

We study identification of parametric and semiparametric models with missing covariate data. When covariate data are missing not at random, identification is not guaranteed even under fairly restrictive parametric assumptions, a fact that is illustrated with several examples. We propose a general approach to establish identification of parametric and semiparametric models when a covariate is missing not at random. Without auxiliary information about the missingness process, identification of parametric models is strongly dependent on model specification. However, in the presence of a fully observed shadow variable that is correlated with the missing covariate but otherwise independent of the missingness conditional on the covariate, identification is more broadly achievable, including in fairly large semiparametric models. Special consideration is given to the generalized linear models with the missingness process unrestricted. Under such a setting, the outcome model is identified for a number of familiar generalized linear models, and we provide counterexamples when identification fails. For estimation, we describe an inverse probability weighted estimator that incorporates the shadow variable to estimate the propensity score model, and we evaluate its performance via simulations. We further illustrate the shadow variable approach with a real data example about home price in China.