Web Appendix Materials for:

Using Auxiliary Data to Estimate Selection Bias Models,

With an Application to Interest Group Use of the Direct Initiative Process

Frederick J. Boehmke [1]

University of Iowa

June 16, 2003

[1]frederick-boehmke@uiowa.edu. These materials are supporting documents for an article published in *Political Analysis*.

Abstract

This document contains supplemental materials for the article published in *Political Analysis*, including evidence relating to the difficulties associated with estimating the Full Information Maximum Likelihood version of the stochastic truncation model and the results of additional Monte Carlo analyses run to demonstrate the robustness of the model to the normality assumption of the error distribution. Readers interested in the derivation of the adjusted covariance matrix for the two-stage correction should see the article by Murphy and Topel (1985), which uses the Taylor series approach to derive the correction. Sample programs used in the paper and this appendix available from the author's web page: http://rubagalo.polisci.uiowa.edu/∼fredb/.

1 Failure of FIML Convergence

Table 1: Frequency of Failure to Converge for Stochastic Truncation Model

Seed	Completed Estimations Until Failure
1	0
2	0
3	1
4	3
5	4
6	1
7	1
8	1
All Trials	1.38 (42% failure)

Trials done in GAUSS with 10,000 observations per trial. The program was initialized with the given seed, data was then generated and the model estimated. If it converged, it was allowed to generate another data set and estimate the model. This continued until it failed to converge, when the number of successful trials was noted. The seed was then increased by one and the process repeated. The parameter values are the same as those used in the simulations discussed in the text.

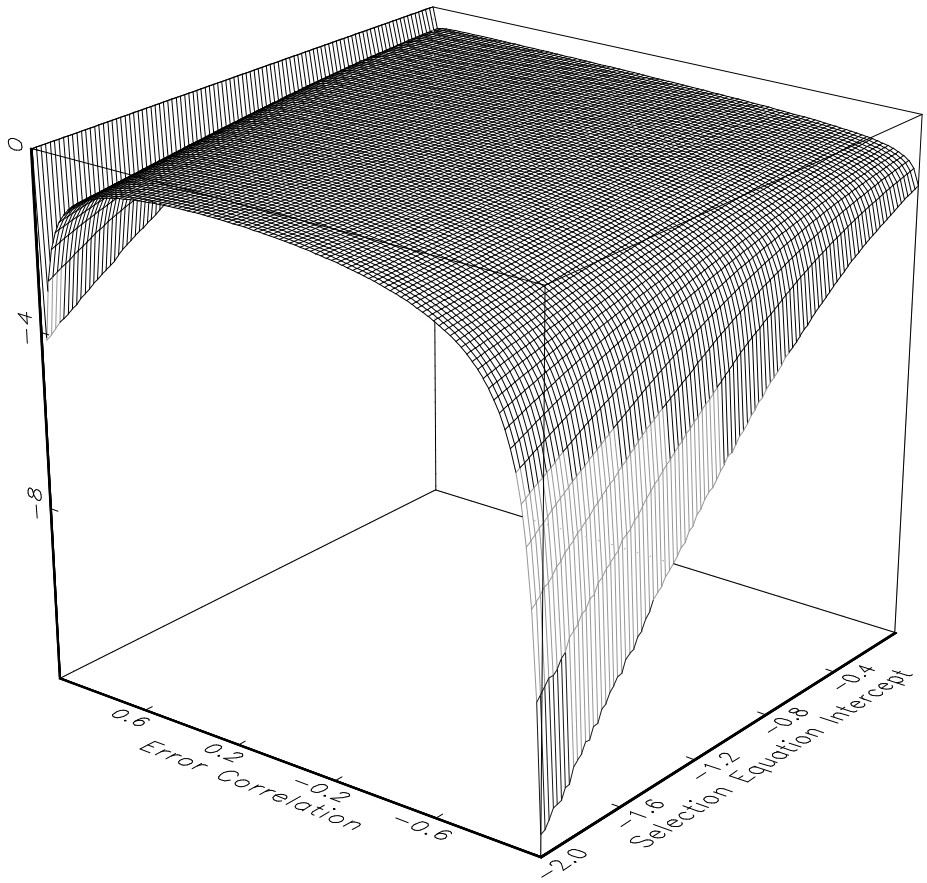

Figure 1: Log Likelihood Function for Stochastic Truncation Model, Varying ρ and Selection Equation Intercept (other parameters fixed at true values)

2 Additional Monte Carlo Analyses

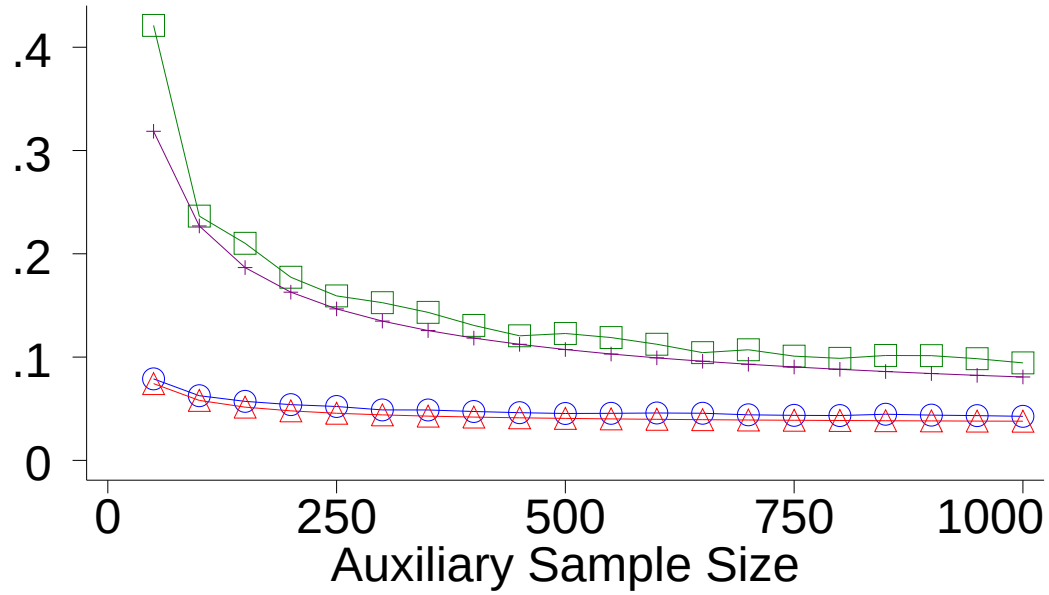

Figure 2: Comparison of True Delta Method and Sampling Distribution Standard Errors; N=2000, $\rho = 0.5$. Key: ◯ Delta method SE($\hat{\alpha}$); △ Sampling distribution SD($\hat{\alpha}$); □ Delta method SE($\hat{\beta}$); + Sampling distribution SD($\hat{\beta}$).

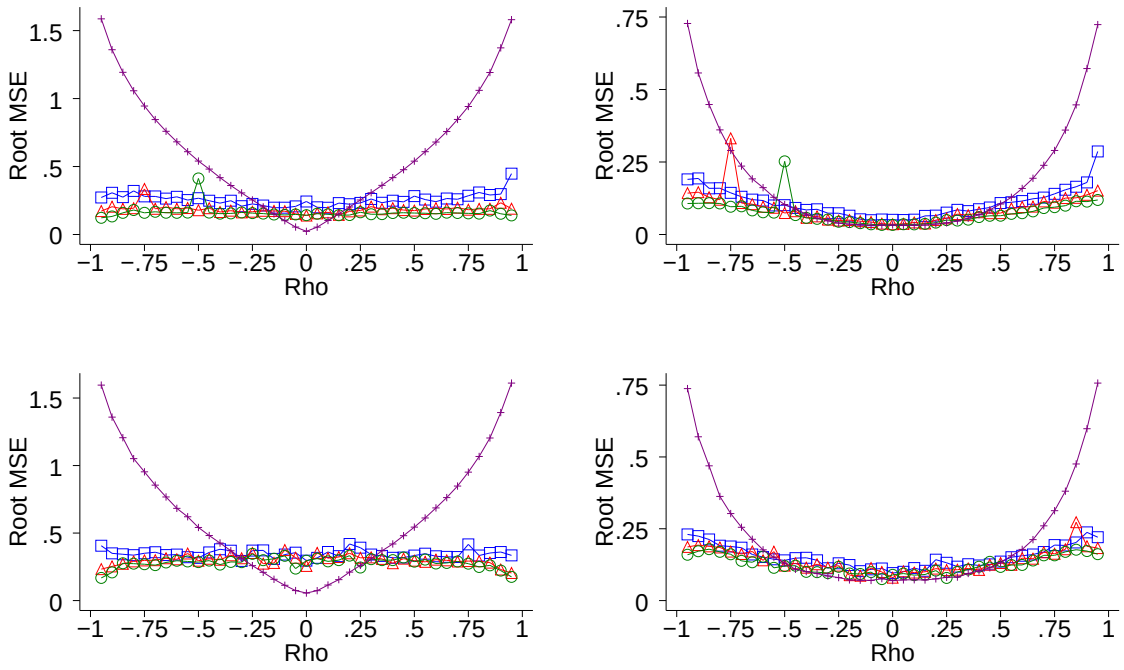

Figure 3: RMSE of the equation of interest parameters when selection equation error is distributed $N(0, 4)$, varying ρ; intercept (left) and slope (right); N=10,000 (top) and N=2000 (bottom). Key: △ Intercept; □ Slope; ◯ Correlation.

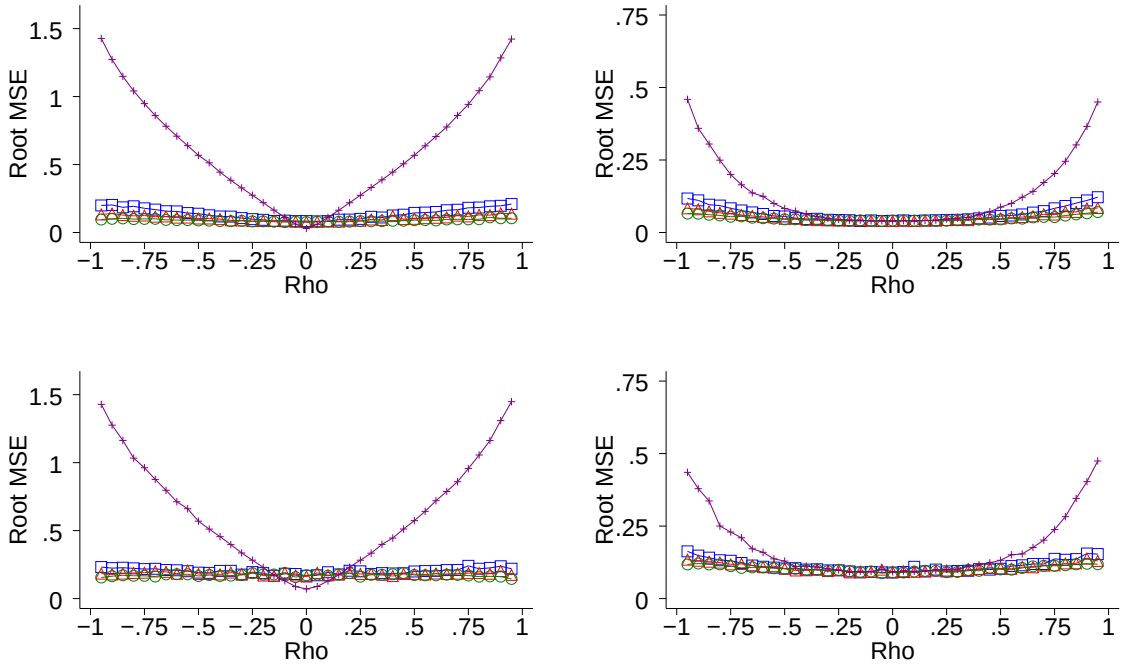

Figure 4: RMSE of the equation of interest parameters when selection equation error has logarithmic distribution, varying ρ; intercept (left) and slope (right); N=10,000 (top) and N=2000 (bottom). Key: △ Intercept; □ Slope; ◯ Correlation.

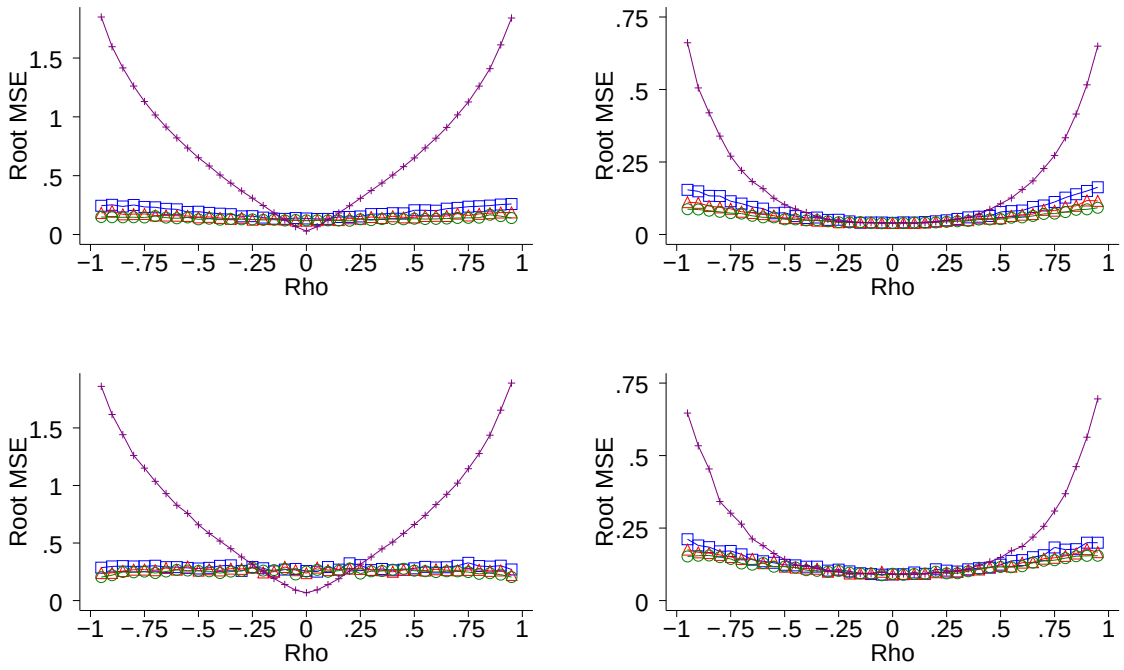

Figure 5: RMSE of the equation of interest parameters when selection equation error has χ^2_5 distribution, varying ρ; intercept (left) and slope (right); N=10,000 (top) and N=2000 (bottom). Key: △ Intercept; □ Slope; ◯ Correlation.

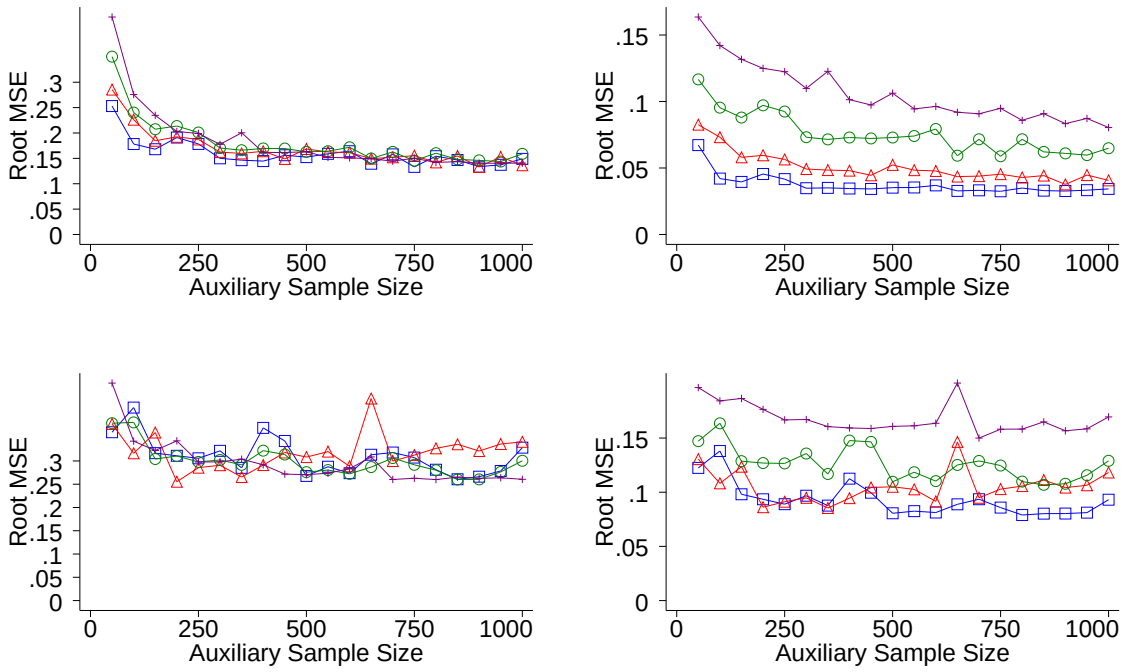

Figure 6: RMSE of the equation of interest parameters when selection equation error is distributed $N(0, 4)$, varying auxiliary sample size; intercept (left) and slope (right); N=10,000 (top) and N=2000 (bottom). Key: $\square$ $\rho = 0$; $\triangle$ $\rho = 0.25$; $\bigcirc$ $\rho = 0.5$; $+$ $\rho = 0.75$.

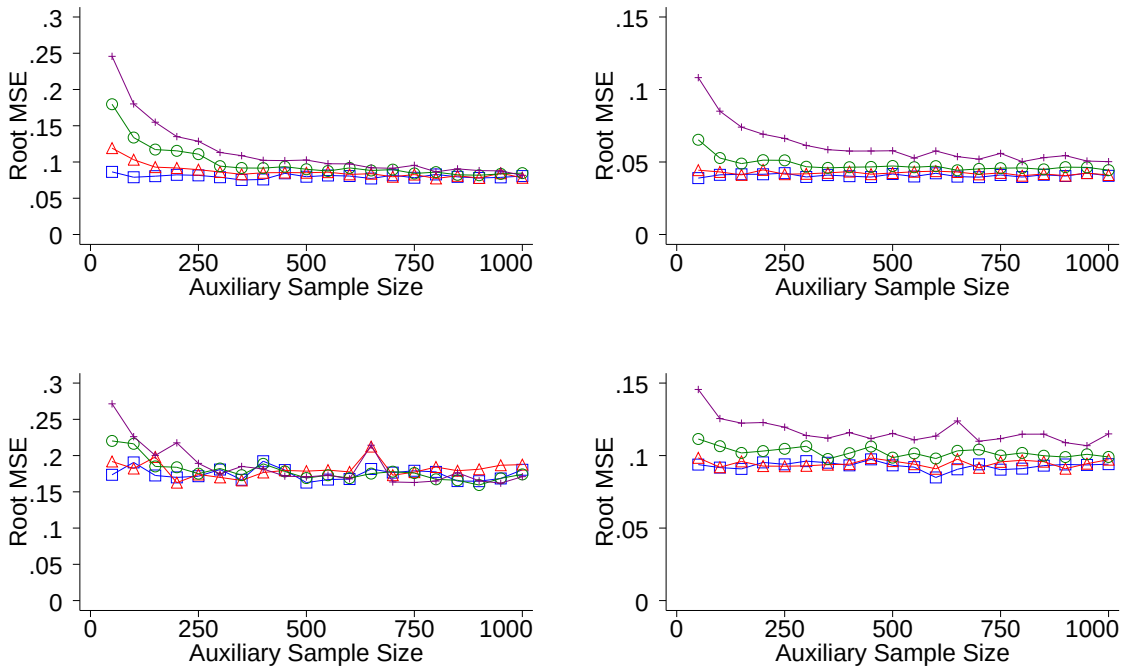

Figure 7: RMSE of the equation of interest parameters when selection equation error has logarithmic distribution, varying auxiliary sample size; intercept (left) and slope (right); N=10,000 (top) and N=2000 (bottom). Key: $\square$ $\rho = 0$; $\triangle$ $\rho = 0.25$; $\bigcirc$ $\rho = 0.5$; $+$ $\rho = 0.75$.

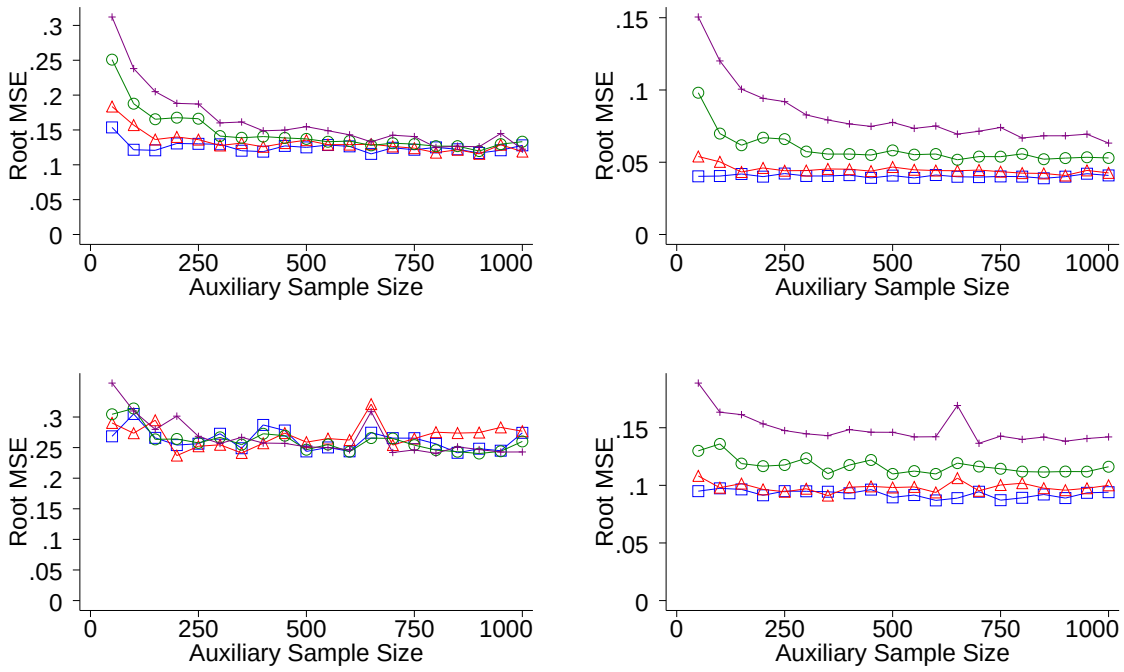

Figure 8: RMSE of the equation of interest parameters when selection equation error has χ^2_5 distribution, varying auxiliary sample size; intercept (left) and slope (right); N=10,000 (top) and N=2000 (bottom). Key: $\square$ $\rho = 0$; $\triangle$ $\rho = 0.25$; $\bigcirc$ $\rho = 0.5$; $+$ $\rho = 0.75$.