



香港中文大學統計學系

Department of Statistics

THE CHINESE UNIVERSITY OF HONG KONG

## DISTINGUISHED LECTURE

# Sample Splitting and Assessing Goodness-of-fit of Time Series



Professor **Richard A. DAVIS**

Howard Levene Professor of Statistics

Department of Statistics, Columbia University

President, Institute of Mathematical Statistics (2016)

Editor-in-Chief, Bernoulli Journal (2010-2012)

Fellow: American Statistical Association (2002)

Koopmans Econometric Theory Prize (1997) (with William Dunsmuir)

Fellow: Institute of Mathematical Statistics (1995)

Elected member: International Statistical Institute (1992)

Date: 17 April 2024 (Wednesday)

Time: 10:30 am — 11:30 am

Venue: LT5, Lady Shaw Building,  
The Chinese University of Hong Kong

### Abstract

A fundamental and often final step in time series modeling is to assess the quality of fit of a proposed model to the data. Since the underlying distribution of the innovations that generate a model is often not prescribed, goodness-of-fit tests typically take the form of testing the fitted residuals for serial independence.

However, these fitted residuals are inherently dependent since they are based on the same parameter estimates and thus standard tests of serial independence, such as those based on the autocorrelation function (ACF) or distance correlation function (ADCF) of the fitted residuals need to be adjusted. The sample splitting procedure in Pfister et al. (2018) is one such fix for the case of models for independent data, but fails to work in the dependent setting. In this paper sample splitting is leveraged in the time series setting to perform tests of serial dependence of fitted residuals using the ACF and ADCF. Here the first  $f_n$  of the data points are used to estimate the parameters of the model and then using these parameter estimates, the last  $l_n$  of the data points are used to compute the estimated residuals. Tests for serial independence are then based on these  $l_n$  residuals. As long as the overlap between the  $f_n$  and  $l_n$  data splits is asymptotically  $1/2$ , the ACF and ADCF tests of serial independence tests often have the same limit distributions as though the underlying residuals are indeed iid. In particular if the first half of the data is used to estimate the parameters and the estimated residuals are computed for the entire data set based on these parameter estimates, then the ACF and ADCF can have the same limit distributions as though the residuals were iid. This procedure ameliorates the need for adjustment in the construction of confidence bounds for both the ACF and ADCF in goodness-of-fit testing.

(This is joint work with Leon Fernandes.)

★★★★★ All are welcome ★★★★★

For enquiries please contact Miss Esther TAM (Tel: 3943 7931)  
<https://www.sta.cuhk.edu.hk/news-and-events/>

