

## Microdata Disclosure Control by Resampling – Effects on Regression Results

By Sandra Gottschalk, Mannheim

JEL C13, C15, C81

Nonparametric resampling, Monte Carlo simulations, Mannheim Innovation Panel,  
linear and nonlinear regression.

---

### Summary

Nonparametric resampling is a method for generating synthetic microdata and is introduced as a procedure for microdata disclosure limitation. Theoretically, re-identification of individuals or firms is not possible with synthetic data. The resampling procedure creates datasets – the resample – which nearly have the same empirical cumulative distribution functions as the original survey data and thus permit econometricians to calculate meaningful regression results. The idea of nonparametric resampling, especially, is to draw from univariate or multivariate empirical distribution functions without having to estimate these explicitly. Until now, the resampling procedure shown here has only been applicable to variables with continuous distribution functions. Monte Carlo simulations and applications with data from the Mannheim Innovation Panel show that results of linear and nonlinear regression analyses can be reproduced quite precisely by nonparametric resamples. A univariate and a multivariate resampling version are examined. The univariate version as well as the multivariate version which is using the correlation structure of the original data as a scaling instrument turn out to be able to retain the coefficients of model estimations. Furthermore, multivariate resampling best reproduces regression results if all variables are anonymised.