The Effect of Spectral Line Components on Coefficients in Time Series Regression Models

Jordan Kokocinski [1], Wesley S. Burr [2], Glen Takahara [1]
[1] Department of Mathematics & Statistics, Queen's University
99 University Ave, Kingston, Canada
11jdk13@queensu.ca; takahara@queensu.ca
[2] Department of Mathematics, Trent University
1600 West Bank Drive, Peterborough, ON
wesleyburr@trentu.ca

Extended Abstract

Time series regression models are useful for determining association between a response quantity and any number of explanatory variables when both response and explanatory variables vary over time. In this work, we consider the presence of deterministic periodic components and their effect on the covariance between fitted regression coefficients which are the result of two regressions on different, possibly overlapping, time segments of common response and predictor series.

Two covariance estimators are compared: one based on the regular Bartlett cross-covariance, and the other based on a cross-covariance function obtained by taking the inverse Fourier transform of a multitaper [1] cross-spectral estimate. Attention is constrained here to the framework of the simple linear regression model, however the extension to more general linear models is straightforward. For illustration, we focus on models with only a slope coefficient, since we assume that the response and predictor time series have been zero-meaned. To assess the performance of these covariance estimators, simulations of bivariate autoregressive processes are generated and used to compute covariance estimates which can then be compared to the theoretical covariances, since the theoretical cross-covariance function (CCVF) of a bivariate AR process has a known form.

Differences in the performance of these estimators are observed both when the data involved is (i) purely random and (ii) random with deterministic sinusoids common to both response and predictor. In both (i) and (ii), both estimators tend to perform similarly, but the Bartlett CCVF estimate consistently shows a pattern of slow decay as the lag increases, as compared to the theoretical CCVF and the estimate based on the multitaper method (MTM); this agrees with the behaviour of the autocovariance function (ACVF) estimate, as discussed in [2]. Both estimators appear to be badly biased in (ii), thus we seek to remove the common deterministic sinusoids (also called *line components*). Thomson's harmonic *F*-test [1] is used for the detection of these line components, and ones common to both response and predictor are reconstructed and then subtracted in the time domain, where we take "common" to mean that two line components' detected frequencies are within a specified threshold, chosen to be the MTM spectral bandwidth. Both types of covariance estimates computed using this new closer-to-stationary series show less bias when line components are removed than when they were left present, but the MTM-based estimator consistently has a smaller variance. For fitted regression coefficients from each time segment, we estimate a joint distribution and include estimated covariance information in order to obtain more informative approximate confidence intervals for each regression coefficient.

The methods developed here are applied to an hourly electricity demand and market price dataset retrieved from the Independent Electricity System Operator (IESO) in Ontario, Canada. Previous studies, such as [3] and [4], make use of regression models for short- and long-term load forecasting, and explicit use of the kind of covariance information examined in this talk may be used to improve the accuracy of confidence bands for trend estimation and improve prediction accuracy in future developed models.

References

[1] D. J. Thomson, "Spectrum estimation and harmonic analysis," in *Proceedings of the IEEE*, 1982, vol. 70, pp. 1055–1096.

[2] D. J. Thomson, "Some comments on multitaper estimates of autocorrelation," in *IEEE Statistical Signal Processing Workshop (SSP)*, 2012, pp. 656–659.

[3] H. Zareipour, K. Bhattacharya, and C. A. Canizares, "Forecasting the hourly Ontario energy price by multivariate adaptive regression splines," in *IEEE Power Engineering Society General Meeting*, 2006, doi: 10.1109/PES.2006.1709474.

[4] F. Nogales, J. Contreras, A. Conejo, and R. Espinola, "Forecasting next-day electricity prices by time series models," in *IEEE Transactions on Power Systems*, 2002, vol. 17, no. 2, pp. 342–348, doi: 10.1109/TPWRS.2002.1007902.