Methods for estimating a time series of densities

Thilaksha Tharanganie & Rob J. Hyndman
Department of Econometrics and Business Statistics, Monash University

Abstract

We consider the problem of estimating a time series of density functions. A dataset comprising many observations is recorded at each time period, and the associated probability density function is to be estimated for each time period. It is assumed that the densities change slowly over time and that neighbouring densities are similar but not identical.

We consider several methods for estimating a time series of densities: (1) A logspline approach applied to each dataset separately, where each estimated density has common knots but different coefficients; (2) A conditional logspline approach with a full splines framework to account for the time variation; (3) A conditional logspline approach applied to all data simultaneously with common knots and kernel weights to account for the time variation; (4) A conditional kernel approach with kernel weights to account for the time variation.

In the full splines framework, we allow the degree of the splines to change from 2 to 10, and select the optimal value based on Bayesian information criterion (BIC). For the conditional kernel estimator, we compute two new bandwidth selection approaches that explicitly account for the discrete nature of the time conditioning.

We apply our methods to two simulated examples (unimodal and bimodal) and four real examples. The four datasets comprise UK and Australian income and age data over many years with thousands of observations per year. Probability integral transforms and proper scoring rules such as logarithmic score, quadratic score and spherical score are used to evaluate the density estimates.

Keywords

logspline, conditional kernel, density estimation, density evaluation, probability scoring.