SEMI-PARAMETRIC ESTIMATION OF WEIGHTED AVERAGE DERIVATIVES OF MODAL REGRESSION

HIROFUMI OHTA

Graduate School of Economics, Department of Statistics, University of Tokyo

ABSTRACT. In this talk, I will discuss a single index model for modal regression function, and propose a novel estimator of its weighted average derivatives. I construct the estimator based on sample-splitting/cross-fitting techniques. (cf. [1]). In general semi-parametric estimation, these approaches are expected to reduce "own-estimating bias" from estimating nuisance functions and parameters of interest based on full sample. Under the sample-splitting scheme, I derive some asymptotic properties of the estimator and a valid inference method.

Keywords: single index model, conditional mode, average derivatives, U-statistics, cross-fitting.

This paper studies a regression model which estimates the conditional mode of an outcome Y given regressors X, called *modal regression*. I consider a single index model of the form

$$\begin{aligned} \mathrm{Mode}(Y|X) &= m_0(X), \\ m_0(X) &= G(X^\top \beta), \end{aligned}$$

where X is a d-dimensional vector of continuous regressors, Y is a continuous scalar outcome variable, m_0 is an unknown regression function, and G is an unknown link function. Let $f(y|x)$ be a conditional density function of Y given X, and assume the global mode of $f(y|x)$, denoted by $\mathrm{Mode}(Y|X) = \arg\max_y f(y|x)$ is unique. The focus of this paper is on semi-parametric estimation of the index coefficient $\beta \in \mathbb{R}^d$, when m_0, G and $f(y|x)$ are unspecified.

Compared to mean regression, the modal regression has particularly useful features when the data distribution is highly skewed or has fat tails. In such situations, the mean regression, which estimates the conditional mean of the distribution, fails to capture the major trend underlying the data. This is because the conditional mean is not necessarily the point where the data points distribute densely. Then, the conditional mode is a convenient alternative in this situation as it can capture the majority of the data.

Recently, several modal regression models and their estimation methods, including fully non-parametric model and linear model, have been developed. On the other hand, there is no study on estimating the single index model for the modal regression. The single index model is known as a semi-parametric model, which restricts the regression function to be the transformation G of the linear term $X^\top \beta$. Define $\varepsilon := Y - G(X^\top \beta)$ as a disturbance, then, $\mathrm{Mode}(\varepsilon|X) = 0$ is the restriction on the noise. The single index model is semi-parametric because G is unknown function and the conditional density of ε given X is not specified. Thus, the lack of information of G makes the estimation of β difficult and nontrivial. I construct the estimator based on [2] and the sample-splitting allows to use asymptotic analysis for U-statistics.

REFERENCES

[1] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 2017.

[2] J. L. Powell, J. H. Stock, and T. M. Stoker. Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, pages 1403–1430, 1989.