Some notes on Bayesian time series analysis in psychology

Krone, Tanja

Chapter 7

Conclusion

The aim of this thesis was to study time series analysis in psychology. To narrow the focus in such an extensive field, the focus was lain on AR(1) models and the challenges of empirical data analysis. In this conclusive chapter, I will discuss the issues raised in the first chapter. After this, I will close this thesis with some recommendations for further studies.

7.1 Estimation of the autoregressive model

7.1.1 The difference between the estimators

The first question posed was: which estimator is preferred for AR(1) data? To answer this question, I compared several estimators for univariate, single individual data using a simulation study. I analyzed the data using six estimators: the r_1 estimator (Yule, 1927; Walker, 1931; Box & Jenkins, 1976), C-statistic (Young, 1941), ordinary least squares estimator (OLS), maximum likelihood estimator (MLE) and Bayesian Markov Chain Monte Carlo (MCMC) estimator using either a flat prior (B_f), or a symmetrized reference prior (B_{sr}).

I compared the estimators under several conditions, varying the length of the time series between 10 and 100 time points, and varying the true autocorrelation between -0.90 and 0.90. This showed that the distinction is not between Bayesian and frequentistic methods, but between iterative and closed form methods. The iterative methods, i.e., MLE, B_f and B_{sr}, showed better results with regard to the bias of the estimated autocorrelation. The closed form estimators showed a smaller variability for the estimators. However, the variability of the iterative estimators decreased strongly as the number of time points increased, decreasing the difference between iterative and closed form estimators for longer time series. Comparing the two Bayesian estimators showed that B_{sr} is to be preferred over

B_f. This led to the conclusion that the MLE and B_{sr} merit further study.

In two subsequent studies I compared the MLE and the B_{sr}. The first study was aimed at the robustness of the estimators, examining the effect of underspecification by using an AR(1) to analyze ARMA(1,1) data. Again, I varied several parameters, being the number of time points (25 or 50) and the size of both the AR(1) and MA(1) parameters (between −0.90 and 0.90). The differences between the two estimators were small and inconsistent over the conditions. In general, the B_{sr} showed a slightly larger variability and bias then the MLE. However, the difference is too small to draw strong conclusions as to the difference in robustness of the estimators.

The second study compared the MLE and B_{sr} for multiple individuals using a simulation study. Here, I compared a total of four estimators, analyzing the data with both estimation methods under a random coefficients model and a fixed effects model. As with single individual data, the multiple individual data was simulated under several conditions, varying the number of individuals (10 or 25), the length of the time series (also 10 or 25), and the mean (−0.60 to 0.60) and standard deviation (0.25 or 0.40) of the true autocorrelation distribution. The differences between the estimation methods were small to negligible. However, the differences between the random coefficients model and the fixed model were substantial. The random coefficients model shows better results for five out of the six measurements, making it the preferred model over the fixed model where possible, be it estimated with MLE or B_{sr}.

In conclusion, both the MLE and the B_{sr} show promising results in estimating the autocorrelation of AR(1) data. The difference between the two is small, where the B_{sr} shows a smaller bias and the MLE a smaller variability. When analyzing multiple individuals, it is advised to use the random coefficients model instead of the fixed coefficients model where possible.

7.1.2 The effect of data properties on the estimation of the AR(1) model

The second question I sought to answer is: what is the influence of data properties on the estimation of the AR(1) model? In the first two studies, the estimators were compared using data simulated under different conditions. This allows for a comparison of the bias, variability and power with regard to the length of the time series, the number of individuals in a dataset, and the mean and standard deviation of the true autocorrelation.

The most prominent effect is found for the length of the time series, especially for single individual data. As expected, the bias and the variability become smaller as the time series becomes longer. An often asked question is how long the time series must be to get a fairly trustworthy estimation. For an univariate, single individual dataset, a length of 50 time points is generally advised for an AR(1) model (Box & Jenkins, 1976). While the B_{sr} shows a small bias for time series as short as 25 time points for most autocorrelations, 40 and preferably 50 still is the advised number of time points. This is due to the variability, which is still rather high at 25 time points.

For the random model used on multilevel data sets, the length of the time series is about as important as the number of individuals included. As either the number of individuals or the number of time points increases, the bias and the variability decrease. The required sample size depends on the size of the standard deviation of the autocorrelation. In my studies I found that for samples with little variability in the autocorrelation, i.e., a standard deviation of 0.25 or less, the random model may produce results with an acceptable size of bias when either the number of time points or the number of individuals is at least 25, while the other one of the two is at least 10. For samples with more variability in the autocorrelation autocorrelation, i.e., a standard deviation of 0.40 or less, the random model may produce results with an acceptable size of bias when both the number of time points and the number of individuals is at least 25.

For the fixed model, the effect of the number of individuals is small, while the effect of the number of time points is similar to that for the single individual design. The results of the fixed model are comparable with those of the univariate, single individual data. Taking this into account, it is advisable to follow the guideline of 50 time points for the fixed model, regardless of the number of individuals included.

The mean and standard deviation across individuals for the true autocorrelation influence the estimation of the model, but are outside the influence of the researcher. For the single individual study we varied the true autocorrelation of the series between -0.90 and 0.90, with a fixed standard deviation of zero within the population. For the multilevel study we varied the autocorrelation between -0.60 and 0.60, with a standard deviation of either 0.25 or 0.40. Over both studies, the bias decreases and the variability increases when the autocorrelation becomes closer to zero. The power increases as the autocorrelation is further from zero. For the standard deviation of the autocorrelation, as compared in the multilevel study, the variability and the bias increase as the standard deviation becomes higher. The power decreases for a higher standard deviation of the autocorrelation.

Finally, a short robustness study was done on univariate, single individual data, to see the effect of underspecifying a model. The results show that the effect of underspecification increases with the size of the unmodeled parameter, and affects the bias, variability, rejection rate and power when compared to the results for the justly specified AR(1) model. The estimated variability of the autocorrelation in the data used for this study became smaller for a longer time series. The effect of a longer time series on the bias was small and dependent on the estimation method and the size of the autocorrelation. With regard to the data properties, detecting misspecification is more important than the length of the time series.

In conclusion, the properties of the data strongly influence the estimation of the autocorrelation. As the number of time points and individuals are the only factors which can be influenced by the researcher, we focus our answer on these. First, for univariate, single subject studies as for multilevel studies following the fixed model, at least 50 time points are advised. However, this may still provide results with a low power and high variability if the true autocorrelation is small. For univariate, multiple individual data sets which are analyzed using the random coefficients model, it is advisable to strive for 25 individuals with 25 time points each. Especially when the expected mean of the autocorrelation is close to zero, and the standard deviation of the autocorrelation in the population is expected to be high.

7.2 Empirical data analysis using the BDM

7.2.1 Practical issues in time series analysis

Studying empirical time series data poses certain challenges, such as properly including in the model trends, dynamics and external variables and dealing with missing data and non-normally distributed residuals. While the other possible models were unable to deal with some or all of these issues, the BDM can implement all elements needed to handle this in one model, as shown in Chapter 4. In the following section, each of the issues found is posed and then I show how this may be handled using the BDM.

As discussed in the introduction, the trend and the dynamics of the data must be studied in order to create a fitting model. The BDM has shown to be capable of both. The trend may be studied by inclusion of a slope, as shown in Chapter 4 pertaining to the panic attack data. The model dynamics, such as the autregression for the score and the white noise, may be implemented as a model element. The BDM also allows for a combination of the trend and the dynamics, which is shown in Chapter 4. Here we complement the linear model with an autregression element for the noise. It is important to keep in mind that adjusting the model may influence the identifiability of the model.

An important issue for empirical data is missing data, especially in time series analysis. Commonly, the amount of incomplete data substantially increases with larger numbers of scheduled time points. The BDM can handle both incidental missing data and drop outs. The incidental missing data is handled through application of the link function. However, large amounts of missing data introduce uncertainty in the parameter estimates. When a drop-out occurs, the analysis is stopped for that series.

The BDM can deal with non-normally distributed residuals. For an observed variable, the link function can be implemented in the same manner as used in generalized linear models (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989). It must be noted that using a link function may increase the amount of data needed to obtain a model that can be estimated.

Furthermore, the BDM allows for the inclusion of external variables in two ways. First, the external variable can be included as an active covariate. As an active covariate, it may influence the estimation of the model. This is done in Chapter 4, where the slope for each individual is dependent on the treatment and the presence of agoraphobia. Second, the external variable can be implemented post-hoc. Using a variable post-hoc does not influence the estimation, but allows to compare the results with regard to an external variable. Again, this was shown in Chapter 4. The external variable for both active and inactive covariates may be time dependent or independent.

Finally, the BDM allows for easy comparison of different models. Often a researcher has a general idea whether to look for a trend, or which dynamic is most likely present in the data. However, this still allows for some fine-tuning of the model, such as adding an autoregression for the white noise or including an external variable. To study the result of these modifications, one can compare different models. To compare the likelihood, the Watanabe-Aikake information criterion (WAIC) (Watanabe, 2010) may be used. The WAIC is very well adapted for comparing the likelihood of Bayesian estimated models, as it uses the whole distribution instead of only the point estimation of the parameter. Further, as we like to have as little noise as possible, the white noise and the innovation noise may be compared between models. Finally, the distribution of the model parameters may be studied to exclude superfluous variables.

In conclusion, several challenges may be found in empirical data analysis. The challenges I found were defining both trend and dynamics, the handling of missing data, the presence of non-normally distributed residuals, the inclusion of external variables and the model selection process. As shown, the BDM can handle all of these challenges. However, the presence of these challenges may result in convergence problems due to lack of data or identifiability issues.

7.2.2 Integrating psychological hypotheses into the BDM

In psychological research, models are used to quantify hypotheses. In Chapters 4, 5 and 6 I explored the possibilities of the BDM with regard to the quantification of hypotheses, and the integration between a psychological hypotheses and frameworks, and the statistical model.

In Chapter 4, I created a model which would fit the hypothesized effects in the data. Here it was hypothesized that the slope was influenced by the presence of agoraphobia and the treatment the individual received. By comparing two models, I was able to study the effect of the pre-treatment symptoms on the intercept of an individual. Using the BDM, I found that there is a difference between treatments with regard to the decrease in symptoms over time. Individuals who received only behavioral therapy showed a slower decrease in symptoms than individuals receiving medication or both medication and behavioral therapy. The presence of agoraphobia had no clear effect on the decrease in symptoms over the different treatments.

In Chapter 5, I put the emotion dynamic features framework as discussed by Kuppens and Verduyn (2015) into one model. To this end I quantified within person and innovation variability, granularity, inertia, cross-lag correlation and the intensity of the emotions in one model. To this end, I used the VAR-BDM, which can handle the multi-individual, multivariate dataset with missing data. Thus the link was lain between a psychological framework, represented by the six emotion dynamic features, and the statistical model used to analyze the data. All of the features were quantified in one model, allowing for an easy comparison of the six features between the different individuals and emotions included in the dataset.

In Chapter 6, I adapted the framework of Kuppens and Verduyn (2015) for affect data. Again, I used the VAR-BDM to analze the multi-individual, bivariate data. Apart from quantifying the six affect features defined, I aimed to see whether there was a weekday effect in the data. This weekday effect is hypothesized on the basis of earlier studies, where there is no clear consensus whether it exists or not. Furthermore, I checked the presence of a moving average effect, i.e., an autoregression in the error. While this is not firmly based in affect theory, it is known that in time series the error is often autocorrelated. The weekday effect was not found in this data, but including the moving average effect improved the model. This combined into a model which included both the psychological framework on affect, given by the interpretation of the parameter of the VAR-BDM, and the statistical theory on modeling time series data, given by the inclusion of the autoregression in the white noise.

7.3 Important points when studying time series

The aim of this thesis was to discuss the limitations and explore the possibilities of psychological time series analysis. However, through the writing of this thesis several points sprang out that beg for attention but I could not elaborate on. These points are, in general, applicable to the analysis of time series data with any model.

The advised data points needed to analyze a data set with a certain model are specific for that model. In this thesis I only structurally studied a simple AR(1) model, with univariate data. The demands for a more complex model may be very much larger. However, little is known about this, apart from the notion that an increase in parameters demands an increase in data points. As such, it is important to check for convergence when analyzing any time series data set and to take into account the uncertainty of the estimated parameters. This should also be taken into account when designing and planning an empirical study. A wise strategy may be to perform a simulation study to find the minimal sample size needed to estimate the preferred model, before deciding on the preferred sample size.

Further, the model comparison is heavily dependent on the choices made by the researcher. While there are several ways to compare models, it is impossible to say whether the optimal model has been found. When all models misspecify the data, the best fitting model will just be the least bad model out of several bad models. It is thus advisable to check for the presence of important trend and dynamic elements before modifying the lesser elements, such as the dependent on external variables.

I found that a model that fits most of the data, may not fit all of the data. For example, in Chapter 6, a general fit comparison shows that Model 3 fits the data better than Model 1 and 2. However, one individual out of the 12 individuals shows strong convergence problems for this model. For this individual, the estimation of Model 3 does not converge, indicating that the data of this individual may better fit Model 1 or 2. Taking this into account, it is important to not only compare a model with a general fit measure, but to study the fit for different individuals. A wise approach may be to compare models per individual.

Finally, the VAR-BDM may be implemented in another modeling framework, such as a network model. A network model aims to charter the relations between several observed variables. In a longitudinal setting, a VAR network model may be used (e.g., Bringmann et al., 2013; Bulteel, Tuerlinckx, Brose, & Ceulemans, 2016). In a VAR network model the autoregressive and cross-regressive relations are represented as the edges between the nodes, which represent the observed variables. As such, it offers an appealing visualization of the relations between the variables studies. A network model generally encompasses a large number of variables, which complicates estimation. As a VAR network is calculated using a VAR model on all dependent variables, one may see the possibilities of using the VAR-BDM to estimate a VAR network model.