A.IX. Estimation of Conditional Mean

Olivier Scaillet

University of Geneva and Swiss Finance Institute

Outline

Introduction

Recall the linear regression

$$\begin{aligned} Y_t &= X_t\beta + \varepsilon_t \\ &= \sum_{k=1}^{K} X_{k,t}\beta_k + \varepsilon_t \end{aligned}$$

where

$$\begin{aligned} \varepsilon_t &= \text{noise with mean zero} \\ &= \text{innovation or error term,} \end{aligned}$$

$$X_t\beta = E\left[Y_t \,|X_t\right]$$

$$\begin{aligned} &= \text{conditional mean of } Y_t \\ &= \text{linear function of } X_t, \end{aligned}$$

β = parameter (estimation by OLS).

Extension: Nonparametric Regression

Y_t now depends *nonlinearly* on the explanatory variables X_t

$$Y_t = m(X_t) + \varepsilon_t,$$

$$\begin{aligned} \varepsilon_t &= \text{noise with mean zero} \\ &= \text{innovation or error term,} \end{aligned}$$

$$\begin{aligned} m(X_t) &= E\left[Y_t \,|X_t\right] \\ &= \text{conditional mean of } Y_t \\ &= \text{unspecified nonlinear function of } X_t. \end{aligned}$$

We may conduct estimation of the function m by a kernel method.

Univariate Case

We consider a single variable: $X_t = X_{1,t}$

$$Y_t = m(X_{1,t}) + \varepsilon_t.$$

The estimation of the function m at point x_1:

$$m(x_1) = E\left[Y_t \,|X_{1,t} = x_1\right]$$

by a kernel method is given by

$$\hat{m}(x_1) = \frac{\frac{1}{Th}\sum_{t=1}^{T} y_t K\left(\frac{x_{1,t}-x_1}{h}\right)}{\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{x_{1,t}-x_1}{h}\right)}.$$

It is very easy to implement since it only involves empirical averages computed from a grid of distinct points.

Multivariate Case

We consider K variables: $X_t = (X_{1,t}, ..., X_{K,t})'$

$$\begin{aligned} Y_t &= m(X_t) + \varepsilon_t \\ &= m(X_{1,t}, ..., X_{K,t}) + \varepsilon_t \end{aligned}$$

The estimation of function m at point $x = (x_1, ..., x_K)'$:

$$\begin{aligned} m(x) &= E\left[Y_t \,|X_t = x\right] \\ &= E\left[Y_t \,|X_{1,t} = x_1, ..., X_{K,t} = x_K\right] \end{aligned}$$

by a kernel method is given by

$$\hat{m}(x) = \frac{\frac{1}{Th^K} \sum_{t=1}^{T} y_t \prod_{k=1}^{K} K\left(\frac{x_{k,t} - x_k}{h}\right)}{\frac{1}{Th^K} \sum_{t=1}^{T} \prod_{k=1}^{K} K\left(\frac{x_{k,t} - x_k}{h}\right)}$$

For dimension $K > 5$, we get poor properties of the estimation procedure (curse of dimensionality).

Further Extensions

The method remains valid for estimation of conditional expectations of *functions* $g(Y_t)$ of Y_t, namely $E\left[g(Y_t)\,|X_t\right]$.

The kernel estimator is simply

$$\frac{\frac{1}{Th^K}\sum_{t=1}^{T} g(y_t)\prod_{k=1}^{K} K\left(\frac{x_{k,t}-x_k}{h}\right)}{\frac{1}{Th^K}\sum_{t=1}^{T}\prod_{k=1}^{K} K\left(\frac{x_{k,t}-x_k}{h}\right)}$$

Example

Estimation of a conditional variance

$$V(x) = V\left[Y_t \,|X_t = x\right] = E\left[(Y_t)^2 \,|X_t = x\right] - \left(E\left[Y_t \,|X_t = x\right]\right)^2$$

here $g(y) = y^2$:

$$\hat{V}(x) = \frac{\frac{1}{Th^K}\sum_{t=1}^{T}(y_t)^2 \prod_{k=1}^{K} K\left(\frac{x_{k,t}-x_k}{h}\right)}{\frac{1}{Th^K}\sum_{t=1}^{T} \prod_{k=1}^{K} K\left(\frac{x_{k,t}-x_k}{h}\right)} - (\hat{m}(x))^2 \,.$$