Smoothing and interpolation of irregularly sampled time series using a continuous time state space model

Granville Tunnicliffe Wilson*

July 29, 2019

Keywords:

1 Introduction: the model and an example

The aim of this document is two-fold: firstly to document the MATLAB function `VCZstateFilterSmooth.m` for state-space filtering and smoothing of an irregularly sampled vector continuous time series represented by a VCZAR model, and secondly to illustrate this by application to the Extinctions - Originations series. This application is held in the script `ExtOrigBookModelInterp.m` and includes interpolating the series at regularly spaced time points. A further script `VCZARsmoothtest.m` provides a test of the method by using the alternative direct calculation of the smoothed values from the model covariances.

2 The function `VCZstateFilterSmooth.m`

The usage and help lines of the function are

```
function [statfil,statsmo,trend,Vstatfil,Vstatsmo,cTranMat] =...
    VCZstateFilterSmooth(yRef,m,p,kappa,obsnoise,xreg,xregnum,...
    VCZARcoeff,VCZARSig,NoiseSig,xregcoeff,Strucoeff,parind,struc)
  Uses State Space formulation of the Continuous Vector ZAR model
  to estimate filtered and smoothed states including at unobserved time points
  these times are specified in yRef but with NaN for observations
  xreg values will also need to be given at interpolated points
  states are estimated in the time sequence given in yRef including simultaneity
  statfil and statsmo are respectively the filtered and smoothed states at these times
  trend is the correction applied to the series value of current observation
  Vstatfil and Vstatsmo are respectively their variance matrices
  observations noise is appended to the state vector
  cTranMat is the continuous time transition matrix for CZAR states
  useful for transforming state estimates to decoupled eigencomponents
```

The arguments `m, p, kappa, obsnoise, xreg, parind, struc` of the function are used as in the arguments of `VCZARestimate.m` to specify the estimated VCZAR model, and

*Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.

are described in the documentation *VCZARestimation.pdf* of that function. The arguments `VCZARcoeff, VCZARSig, NoiseSig, xregcoeff, Strucoeff` are the estimated model parameters returned by that function as components of the cell structure `results`, and are described in the same document.

The argument `yRef` is also returned as a component of the `results` of estimation. It is assembled in the function `VCZARestimate.m` from the time series data set `y` supplied as an argument to that function. We reproduce here its description as given in the documention.

`yRef` is an array with 4 columns of length equal to the total number of time series values in the input structure `y`. Each time series value `y{i}[t]`is listed in column 4 in the order of its time `t` which is listed in column 3 of the same row. Column 1 lists the series number `i` and column 2 lists the index (row number) of this value in its original series.

We use `yRef` as the argument of `VCZstateFilterSmooth.m` rather than the original series structure `y` in order to allow application of this function to the interpolation of series values at time points extra to those used in the original series. This is done by adding corresponding extra rows to `yRef`, as described in the documentation below of the applications.

The remaining argument `xregnum` is a vector listing the number of regression components for the respective time series, and is simply constructed from the argument `xreg`. Note that if the series are to be interpolated at extra time points, any regressors must also have extra values supplied in `xreg` at the times appropriate for the respective series.

The quantities `statfil, statsmo` returned by the function are respectively the filtered and smoothed estimated states of the the model at the times listed in column 3 of `yRef`, and `Vstatfil, Vstatsmo` are their respective variance matrices. Thus `statfil(i,t)` is the filtered state `i` at time `t` and `Vstatfil(i,j,t)` is the covariance between filtered states `i` and `j` at time `t`. Similarly for the smoothed estimates.

The returned quantity `trend` is the vector of trend components of the respective series, constructed from the supplied regression vectors in `xreg` and the model regression coefficients in `xregcoeff`. Estimates of the filtered and smoothed series values can be derived from this vector together with the estimated states.

The returned quantity `cTranMat` is the continuous time transition matrix of the VCZAR model.

The function uses the method of estimating smoothed and filtered states described in the document titled *State space filtering and smoothing using square root methods* in the file `SSFilterSmooth.pdf`. This is necessarily modified to allow for the time varying nature of the integrated transition equation due to variation of the discrete time steps between observations. However, the modification involves little more than notational indexing with no change of algorithmic procedure, which uses the state space formulation of the VCZAR model found in (7.45) of the book.

3 Application to the Extinctions - Originations series

The script `ExtOrigBookModelInterp.m` loads the saved VCZAR model and the time series to which this model is fitted. Extra time points are then specified, giving 200 equally spaced intervals between the first and last observation times of these series. The structure `yRef`,

which has been loaded as a component of the save structure of the fitted model, is then extended and redefined to represent the same time series but with missing values at the extra time points. The trend regressors are similarly extended but with the extra values set appropriate to the extra time points.

The function `VCZstateFilterSmooth.m` is then used to generate the filtered and smoothed states.. The filtered and smoothed values of the two series are extracted as the first and second states respectively. They are plotted with their 2 standard error limits against the time before present.

It should be noted that each filtered series term is an estimate using all observed series values that are prior in the sequence in which they are held in `yRef`. If several series are observed at the same actual time point, some of those will be used to estimate the filtered values of others at the same time, but subsequent in the ordering in `yRef`. This ordering depends on the order of the component series of the argument `y` of `VCZARestimate.m`. In our example the two series have common time points so the filtered state values of the second series of Originations will be estimated using the observed value of the Extinctions series at the same time. Of course, the smoothed state values of both series are estimated from the same set of all observed series values.

The script `VCZARsmoothtest.m` is similar but only generates the smoothed and filtered states at the time points of the observed series. But in addition it generates the smoothed series values based on the calculated model covariances between all observations. using the general procedure described in the section 2.13 of the book. The covariances are readily calculated using the function `VCZARCov.m` described in the document *VCZARutilities.pdf*.

The smoothed values using both methods are plotted together with a small offset to allow comparison, and the total absolute differences between the results from the two methods are shown to be negligible.