ESTIMATING AN UNKNOWN FUNCTION BY LOCAL LINEAR REGRESSION WHEN THE ERRORS ARE CORRELATED

by

Jean Opsomer
Iowa State University

Abstract

Automated bandwidth selection methods for nonparametric regression break down in the presence of correlated errors. While this problem has been previously studied in the context of kernel regression, the results to date have only been applicable to univariate observations following an equidistant design. This article significantly generalizes these results by addressing the problem for local linear regression and considering both univariate and bivariate observations following a random design. In the bivariate case, we analyze both the general bivariate model and the additive model. In this more general setting, we show that when the errors are correlated, the asymptotically optimal bandwidth for local linear regression depends on the integrated covariance function. Nonparametric plug-in bandwidth estimators which take this effect into account are proposed for the univariate and bivariate models, and estimators of regression functionals, variance and integrated covariance function are developed. In particular, the estimate of the integrated covariance function is constructed by estimating the spectral density at 0 through binning of the data and periodogram smoothing of the residuals of a local linear regression. For the univariate and additive model, the resulting bandwidth estimators achieve the same rate of convergence as when the errors are independent. The general bivariate model is shown to suffer from the "curse of dimensionality," making plug-in estimation of the integrated covariance function difficult. Simulation experiments and real-life examples are used to demonstrate the proposed bandwidth selection algorithm.