Tucker McElroy

Treatment of Time Series Outliers via Maximum Entropy

Tucker McElroy[1]

[1] U.S. Census Bureau

Abstract:
The Covid-19 epidemic has greatly impacted economic time series, generating sustained strings of extreme values in the data. We extend the maximum entropy framework – for identifying and adjusting for additive outliers – to a generalized class of outliers, including level shifts, temporary changes, and seasonal outliers. These are described as a particular type of stochastic process that is latent, or unobserved, such that its removal increases the time series entropy, i.e., makes the data more closely resemble a Gaussian process. Methodology for simultaneously modeling missing values and such generalized outliers is developed, and the issues of identification, testing, and shrinkage are discussed. Shrinkage ensures a smoother modification of a given time series, making it less sensitive to data revisions. We illustrate the expanded methodology on initial weekly claims data. The application of these methods to economic data impacted by the Covid-19 epidemic is developed.

Keywords:
Covid-19; Extreme-value Adjustment; Level Shifts; Shrinkage

1. Introduction:

The onset of the COVID-19 epidemic has inflicted massive effects on broad portions of the world economy. Statistical agencies are tasked with publishing economic data along with their seasonal adjustments, which are important in view of public consumption of data products. One series of particular concern is the National Unemployment Insurance Weekly Claims Data. (Hereafter, denoted "Claims.") In March 2020 the claims surged to levels never seen before.

The objective of this research is to generate suitable models for time series data that mingles regular and crisis epochs, and also to produce signal extraction results that have a suitable interpretation. Towards that end, we propose to extend the maximum entropy extreme-value adjustment framework of McElroy and Penny (2019) to using shrinkage for a much broader scope of intervention effects. The entropy framework, as opposed to regression, treats the outliers as stochastic.

2. Methodology:

The data process is written as the sum of two latent processes: a regular Gaussian process and an extreme process, which itself is composed of all the types of outliers present in the data. Each is described as a certain kind of fixed regressor multiplied by a heavy-tailed random variable coefficient; these regressors can correspond to additive outliers, level shifts, temporary changes, seasonal outliers, or more general effects. Supposing there are r of these outlier effects, and the sample has length n, we obtain an n x r matrix of regression effects, called X. So-called regular values are defined as a transformation of the sample via a n-r x n dimensional matrix L. This can be constructed in such a way that L X = 0.

As a result, L multiplying the data vector annihilates the extremes, and therefore represents a transformation to a regular process. A time series model can be formulated for the regular process, and using L we can compute the Gaussian likelihood, and fit this model. We can also compute the expected value of the data vector conditional on having applied L, and this gives an extreme-value adjustment. This methodology is a direct generalization of McElroy and Penny (2019), which considers the special case that X corresponds to additive outliers (so the columns are unit vectors).

Computing the conditional variance of the extreme-value adjustment, we can modify this operation to scale the amount of attenuation, accomplishing a partial shrinkage of extremes. This can be done in conjunction with test statistics that assess the overall improvement to time series entropy as a function of the degree of shrinkage.

3. Result:

Claims data is modelled with various types of extremes, including additive outliers and level shifts. The method is used to produce extreme-value adjusted data with shrinkage. Seasonal adjustment filters appropriate for weekly data are applied after to the regularized data; the extremes can be added back afterwards, as they correspond to trend and irregular effects.

4. Discussion and Conclusion:

Current modelling approaches to extreme value adjustment use regression, which posits extremes as deterministic phenomena. In contrast, the entropy approach views extremes as being generated by heavy-tailed random variables. Future work will examine how to incorporate missing values with the extreme-value framework, and make generalizations to multivariate time series.

References:

1. McElroy, T. and Penny, R. (2019). Maximum entropy extreme-value seasonal adjustment. Australian New Zealand Journal of Statistics 61(2), 152-174.

NOTE: THE MAXIMUM NUMBER OF PAGES FOR THE PAPER IS SIX PAGES