

Fitting a CZAR model to the Extinction - Origination series

Granville Tunnicliffe Wilson*

June 17, 2019

Abstract

Keywords: Irregularly sampled time series, bivariate autoregressive model, continuous time impulse response estimation

1 The time series data and model

The aim of this document is to show how MATLAB functions can be applied to carry out the modeling of the irregularly sampled series of extinction and origination rates described in Chapter 8, and to derive the model properties. The Figure below is a reproduction of part of Figure 8.2 of the book and shows the series after they have been transformed logarithmically. These are stored in the files `lgeps.txt` and `lgops.txt` which are input by the scripts described below.

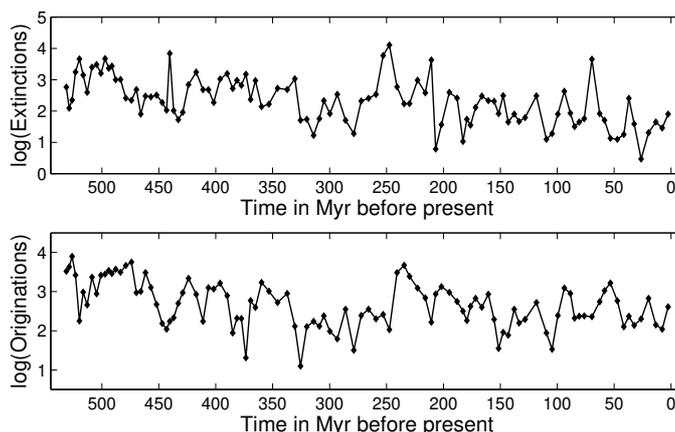


Figure 1: Irregularly observed time series of transformed extinction and origination rates of species with sampling time points set as the centers of stratigraphic intervals.

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

The examples we give here differ slightly in some aspects from those in the book, such as only using a single value of the time constant $\kappa = 0.3$ when selecting the model order and using the natural and structural model form up to the final stage of model estimation. Three MATLAB scripts are described in this section:

`ExtOrigOrdersAIC.m`, `ExtOrigOrdersDEV.m`, `ExtOrigBookModelFit.m`.

These develop the final model and enable anyone to successfully investigate the effect of using different options and to apply the methods to different series.

1.1 Order selection

The first step is to use the script `ExtOrigOrdersAIC.m` to investigate and select the model order. This requires an initial value of the rate parameter κ to be set. The value chosen is $\kappa = 0.3$, following consideration given in the book. Models up to order 4 are fitted and an order selected using the AIC. In the book this is repeated for the values of $\kappa = 0.25$ and 1.0 with very similar results.

The series are mean corrected before estimation to gain a slight improvement in numerical conditioning, but constant terms are still estimated, so that the MLEs of the series mean values can be recovered. Trends are also fitted to both series.

The natural form of model is specified, as its parametrization seems to lead to some improvement in convergence. Also a structural error model is specified, which replaces the estimation of the covariance between the errors. The constraints for that option are set for dependence of the disturbance in the origination series on that of the extinctions series. The default (with no constraints set) would be the reverse of this.

Whichever direction is set for this dependence by the choice of the constraint or the default, the models are equivalent in theory. However, the former might be preferred as corresponding to the expected direction of dependence between the series and gives a very slightly better fit for the model of order 4 only - the fit is identical for the lower orders. In fact, for the model of order 4, the noise disturbance variance is close to zero for the originations series, with the constraint as set (and for the extinctions series with the default constraint). A zero variance would still give a valid model, because observation noise is included. It simply implies dependence of one unobserved component upon the other, much as the predicted response depends upon the explanatory variable in standard regression. But a numerical difficulty in the iterative estimation does arise with a variance parameter close to its boundary value of zero. This explains the slight difference in results for order 4, depending on the specified constraints. For models of order 1 and 2 this difficulty does not arise.

Figure 2 shows the AIC derived in this way. Order 2 is clearly selected

In the script that produces these results it will be seen that the model for order 1 is first fitted without observation noise. The parameter estimates from this are then used to provide initial parameters for the model for order 1 with observation noise. This overcomes numerical difficulties due to the non-linearity of observation noise models when initial parameters are used which are not close to those of a good fitting model. Furthermore, at higher orders, initial parameter value are set from the model with the immediately lower order. This

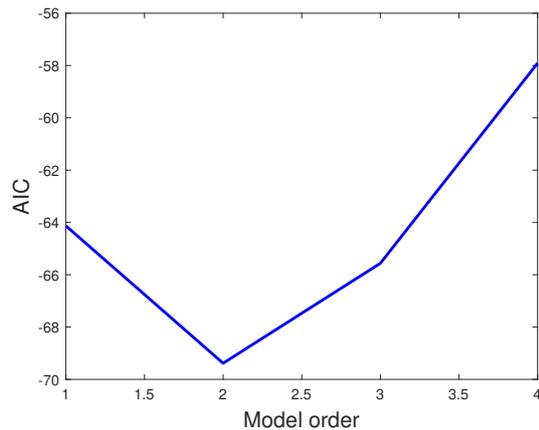


Figure 2: The AIC up to order 4 of a CZAR model for the Extinctions and Originations series.

is a pragmatic procedure based on experience, but models including observation noise are substantially non-linear and may have more than one local minimum to their deviance. It may be wise to check the fitted models by using default starting values, or ones devised in some other way.

The results of running this script may be saved in the file `ExtOrigOrdersModels.mat` by a command close to the end of the script. This file is not included with the scripts and functions available from the web site. The save command is commented out so that the user may experiment with the script before deciding to save the results.

1.2 Rate parameter estimation

The next step is to determine the best value of the rate parameter κ with the model order fixed at $p = 2$ determined by the minimum AIC. This is done by using the script `ExtOrigOrdersDEV.m` to fit the model for a range of values from $\kappa = 0.05$ to $\kappa = 0.5$, enclosing the initially chosen value $\kappa = 0$. The same specification of model is used, i.e. the natural model parametrization with structural noise for dependence of the disturbance in the origination series on that of the extinctions series. However, we now use the residual maximum likelihood criterion to estimate the model, for the reason explained in the book. For small values of κ this avoids the model degenerating to a low frequency discrete spectral component when trend regressors are fitted. This script reads the series text files and loads the file `ExtOrigOrdersModels.mat` to obtain initial parameter values for the models to be fitted.

The scripts produces the plot shown in Figure 3, from which is is seen that the value of $\kappa = 0.125$ minimizes the model deviance.

The models resulting from running this script may be saved in the file `ExtOrigKappasModels.mat` by a command near the end of the script. This is commented out for the same reason as before.

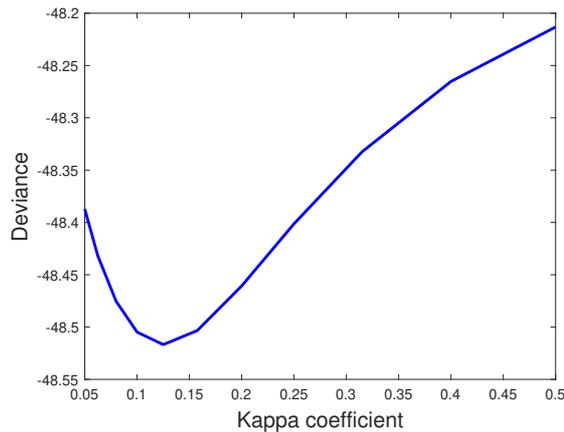


Figure 3: The deviance of the CZAR model for the Extinctions and Originations series over a range of κ values.

1.3 Final model development

The model with the specified order $p = 2$ and $\kappa = 0.125$ is then re-estimated using the script `ExtOrigBookModelFit.m`. This reads the time series text files and loads the file `ExtOrigKappasModels.mat` with the saved models.

However, the model now to be fitted is specified to be of the predictive form, which is more readily interpreted than the natural form. The saved model of order 2 with rate parameter $\kappa = 0.125$ is therefore transformed to the predictive form to provide starting values. The script prints the estimated parameters and their t values, and saves the fitted model. This has no structural noise parameters, corresponding to that presented in Table 8.3 of the book. The estimated autoregressive parameters differ slightly from those in the book because those were estimated following selection of κ with the predictive model form. Reducing the convergence criterion leads to very similar estimates.

There is a line, clearly indicted in the script, which may be commented out to allow estimation of the single structural noise parameter. This then has a t value of -0.16, explaining why this parameter was removed from the final model. This makes sense with the predictive model form as it suggests that the disturbance process is independent for the two series.

Also indicated in the script, is how two of the other parameters with low t values may be constrained to zero and the model refitted and saved.

A commented line in the script provides for saving the fitted model with the form of that in Table 8.3 in the file `ExtOrigBookModelWeb.mat`. This is used in the further script `ExtOrigBookModelProperties.m` to display the model properties, as described in the next section.

2 The fitted model properties

The properties of interest of VCZAR models are the lagged responses and covariances, and spectra. These are derived using functions described in the document *VCZARutilities.pdf*. For an estimated model, error limits on these quantities may be derived using the simulated delta method which is described on p.112 of the book. This is implemented for the estimated model of the Extinctions - Originations series by the script `ExtOrigBookModelProperties.m` which loads the model from the file `ExtOrigBookModelWeb.mat`. In the script the lag and frequency ranges and values for the model properties are specified. Also the option of whether to compute the impulse and step responses for the model with or without feedback specified. The number of simulations for the delta method is also specified, as is the scale factor for that method. The model parameters and t values are displayed and the model is then transformed; the disturbance structure from the structural to the variance matrix (standard) form, and the autoregressive coefficients from the predictive to the natural form. These forms are used by the functions which compute the model properties.

The simulated delta method is implemented in this script - no general function for this purpose has been written. This section is quite long because various model options must be allowed for.

A total of 11 plots are produced, 10 of which are shown in Figures 8.11 and 8.12 in the book, the exception being the warped spectrum of the extinctions series. However, the lagged impulse and step responses can be shown with or without the option of model feedback, giving 4 plots. Only 2 are shown in the book, so we show all 4 below.

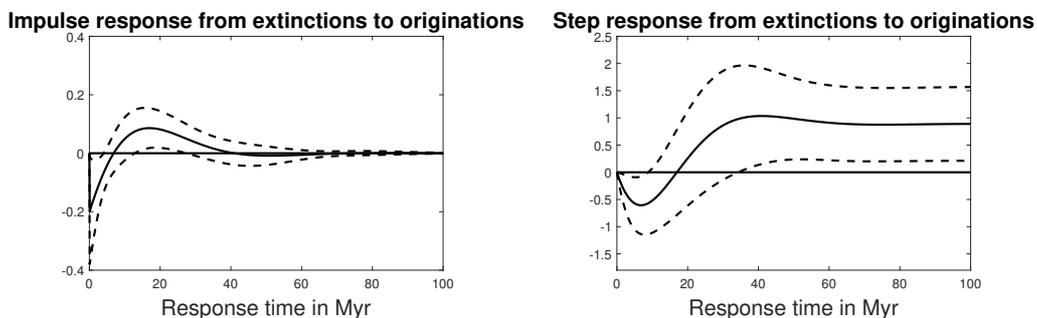


Figure 4: Open loop impulse and step responses from extinctions to originations

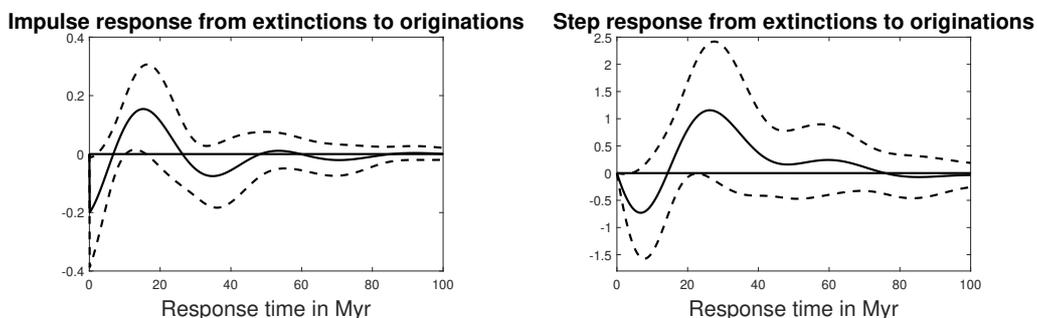


Figure 5: Closed loop impulse and step responses from extinctions to originations

We now point out that the plot of Figure 8.12 (f) in the book was mistakenly copied as

the closed loop step response from the extinctions to originations. The intention was that it should be the closed loop impulse response which is rather similar in shape, though not in magnitude, and also shows the zero-delay drop in originations which derives from the immediate reduction in the number of genera due to the extinctions. After a period, the originations increase to replace the extinctions.