The Stata Journal (2009)
9, Number 3, pp. 388–397

Improved degrees of freedom for multivariate significance tests obtained from multiply imputed, small-sample data

Yulia V. Marchenko
StataCorp
College Station, TX
ymarchenko@stata.com

Jerome P. Reiter
Department of Statistical Science
Duke University
Durham, NC
jerry@stat.duke.edu

Abstract. We propose improvements to existing degrees of freedom used for significance testing of multivariate hypotheses in small samples when missing data are handled using multiple imputation. The improvements are for 1) tests based on unrestricted fractions of missing information and 2) tests based on equal fractions of missing information with $M(p-1) \leq 4$, where M is the number of imputations and p is the number of tested parameters. Using the `mi` command available as of Stata 11, we demonstrate via simulation that using these adjustments can result in a more sensible degrees of freedom (and hence closer-to-nominal rejection rates) than existing degrees of freedom.

Keywords: st0170, multiple imputation, degrees of freedom, sample, missing, testing, multivariate

1 Introduction

Multiple imputation developed by Rubin (1987) is a popular approach for handling missing data. The basic idea is for the data collector or imputer to simulate values for the missing data repeatedly by sampling from predictive distributions of the missing values. The data analyst, who may be the same person as the imputer or may be a secondary data user, performs the desired analysis on each completed dataset and combines the results using simple formulas (Rubin 1987, 76–77). As of Stata 11, the `mi` command provides methods for generating multiple imputations and implements the formulas for combining results (StataCorp 2009). Users also can perform multiple imputation by using `ice` and `mim` (Royston 2004, 2005a, 2005b, 2007; Carlin, Galati, and Royston 2008; and Royston, Carlin, and White 2009). For reviews of multiple imputation, see Schafer (1997), Little and Rubin (2002), and Reiter and Raghunathan (2007).

Often analysts seek to test multivariate hypotheses, for example, if several regression coefficients are equal to zero. Rubin (1987) suggests two approaches to doing so with multiply imputed data. The first approach, which is the most widely used method, presumes that the fractions of missing information (FMI) are equal across the parameters of interest. A reference F distribution for this method was derived by Li, Raghunathan, and Rubin (1991). The second approach does not presume equal

FMIs; however, it may not give well calibrated p-values unless the number of imputed datasets is large (Li, Raghunathan, and Rubin 1991).

The derivations of these test statistics and their reference distributions presume infinite sample size. However, Reiter (2007) demonstrates that, for the equal FMI test, the infinite sample-size assumption can result in nonsensical procedures. For example, in modest samples, the computed degrees of freedom for the reference distributions can exceed the number of cases in the dataset, which should not be possible. A related phenomenon is illustrated by Barnard and Rubin (1999), who derive small-sample degrees of freedom for univariate inferences.

Reiter (2007) goes on to develop small-sample degrees of freedom for the equal FMI test that results in better performance than the infinite sample degrees of freedom of Li, Raghunathan, and Rubin (1991). However, Reiter's (2007) degrees of freedom requires $M(p-1) > 4$, where M is the number of imputations and p is the number of tested parameters. While this case may not be a concern in practice because analysts can set M to be large, it nonetheless must be accounted for when designing software to implement multiple imputation. For multivariate tests based on unrestricted FMIs, we are not aware of any published research on small-sample adjustments to the degrees of freedom.

Motivated by the development of `mi`, we propose to fill these gaps in the literature. Specifically, we present small-sample degrees of freedom for the unrestricted FMI test and for the equal FMI test with $M(p-1) \leq 4$. We demonstrate with simulation results that using the adjusted degrees of freedom can result in more sensible reference distributions (and hence closer-to-nominal rejection rates) than using degrees of freedom based on infinite sample sizes.

2 Significance tests with multiple imputation

We first review the unrestricted and equal FMI tests. Let $\mathbf{q}$ be the $p \times 1$ vector of parameters of interest, such as p regression coefficients. In each completed dataset i, where $i = 1, \ldots, M$, let $\widehat{\mathbf{q}}_i$ be the completed-data estimate of $\mathbf{q}$, and let $\widehat{\mathbf{U}}_i$ be its associated completed-data variance estimate. The analyst combines each $\widehat{\mathbf{q}}_i$ and $\widehat{\mathbf{U}}_i$ using

$$\overline{\mathbf{q}} = \frac{1}{M}\sum_{i=1}^{M}\widehat{\mathbf{q}}_i$$

$$\mathbf{T} = \overline{\mathbf{U}} + (1+\frac{1}{M})\mathbf{B}$$

Here $\overline{\mathbf{U}} = \sum_{i=1}^{M}\widehat{\mathbf{U}}_i/M$ is the within-imputation variance–covariance matrix, and $\mathbf{B} = \sum_{i=1}^{M}(\mathbf{q}_i - \overline{\mathbf{q}})(\mathbf{q}_i - \overline{\mathbf{q}})'/(M-1)$ is the between-imputation variance–covariance matrix. The analyst can use $\overline{\mathbf{q}}$ as a point estimate of $\mathbf{q}$ and $\mathbf{T}$ as an estimate of the variance of $\overline{\mathbf{q}}$.

We now suppose that the analyst seeks to test the null hypothesis, H_0: $\mathbf{q} = \mathbf{q}_0$. The *unrestricted FMI test* proposed by Rubin (1987) is

$$(\mathbf{q}_0 - \overline{\mathbf{q}})\mathbf{T}^{-1}(\mathbf{q}_0 - \overline{\mathbf{q}})'/p \sim F_{p,\nu} \tag{1}$$

where

$$\begin{aligned} \nu &= (M-1)(1+1/r_{\text{ave}})^2 \\ r_{\text{ave}} &= (1+1/M)\text{tr}(\mathbf{B}\overline{\mathbf{U}}^{-1})/p \end{aligned}$$

Here r_{ave} is the average relative variance increase due to missing data.

Note that even under the assumption of infinite sample size, multivariate testing uses an F reference distribution rather than a chi-squared distribution. This is because of the fact that the variance $\mathbf{T}$ in the test statistic (1) involves estimates of the within- and between-imputations variances based on the finite number of M imputations. Therefore, the denominator degrees-of-freedom parameter ν in (1) represents the amount of independent information used to estimate the variance after accounting for a finite number of imputations. In standard multiple-imputation contexts, this amount of information theoretically cannot exceed the number of cases in the dataset, which sometimes happens with the approximations of Rubin (1987) and Li, Raghunathan, and Rubin (1991).

The *equal FMI test* originally suggested by Rubin (1987) is

$$(1+r_{\text{ave}})^{-1}(\mathbf{q}_0 - \overline{\mathbf{q}})\overline{\mathbf{U}}^{-1}(\mathbf{q}_0 - \overline{\mathbf{q}})'/p \sim F_{p,(p+1)\nu/2} \tag{2}$$

Reiter (2007) uses the same test statistic as (2) with an alternative denominator degrees of freedom for the F distribution appropriate when $M(p-1) > 4$.

The key distinction between the two test statistics is the variance inside the quadratic form. The unrestricted FMI test uses $\mathbf{T}$, whereas the equal FMI test uses $(1+r_{\text{ave}})\overline{\mathbf{U}}$. This difference arises because of the equal FMI condition. To see this, define $\mathbf{B}_\infty = \lim \mathbf{B}$ as $M \to \infty$, and define $\mathbf{T}_\infty = \overline{\mathbf{U}} + (1+1/M)\mathbf{B}_\infty$; we could obtain these values if we had an infinite number of datasets to estimate $\mathbf{B}$ and $\mathbf{T}$. Under equal FMIs, $\overline{\mathbf{U}} = \rho\mathbf{B}_\infty$ for some constant ρ, and thus $\mathbf{T}_\infty = (1+\rho)\overline{\mathbf{U}}$. The relative variance increase, r_{ave}, in (2) is an estimate of ρ.

At first glance, the unrestricted test would seem to be always preferable because it is derived under more general conditions. However, Rubin (1987) shows that the unrestricted test can perform poorly when M is small relative to p because $\mathbf{B}$ can be unreliable. Essentially, using $\mathbf{B}$ to estimate $\mathbf{B}_\infty$ from the M datasets is akin to estimating a $p \times p$ covariance matrix with only M observations, which can be problematic when $M < p$. Using the equal FMI test mitigates these difficulties because the analyst estimates only one parameter, ρ, rather than $p^2 + p(p-1)/2$ parameters. Li, Raghunathan, and Rubin (1991) demonstrate that testing procedures based on the assumption of equal FMIs perform well as long as the fractions do not vary substantially.

3 Small-sample degrees-of-freedom adjustments

We now consider adjustments for the denominator degrees of freedom in the reference distributions in (1) and, for cases with $M(p-1) \leq 4$, in (2) to reflect small samples.

For the unrestricted FMI test, we propose to use the small-sample degrees of freedom of Barnard and Rubin (1999) in place of ν in (1). That is, we use

$$\nu_{\rm br} = \left(\nu_\star^{-1} + \widehat{\nu}_{\rm obs}^{-1}\right)^{-1}$$

where

$$\begin{aligned} \nu_\star &= (M-1)\gamma_{\rm ave}^{-2} \\ \widehat{\nu}_{\rm obs} &= (1-\gamma_{\rm ave})\nu_{\rm com}(\nu_{\rm com}+1)/(\nu_{\rm com}+3) \\ \gamma_{\rm ave} &= (1+1/M){\rm tr}(\mathbf{B}\mathbf{T}^{-1})/p \end{aligned}$$

Here $\nu_{\rm com}$ is the degrees of freedom if the data were complete, and $\gamma_{\rm ave}$ is the approximate average FMI. The quantity $\nu_{\rm br}$ has several features that led Barnard and Rubin (1999) to recommend its general use, regardless of the sample size. First, $\nu_{\rm br} \leq \nu_{\rm com}$, whereas ν can exceed $\nu_{\rm com}$. This property of $\nu_{\rm br}$ is desirable because the presence of missing data should reduce the degrees of freedom rather than increase it. Second, $\nu_{\rm br} < \nu$ with approximate equality when the sample size is large, so using $\nu_{\rm br}$ instead of ν is slightly conservative in large samples. Third, $\nu_{\rm br}$ is always between $\nu_{\rm com}$ and ν, making it a compromise degrees of freedom.

Barnard and Rubin (1999) illustrate the effectiveness of this degrees of freedom for univariate inferences. To our knowledge, $\nu_{\rm br}$ is rarely, if ever, used for multivariate inferences. However, Barnard and Rubin (1999) note that the steps in the derivation of $\nu_{\rm br}$ for multivariate $\mathbf{q}$ follow immediately under equal FMIs. Hence, by using $\nu_{\rm br}$ for the degrees of freedom in the unrestricted test, we lean on the equal FMI assumption to avoid unrealistic degrees of freedom, but we do allow the variance in the quadratic form to be estimated without the restriction.

For the equal FMI test, we suggest a refinement to the degrees of freedom of Reiter (2007) for cases when $M(p-1) \leq 4$. Here we again propose to substitute $\nu_{\rm br}$ for ν in (2). This is similar in spirit to the suggestion of Li, Raghunathan, and Rubin (1991), who use (2) for cases when $M(p-1) \leq 4$ for their large-sample tests. The primary difference is that we use a degrees of freedom, $\nu_{\rm br}$, that has more desirable properties in small samples.

4 Simulation studies of properties of adjustments

The proposed adjusted degrees of freedom are ad hoc in nature. As noted by Rubin (1987), there is little way around such constructions, because we are approximating complicated Bayesian integrals with simple distributions. Thus it is imperative to evaluate the operating characteristics of tests based on these procedures by using simulation studies.

In all simulations studies, we generate an outcome, Y, and covariates, $(X_1, X_2, \ldots, X_p)$, where p depends on the simulation scenario, for 50 observations. The covariates are sampled from a multivariate normal distribution with means equal to zero, variances equal to one, and all pairwise correlations equal to 0.5. The outcome is sampled from a normal distribution with mean equal to zero and variance equal to one independently of covariates. The `simulate` command is used to generate the data. We investigate the empirical significance levels of the procedures when testing if all coefficients in the regression of Y on $(X_1, X_2, \ldots, X_p)$ are equal to zero; that is, we test H_0: $\beta_1 = \beta_2 = \cdots = \beta_p = 0$. The empirical significance levels are based on 10,000 replications.

We consider four simulation scenarios in which we vary FMIs; these are described in sections 4.1 and 4.2. Multiple imputations are performed using `mi impute mvn`, which implements multivariate normal imputation. The estimation step is performed using `mi estimate`. The results of the equal FMI test are obtained from the default settings of `mi test`. The results of the unrestricted FMI test are obtained by specifying the `ufmitest` option with `mi test`. The results from the corresponding large-sample tests are obtained by specifying the `nosmall` option with `mi test`.

4.1 Small-sample adjustment for the unrestricted FMI test

Scenarios 1, 2, and 3 use $p = 4$ covariates, and scenario 4 uses $p = 5$ covariates. In scenario 1, we randomly delete 10% of the 50 observations, which corresponds to approximately equal fractions of information missing due to nonresponse. Scenario 2 is similar to scenario 1 but with 30% of the observations deleted. In scenario 3, we introduce variation among the FMIs by randomly deleting 10% of the data from X_2, 20% of the data from X_4, and 35% of the data from X_3; here X_1 and Y are complete. Scenario 4 represents a relatively large deviation from equal FMI with increased missingness: 10% of the data are deleted from X_4, 30% of the data are deleted from X_2, 50% of the data are deleted from X_1 and X_3, and X_5 and Y are complete. We use $M = 20$ multiple imputations.

Table 1 displays key results from the 10,000 replications. Across all scenarios, the small-sample degrees of freedom, ν_{br}, is more sensible than the large-sample degrees of freedom in (2), ν, which always greatly exceeds the sample size of 50. The unrestricted FMI test using ν_{br} provides close-to-nominal significance levels and is somewhat conservative. In contrast, the unrestricted FMI test using ν is anticonservative; its empirical significance levels always exceed the corresponding nominal significance levels. The difference between the empirical and nominal levels is always smaller for the test based on ν_{br}. Thus we recommend ν_{br} over ν for the unrestricted FMI test.

Table 1. Simulated significance levels for the unrestricted FMI test of all coefficients equal to zero. $\bar{\nu}$ denotes the denominator degrees of freedom averaged over replications.

Scenario	DF	$\bar{\nu}$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
10% missing, equal FMI (1)	small	36.22	0.0941	0.0455	0.0081
	large	1180.76	0.1164	0.0636	0.0168
30% missing, equal FMI (2)	small	22.57	0.0819	0.0361	0.0069
	large	130.58	0.1103	0.0587	0.0162
max 35% missing, unequal FMI (3)	small	33.04	0.0965	0.0477	0.0082
	large	660.37	0.1209	0.0674	0.0185
max 50% missing, unequal FMI (4)	small	23.57	0.0892	0.0452	0.0089
	large	151.24	0.1224	0.0725	0.023

4.2 Small-sample adjustment for the equal FMI test

To evaluate the performance of the testing procedure under the equal FMI assumption for the case when $M(p-1) \leq 4$, we consider four simulation scenarios similar to those used for the unrestricted test. We use $p = 2$ covariates, (X_1, X_2), and $M = 3$ imputations so that $M(p-1) = 3$. In scenario 1, we randomly delete 10% of all observations. In scenario 2, we randomly delete 30% of all observations. In scenario 3, we randomly delete 10% of the data from X_1 and 35% of the data from X_2. In scenario 4, we randomly delete 30% of the data from X_1 and 50% of the data from X_2.

Table 2 displays the key results from the 10,000 replications. In all cases, ν_{br} is less than the sample size of 50, whereas the degrees of freedom in (2) far exceeds 50. For the scenarios with modest FMIs (scenarios 1 and 3), the test based on ν_{br} generally has closer-to-nominal empirical significance levels than the test based on the degrees of freedom in (2). However, the picture is less clear with large FMIs (scenarios 2 and 4): the levels for the test based on ν_{br} are closer to nominal when $\alpha = 0.01$ but not when $\alpha \in (0.05, 0.10)$. For the scenarios with equal FMIs, the test based on ν_{br} is conservative, whereas the test based on the degrees of freedom in (2) can be anticonservative. For both degrees of freedom, the tests in scenarios 3 and 4 are reasonably well calibrated despite the unequal FMI, although the levels for the test based on ν_{br} can exceed the nominal α in this case.

(*Continued on next page*)

Table 2. Simulated significance levels for the equal FMI test of two coefficients equal to zero. $\overline{\nu}$ denotes the denominator degrees of freedom averaged over replications.

Scenario	DF	$\overline{\nu}$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
10% missing, equal FMI (1)	small	40.09	0.0933	0.0485	0.0099
	large	2025.32	0.1017	0.0558	0.0123
30% missing, equal FMI (2)	small	15.65	0.0795	0.0406	0.0096
	large	97.74	0.0886	0.0480	0.0127
max 35% missing, unequal FMI (3)	small	31.08	0.1023	0.0536	0.0132
	large	888.59	0.1111	0.0625	0.0170
max 50% missing, unequal FMI (4)	small	17.22	0.0872	0.0435	0.011
	large	154.50	0.0967	0.0509	0.0147

Taking these results as a whole, we recommend using the adjusted degrees of freedom when $M(p-1) \leq 4$. The test based on ν_{br} tends to be conservative when the assumption of equal FMI is true or nearly true, which is when these tests perform best. Of course, data analysts need not force themselves into choosing between these two degrees of freedom. They can increase M sufficiently so that $M(p-1) > 4$ and use the degrees of freedom developed by Reiter (2007) for the equal FMI test, which has been shown to perform well with approximately equal FMIs. For small sample sizes, using a large M should not be a computational burden and can greatly improve analyses.

5 Illustration of testing with multiple imputation in Stata

As an example of testing multivariate hypotheses, we use multiply imputed data on house resale prices, `mhouses1993s30.dta`, from example 2 in the Stata manual for the `mi estimate` command, [MI] **mi estimate**. The original data are provided by the Albuquerque Board of Realtors and distributed by the Data and Story Library (http://lib.stat.cmu.edu/DASL/Stories/homeprice.html).

We are interested in the effect of house characteristics like square footage, age of house, and amount of taxes paid on house prices, which we estimate with a linear regression. The data contain missing values on age and taxes. `mhouses1993s30.dta` contains $M = 30$ imputations created using `mi impute mvn`, which invokes multivariate normal imputations. The imputation strategies are described in detail in example 3 of the Stata manual entry [MI] **mi impute mvn**.

Below we present the results of the regression on the multiply imputed data. These results are obtained by using `mi estimate`. We specify the `vartable` option to display the estimated FMIs. The test statistic and p-value for the test of all coefficients equaling zero are displayed in the regression output header. By default, this test is based on the equal FMI test with the degrees of freedom of Reiter (2007). Based on this test, there is significant evidence to reject the null hypothesis that all coefficients equal zero.

```
. use http://www.stata-press.com/data/r11/mhouses1993s30
(Albuquerque Home Prices Feb15-Apr30, 1993)
. mi estimate, vartable: regress price sqft age nfeatures ne custom corner tax
Multiple-imputation estimates                    Imputations     =         30
Variance information

                     Imputation variance                             Relative
               Within   Between     Total       RVI       FMI    efficiency

        sqft  .004442   .003623   .008186   .842713   .464984       .984737
         age  .277762   .896309   1.20395   3.33446   .778164       .974717
   nfeatures  157.333   26.7139   184.937   .175452   .150568       .995006
          ne  1104.74   114.734   1223.29   .107319   .097502        .99676
      custom  1783.12   85.8858   1871.87   .049772    .04756       .998417
      corner  1548.13   93.6976   1644.95    .06254   .059084       .998034
         tax  .012421    .00814   .020832   .677183   .410355       .986506
       _cons  3834.84   257.487   4100.91   .069382   .065152       .997833

Linear regression                                Number of obs   =        117
                                                 Average RVI     =     0.5415
                                                 Complete DF     =        109
DF adjustment:   Small sample                    DF:     min     =      16.42
                                                         avg     =      72.83
                                                         max     =     101.18
Model F test:       Equal FMI                    F(   7,   96.3) =      45.63
Within VCE type:          OLS                    Prob > F        =     0.0000

       price      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

        sqft   .2900879   .0904748     3.21   0.003      .1073624    .4728134
         age  -.7524605   1.097246    -0.69   0.502     -3.073675    1.568754
   nfeatures   4.361055   13.59917     0.32   0.749     -22.67719     31.3993
          ne   5.495913   34.97562     0.16   0.875     -63.95148    74.94331
      custom   132.3453   43.26507     3.06   0.003      46.52087    218.1697
      corner  -66.95606   40.55801    -1.65   0.102     -147.4264    13.51429
         tax   .5516444   .1443319     3.82   0.000      .2612817     .842007
       _cons   130.3491   64.03837     2.04   0.044      3.277868    257.4203
```

We can use `mi test` to test hypotheses about subsets of coefficients. Suppose that we seek to test the null hypothesis that the coefficients for `age`, `nfeatures` (number of certain features), and `ne` (whether the city is located in the northeast, largest residential, area) all equal zero. By default, `mi test` performs the equal FMI test, as illustrated below.

```
. mi test age nfeatures ne
note: assuming equal fractions of missing information
 ( 1)  age = 0
 ( 2)  nfeatures = 0
 ( 3)  ne = 0
       F(  3,  70.4) =    0.39
            Prob > F =    0.7639
```

However, from the output of `mi estimate`, the assumption of equal FMIs for `age`, `nfeatures`, and `ne` does not seem plausible: the estimated FMIs range from 0.10 for `ne` to 0.78 for `age`. We therefore perform the unrestricted FMI test with the `ufmitest` option, as follows.

```
. mi test age nfeatures ne, ufmitest
 ( 1)  age = 0
 ( 2)  nfeatures = 0
 ( 3)  ne = 0
       F(  3,  41.8) =    0.28
            Prob > F =    0.8376
```

The unrestricted FMI test results in a larger p-value than the equal FMI test. However, both tests indicate that these three variables are not strong predictors of house resale prices, at least according to the model we fit here.

6 Conclusion

We proposed improvements to the existing degrees of freedom for multivariate tests for multiply imputed data. In particular, we proposed a small-sample adjustment to the degrees of freedom of the unrestricted FMI test, and we refined the small-sample adjustment for the equal FMI test when $M(p-1) \leq 4$. Empirical evaluations of these adjustments, while admittedly limited in scope as all such evaluations must be, demonstrated that using tests based on the proposed small-sample adjustments can improve performance over using tests based on the large-sample analogues. Simulations also showed that the proposed testing procedures become more conservative as FMIs increase or start varying substantially. The deviations from nominal significance result because the adjustments are, as noted previously, unavoidably ad hoc in nature. For example, the derivation of the proposed degrees of freedom presumes that FMIs are approximately equal even though this assumption is not used in the test statistic. Additionally, estimates of the within-imputation and between-imputations variance components can be unreliable for small sample sizes and modest numbers of imputations.

We also considered using the denominator degrees of freedom suggested by Reiter (2007) for the unrestricted FMI test. This led to a slightly more conservative test than the one using the degrees of freedom from Barnard and Rubin (1999).

Other simulations not shown here suggested that the small-sample unrestricted FMI test performs better than the small-sample equal FMI test when the FMIs vary noticeably, and that the small-sample equal FMI test performs better when the FMIs are approximately equal. Further research is needed to compare the properties of these two tests in a wide range of plausible scenarios.

7 References

Barnard, J., and D. B. Rubin. 1999. Small-sample degrees of freedom with multiple imputation. *Biometrika* 86: 948–955.

Carlin, J. B., J. C. Galati, and P. Royston. 2008. A new framework for managing and analyzing multiply imputed data in Stata. *Stata Journal* 8: 49–67.

Li, K.-H., T. E. Raghunathan, and D. B. Rubin. 1991. Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association* 86: 1065–1073.

Little, R. J. A., and D. B. Rubin. 2002. *Statistical Analysis with Missing Data*. 2nd ed. Hoboken, NJ: Wiley.

Reiter, J. P. 2007. Small-sample degrees of freedom for multi-component significance tests with multiple imputation for missing data. *Biometrika* 94: 502–508.

Reiter, J. P., and T. E. Raghunathan. 2007. The multiple adaptations of multiple imputation. *Journal of the American Statistical Association* 102: 1462–1471.

Royston, P. 2004. Multiple imputation of missing values. *Stata Journal* 4: 227–241.

———. 2005a. Multiple imputation of missing values: Update. *Stata Journal* 5: 188–201.

———. 2005b. Multiple imputation of missing values: Update of ice. *Stata Journal* 5: 527–536.

———. 2007. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. *Stata Journal* 7: 445–464.

Royston, P., J. B. Carlin, and I. R. White. 2009. Multiple imputation of missing values: New features for mim. *Stata Journal* 9: 252–264.

Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Schafer, J. L. 1997. *Analysis of Incomplete Multivariate Data*. Boca Raton, FL: Chapman & Hall/CRC.

StataCorp. 2009. *Stata 11 Multiple-Imputation Reference Manual*. College Station, TX: Stata Press.

About the authors

Yulia V. Marchenko is a senior statistician at StataCorp.

Jerome P. Reiter is an associate professor in the Department of Statistical Science at Duke University in Durham, North Carolina. His research interests include methodology for missing data, methodology for protecting confidentiality in public use data, methodology for causal inference, and applied statistical modeling.