A Model of Unordered Multiple Choice with Unobserved Choice Set Selection

Scott Desposato*
Department of Political Science
University of California, San Diego
swd@ucsd.edu

Version 1.1

May 25, 2005

Abstract

Models of unordered choice typically assume that all agents can choose freely among a known set of choices. Such models have been applied to many problems, including job choice, migration, party switching by politicians, and most frequently, consumer choice. However, in many cases, some choices are excluded from the choice set, though this exclusion is unobserved by the analyst. In this paper, I show how unobserved variation in choice sets can bias estimates from choice models and offer a method that will correct for this bias. The method estimates both predictors of the components of the choice set as well as predictors of choice from whatever options are available.

*Thanks to Alex Braithwaite for useful suggestions.

Introduction

Unordered choice models have been widely implemented in the social sciences since McFadden (1973)'s seminal paper on transportation choice. This paper presents a method to estimate parameters for choice models with unobserved choice set selection. With such models, individuals choose among several options, but each individual has a different choice set, as a function of their characteristics. The analyst is aware of the maximal choice set, but does not know which options are in any particular individual's choice set. Although unobserved choice set variation is effectively ignored in the literature, it is present (and biases estimates) in many current applications of choice models. For example, traditional methods might be used to model workers' employment decisions, students' university choices, and prime ministers' choice of government partners. But unbeknownst to the analyst, some choices simply aren't in the choice set. Would-be employees can't take a job not offered to them. Students cannot attend UC Berkeley (at least not for credit) if not admitted. A prime minister can't add an extremist party to her governing coalition if that party rejects the invitation.

Previous empirical work examining such choices usually uses standard choice models, especially conditional logit. Related alternatives include multinomial probit and more generally, mixed logit. If the choice sets don't have unobserved variation, such methods are appropriate. But when the analyst cannot observe choice set variation, traditional methods lead to biased results. Intuitively, the agent might prefer option "A", but if only B and C are available, the analyst would incorrectly infer that B or C offers higher utility, when in fact they do not. This unobserved choice set variation will bias estimated coefficients toward zero, as we will see.

A few scholars have examined some related issues, but not directly tackled variance in choice sets. Previous work on selectivity focuses on the fact that the expected payoffs for choices are unobserved for options not chosen. For example, if studying labor market migrations, we don't observe agent i's wages in a city that she did not move to. Lee (1983) offer a solution, which has been implemented by labor economists for migration and employment models, but do not consider the problem of choice set variation.

In this paper I present a method for studying consumer choice when choice set selection exists. I proceed by illustrating the problem, proposing a solution, and examining its properties in simulations, and offer recommendations for model building when analysts suspect unobserved choice set variation.

Standard Conditional Logit

In the classic conditional logit model (McFadden, 1973), consumers choose among options to maximize their utility. In the original application, McFadden (1973) examined subjects' choice of transportation medium: private automobile versus several public transportation options. An individual i chooses the option j of m available choices that maximizes her utility U_{ij}:

$$U_{ij} = Max(U_{i1}, ..., U_{im}) \tag{1}$$

The utility associated with a choice is a function of choice-specific covariates plus a random error:

$$U_{ij} = x_{ij}\beta + \epsilon_{ij}$$

where the ϵ_{ij} are distributed iid extreme value with expected value 0 and variance 1.

The utility U_{ij} is unobserved; we only observe the ultimate choice of the agents. The probablity that individual i chooses option j is:

$$P(U_{ij} = MAX(U_{i1}, ..., U_{im})) = \frac{e^{x_{ij}\beta}}{\sum_{j=1}^{m} e^{x_{ij}\beta}} \tag{2}$$

the likelihood function is simply the product of the probability of each observed outcome:

$$L = \prod_{i=1}^{n} \prod_{j=1}^{m} P(y_{ij} = 1)^{y_{ij}} \tag{3}$$

where n is the number of individuals in the dataset and y_{ij} is an indicator variable for the observed choice, coded "1" if i chose j, and "0" otherwise. Estimation is straightforward and canned routines exist in most software packages. The properties of this method and resulting estimator b are well-studied and documented elsewhere.

Unobserved Variation in Choice Sets

Conditional logit and recent extensions have been widely applied in economics, geography, and political science to study diverse questions include vote choice, firms' location decisions, labor market and migration patterns, and family size (Morgan and Rindfuss, 1985; Adams and Samuel Merril, 1999; Fox, 1996; Stafford, 2000; Whittington, 1992; Berger, 1988; Woodward and Rolfe, 1993; Boskin, 1974; Alvarez and Nagler, 1998; Hartman, 1982). Ultimately, the classic conditional logit model assumes that all choices are indeed available to the agent. But in many situations, each individual's choice set may vary, and such variance may be unobserved by the analyst. A firm might prefer to locate in a particular municipality, but be excluded by local ordinances. A President may want to invite a large party into a coalition government, but said party may refuse on ideological grounds. A couple may want a large family but be unable to have children. A graduate student might want a tenure-track job at Harvard, but not have that option. In effect, each agent's choice set is first determined, then agents choose from their available options, but the analyst does not observe the determination of the choice set and is unaware which options are available and which are not.[1]

Previous work has at least acknowledged the possibility of unobserved choice set variation, but not offered any solutions. Parsons and Kealy (1992) examine consumers' choice of recreational alternatives - lakes - in Wisconsin. Their concern is different from mine; they seek an estimation solution to the problem of having huge choice sets. However, given that there are thousands of possible lakes for recreation in that region, they acknowledge that if individuals do not have full knowledge about all options, their choice set may be smaller than that included in the empirical analysis. They dismiss choice set variation as not important for their analysis, since "no information" options are likely to be those that consumers would not choose anyway: an avid fisherman would certainly be aware of a large, unpolluted, well-stocked fishing lake close to home. Effectively, they assume that strong correlation between choice predictors and choice set predictors will alleviate any bias induced by unobserved choice set variance. As we will see, their approach diminishes but does not eliminate the bias associated with ignoring choice set variation.

I propose explicitly modeling the unobserved choice set selection. This alternative model of choice, which I call "conditional logit with choice set selection", can be formalized as follows. First, the presence of option j in the choice set is a function of a standard random utility function. For individual i, option j's presence or absence in her choice set is determined by the value of an underlying variable, $z*$:

j is included in i's choice set iff: $z*_{ij} \geq 0$

j is excluded from i's choice set iff: $z*_{ij} < 0$

$z*$ is a linear function of a set of observed variables w, their coefficients, and a random error γ:

$$z*_{ij} = \alpha w_{ij} + \gamma_{ij}$$

where all γ_{ij} are identically and independently distributed logistic with variance 1. Thus, the probability that j is included in i's choice set is $P(z* > 1)$ or $\frac{e^{\alpha w}}{1+e^{\alpha w}}$. Intuitively, one can imagine an employer deciding whether or not to make a job offer, a party deciding whether or not they are willing to be part of a coalition, or a university deciding whether or not they would accept prospective student i

[1]Such choice set selection might even apply to voting decisions. Conditional logit has been used to examine voting decisions in multi-party elections, sometimes including "not voting" as a choice Palfrey and Poole (1987). However, some respondents may not be registered to vote and consequently have a restricted choice set, though unobserved by the analyst. Not being registered to vote could be voluntary, but might also be involuntary, imposed by institutional bias, recent moves, or felony convictions. Many other examples exist.

as a member. The analyst does not observe this choice directly, just the ultimate decision made by the agent.

Let option $j = 1$ be available to every individual i. This eliminates the possibility that individual i winds up with no choices in the final set. This notion of a default option makes sense: an individual can choose not to enter the labor market, not to go to college, not to join any party, and so on. Thus given m options, there are 2^{m-1} possible choice sets. Suppose there are three choices, labeled A, B, and C, and the default option that is present in every choice set is A. There are thus $2^{3-1} = 2^2 = 4$ possible choice sets, listed below:

Table 1: All Possible Choices Sets for $m = 3$ and $j = A$ default

	Choice Sets		
1	A	B	C
2	A	B	
3	A		C
4	A		

In set 1, all choices are available, in set 2, just A and B, and so on. In set 4, only the default option, A, is available. Note that the choices observed provide information about which choice sets i did or did not face. Choosing "B" means that the choice set is 1 or 2; B is not an option for the other choice sets. The exception is the default choice, "A" - it is available in every choice set and thus does not restrict the possible sets.

For individual i, the probability of observing a particular choice set is:

$$\prod_{j=1}^{M} P(\alpha w_{ij} > 0)^{z_{ij}} (1 - P(\alpha w_{ij} > 0)^{1-z_{ij}}$$

$$= \prod \left(\frac{e^{\alpha w_{ij}}}{1 + e^{\alpha w_{ij}}} \right)^{z_{ij}} \left(1 - \frac{e^{\alpha w_{ij}}}{1 + e^{\alpha w_{ij}}} \right)^{1-z_{ij}} \tag{4}$$

where z_{ij} is an indicator variable for the presence of option j in individual i's choice set. For the example given above, the probability of choice set number 2 is (recalling that A, the default option, is present in every set with probability 1):

$$P(set2) = P(B) * (1 - P(C))$$

or

$$P(\alpha w_{iB} > 0) * (1 - P(\alpha w_{iC} > 0)$$
$$= \left(\frac{e^{\alpha w_{iB}}}{1 + e^{\alpha w_{iB}}} \right) * \left(1 - \frac{e^{\alpha w_{iC}}}{1 + e^{\alpha w_{iC}}} \right)$$

When we observe a choice of "B", as mentioned, this could reflect either of two choice sets (1 or 2). Given the choice set, calculating the probability of observing choice B is easy:

$$P(B|1) = P(U_{iB} = Max(U_{iA}, U_{iB}, U_{iC}))$$

$$P(B|2) = P(U_{iB} = Max(U_{iA}, U_{iB}))$$

$$P(B|3) = 0$$

$$P(B|4) = 0$$

Since we do not observe the choice set, only the predictors of inclusion/exclusion from the choice set, we write the probability of observing individual i choose option B as the sum of the conditional probabilities of choosing B in each of the choice sets, times the probability of each set:

$$\begin{gathered} P(Y_{ij} = B) = P(U_{iB} = Max(U_{iA}, U_{iB}, U_{iC})) * P(Set1) \\ +P(U_{iB} = Max(U_{iA}, U_{iB})) * P(Set2) \end{gathered}$$

dropping the i's for clarity and writing out the probabilities:

$$= \left(\frac{e^{x_B\beta}}{e^{x_A\beta} + e^{x_B\beta} + e^{x_C\beta}} \frac{e^{\alpha w_B}}{1 + e^{\alpha w_B}} \frac{e^{\alpha w_C}}{1 + e^{\alpha w_C}} \right) + \left(\frac{e^{x_B\beta}}{e^{x_A\beta} + e^{x_B\beta}} * \frac{e^{\alpha w_B}}{1 + e^{\alpha w_B}} \left(1 - \frac{e^{\alpha w_C}}{1 + e^{\alpha w_C}}\right)\right)$$

More generally, the probability of observing individual i choosing choice $j = J$ is:

$$\sum_{k=1}^{S} \frac{e^{x_{iJ}\beta}}{\sum_{j=1}^{m} e^{x_{ij_s}\beta I_{ijk}}} P(\alpha w_{iJ} > 0) \prod_{j \neq 1} P(\alpha w_{ij} > 0)^{I_{ijk}} (1 - P(\alpha w_{ij} > 0))^{1-I_{ijk}}$$

where k indexes the choice sets, and I_{ijk} is an indicator variable coded "1" if option j is included in the k^{th} choice set of individual i, and "0" otherwise. If J is the default choice that is available in every choice set, then the term immediately before the product ($P(\alpha w_{iJ} > 0)$) is "1" - because J is always available - and $S = 2^{m-1}$. If J is not the default choice, then the term before the product is less than one, and $S = 2^{m-2}$.

The log-likelihood of observed choice Y_{ij} given covariates x and w is:

$$\begin{aligned} LL &= \sum_{i=1}^{n} x_{iJ}\beta + log(P(\alpha w_{iJ})) + \\ & log\left(\sum_{k=1}^{S} \frac{1}{\sum_{j=1}^{m} e^{x_{ijs}\beta I_{ijk}}} \prod P(\alpha w_{ij} > 0)^{I_{ijk}} (1 - P(\alpha w_{ij} > 0))^{1-I_{ijk}} \right) \end{aligned}$$

Estimation and Examples

Estimation of β and α is straightforward, though much more computationally intensive than classic conditional logit, because one must sum the probabilities across all choice sets for each individual. With $M = 3$ this is not so bad; there are only 4 possible choice sets. But with $M = 7$, there are $2^{7-1} = 64$ possible choice sets; with $M = 20$ there are over 500,000.

In the following sections, I examine the properties of CLS via simulation, first comparing CLS with classic conditional logit, then considering CLS' weaknesses and limitations.

Simulated Comparison: Classic Conditional Logit versus Conditional Logit with Unobserved Choice Set Selection

Figures 1, 2, and 3 compare simulated coefficient estimates for three models. In each graph, the x-axis is the sample size from the simulated dataset (100 to 1000), and the y-axis shows the estimated value of β. The shaded area is the range of 95% of the simulated estimates - the area between the 2.5th quantile and the 97.5th quantile. The dashed line shows the mean estimated value and the dotted line the median estimated value over all 1,000 simulations for each sample size.

The first column in each figure shows classic conditional logit, ignoring the choice set variation. The second column shows results for classic conditional logit, but including the selection covariates w in the matrix of choice covariates, without modeling choice set selection.[2] The third column shows estimates of β using my proposed method - explicitly accounting for choice set selection.

In each case, 1,000 datasets were generated for each sample size with a single covariate at the choice level, with a parameter value of "1", and an intercept and single covariate at the selection level, both with parameter values of "1".[3] The three graphs show how manipulating different parameters affects estimation.

In Figures 1, each row shows results for different values of α_0, the intercept in the choice set model. Effectively, manipulating α_0 affects the baseline probability of inclusion in the choice set. In the top row, the probability that a choice is included is .2, in the second row, the probability is .5, and in the bottom row the probability is .8.

[2]This last option deserves additional explanation. For example, suppose we are modeling Presidents' choices of coalition partners in government formation, but we believe that some parties are not in the choice set, because they will refuse to join an ideologically incompatible President. A "quick and dirty" fix to the selection problem might be to simply include a measure of ideological compatibility in the model of Presidential choice, expecting that Presidents' won't choose incompatible parties - because they can't. Ultimately, this solution fails to fix the bias problem.

[3]x and w were simultaneously generated from a bivariate normal random variable. In most cases, I set the correlation between x and w to 0, but I experiment with collinearity, shown in Figure 2.

Three results are immediately apparent. First, the conditional logit with unobserved selection (henceforth CLWS) model will provide unbiased estimates while both of the other approaches will not. The median simulated value for conditional logit with selection is consistently close to the true value 1, and its variance around that value falls as sample size grows. In contrast, for both of the alternatives, estimates are consistently well below the true value. In almost every case, the true value is well *outside* the 95% interval of simulated values. Specifically, when there is choice set selection, coefficient estimates from classical conditional logit are biased toward zero.

Intuitively, the classic conditional logit has a harder time making correct predictions, because it is unaware of the varying choice sets. The additional errors are taken as evidence that the options j are more similar than they really are, or that β is closer to zero. Related, the bias in traditional methods is greatest when $P(include)$ is low. When there is a significant amount of choice set variation, traditional methods will be furthest from the true parameter value. Where such unobserved variation is minimized, the classic model's estimates approach the true value, though they are still biased downward.

The second core result is that using conditional logit when there is unobserved choice set variance will produce misleadingly small standard errors and inflate significance tests. CLWS gives estimates have significantly more variance than the classical approach. For larger sample sizes, the range of estimates narrows around the true value to acceptable limits, but for small samples the range of estimates is more than three times that of the classical model. This, too, makes sense. The selection model has to incorporate additional variance associated with the uncertainty about which choice set is available; the classic model assumes away this additional variance, resulting in much tighter (and biased) confidence bands. Note, however, that as the probability of a full choice set increases (moving to the bottom row), CLWS is nearly as efficient as conditional logit and still the only unbiased approach. The third finding from Figure 1 is that adding the selection covariates to a traditional conditional logit (aka, the 2nd column) has effectively no impact on the bias, indicating that this approach is not a solution to choice set variation.

Figure 2 repeats the procedure, but examines the impact of variation in the correlation (ρ_{xw}) between the choice and selection covariates (x and w). The correlation between the two variables is 0 in the first row, .45 in the second row, and .90 in the third. The figure shows that traditional approaches are still biased with correlation between x and w, though that bias fades for both alternatives (1st and 2nd columns) as the correlation increases. Including the selection covariates in a traditional conditional logit, however, does not work very well. The bias only fades slightly as ρ increases, and the variance quickly grows.

CLWS remains unbiased regardless of the value of ρ_{xw}. In addition, increasing ρ_{xw} increases the variance of estimates of β for all methods. For small samples, CLWS is again very imprecise: the range of estimates for a sample of size 100 with

$\rho_{xw} = .90$ is more than three times that of the alternative methods. Precision increases as sample size grows.

Finally, Figure 3 compares bias when increasing the size of the choice set from 3 (top row) to 6 (middle row) and 9 (bottom row) choices. In each case, the selection model (CLWS) continues to return unbiased estimates, and increasing the number of choices decreases the variance in estimates: larger choice sets provide more information about β. For classic results, increasing the number of choices decreases the bias, with confidence intervals capturing the true value for the smallest (highest variance) samples.

Discussion

Combined, these observations suggest that conditional logit with selection should be used whenever the analyst suspects unobserved choice set variation, with one important qualification. Traditional approaches lead to biased coefficient estimates and smaller standard errors than are in fact the case. CLWS will produce unbiased estimates, though it is extremely inefficient for small sample sizes.

The important qualification is a practical one. As discussed, the method does have an important computational problem: it is constrained by the size of the choice set. As the number of choices increases, the number of possible choice sets increases exponentially, and becomes cumbersome to compute for large choice sets. Estimating β for a sample of size 1,000 with 3 choices takes 10 seconds (including matrix setups and transformations) on a desktop computer with a Pentium 4© processor. With a choice set of 6, the time increases 300% to 40 seconds. And with a choice set of 9, the time increased another 200% to 205 seconds, and with a choice set of 12, estimation time increased to 1,516 seconds, or about 25 minutes. In contrast, traditional conditional logit without selection reached convergence in less than 2 seconds for all these choice set sizes.

When choice sets are of a manageable size and the analyst suspects choice set variation, CLWS is clearly superior - it elminates bias and avoids understating standard errors. What should one do when agents are faced with many options? Estimates will still be biased, but not by very much. One solution would be to use the selection model, but sample randomly from the available choice sets to increase computational ease. Further, in cases where there is public information about the choice set, the analyst can incorporate such information into the likelihood function. An example is the study of party-switching by legislators, common in many new democracies. Some parties may refuse ideologically incompatible new members; other parties don't care about ideology, and just want to maximize their size. Unfortunately, the analyst usually doesn't observe private offers or denials of membership between parties and politicians, so there is unobserved choice set variance. However, occasionally parties make public statements about their willingness to accept a new member, or their refusal of membership to one they deem incompatible. Such information can be incorporated directly into the likelihood function, reducing the number of choice sets and providing additional information about α.

Finally, while I have focused on extending the conditional logit model, the method can easily be extended to other variants of choice models, including multinomial probit, or mixed logit.

References

Adams, J. and I. Samuel Merril (1999, July). Modeling Party Strategies and Policy Representation in Multiparty Elections: Why Are Strategies so Extreme? *AJPS 43*(3), 765–791.

Alvarez, R. M. A. and J. Nagler (1998, January). When Politics and Models Collide: Estimating Models of Multiparty Elections. *American Journal of Political Science 42*(1), 55–96.

Berger, M. C. (1988, April). Predicted future earnings and choice of college major. *Industrial and Labor Relations Review 41*(3), 418–429.

Boskin, M. J. (1974, March). A Conditional Logit Model of Occupational Choice. *The Journal of Political Economy 82*(2), 389–398.

Fox, S. E. (1996, March). The Influence of Political Conditions on Foreign Firm Location Decisions in the American States (1974-1989). *Political Research Quarterly 49*(1), 51–75.

Glasgow, G. (2001). Mixed Logit Models for Multiparty Elections. *Political Analysis*.

Gschwend, T. and C. M. King (2002). The Politics of Opinion Assignment: A Conditional Logit Model with varying Choice Set. *Working Paper*.

Hartman, R. S. (1982, November). The Appropriateness of Conditional Logit for the Modeling of Residential Fuel Choice. *Land Economics 58*(4), 478–487.

Lee, L.-F. (1983, March). Generalized Econometric Models with Selectivity. *Econometrica 51*(2), 507–512.

Manski, C. F. and S. R. Lerman (1977, November). The Estimation of Choice Probabilities from Choice Based Samples. *Econometrica 45*(8), 1977–1988.

Martin, L. W. and R. T. Stevenson (2001). Government Formation in Parliamentary Democracies. *AJPS 45*(1), 33–50.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic.

Morgan, S. P. and R. R. Rindfuss (1985). Marital Disruption: Structual and Temporal Dimensions. *The American Journal of Sociology 90*(5), 1055–1077.

O Brien, D. M. (2003). *Storm Center: The Supreme Court in American Politics* (6th ed.). New York: W. W. Norton & Company.

Palfrey, T. R. and K. T. Poole (1987, August). The Relationship between Information, Ideology, andVoting Behavior. *AJPS 31*(3), 511–530.

Parsons, G. R. and M. J. Kealy (1992). Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation. *Land Economics 68*(1), 93–106.

Stafford, S. L. (2000, November). The Impact of Environmental Regulations on the Location of Firms in the Hazardous Waste Management Industry. *Land Economics 76*(4), 569–589.

Whittington, L. A. (1992, May). Taxes and the family: The impact of the tax exemption for dependents on marital fertility. *Demography 29*(2), 215–226.

Woodward, D. P. and R. J. Rolfe (1993, 1st Quarter). The Location of Export-Oriented Foreign Direct Investment in the Caribbean Basin. *Journal of International Business Studies 24*(1), 121–144.

Figure 1: Impact of P(In Choice Set) on Estimates of β: Rows are P(I) = .2,.5,.8

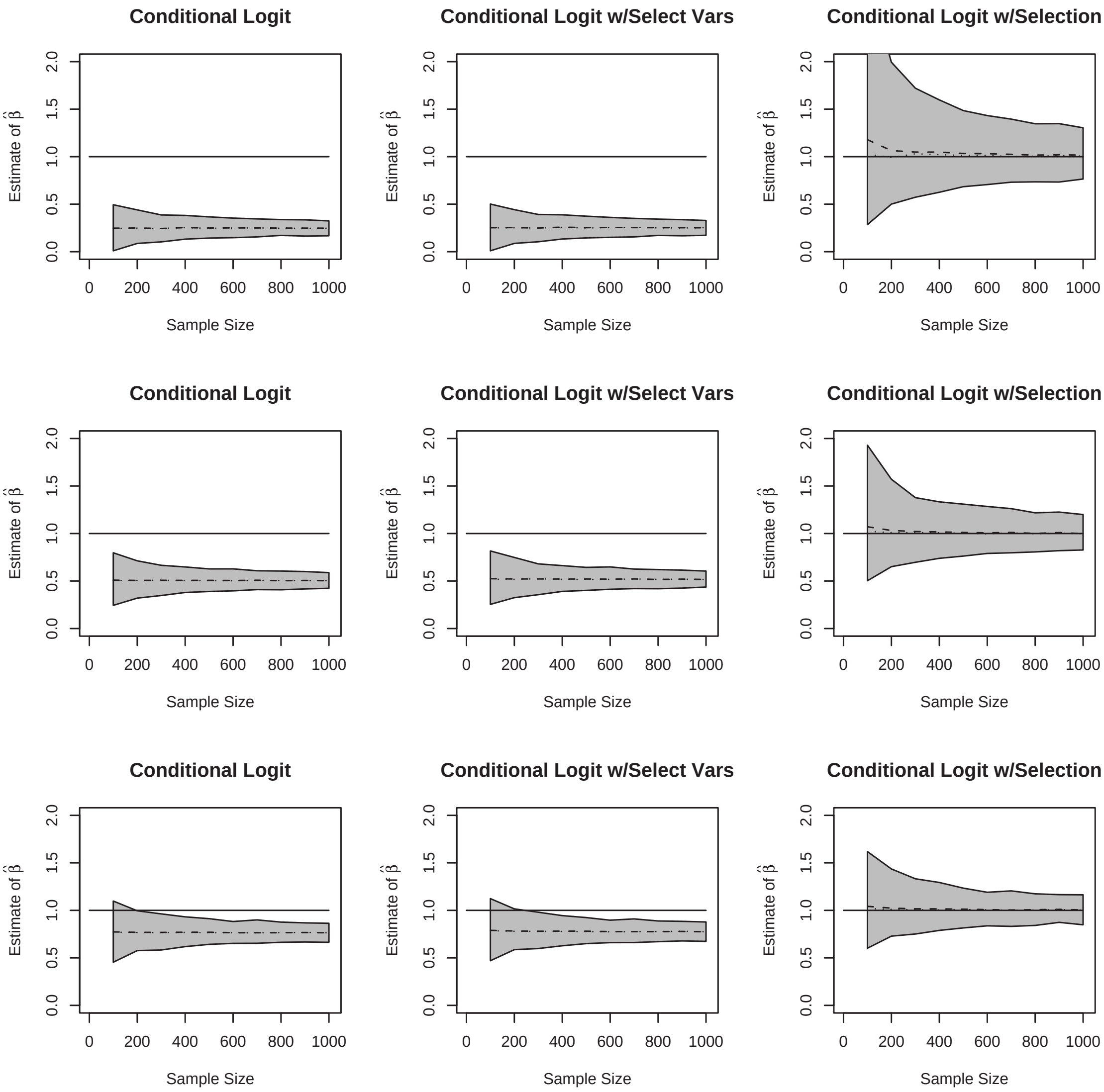

Figure 2: Impact of Corr(x,w) on Estimates of β: Rows are $\rho_{xw} = 0,.45,.90$

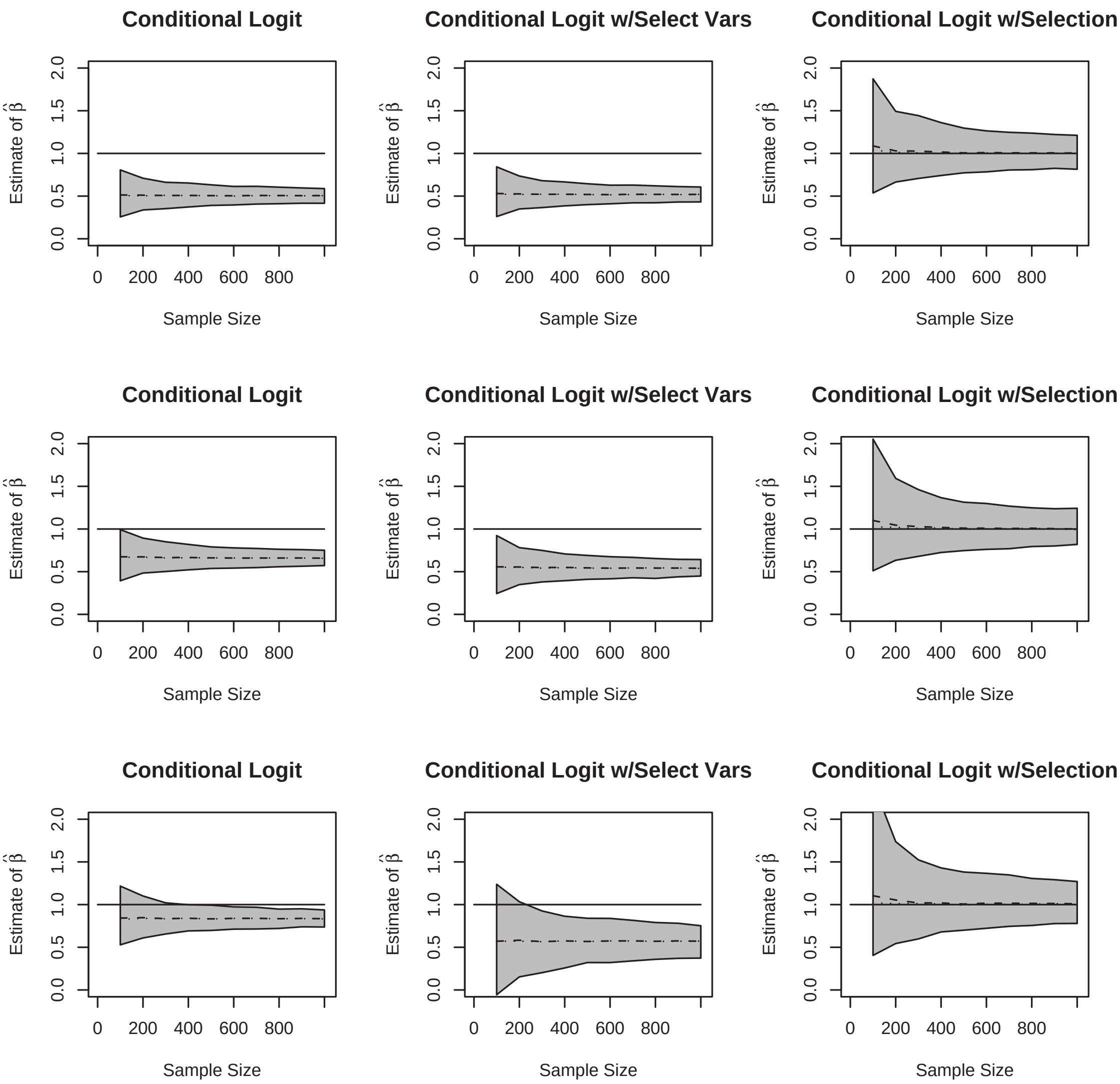

Figure 3: Impact of Choice Set Size on Estimates of β: Rows are M = 3,6,9