

Estimation Method	Model/Parameter	Inputs	Assumptions	Description
Anderson et al. (2019)	GMM	Multiple outcomes $\{y_{i\ell}\}_{\ell=1}^{L_i}$ linked to same individual x_i	Datasets are random samples conditional on the matching variables; each x_i has exactly one true link among its matched outcomes $\{y_{i\ell}\}_{\ell=1}^{L_i}$, and each $y_{i\ell}$ is equally likely to be the correct match	Rewrite the moments as a difference between the sum of the moments for all possible links and a conditional expectation of the moment function based on i 's identifying variables; estimation and variance estimation proceed as standard 2-step semiparametric GMM problem
Lahiri and Larsen (2005)	GLM	Multiple outcomes $\{y_{i\ell}\}_{\ell=1}^{L_i}$ linked to same x_i with probabilities $p_{i\ell}$	Each record in File A has exactly one true link in File B; distribution of matching variables is independent of the response variable	Apply OLS to a transformed regression model constructed using parameter estimates from the Fellegi-Sunter/EM algorithm method; estimate the variance using parametric bootstrap
Scheuren and Winkler (1997)	GLM	Linked dataset	Analysts are able to produce a reasonable model for the relationships between the noncommon quantitative terms(?), no misspecification	Iterate over the following steps: (i) fit regression to a configuration of matched data; (ii) drop outliers and fit regression using only highest weight matches; (iii) use these updated estimates to impute values for dropped observations
Nix and Qian (2015)	Population Mean	Multiple $\{x_{i\ell}\}_{\ell=1}^{L_i}$	True match is included among potential matches; multiple matches are equally likely to be correct	Construct bounds on parameters by estimating model with different configurations of matched data, then impute values for unmatched individuals to construct absolute bounds for the population statistic of interest
Bleakley and Ferrie (2016)	GLM/ATE	Multiple $\{x_{i\ell}\}_{\ell=1}^{L_i}$ linked to same treatment T_i	True match is included among potential matches; each of the $\{x_{i\ell}\}_{\ell=1}^{L_i}$ is equally likely to be the true match; treatment T_i is randomly assigned	For $\{x_{i\ell}\}_{\ell=1}^{L_i}$ linked to the same binary treatment T_i , impute $i\ell$'s treatment as $T_{i\ell} = T_i/L_i$; add a fixed effect for each surname, and use heteroskedasticity robust standard errors, clustered by $\{x_{i\ell}\}$ groups that share a common T_i
Hirukawa and Prokhorov (2018)	GLM	Linked dataset using NN/highest probability matching	The samples “jointly identify” the regression models	Bias correction for nearest-neighbor matched samples; could be used if we define an appropriate metric for matching variables and use a nearest neighbor matching rule (i.e. the Fellegi-Sunter optimal matching rule)
Goldstein et al. (2012)	GLM	Record pairs (x_i, y_j) and probability p_{ij} that (i, j) refer to the same entity	All variables follow a joint multivariate normal distribution	Construct modified match probabilities π_{ij} equal to p_{ij} times a likelihood term $f(y_j)$ based on an imputation model; accept as true links highest values of π_{ij} exceeding a threshold; finally, impute values of y for records that are not assigned a match
Anderson (2019)	GMM	Record pairs (x_i, y_j) and probability p_{ij} that (i, j) refer to the same entity	Goals: (i) allow matching variables to be correlated with variables of interest; (ii) allow for true link to not be included among possible matches	Ideas: Combine techniques such as PII with multiple matches; use equal probability weights (exogenously determined)