prob}(j \mid \beta_0, x, z) \sum_{k \in \mathcal{K} \setminus \{j\}} \mathbb{1}\left[x'_{ij}\beta - z'_j \gamma_0 > x'_{ik}\beta - z'_k \gamma_0\right]\right],$$

where the expectation is over the sampling distribution of $x = \{x_{ij}\}_{j=1}^{J}$ and $z = \{z_j\}_{j=1}^{J}$.[27] Note that the choice probability $\mathrm{Prob}(j \mid \beta_0, x, z)$ is the probability of choice using the true choice set, $\mathcal{J}$, and the true parameter vector, β_0. Also note that by the properties of OLS, $\hat{\gamma}$

[26] The notion of consistency in Amemiya's Theorem 9.6.1 is convergence in "the generalized sense", as mentioned in his Definition 9.6.1. This alternative notion of consistency is necessary because the subset maximum score objective function is not continuous in the parameter vector β.

[27] The limit of the objective function is taken over all observations, not only those in $\mathcal{N}(\mathcal{K})$.

converges to the true γ_0.

Fix x and z. Each choice pair j and k in $\mathcal{K}$ appears twice in $Q^{\mathcal{K}}(\beta)$: once multiplying $\mathrm{Prob}(j \mid \beta_0, x, z)$ and once multiplying $\mathrm{Prob}(k \mid \beta_0, x, z)$. Only zeros contribute to the objective function if $x_{ij}'\beta - z_j'\gamma_0 = x_{ik}'\beta - z_k'\gamma_0$. Otherwise, either $1\left[x_{ij}'\beta - z_j'\gamma_0 > x_{ik}'\beta - z_k'\gamma_0\right]$ or $1\left[x_{ik}'\beta - z_k'\gamma_0 > x_{ij}'\beta - z_j'\gamma_0\right]$, so one of the two indicator functions has value 1 and the other has value 0. An assignment where the value of 1 multiplies the higher of the two probabilities for all x and z would be a global maximum of $Q^{\mathcal{K}}(\beta)$. By Lemma 1, the parameter vector β_0 implements this assignment.

A question remains about the uniqueness of the global maximum. Let $\tilde{\beta}$ be another global maximum of $Q^{\mathcal{K}}(\beta)$, that is distinct from β_0. By Assumption 1 there exists a continuous covariate x_{i1} that varies separately for all choices. For any covariates x_{-1} and z, we can find a value $x_{i1j} - x_{i1k}$ such that $(x_{i1j} - x_{i1k})\,\tilde{\beta}_1 = x_{ik,-1}'\tilde{\beta}_{-1} - z_k'\gamma_0 - \left(x_{ij,-1}'\tilde{\beta}_{-1} - z_j'\gamma_0\right)$. There are as many such linear equations (in $\tilde{\beta}$) as there are distinct values of $\{x_{ij,-1}, x_{ik,-1}, z_k, z_j\}$. As long as there are more such values than parameters in $\tilde{\beta}$, then only $\tilde{\beta}$ can simultaneously solve all the equations. So for the true parameter vector β_0, there must exist one such point where $(x_{i1j} - x_{i1k})\,\beta_{0,1} \neq x_{ik,-1}'\beta_{0,-1} - z_k'\gamma_0 - \left(x_{ij,-1}'\beta_{0,-1} - z_j'\gamma_0\right)$.

Consider the case where $(x_{i1j} - x_{i1k})\,\beta_{0,1} > x_{ik,-1}'\beta_{0,-1} - z_k'\gamma_0 - \left(x_{ij,-1}'\beta_{0,-1} - z_j'\gamma_0\right)$, $\tilde{\beta}_1 > 0$ and $\beta_{0,1} > 0$. Equivalent arguments can be made when the inequalities that define this case switch. Then there exists some $\Delta > 0$ where $(x_{i1j} - x_{i1k})\,\tilde{\beta}_1 - \Delta\tilde{\beta}_1 < x_{ik,-1}'\tilde{\beta}_{-1} - z_k'\gamma_0 - \left(x_{ij,-1}'\tilde{\beta}_{-1} - z_j'\gamma_0\right)$ and $(x_{i1j} - x_{i1k})\,\beta_{0,1} - \Delta\beta_{0,1} > x_{ik,-1}'\beta_{0,-1} - z_k'\gamma_0 - \left(x_{ij,-1}'\beta_{0,-1} - z_j'\gamma_0\right)$. So at this point (where x_{ij1} has Δ subtracted from it), the alternative $\tilde{\beta}$ implies that k has the highest deterministic payoff, and the score 1 would multiply $\mathrm{Prob}(k \mid \beta_0, x_{j,1} - \Delta, x_{-1,j}, x_{-j}, z)$, which according to the true model is less than $\mathrm{Prob}(j \mid \beta_0, x_{j,1} - \Delta, x_{-1,j}, x_{-j}, z)$. Therefore the objective function is not maximized at $\tilde{\beta}$, and we have a contradiction.

Note that there is an strong inequality in the indicator function in the objective function, so that $\beta = 0$ is not a global maximum.

$\square$

Lemma 3. *The maximum score objective function converges uniformly to its limit.*

Proof. The proof does not contain any economics, so is placed in the appendix. ☐

5 Identification of Continuation Values in the C Block

Lemma 2 is a formal constructive proof of the identifiability of a semiparametric discrete choice model using a subset of alternatives. This section focuses on an intuitive discussion of identification of continuation values in the C block spectrum auction, by relating observed variables to the economic questions outlined in the introduction.

5.1 License Characteristics

In our application, we suppose that bidder i's continuation value at the end of the auction can be written as a function of the following variables:

- the total population covered by the licenses in the package $\mathcal{L}_{i,T}$,
- the number of people with household incomes over \$35,000 in the area covered by the licenses in the package $\mathcal{L}_{i,T}$,
- the initial eligibility of bidder i (which is interacted with other characteristics),
- various proxies for the potential complementarities between licenses in $\mathcal{L}_{i,T}$ (explained below), and
- the (population weighted) initial eligibility of the bidders who won the licenses in $\mathcal{L}_{i,T}$.

Some explanations are in order. As discussed in Section 2.4, the major characteristic predicting the closing bid price of a license is the population of that license. It follows quite naturally that

the income level of the potential customers in that license should also be a major determination of price.[28]

The only bidder-specific characteristic we consider (or have data on) is the initial eligibility of the bidder, which was discussed in Section 2.2. We use initial eligibility as a proxy for the financial resources of the bidder, while remembering that most bidders planned to negotiate for financing after the close of the auction. The mean of initial eligibility across 254 C block bidders is 6.7 million people, and the largest value was 181 million people. The largest eventual winner, NextWave, has an initial eligibility of 176.1 million people. Zhèng (2001) suggests that because of the ability of winners to default if they cannot raise financing for their bids, bidders with fewer financial resources should bid more aggressively. If initial eligibility is a proxy for financial resources, then we might expect to see continuation values increasing in interactions of initial eligibility and license characteristics.

Finally, the population-weighted eligibility of winning bidders proxies for the possibility that a bidder might not want to counterbid on a license where the eventual winner was a bidder with a lot of financial resources. **?** suggest that there are many collusive equilibria in an ascending bid auction. If the punishment regime in a collusive equilibrium is a price war, then in such an equilibrium bidders should be especially reticent to challenge high-bidders who have the financial resources to wage a long price war.

5.2 Complementarities vs. Correlated Preferences

As discussed in the introduction, the FCC spectrum auction design allows bidders to assemble a package of licenses that have the potential for complementarities. The C block bidder Carolina PCS won a package of licenses for most of South Carolina. Our discrete choice estimator

[28] The cutoff level of household income of $35,000 is the same measure used in Ausubel et al. (1997). We simply multiply the total population of the area by the fraction of households with incomes over $35,000 a year. Strictly speaking, we assume the number of household members does not vary with income. Magazine articles from that mid-1990s show that the level of penetration of mobile phones into lower-income groups that we see in 2004 was not predicted by many analysts, who considered higher-income groups to be the main market for mobile phones.

will find complementarities contribute positively to payoffs when this bidder's package features more potential for complementarities than alternative packages in the estimation choice set. An alternative explanation is that Carolina PCS has preferences for licenses in the southeastern United States. In other words, an explanation for the data is clusters of winning licenses are driven not by complementarities, but by heterogeneous bidder payoffs that are correlated across licenses. Distinguishing between true complementarities and correlated preferences is important for auction design. Without complementarities and ignoring strategic behavior, a sequence of separate ascending-bid auctions for each license ensures that all licenses are awarded to the bidders with the highest idiosyncratic payoff. The simultaneous auctioning of all licenses, as implemented by the FCC, is important mainly because of the potential for true complementarities between licenses in a package.

Because our computationally tractable discrete choice estimator is consistent only when the true data-generating process involves payoff shocks that are iid across packages, our structural approach cannot test between complementarities and correlated preferences.[29] Further, we recognize that any choice data can be explained by an arbitrarily complex distribution of stochastic errors without regard to deterministic payoff terms. However, by limiting ourselves to a reasonable class of correlated payoffs, we can argue why we believe true complementarities are more important.

The class of correlated preferences we consider is based upon the notion that each bidder has a spatial bliss point. If this is true, we expect bidders to purchase licenses mainly in the same region of the United States. Only two of the top ten winning C block bidders won licenses in only one region. One bidder purchased licenses to cover all of Tennessee and some bordering areas, and another bidder (Carolina PCS) won licenses for most of South Carolina. Both companies sold their licenses to national carriers before establishing serious market posi-

[29] To our knowledge, there are no identification proofs for semiparametric discrete choice models with correlated errors. McFadden and Train (2000) prove that the mixed logit model (which cannot be estimated with only a subset of choices) can arbitrarily approximate any system of choice probabilities. This flexibility property of a parametric model would seem to suggest that the deterministic part of payoffs are not semiparametrically identified when correlated errors are allowed.

tions as independent carriers. A third top ten bidder won licenses in a narrow, geographically contiguous band stretching from Detroit to Dallas, although this area is so diverse it is hard to explain with a geographic preferences explanation. The other top ten bidders that did win licenses in clusters did so in multiple areas of the country. The carrier now known as MetroPCS won licenses in southern Florida, the greater Atlanta area, and northern California. While it is possible MetroPCS had correlated payoffs for only those three regions, we (subjectively) feel complementarities are a much more likely explanation for MetroPCS's license clusters.

5.3 Proxies for Potential Complementarities

Our discrete choice model allows us to measure how much bidders' continuation values depend on complementarities when other license characteristics are factored in. One approach would be to estimate complementarities using abstract functional forms. Instead, we prefer to measure the valuation of complementarities by collecting data on variables that proxy for potential complementarities. We use four separate proxies for complementarities, in order to examine the robustness of our estimates to different measures.

Our first measure of potential complementarities between licenses in a package is the one used in the previous literature: the population of geographically adjacent licenses within a package. For a package $\mathcal{L}_{i,T}$, potential complementarities are

$$\sum_{j \in \mathcal{L}_{i,T}} \sum_{k \in \mathcal{L}_{i,T} \setminus \{j\}} \text{population}\,(k) \cdot \mathtt{1}[j, k \text{ adjacent}]\,,$$

where $\mathtt{1}[\cdot]$ is an indicator function and population is measured in millions of residents. Visually, geographic adjacency is interesting because it is evident from looking at a map of winning bids that some bidders purchase clusters of adjacent licenses. Our second proxy for potential complementarities is based on the geographic distance between pairs of licenses within a package.[30]

[30] We measure distance between two licenses using the population-weighted centroid of each license. The population-weighted centroid is calculated using a rasterized smoothing procedure using county-level population

For a package $\mathcal{L}_{i,T}$, potential complementarities are

$$\sum_{j \in \mathcal{L}_{i,T}} \sum_{k \in \mathcal{L}_{i,T} \setminus \{j\}} \frac{\text{population}(k)}{\text{distance}(k,j)},$$

where population is measured in millions and distance is measured in kilometers. We believe geographic distance is more accurate than the adjacency measure used in previous studies. Adjacency is especially poor in the western United States, where many C block license boundaries are in the desert or other sparsely populated areas.[31]

Geographic measures of distance may not capture all possible complementarities. In particular, a mobile phone customer may travel a long distance, and may still want to use his or her mobile phone in the destination. For example, many business users travel between Los Angeles and New York. In fact, the C block bidder NextWave won both the New York and Los Angeles licenses. We have two complementarity proxies based upon travel between two licenses. The first measure, from the 1995 American Travel Survey (ATS), is proportionate to the number of trips longer than 100 km between major cities. All forms of transportation are covered. The downside of this measure is that for privacy reasons the ATS does not provide enough information about rural origin and destinations to tie those to particular mobile phone licenses. Our second measure, from the Airline Origin and Destination Survey for the calendar year 1994, is the projected number of passengers flying between two mobile phone license areas.[32] The drawback of the air travel measure is that it assumes all passengers stay in the mobile phone

data from the US Census Bureau.

[31] Adjacency says, for example, that the Reno license is next to Los Angeles license, while those two cities are actually 615 km apart. Our population-weighted centroid measure says the Reno license is 510 km away from the Los Angeles license. Note that the previous descriptive empirical work only considered observed winning packages, and as Reno and Los Angeles were won by different bidders, this example is not relevant for their analysis.

[32] Intermediate stops are not counted for either dataset. For both datasets, geographic information software (GIS) was used to match origins and destinations with mobile phone licenses. For airports, the origin and destination license area is easy to calculate. For the MSAs (Metropolitan Statistical Areas) used in the ATS., the equivalent license area was found using the centroid of the origin or destination MSA. The C block license boundaries for urban areas typically correspond to MSAs.

license area where their destination airport is located.[33] Both travel measures take the form of

$$\sum_{j \in \mathcal{L}_{i,T}} \sum_{k \in \mathcal{L}_{i,T} \setminus \{j\}} \text{trips} \left(\text{origin is } j, \text{destination is } k \right),$$

where our ATS measure uses the count of raw trips in the survey, and the air travel count is inflated to approximate the total number of trips during 1994.[34]

6 Empirical Implementation and Results

6.1 Estimation Choice Set

The discrete choice estimation procedure in Section 4 allows us to consistently estimate the parameters of a discrete choice model while using only a subset of choices. This procedure works by using the monotonicity property (Lemma 1) of choice probabilities in discrete choice models when we assume that the error terms are iid across choices for a given bidder. The iid assumption is unlikely to hold when the choices (packages) are really comprised of individual licenses. However, the subset estimation procedure we introduce is the only known computationally tractable method (other than the logit sampling estimator of McFadden (1978), which also makes the iid errors assumption) to consistently estimate continuation value parameters.

Our chosen estimation subset has 336 packages in it. We consider only packages comprised of licenses in the continental United States, because of the various unique characteristics of licenses in Alaska, Hawaii and the overseas territories of the United States.[35] Our 336 packages are comprised of the outside good of not winning any license, 85 usable packages that were

[33]We effectively code that there are zero potential complementarities between rural licenses for both travel measures.

[34]Our airline passenger measure does not distinguish between origins and destinations, so we simply divide the formula for the complementarity proxy by 2. If all airline travel is round-trips during the same calendar year, this measure should be exactly correct.

[35]When constructing the observed choices of bidders, we also ignore licenses not in the continental United States. For example, a bidder winning only American Samoa would be coded as not winning any licenses in our estimation.

actually won in the C block auction, and 250 randomly sampled packages. To draw a random package of licenses, we first draw the number of licenses in that package, which we assume to have a discrete geometric distribution with the probability of first success equal to $1/15$, with the distribution truncated between 1 and 98 licenses. The FCC limits C block bidders to win no more than 98 total licenses. Given the number of licenses in a package, we draw the individual licenses to form packages with more complementarities than packages with uniformly drawn licenses would have. In particular, given an uniformly-drawn initial seed license in a package, a potential second license adjacent to the first is 25 times more likely to be picked than a non-adjacent license, and a potential third license is 25 times more likely to be picked for each of the first two licenses that it is adjacent to, and so on.

6.2 Continuation Value Estimates

We will use our estimates of continuation values to gain some insight into whether the allocation is efficient, whether bidders fear setting off a price war and finally whether there is an adverse selection problem due to bidders with low collateral bidding aggressively.

Table 3 lists estimates of β in payoffs, equation (2), from using the subset multinomial maximum score estimator for the estimation choice set of 335 choices described above.[36] Price is normalized to $\$-1$ (million), so these payoff parameters are already in monetary units. These estimates and confidence intervals were calculated using the median Laplace Type Estimator (LTE) of Chernozhukov and Hong (2003).[37] The most basic need for a carrier is potential cus-

[36]These estimates use only the 85 winning bidders. If non-winners are included, the results do not make economic sense unless a constant that only equals 0 when no licenses are won is included to represent a fixed cost of winning. The maximum score objective function, which counts the number of correct predictions given the deterministic payoffs, is then maximized by setting the fixed cost to $-50 billion, which gives correct predictions for all 336 choices for all 170 non-winners, and at least one wrong prediction for the 85 winners. The other parameter estimates are equal to the same values as when only the 85 winners are included with no fixed cost of winning a package. For this sample, including non-winners does not provide useful information to identify the parameters of interest.

[37]The LTE uses a Metropolis Hastings algorithm to simulate a quasi-Bayesian posterior for a classical extremum estimator. The LTE is useful for objective functions that have many local maxima or are otherwise hard to numerically maximize, such as (even the smoothed) maximum score. Unfortunately, the theorems in Chernozhukov and Hong apply only to $\sqrt{n}$-consistent estimators. While the sampling distribution of the multinomial maximum score

tomers. Column (1) presents a base specification where deterministic payoffs are linear in the population covered by a package, the population of higher-income people (a group initially targeted for mobile phone sales) in a package, and an interaction between high-income population and the initial eligibility of a bidder. An additional low income resident contributes $21.3 to payoffs, while at a typical (for winners) initial eligibility of 10 (million), an additional high income resident contributes $21.3 - 46.9 + 7.87 * 10 = \$53.1$.[38]

Column (2) adds a proxy for the geographic scope of a package. The estimated payoffs from population remain similar. For scope economies, for a bidder with an initial eligibility of 10, adding an additional resident that is 100 km (62 miles) away from a different market in the same package contributes $0.85 to payoffs, which seems small relative to the value of population by itself. The mean value for the scope proxy is 0.175 (millions/kms), so geographic scope economies for the mean package only contribute to payoffs a value equal to around 280,000 high income residents. However, scope economies for the largest winners may be substantial. For the second-largest winner, for example, scope economies for its winning package correspond to the value of 5 million high-income residents.

Column (3) adds squared high income population, both alone and interacted with bidder eligibility, in order to investigate operating scale economies. The derivative of payoffs with

is not known, it is known the binary choice maximum score estimator converges as the rate $n^{1/3}$ (Kim and Pollard, 1990). Because the median of the posterior distribution of the multinomial maximum score comes perilously close to numerically maximizing the maximum score objective function, it is highly likely that the LTE estimator is consistent for this problem. It is more speculative that the quantiles of the posterior distribution can be used for a 95% confidence interval, as can some of the $\sqrt{n}$-consistent cases described by Chernozhukov and Hong. For all runs of the Metropolis-Hastings algorithm, serious care was taken to let the algorithm run for hundreds of thousands of iterations after multiple burn-in runs. Plots were made of the random draws in order to see whether the Markov chain looks like an ergodic distribution.

[38]One concern with these payoff estimates is price endogeneity. Intuition from the omitted variable bias formula for ordinary least squares regression suggests that a positive correlation of price with an omitted package characteristic should cause an upward bias for the price coefficient, meaning that the (currently normalized to -1) price coefficient would be more negative. Re-normalizing the coefficients to dollar values suggests that the parameter estimates in Table 3 are overestimates of the payoffs from potential customers, geographic scope and geographic scale. Bajari and Fox (2004) develop an instrumental variables version of the subset maximum score estimator, where prices are allowed to be correlated with omitted product characteristics, although the numbers in Table 3 are not IV estimates.

respect to increasing the high-income population is

$$\$ - 34.9 + 12.8 \cdot \text{eligibility} - 28.4 \cdot \text{highincpop} + 1.37 \cdot \text{eligibility} \cdot \text{highincpop},$$

which corresponds to economies of scale for bidders who have initial eligibilities greater than 20.7. There were six winning bidders with eligibility greater than 20.7 million people, and together they won licenses corresponding to 54% of the US population. Adding potential scale economies to the specification increases the importance of scope economies, suggesting the lower estimates in Column (2) may have been due to omitting scale economies. The scope economies in Column (3) are large, with an additional resident located 100 km from another market in the same package being worth $8.67 for a bidder with an initial eligibility of 10 million residents. The maximum score estimates of bidder payoffs certainly are consistent with the presence of large scale and scope economies.

7 Conclusions

Our work is preliminary, and we do not want to make strong conclusions at this time. However, it appears the maximum score estimates of bidder preferences are consistent with the presence of strong returns to geographic scale and scope. Future versions of this paper will address concerns with collusion, limited liability, and other aspects of the auction that may have reduced allocative efficiency.

References

Amemiya, Takeshi, *Advanced Econometrics*, Harvard University Press, 1985.

Ausubel, Lawrence M., Peter Cramton, R. Preston McAfee, and John McMillan, "Synergies in Wireless Telephony: Evidence from the Broadband PCS Auctions," *Journal of Eco-*

nomics and Management Strategy, Fall 1997, *6* (3), 497–527.

Bajari, Patrick and Jeremy T. Fox, "Estimating Continuation Values in an FCC Spectrum Auction," November 2004.

Brusco, Sandro and Giuseppe Lopomo, "Collusion via Signalling in Simultaneous Ascending Bid Auctions with Heterogenous Objects, with and without Complementarities," *Review of Economic Studies*, April 2002, *69* (2), 407–436.

Chernozhukov, Victor and Han Hong, "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 2003, *115* (2), 293–346.

Cramton, Peter and Jesse A. Schwartz, "Collusive Bidding in the FCC Spectrum Auctions," *Contributions to Economic Analysis & Policy*, 2002, *1* (1), Article 11.

Gul, Faruk and Ennio Stacchetti, "The English Auction with Differentiated Commodities," *Journal of Economic Theory*, May 2000, *92*, 66–95.

Haile, Phil and Elie Tamer, "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, February 2003, *111* (1), 1–51.

Hatfield, John William and Paul Milgrom, "Matching with Contracts," November 2003. working paper.

Horowitz, Joel L., *Semiparametric Methods in Econometrics*, Vol. 131 of *Lecture Notes in Statistics*, Springer, 1998.

Kim, Jeankyung and David Pollard, "Cube Root Asymptotics," *The Annals of Statistics*, 1990, *18* (1), 191–219.

Manski, Charles F., "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 1975, *3*, 205–228.

Matzkin, Rosa L., "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 1993, *58*, 137–168.

McFadden, Daniel and Kenneth Train, "Mixed MNL Models for Discrete Response,," *Journal of Applied Econometrics*, 2000, *15* (5), 447–470.

McFadden, Daniel L., "Modelling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, Vol. I, North Holland, 1978, pp. 75–96.

Moreton, Patrick S. and Pablo T. Spiller, "What's In the Air: Interlicense Synergies in the Federal Communications Commission's Broadband Personal Communication Service Spectrum Auctions," *Journal of Law and Economics*, October 1998, *41* (2, Part 2), 677–725.

Murray, Jr. James B., *Wireless Nation*, Perseus Publishing, 2001.

Rosenthal, Robert W. and Balàzs Szentes, "Three-Object Two-Bidder Simultaneous Auctions: Chopsticks and Tetrahedra," *Games and Economic Behavior*, 2003, *44*, 114–133.

Salant, David J., "Up in the Air: GTE's Experience in the MTA Auction for Personal Communication Services Licenses," *Journal of Economics and Management Strategy*, Fall 1997, *6* (3), 549–572.

Szentes, Balàzs, "Two-Object Two-Bidder Simultaneous Auctions," May 2004.

Thompson, T. Scott, "Identification of Semiparametric Discrete Choice Models," 1989. University of Minnesota Discussion Paper 249.

Zhèng, Charles Z., "High Bids and Broke Winners," *Journal of Economic Theory*, September 2001, *100* (1), 129–171.

A Proof of Lemma 3

This is the proof of Lemma 3.

Proof. Amemiya has a straightforward proof of uniform convergence in probability for a binary ($J = 2$) choice objective function that can easily be generalized to the present multinomial case. By the triangle inequality and for any $\lambda > 0$,

$$\begin{aligned}\sup_{\beta}\left|\frac{1}{N}Q_N^{\mathcal{K}}(\beta) - Q^{\mathcal{K}}(\beta)\right| &\leq \sup_{\beta}\left|\frac{1}{N}Q_N^{\mathcal{K}}(\beta) - \frac{1}{N}Q_{\lambda N}^{\mathcal{K}}(\beta)\right| \\ &\quad + \sup_{\beta}\left|\frac{1}{N}Q_{\lambda N}^{\mathcal{K}}(\beta) - Q_{\lambda}^{\mathcal{K}}(\beta)\right| \\ &\quad + \sup_{\beta}\left|Q_{\lambda}^{\mathcal{K}}(\beta) - Q^{\mathcal{K}}(\beta)\right|,\end{aligned}$$

where

$$\frac{1}{N}Q_{\lambda N}^{\mathcal{K}}(\beta) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j\in\mathcal{K}} y_{ij}\sum_{k\in\mathcal{K}\setminus\{j\}}\psi_\lambda\big(x_{ij}'\beta - z_j'\hat{\gamma} - (x_{ik}'\beta - z_k'\hat{\gamma})\big)$$

and

$$Q_{\lambda}^{\mathcal{K}}(\beta) = E\left[\sum_{j\in\mathcal{K}}\mathrm{Prob}(j \mid \beta_0, x, z)\sum_{k\in\mathcal{K}\setminus\{j\}}\psi_\lambda\big(x_{ij}'\beta - z_j'\gamma_0 - (x_{ik}'\beta - z_k'\gamma_0)\big)\right],$$

where the function ψ_λ is defined as

$$\psi_\lambda(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ \lambda z & \text{if } 0 < z \leq \frac{1}{\lambda} \\ 1 & \text{if } z > \frac{1}{\lambda}\end{cases}.$$

The argument proceeds by working with each of the three terms generated by the triangle inequality, and showing that for any pair $\epsilon, \delta > 0$, there is a significantly large λ and n such that

for each term,

$$P\left[\text{term} > \frac{\epsilon}{3}\right] < \frac{\delta}{3}.$$

The middle term, $\sup_{\beta}\left|\frac{1}{N}Q^{K}_{\lambda N}\left(\beta\right) - Q^{\mathcal{K}}_{\lambda}(\beta)\right|$, involves a straightforward uniform convergence of a continuous (in b) finite-sample objective function to a continuous limiting function, and standard laws of large numbers apply. The only special complication is the convergence of the first-sage OLS estimator $\hat{\gamma}$ to γ_0. The triangle inequality can be reapplied to this term to distinguish between the limiting properties of the first and second stages. There is a sample size $N_1\left(\lambda\right)$, which is the max of the sample size corresponding to each of the two stages, that makes the probability of the sup difference arbitrarily small.

The first term involves both data and approximating a discontinuous function with a continuous function. The term is

$$\sup_{\beta}\left|\frac{1}{N}Q^{K}_{N}\left(\beta\right) - \frac{1}{N}Q^{K}_{\lambda N}\left(\beta\right)\right| = \sup_{\beta}\left|\frac{1}{N}\sum_{i=1}^{N}\eta_{\lambda}\big(x'_{ij}\beta - z'_{j}\hat{\gamma} - (x'_{ik}\beta - z'_{k}\hat{\gamma})\big)\right|,$$

where

$$\eta_{\lambda}(z) = \begin{cases} 0 & \text{if } z \leq 0 \text{ or } z > \frac{1}{\lambda} \\ 1 - \lambda z & \text{if } 0 < z < \frac{1}{\lambda} \end{cases}.$$

The triangle inequality can be reapplied to break the term into two parts, one focusing on convergence in the data and the other on approximation:

$$\begin{aligned} &\sup_{\beta}\left|\frac{1}{N}\sum_{i=1}^{N}\eta_{\lambda}\big(x'_{ij}\beta - z'_{j}\hat{\gamma} - (x'_{ik}\beta - z'_{k}\hat{\gamma})\big)\right| \leq \\ &\quad \sup_{\beta}\left|\frac{1}{N}\sum_{i=1}^{N}\eta_{\lambda}\big(x'_{ij}\beta - z'_{j}\hat{\gamma} - (x'_{ik}\beta - z'_{k}\hat{\gamma})\big) - E\left[\eta_{\lambda}\big(x'_{ij}\beta - z'_{j}\hat{\gamma} - (x'_{ik}\beta - z'_{k}\hat{\gamma})\big)\right]\right| + \\ &\qquad \sup_{\beta}\left|E\left[\eta_{\lambda}\big(x'_{ij}\beta - z'_{j}\hat{\gamma} - (x'_{ik}\beta - z'_{k}\hat{\gamma})\big)\right]\right|. \end{aligned}$$

The first of the two terms above involves the convergence of a continuous finite-sample

objective function to a continuous limit function, and standard laws of large numbers apply. A sample size $N_2(\lambda)$ that makes the probability of the sup difference between large arbitrarily small exists. The second term involves the approximation of a discontinuous function to a continuous function. A $\lambda_1 > 0$ can be chosen to make the approximation arbitrarily close. More formally, $\sup_\beta \left| E\left[\eta_\lambda\left(x'_{ij}\beta - z'_j\hat{\gamma} - (x'_{ik}\beta - z'_k\hat{\gamma})\right)\right]\right|$ is bounded above by $\sup_\beta P\left[\left(x'_{ij}\beta - z'_j\hat{\gamma} - (x'_{ik}\beta - z'_k\hat{\gamma})\right)^2 < \frac{1}{\lambda^2}\right]$, which converges to 0 as $\lambda \to \infty$ if there is not a mass point at the collection of x's such that $x'_{ij}\beta - z'_j\hat{\gamma} = x'_{ik}\beta - z'_k\hat{\gamma}$, for some β. Assumption 1 implies ties in linear index values and hence choice probabilities occur with probability 0.

The third term, $\sup_\beta \left|Q^{\mathcal{K}}_\lambda(\beta) - Q^{\mathcal{K}}(\beta)\right|$, does not even involve data and just involves approximating a discontinuous function with a continuous function. A $\lambda_2 > 0$ can be chosen to make the approximation arbitrarily small, again as long the covariate support is flexible enough that equal linear indices and hence equal choice probabilities occur with probability 0.

Thus, by choosing $\lambda = \max\{\lambda_1, \lambda_2\}$ and then $N = \max\{N_1(\lambda), N_2(\lambda)\}$, the original triangle inequality and hence the definition of uniform convergence in probability are satisfied.

□

B More Details on Subset Estimation Using Maximum Score

There are many theoretical advantages to using maximum score. A major computational novelty is that maximum score is consistent for a fixed subset of choices. This section explores this feature in more detail.

B.1 Using All the Data

A downside of the maximum score objective function in equation (7) is that it only uses observations in $\mathcal{N}(\mathcal{K})$, or the set of observations where the agent picked a choice in $\mathcal{K}$. Throwing away observations is likely to result in a loss of efficiency. If only a small subset of choices in

$\mathcal{J}$ are actually observed in the data, a researcher should pick $\mathcal{K}$ to include all observed choices.

If an otherwise intractable number of choices are observed in the data, a feasible estimator splits all choices into mutually exclusive categories. For example, if J is 1 million and there are 1,000 choices observed in the data, then the researcher could create 100 subgroups, each with 10 observed choices and 10 unobserved choices. Say the researcher has chosen L distinct choice sets $\mathcal{K}_1, \ldots, \mathcal{K}_L$. Define $\mathcal{K}(i)$ to be the choice set that includes the agent i's observed choice, j_i. For i, the estimator will only relate the linear index of j_i to the linear indices of other options in $\mathcal{K}(i)$. In particular, the new maximum score objective function is

$$Q_N^{\mathcal{K}}(\beta) = \sum_{i=1}^{N} \sum_{k \in \mathcal{K}(i) \setminus \{j_i\}} \mathbb{1}\left[x'_{ij}\beta - z'_j\hat{\gamma} > x'_{ik}\beta - z'_k\hat{\gamma}\right].$$

The previous consistency proof easily extends to the new estimator. The choice sets are distinct, and therefore the argument that the limit of the objective function has a unique maximum at the true parameter value can be applied separately to the indicator functions corresponding to each choice set. The uniform convergence property trivially extends to this case as well.

B.2 Monte Carlo Experiments

This section presents Monte Carlo experiments to study the finite-sample properties of the subset maximum score estimator. This section considers the case where the true model does not have endogenous covariates, in order to highlight the computational savings of considering only a subset of the choices in estimation.

For all experiments, for an observation i, data is generated from the random utility model

$$x_{ij1} + \beta_2 x_{ij2} + \epsilon_{ij}, \quad \text{for } j = 1, \ldots, J,$$

where ϵ_{ij} is an i.i.d. across observations and choice random variable with a given distribution. The covariates x_{ij1} and x_{ij2} are i.i.d. $N(0,2)$. The true parameter value is always $\beta_2 = 1$.

As choice behavior alone cannot identify the cardinalization of utility functions, discrete choice methods require location and scale normalizations. We impose the normalizations that the mean of ϵ_{ij} is equal to Euler's constant ($\gamma \approx 0.577$), the variance of ϵ_{ij} is equal to the logit variance ($\pi^2/6 \approx 1.65$), and the coefficient on x_{j1} is 1. The first two normalizations correspond to those made in the traditional logit maximum likelihood estimator.[39] The coefficient normalization is more common in semiparametric estimators.[40]

The distribution of the error term ϵ_{ij} is mixed normal, which in this case is a bimodal, asymmetric distribution. Figure 3 displays a plot of the probability density function, and gives details of the exact parameters chosen. Again, the parameters are chosen to match the mean and variance of the extreme value distribution in the logit. The mixed normal density is chosen specifically so the parametric logit estimator will be misspecified, to highlight the greater robustness of semiparametric estimation.

Table 4 reports the mean finite-sample biases, mean-squared errors (MSE), and execution times from the Monte Carlo experiments.[41] In all experiments, only the parameter β_2 is estimated. The first estimator in Table 4 is the logit maximum likelihood estimator. The second estimator is maximum score estimator when all choices are included. This means there is one choice set equal to the true choice set, $\mathcal{J}$. The table also includes the two estimators that are computationally feasible when the true number of choices is large. The logit sampling estimator was introduced by McFadden (1978). It exploits the independence from irrelevant alternatives property of the logit to allow estimation using agent-specific random choice sets. The logit sampling estimator is compared to the new subset maximum score estimator, where observations are grouped into smaller choice sets. We use the version of the subset maximum score

[39] An alternative normalization is to make the median (not the mean) of ϵ_{ij} equal to its value for the logit ($-\log\log 2 \approx 0.367$), while preserving the variance of $\pi^2/6$. Under one such mixed normal satisfying these alternative normalizations, the finite-sample bias of the logit maximum likelihood estimator is 0.01 farther from 0.0 than the logit ML bias reported in the first row ($N = 100$, $J = 10$) of Table 4.

[40] Technically, a semiparametric estimator can estimate the sign on x_{j1} (whether the coefficient is 1 or -1), although the estimate of the sign would converge at such a fast rate that further analysis of its finite sample properties is unhelpful.

[41] The bias of an estimator $\hat{\beta}_2$ is $E\left[\hat{\beta}_2 - \beta_2\right]$, and the mean squared error is $E\left[\left(\hat{\beta}_2 - \beta_2\right)^2\right]$.

estimator that uses information from all observation by creating distinct estimation sets. In the Monte Carlo study, the estimation choice set sizes (K) for both the logit sampling and subset maximum score estimators are equal to 5 or 10, as marked in Table 4.[42]

When examining the execution times, it is clear that both the logit sampling estimator and the subset maximum score estimator with choice sampling fulfill their primary purpose of requiring less computer time for estimation than methods that require itemization of the entire choice set. In fact, performing 1000 replications of the subset maximum score with the full-choice set is burdensome enough that we do not investigate its small-sample properties when the number of choices is greater than 10. For all estimators, and for a given number of observations, execution time is roughly equal to a constant times the number of choices considered in estimation for each observation. The logit without choice sampling estimator takes 19 seconds with 500 observations and 1000 choices. By extrapolation, a model where there are 1 million choices and only 1 parameter might take five hours. Optimization routines suffer from a curse of dimensionality in the number of parameters, so a multivariate logit model with 1 million choices and five covariates might take days or weeks to estimate.[43]

Table 4 shows that the maximum score estimators have lower levels of finite-sample bias than the logit estimators. For example, in the first row ($N = 100,\ J = 100$), the logit sam-

[42] In the sampling logit, the choice set for an agent is its observed choice plus 4 or 9 other choices, randomly sampled from all possible choices. The other choices are included with uniform probabilities. In the subset maximum score with distinct choice sets estimator, the total set of observed (in the data) choices is carved up into J/K distinct estimation choice sets. Half of the observations in each choice set are (deterministically) taken from choices seen in the data, and half are randomly sampled from all choices. Estimation proceeds by assigning each agent the estimation choice set that includes the agent's observed choice.

[43] The Monte Carlo experiments are not designed to compare the execution time of the sampling logit and subset maximum score estimators. The logit maximum likelihood estimator uses a gradient-based method compared to the global optimization routine (a genetic algorithm) employed by the subset maximum score estimator. For the same optimization problem, the global routine typically takes longer, although execution time is sensitive to the starting values for the local routine used by the logit. In practice, the logit model will require the estimation of one more linear index parameter than the subset maximum score estimator, as semiparametric estimators make location and scale normalizations on the parameter space, rather than the distribution of the unobservables. As optimization routines suffer from a curse of dimensionality in the number of parameters, in practice the logit might take longer than the maximum score. The maximum score estimator should be computed twice, once for the first parameter $b_1 = 1$, and again for $b_1 = -1$. The set of estimates corresponding to the minimum of the the two sets of estimates should be kept. However, there is no curse of dimensionality in executing an optimization routine multiple times. The execution time of both programs would be improved by coding them in C or Fortran. Further, hardware-specific optimizations to the $\exp$ function and hand-coding analytic derivatives should speed up the logit.

pling estimator has a bias of 0.09, while the bias of the semiparametric estimator is an order of magnitude lower, at 0.01. The most likely explanation is that the semiparametric estimator is consistent under the mixed normal distribution, while the logit is inconsistent. The mean squared errors of the logit is sometimes smaller (0.021 vs. 0.039 in the first row), probably because the logit assumes additional structure that makes continuous choice-probability predictions, rather than just discrete yes-or-no predictions about choices in the semiparametric case. However, the any additional precision of the logit does not contribute much to accuracy, as the logit is precisely estimating a biased coefficient.

Figure 1: Winning Bids by the Population of the 493 C Block Licenses

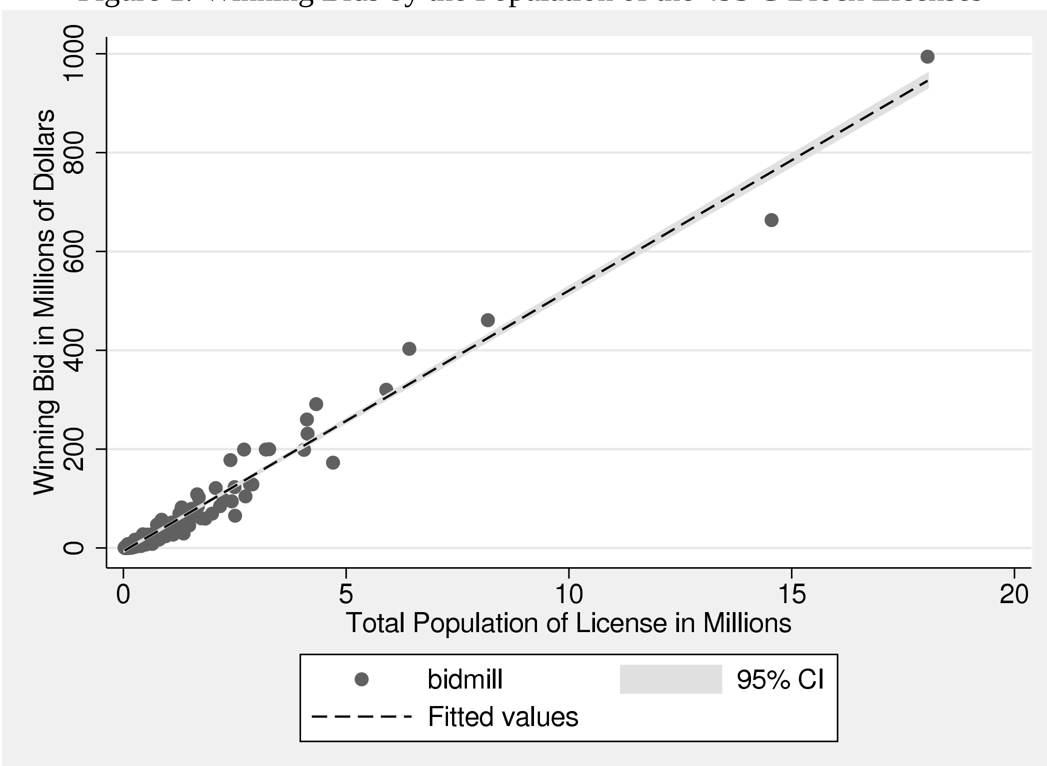


Figure 2: Winning Bids per Resident by the Population of Licenses with Fewer than 5 Million Residents

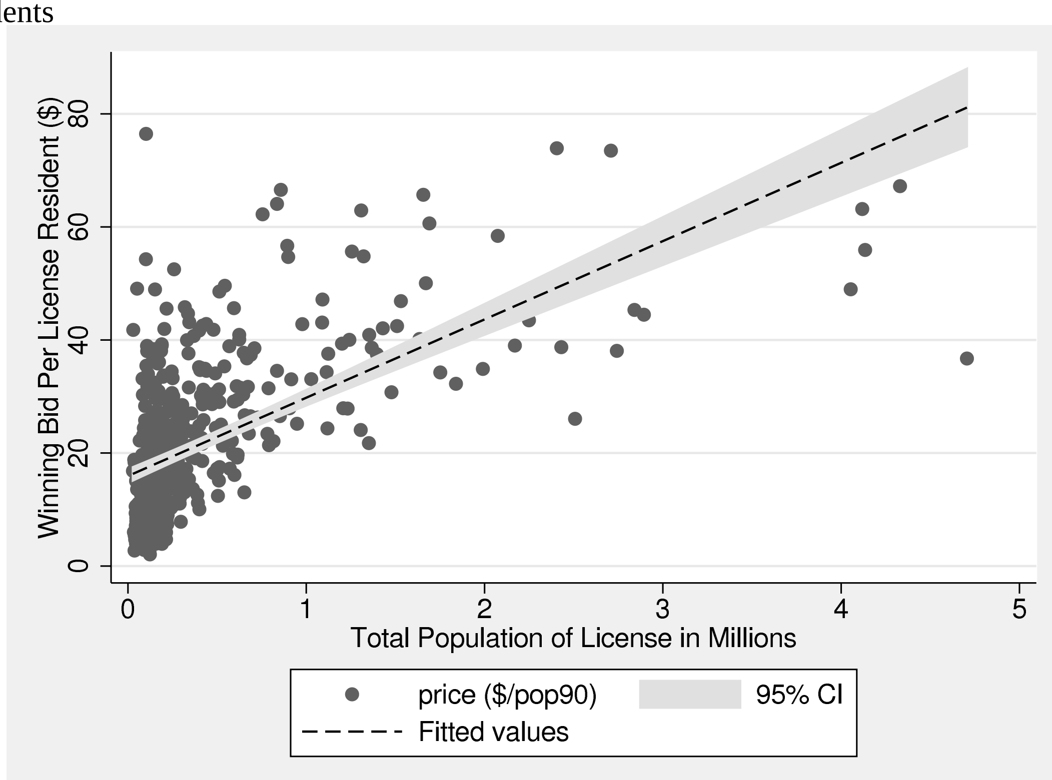
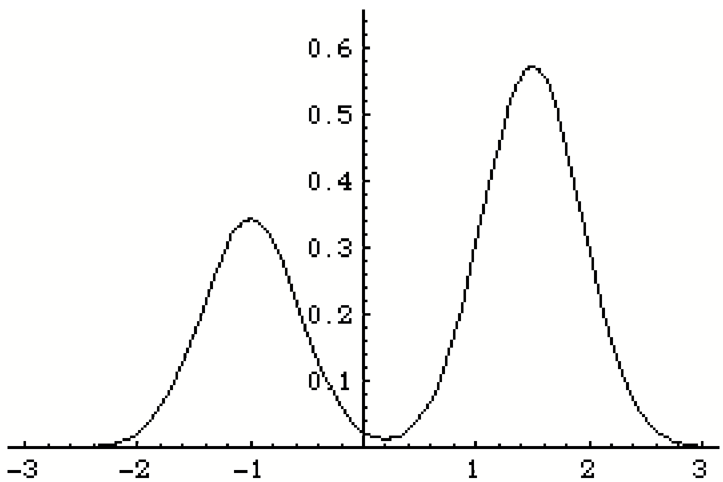


Figure 3: Mixed Normal Density Function Used in Monte Carlo Experiments



This figure is a plot of the density function used in the Monte Carlo experiments in Table 4. The density is a mixed normal with the specific form $0.369 \cdot N(-1.0, 0.184) + (1 - 0.369) \cdot N(1.5, 0.193)$. These parameters are chosen to create an asymmetric, bimodal distribution with the same mean and variance as the extreme value distribution assumed by the logit.

Table 1: Characteristics of Winners and non-Winners of Packages in the Continental United States for the C Block Auction

	Winners		non-Winners	
Characteristic	Mean	Stand. Dev.	Mean	Stand. Dev.
Initial Eligibility (millions of residents)	9.77	27.2	5.15	18.5
Assets ($ millions)	13.1	21.8	12.3	18.8
Revenues ($ millions)	40.7	67.8	39.9	72.3
# of bidders	85		170	

Table 2: Total Closing Prices and Scale and Scope Proxies of 85 Winning Packages in the Continental United States for the C Block Auction

Characteristic	Mean	Standard Deviation	Min	Max
Total price ($millions)	116.2	496.1	0.102	4,201
Total population in 1994 (millions)	2.91	10.93	0.027	93.8
Population with income >$35K (millions)	1.60	6.39	0.013	55.0
Population / distance two markets in a package (millions of people/distance in km)	0.175	0.866	0	7.37
Trips between markets in a package in the American Travel Survey	1,240	8,770	0	79,500
Total trips between airports in markets in a package (thousands)	321	2010	0	18,130

Table 3: Median LTE Subset Maximum Score Estimates of Terminal Node Payoffs in the C Block Auction, with 95% "Confidence Intervals"

Characteristic	(1)	(2)	(3)
Total population in 1994 (millions)	21.3	23.1	19.8
	[21.0,21.6]	[23.0,23.2]	[19.7,20.2]
Population with income >$35K (millions)	-46.9	-58.5	-54.7
	[-47.4,-46.2]	[-58.9,-57.5]	[-55.8,-54.3]
Population with income >$35K	7.87	7.64	6.42
* initial eligibility (millions)	[7.78,7.90]	[7.53,7.68]	[6.30,6.49]
Population with income >$35K squared			-14.2
			[-14,8,-13.7]
Population with income >$35K squared			0.685
* initial eligibility			[0.653,0.723]
Population * distance two markets in a package		-63.0	153.9
(millions / kms)		[-69.9,-59.0]	[148.3,157.4]
Population * distance two markets in a package		14.8	71.3
* initial eligibility		[14.5,15.7]	[69.9,75.8]
% Score of Correct Predictions in Objective Function	80.8%	81.0%	83.2%

Table 4: Monte Carlo Calculations of Finite-Sample Properties of Logit and Maximum Score Multinomial Discrete Choice Estimators under Full Choice Sets and Estimation Subsets: True Model is i.i.d. Multinomial Mixed Normal

N	# of True Choices	Logit Maximum Likelihood			Maximum Score			# of Est. Choices	Logit Sampling			Subset Maximum Score		
	(J)	Bias	MSE	Time	Bias	MSE	Time	(K)	Bias	MSE	Time	Bias	MSE	Time
100	10	0.072	0.014	0.035	0.006	0.019	1.09	5	0.089	0.021	0.018	0.008	0.039	0.480
	100	0.178	0.036	0.314	-	-	-	10	0.178	0.046	0.039	0.025	0.046	1.16
	1000	0.257	0.067	3.25	-	-	-	10	0.198	0.072	0.043	0.084	0.142	1.27
500	10	0.071	0.007	0.207	0.004	0.006	6.85	5	0.082	0.009	0.121	0.009	0.010	2.93
	100	0.178	0.033	1.72	-	-	-	10	0.169	0.031	0.219	0.008	0.013	6.76
	1000	0.253	0.065	19.3	-	-	-	10	0.151	0.0287	0.233	0.015	0.030	7.91

In all cases the data generating process has an i.i.d. error term that is mixed normal with mean 0.5772 and variance $\frac{\pi^2}{6}$, which are the mean and variance of the extreme value distribution used by the logit. 1000 replications are performed for all experiments the reported time is the mean number of seconds to perform one replication for that estimator. For each replication, the four estimators use the same fake data.

The logit estimates are computed using a gradient-based maximization routine. The maximum score estimates are computed using a genetic algorithm because the objective function is a step function. Computations are done in Mathematica 5 for (32-bit) Windows on a machine with a 2.2 GHz Athlon 64 CPU.