

**Workshop on Missing Data**  
**Frederick J. Boehmke**  
**May 23-24, 2012**  
**University of Kentucky**

**Readings**

These articles are intended to touch on various aspects of missing data. Starred books and articles are recommended as good background sources while the others are included for more complete treatments or for a focus on more specific issues.

*Background on Missing Data*

\*Allison, Paul D. 2001. *Missing data: Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: Sage (Vol. 136).

Little, Roderick J. A. and Donald B. Rubin. 1987. *Statistical Analysis with Missing Data*. John Wiley & Sons.

\*Joseph L. Schafer and John W. Graham. 2002. "Missing Data: Our View of the State of the Art." *Psychological Methods* 7 (2): 147–177.

*Addressing Missing Data*

\*Abayomi, Kobi; Andrew Gelman; and Marc Levy. 2008. "Diagnostics for Multivariate Imputations." *Applied Statistics* 57 (3): 273–291.

Allison, Paul D. 2003. "Missing Data Techniques for Structural Equation Modeling." *Journal of Abnormal Psychology* 112 (4): 545–557.

Franklin, Charles H. 1989. "Estimation Across Data Sets: Two-Stage Auxiliary Instrumental Variables Estimation (2SAIV)." *Political Analysis* 1: 1-23.

\*Honaker, James; Gary King; and Matthew Blackwell. 2012. *Amelia II: A Program for Missing Data* (Windows Version 1.6.2). Cambridge, MA: Harvard University, <http://GKing.harvard.edu>.

Read the documentation at <http://gking.harvard.edu/amelia/>.

Honaker, James and Gary King. 2010. "What to do About Missing Values in Time Series Cross-Section Data." *American Journal of Political Science* 54 (2): 561-581.

\*King, Gary; James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95: 49-69.

Meng, Xiao-Li and Donald B. Rubin. 1992. "Performing Likelihood Ratio Tests with Multiply-Imputed Data." *Biometrika* 79: 103-111.

Murphy, Kevin M. and Robert H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business and Economic Statistics* 3: 370-379.

White, Ian R.; Patrick Royston; and Angela M. Wood. 2011. "Multiple Imputation Using Chained Equations: Issues and Guidance for Practice." *Statistics in Medicine* 30 (4): 377-399.

### Non-Random Sample Selection

Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: The University of California Press.

Bloom, David E. and Mark R. Killingsworth. 1985. "Correcting for Truncation Bias Caused by a Latent Truncation Variable." *Journal of Econometrics* 27:131-135.

\*Dubin, Jeffrey A. and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit and Probit Models." *Sociological Methods and Research* 18: 360-390.

\*Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5/4:475-492.

\*Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153-161.

Winship, Christopher and Robert D. Mare. 1992. "Models for Sample Selection Bias." *Annual Review of Sociology* 18:327-350.