Assessing Bayesian Covariance Structure Modelling for Time Series Data

Linus Held (s2798026)

Department of Psychology, Section Cognition, Data and Education (CoDE),

Behavioural Management and Social Sciences, University of Twente

M12 BSc Thesis PSY (2023-2A)

First Supervisor: Dr. G.J.A. Fox

Second Supervisor: Dr. R.H.J. van der Lubbe

June 25th, 2024

Abstract

Time series prediction has important implications for psychological research, for instance when assessing longitudinal change in outcomes or modelling electroencephalogram (EEG) data. One relatively new tool that can be used for this purpose is the Bayesian Covariance Structure Model (BCSM), which can model serial dependences within time series through a covariance model. The suitability of the BCSM for predicting time series has been investigated and in performance compared to the commonly used autoregressive integrated moving average (ARIMA) model. For this purpose, three simulation studies were designed, in which the sample size and properties of the simulated time series were varied. The results show that for time series of shorter lengths, the BCSM provides more accurate predictions than the ARIMA model according to which the data was generated. However, the autocorrelation function (ACF) of the BCSM's residuals indicated that there was still significant autocorrelation, suggesting that not all available information in the time series is used efficiently. The BCSM also failed to describe MA dependences in the data. Nonetheless, the BCSM has shown high predictive accuracy, while remaining relatively simple to interpret.

Keywords: Bayesian Covariance Structure Model (BCSM), Autoregressive Integrated Moving Average (ARIMA), time series prediction, serial correlation, prediction-explanation fallacy.

Table of Contents

Introduction

Longitudinal research in psychology can give researchers insight in complex psychological processes over time. However, large amounts of time series data can be complex to analyse because of serial dependence in the data that violates the assumption of independence of linear regression models and can therefore lead to imprecise estimates of fixed and random effects and their standard errors (Fieberg et al., 2010; Schielzeth et al., 2020). Therefore, literature points to different approaches for time series prediction and explanation, such as autoregressive integrated moving average (ARIMA) models (Jebb et al., 2015). However, the ARIMA method requires that the order of the AR, I and MA components is set a priori, which can be challenging for complex time series (Koreisha & Pokkila, 1995; Liu, 2024). Alternatively, Bayesian Covariance Structure Modelling (BCSM) can be used, which has mainly been tested in the context of random effect modelling for testlets (Fox et al., 2020, 2021; Klotzke & Fox, 2019) and analysis of electroencephalogram (EEG) data (Bazen, 2023). In BCSM, dependences between time series observations are modelled directly, using a covariance matrix, which displays dependences implied by random factor variables and their interactions (Fox et al., 2020; Klotzke & Fox, 2019). By directly modelling these dependences, problems with the assumption of independence can be overcome. Thus, the BCSM might provide an alternative to time series prediction and explanation by reducing the model complexity. Next to that, the BCSM is not limited to positive covariances, because they can be positive, zero or negative (Fox et al., 2020, 2021; Fox & Smink, 2023).

The flexibility of the BCSM to describe complex dependences in the data is examined by assessing its performance on predicting time series data. Hence, the aim of this thesis is to examine whether BCSM is applicable to predict time series data for which the dependence structure is unknown in comparison to an ARIMA model. For this purpose, both models are fitted to simulated ARIMA data and different model parameters indicative of prediction accuracy and model adequacy are compared to each other.

Time Series Data and ARIMA

Time series data can be described as any kind of data that is successively collected over a certain period of time, includes measurements from many time points and is usually univariate (Cryer, 1986, chapter 1). Thus, longitudinal data of psychological processes over time (Ariens et al., 2020; Jebb et al., 2015) or EEG data (Barlow, 1959; Scott & Schiff, 1995) can be considered as time series data relevant to psychological research. Within this kind of data, autocorrelation may occur, meaning that the observations are correlated with each other based on their temporal distance or lag. This autocorrelation needs to be accounted for by the statistical model used to effectively predict a time series, as otherwise the independence assumption is violated.

Time series are frequently predicted using ARIMA models. In a stationary ARIMA model, the time series data revolves around a constant mean with roughly equal variance and a similar pattern of variation over time (Hyndman & Athanasopoulos, 2018, chapter 8.1; Nau, 2014). Furthermore, the ARIMA model consists of an autoregressive (AR) component, which specifies the order of the lagged correlation between observations, an integrated (I) component that keeps the model stationary, and a moving average (MA) component, that captures short-term fluctuations in the data (Cryer, 1986, chapter 2; Nau, 2014). ARIMA models can be used to describe time series data and are fitted by defining the AR, I and MA parameters a priori (Gulnara et al., 2022; See Appendix A for an example). Statistical models that can take autocorrelations into account, such as ARIMA, have been shown to be generally effective in predicting time series data (Korevaar et al., 2022; Turner et al., 2021). Therefore, in this study, ARIMA modelling will be used to compare the efficacy of BCSM in time series modelling.

Bayesian Covariance Structure Modelling

In BCSM, a covariance matrix is used to directly model dependences within data. This covariance structure is implied by random factor variables, where the levels of random factors are set a priori. For predicting a time series, sequential observations of the time series can be clustered, thereby accounting for serial dependence, as their serial correlations are directly modelled within the covariance structure. Therefore, one covariance parameter is associated with each clustering of observations determined by a random factor variable. For this, the number of observations to be clustered can be set a priori, so that several cluster levels are defined to account for dependencies between observations of a time series.

However, not only dependences within the clusters of observations are modelled, but also dependences between clusters. Therefore, the clustering of observations determined by an interaction of random factors can also be included in the covariance structure. Nevertheless, because the BCSM does not require estimates of random effects but directly models dependencies through the covariance structure, it is less complex than random effect models that would require such estimates (Klotzke & Fox, 2019). Consequently, the BCSM might be an efficient alternative for predicting time series with an unknown dependence structure. For drawing the Bayesian inference for the BCSM model parameters, the Markov Chain Monte Carlo (MCMC) method is applied (Gilks et al., 1995; Klotzke & Fox, 2019).

Prediction-Explanation Fallacy

ARIMA models can be used to make predictions about the course of time series data, even in dynamic and complex scenarios. For instance, research has shown a high accuracy of ARIMA forecasts of the spread of COVID-19 (Alabdulrazzaq et al., 2021), predicting traffic volume (Dong et al., 2009), energy output of wind turbines (Biswas et al., 2021) and network traffic (Gulnara et al., 2022). However, ARIMA is also applied in social sciences, for instance to analyse and explain longitudinal trends (Jebb et al., 2015; Vasileiadou & Vliegenthart, 2015). Therefore, literature points towards making a distinction between predicting and explaining when it comes to forecasting models.

In social sciences like psychology, forecasting methods are rather used to investigate causal theories based on data instead of making predictions (Jebb et al., 2015; Shmueli, 2011). This distinction is important to avoid the prediction-explanation fallacy, in which a complex prediction model is used to explain certain phenomena or a simplistic and biased model is used to make predictions (Del Giudice, 2022). Del Giudice (2022) argues that in order to prevent this fallacy, researchers must be aware of a trade-off between the predictive accuracy and the explanatory value of a model. Models that aim to achieve a high predictive accuracy are usually overfitted, which leads to an increased model complexity and makes it difficult to interpret which parameters caused an effect. On the other hand, models that aim to have high explanatory value deliberately fail to accurately describe the predictions but allow for an objective interpretation. BCSM might offer an advantage in this respect, as it is able to make accurate predictions, while the way the data is structured within the covariance matrix is set a priori and remains relatively simple to interpret. Hence, the BCSM has the potential to combine explanatory value and predictive accuracy, thereby circumventing this fallacy.

The Current Study

As there is currently no research on the efficacy of BCSM for modelling time series data, the aim of this thesis is to compare the performance of ARIMA and BCSM models in describing time series data. For this reason, different time series are simulated according to an ARIMA model, and statistical results of the fitted ARIMA and BCSM models are used to assess their performance in describing the simulated time series. The efficacy of a model is assessed by comparing root-mean-square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) of fitted values under the model. The RMSE, MAE and MAPE indicate to which extent the predictions of a model deviate from the observed data and can therefore provide an estimate on how accurate the model describes a time series (Montaño et al., 2013; Willmott & Matsuura, 2005). Furthermore, the autocorrelation function (ACF) of the residuals is computed in order to assess the model's capability to account for autocorrelation of the time series. The ACF of residuals is the correlation of residuals over the time lags, which need to be non-significant for an appropriate statistical model (Box & Pierce, 1970; Cryer, 1986, chapter 3.6; Jebb et al., 2015).

Methods

Materials

Three simulation studies were conducted using R (version 4.2.3), in which different stationary time series were generated according to an ARIMA model. To run and fit the BCSM on the time series, developed functions for MCMC estimation were used.

Simulation Study Design

Simulation Study 1

For the first simulation study, three conditions were considered in which the length of the time series was varied. In the three conditions the AR components of the simulated time series were fixed and set to .4, .2, .1 and -.1 for lags 1, 2, 3 and 4, respectively. The ARIMA for this study design had the following expression

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3} + \varphi_4 Y_{t-4} + e_t,$$

where φ_t are the AR coefficients and e_t the normally distributed errors (Jebb et al., 2015; Nau, 2014). In the first condition, the simulated time series were 128 observations long, in the second condition 512 observations long and in the third condition 1,024 observations long. The length of the time series was varied in these conditions to test how well the BCSM was able to predict time series of different lengths.

Simulation Study 2

In the second study design, another time series consisting of 128 observations was specified, in which the AR coefficients were equal to the ones in the first study. Additionally, an MA component was included and set to 1 in order to assess the capability of a BCSM to model the MA component of a time series. Therefore, this specified ARIMA was expressed as

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3} + \varphi_4 Y_{t-4} + \theta_t e_{t-1} + e_t,$$

where φ_t represent the AR coefficients, θ the MA coefficient and e_t the normally distributed errors (Jebb et al., 2015; Nau, 2014).

Simulation Study 3

In the first two simulation study designs, the AR and MA components of the simulated time series were fixed. Hence, in this study design, the AR coefficients of the time series were varied across replications. For this purpose, AR components of the time series were randomly drawn from a uniform distribution between -.5 and .5 for a fifth order AR time series, consisting of 5,120 observations, with the restriction that a stationary time series was generated. Therefore, the ARIMA for this simulation study had the expression

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3} + \varphi_4 Y_{t-4} + \varphi_5 Y_{t-5} + e_t,$$

For all conditions and every study design, 1,000 data sets were generated. In the first two simulation studies, two BCSM models with different components and an ARIMA model were fitted to the generated time series. In the third simulation study, however, only one BCSM model with three components was compared to the ARIMA model due to computational constraints. The ARIMA model served as the baseline model to which the BCSMs were compared for all conditions.

Model Specification

To predict the simulated time series data in the first simulation study, a fourth order AR model was fitted. For the second simulation study, an ARIMA model with a fourth order AR and a first order MA component was specified. In the third study design, the fitted ARIMA model was specified as a fifth order AR model. Thus, for all generated time series, the ARIMA model according to which the data was generated was fitted. To estimate the time series coefficients, the maximum likelihood estimation method was used.

Moreover, two BCSMs were fitted to the time series data in the first two simulation studies. The first BCSM consisted of eight factors with three components (m=[2,2,4]), in which two observations were clustered on the first level, four on the second level and 16 on the third level. Hence, in the BCSM's covariance matrix, the first level represents a correlation between two sequential observations of the time series and thereby accounting for the first lag, while the second and third level represent correlations between four and 16 observations respectively. As the specified covariance matrix structure only clusters 16 observations, it was replicated eight times for the time series of 128 observations, 32 times for the time series of 512 observations and 64 times for the time series with 1,024 observations. In the third simulation study, this model was compared to the fitted ARIMA model. As the time series consisted of 5,120 observations in this study, the covariance matrix was replicated 320 times.

The second BCSM that was fitted to the time series in simulation studies 1 and 2 consisted of 32 factors with five components (m=[2,2,2,4,4]), so that two, four, eight, 32 and 128 observations were clustered on the respective levels. For the time series with 128 observations, this covariance matrix was only replicated once, for the time series of 512 observations four times and for the time series of 1,024 observations eight times. This model was not fitted to the time series in the third study design.

The dependences of the time series data are accounted for by the BCSM by activating all nested and crossed factor components in the covariance matrix, so that both fitted BCSMs had a full factorial design. The number of MCMC iterations for the BCSM models was set to 5,000 in all conditions. The presented results were averaged over 1,000 simulated time series.

Data Analysis

For data analysis, descriptive statistics of the simulated time series and the residuals of each fitted model were computed. After that, estimates of prediction accuracy were calculated to assess the efficacy of the BCSM fit compared to the ARIMA fit. To assess the prediction accuracy of the different models, their RMSE, MAPE and MAE were compared. The RMSE is denoted as $\sqrt{\frac{1}{n}\sum_{i=1}^{n}|Pi - Oi|^2}$, in which *n* are the number of observations, *Pi* the i-th predicted value of the fitted model and *Oi* the i-th observed value of the simulated time series. Hence, the RMSE is the square root of the average squared difference of the model's estimates and indicates its deviance from the simulated data (Willmott & Matsuura, 2005). The RMSE is frequently used as a measure of deviation and is optimally applied when the standard error of the model is normally distributed (Hodson, 2022).

Additionally, the MAPE was calculated as an indicator of the models' accuracy. The mathematical expression of the MAPE is $\frac{1}{n}\sum_{i=1}^{n}|\frac{Pi-Oi}{Oi}|$, where *n* is the number of observations in the time series, *Oi* the i-th observed value of the simulated time series and *Pi* the i-th predicted value of the fitted model (Ji & Gallo, 2006; Montaño et al., 2013). Therefore, the MAPE is the average absolute standardized difference between observed and predicted values. The MAPE is a percentage deviation of the predicted model from the actual time series data, which provides a standardized measure of deviation.

Similarly, the MAE was calculated, which is expressed as $\frac{1}{n}\sum_{i=1}^{n}|Pi - Oi|$, where *n* is the number of observations in the time series, *Pi* the i-th predicted value and *Oi* the i-th observed value of the time series (Hodson, 2022; Willmott & Matsuura, 2005). The MAE indicates the mean of the absolute differences between predicted and observed values, which provides a third measure of the models' accuracy. Willmott and Matsuura (2005) further argue that the MAE provides a clearer measure of the models' error than the RMSE.

Lastly, the ACF of the different models' residuals was analysed in order to assess whether there was correlation between residuals at different lags. The significance boundary of the residual ACF was calculated in order to assess the fit of the model. A significant residual ACF indicates that the model fails to capture all residual correlation in the data, as the residuals show a correlation where it is assumed that they are independently distributed (Box & Pierce, 1970).

Results

Descriptive Statistics

Descriptive statistics of the partial ACF (PACF) for the first five lags of the simulated data of both simulation studies are described in Table 1. Table 2 contains the descriptive statistics of the residuals of the fitted models for all three simulation studies. From Table 2, it can be inferred that the residuals of all fitted models are centred around a mean and median of zero in all conditions of every simulation study.

Table 1

PACF of Lags 1–5 of the Simulated Time Series Data for all Conditions

Lag	Statistic	Sim. Study 1			Sim. Study 2	Sim. Study 3
		n=128	n=512	n=1,024	n=128	n=5,120
1	Mean	.542	.578	.581	.829	.010
	SD	.102	.047	.034	.043	.366
2	Mean	.274	.290	.292	-.256	-.086
	SD	.076	.038	.026	.079	.358
3	Mean	.046	.058	.058	.157	-.002
	SD	.087	.044	.033	.079	.324
4	Mean	-.109	-.105	-.101	-.224	-.023
	SD	.086	.043	.031	.077	.297
5	Mean	-.008	-.000	-.002	.153	.011
	SD	.080	.044	.031	.080	.269

Note. Results are averaged over 1,000 replications and rounded to the third decimal.

Table 2

Residuals of the fitted BCSM and ARIMA Models in all Simulation Studies

Time Series Length	Statistic	ARIMA	BCSM 1	BCSM 2
		Simulation Study 1		
n=128	Mean	-0.000	0.000	-0.000
	Median	-0.002	0.001	-0.001
	SD	0.983	0.724	0.713
n=512	Mean	0.000	0.000	0.000
	Median	-0.000	0.000	0.000
	SD	0.996	0.721	0.739
n=1,024	Mean	0.000	0.000	0.000
	Median	-0.000	0.000	-0.000
	SD	0.997	0.720	0.742
		Simulation Study 2		
n=128	Mean	0.001	0.000	0.000
	Median	-0.001	-0.001	-0.000
	SD	0.966	0.933	0.842
		Simulation Study 3		
n=5,120	Mean	0.000	-0.008	- [a]
	Median	0.000	-0.006	-
	SD	1.000	2.036	-

Note. BCSM 1 has a structure of m=(2,2,4). BCSM 2 has a structure of m=(2,2,2,4,4). Results are averaged over 1,000 replications and rounded to the third decimal.

[a] In the third simulation study, only one BCSM was fitted.

Performative Statistics

Next to the descriptive statistics, performative analyses were conducted to compare the models' efficacy in predicting a time series. Table 3 shows the average RMSE, MAE and MAPE of each fitted model for each simulation study. In the first simulation study, both BCSMs show a higher accuracy than the ARIMA model as indicated by lower scores of RMSE, MAE and MAPE. Varying the length of the time series did not greatly influence the measures of accuracy for the fitted models. In the second simulation study, the BCSM slightly outperformed the ARIMA model as well. However, introducing an MA component to the time series led to overall worse accuracy when compared to the results of the first simulation study. In the third simulation study with varying AR components, the BCSM shows greater deviations from the simulated time series than the ARIMA model, indicating worse predictions.

Moreover, when comparing both BCSMs to each other of simulation study 1 and 2 there is only a marginal difference between their prediction accuracy. In the first simulation study, the MAPE of the three component BCSM is lower than that of the five component BCSM for the two conditions with sample sizes 128 and 512, but not for the largest sample size of 1,024. The RMSE on the other hand is lower for the largest sample size for the three component BCSM. The MAE is consistently lower across all three conditions for the BCSM with three components. In the second simulation study, the five components BCSM shows slightly better performance for all computed statistics than the three component BCSM.

Table 3

Performative Statistics of the fitted BCSM and ARIMA Models for all Conditions

Time Series Length and Statistic		ARIMA	BCSM 1	BCSM 2
		Simulation Study 1		
n=128				
	RMSE	0.983 [0.979, 0.987]	0.721 [0.709, 0.734]	0.710 [0.689, 0.731]
	MAPE	3.488 [2.975, 4.001]	2.658 [2.347, 2.969]	3.139 [2.562, 3.716]
	MAE	0.787 [0.784, 0.790]	0.577 [0.568, 0.587]	0.569 [0.552, 0.586]
n=512				
	RMSE	0.996 [0.994, 0.998]	0.720 [0.714, 0.726]	0.738 [0.728, 0.748]
	MAPE	6.924 [2.612, 11.236]	4.481 [2.519, 6.443]	4.551 [2.829, 6.273]
	MAE	0.795 [0.793, 0.796]	0.575 [0.570, 0.579]	0.589 [0.581, 0.597]
n=1,024				
	RMSE	0.997 [0.996, 0.999]	0.719 [0.715, 0.724]	0.742 [0.734, 0.749]
	MAPE	4.987 [3.484, 6.490]	4.409 [1.754, 7.063]	4.033 [2.546, 5.519]
	MAE	0.796 [0.794, 0.797]	0.574 [0.570, 0.577]	0.592 [0.586, 0.598]
		Simulation Study 2		
n=128				
	RMSE	0.983 [0.980, 0.987]	0.929 [0.912, 0.947]	0.838 [0.812, 0.864]
	MAPE	3.102 [2.563, 3.642]	2.750 [2.296, 3.205]	2.462 [2.020, 2.904]
	MAE	0.787 [0.784, 0.790]	0.738 [0.724, 0.751]	0.668 [0.648, 0.689]
		Simulation Study 3		
n=5,120				
	RMSE	1.000 [0.999, 1.000]	2.037 [1.672, 2.402]	- [a]
	MAPE	4.801 [4.155, 5.447]	7.005 [5.907, 8.103]	-
	MAE	0.797 [0.797, 0.798]	1.622 [1.332, 1.912]	-

Note. BCSM 1 has a structure of m=(2,2,4). BCSM 2 has a structure of m=(2,2,2,4,4). Results are averaged over 1,000 replications and rounded to the third decimal.

[a] In the third simulation study, only one BCSM was fitted.

Model Adequacy

Lastly, the adequacy of the models for predicting time series data was assessed by analysing the ACF of the models' residuals over the 1,000 replications for the first five lags for all simulation studies. The residual ACF of the ARIMA model was consistently closer to zero over lags than the BCSM in each study and across all conditions, indicating that the ARIMA model generally provides a better model fit than the BCSM for the time series (Table 4; For a graphical representation, see Figures B1 - B3).

Regarding the BCSM, two main findings appeared. Firstly, it became apparent that the longer the time series were, the more often the residual ACF of the first five lags were significant in simulation study 1. For the first lag of the longest time series condition in the first simulation study (1,024 observations) this was the case for all 1,000 generated time series for both fitted BCSMs. The residual ACF value of the fitted BCSMs, however, did not differ greatly across conditions in simulation study 1.

Secondly, it was found that the BCSM is better able to capture AR dependencies than MA dependencies in time series. When an MA component was included in the time series, the BCSM had higher residual ACF values and showed significant autocorrelations of residuals more than 90 percent of the time for the first lag. Comparing the residual ACF of Table 4 to the time series PACF in Table 1, it can be inferred that the ARIMA fit captured most of the autocorrelation in the data, so that the residual ACF was approximately zero. The BCSM fit, on the other hand, only marginally captured the autocorrelation of the time series.

Table 4

Residual ACF of the fitted BCSM and ARIMA Models for all conditions averaged over all replications

Model	Lag 1		Lag 2		Lag 3		Lag 4		Lag 5	
	M (SE)	Significant	M (SE)	Significant	M (SE)	Significant	M (SE)	Significant	M (SE)	Significant
Simulation Study 1										
n=128										
ARIMA	-.007 (.015)	0.0%	-.007 (.019)	0.0%	-.008 (.032)	0.0%	-.007 (.053)	0.2%	-.002 (.076)	2.0%
BCSM1	-.156 (.081)	41.5%	-.065 (.104)	15.1%	-.001 (.088)	5.1%	-.025 (.104)	10.1%	-.017 (.084)	4.9%
BCSM2	-.163 (.102)	46.5%	.085 (.122)	23.8%	-.020 (.102)	8.5%	.001 (.126)	13.8%	-.029 (.099)	9.3%
n=512										
ARIMA	-.002 (.006)	0.0%	-.002 (.006)	0.0%	-.002 (.014)	0.0%	-.002 (.028)	0.1%	-.000 (.040)	2.5%
BCSM1	-.163 (.038)	97.6%	-.059 (.048)	29.0%	-.000 (.041)	3.3%	-.002 (.049)	7.8%	-.019 (.042)	5.9%
BCSM2	-.162 (0.41)	96.8%	-.076 (.055)	42.2%	-.023 (.043)	7.6%	.016 (.052)	10.1%	-.044 (.045)	17.0%
n=1,024										
ARIMA	-.001 (.004)	0.0%	-.000 (.004)	0.0%	-.000 (.010)	0.0%	-.000 (.019)	0.1%	-.001 (.028)	2.7%
BCSM1	-.164 (.027)	100%	-.058 (.035)	47.3%	-.000 (.029)	3.1%	.000 (.037)	10.6%	-.019 (.030)	7.6%
BCSM2	-.160 (.028)	100%	-.075 (.037)	65.5%	-.024 (.031)	10.0%	.017 (.036)	12.1%	-.046 (.031)	30.6%
Simulation Study 2										
n=128										
ARIMA	-.006 (.018)	0.0%	-.010 (.206)	0.0%	-.010 (.032)	0.0%	-.012 (.055)	0.3%	.011 (.074)	2.1%
BCSM1	-.327 (.112)	94.2%	.099 (.152)	30.4%	-.148 (.122)	36.9%	.013 (.152)	23.4%	-.061 (.114)	16.0%
BCSM2	-.328 (.116)	93.8%	.074 (.170)	31.7%	-.142 (.131)	35.5%	.041 (.154)	24.7%	-.037 (.120)	12.2%
Simulation Study 3										
n=5,120										
ARIMA	-.000 (.004)	0.0%	-.000 (.006)	0.1%	-.000 (.006)	0.0%	.000 (.008)	0.3%	-.000 (.008)	0.4%
BCSM1	-.002 (.122)	82.9%	.007 (.094)	76.2%	-.005 (.145)	85.7%	.043 (.096)	72.3%	.004 (.151)	87.4%

Note. BCSM 1 is a three component BCSM with the structure m=(2,2,4). BCSM 2 is a five component BCSM with the structure m=(2,2,2,4,4). Results are averaged over 1,000 replications and rounded to the third decimal. The significance values indicate how many times the residual ACF was significant out of the 1,000 replications.

Discussion

Summary and Explanation of Findings

The aim of this study was to assess whether BCSM is an effective approach for predicting time series. It was investigated whether the BCSM is able to describe dependences in time series through a clustering method, thereby accounting for serial correlations. For this reason, three simulation studies were conducted in which the performance to predict simulated time series of the BCSM was compared to an ARIMA model. In the first simulation study, the effectiveness of the BCSM was assessed for predicting time series of varying lengths with fixed AR components. In the second simulation study, an additional MA component was added to the time series and in the third simulation study, the AR components of the time series were varied across replications.

When the AR and MA components of the predicted time series were fixed in the first two simulation studies, the BCSM outperformed the ARIMA model in terms of prediction accuracy. Contrarily, in the third simulation study, the BCSM showed a lower prediction accuracy than the ARIMA model. Thus, when the AR and MA components were fixed, the BCSM made more accurate predictions than ARIMA but not when the AR components were varied. Nevertheless, according to Montaño et al. (2013), the fitted BCSM and ARIMA models in all conditions and in every study design were highly accurate, as their MAPE was consistently lower than 10. This indicates that the BCSM is generally capable of making accurate predictions. Moreover, when comparing the predictive performance of the two BCSMs with different covariance matrix structures, there was only a marginal difference. Hence, because of their similar covariance matrix structure, the predictive performance of both BCSMs was approximately equal to each other in the first two simulation studies.

When looking at the residual ACF, however, it was found that the specified BCSMs were unable to account for all the remaining autocorrelation of the residuals. Thus, the residuals of the fitted model displayed a systematic relationship, indicating that they are not independent of each other (Hyndman & Athanasopoulos, 2018, chapter 3.3). That means that the assumption of independently distributed residuals is violated, which causes problems for further methods that can be applied (e.g. Kenny & Judd, 1986). The dependence of the residuals might be explained by the specified structure of the BCSM. For all fitted BCSMs in every study, a full factorial design was applied so that the dependence structure was based on all crossed and nested factor components. This, however, might have led to overfitting of the model so that random noise within the data was modelled too closely (Montesinos-López et al., 2022). Thereby, the serial correlation within the data might not have been accurately accounted for. Moreover, the residual ACF became more often significant with increasing observations of the time series, even though the residual ACF did not greatly change across conditions in the first simulation study. This might have happened because the significance criterion of the ACF decreases with increasing time series length, leading to more significant results (Box & Pierce, 1970; Cryer, 1986, chapter 3.6).

Furthermore, it was found that the residual ACF of the BCSMs greatly increased when an MA component was included in the time series. This can be explained by the fact that the serial correlation of a time series can be modelled by using the covariances of lagged variables, which were directly modelled in the covariance matrix. An MA component, however, consists of lagged errors (Hyndman & Athanasopoulos, 2018, chapter 8.4; Nau, 2014), which cannot be predicted directly by covariances between clustered observations. Hence, when an MA component is included in the time series, the BCSM fails to account for the MA process, resulting in remaining autocorrelation between residuals.

Relation to Previous Work

Previous research on the BCSM made comparisons to mixed or random effect models in the contexts of analysing testlet data (Fox et al., 2020, 2021; Klotzke & Fox, 2019) and EEG data (Bazen, 2023). In those studies, the BCSM has shown that dependences within the data can directly be modelled with a covariance structure. Moreover, the BCSM has been shown to be an efficient alternative to mixed and random effect models as it is applicable to small sample sizes, not limited to positive covariances and less complex than models that need to include many random effect variables. The current study has further shown that the BCSM is not limited to random and mixed effect modelling, but that it can also be applied to time series. This is done by clustering sequential observations and directly modelling their dependence structure within a covariance matrix.

Limitations and Suggestions for Future Research

The generated time series in each simulation study consisted of multiple AR components and in one case an MA component. However, the BCSM has not been tested for time series with a trend or seasonality. Nonetheless, time series with a trend also commonly occur and are relevant in research. Examples for this in the field of psychology might be EEG data that varies as a function of a certain treatment or stimulation (Hinrichs et al., 1996; Vayá Abad, 2023) or longitudinal research in how psychological processes change over time (Ariens et al., 2020). Hence, the presented results are limited to time series that are stationary, which have an equal variance and mean over time (Nau, 2014). For this reason, future research should also address whether the BCSM is applicable to non-stationary time series.

Furthermore, only two BCSMs with relatively similar component structure have been compared to each other in the first two simulation studies. In every case, a full factorial design was applied so that all nested and crossed covariances were modelled. The results of both fitted BCSMs only marginally differed from each other in terms of accuracy and their capability to account for the autocorrelation of the data. Hence, both specified BCSMs might have been too similar to make a meaningful comparison between each other. For this reason, it was not expected that in the third simulation study, the inclusion of the second BCSM would have resulted in a great difference compared to the BCSM that was fitted. For a better understanding of how the specification of the covariance matrix and the factorial design of the BCSM models influences the accuracy of time series prediction, more diverse testing is needed. Therefore, in future research the factorial structure of the BCSM could be adjusted, so that not all crossed and nested terms are activated to prevent overfitting. If a design is applied that is capable of taking all the remaining serial correlations of the residuals into account, this could increase the prediction error as fewer components are used to model the data. However, this might also improve the model fit since the residuals might be less correlated.

An additional suggestion for future research that could give insights into the benefits and shortcomings of the BCSM is to compare the fitted covariance matrix to the sample correlation matrix of the time series. In the current study, only the properties of the fitted models' residuals were analysed. The correlation matrix of a time series contains further information about the underlying dynamics of a time series (Easaw et al., 2023) and has, for instance, been used in research to model and explain EEG data (Schindler et al., 2007). Therefore, comparing the sample correlation matrix of a time series to the fitted covariance matrix of the BCSM might provide further information on how well the BCSM is able to model time series processes and has implications for psychological research.

Finally, next to the residual ACF, the PACF could be investigated. The residual PACF shows correlations of the residuals over lags, while removing the influence of earlier lags, and can therefore be used to examine whether AR components in the data are correctly modelled (Weiß et al., 2023). Thus, in future research, the PACF can also be assessed next to the ACF.

Practical Application of Findings

In the future, BCSM might provide a powerful tool for predicting time series. Especially in social sciences like psychology, BCSM might offer advantages as it does not make a large trade-off between interpretability and predictive accuracy, which has the potential to diminish problems associated with the prediction-explanation fallacy (Del Giudice, 2022). The way how observations are clustered is set a priori and easily interpretable, so that the BCSM can be used for explanatory modelling. On the other hand, it has been shown that the predictive performance of the BCSM was in most cases better than the ARIMA model, so that a high predictive accuracy is given as well.

Furthermore, in practice, the time series data is not simulated according to an ARIMA model. For this reason, it can be more challenging to find the appropriate order of the ARIMA components for real-world applications (e.g. Liu, 2024). Because of this, it is also likely that the models would perform differently in a real setting. Thus, comparing the BCSM to an optimally fitted ARIMA model on real data would yield a more realistic comparison between both models. Nevertheless, in any case, diagnostics, such as the residual ACF and PACF should be examined, to ensure that the model is able to adequately account for dependences within the data (Box & Pierce, 1970; Jebb et al., 2015). If the fitted model does not accurately account for remaining residual autocorrelation or underlying dependence structures, the factorial structure of the BCSM should be refined to avoid over- or underfitting.

Conclusion

In this thesis, the ability of the BCSM to make predictions about the trajectory of a time series has been assessed in three simulation studies. The results indicate that the accuracy of the method is high compared to an ARIMA model, especially for shorter time series. However, there are also some drawbacks to the BCSM, as it does not account for all the correlation between the residuals, which could indicate a structural bias. Therefore, future research needs to address how the BCSM can account for serial correlation within the residuals of the predicted time series, which will allow making more adequate predictions. This could, for instance, be done by testing and refining the factorial structure of the BCSM, as in the current study only a full factorial design was applied. When the problem regarding the significant autocorrelation of the model's residuals can be resolved, the BCSM could become a powerful and effective tool for predicting time series in the future. Next to that, the BCSM offers benefits in terms of explanatory and predictive value, without making large trade-offs between them. Therefore, the BCSM might offer an alternative to overly complex machine learning tools, which are often not easily interpretable.

References

Alabdulrazzaq, H., Alenezi, M. N., Rawajfih, Y., Alghannam, B. A., Al-Hassan, A. A., & Al-Anzi, F. S. (2021). On the accuracy of ARIMA based prediction of COVID-19 spread. *Results in Physics*, *27*. Scopus. https://doi.org/10.1016/j.rinp.2021.104509

Ariens, S., Ceulemans, E., & Adolf, J. K. (2020). Time series analysis of intensive longitudinal data in psychosomatic research: A methodological overview. *Journal of Psychosomatic Research*, *137*, 110191. https://doi.org/10.1016/j.jpsychores.2020.110191

Barlow, J. S. (1959). Autocorrelation and Crosscorrelation Analysis in Electroencephalography. *IRE Transactions on Medical Electronics*, *ME-6*(3), 179–183. https://doi.org/10.1109/IRET-ME.1959.5007949

Bazen, J. L. (2023). *Statistical Analysis of EEG (Trial) Data through Covariance Structure Modelling* (public). http://essay.utwente.nl/95279/

Biswas, A. K., Ahmed, S. I., Bankefa, T., Ranganathan, P., & Salehfar, H. (2021). *Performance Analysis of Short and Mid-Term Wind Power Prediction using ARIMA and Hybrid Models*. 2021 IEEE Power and Energy Conference at Illinois, PECI 2021. Scopus. https://doi.org/10.1109/PECI51586.2021.9435209

Box, G. E. P., & Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*, *65*(332), 1509–1526. JSTOR. https://doi.org/10.2307/2284333

Cryer, J. D. (1986). *Time series analysis* (Bd. 286). Springer. https://mybiostats.wordpress.com/wp-content/uploads/2015/03/time-series-analysis-with-applications-in-r-cryer-and-chan.pdf

Del Giudice, M. (2022). *The Prediction-Explanation Fallacy: A Pervasive Problem in Scientific Applications of Machine Learning*. https://doi.org/10.31234/osf.io/4vq8f

Dong, H., Jia, L., Sun, X., Li, C., & Qin, Y. (2009). *Road traffic flow prediction with a time-oriented ARIMA model*. 1649–1652. Scopus. https://doi.org/10.1109/NCM.2009.224

Easaw, N., Lee, W. S., Lohiya, P. S., Jalan, S., & Pradhan, P. (2023). Estimation of correlation matrices from limited time series data using machine learning. Journal of Computational Science, 71, 102053. https://doi.org/10.1016/j.jocs.2023.102053

Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M. S., & Frair, J. L. (2010). Correlation and studies of habitat selection: Problem, red herring or opportunity? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2233–2244. Scopus. https://doi.org/10.1098/rstb.2010.0079

Fox, J.-P., Koops, J., Feskens, R., & Beinhauer, L. (2020). Bayesian covariance structure modelling for measurement invariance testing. *Behaviormetrika*, *47*, 385–410. https://doi.org/10.1007/s41237-020-00119-3

Fox, J.-P., & Smink, W. A. C. (2023). Assessing an alternative for "negative variance components": A gentle introduction to Bayesian covariance structure modeling for negative associations among patients with personalized treatments. *Psychological Methods*, *28*(1), 1–20. https://doi.org/10.1037/met0000442

Fox, J.-P., Wenzel, J., & Klotzke, K. (2021). The Bayesian Covariance Structure Model for Testlets. *Journal of Educational and Behavioral Statistics*, *46*(2), 219–243. https://doi.org/10.3102/1076998620941204

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice* (1st ed.). Chapman and Hall/CRC. https://doi-org.ezproxy2.utwente.nl/10.1201/b14835

Gulnara, B., Ahmad, A., Mirzakulova, S., & Ибраева, Ж. (2022). TIME SERIES FORECASTING BY THE ARIMA METHOD. *Scientific Journal of Astana IT University*, 14–23. https://doi.org/10.37943/HFCH4395

Hinrichs, H., Feistner, H., & Heinze, H. J. (1996). A trend-detection algorithm for intraoperative EEG monitoring. *Medical Engineering & Physics*, *18*(8), 626–631. https://doi.org/10.1016/s1350-4533(96)00023-9

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts. http://otexts.org/fpp/

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, *6*, 727. https://doi.org/10.3389/fpsyg.2015.00727

Ji, L., & Gallo, K. (2006). An Agreement Coefficient for Image Comparison. *Photogrammetric Engineering and Remote Sensing.*, *73*, 823–833. https://doi.org/10.14358/PERS.72.7.823

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. Psychological Bulletin, 99(3), 422–431. https://doi.org/10.1037/0033-2909.99.3.422

Klotzke, K., & Fox, J.-P. (2019). Bayesian covariance structure modelling of responses and process data. *Frontiers in Psychology*, *10*(JULY). Scopus. https://doi.org/10.3389/fpsyg.2019.01675

Koreisha, S. G., & Pokkila, T. (1995). A comparison between different order-determination criteria for identification of arima models. *Journal of Business and Economic Statistics*, *13*(1), 127–131. Scopus. https://doi.org/10.1080/07350015.1995.10524586

Korevaar, E., Karahalios, A., Turner, S. L., Forbes, A. B., Taljaard, M., Cheng, A. C., Grimshaw, J. M., Bero, L., & McKenzie, J. E. (2022). Methodological systematic review recommends improvements to conduct and reporting when meta-analyzing interrupted time series studies. *Journal of Clinical Epidemiology*, *145*, 55–69. https://doi.org/10.1016/j.jclinepi.2022.01.010

Liu, J. (2024). Navigating the Financial Landscape: The Power and Limitations of the ARIMA Model. *Highlights in Science, Engineering and Technology*, *88*, 747–752. https://doi.org/10.54097/9zf6kd91

Montaño, J., Palmer, A., Sesé, A., & Cajal, B. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, *25*, 500–506. https://doi.org/10.7334/psicothema2013.23

Montesinos-López, O., Montesinos, A., & Crossa, J. (2022). Overfitting, Model Tuning, and Evaluation of Prediction Performance (pp. 109–139). https://doi.org/10.1007/978-3-030-89010-0_4

Nau, R. (2014). Notes on nonseasonal ARIMA models. *Fuqua School of Business, Duke University*. https://people.duke.edu/~rnau/Notes_on_nonseasonal_ARIMA_models--Robert_Nau.pdf

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Schindler, K., Leung, H., Elger, C. E., & Lehnertz, K. (2007). Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG. Brain : A Journal of Neurology, 130(Pt 1), 65–77. https://doi.org/10.1093/brain/awl304

Scott, D. A., & Schiff, S. J. (1995). Predictability of EEG interictal spikes. *Biophysical Journal*, *69*(5), 1748–1757. https://doi.org/10.1016/S0006-3495(95)80044-5

Shmueli, G. (2011). To Explain or to Predict? *Statistical Science*, *25*. https://doi.org/10.1214/10-STS330

Turner, S. L., Forbes, A. B., Karahalios, A., Taljaard, M., & McKenzie, J. E. (2021). Evaluation of statistical methods used in the analysis of interrupted time series studies: A simulation study. *BMC Medical Research Methodology*, *21*(1), 181. https://doi.org/10.1186/s12874-021-01364-0

Vasileiadou, E., & Vliegenthart, R. (2015). Studying dynamic social processes with ARIMA modeling. *International Journal of Social Research Methodology*, *17*. https://doi.org/10.1080/13645579.2013.816257

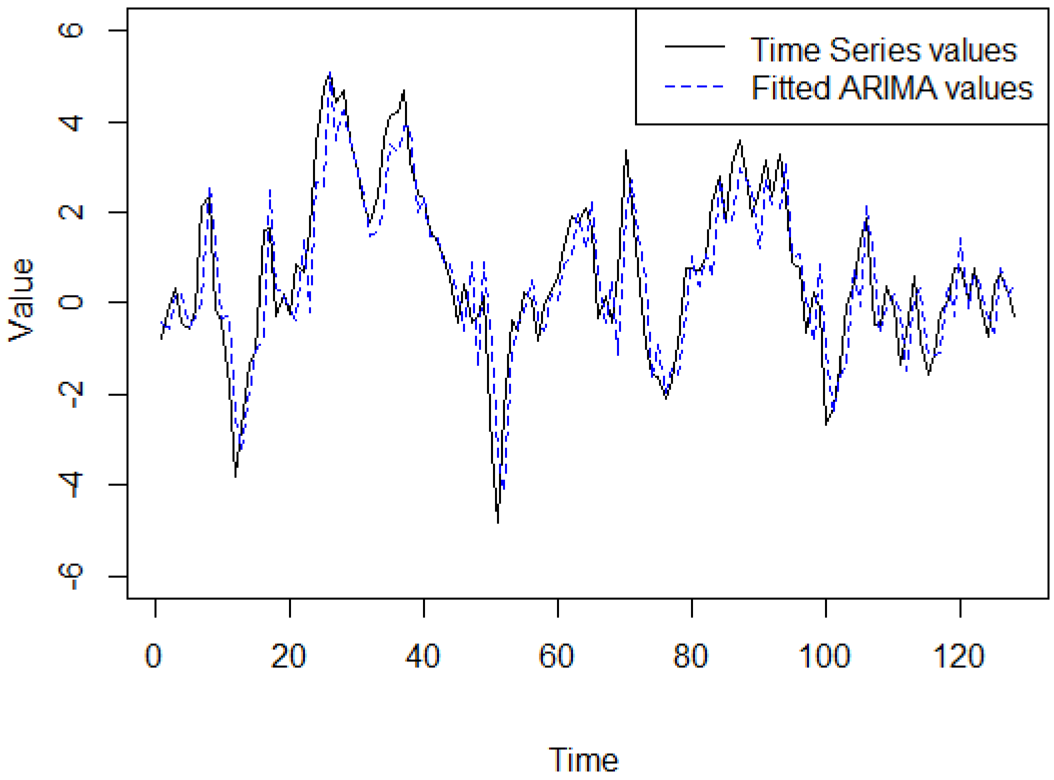
Vayá Abad, G. (2023). ARIMA processes for EEG modeling. *Universitat de Barcelona*. http://hdl.handle.net/2445/203545

Weiß, C. H., Aleksandrov, B., Faymonville, M., & Jentsch, C. (2023). Partial Autocorrelation Diagnostics for Count Time Series. Entropy, 25(1). https://doi.org/10.3390/e25010105

Willmott, C., & Matsuura, K. (2005). Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Climate Research*, *30*, 79. https://doi.org/10.3354/cr030079

Appendix A

Example Figure of a Time Series and a fitted ARIMA Model



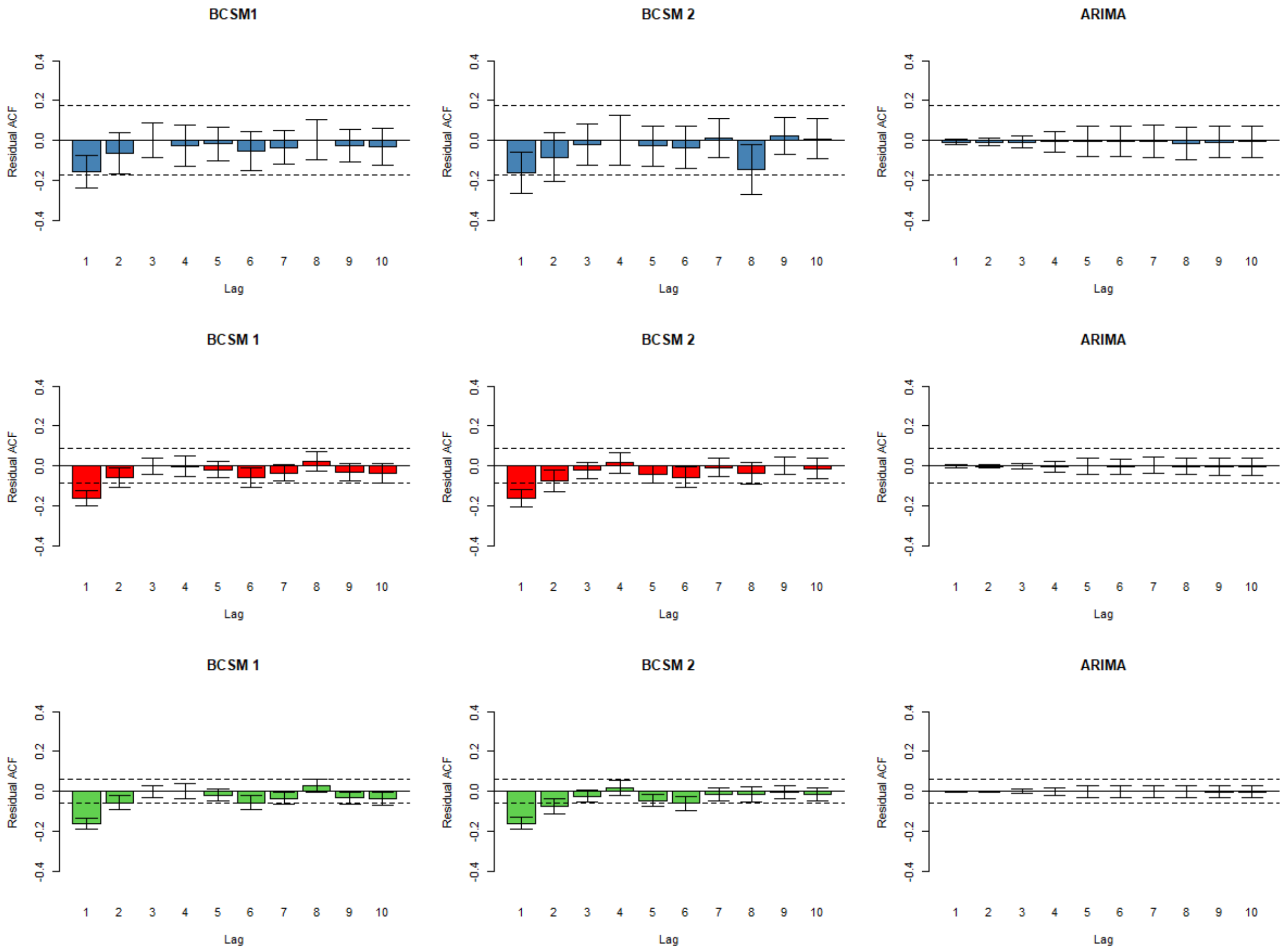
Note. Example plot of an autoregressive moving average time series (n = 128; AR components = .4, .3, .1, -.1; MA component = 1) and fitted values of an ARIMA (4,0,1) model.

Appendix B

Figures of the residual ACF of all fitted Models in Simulation Studies 1 to 3

Figure B1

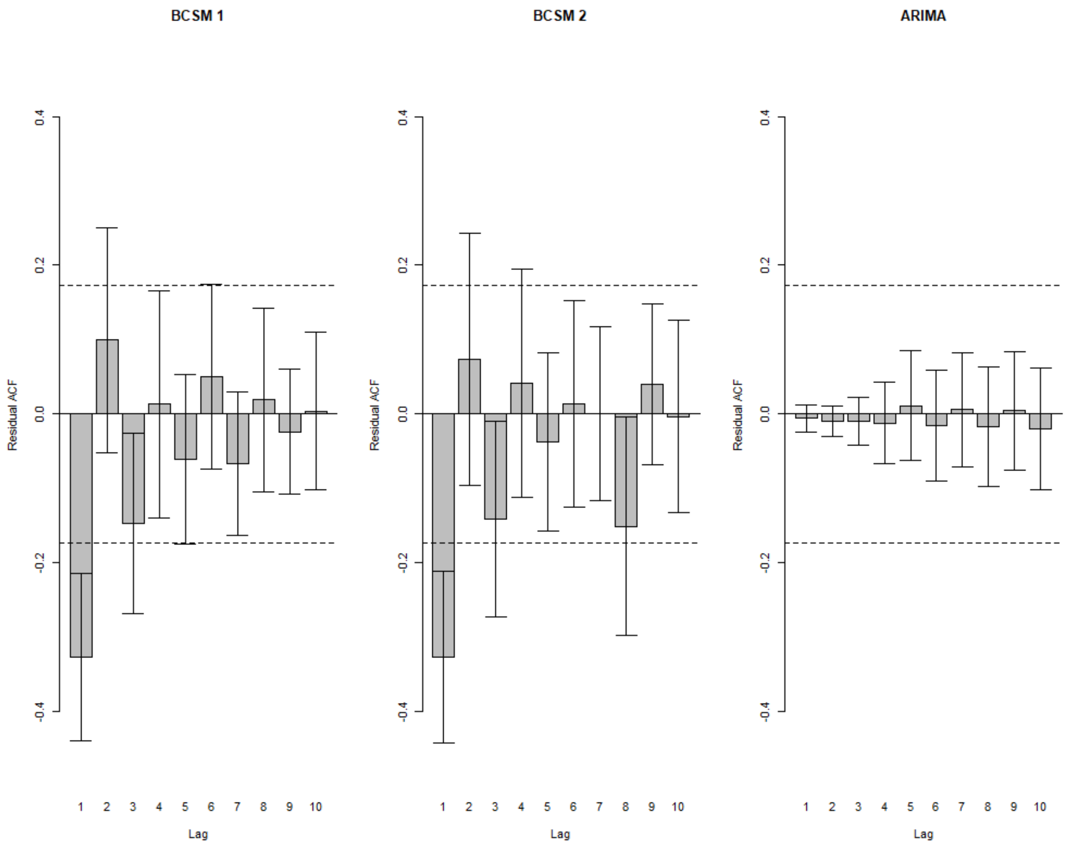
Average residual ACF over 1,000 replication for Simulation Study 1



Note. The first, second, and third row respectively represent the residual ACF of the three fitted models in the first condition (n=128), the second condition (n=512) and the third condition (n=1,024). BCSM 1 has a structure of m=(2,2,4), BCSM 2 has a structure of m=(2,2,2,4,4). The figures represent the average ACF over 1,000 data sets, with an error bar indicating a 95% confidence interval.

Figure B2

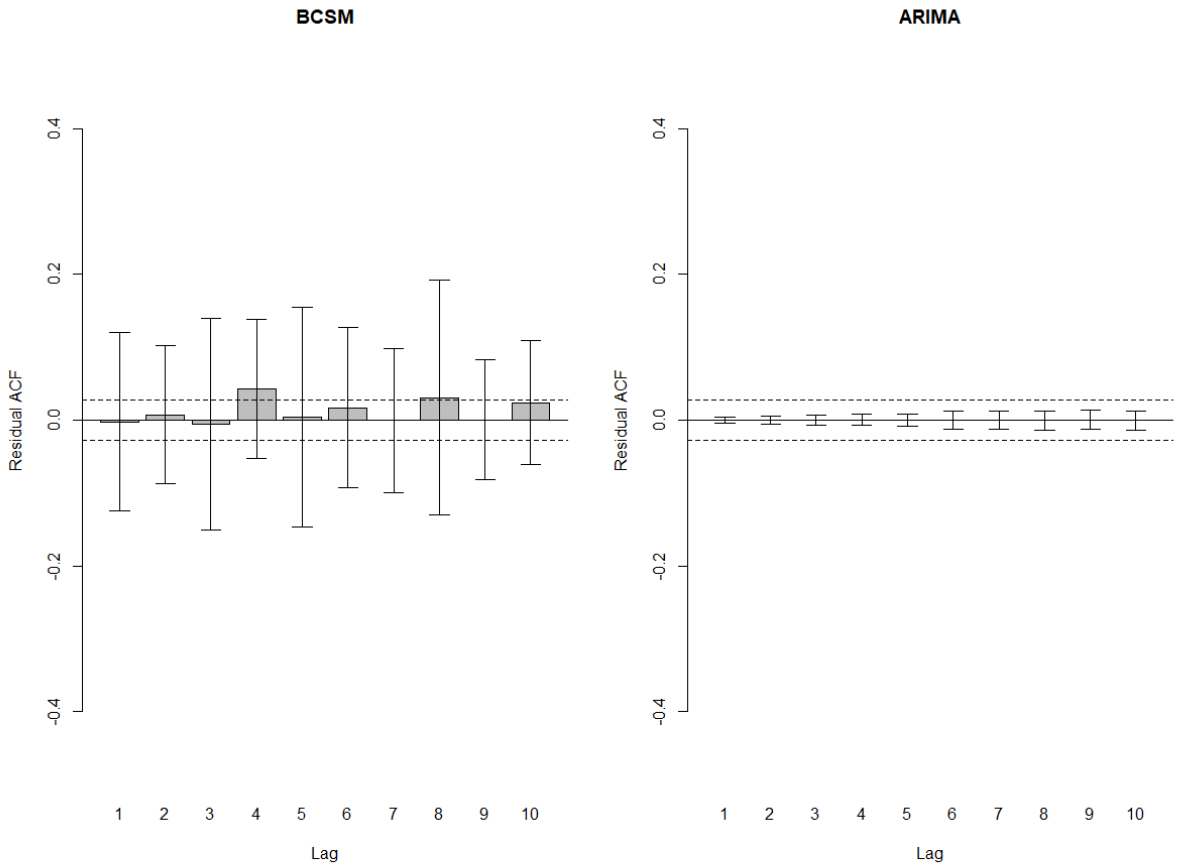
Average residual ACF over 1,000 replications for Simulation Study 2



Note. BCSM 1 has a structure of m=(2,2,4), BCSM 2 has a structure of m=(2,2,2,4,4). The figures represent the average ACF over 1,000 data sets with an error bar indicating a 95% confidence interval.

Figure B3

Average residual ACF over 1,000 replications for Simulation Study 3



Note. The BCSM has a structure of m=(2,2,4). The figures represent the average ACF over 1,000 data sets, with an error bar indicating a 95% confidence interval.