Evolving Multivariate Regression

Problem 02-004, *by* Peter Thejll (Solar-Terrestrial Physics Division, Danish Meteorological Institute).

Consider a time series $T(t)$ produced in the following way:

$$T(t) = aA(t) + bB(t) + n(t),$$

where $A(t), B(t)$ are some time series, a, b represent constant *coupling factors*, and $n(t)$ is a noise series of some sort ("white" (no serial correlation) or "red" (serially correlated)). Somehow series $A(t), B(t)$ are known and it is desired to find the values of a, b in order to perform some *attribution study*, such as in climate research, where $T(t)$ may be some observed temperature series and $A(t), B(t)$ may represent solar irradiance and forcing due to greenhouse gases. It is noted that the statistical natures of $A(t), B(t)$ are quite general. Specifically, we do not guarantee that they are stationary and they may not be of the same nature at all. For example, $A(t)$ may be dominated by nonstationary white noise while $B(t)$ may be dominated by low frequencies. An *evolving multivariate regression* technique is applied wherein multivariate regression is performed on subsections of the series, for example in sliding windows that progress one time step at a time but are many time steps wide, thus providing, as it were, series of values for a, b.

It is seen that in studies of examples where $T(t)$ series are directly generated via the above equation, evolving multivariate regression does not recover constant coefficients a and b. The recovered values appear to be biased in proportion to the partial correlation coefficients between $T(t)$ and $A(t)$ and $B(t)$. How should the analysis be performed to retrieve values of a, b that are not biased, if that indeed is the problem above?

Status. This problem is open.