What to do with Incomplete Nominal Variables?

A Monte Carlo Simulation Study Comparing Methods for Creating Multiple Imputations of Unordered Factors

Kyle M. Lang

Quantitative Psychology Training Program
Department of Psychology
University of Kansas
Lawrence, KS

May 20, 2014

Outline

- Introduce the problem
- Describe the simulation study
- Present the results of the simulation study
- Give recommendations based on the findings

Motivation for the Current Work

THE MOTIVATING PROBLEM: The vast majority (all?) of the research on imputing categorical items focuses on binary or ordinal responses. How do the recommended techniques perform when applied to nominal variables?

Motivation for the Current Work

THE MOTIVATING PROBLEM: The vast majority (all?) of the research on imputing categorical items focuses on binary or ordinal responses. How do the recommended techniques perform when applied to nominal variables?

To answer this question, I conducted a Monte Carlo simulation study comparing the performance of five leading multiple imputation (MI) approaches that can accommodate unordered factors.

Motivation for the Current Work

THE MOTIVATING PROBLEM: The vast majority (all?) of the research on imputing categorical items focuses on binary or ordinal responses. How do the recommended techniques perform when applied to nominal variables?

To answer this question, I conducted a Monte Carlo simulation study comparing the performance of five leading multiple imputation (MI) approaches that can accommodate unordered factors.

Two Crucial Requirements:

- The imputation algorithm must produce “legal” values for the nominal variables.
- The imputation algorithm must be able to simultaneously treat mixed variable types (i.e., continuous and categorical items).

The Chosen Imputation Methods

Five MI strategies were chosen for comparison:

The Chosen Imputation Methods

Five MI strategies were chosen for comparison:

- MICE with multinomial logistic regression (MNOM) as the elementary method
- MICE with classification and regression trees (CART) as the elementary method
- MICE with predictive mean matching (PMM) as the elementary method

The Chosen Imputation Methods

Five MI strategies were chosen for comparison:

- MICE with multinomial logistic regression (MNOM) as the elementary method
- MICE with classification and regression trees (CART) as the elementary method
- MICE with predictive mean matching (PMM) as the elementary method
- The General Location Model (GLOC; Little & Schluchter, 1985)

The Chosen Imputation Methods

Five MI strategies were chosen for comparison:

- MICE with multinomial logistic regression (MNOM) as the elementary method
- MICE with classification and regression trees (CART) as the elementary method
- MICE with predictive mean matching (PMM) as the elementary method
- The General Location Model (GLOC; Little & Schluchter, 1985)
- The ranking method (RANK) suggested by Allison (2002)
    - Impute the dummy-coded factors under the normal model.
    - Calculate the imputed value for the reference group as: $y_K^* = 1 - (y_1^* + y_2^* + \ldots + y_{(K-1)}^*)$.
    - Assign the imputed dummy codes as: $y_{imp} = \mathrm{I}\left(max\{y_k^*\}_{k=1}^{K}\right)$.

Simulation Design

Three simulation parameter were varied:

- Sample Size $(N) = \{250, 500, 1000\}$
- Percent Missing $(PM) = \{10, 20, 30, 40, 50\}$
- Factor Levels $(K) = \{3, 5, 7, 10\}$

Simulation Design

Three simulation parameter were varied:

- Sample Size $(N) = \{250, 500, 1000\}$
- Percent Missing $(PM) = \{10, 20, 30, 40, 50\}$
- Factor Levels $(K) = \{3, 5, 7, 10\}$

A baseline-category multinomial logistic regression (MLogit) model was used for data generation and analysis:

- $\text{logit}(y_{nom}) = \boldsymbol{\alpha} + \boldsymbol{\beta} x^T + \varepsilon$

Simulation Design

Three simulation parameter were varied:

- Sample Size $(N) = \{250, 500, 1000\}$
- Percent Missing $(PM) = \{10, 20, 30, 40, 50\}$
- Factor Levels $(K) = \{3, 5, 7, 10\}$

A baseline-category multinomial logistic regression (MLogit) model was used for data generation and analysis:

- $\text{logit}(y_{nom}) = \boldsymbol{\alpha} + \boldsymbol{\beta} x^T + \varepsilon$

Fully crossed design:

- 3(N) $\times$ 5(PM) $\times$ 4(K) $\times$ 5(MI Method) = 300
- 500 replications within each cell

Data Generation I: Complete Data

A single nominal variable was simulated. Its response probabilities were given by:

$$P(y_{nom,n} = k|x_n) = \frac{\exp(\alpha_k + \beta x_n)}{1 + \sum_{k=1}^{(K-1)} \exp(\alpha_k + \beta x_n)},$$

where

- $n = 1, 2, \ldots, N$ indexes observation
- $k = 1, 2, \ldots, K$ indexes nominal response category
- $\alpha_k = \ln(3) \left(\frac{1}{2}\right)^{k-1}$, $k = 1, 2, \ldots, (K - 1)$
- $\beta = \ln(2)$
- $x \sim N(.25, 1)$

These probabilities were supplied to the R function `rmultinom()` to create the nominal responses.

Data Generation I: Complete Data

A single nominal variable was simulated. Its response probabilities were given by:

$$P(y_{nom,n} = k|x_n) = \frac{\exp(\alpha_k + \beta x_n)}{1 + \sum_{k=1}^{(K-1)} \exp(\alpha_k + \beta x_n)},$$

where

- $n = 1, 2, \ldots, N$ indexes observation
- $k = 1, 2, \ldots, K$ indexes nominal response category
- $\alpha_k = \ln(3) \left(\frac{1}{2}\right)^{k-1}$, $k = 1, 2, \ldots, (K-1)$
- $\beta = \ln(2)$
- $x \sim N(.25, 1)$

These probabilities were supplied to the R function `rmultinom()` to create the nominal responses.

Two additional continuous variables were also created:

- Auxiliary variable: $x_{aux} \sim N(0, 1)$
- Nuisance covariate: $x_{cov} \sim N(0, 1)$

Data Generation II: Missing Data Imposition

Missing data were imposed via a MAR process by applying the following functions variable-wise:

$$R_n = \mathrm{I}(\Phi^{-1}(x_{aux,n}) \leq PM)$$
$$R'_n = \mathrm{I}(\Phi^{-1}(x_{aux,n}) \geq (1 - PM))$$

where

- R and R' are $N \times 1$ indicator vectors with 1's indicating missing data and 0's indicated observed data
- $\Phi^{-1}(\cdot)$ is the inverse normal CDF
- $\mathrm{I}(\cdot)$ is an indicator function that equals 1 when its argument is true and 0 otherwise
- PM is the desired proportion of nonresponse

Data Generation II: Missing Data Imposition

Missing data were imposed via a MAR process by applying the following functions variable-wise:

$$R_n = \mathrm{I}(\Phi^{-1}(x_{aux,n}) \leq PM)$$
$$R'_n = \mathrm{I}(\Phi^{-1}(x_{aux,n}) \geq (1 - PM))$$

where

- R and R' are $N \times 1$ indicator vectors with 1's indicating missing data and 0's indicated observed data
- $\Phi^{-1}(\cdot)$ is the inverse normal CDF
- $\mathrm{I}(\cdot)$ is an indicator function that equals 1 when its argument is true and 0 otherwise
- PM is the desired proportion of nonresponse

Missing Data were only imposed on y_{nom} and x_{cov}.

Missing Data Imputation I: Software

The imputation was conducted with three different R packages:

- MNOM, CART, and PMM: `mice` (van Buuren & Groothuis-Oudshoorn, 2011)
- General Location Model: `mix` (Schafer, 2013)
- Allison (2002) Ranking Method: `Amelia II` (Honaker, King, & Blackwell, 2012)

Missing Data Imputation II: Settings

- Each of the five algorithms were used to create 100 imputations.
- All MICE-based conditions used 10 iterations per chain.
- For CART and PMM, all missingness was imputed via the same elementary method.
- For MNOM, the continuous covariate was imputed via a normal-theory Bayesian regression.
- For RANK, an adaptive ridge prior, with a maximum value of $N/10$, was employed.
- For GLOC, EM estimates were used as starting values, and each Markov Chain was run for 1000 iterations before drawing imputations.

Analysis & Evaluation

The analysis model was a multinomial logistic regression model:

- $\text{logit}(y_{nom}) = \boldsymbol{\alpha} + \boldsymbol{\beta} x^T + \varepsilon$

Analysis & Evaluation

The analysis model was a multinomial logistic regression model:

- $\text{logit}(y_{nom}) = \boldsymbol{\alpha} + \boldsymbol{\beta} x^T + \varepsilon$

The accuracy of the estimated regression coefficients was the focal outcome. This was assessed via two statistics:

- Percentage Relative Bias (PRB)
- Confidence Interval (CI) Coverage

The analysis model was a multinomial logistic regression model:

- $\text{logit}(y_{nom}) = \boldsymbol{\alpha} + \boldsymbol{\beta}x^T + \varepsilon$

The accuracy of the estimated regression coefficients was the focal outcome. This was assessed via two statistics:

- Percentage Relative Bias (PRB)
- Confidence Interval (CI) Coverage

PRB was calculated from the exponentiated regression coefficients, and the CIs were constructed by exponentiating the bounds computed from the raw coefficients, as recommended by Agresti (2007).

Results I: PRB

Percentage Relative Bias for Slope Coefficients:
Multinomial Logistic Regression (Averaged Over *K* - 1 Coefficients)

		N = 250					N = 500					N = 1000				
K	*PM*	*MNOM*	*CART*	*PMM*	*RANK*	*GLOC*	*MNOM*	*CART*	*PMM*	*RANK*	*GLOC*	*MNOM*	*CART*	*PMM*	*RANK*	*GLOC*
3	*0.1*	0.56	-0.54	-2.12	-0.95	0.42	0.30	-0.60	-2.63	-1.42	0.16	-0.10	-0.70	-3.05	-1.87	-0.21
3	*0.2*	0.39	-1.63	-4.55	-2.51	0.15	-0.07	-1.80	-5.35	-3.27	-0.33	-0.24	-1.94	-5.76	-3.62	-0.42
3	*0.3*	0.35	-3.03	-6.60	-3.43	0.25	0.01	-2.72	-7.55	-4.49	-0.29	-0.33	-2.71	-8.10	-5.21	-0.54
3	*0.4*	0.29	-3.74	-8.20	-3.95	0.78	-0.09	-3.26	-9.43	-5.54	-0.18	-0.46	-3.71	**-10.19**	-6.48	-0.58
3	*0.5*	-3.56	-9.84	**-11.78**	-6.02	-1.68	-2.14	**-11.54**	**-11.99**	-7.37	-1.88	-1.41	**-13.45**	**-12.20**	-8.12	-1.77
5	*0.1*	2.65	0.07	-3.14	0.00	2.02	1.24	-0.60	-4.29	-1.23	0.88	0.67	-1.01	-4.84	-1.78	0.49
5	*0.2*	2.02	-2.74	-8.69	-3.03	0.64	1.19	-2.47	-9.36	-3.67	0.45	0.81	-2.17	-9.64	-3.99	0.32
5	*0.3*	2.34	-5.75	**-12.91**	-4.97	0.27	1.49	-4.19	**-13.68**	-5.72	0.24	0.81	-3.99	**-14.08**	-6.26	-0.03
5	*0.4*	2.12	-9.74	**-16.83**	-7.42	-0.90	1.36	-7.38	**-17.65**	-8.00	-0.31	0.74	-6.32	**-18.05**	-8.42	-0.23
5	*0.5*	2.87	**-13.70**	**-19.73**	-8.88	-1.36	1.48	**-14.01**	**-20.79**	-9.82	-1.09	0.89	**-15.02**	**-21.56**	**-10.30**	-0.66
7	*0.1*	2.17	-0.32	-3.95	-0.52	1.35	1.37	-0.75	-4.76	-1.41	0.84	0.75	-0.66	-5.21	-1.94	0.50
7	*0.2*	2.98	-2.33	-9.08	-2.32	1.33	2.08	-2.13	-9.78	-3.42	1.10	1.11	-2.14	**-10.28**	-4.24	0.62
7	*0.3*	4.06	-4.17	**-13.91**	-4.48	1.22	2.80	-3.55	**-14.26**	-5.44	1.28	1.34	-3.73	**-15.04**	-6.66	0.51
7	*0.4*	5.14	-7.22	**-18.23**	-6.18	1.16	3.11	-6.27	**-18.87**	-7.69	0.83	1.56	-5.84	**-19.46**	-8.99	0.28
7	*0.5*	6.84	**-11.97**	**-22.58**	-8.17	1.74	3.58	**-12.43**	**-22.97**	-9.77	0.63	1.40	**-13.16**	**-23.63**	**-11.31**	-0.26
10	*0.1*	5.07	2.59	-1.80	1.65	4.18	2.42	0.25	-3.81	-0.71	1.92	1.78	0.23	-4.41	-1.37	1.46
10	*0.2*	6.51	1.29	-7.14	-0.62	4.42	3.15	-1.04	-9.08	-3.00	2.03	2.30	-0.96	-9.54	-3.72	1.71
10	*0.3*	8.01	-1.19	**-12.10**	-2.94	4.58	3.46	-2.95	**-14.29**	-5.87	1.79	2.64	-2.80	**-14.71**	-6.41	1.61
10	*0.4*	8.74	-4.39	**-17.74**	-4.79	3.61	3.54	-6.65	**-19.30**	-8.60	1.04	2.62	-4.57	**-19.40**	-9.26	1.12
10	*0.5*	**10.28**	**-10.90**	**-22.90**	-7.89	1.85	3.58	**-11.69**	**-23.86**	**-11.49**	-0.15	2.44	**-11.33**	**-23.93**	**-12.24**	0.27

Results II: CI Coverage

Confidence Interval Coverage for Slope Coefficients:
Multinomial Logistic Regression (Averaged Over *K* – 1 Coefficients)

		N = 250					N = 500					N = 1000				
K	*PM*	*MNOM*	*CART*	*PMM*	*RANK*	*GLOC*	*MNOM*	*CART*	*PMM*	*RANK*	*GLOC*	*MNOM*	*CART*	*PMM*	*RANK*	*GLOC*
3	0.1	0.945	0.935	0.948	0.949	0.951	0.939	0.926	0.933	0.945	0.943	0.941	**0.920**	**0.928**	0.937	0.945
3	0.2	0.933	**0.921**	0.940	0.953	0.949	0.941	**0.927**	**0.911**	0.948	0.951	0.937	**0.911**	**0.884**	0.935	0.951
3	0.3	0.931	**0.913**	0.936	0.955	0.957	**0.923**	**0.903**	**0.885**	0.949	0.958	**0.917**	**0.881**	**0.808**	**0.914**	0.953
3	0.4	**0.891**	**0.901**	**0.925**	0.946	0.951	**0.905**	**0.907**	**0.874**	0.936	0.961	**0.890**	**0.891**	**0.747**	**0.909**	0.950
3	0.5	**0.873**	**0.926**	**0.923**	0.944	0.939	**0.859**	**0.919**	**0.849**	**0.921**	0.950	**0.855**	**0.909**	**0.706**	**0.869**	0.947
5	0.1	0.952	0.957	0.965	0.957	0.953	0.948	0.946	0.953	0.955	0.950	0.949	0.941	0.942	0.953	0.951
5	0.2	0.947	0.945	0.956	0.958	0.955	0.940	0.937	0.934	0.956	0.953	0.936	**0.917**	**0.878**	0.944	0.946
5	0.3	**0.924**	0.932	0.948	0.956	0.951	**0.923**	**0.924**	**0.898**	0.945	0.952	**0.924**	**0.907**	**0.779**	0.933	0.944
5	0.4	**0.910**	**0.926**	0.936	0.957	0.955	**0.909**	**0.910**	**0.838**	0.945	0.953	**0.898**	**0.892**	**0.637**	**0.914**	0.948
5	0.5	**0.905**	0.939	**0.928**	0.956	0.950	**0.895**	**0.916**	**0.779**	0.939	0.952	**0.876**	**0.880**	**0.512**	**0.886**	0.945
7	0.1	0.945	0.947	0.955	0.954	0.949	0.944	0.945	0.956	0.957	0.948	0.952	0.953	0.953	0.961	0.957
7	0.2	0.936	0.942	0.955	0.953	0.947	0.938	0.935	0.941	0.955	0.947	0.943	0.937	**0.915**	0.953	0.948
7	0.3	**0.929**	0.935	0.952	0.957	0.949	0.931	0.931	**0.927**	0.955	0.949	0.931	**0.917**	**0.846**	0.946	0.952
7	0.4	**0.922**	0.933	0.949	0.958	0.951	**0.913**	**0.923**	**0.901**	0.951	0.952	**0.921**	**0.915**	**0.723**	0.931	0.951
7	0.5	**0.902**	0.938	0.940	0.957	0.939	**0.901**	0.930	**0.854**	0.949	0.955	**0.898**	**0.909**	**0.584**	**0.912**	0.949
10	0.1	0.935	0.941	0.958	0.949	0.939	0.944	0.941	0.959	0.955	0.947	0.945	0.947	0.958	0.955	0.950
10	0.2	**0.929**	0.938	0.963	0.956	0.944	0.940	0.945	0.960	0.959	0.949	0.940	0.941	0.946	0.960	0.951
10	0.3	**0.918**	0.940	0.966	0.959	0.943	0.932	0.943	0.949	0.964	0.952	0.932	0.933	**0.895**	0.954	0.951
10	0.4	**0.911**	0.941	0.965	0.964	0.946	**0.919**	0.936	0.932	0.962	0.951	**0.922**	0.931	**0.835**	0.945	0.952
10	0.5	**0.908**	0.949	0.959	0.965	0.940	**0.908**	0.948	**0.909**	0.955	0.949	**0.912**	0.943	**0.731**	0.936	0.953

Recommendations I: What Doesn't Work

GLOC performed very well here, but it is highly sensitive to model complexity.

- Previous authors have shown the General Location Model to break with *real-world-sized* problems (Belin, Hu, Young, & Grusky, 1999).
- A follow-up confirms that the performance of GLOC rapidly declines as the contingency table grows.

Recommendations I: What Doesn't Work

GLOC performed very well here, but it is highly sensitive to model complexity.

- Previous authors have shown the General Location Model to break with *real-world-sized* problems (Belin et al., 1999).
- A follow-up confirms that the performance of GLOC rapidly declines as the contingency table grows.

PMM should not be used in this context.

- It produces substantially biased point estimates and very low CI coverage in many conditions.
  - The problem is especially bad for larger N.

Recommendations I: What Doesn't Work

GLOC performed very well here, but it is highly sensitive to model complexity.

- Previous authors have shown the General Location Model to break with *real-world-sized* problems (Belin et al., 1999).
- A follow-up confirms that the performance of GLOC rapidly declines as the contingency table grows.

PMM should not be used in this context.

- It produces substantially biased point estimates and very low CI coverage in many conditions.
  - The problem is especially bad for larger N.

MICE with CART is not recommended for general use.

- Its CI coverage rates were too unstable.

Recommendations II: What Works

– Best In Show –

For the best balance of point estimation performance and CI coverage, the Allison (2002) Ranking method is preferred.

- The SEs may not be trustworthy when $N \gg 500$.

Recommendations II: What Works

– Best In Show –

For the best balance of point estimation performance and CI coverage, the Allison (2002) Ranking method is preferred.

- The SEs may not be trustworthy when $N \gg 500$.

– Runner Up –

If accurate standard errors are not a concern, then MICE with MNOM is the best option.

Conclusions

Unfortunately, there is no "magic bullet" technique.

Conclusions

Unfortunately, there is no "magic bullet" technique.

- GLOC performed well but is prohibitively sensitive to model complexity.

Conclusions

Unfortunately, there is no “magic bullet” technique.

- GLOC performed well but is prohibitively sensitive to model complexity.
- Most of the examined methods are subject to a considerable bias variance trade-off.

Conclusions

Unfortunately, there is no "magic bullet" technique.

- GLOC performed well but is prohibitively sensitive to model complexity.
- Most of the examined methods are subject to a considerable bias variance trade-off.
  - MNOM imputation was consistently unbiased but had some of the poorest CI coverage.

Conclusions

Unfortunately, there is no "magic bullet" technique.

- GLOC performed well but is prohibitively sensitive to model complexity.
- Most of the examined methods are subject to a considerable bias variance trade-off.
  - MNOM imputation was consistently unbiased but had some of the poorest CI coverage.
  - The optimal choice will probably depend on whether accurate point estimates or correct standard errors are more important.

Conclusions

Unfortunately, there is no "magic bullet" technique.

- GLOC performed well but is prohibitively sensitive to model complexity.
- Most of the examined methods are subject to a considerable bias variance trade-off.
  - MNOM imputation was consistently unbiased but had some of the poorest CI coverage.
  - The optimal choice will probably depend on whether accurate point estimates or correct standard errors are more important.

Fortunately, there are good solutions for the type of problems most frequently encountered in social science research.

Conclusions

Unfortunately, there is no "magic bullet" technique.

- GLOC performed well but is prohibitively sensitive to model complexity.
- Most of the examined methods are subject to a considerable bias variance trade-off.
  - MNOM imputation was consistently unbiased but had some of the poorest CI coverage.
  - The optimal choice will probably depend on whether accurate point estimates or correct standard errors are more important.

Fortunately, there are good solutions for the type of problems most frequently encountered in social science research.

- When imputing the DV in MLogit models with $N \leq 500$, the Allison (2002) Ranking method should perform well.

Limitations

- The models used here were trivially simple.
- Since the data were simulated via a MLogit model, the positive results for MICE with MNOM may be too optimistic.
- The results reported here could be somewhat software dependent.
- The recommendations differ when using the imputed nominal variables as IVs in multiple linear regression.

References

Agresti, A. (2007). *An introduction to categorical data analysis*. Hoboken, NJ: Wiley-Interscience.

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage Publications.

Belin, T. R., Hu, M.-Y., Young, A. S., & Grusky, O. (1999). Performance of a general location model with an ignorable misssing-data assumption in a multivariate mental health services study. *Statistics in Medicine*, *18*, 3123–3135.

Honaker, J., King, G., & Blackwell, M. (2012, June). Amelia ii: A program for missing data [Computer software manual]. Retrieved from `http://cran.r-project.org/web/packages/Amelia/index.html` (R package version 1.6-3)

Little, R. J. A., & Schluchter, M. D. (1985). Maximum likelihood estimation for mixed continuous and categorical variables with missing values. *Biometrika*, *72*, 497–512.

Schafer, J. L. (2013, August). Estimation/multiple imputation for mixed categorical and continuous data [Computer software manual]. Retrieved from `http://sites.stat.psu.edu/~jls/misoftwa.html` (R package version 1.0-8)

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*(3), 1–67. Retrieved from `http://www.jstatsoft.org/v45/i03/`

Thank you for your time.

QUESTIONS/COMMENTS?

Percentage Relative Bias for Slope Coefficients:
Multiple Linear Regression (Averaged Over *K* – 1 Coefficients)

		N = 250					N = 500					N = 1000				
K	*PM*	*MNOM*	*CART*	*PMM*	*RANK*	*GLOC*	*MNOM*	*CART*	*PMM*	*RANK*	*GLOC*	*MNOM*	*CART*	*PMM*	*RANK*	*GLOC*
3	0.1	0.33	-0.42	-3.69	-1.99	-0.09	-0.56	-1.08	-4.28	-2.90	-0.78	-0.33	-1.03	-3.95	-2.67	-0.43
3	0.2	0.47	-0.53	-7.63	-4.25	-0.36	-0.47	-1.38	-7.82	-5.05	-0.91	-0.23	-1.62	-7.48	-4.90	-0.49
3	0.3	-0.02	-2.57	**-12.31**	-6.91	-1.45	-0.61	-2.29	**-11.89**	-7.49	-1.40	-0.35	-2.55	**-11.35**	-7.31	-0.79
3	0.4	0.37	-2.61	**-16.92**	-8.83	-1.75	-0.68	-4.43	**-16.26**	-9.88	-1.83	-0.25	-3.74	**-15.27**	-9.49	-0.90
3	0.5	-0.26	**-13.22**	**-25.59**	**-10.66**	-3.93	-1.09	**-14.67**	**-24.31**	**-12.00**	-3.21	-0.32	**-15.74**	**-22.57**	**-11.57**	-1.51
5	0.1	-0.89	-1.00	-1.67	-3.91	-0.92	-0.51	-0.55	-1.17	-3.65	-0.53	-0.26	-0.23	-0.81	-3.51	-0.27
5	0.2	-0.63	-0.83	-2.33	-6.75	-0.71	-0.43	-0.32	-1.71	-6.77	-0.40	-0.05	0.25	-1.19	-6.57	-0.03
5	0.3	-0.87	-0.93	-3.53	-9.83	-0.83	-0.60	-0.19	-2.58	-9.94	-0.52	-0.20	-0.06	-2.00	-9.95	-0.14
5	0.4	-1.44	-2.22	-5.28	**-12.98**	-1.14	-0.94	-1.23	-3.79	**-13.22**	-0.69	-0.34	-0.56	-2.88	**-13.19**	-0.25
5	0.5	-2.39	-5.90	-9.19	**-15.95**	-1.00	-1.32	-4.31	-6.66	**-16.27**	-0.21	-0.47	-3.49	-5.11	**-16.34**	0.07
7	0.1	0.36	0.52	0.13	-2.88	0.52	-0.15	-0.10	-0.33	-3.55	-0.06	-0.21	-0.13	-0.44	-3.73	-0.19
7	0.2	0.31	0.70	-0.08	-6.06	0.75	-0.10	0.07	-0.52	-6.83	0.08	-0.19	-0.13	-0.67	-7.18	-0.10
7	0.3	-0.34	0.31	-0.91	-9.63	0.53	-0.36	-0.03	-0.93	**-10.34**	0.08	-0.42	-0.20	-1.04	**-10.80**	-0.21
7	0.4	-1.00	0.30	-1.68	**-12.62**	0.67	-0.42	0.35	-1.24	**-13.56**	0.37	-0.55	-0.17	-1.34	**-14.32**	-0.19
7	0.5	-2.77	-1.79	-4.04	**-16.16**	0.66	-1.17	-0.96	-2.47	**-16.98**	0.50	-0.90	-0.31	-2.15	**-17.89**	-0.04
10	0.1	-0.09	0.00	-0.01	-3.51	0.18	0.05	0.11	0.05	-3.61	0.18	0.03	0.09	0.00	-3.76	0.10
10	0.2	-0.47	0.06	-0.14	-7.02	0.22	-0.12	0.13	0.02	-7.28	0.22	-0.05	0.09	-0.05	-7.55	0.14
10	0.3	-1.04	-0.20	-0.36	**-10.46**	0.19	-0.45	0.02	-0.16	**-11.00**	0.19	-0.29	-0.02	-0.17	**-11.42**	0.10
10	0.4	-2.21	-0.57	-0.79	**-13.78**	0.20	-0.91	-0.12	-0.41	**-14.64**	0.21	-0.44	0.07	-0.25	**-15.13**	0.18
10	0.5	-4.48	-1.19	-1.21	**-16.89**	0.09	-2.05	-0.80	-0.64	**-18.12**	0.32	-1.03	-0.36	-0.37	**-18.91**	0.23

Results IV: MLR CI Coverage

Confidence Interval Coverage for Slope Coefficients:
Multiple Linear Regression (Averaged Over *K* – 1 Coefficients)

		N = 250					N = 500					N = 1000				
K	***PM***	***MNOM***	***CART***	***PMM***	***RANK***	***GLOC***	***MNOM***	***CART***	***PMM***	***RANK***	***GLOC***	***MNOM***	***CART***	***PMM***	***RANK***	***GLOC***
3	*0.1*	0.952	0.958	0.962	0.968	0.957	0.952	0.944	0.955	0.962	0.953	0.938	0.933	0.935	0.950	0.938
3	*0.2*	0.952	0.952	0.952	**0.972**	0.961	0.954	0.937	0.947	**0.970**	0.958	0.938	**0.901**	**0.896**	0.957	0.946
3	*0.3*	0.937	**0.924**	0.940	0.961	0.950	0.941	**0.916**	**0.916**	0.969	0.947	**0.926**	**0.891**	**0.856**	0.961	0.939
3	*0.4*	**0.924**	**0.919**	**0.923**	**0.975**	0.956	0.938	**0.913**	**0.864**	**0.972**	0.958	**0.926**	**0.895**	**0.767**	0.958	0.948
3	*0.5*	0.930	0.969	0.931	**0.979**	0.960	0.939	**0.977**	**0.850**	**0.977**	0.961	**0.926**	**0.980**	**0.709**	0.952	0.956
5	*0.1*	0.951	0.954	0.954	**0.971**	0.954	0.950	0.950	0.953	0.963	0.953	0.950	0.945	0.951	0.960	0.952
5	*0.2*	0.948	0.941	0.954	**0.977**	0.958	0.952	0.946	0.948	0.967	0.959	0.942	**0.929**	0.942	0.951	0.952
5	*0.3*	0.940	0.938	0.947	**0.975**	0.956	0.948	0.932	0.940	0.963	0.959	0.939	**0.917**	**0.926**	0.939	0.954
5	*0.4*	0.930	**0.920**	0.942	**0.977**	0.959	**0.926**	**0.918**	**0.924**	0.951	0.955	0.934	**0.911**	**0.903**	**0.923**	0.954
5	*0.5*	**0.923**	0.947	0.946	**0.979**	0.961	**0.920**	**0.971**	**0.913**	0.945	0.957	**0.926**	**0.984**	**0.895**	**0.895**	0.964
7	*0.1*	0.949	0.947	0.957	**0.975**	0.950	0.936	0.936	0.941	0.965	0.938	0.953	0.947	0.951	0.962	0.956
7	*0.2*	0.946	0.943	0.954	**0.982**	0.951	**0.928**	**0.922**	0.936	0.969	0.940	0.945	0.933	0.936	0.952	0.950
7	*0.3*	0.933	**0.929**	0.947	**0.984**	0.952	**0.927**	**0.917**	0.932	0.967	0.940	0.933	**0.915**	**0.923**	**0.924**	0.949
7	*0.4*	**0.918**	**0.909**	0.939	**0.980**	0.948	**0.917**	**0.915**	**0.928**	0.956	0.941	**0.922**	**0.911**	**0.907**	**0.889**	0.953
7	*0.5*	**0.915**	0.934	0.943	**0.978**	0.948	**0.897**	0.948	**0.922**	0.945	0.948	**0.907**	0.970	**0.896**	**0.849**	0.942
10	*0.1*	0.957	0.949	0.958	**0.991**	0.957	0.947	0.945	0.951	**0.987**	0.949	0.952	0.941	0.950	0.982	0.951
10	*0.2*	0.949	0.940	0.952	**0.997**	0.953	0.949	0.940	0.956	**0.989**	0.956	0.947	0.933	0.949	0.966	0.954
10	*0.3*	0.945	0.938	0.949	**0.997**	0.952	0.943	0.931	0.945	**0.983**	0.957	0.944	**0.918**	0.945	0.935	0.955
10	*0.4*	0.931	**0.916**	0.939	**0.990**	0.949	**0.929**	**0.929**	0.938	**0.975**	0.954	**0.928**	**0.914**	**0.929**	**0.894**	0.954
10	*0.5*	0.942	0.936	0.942	**0.993**	0.949	**0.921**	0.960	0.934	0.963	0.956	**0.915**	**0.975**	**0.919**	**0.857**	0.953