國立中山大學應用數學系
學術演講

講　者：Prof. Wolfgang Karl Härdle

(Humboldt-Universit¨at zu Berlin, Berlin, Germany)

講　題：Kernels, Pricing, Attention!

時　間：2026/11/06（Friday）14:10pm~15:00pm

地　點：理 SC4011 教室

茶　會：13:30

Abstract

Consider three seemingly distinct ideas: kernel estimation, financial option pricing, and attention mechanisms. Economic data should speak to model builders through flexible methods. A classical Engel curve example illustrates how nonparametric kernel techniques uncover nonlinear relationships between income and consumption. Uniform confidence bands provide inferential tools for distinguishing genuine economic structure from sampling variation. Kernel density estimators further reveal how entire income distributions evolve across time. Asset pricing and the stochastic discount factor underlying the valuation of contingent claims also involve „kernels“, as well: Under no arbitrage, the pricing kernel links physical probabilities to risk-neutral probabilities extracted from option prices. Its empirical estimation therefore requires both historical return distributions and option-implied risk-neutral densities. For a risk-averse representative investor, conventional theory predicts a pricing kernel decreasing with market returns. Empirical estimates for S&P 500 and Bitcoin options, however, frequently display the non-monotonic pricing kernel puzzle that may reflect investor heterogeneity, behavioural effects, market imperfections, or estimation uncertainty. Flexible B-splines offer one possible remedy, although their conclusions depend on the chosen knots and moment restrictions. More general smoothing operators and random forests reveal that many modern learning procedures implicitly construct adaptive kernels. This observation provides a conceptual bridge from flexible econometrics to contemporary machine learning, e.g. attention mechanism underlying transformer models. Attention assigns data-dependent weights to values according to the

similarity between queries and keys. Single-head attention can consequently be interpreted as a learned kernel-smoothing estimator in a transformed geometry. Multi-head attention extends this principle by combining multiple geometries, scales, and smoothing operators. Transformers thus replace manually selected kernels and bandwidths with kernels learned jointly from large-scale data. The resulting synthesis shows that kernels, pricing, and attention share one central principle: adaptive weighting of relevant information.

敬 請 公 告！ 歡 迎 參 加！

應用數學系：http://math.nsysu.edu.tw

校園地圖：http://math.nsysu.edu.tw/var/file/183/1183/img/779/nsysu_math_map.jpg

交通資訊：https://www.nsysu.edu.tw/p/412-1000-4132.php?Lang=zh-tw

用數學系 校園地圖 交通資訊