A Semiparametric Discrete Choice Model: An Application to Hospital Mergers*

Devesh Raval Federal Trade Commission

Ted Rosenbaum Federal Trade Commission

Steven A. Tenn Charles River Associates

March 7, 2017

Abstract

We propose a computationally simple semiparametric discrete choice estimator to model rich consumer heterogeneity. We assume groups of observably similar consumers have similar preferences, but allow preferences to vary freely across these groups. Model flexibility is easily adjusted by setting a single tuning parameter; we suggest a cross-validation method to do so. We analyze the model's properties in the context of hospital mergers, both analytically and via a Monte Carlo analysis. The model performs well for policy relevant substitution and welfare measures, even if misspecified, when the tuning parameter is set within the neighborhood of the value chosen by cross-validation.

JEL Codes: C14, D12, I11, L41

Keywords: hospitals, patient choice, demand models, semiparametric, mergers

*Thanks to Liz Swigert for her excellent research assistance and Mark Shepard and Eduardo Souza-Rodrigues for very helpful comments. The views expressed in this paper are those of the authors. They do not necessarily represent those of the Federal Trade Commission, any of its Commissioners, or Charles River Associates.

Raval: Economist, Federal Trade Commission, 600 Pennsylvania Avenue NW, Washington, DC 20580. Phone 1-202-326-2260, E-mail draval@ftc.gov

Rosenbaum: Economist, Federal Trade Commission, 600 Pennsylvania Avenue NW, Washington, DC 20580. Phone 1-202-326-2275, E-mail trosenbaum@ftc.gov

Tenn (Corresponding Author): Vice President, Charles River Associates, 1201 F Street NW, Suite 800 Washington, DC 20004. Phone 1-202-662-3806, E-mail stenn@crai.com

1 Introduction

Individual level data on consumer choices can significantly improve predictions of consumer substitution patterns (Berry et al. (2004)). Over the past twenty years, these data have become increasingly available to researchers, a trend that seems likely to continue into the future. Individual level data facilitate the estimation of models with rich individual heterogeneity, which can yield more accurate out-of-sample predictions of individual choices than models that are less flexible along this dimension (Raval et al. (2015)).

However, the estimation of more flexible models can be fraught with statistical and computational concerns. In a context similar to ours, Ho and Pakes (2014) note that estimating models with many dummy variables via maximum likelihood can lead to an incidental parameters problem. Beyond these econometric concerns, estimating this type of model involves computational issues that are likely to consume significant researcher time (Greene (2004)).

We outline a computationally light semiparametric estimator to flexibly estimate consumer choice probabilities, substitution patterns, and welfare.[1] In our estimator, an iterative procedure groups consumers with similar characteristics across multiple dimensions. In order to estimate choice probabilities, we assume that agents' choice probabilities are proportional to the share of the chosen option within a group. To compute substitution patterns and welfare, we further assume that agents substitute in proportion to choice shares within the group. After imposing this additional assumption, our estimator can be viewed as equivalent to estimating a highly flexible multinomial logit model.

[1] This estimator is also easy to implement via the MapReduce algorithm for parallelization, which would reduce computational costs even further.

Our estimator has one major tuning parameter, the minimum group size. The minimum group size parameter balances the bias-variance trade-off between greater bias from larger groups and greater variance from smaller groups. This parameter can be adjusted easily and transparently to make the estimator more or less flexible depending on the characteristics of a given dataset. If we set this parameter equal to one, our estimator collapses to a standard frequency estimator that would group all individuals with the same characteristics. To set the minimum group size, we propose using "leave one out" cross-validation. With large numbers of observations relative to the number of characteristics, the number of people in each group is likely to be sufficiently large such that the share of each group that selects a given choice can be precisely estimated for a large number of groups.

We apply the estimator to hospital discharge data and estimate both choice probabilities and proxies for the change in market power after a merger. In particular, we use the estimator to compute two measures that are widely used as proxies for insurer leverage in merger analysis: diversion ratios, which summarize consumer substitution between the merging providers, and willingness to pay, which is a measure of welfare (Shapiro (1996), Capps et al. (2003), Farrell et al. (2011), Gowrisankaran et al. (2015)). In our application, our estimates of the post-merger change in market power are both large and not sensitive to the group size tuning parameter so long as it is set within an intermediate range. However, the percent change in willingness to pay is more sensitive to the minimum group size than the diversion ratio.

We then examine our estimator's robustness to misspecification using a Monte Carlo exercise. For an intermediate range of the minimum group size, the estimator has a low rate of error for the diversion ratio, percent change in willingness to pay, and individual choice

probabilities across different specifications of the "true" model, including a semiparametric model with extremely large preference heterogeneity and a parametric logit model. However, we find greater sensitivity to the minimum group size parameter for predicting individual choice probabilities and the percent change in WTP. We also find only modest increases in error for much lower sample sizes than in our main specification. Thus, the estimator performs well if it is not implemented in an overly flexible or inflexible manner.

We do find, however, that the estimator's performance depends upon the order of grouping variables. The performance is considerably worse when, given a true model with location variables as the first variables used for grouping, location variables are the last variables to be used for grouping. Therefore, in order to implement the estimator, researchers will need information on the best ordering of these grouping variables. While researchers will need some knowledge of the data to ensure that the covariates that are likely to be most important in explaining consumer choice are given priority, we show that cross-validation can assist in picking the best grouping order from a set of alternatives. In addition, a lower minimum group size can partially compensate for an incorrect variable ordering.

Our estimator builds on both frequency and multinomial logit estimators. Compared to a simple nonparametric frequency estimator, we provide an algorithm to allocate individuals with different characteristics into groups. Compared to the empirical literature using multinomial logit estimators, we provide a computationally simple way to estimate flexible substitution patterns with rich microdata. Our estimator allows researchers to allow a large number of dummy variables in a logit framework without the computational concerns outlined in Greene (2004). In this sense, our estimator is semiparametric; while we assume a logit error in order to predict substitution patterns and welfare, the estimator provides

nonparametric estimates of choice probabilities.

Our estimator is also complementary to a machine learning decision tree approach (Breiman et al. (1984)). Both approaches segment the data into a large number of groups, and estimate predicted probabilities for each group. A decision tree approach does not require knowledge of the correct order of variables for grouping. However, the decision tree's use of in-sample model fit to create groups creates a pre-test bias, which invalidates conventional inference procedures.[2] Because measures of variable importance from a decision tree could be used as an alternative to cross-validation or domain knowledge to select the ex-ante order of variables for our approach, the decision tree approach provides a useful complement to our estimator.

Some previous work has used the estimator we outline here. Carlson et al. (2013) discuss how a version of our estimator has been used for policy analysis, but do not examine its sensitivity. Raval et al. (2015) examine the performance of a version of the estimator using a set of natural experiments. In particular, they find that following the exogenous elimination of a choice from the choice set, a version of our approach does a better job of predicting consumer substitution patterns than the parametric specifications that are frequently used in the literature. In contrast to those papers, the main contribution of this paper is to formally outline the estimator and examine its sensitivity to the tuning parameter, model misspecification, and the size of the dataset.

Beyond health care, our estimator can be used in any situation where a researcher would use a discrete choice approach to model the counterfactual elimination of an option from

[2]Econometricians are actively developing techniques to allow inference for regression tree models; see Athey and Imbens (2015), for example.

a choice set. It can be directly applied to merger analysis in industries, such as medical devices or television, where prices are determined by bargaining between a supplier and an intermediary (Grennan (2013), Crawford and Yurukoglu (2012)). More broadly, the estimation of the Upward Pricing Pressure ("UPP") of a merger, which requires the diversion ratio in response to a small price change, can use this counterfactual diversion ratio given certain assumptions on demand (Farrell and Shapiro (2010), Conlon and Mortimer (2013), Jaffe and Weyl (2013)).[3] In addition to modeling the elimination of an option from a choice set, when researchers have access to data on product characteristics, they can project these characteristics off of the estimated mean utilities in order to conduct counterfactuals in which product characteristics change.

Our approach to demand estimation is most useful when researchers have available detailed information on the demographics relevant for consumer decision-making, when the dataset is large, and when product characteristics are not particularly informative of choice. Many markets outside of health care fit this description. For example, online markets for goods may have limited information on product attributes other than a "star rating," but have detailed information on consumers based on their browsing behavior or their previous purchases. In retail markets such as grocery stores or pharmacies, geographic differences are the main source of individual heterogeneity in purchasing patterns and are often available through loyalty card data, whereas little information may exist on differences in store quality. In addition, our approach allows for a more granular conception of distance – through zip codes or census blocks – which may better represent preferences than travel time, as is

[3]The diversion ratio calculated based on eliminating a choice from the choice set is equal to the diversion ratio in response to a small price change under linear demand or the representative consumer logit model (Conlon and Mortimer (2013)).

typically used in the health care and environmental travel cost literatures.

The paper proceeds as follows. In Section 2 we outline our approach, in Section 3 we introduce our empirical application, and in Section 4 we show results for our Monte Carlo simulation. We conclude in Section 5.

2 Logit Choice Models

In this section, we first detail the parametric logit choice model typically used in the hospital choice literature, and then introduce our semiparametric estimator. We discuss our approach in the context of a patient's choice of provider, but it can be applied in any multinomial choice context where one has access to individual level data.[4]

2.1 Parametric

Each patient i chooses the specific hospital j from the set of hospitals J. The utility u_{ij} that patient i receives from choosing hospital j is specified as follows:

$$u_{ij} = \delta_{ij} + \epsilon_{ij}. \tag{1}$$

Utility u_{ij} is determined by mean utility δ_{ij} and an i.i.d. error term ϵ_{ij} that is distributed Type I extreme-value. Each patient selects the utility maximizing option from the set of hospital choices J.

While mean utility δ_{ij} may be determined by numerous factors, the literature usually

[4]See Ackerberg et al. (2007) for further discussion of these models in general and Gaynor et al. (2015) for further elaboration in the health care context.

assumes a functional form in which δ_{ij} depends on patient characteristics, hospital characteristics, and travel costs between patient i and hospital j. Hospital characteristics control for the set of services offered at a given hospital and the overall quality of those services. Travel costs capture the fact that patients are, all else equal, more likely to select hospitals close to where they live. Patient characteristics, such as an individual's medical condition and demographics, are interacted with travel costs and hospital characteristics. For example, a patient in labor may be more likely to go to a hospital with a labor and delivery room. Under the assumed error structure, the ex-ante probability s_{ij} that patient i selects hospital j follows the familiar logistic form:

$$s_{ij} = \frac{\exp(\delta_{ij})}{\sum_{k \in J} \exp(\delta_{ik})}. \tag{2}$$

Given a set of patient and hospital interactions, a functional form for these interactions, and the observed hospital choice for each patient, the parametric logit model can be estimated via maximum likelihood. Capps et al. (2003), Gowrisankaran et al. (2015), and Ho (2006) all estimate parametric logit models with different specifications for δ_{ij}. After recovering the vector of δ_{ij}'s, it is straightforward to compute substitution patterns if any option is eliminated from the choice set.

The difficulty in designing such a hospital choice model lies in how to allow for heterogeneity in preferences across patients. Any two individuals with the same δ_{ij} for all j will have the same substitution patterns following the elimination of a choice from the choice set. Therefore, the degree of flexibility in δ_{ij} determines the flexibility of the allowed substitution patterns. Even with rich individual data, as is available in the hospital context, the

researcher typically specifies a parameterized function.

The degree of flexibility in δ_{ij} affects the accuracy of predicted substitution patterns. Raval et al. (2015) compare parameterizations of δ_{ij} from the literature that allow for varying levels of heterogeneity in the population. Their results show that models that richly account for heterogeneity based on patient observables (more allowed heterogeneity in δ_{ij}) provide for more accurate out-of-sample predictions of individual choices than models that are less flexible along this dimension.

2.2 Semiparametric

With our estimator, we are able to move away from the restrictive parameterizations of δ_{ij} by employing two assumptions. First, we assume that an individual's choice probability for a given facility is equal to that of a group of similar patients. Second, we assume that, within their group, patients substitute proportionally to all alternatives. However, we allow for substitution patterns to freely vary across groups. Therefore, defining these groups of similar patients in a way that there is sufficient power in each group to estimate choice probabilities is key to our approach. We begin with a discussion of how we partition individuals into groups and then discuss how those groups allow us to estimate choice probabilities and substitution patterns.

We consider a case where the researcher has an individual level dataset with c individual characteristics and consumer choices. For example, in our application these individual characteristics include information on demographics, location of residence, and reason for hospital admission. The consumer choice is patients' choice of hospital.

In order to partition individuals into G groups (indexed by g), we place patients having the same values for all c characteristics into the same group. We keep groups that have at least S_{min} observations, where S_{min} is a tuning parameter. For the remaining individuals who are not in groups with at least S_{min} observations, we then repeat this procedure using only the first $c-1$ characteristics. The excluded characteristic is determined by the predetermined ordering of characteristics. In particular, we order all characteristics according to our beliefs about which characteristics are most likely to predict choices. The first characteristic we eliminate is the one that we think is least important in predicting patient choices. We then iterate on this procedure, reducing the number of characteristics by one each time, until all patients are allocated into groups. For each iteration, we eliminate the characteristic that we believe is least likely to predict choices from the set of remaining characteristics.

In order to compute choice probabilities, we first assume that individuals' choice probabilities are equal to that of the other individuals in their group. Therefore, we can compute an estimate of these probabilities by computing the share of individuals within each group that go to each facility. If one is only interested in choice probabilities, no stronger assumptions are required.

However, more structure is required to estimate substitution patterns or welfare. In order to do so, we assume that a patient's utility for a hospital is equal to mean utility δ_j^g for all individuals i in group g plus an i.i.d. logit error draw, so our approach is equivalent to estimating a multinomial logit model with a dummy variable for each hospital-group combination. When viewed in this light, our key assumption is that observably similar patients – patients within a group – are also unobservably similar except for an i.i.d. logit error. This logit error makes the approach semiparametric. Given the logit assumptions and

a vector of δ_j^g's, the probability that patient i in group g selects hospital j is as follows:

$$s_{i(g)j} = s_j^g = \frac{\exp(\delta_j^g)}{\sum_{k \in J} \exp(\delta_k^g)}. \tag{3}$$

As noted above, we estimate $\hat{s}_j^g$ directly rather than estimating each $\hat{\delta}_j^g$ and using them to compute $\hat{s}_j^g$.

Thus, to apply our approach one needs to set a minimum group size (S_{min}) and have an ordering of characteristics. We discuss the choice of S_{min} here, and defer our discussion of how to order characteristics until Section 3.

The choice of minimum group size, S_{min}, must be set to balance a bias-variance trade-off, which we discuss in Appendix A. A smaller value of S_{min} leads to a more flexible model with possibly many groups of small size, which will lead to large variance but low bias. A higher value of S_{min} leads to a coarser grouping, which will mean lower variance but higher bias. Thus, the minimum group size functions analogously to a bandwidth parameter in kernel density estimation.[5] In our empirical application, we calibrate this parameter using "leave one out" cross-validation.

We have to confront two other issues when implementing this estimator. First, our approach requires discrete variables; therefore non discrete variables, such as age, must be discretized into categories. Second, a small number of ungrouped individuals may remain after iterating across characteristics as described above. If so, the remaining observations can either be grouped together or simply omitted from the analysis.[6]

[5] See Pagan and Ullah (1999) for a comprehensive treatment of kernel density estimation.

[6] In practice, only a few patients are left ungrouped, with only 5 patients (0.004%) remaining ungrouped in our empirical application for a group size of 3 and 1334 patients (1%) for a group size of 50. In our application, we put these individuals together into their own group.

We can calculate standard errors for our estimates using the bootstrap, which is appropriate as long as the minimum group size is not too large. The bootstrap may not be valid for nonparametric and semiparametric estimates because of a potentially slower than $\sqrt{n}$ rate of convergence (Horowitz, 2001). Bootstrapped standard errors are appropriate here since there is a fixed maximum number of groups. Therefore, as the number of observations n goes to infinity, the number of groups remains constant once it reaches this maximum. However, in finite samples, extremely large minimum group sizes may over smooth the data compared to the true distribution of groups, and so lead to a biased estimate of the statistic of interest and the standard errors. We thus do not recommend the use of extremely large minimum group sizes.

2.2.1 Cross-Validation

We propose a "leave one out" cross-validation approach (Stone (1974)) to select the minimum group size.[7] In this approach, the researcher estimates a candidate model m on all of the data, except for one observation. Then the researcher compares the predicted value for that observation from model m to the observed value for that observation. Iterating over all observations yields a vector of predicted values from a given model and a vector of observed values corresponding to each prediction. Using these two vectors, one can compute a measure of the out-of-sample fit for model m.

For our context, we examine model fit for different minimum group sizes. For a given minimum group size, we estimate the model using all of the data except for observation i. This assigns all observations, except for i, to groups and gives a predicted share for all groups

[7] As we show, this cross-validation approach can also be used for variable ordering.

and hospitals. We then regroup all observations, including observation i, into a new set of groups. We take the set of individuals that are in i's group in the "including i" grouping and compute their predicted values from the "excluding i" estimation. We use the average of these predicted values as individual i's predicted choice probability for the cross-validation.

One can use different loss functions to measure the goodness of model fit in the cross-validation. For illustrative purposes, we use root mean squared error ("RMSE") and McFadden's pseudo R^2.

To compute the RMSE for minimum group size m, we take the difference between the predicted and actual choices for each individual and sum over all individuals in the dataset:

$$RMSE^m = \sqrt{\frac{1}{NJ}\sum_{i=1}^{N}\sum_{j=1}^{J}(\hat{s}_{ij}^m - y_{ij})^2}, \tag{4}$$

where y_{ij} equals 1 if individual i chose hospital j and 0 if not, and $\hat{s}_{ij}^m$ is the predicted choice probability using a minimum group size of m.

McFadden's pseudo R^2 is inversely proportional to the log likelihood of the model. To compute the log likelihood for minimum group size m, we take the log of the predicted choice probability for the actual choice:

$$E^m = \frac{1}{N}\sum_{i=1}^{N}\log(\hat{s}_{ij^*}^m), \tag{5}$$

where $\hat{s}_{ij^*}^m$ is the predicted choice probability of the chosen option using a minimum group size of m. This statistic is also called the relative entropy of the model.[8] McFadden's pseudo

[8]Since for many minimum group sizes there are observations where $\hat{s}_{ij^*}^m$ is zero, we use a bottom code (e.g., assign values below a specific threshold to the value at that threshold) when computing the log likelihood.

R^2 then scales the log likelihood to range between 0 and 1 as follows:

$$R^2 = 1 - \frac{E^m}{E^{Intercept}}, \tag{6}$$

where $E^{Intercept}$ is the log likelihood of a model that only includes an intercept, so each hospital is predicted to have a $\frac{1}{J}$ share of the market. An R^2 value of zero indicates a model with the same log likelihood as an intercept only model, and an R^2 value of one perfectly predicts the data.

In our empirical application, we use two different approaches to compute goodness of fit statistics. First, we compute goodness of fit as described above, and use every observation in the dataset for cross-validation (i.e., at some point, every person in the dataset is "left out"). We also apply an alternative approach in which we leave out only a random sample of observations in the data to reduce the computational time required.[9]

3 Application to Hospital Mergers

3.1 Hospital Merger Setting

We examine data from two hospital systems in a mid-sized metropolitan area, where one of the larger systems in the area ("System 1") proposed acquiring one of the smaller systems ("System 2"). For confidentiality reasons we do not reveal the identity of the firms.

Our empirical analysis relies on inpatient discharge data for patients living in the metropoli-

[9]In this approach, we use the full dataset, except for the excluded individual, in the estimation. The only difference is that we do not use all individuals in the dataset for validation.

tan area where the merging parties are located.[10] This dataset contains 124,237 adult commercial admissions.[11] For each hospital admission, we observe patient age, gender, zip code, Diagnostic Related Group ("DRG"),[12] Major Diagnostic Category ("MDC"),[13] and whether the admission was an emergency.

We group admissions using the iterative procedure detailed in Section 2.2. For our baseline specification, we select the variable order based on two criteria. First, we put each type of variable in descending order of its likely importance in determining hospital choice. That way, to the extent necessary to maintain sufficient group sizes, individuals that differ with respect to less important types of variables are pooled together first. We assume patient location is the most important predictor, followed by admission type and patient demographics. Our second criterion is that, within each variable type, we order the characteristics from the least to most detail. This allows a finer measure to be employed when group sizes are sufficiently large (e.g., DRG), but a coarser measure for smaller groups (e.g., MDC).

In order, the variables used to group admissions are as follows:

1. Patient Location (L)
   (a) County
   (b) Zip code
2. Admission Type (A)
   (a) MDC
   (b) Emergency admission indicator

[10] This area is largely self-contained. The vast majority of patients living in the area are treated there, and few patients who are treated at area hospitals come from outside the region.

[11] Children are omitted because the merging parties rarely admit them. We also remove patients with psychiatric, substance abuse, or rehabilitation diagnoses, and patients transferred out of a hospital to another acute care facility.

[12] The DRG system is a widely employed method of classifying hospital cases which contains hundreds of different "services" that a hospital may offer.

[13] The MDC system groups DRGs into 25 mutually exclusive categories. We aggregate a small number of MDC groups with very few admissions.

   (c) DRG type (medical vs. surgical)
   (d) DRG weight quartile[14]
   (e) DRG
3. Patient Demographics (D)
   (a) Age group (18-45, 46-62, and 62+)
   (b) Gender

3.2 Statistics of Interest

Antitrust agencies assessing the likely competitive impact of a proposed merger use hospital choice models to calculate measures of substitution between the hospitals (Farrell et al. (2011)). We focus on two widely employed statistics. The first is a substitution measure known as the "diversion ratio" (Shapiro (1996)). The diversion ratio from hospital h to hospital j measures the fraction of hospital h's patients who would switch to hospital j if hospital h were removed from the choice set. In the logit context, the diversion from hospital h to hospital j is proportional to hospital j's share relative to the other hospitals in the market:

$$div_{ihj} = \frac{s_{ij}}{1 - s_{ih}}. \tag{7}$$

The overall diversion div_{hj} from hospital h to hospital j is obtained by computing the average patient-level diversion across the set of patients that select hospital h:

$$div_{hj} = \frac{1}{N_h} \sum_i div_{ihj}, \tag{8}$$

[14]DRG weights are a resource intensity measure used by Medicare to calculate hospital reimbursement. DRG weights are a relative measure, defined such that the resource intensity of the average admission equals one. We group highly complex tertiary admissions separately, which we define as those with a DRG weight greater than 2.

where the sum is only over the N_h patients who chose hospital h.

The second statistic that we consider is the post-merger percent change in a metric known as "willingness to pay" ("WTP"). Developed by Capps et al. (2003), WTP measures the reduction in consumers' expected utility from removing a set of hospitals from the choice set. In the logit model, the ex-ante expected decline in patient i's welfare from excluding a set of hospitals $S \subset J$ is as follows:

$$WTP_{iS} = -\ln(1 - \sum_{j \in S} s_{ij}). \tag{9}$$

A patient's WTP is an increasing function of the probability s/he will select a hospital in set S, and equals zero when that probability is zero. Overall WTP is obtained by adding up patient-level WTP across all patients.

The antitrust agencies have used WTP to assess the expected harm from a merger of two hospital systems (Farrell et al. (2011)). The combined system's bargaining position changes post-merger, since it can now threaten to exclude both systems simultaneously from the provider's network. Let WTP_{12} represent the WTP for the combined system, and WTP_1 and WTP_2 for System 1 and System 2 individually. If the two systems are substitutes, then the loss in welfare from simultaneously excluding both systems exceeds the sum of the losses from individually excluding each system. The percentage increase in WTP resulting from a merger between the two systems can then be calculated as follows:

$$\Delta WTP_{12} = \frac{WTP_{12}}{WTP_1 + WTP_2} - 1. \tag{10}$$

This measure has the property that it equals zero when the two systems are not substitutes, and is an increasing function of the level of substitution between the two systems.

In order to estimate a diversion ratio or obtain a finite WTP measure using our approach, group members must differ in the system they chose. While this is generally true for large groups, hospital choice homogeneity becomes more likely in a group consisting of only a handful of individuals. In practice, we exclude admissions where a diversion ratio cannot be calculated from estimates of the choice probability for that group since there was no variation in choices within the group. Further, we bound any measure of WTP by imposing a top code at a share of 95%; if one of the merging partners hits this bound, the bound will imply a zero change in WTP.

3.3 Merger Estimates

We use the 124,237 observations of the hospital discharge data to estimate the semiparametric choice model for a minimum group size S_{min} from 3 to 50,000. Selecting such a wide range of values for S_{min} allows us to compare results for extremely flexible and inflexible models, as well as intermediate specifications. Depending on the choice of S_{min}, admissions are put in between one and 23,157 groups. The most flexible specification has a pseudo R^2 of 0.70, while the least flexible specification has a pseudo R^2 of only 0.18. We then estimate diversion ratios and the change in WTP for the two hospital systems for each value of S_{min}. Given the results of prior research, we assume the variable ordering of LAD, or Location, Admission Type, and then Demographics, as in Table 3.1.

We apply the cross-validation approach described above in Section 2.2.1 to select a value

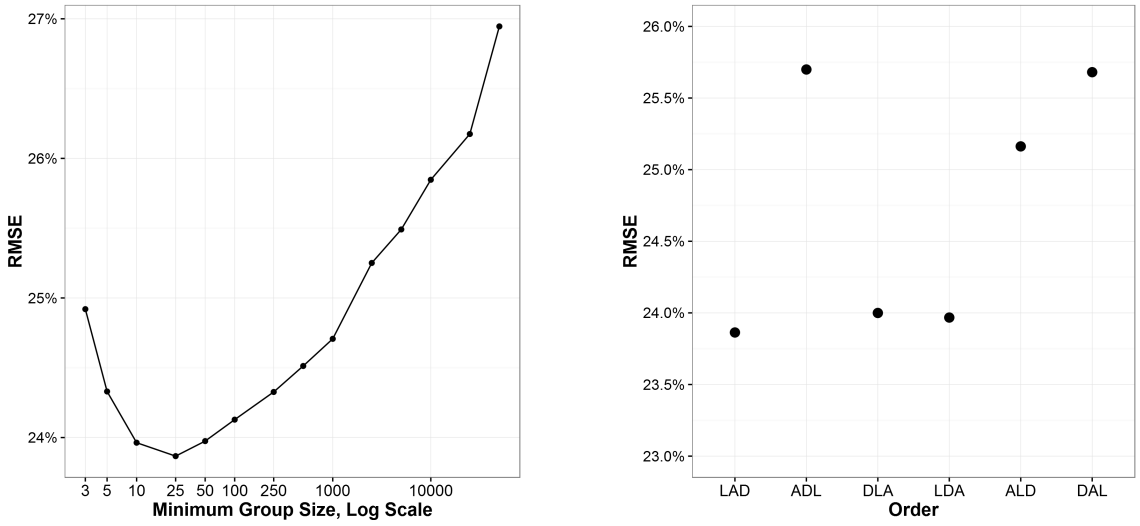
for S_{min} given the LAD ordering. Figure 1a displays the cross-validation results based upon the RMSE when we use all the observations in the data for cross-validation for S_{min}. RMSE is minimized at a minimum group size of 25.

The optimal minimum group size does depend upon the loss function that we use. For example, McFadden's pseudo R^2 is maximized at a minimum group size of 10.[15] For models with a minimum group size under the optimum, the benefit of the increased model flexibility is outweighed by the cost of the lower power of the $\hat{s}_j^g$ estimates. Conversely, for a minimum group size above the optimum, the benefits of the increased model flexibility outweigh the cost associated with lower statistical power. Since the RMSE and McFadden's pseudo R^2 approaches penalize errors differently, they weigh this trade-off differently.

Cross-validation on a very large dataset could take hours of computational time. Therefore, we also conduct a cross-validation for the choice of minimum group size using a random sample of the data, as described above in Section 2.2.1. In this alternative cross-validation, we randomly select 1,000 observations from the data to serve as our validation points. This cross-validation takes only minutes of computational time on a standard desktop computer and yields similar results to cross-validation on the full dataset. This result is reassuring, since it suggests that a cross-validation exercise to suggest the minimum group size can be done quickly.[16]

[15] The McFadden's pseudo R^2 statistic reported in the text uses a "bottom code" of .05, which is consistent with our approach in setting a top code for WTP. A smaller bottom code will heavily penalize choice probabilities at or near zero and is maximized at a larger minimum group size.

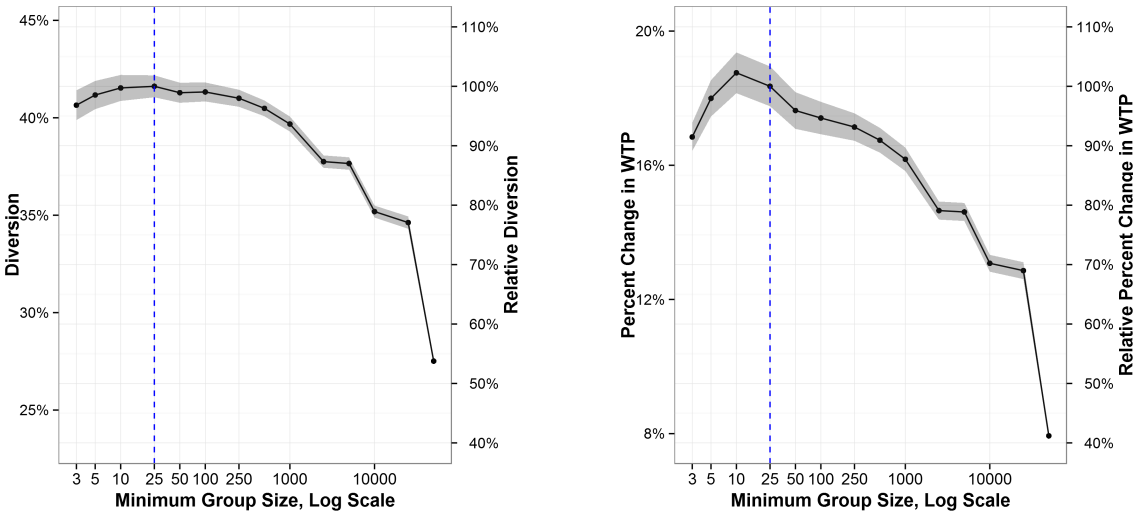
[16] We sample one thousand people from the data fifty times. The results obtained using sampling are similar to those from using the full data set. For RMSE, in 82% of the data samples, a minimum group size of 25 minimizes the RMSE. For the remainder, the RMSE is smallest for a minimum group size of either 10 or 50 (14% and 4% of the data samples, respectively). For the pseudo R^2 measure, in 52% of the samples, a minimum group size of 10 maximizes the statistic. For the remainder, the minimum group size is either 3 (2%), 5 (40%), or 25 (6%).



(a) S_{min} (b) Variable Order

Figure 1 RMSE from Leave One Out Cross-Validation by S_{min} and Variable Order

Note: Cross-validation for the minimum group size uses the variable ordering LAD and cross-validation for the variable ordering uses a minimum group size of 25.



(a) Diversion Ratio (b) Percent Change in WTP

Figure 2 Estimated Diversion Ratio and Percent Change in WTP by S_{min}

Note: Diversion ratio is from System 2 to System 1. The left axis indicates the estimated value of the statistic, and the right axis the estimates scaled by the value of the statistic at a minimum group size of 25. Table I presents these estimates numerically.

We then examine the diversion ratio from System 2 to System 1 and the percent change in WTP for different values of the tuning parameter.[17] Figure 2a depicts this diversion ratio for different values of the minimum group size, while Figure 2b displays the estimated post-merger percent change in WTP. S_{min} is plotted on a log scale. The left axis indicates the estimated value of the statistic, while the right axis indicates the result scaled by the value of the statistic for a minimum group size of 25, the value selected by the RMSE cross-validation approach. For both figures, the shaded region indicates 95% confidence intervals based upon bootstrapped standard errors from 50,000 draws.

Overall, these results suggest that the statistics of interest are relatively insensitive across a range of values for S_{min} in the neighborhood of 25. For a minimum group size between 5 and 500, the diversion ratio is within 4% of the value at 25, and the change in WTP is within 10% of the value at 25. The diversion ratio is less sensitive to the choice of minimum group size than the willingness to pay measure.

As expected, the use of a more flexible specification generally leads to larger standard errors for the diversion ratio. However, since fairly precise estimates are obtained even for small values of S_{min}, the loss of precision from using a more flexible model appears to be relatively small. For the change in WTP, the very low S_{min} specifications also have a fairly low standard error, because the top code on WTP implies that many groups have a zero percent change in WTP.

The fraction of patients which are in groups without system choice heterogeneity is only large when S_{min} is 3, at 12%. Only 5% of patients are in groups without system choice

[17] Results for the diversions from System 1 to System 2 are in Table I. Those results are consistent with our findings.

heterogeneity when the minimum group size is 5, and 1% are when the minimum group size is 10. No patients are in such groups for a minimum group size of 25 or above. Thus, so long as S_{min} is set to at least 5 patients, groups without system choice heterogeneity do not seem to be a major issue.

For all of the above results, we have assumed the LAD variable ordering. In Figure 1b, we conduct cross-validation to confirm that the LAD ordering is appropriate. We examine all six permutations of the major variable groupings of Location, Admission Type, and Demographics (i.e., LAD, LDA, ALD, ADL, DLA, and DAL) given a minimum group size of 25. Reassuringly, we find that LAD is the preferred ordering; the RMSE is similar for LDA and DLA, but much higher for the other three variable orderings for which location variables are relatively late in the ordering of characteristics. This cross-validation exercise thus demonstrates the importance of ordering the characteristics correctly. This ordering can be set using a combination of prior knowledge about characteristics' expected importance in predicting choices together with cross-validation.

4 Monte Carlo Analysis

The results presented in the previous section suggest that estimates of the diversion ratio and percent change in WTP are relatively robust to the minimum group size, and that the diversion ratio is more robust than the percent change in WTP. However, it is impossible to analyze the performance of the semiparametric model from the estimated results since we do not know the true value for these statistics. In this section, we use the obtained estimates to calibrate a model where we know the true value for the diversion ratios and WTP, and assess

whether the semiparametric model can accurately estimate these statistics. In addition, we examine how well the semiparametric model estimates individual choice probabilities.

4.1 Monte Carlo Approach

We undertake the following Monte Carlo analysis to assess the semiparametric model's performance in a real-world setting. Admissions are randomly sampled with replacement from the hospital discharge data employed earlier. For each sampled admission, we randomly generate a hospital choice using the admission's predicted choice probabilities from the semiparametric model estimated in Section 3.3 for a given choice of minimum group size. This procedure results in a realistically calibrated data sample for which we know the true probability that a given patient will select any given hospital, and thus the true diversion ratios and percent change in WTP.

For various model specifications, we use this data generating process to simulate 50,000 datasets that contain the same number of admissions as the original data. Each simulated dataset is used to estimate the semiparametric model for a given choice of S_{min}, as well as choice probabilities of each system for each individual, the diversion ratio, and the percent change in WTP.[18] For the diversion ratio and change in WTP, we then report the RMSE of the percent difference between each estimated statistic and its true value across the Monte Carlo simulations. For the individual choice probabilities, we report the absolute value of the RMSE. Since we employ a large number of simulations, the estimated RMSE should accurately represent the magnitude of the semiparametric estimator's RMSE.

[18]Since we find very similar patterns for the choice probabilities of each system, we depict only the RMSE for the choice probabilities of System 1.

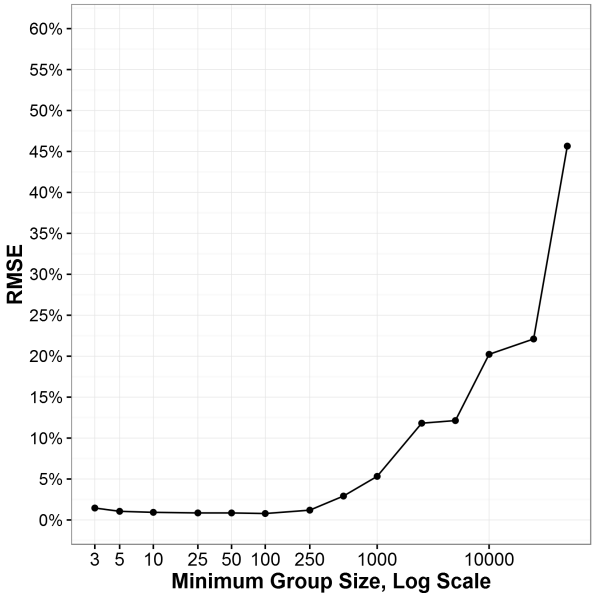
4.2 Different Group Sizes

We start with the true model calibrated based on the $S_{min} = 50$ specification. For all of the Monte Carlo simulations with different group sizes, we assume that the variable ordering is LAD for both the true model and all simulated datasets.

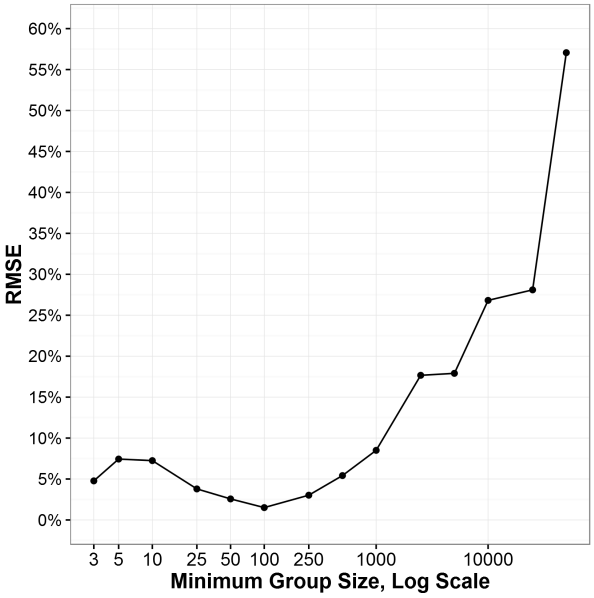
Figure 3 displays the RMSE for the diversion ratio, percent change in WTP, and choice probabilities of System 1 for different values of S_{min}. For the diversion ratio, the RMSE is still fairly low for very flexible models, with only small changes before a minimum group size of 250, but does increase considerably for sufficiently inflexible models. The percent change in WTP and choice probability of System 1 exhibit a U shape, with higher RMSE for very low and very high values of S_{min}. Intuitively, estimates are biased for very inflexible models, while estimates for very flexible models both have high variance, as well as bias due to groups without heterogeneity for the percent change in WTP. For both the diversion ratio and percent change in WTP, the RMSE is at its lowest when S_{min} is 100, while for the choice probability, the RMSE is at its lowest at the true value for S_{min} of 50.

However, all values of S_{min} below 500 have an RMSE less than 1.5% for the diversion ratio, and all values between 3 and 500 have an RMSE below 8% for the percent change in WTP. The RMSE for individual choice probabilities is below 10% for group sizes between 10 and 1,000, compared to an RMSE of 4.7% at the true group size of 50. Thus, errors may be fairly small so long as S_{min} is set within an intermediate range, although the percent change in WTP has a significantly higher error rate than the diversion ratio and the individual choice probabilities have a greater error rate than the other two metrics.

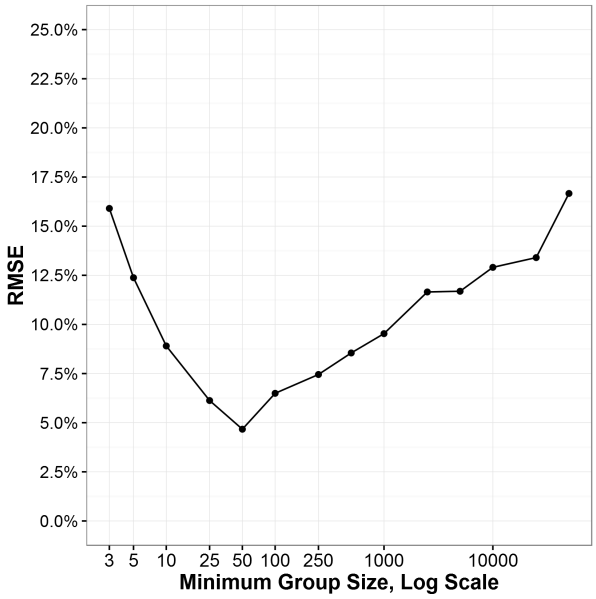
When patient heterogeneity is sufficiently prevalent, it may not be possible to choose a



(a) Diversion Ratio



(b) Percent Change in WTP



(c) Choice Probability

Figure 3 Estimated RMSE by S_{min}, When True Value of S_{min} is 50

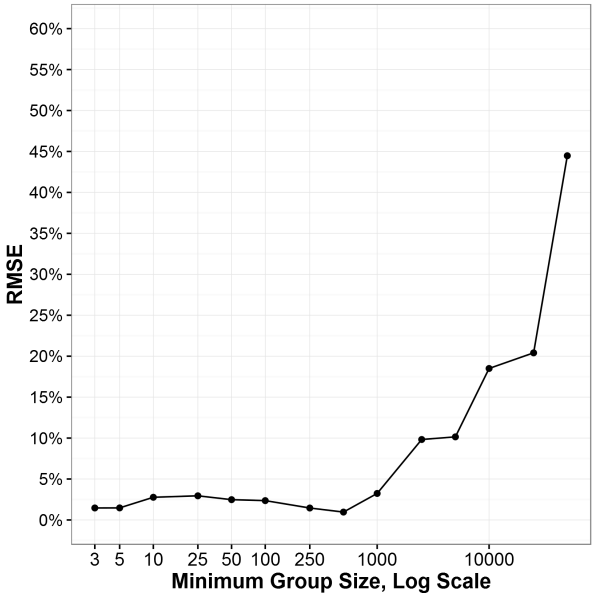
Note: Diversion ratio is from System 2 to System 1. Choice Probability is for System 1. Table II presents these estimates numerically.

value for the minimum group size that is both large enough to avoid the variance associated with small groups but small enough to avoid bias from being insufficiently flexible. To examine this issue, we consider the results of a Monte Carlo analysis in which the data generating process corresponds to the $S_{min} = 3$ calibration.

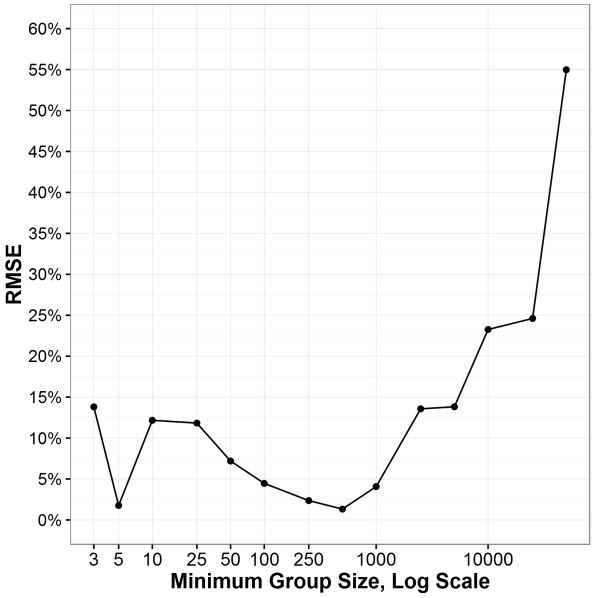
The results from this analysis are displayed in Figure 4. In this case, the RMSE for both the diversion ratio and change in WTP has two local minima, at a minimum group size of 3 and 500, with the global minimum at 500. Again, the RMSE for the diversion ratio is fairly flat below a minimum group size of 1,000, with the RMSE below 3.5% for all values between 3 and 1,000. The RMSE for the percent change in WTP is less stable. It is below 15% for S_{min} between 3 and 5,000, although it is much lower, at 1.33%, at the minimum point of 500. These results suggest that, even if one is concerned that patients may have very heterogeneous hospital preferences, one may not need to use extremely flexible specifications to avoid a high degree of error. The choice of S_{min} appears to be more important for the percent change of WTP than for the diversion ratio.

For the individual choice probability, the RMSE is lowest at a minimum group size of 5 and steadily increases after that value, rising from 14.2% at 5 to 20.0% at 1,000. Thus, for the estimation of individual choice probabilities, it may be more important to set the minimum group size accurately.

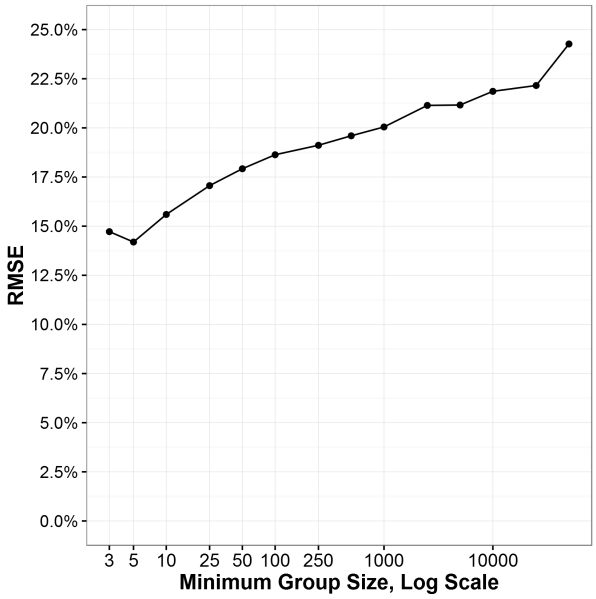
Next, we analyze the efficiency loss from using a semiparametric model. The use of a properly specified parametric logit model will generally provide more precise estimates than the semiparametric model. The degree of inefficiency will depend on the functional form of the parametric specification. The semiparametric estimator is particularly inefficient when the true model has a simple, known parameterization. We consider a Monte Carlo analysis



(a) Diversion Ratio



(b) Percent Change in WTP



(c) Choice Probability

Figure 4 Estimated RMSE by S_{min}, When True Value of S_{min} is 3

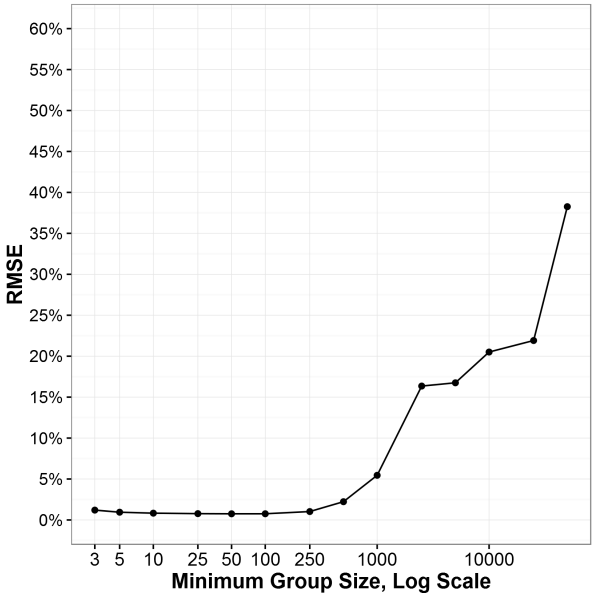
Note: Diversion ratio is from System 2 to System 1. Choice Probability is for System 1. Table III presents these estimates numerically.

where the data generating process is a particularly simple specification to gain a better understanding of the "worst case" scenario for the semiparametric model. First, we use the dataset employed earlier to estimate a parametric logit model that controls only for travel time, its square, and a set of hospital fixed effects. We predict choice probabilities for each patient from the model estimates, which are then used to generate new hospital choices for admissions that are randomly sampled with replacement from the data.

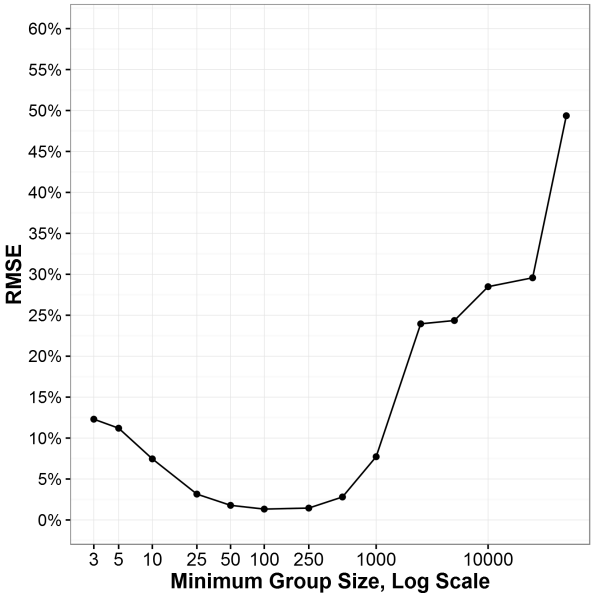
The results from this analysis are presented in Figure 5. We again see a U shaped curve for the RMSE for the percent change in WTP, and a flat RMSE that only rises with high values of S_{min} for the diversion ratio. The RMSE is at its lowest when the minimum group size is 50 for the diversion ratio, and 100 for the percent change in WTP. The RMSE is approximately 1% or less for the diversion ratio for all values of S_{min} below 250. For the percent change in WTP, the RMSE is 3.5% or below for S_{min} between 25 and 500.

We also estimate the RMSE when the model is estimated using the correctly specified simple parametric logit.[19] The RMSE of the parametric logit is 0.5% for the diversion ratio and 0.8% for the percent change in WTP; the RMSE for the semiparametric logit is very close to the RMSE for the parametric logit for intermediate values of S_{min}. Thus, the efficiency loss from using a semiparametric model is low so long as the minimum group size is set within an intermediate range. This analysis considers the performance of the semiparametric model when the true model is an extremely simple parametric alternative. In a realistic setting where the data generating process is more complex, the inefficiency from using a semiparametric estimator is presumably smaller.

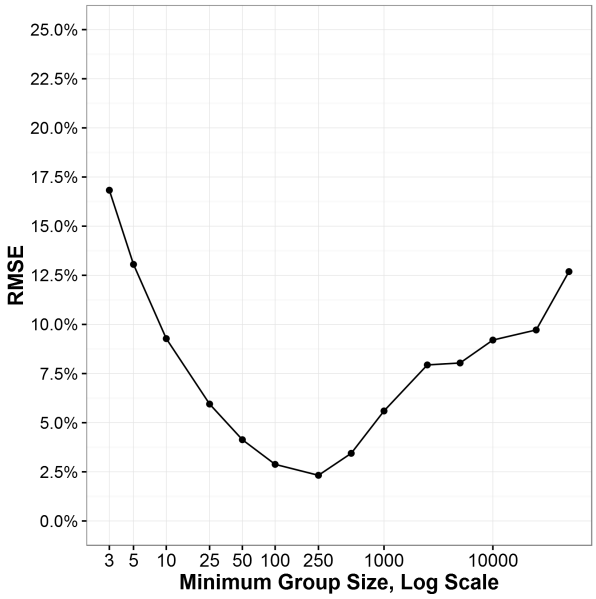
[19]Due to the much higher computation burden, we employ only 1,000 simulations, rather than the 50,000 used for the semiparametric model.



(a) Diversion Ratio



(b) Percent Change in WTP



(c) Choice Probability

Figure 5 Estimated RMSE by S_{min}, When True Model is Parametric Logit

Note: Diversion ratio is from System 2 to System 1. Choice Probability is for System 1. Table IV presents these estimates numerically.

For the individual choice probabilities, the RMSE has a U shaped pattern with a minimum of 2.3% for a minimum group size of 250. However, the RMSE is below 10% for S_{min} between 10 and 25,000. Thus, for estimating individual choice probabilities, the semiparametric model is more sensitive to the choice of minimum group size, but the increase in RMSE for values of S_{min} within an intermediate range of values is moderate.

4.3 Different Grouping Order

In contrast to the choice of the minimum group size, the performance of the semiparametric model is sensitive to the order of the grouping variables. Given that we use nine grouping variables (see Table 3.1), there are a large number of alternative ways to group the variables. We thus examine all six ways to group the broader subheadings of Location (L), Admission Type (A), and Demographics (D), and compare these to a true order in which the order is Location first, then Admission Type, then Demographics, as in Table 3.1, and the true S_{min} is either 50 or 3. Within each of these, we keep constant the ordering of the covariates within each of the broad groupings.[20]

The results from this analysis for each statistic for a true S_{min} of 50 are presented in Figure 6. The RMSE is fairly low for the true LAD order, as well as for DLA and LDA orders, but is an order of magnitude higher for the other three orders. It appears to be crucial for the grouping to place Location variables either first or relatively early in the ordering; the orderings that do badly either place Location last or, for ALD, second with several different variables before the Location variables. Since both the Location and Admission variables are high dimension, it is difficult to control for both when S_{min} is high. By contrast, the

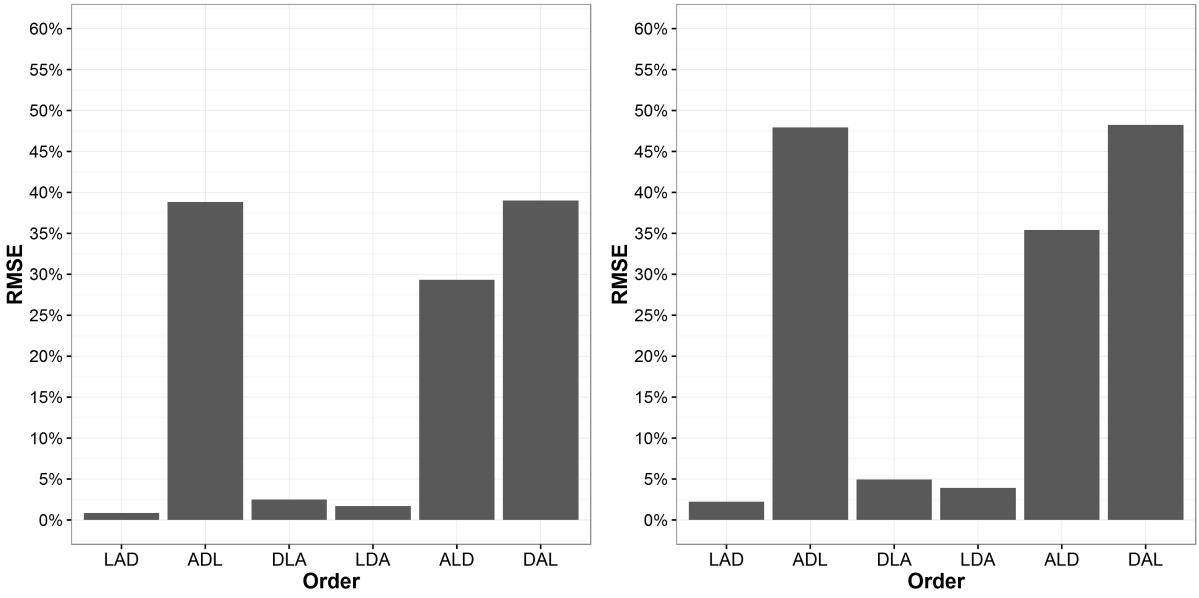
[20]For computational reasons, we only include 5,000 simulations for estimates varying the grouping order.

Demographic variables are low dimension, so the performance of the algorithm is much less affected if Demographic variables come before Location variables. We find a similar pattern for choice probabilities for System 1. Thus, the semiparametric model does require some knowledge of the appropriate grouping variables to function well.

These statistics are significantly less sensitive to the choice of grouping variables when S_{min} is small. Figure 7 depicts the differences in RMSE for all of our statistics for all six orders for a S_{min} of 3. For both the diversion ratio and change in WTP, the true LAD order continues to have the lowest RMSE and the difference between the RMSE for the worst order and the true LAD order falls by slightly more than half compared to a true S_{min} of 50. For the choice probability of System 1, while the true LAD order does have the lowest RMSE, the difference between the RMSE for the worst order and the true LAD order falls by half compared to a true S_{min} of 50. This fall is intuitive. Because a smaller S_{min} size builds groups based on a larger number of variables, the order of the variables matters less. Thus, misspecification of the variable ordering matters less when S_{min} is small.

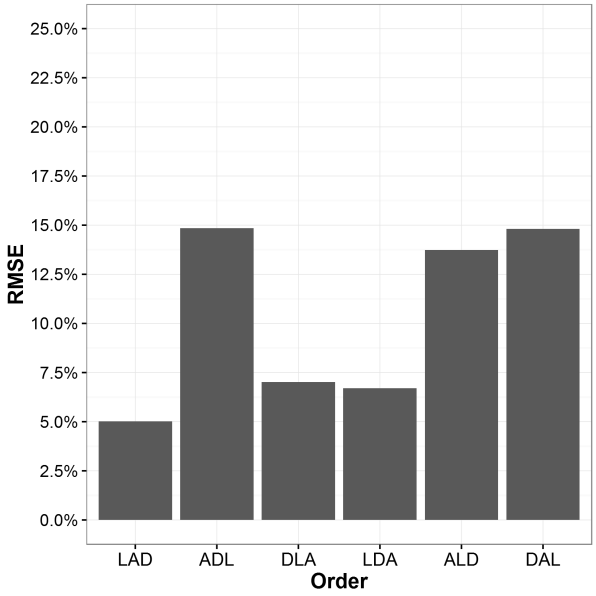
In addition, the optimal S_{min} is lower when the variable ordering is misspecified. In order to illustrate how the error from misspecifying the variable ordering changes as we lower the minimum group size, we set the minimum group size to 3 when the true value is 50 and then examine the performance of each variable ordering.[21] Figure 8 displays the results from this simulation. The differences in RMSE across the different orderings are much smaller; for the diversion ratio, the worst orderings have a RMSE of 13%, compared to an RMSE of 39% if estimated under the true minimum group size. Similarly, the RMSE for the percent

[21] We have also examined the lowest RMSE across all possible values of S_{min} for each variable ordering. For the diversion ratio and percent change in WTP, the lowest RMSE across S_{min} values is similar to the case where S_{min} is 3, except for the true LAD ordering. For the individual choice probabilities, the lowest RMSE across S_{min} values is similar to the case where S_{min} is set to the true value of 50.



(a) Diversion Ratio

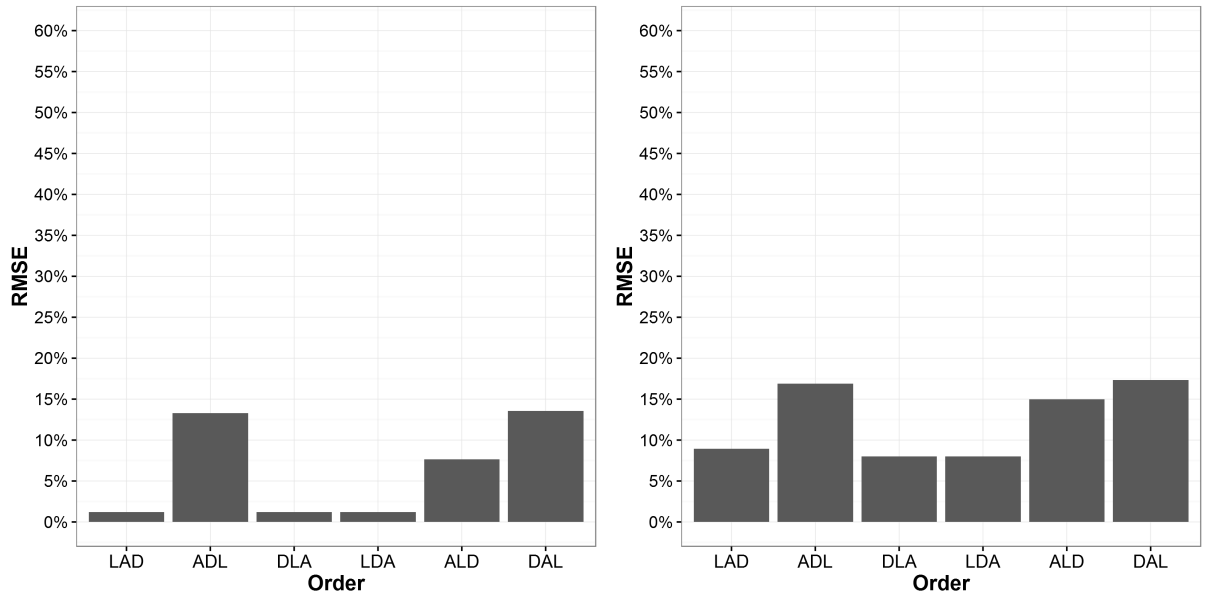
(b) Percent Change in WTP



(c) Choice Probability

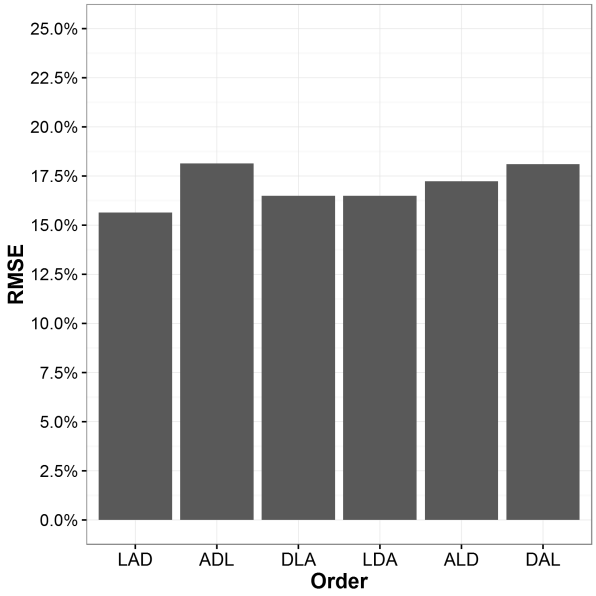
Figure 6 Estimated RMSE by Order of Grouping Variables, $S_{min} = 50$

Note: Diversion ratio is from System 2 to System 1. The true minimum group size is 50. Choice Probability is for System 1. The x axis provides the group order, with L representing Patient Location variables, A Admission Type variables, and D Patient Demographic variables as described in Table 3.1. Table V presents these estimates numerically.



(a) Diversion Ratio

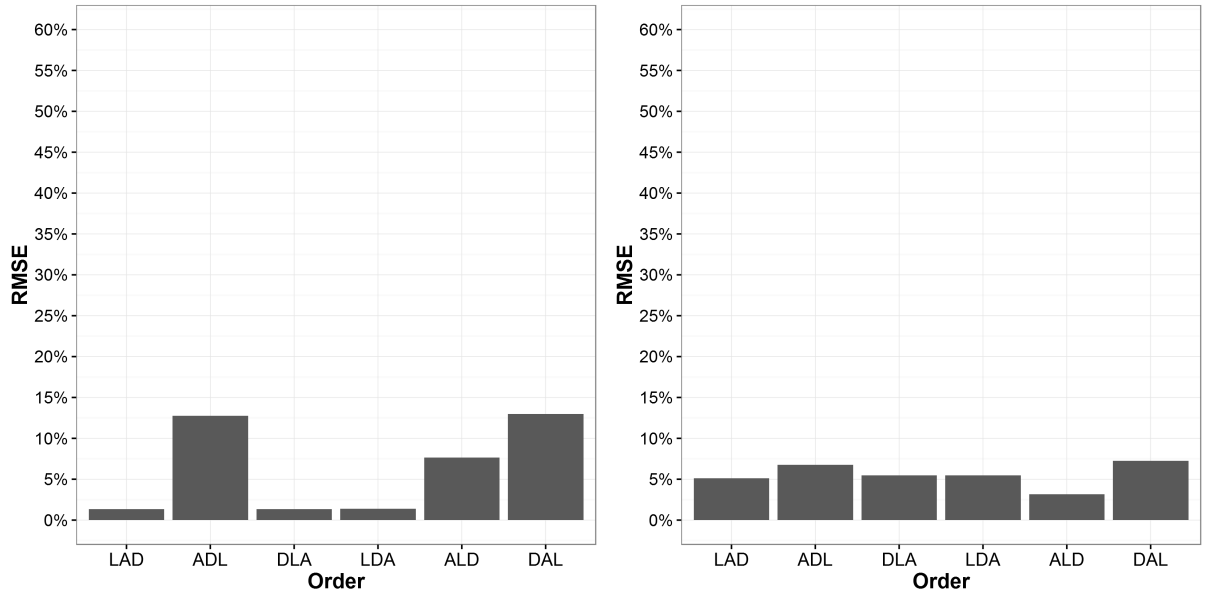
(b) Percent Change in WTP



(c) Choice Probability

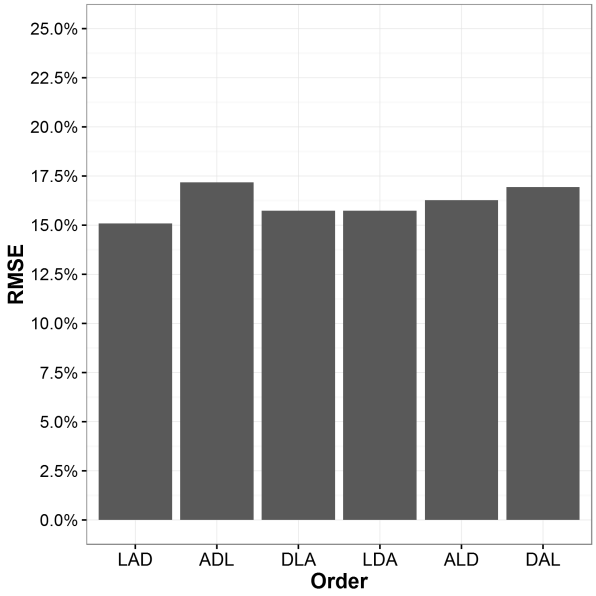
Figure 7 Estimated RMSE by Order of Grouping Variables, $S_{min} = 3$

Note: Diversion ratio is from System 2 to System 1. The true minimum group size is 3. The x axis provides the group order, with L representing Patient Location variables, A Admission Type variables, and D Patient Demographic variables as described in Table 3.1. Table VI presents these estimates numerically.



(a) Diversion Ratio

(b) Percent Change in WTP



(c) Choice Probability

Figure 8 Estimated RMSE by Order of Grouping Variables Under $S_{min} = 3$ When the True $S_{min} = 50$

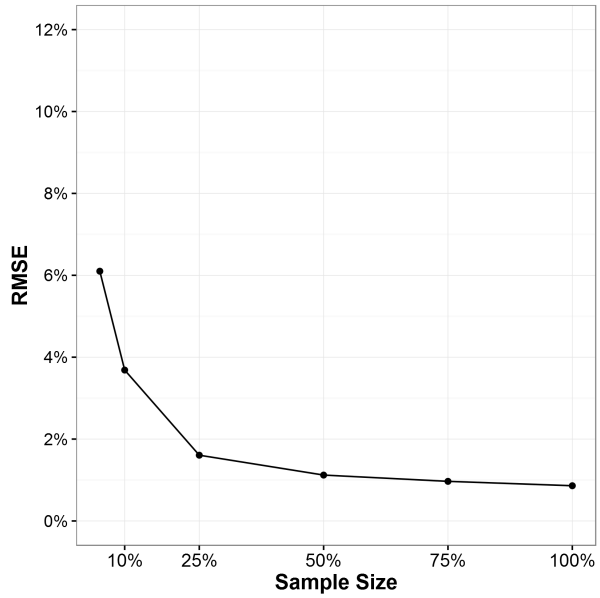
Note: Diversion ratio is from System 2 to System 1. Choice Probability is for System 1. The x axis provides the group order, with L representing Patient Location variables, A Admission Type variables, and D Patient Demographic variables as described in Table 3.1. Table VII presents these estimates numerically.

change in WTP is less than 8% under the worst ordering, much lower than the 48% when estimated under the true minimum group size of 50. Thus, a smaller value of S_{min} can partially compensate for an incorrect grouping order.

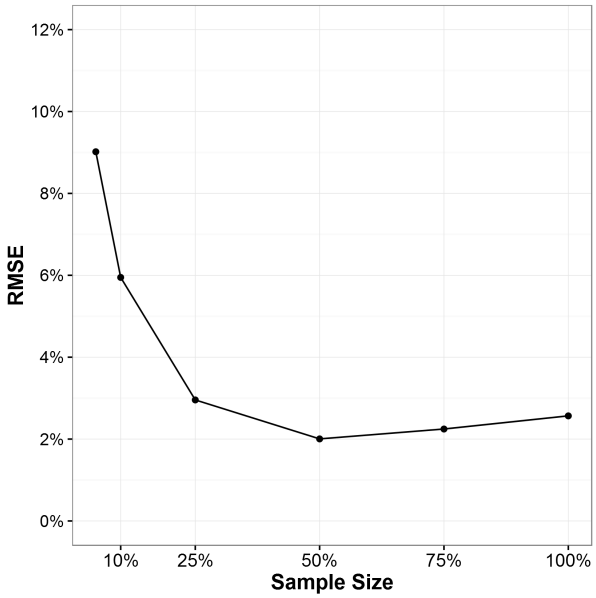
4.4 Different Sample Size

We conclude by considering the performance of the model for different sample sizes. We generate data samples between 5 percent and 100 percent of the original sample size of 124,237. In this Monte Carlo analysis, we estimate a correctly specified model where the data generating process corresponds to the $S_{min} = 50$ specification. The Monte Carlo results presented in Figure 9 suggest that the semiparametric model performs relatively well even when the number of admissions is quite small. The RMSE does rise as the sample size shrinks, but the RMSE is below 4% when the sample size is at or above 10% of the original for the diversion ratio, below 6% for the percent change in WTP, and below 10% for the choice probability of System 1.

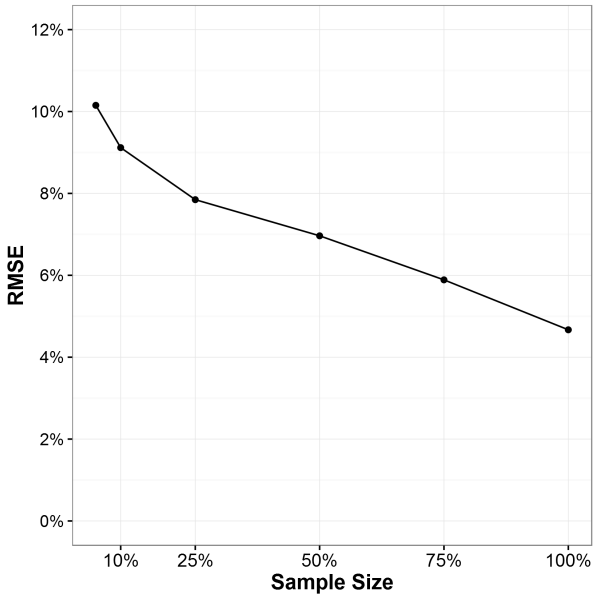
These results suggest that the semiparametric model can be usefully applied in a wide range of settings, even when the dataset contains a small number of observations. One reason why the semiparametric model performs relatively well even for small sample sizes is that the iterative grouping procedure automatically adjusts model flexibility to the size of the data sample. For a fixed value for the minimum group size S_{min}, the grouping procedure puts patients into a smaller number of groups when the sample size is smaller, which highlights an advantage of our approach over a standard frequency estimator. This avoids potential biases associated with using very small group sizes in the semiparametric model, although



(a) Diversion Ratio



(b) Percent Change in WTP



(c) Choice Probability

Figure 9 Estimated RMSE by Sample Size

Note: Diversion ratio is from System 2 to System 1. The true minimum group size is 50. Choice Probability is for System 1. The x axis is the fraction of the overall sample size. Table VIII presents these estimates numerically.

it can lead to bias from model inflexibility if the sample size becomes too small.

5 Conclusion

When presented with rich microdata, researchers must balance the competing objectives of allowing for significant individual level heterogeneity while ensuring statistical power. In the parametric logit models that are typically used, the extent of the permitted heterogeneity is limited by the parametric specification of the model. To complement these methods, we developed a semiparametric discrete choice estimator that allows for rich heterogeneity across the population. Highlighting the importance of allowing for such heterogeneity in choice patterns, Raval et al. (2015) find that our proposed estimator outperforms many parametric multinomial choice models previously used in the literature in predicting choices after a change in the choice set.

In our estimator, the trade-off between heterogeneity and power is determined by a single tuning parameter, the minimum group size. We applied our semiparametric method to patient discharge data and simulated a merger of two hospital systems to test the estimators' sensitivity to this parameter. While we suggested a possible cross-validation approach to choosing this parameter, we found that the main substitution measures are relatively insensitive to the choice of minimum group size. Of the two measures, the change in willingness to pay was more sensitive than the diversion ratio to the choice of minimum group size, while individual choice probabilities were more sensitive to the choice of minimum group size than both substitution measures. The performance of the semiparametric approach was, however, sensitive to the order of the grouping variables.

These results should give researchers confidence to pursue this approach in health care as well as other settings where rich microdata are available. Further, they should encourage research in other methods that relax the functional form restrictions that underlie most empirical work in discrete choice demand modeling. The increased availability of large datasets and recent research in "machine learning" approaches suggest that advances in this area may be on the horizon.

References

Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes, "Econometric Tools for Analyzing Market Outcomes," *Handbook of Econometrics*, 2007, *6*, 4171–4276.

Athey, Susan and Guido Imbens, "Recursive Partitioning for Heterogeneous Causal Effects," *arXiv preprint arXiv:1504.01132*, 2015.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 2004, *112* (1), 68–105.

Breiman, Leo, Jerome Friedman, Charles J. Stone, and R.A. Olshen, *Classification and Regression Trees*, Chapman and Hall, 1984.

Capps, Cory, David Dranove, and Mark Satterthwaite, "Competition and Market Power in Option Demand Markets," *RAND Journal of Economics*, 2003, *34* (4), 737–763.

Carlson, Julie A., Leemore S. Dafny, Beth A. Freeborn, Pauline M. Ippolito, and Brett W. Wendling, "Economics at the FTC: Physician Acquisitions, Standard Essential Patents, and Accuracy of Credit Reporting," *Review of Industrial Organization*, 2013, *43* (4), 303–326.

Conlon, Christopher T. and Julie Holland Mortimer, "An Experimental Approach to Merger Evaluation," *NBER Working Paper*, 2013.

Crawford, Gregory S. and Ali Yurukoglu, "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 2012, *102* (2), 643–685.

Farrell, Joseph and Carl Shapiro, "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition," *The BE Journal of Theoretical Economics*, 2010, *10* (1), 1–39.

_ , David J. Balan, Keith Brand, and Brett W. Wendling, "Economics at the FTC: Hospital Mergers, Authorized Generic Drugs, and Consumer Credit Markets," *Review of Industrial Organization*, 2011, *39* (4), 271–296.

Gaynor, Martin, Kate Ho, and Robert J. Town, "The Industrial Organization of Health-Care Markets," *Journal of Economic Literature*, 2015, *53* (2), 235–284.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town, "Mergers when Prices are Negotiated: Evidence from the Hospital Industry," *American Economic Review*, 2015, *105* (1), 172–203.

Greene, William, "The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects," *Econometrics Journal*, 2004, *7* (1), 98–119.

Grennan, Matthew, "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices," *American Economic Review*, 2013, *103* (1), 145–177.

Ho, Kate and Ariel Pakes, "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, 2014, *104* (12), 3841–3884.

Ho, Katherine, "The Welfare Effects of Restricted Hospital Choice in the US Medical Care Market," *Journal of Applied Econometrics*, 2006, *21* (7), 1039–1079.

Horowitz, Joel L, "The Bootstrap," *Handbook of Econometrics*, 2001, *5*, 3159–3228.

Jaffe, Sonia and E. Glen Weyl, "The First-Order Approach to Merger Analysis," *American Economic Journal: Microeconomics*, 2013, *5* (4), 188–218.

Pagan, Adrian and Aman Ullah, *Nonparametric Econometrics*, Cambridge University Press, 1999.

Raval, Devesh, Ted Rosenbaum, and Nathan E. Wilson, "Industrial Reorganization: Learning about Patient Substitution Patterns from Natural Experiments," *mimeo*, 2015.

Shapiro, Carl, "Mergers with Differentiated Products," *Antitrust*, 1996, *10* (2), 23–30.

Stone, M., "Cross-Validatory Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1974, *36* (2), 111–147.

Table I Semiparametric Logit Estimates for Alternative Minimum Group Sizes

	3	5	10	25	50	100	250	500	1,000	2,500	5,000	10,000	25,000	50,000
Diversion, System 1 to 2	19.7%	19.9%	20.0%	19.9%	19.7%	19.7%	19.9%	19.9%	19.6%	18.7%	18.6%	16.8%	16.4%	10.9%
	(0.32%)	(0.30%)	(0.29%)	(0.26%)	(0.25%)	(0.23%)	(0.21%)	(0.20%)	(0.20%)	(0.18%)	(0.18%)	(0.17%)	(0.16%)	(0.10%)
Diversion, System 2 to 1	42.4%	43.2%	43.7%	43.8%	43.4%	43.4%	42.9%	42.2%	41.0%	38.3%	38.1%	34.6%	33.8%	23.6%
	(0.56%)	(0.53%)	(0.49%)	(0.41%)	(0.38%)	(0.36%)	(0.32%)	(0.30%)	(0.28%)	(0.23%)	(0.23%)	(0.23%)	(0.23%)	(0.13%)
WTP, post-merger % chng	17.4%	18.7%	19.5%	19.0%	18.3%	18.0%	17.7%	17.3%	16.7%	15.1%	15.0%	13.4%	13.1%	7.8%
	(0.23%)	(0.30%)	(0.33%)	(0.33%)	(0.30%)	(0.27%)	(0.22%)	(0.20%)	(0.19%)	(0.15%)	(0.15%)	(0.14%)	(0.14%)	(0.06%)
Number of groups	23,157	13,709	6,932	2,858	1,334	604	209	110	62	12	7	4	3	1
Average group size	5	9	18	43	93	206	594	1,129	2,004	10,353	17,748	31,059	41,412	124,237
McFadden's pseudo R^2	0.70	0.65	0.60	0.54	0.51	0.48	0.46	0.44	0.42	0.37	0.36	0.31	0.29	0.18
% observations without intra-group system choice heterogeneity	12%	5%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Notes: The semiparametric model is estimated using hospital discharge data for 124,237 admissions. Each column reports results for a separate specification that uses the indicated minimum group size, and a variable ordering of LAD. Bootstrapped standard errors based on 50,000 simulations are reported in parentheses.

Table II Monte Carlo Results Varying Group Size, Minimum Group Size of 50 Patients is the True Model

		3	5	10	25	50	100	250	500	1,000	2,500	5,000	10,000	25,000	50,000
Diversion, System 2 to 1	Avg estimate	42.94%	43.26%	43.35%	43.39%	43.42%	43.36%	42.95%	42.12%	41.06%	38.24%	38.10%	34.59%	33.78%	23.56%
	Avg % Error	-0.96%	-0.21%	-0.01%	0.09%	0.15%	0.01%	-0.94%	-2.84%	-5.29%	-11.80%	-12.13%	-20.22%	-22.09%	-45.65%
	SD of estimate	0.48%	0.45%	0.40%	0.37%	0.37%	0.34%	0.32%	0.30%	0.28%	0.23%	0.23%	0.23%	0.23%	0.13%
	SD of % Error	1.10%	1.04%	0.93%	0.86%	0.84%	0.79%	0.74%	0.69%	0.65%	0.54%	0.54%	0.52%	0.53%	0.29%
	RMSE, absolute error	0.63%	0.46%	0.40%	0.37%	0.37%	0.34%	0.52%	1.27%	2.31%	5.12%	5.26%	8.77%	9.58%	19.79%
	RMSE, % error	1.46%	1.06%	0.93%	0.86%	0.86%	0.79%	1.20%	2.92%	5.32%	11.82%	12.14%	20.23%	22.10%	45.65%
WTP, post-merger % chng	Avg estimate	19.08%	19.59%	19.55%	18.89%	18.63%	18.15%	17.76%	17.30%	16.72%	15.04%	15.00%	13.37%	13.13%	7.84%
	Avg % Error	4.48%	7.26%	7.06%	3.45%	2.02%	-0.63%	-2.77%	-5.30%	-8.44%	-17.65%	-17.89%	-26.80%	-28.09%	-57.06%
	SD of estimate	0.30%	0.30%	0.30%	0.29%	0.29%	0.25%	0.22%	0.20%	0.19%	0.15%	0.15%	0.14%	0.14%	0.06%
	SD of % Error	1.64%	1.63%	1.66%	1.58%	1.59%	1.38%	1.22%	1.12%	1.05%	0.80%	0.81%	0.75%	0.76%	0.33%
	RMSE, absolute error	0.87%	1.36%	1.32%	0.69%	0.47%	0.28%	0.55%	0.99%	1.55%	3.23%	3.27%	4.90%	5.13%	10.42%
	RMSE, % error	4.77%	7.44%	7.25%	3.79%	2.57%	1.51%	3.02%	5.42%	8.50%	17.67%	17.91%	26.82%	28.10%	57.06%
Choice Probability, System 1	Mean absolute value of bias	0.10%	0.12%	0.20%	0.52%	0.64%	3.31%	4.60%	5.84%	6.71%	8.45%	8.48%	9.75%	10.26%	13.99%
	SD of estimate	13.53%	10.72%	7.84%	5.27%	3.87%	2.72%	1.62%	1.18%	0.87%	0.31%	0.25%	0.21%	0.19%	0.12%
	RMSE, absolute error	15.90%	12.38%	8.90%	6.13%	4.67%	6.49%	7.45%	8.55%	9.53%	11.65%	11.69%	12.90%	13.40%	16.66%
Choice Probability, System 2	Mean absolute value of bias	0.06%	0.07%	0.11%	0.29%	0.34%	1.81%	2.67%	3.37%	4.20%	5.71%	5.81%	6.80%	6.96%	9.15%
	SD of estimate	8.08%	6.35	4.60%	3.04%	2.20%	1.56%	0.95%	0.69%	0.51%	0.18%	0.16%	0.13%	0.13%	0.08%
	RMSE, absolute error	11.28%	8.72%	6.21%	4.18%	3.14%	4.38%	5.31%	6.52%	7.60%	9.44%	9.50%	10.26%	10.40%	12.50%

Notes: Each column reports results from a different Monte Carlo specification consisting of 50,000 simulations. A random sample of 124,237 admissions is generated for each simulation. The hospital choice for each admission is randomly generated based on estimates from the model reported in Table I for a minimum group size of 50 patients and a variable ordering of LAD. Each simulated dataset is used to estimate the semiparametric model for the indicated minimum group size and a variable ordering of LAD.
% of estimate is the average of the statistic (i.e., % change in WTP or diversion ratio) across simulations divided by the true value of the statistic in the data. SD of % error is the standard deviation of the statistic across simulations divided by the true value of the statistic in the data. RMSE, % error is the root mean squared error of the statistic, computed across simulations, also divided by the true value of the statistic in the data. The mean absolute value of bias is the average of the absolute value of the biases from estimates of individuals' choice probabilities. The reported RMSE for individual choice probabilities is the square root of the average MSE from individuals' choice probabilities.

Table III Monte Carlo Results Varying Group Size, Minimum Group Size of 3 Patients is the True Model

		3	5	10	25	50	100	250	500	1,000	2,500	5,000	10,000	25,000	50,000
Diversion, System 2 to 1	Avg estimate	42.11%	42.78%	43.51%	43.62%	43.42%	43.37%	42.97%	42.15%	41.09%	38.27%	38.13%	34.59%	33.78%	23.56%
	Avg % Error	-0.76%	0.81%	2.52%	2.79%	2.31%	2.21%	1.25%	-0.66%	-3.17%	-9.81%	-10.13%	-18.49%	-20.40%	-44.48%
	SD of estimate	0.53%	0.52%	0.49%	0.41%	0.38%	0.35%	0.32%	0.30%	0.28%	0.23%	0.23%	0.23%	0.23%	0.13%
	SD of % Error	1.26%	1.23%	1.14%	0.97%	0.89%	0.83%	0.76%	0.70%	0.66%	0.55%	0.55%	0.53%	0.55%	0.30%
	RMSE, absolute error	0.62%	0.63%	1.18%	1.25%	1.05%	1.00%	0.62%	0.41%	1.37%	4.17%	4.31%	7.85%	8.66%	18.87%
	RMSE, % error	1.47%	1.47%	2.77%	2.95%	2.48%	2.36%	1.46%	0.96%	3.24%	9.83%	10.15%	18.50%	20.41%	44.48%
WTP, post-merger % chng	Avg estimate	15.03%	17.34%	19.51%	19.45%	18.64%	18.15%	17.76%	17.31%	16.73%	15.06%	15.01%	13.37%	13.13%	7.84%
	Avg % Error	-13.73%	-0.47%	12.02%	11.68%	6.99%	4.20%	1.98%	-0.64%	-3.92%	-13.55%	-13.80%	-23.25%	-24.59%	-54.98%
	SD of estimate	0.26%	0.30%	0.33%	0.32%	0.30%	0.26%	0.22%	0.20%	0.19%	0.15%	0.15%	0.14%	0.14%	0.06%
	SD of % Error	1.50%	1.71%	1.91%	1.86%	1.72%	1.51%	1.29%	1.17%	1.10%	0.85%	0.85%	0.78%	0.79%	0.34%
	RMSE, absolute error	2.41%	0.31%	2.12%	2.06%	1.25%	0.78%	0.41%	0.23%	0.71%	2.36%	2.41%	4.05%	4.29%	9.58%
	RMSE, % error	13.81%	1.77%	12.17%	11.83%	7.20%	4.47%	2.36%	1.33%	4.08%	13.57%	13.83%	23.26%	24.61%	54.98%
Choice Probability, System 1	Mean absolute value of bias	1.61%	3.72%	7.12%	9.69%	11.01%	12.07%	12.94%	13.75%	14.26%	15.34%	15.34%	16.14%	16.70%	19.45%
	SD of estimate	10.90%	9.81%	7.96%	5.68%	4.17%	2.84%	1.65%	1.19%	0.87%	0.31%	0.25%	0.21%	0.19%	0.12%
	RMSE, absolute error	14.72%	14.19%	15.59%	17.06%	17.92%	18.63%	19.11%	19.60%	20.05%	21.14%	21.16%	21.86%	22.16%	24.27%
Choice Probability, System 2	Mean absolute value of bias	0.93%	2.12%	4.04%	5.52%	6.19%	6.70%	7.23%	7.68%	8.17%	9.05%	9.11%	9.92%	10.06%	12.19%
	SD of estimate	5.94%	5.38%	4.42%	3.16%	2.33%	1.60%	0.96%	0.69%	0.51%	0.18%	0.16%	0.13%	0.13%	0.08%
	RMSE, absolute error	10.54%	10.09%	11.03%	11.95%	12.42%	12.85%	13.29%	13.85%	14.40%	15.45%	15.49%	15.97%	16.06%	17.50%

Notes: Each column reports results from a different Monte Carlo specification consisting of 50,000 simulations. A random sample of 124,237 admissions is generated for each simulation. The hospital choice for each admission is randomly generated based on estimates from the model reported in Table I for a minimum group size of 3 patients and a variable ordering of LAD. Each simulated dataset is used to estimate the semiparametric model for the indicated minimum group size and a variable ordering of LAD.
% of estimate is the average of the statistic (i.e., % change in WTP or diversion ratio) across simulations divided by the true value of the statistic in the data. SD of % error is the standard deviation of the statistic across simulations divided by the true value of the statistic in the data. RMSE, % error is the root mean squared error of the statistic, computed across simulations, also divided by the true value of the statistic in the data. The mean absolute value of bias is the average of the absolute value of the biases from estimates of individuals' choice probabilities. The reported RMSE for individual choice probabilities is the square root of the average MSE from individuals' choice probabilities.

Table IV Monte Carlo Results Varying Group Size, Simple Parametric Logit is the True Model

		3	5	10	25	50	100	250	500	1,000	2,500	5,000	10,000	25,000	50,000
Diversion, System 2 to 1	Avg estimate	37.94%	38.14%	38.15%	38.15%	38.13%	38.09%	37.89%	37.36%	36.10%	31.93%	31.77%	30.34%	29.80%	23.56%
	Avg % Error	-0.58%	-0.07%	-0.02%	-0.04%	-0.09%	-0.18%	-0.72%	-2.11%	-5.41%	-16.34%	-16.74%	-20.50%	-21.91%	-38.26%
	SD of estimate	0.40%	0.36%	0.32%	0.29%	0.29%	0.28%	0.28%	0.28%	0.27%	0.19%	0.19%	0.19%	0.19%	0.13%
	SD of % Error	1.05%	0.95%	0.83%	0.77%	0.75%	0.74%	0.74%	0.72%	0.70%	0.51%	0.51%	0.49%	0.50%	0.33%
	RMSE, absolute error	0.46%	0.36%	0.32%	0.29%	0.29%	0.29%	0.39%	0.85%	2.08%	6.24%	6.39%	7.83%	8.36%	14.60%
	RMSE, % error	1.20%	0.95%	0.83%	0.77%	0.75%	0.76%	1.03%	2.23%	5.46%	16.35%	16.75%	20.51%	21.91%	38.26%
WTP, post-merger % chng	Avg estimate	17.37%	17.20%	16.62%	15.92%	15.67%	15.54%	15.38%	15.10%	14.30%	11.78%	11.72%	11.08%	10.91%	7.84%
	Avg % Error	12.18%	11.08%	7.29%	2.82%	1.20%	0.33%	-0.70%	-2.52%	-7.63%	-23.94%	-24.35%	-28.48%	-29.57%	-49.36%
	SD of estimate	0.27%	0.26%	0.24%	0.22%	0.21%	0.20%	0.20%	0.19%	0.19%	0.11%	0.11%	0.10%	0.10%	0.06%
	SD of % Error	1.75%	1.65%	1.57%	1.42%	1.33%	1.29%	1.27%	1.24%	1.20%	0.70%	0.70%	0.67%	0.67%	0.39%
	RMSE, absolute error	1.91%	1.74%	1.16%	0.49%	0.28%	0.21%	0.22%	0.43%	1.20%	3.71%	3.77%	4.41%	4.58%	7.64%
	RMSE, % error	12.31%	11.21%	7.46%	3.16%	1.79%	1.33%	1.45%	2.81%	7.72%	23.95%	24.36%	28.49%	29.57%	49.36%
Choice Probability, System 1	Mean absolute value of bias	0.08%	0.06%	0.05%	0.05%	0.07%	0.12%	0.33%	0.98%	2.43%	5.18%	5.43%	6.61%	7.29%	10.99%
	SD of estimate	15.02%	11.85%	8.57%	5.56%	3.82%	2.54%	1.53%	1.18%	0.94%	0.33%	0.27%	0.22%	0.20%	0.12%
	RMSE, absolute error	16.83%	13.05%	9.28%	5.95%	4.13%	2.88%	2.32%	3.44%	5.60%	7.94%	8.04%	9.20%	9.72%	12.69%
Choice Probability, System 2	Mean absolute value of bias	0.05%	0.04%	0.03%	0.03%	0.03%	0.06%	0.18%	0.58%	1.48%	4.34%	4.49%	5.12%	5.33%	6.44%
	SD of estimate	9.63%	7.59%	5.49%	3.56%	2.45%	1.62%	0.98%	0.76%	0.55%	0.20%	0.17%	0.14%	0.13%	0.08%
	RMSE, absolute error	11.45%	8.88%	6.31%	4.04%	2.80%	1.92%	1.57%	2.46%	3.82%	7.14%	7.18%	7.50%	7.61%	9.13%

Notes: Each column reports results from a different Monte Carlo specification consisting of 50,000 simulations. A random sample of 124,237 admissions is generated for each simulation. The hospital choice for each admission is randomly generated based on estimates from a parametric logit specification that controls for travel time, its square, and a set of hospital fixed effects. Each simulated dataset is used to estimate either the semiparametric model for the indicated minimum group size and a variable ordering of LAD.
% of estimate is the average of the statistic (i.e., % change in WTP or diversion ratio) across simulations divided by the true value of the statistic in the data. SD of % error is the standard deviation of the statistic across simulations divided by the true value of the statistic in the data. RMSE, % error is the root mean squared error of the statistic, computed across simulations, also divided by the true value of the statistic in the data. The mean absolute value of bias is the average of the absolute value of the biases from estimates of individuals' choice probabilities. The reported RMSE for individual choice probabilities is the square root of the average MSE from individuals' choice probabilities.

Table V Monte Carlo Results Varying Variable Ordering, LAD and Minimum Group Size of 50 Patients is the True Model

		LAD	ADL	DLA	LDA	ALD	DAL
Diversion, System 2 to 1	Avg estimate	43.45%	26.53%	42.32%	42.70%	30.65%	26.43%
	Avg % Error	0.23%	-38.81%	-2.39%	-1.51%	-29.30%	-39.03%
	SD of estimate	0.36%	0.20%	0.33%	0.33%	0.25%	0.20%
	SD of % Error	0.83%	0.46%	0.77%	0.76%	0.57%	0.45%
	RMSE, absolute error	0.37%	16.83%	1.09%	0.73%	12.71%	16.92%
	RMSE, % error	0.86%	38.81%	2.51%	1.69%	29.31%	39.03%
WTP, post-merger % chng	Avg estimate	18.56%	9.51%	17.39%	17.59%	11.79%	9.45%
	Avg % Error	1.63%	-47.91%	-4.79%	-3.70%	-35.42%	-48.24%
	SD of estimate	0.28%	0.11%	0.24%	0.23%	0.16%	0.11%
	SD of % Error	1.56%	0.61%	1.30%	1.28%	0.86%	0.60%
	RMSE, absolute error	0.41%	8.75%	0.91%	0.72%	6.47%	8.81%
	RMSE, % error	2.25%	47.92%	4.96%	3.92%	35.43%	48.25%
Choice Probability, System 1	Mean absolute value of bias	1.21%	11.11%	3.56%	3.19%	9.91%	11.19%
	SD of estimate	3.77%	3.79%	3.65%	3.70%	3.61%	3.31%
	RMSE, absolute error	5.03%	14.85%	7.01%	6.70%	13.74%	14.81%
Choice Probability, System 2	Mean absolute value of bias	0.59%	7.49%	2.24%	1.97%	6.73%	7.54%
	SD of estimate	2.13%	2.50%	2.13%	2.13%	2.35%	2.13%
	RMSE, absolute error	3.28%	11.55%	5.31%	4.97%	10.66%	11.52%

Notes: Each column reports results from a different Monte Carlo specification consisting of 5,000 simulations. A random sample of 124,237 admissions is generated for each simulation. The hospital choice for each admission is randomly generated based on estimates from the model reported in Table I for a minimum group size of 50 patients and a variable ordering of LAD. Each simulated dataset is used to estimate the semiparametric model for a minimum group size of 50 and the indicated variable ordering. We examine 6 orderings varying the order of Location (L), Admission Type (A), and Demographics (D) variables, as in Table 3.1.
% of estimate is the average of the statistic (i.e., % change in WTP or diversion ratio) across simulations divided by the true value of the statistic in the data. SD of % error is the standard deviation of the statistic across simulations divided by the true value of the statistic in the data. RMSE, % error is the root mean squared error of the statistic, computed across simulations, also divided by the true value of the statistic in the data. The mean absolute value of bias is the average of the absolute value of the biases from estimates of individuals' choice probabilities. The reported RMSE for individual choice probabilities is the square root of the average MSE from individuals' choice probabilities.

Table VI Monte Carlo Results Varying Variable Ordering, LAD and Minimum Group Size of 3 Patients is the True Model

		LAD	ADL	DLA	LDA	ALD	DAL
Diversion, System 2 to 1	Avg estimate	42.50%	36.81%	42.47%	42.46%	39.22%	36.70%
	Avg % Error	0.14%	-13.26%	0.07%	0.06%	-7.59%	-13.51%
	SD of estimate	0.52%	0.48%	0.52%	0.52%	0.47%	0.46%
	SD of % Error	1.22%	1.12%	1.22%	1.23%	1.11%	1.09%
	RMSE, absolute error	0.52%	5.65%	0.52%	0.52%	3.25%	5.75%
	RMSE, % error	1.23%	13.31%	1.23%	1.23%	7.67%	13.56%
WTP, post-merger % chng	Avg estimate	15.88%	14.49%	16.05%	16.05%	14.82%	14.41%
	Avg % Error	-8.81%	-16.83%	-7.83%	-7.85%	-14.89%	-17.27%
	SD of estimate	0.28%	0.26%	0.27%	0.28%	0.26%	0.26%
	SD of % Error	1.59%	1.52%	1.57%	1.59%	1.50%	1.51%
	RMSE, absolute error	1.56%	2.94%	1.39%	1.40%	2.61%	3.02%
	RMSE, % error	8.95%	16.90%	7.98%	8.01%	14.96%	17.34%
Choice Probability, System 1	Mean absolute value of bias	3.40%	5.51%	3.84%	3.83%	4.77%	5.73%
	SD of estimate	11.32%	13.45%	12.09%	12.08%	12.61%	13.24%
	RMSE, absolute error	15.65%	18.15%	16.49%	16.49%	17.23%	18.10%
Choice Probability, System 2	Mean absolute value of bias	1.98%	3.68%	2.23%	2.22%	3.19%	3.83%
	SD of estimate	6.11%	8.25%	6.66%	6.65%	7.52%	8.08%
	RMSE, absolute error	11.12%	13.41%	11.71%	11.71%	12.75%	13.39%

Notes: Each column reports results from a different Monte Carlo specification consisting of 5,000 simulations. A random sample of 124,237 admissions is generated for each simulation. The hospital choice for each admission is randomly generated based on estimates from the model reported in Table I for a minimum group size of 3 patients and a variable ordering of LAD. Each simulated dataset is used to estimate the semiparametric model for a minimum group size of 3 and the indicated variable ordering. We examine 6 orderings varying the order of Location (L), Admission Type (A), and Demographics (D) variables, as in Table 3.1.
% of estimate is the average of the statistic (i.e., % change in WTP or diversion ratio) across simulations divided by the true value of the statistic in the data. SD of % error is the standard deviation of the statistic across simulations divided by the true value of the statistic in the data. RMSE, % error is the root mean squared error of the statistic, computed across simulations, also divided by the true value of the statistic in the data. The mean absolute value of bias is the average of the absolute value of the biases from estimates of individuals' choice probabilities. The reported RMSE for individual choice probabilities is the square root of the average MSE from individuals' choice probabilities.

Table VII Monte Carlo Results Varying Group Ordering Under Minimum Group Size of 3, LAD and Minimum Group Size of 50 Patients is the True Model

		LAD	ADL	DLA	LDA	ALD	DAL
Diversion, System 2 to 1	Avg estimate	43.01%	37.84%	43.00%	42.99%	40.07%	37.73%
	Avg % Error	-0.80%	-12.72%	-0.81%	-0.83%	-7.57%	-12.96%
	SD of estimate	0.46%	0.42%	0.47%	0.48%	0.43%	0.41%
	SD of % Error	1.07%	0.98%	1.09%	1.11%	0.99%	0.96%
	RMSE, absolute error	0.58%	5.53%	0.59%	0.60%	3.31%	5.64%
	RMSE, % error	1.34%	12.76%	1.36%	1.39%	7.63%	13.00%
WTP, post-merger % chng	Avg estimate	19.15%	17.06%	19.22%	19.21%	17.76%	16.97%
	Avg % Error	4.87%	-6.59%	5.24%	5.20%	-2.76%	-7.07%
	SD of estimate	0.30%	0.28%	0.30%	0.30%	0.28%	0.28%
	SD of % Error	1.66%	1.54%	1.63%	1.64%	1.53%	1.51%
	RMSE, absolute error	0.94%	1.24%	1.00%	1.00%	0.58%	1.32%
	RMSE, % error	5.14%	6.77%	5.49%	5.45%	3.15%	7.23%
Choice Probability, System 1	Mean absolute value of bias	0.26%	3.27%	0.38%	0.38%	2.38%	3.42%
	SD of estimate	13.13%	14.75%	13.74%	13.74%	14.01%	14.45%
	RMSE, absolute error	15.08%	17.17%	15.74%	15.74%	16.28%	16.94%
Choice Probability, System 2	Mean absolute value of bias	0.15%	2.47%	0.23%	0.22%	1.88%	2.57%
	SD of estimate	7.72%	9.47%	8.13%	8.13%	8.75%	9.26%
	RMSE, absolute error	10.58%	12.53%	11.07%	11.07%	11.86%	12.38%

Notes: Each column reports results from a different Monte Carlo specification consisting of 5,000 simulations. A random sample of 124,237 admissions is generated for each simulation. The hospital choice for each admission is randomly generated based on estimates from the model reported in Table I for a minimum group size of 50 patients and a variable ordering of LAD. Each simulated dataset is used to estimate the semiparametric model using the indicated variable ordering, although the minimum group size is set to 3 and not the true value of 50. We examine 6 orderings varying the order of Location (L), Admission Type (A), and Demographics (D) variables, as in Table 3.1.
% of estimate is the average of the statistic (i.e., % change in WTP or diversion ratio) across simulations divided by the true value of the statistic in the data. SD of % error is the standard deviation of the statistic across simulations divided by the true value of the statistic in the data. RMSE, % error is the root mean squared error of the statistic, computed across simulations, also divided by the true value of the statistic in the data. The mean absolute value of bias is the average of the absolute value of the biases from estimates of individuals' choice probabilities. The reported RMSE for individual choice probabilities is the square root of the average MSE from individuals' choice probabilities.

Table VIII Monte Carlo Results, Alternative Sample Sizes

		5%	10%	25%	50%	75%	100%
Diversion, System 2 to 1	Avg estimate	41.04%	42.07%	43.09%	43.37%	43.41%	43.42%
	Avg % Error	-5.35%	-2.96%	-0.60%	0.04%	0.13%	0.15%
	SD of estimate	1.27%	0.95%	0.65%	0.49%	0.42%	0.37%
	SD of % Error	2.94%	2.20%	1.49%	1.12%	0.96%	0.85%
	RMSE, absolute error	2.64%	1.60%	0.70%	0.49%	0.42%	0.37%
	RMSE, % error	6.10%	3.69%	1.61%	1.12%	0.97%	0.86%
WTP, post-merger % chng	Avg estimate	16.88%	17.41%	17.99%	18.30%	18.52%	18.63%
	Avg % Error	-7.58%	-4.70%	-1.51%	0.18%	1.38%	2.01%
	SD of estimate	0.89%	0.67%	0.46%	0.36%	0.32%	0.29%
	SD of % Error	4.89%	3.64%	2.54%	2.00%	1.77%	1.59%
	RMSE, absolute error	1.65%	1.09%	0.54%	0.37%	0.41%	0.47%
	RMSE, % error	9.02%	5.95%	2.96%	2.00%	2.25%	2.57%
Choice Probability, System 1	Mean absolute value of bias	6.54%	5.65%	4.18%	3.22%	1.94%	0.64%
	SD of estimate	3.85%	3.65%	3.60%	3.77%	3.93%	3.87%
	RMSE, absolute error	10.15%	9.12%	7.85%	6.96%	5.89%	4.67%
Choice Probability, System 2	Mean absolute value of bias	4.12%	3.34%	2.41%	1.77%	1.11%	0.34%
	SD of estimate	2.27%	2.13%	2.04%	2.16%	2.25%	2.20%
	RMSE, absolute error	7.95%	6.89%	5.53%	4.70%	3.99%	3.14%

Notes: Each column reports results from a different Monte Carlo specification consisting of 50,000 simulations. A random sample with the indicated number of observations is generated for each simulation. The hospital choice for each admission is randomly generated based on estimates from the model reported in Table I for a minimum group size of 50 patients and a variable ordering of LAD. Each simulated dataset is used to estimate the semiparametric model for the indicated minimum group size and a variable ordering of LAD.

% of estimate is the average of the statistic (i.e., % change in WTP or diversion ratio) across simulations divided by true value of the statistic in the data. SD of % error is the standard deviation of the statistic across simulations divided by the true value of the statistic in the data. RMSE, % error is the root mean squared error of the statistic, computed across simulations, also divided by the true value of the statistic in the data. The mean absolute value of bias is the average of the absolute value of the biases from estimates of individuals' choice probabilities. The reported RMSE for individual choice probabilities is the square root of the average MSE from individuals' choice probabilities.

A Bias-Variance Tradeoffs with S_{min}

In this section, we consider the bias-variance tradeoffs involved in setting the minimum group size, S_{min}, when the semiparametric logit specification is used to estimate diversion and WTP. Starting with a correctly specified model, we first analyze the bias from combining two groups with heterogeneous preferences. We then consider the opposite situation in which the model is unnecessarily flexible.

We start by assuming that the semiparametric model is correctly specified. The overall diversion from hospital h to hospital j across two groups A and B is a weighted average of the group-level diversions where the weight is N_h^g, the number of patients in group g that select hospital h.

$$div_{hj} = (N_h^A div_{hj}^A + N_h^B div_{hj}^B)/(N_h^A + N_h^B). \tag{11}$$

Suppose that the two groups are combined due to the (mistaken) belief that they have the same hospital preferences. The estimated diversion $\hat{div}_{hj}$ from hospital h to j for the combined group is simply the estimated fraction of patients in A and B, after excluding those who choose hospital h, which selects hospital j. The expected value of this estimator does not equal the actual diversion div_{hj} defined in equation (11):

$$E(\hat{div}_{hj}) = (N_{\sim h}^A div_{hj}^A + N_{\sim h}^B div_{hj}^B)/(N_{\sim h}^A + N_{\sim h}^B). \tag{12}$$

The expected value of the estimated diversion from hospital h to j for the combined group is still a weighted average of the group-level diversions, but now the weight is $N_{\sim h}^g$, the number of patients in group g that do not select hospital h. The estimated diversion for the combined group will be unbiased only in special cases. For example, unbiased estimates are obtained when the fraction of each group that selects hospital h is the same (i.e. $\frac{N_h^A}{N_{\sim h}^A} = \frac{N_h^B}{N_{\sim h}^B}$) , or when the group-level diversions are identical, so the weighting difference does not matter (i.e., $div_{hj}^A = div_{hj}^B$). In general, however, the use of an overly restrictive model leads to biased diversion estimates.

Next, consider the potential bias from using an overly flexible model. We start with the assumption that each member of group g has identical preferences, and then divide this homogeneous group into two subgroups A and B. The estimated diversion from hospital h to j across the two groups is as follows:

$$\hat{div}_{hj} = (N_h^A \hat{div}_{hj}^A + N_h^B \hat{div}_{hj}^B)/(N_h^A + N_h^B). \tag{13}$$

Since all members of the two subgroups have the same preferences, $E(\hat{div}_{hj}) = E(\hat{div}_{hj}^A) = E(\hat{div}_{hj}^B) = div_{hj}$. That is, the estimated diversion across the two groups is an unbiased estimate of the true diversion. There is a caveat, however: one must be able to estimate the diversion from hospital h to j for each subgroup. This is not possible when a group is composed solely of individuals that select hospital h.

However, diversion ratios will be less precisely estimated. Returning to the example where a homogeneous group is divided into two subgroups A and B, let α_h denote the fraction of patients, among those who choose hospital h, that are put into group A. Similarly, let $\alpha_{\sim h}$ denote the fraction of those who do not choose hospital h that are put into group A. The variance of the estimated diversion $\hat{div}_{hj}$ defined in equation (13) is as follows:

$$V(\hat{div}_{hj}) = \phi_h \frac{div_{hj}(1 - div_{hj})}{N_{\sim h}}. \tag{14}$$

where $\phi_h = \frac{\alpha_h^2}{\alpha_{\sim h}} + \frac{(1-\alpha_h)^2}{1-\alpha_{\sim h}}$. When $\phi_h = 1$, the variance of the estimated diversion calculated separately for each group equals the variance of the estimated diversion when it is calculated for the combined group. Holding $\alpha_{\sim h}$ fixed, ϕ_h is a convex function of α_h that has a minimum at $\alpha_h = \alpha_{\sim h}$, at which point $\phi_h = 1$. That is, the only time that there is no efficiency loss from dividing a homogeneous group into two subgroups is when those selecting hospital h and those selecting other hospitals are allocated to the two groups in similar proportions. While this condition may approximately hold when a large group is divided, it is less likely to hold when group sizes are small due to random sampling. Thus, the efficiency cost from dividing a large group into medium sized groups is likely to be less than the loss in efficiency from dividing a medium sized group into small groups.

The use of an overly flexible model also affects WTP estimates. As before, we start with a single group with homogeneous preferences. The group's estimated WTP (per person) for hospital h is as follows:

$$\hat{WTP}_h^g = -\ln(1 - \hat{s}_h^g). \tag{15}$$

Next, we divide the group into two. The average WTP (per person) for hospital h across the two groups is estimated as follows:

$$\hat{WTP}_h = -[N^A \ln(1 - \hat{s}_h^A) + N^B \ln(1 - \hat{s}_h^B)]/(N^A + N^B). \tag{16}$$

Both equation (15) and equation (16) provide consistent estimates of the true WTP for the group, although the estimates will not be unbiased since WTP is a nonlinear function. However, the use of a more flexible model can have a significant impact in finite samples. Since WTP is a convex function, and $\hat{s}_h = (N^A \hat{s}_h^A + N^B \hat{s}_h^B)/(N^A + N^B)$, the WTP estimate from equation (16) is weakly larger than the WTP estimate using equation (15). This can lead to an inference problem where it is unclear whether estimated WTP is high because patients strongly value a given hospital, or because an overly flexible model is being employed.

For the change in WTP, both the numerator – the WTP of the combined system – and denominator – the WTP of each individual system – increase with a more flexible model. Thus, the effect of changing the group size on the change in WTP is ambiguous.