

Analysis of sample survey data often requires adjustments for missing values in the variables of interest. Standard adjustments based on item imputation or on propensity weighting factors rely on the availability of auxiliary variables for both responding and non-responding units. Their application can be challenging when the auxiliary variables are numerous and are themselves subject to incomplete-data problems. This paper shows how classification and regression trees and forests can overcome these difficulties and compares them with likelihood methods in terms of bias and mean squared error. The development centers on a component of income data from the U.S. Consumer Expenditure Survey, which has a relatively high rate of item missingness. Classification trees and forests are used to model the unit-level propensity for item missingness in the income component. Regression trees and forests are used to model the conditional mean of the income component. The methods are then used to estimate the mean of the income component, adjusted for item nonresponse. Thirteen methods for estimating a population mean are compared in simulation experiments. The results show that if the number of auxiliary variables with missing values is not small, or if they have substantial missingness rates, likelihood methods can be impracticable or inapplicable. Tree and forest methods are always applicable, are relatively fast, and have higher efficiency than likelihood methods under real-data situations with incomplete-data patterns similar to those in the abovementioned survey. Their efficiency loss under parametric conditions most favorable to likelihood methods is observed to be between 10-25%.