Hybrid ARDL-MIDAS-Transformer Time-Series Regressions for Multi-Topic Crypto Market Sentiment Driven by Price and Technology Factors: Supplementary Appendix

Ioannis Chalkiadakis
School of Mathematical and Computer Sciences,
Heriot-Watt University
ic14@hw.ac.uk

Gareth W. Peters
Department of Statistics & Applied Probability
University of California Santa Barbara
garethpeters@ucsb.edu

Matthew Ames
ResilientML
Victoria, Australia
matt.ames87@gmail.com

March 15, 2023

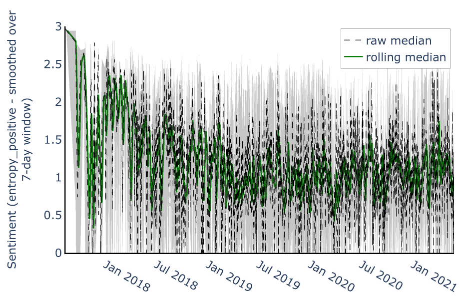
Contents

A Additional Entropy Sentiment time-series: Ethereum

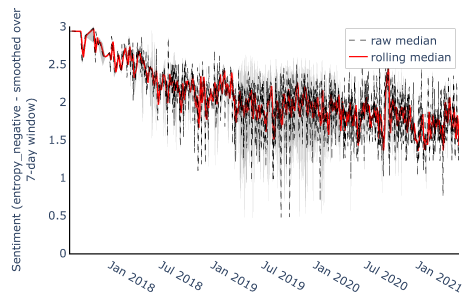
In Figure 1 we plot the smoothed volume-weighted positive and negative sentiment indices as well as the index of absolute sentiment strength, for articles referring to Ethereum.

Figure 1: Sentiment indices with 95% confidence intervals constructed from articles about Ethereum published on Cryptodaily (http://cryptodaily.co.uk) and Cryptoslate (http://cryptoslate.com)

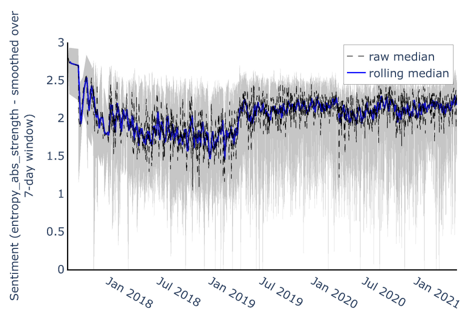
(a) Positive sentiment



(b) Negative sentiment



(c) Absolute sentiment strength

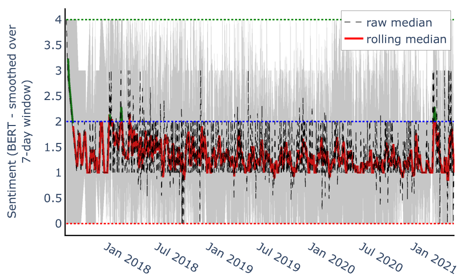


B Additional BERT and VADER sentiment time-series: Ethereum

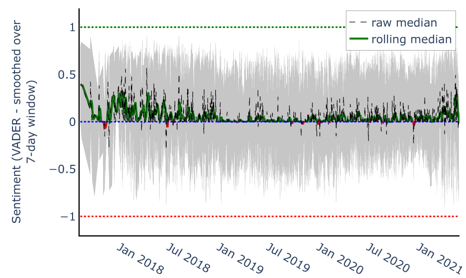
In Figure 2 we plot the sentiment time-series for ETH obtained via the BERT and VADER models.

Figure 2: Sentiment indices based on BERT and VADER with 95% confidence intervals constructed from articles about Ethereum published on Cryptodaily (http://cryptodaily.co.uk) and Cryptoslate (http://cryptoslate.com)

(a) BERT



(b) VADER



C Two-stage Estimation of ARDL(∞) Regression via the classical Koyck transform

In this section we will present the estimation procedure for the classical infinite-lag distributed model, where we consider utilising two covariates. We will consider time-series $\{Y_t\}$ given by the daily sentiment score based on our proposed Entropy Sentiment time-series construction. We will then regress this sentiment scoring method against alternative sentiment extraction methods based on BERT and VADER that we will transform into time-series covariates and denote them by $\{X_t^B\}$ and $\{X_t^V\}$, also constructed from daily measures of sentiment. Then we seek to fit the regression model, which includes the geometric decay parametric form for the coefficients as described in Section 2.1 of the main manuscript:

$$\begin{aligned} Y_t &= \alpha + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{j=0}^{+\infty} \beta_j^B X_{t-j}^B + \sum_{j=0}^{+\infty} \beta_j^V X_{t-j}^V + \epsilon_t \\ &= \alpha + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \beta^B \sum_{j=0}^{+\infty} \phi_B^j X_{t-j}^B + \beta^V \sum_{j=0}^{+\infty} \phi_V^j X_{t-j}^V + \epsilon_t. \end{aligned} \tag{1}$$

First we will include only one of the additional covariates, say BERT. At times t and $t-1$ we have:

$$\begin{aligned} Y_t &= \alpha + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{j=0}^{+\infty} \beta_j^B X_{t-j}^B + \epsilon_t, \\ Y_{t-1} &= \alpha + \sum_{i=1}^{p} \gamma_i Y_{t-1-i} + \beta^B \sum_{j=0}^{\infty} \phi_B^j X_{t-1-j}^B + \epsilon_{t-1}. \end{aligned} \tag{2}$$

Working as before, we begin by multiplying the second row with the geometric decay rate ϕ_B:

$$\phi_B Y_{t-1} = \phi_B \alpha + \phi_B \sum_{i=1}^{p} \gamma_i Y_{t-1-i} + \beta^B \sum_{j=0}^{+\infty} \phi_B^{j+1} X_{t-1-j}^B + \phi_B \epsilon_{t-1} \tag{3}$$

Subtracting the expressions in equations 2 (for Y_t) and 3 we obtain:

$$Y_t - \phi_B Y_{t-1} = \alpha(1-\phi_B) + \epsilon_t - \phi_B \epsilon_{t-1} + \sum_{i=1}^{p} \gamma_i (Y_{t-i} - \phi_B Y_{t-1-i}) + \beta^B X_t^B,$$

and equivalently we have:

$$Y_t = \alpha(1-\phi_B) + \phi_B Y_{t-1} + \sum_{i=1}^{p} \gamma_i (Y_{t-i} - \phi_B Y_{t-1-i}) + \beta^B X_t^B + \epsilon_t - \phi_B \epsilon_{t-1}. \tag{4}$$

If we then work in the same way for the VADER covariate we obtain:

$$Y_t = \alpha(1-\phi_V) + \phi_V Y_{t-1} + \sum_{i=1}^{p} \gamma_i (Y_{t-i} - \phi_V Y_{t-1-i}) + \beta^V X_t^V + \epsilon_t - \phi_V \epsilon_{t-1},$$

and after adding the last two models, one obtains:

$$Y_t = \alpha(1-\phi) + \phi Y_{t-1} + \sum_{i=1}^{p} \gamma_i (Y_{t-i} - \phi Y_{t-1-i}) + \beta^B X_t^B + \beta^V X_t^V + \epsilon_t - \phi \epsilon_{t-1},$$

where we have substituted $\phi = \frac{\phi_B + \phi_V}{2}$.

The final regression model can be estimated with a least squares procedure. Note that we cannot use ordinary least squares (OLS) as it would produce inconsistent estimators since the error term is autocorrelated and therefore correlated with the lagged dependent variable Y_{t-1}. We will instead estimate the model in a two-step procedure using instrumental variables (IV) regression. Because we do not aim to obtain a good model for Y_t for prediction purposes, to simplify our estimation we keep only one lag of the dependent variable in the regression and set $\gamma_i = 0$ for $i = 1, \dots, p$:

$$Y_t = \alpha(1-\phi) + \phi Y_{t-1} + \beta^B X_t^B + \beta^V X_t^V + \epsilon_t - \phi \epsilon_{t-1}. \tag{5}$$

For completeness, we repeat here the two steps of the regression, which have been presented in detail in the main manuscript, Section 6.1:

Stage I : Regress the entropy sentiment index against its lag $1, \dots, p$ values, by fitting an AR(p) model on Y_t.

Stage II : Determine the most suitable model for the instrumental variables of Y_t covariate.

C.1 Stage I

In this first step we remove the autocorrelation from the independent variable by regressing the following:

$$Y_t = \mu_0 + \sum_{j=1}^{p} \mu_j Y_{t-j} + \epsilon_t. \tag{6}$$

We can estimate the parameters of the previous AR(p) model by ordinary least squares, and given those we obtain the raw residuals E_t^y. This first regression will be our reference model for the regressions in Stage II; if we improve the fit with statistical significance in terms of AIC [1] by adding covariates, then the added covariates will be successful in explaining some of the content of Y_t.

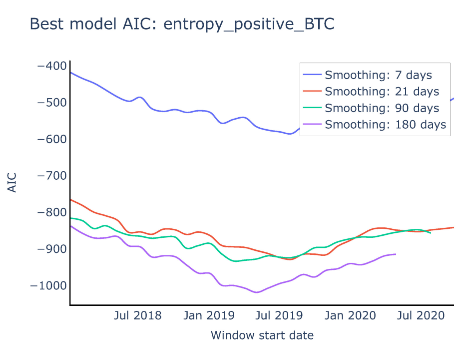
C.1.1 Koyck transform Stage I: model selection and calibration plots for the Positive and Negative Entropy Sentiment indices

As described in the main manuscript, for each configuration we stored the successfully fitted models and ranked them according to the AIC. In Figure 3, we plot the AIC of the best fitting model per window for all smoothing parametrisations. We next plot the lag order per fitting window for the best fitting model (Figure 4 - BTC, 5 - ETH, left panels), and in addition, in the same plots we demonstrate the lag order per fitting window for the best fitting model after dropping the coefficients that were not statistically significant to any level up to 90%.

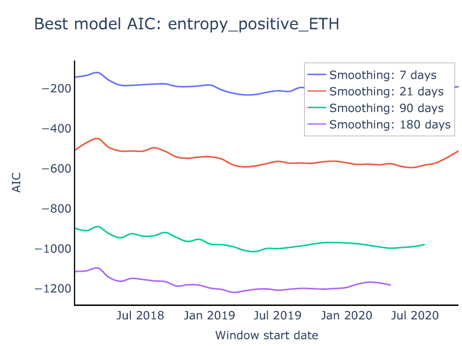
In Figures 4 and 5, we use red colour to plot the worst performing model according to the AIC. Note that for Bitcoin, the importance of selecting a different lag per window exists in the case of statistically significant coefficients, although less prevalent, in the fits for the signals of different polarities (positive vs negative); there the lag structure mainly fluctuates around either one or two lags.

Figure 3: Stage I rolling window fits: best AR(p) model AIC score.

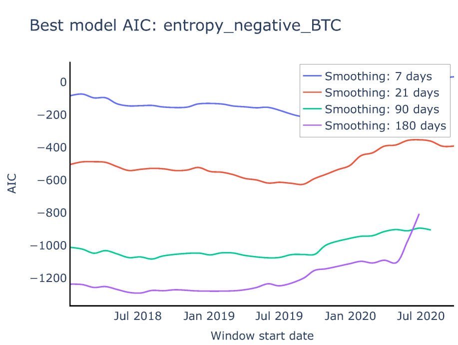
(a) BTC - positive sentiment



(b) ETH - positive sentiment



(c) BTC - negative sentiment



(d) ETH - negative sentiment

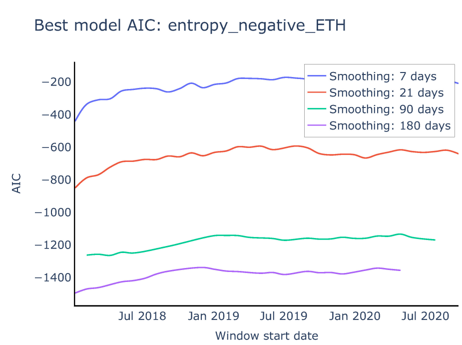
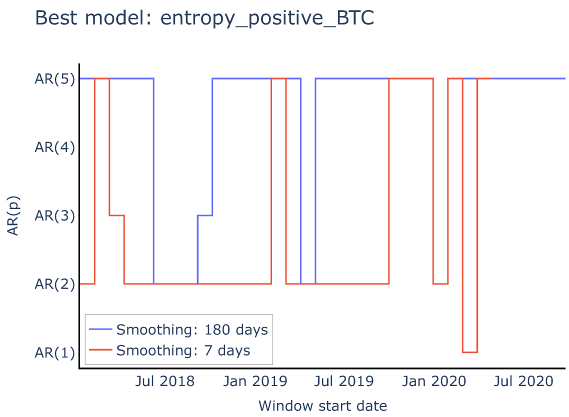
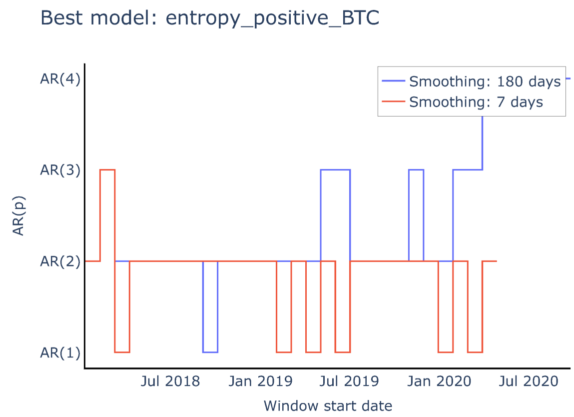


Figure 4: Stage I rolling window fits: best AR lag structure.

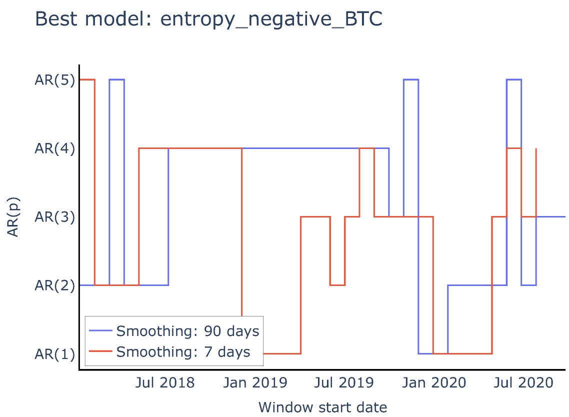
(a) BTC - positive sentiment



(b) BTC - positive sentiment - $\geq 90\%$ coef. significance



(c) BTC - negative sentiment



(d) BTC - negative sentiment - $\geq 90\%$ coef. significance

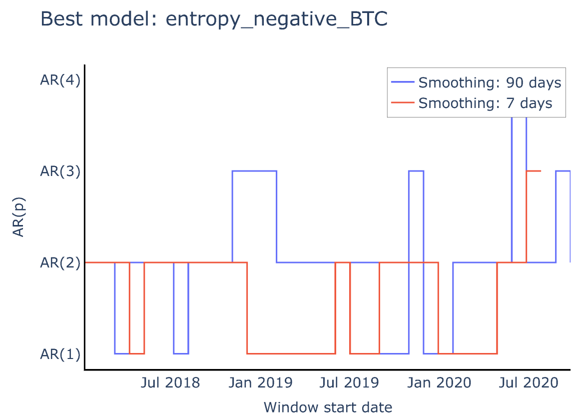
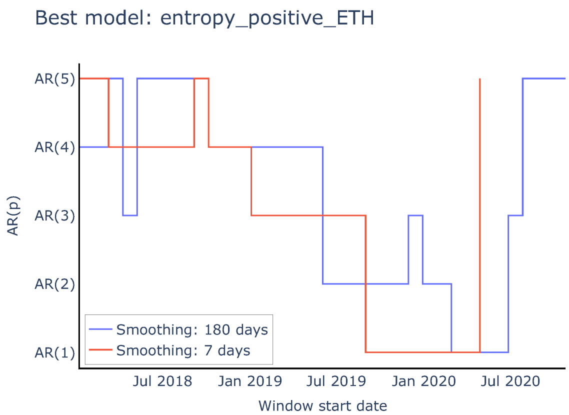
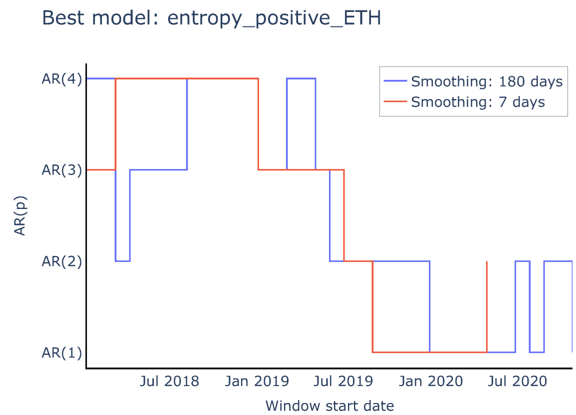


Figure 5: Stage I rolling window fits: best AR lag structure.

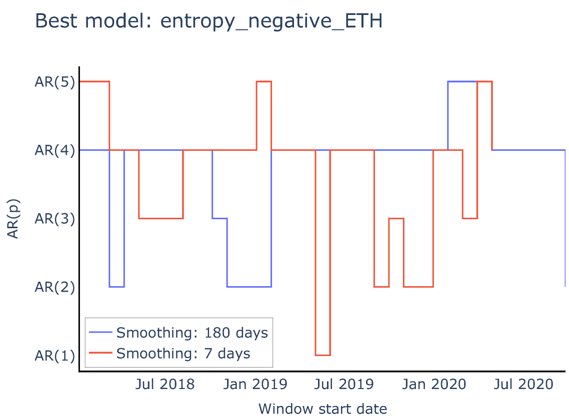
(a) ETH - positive sentiment



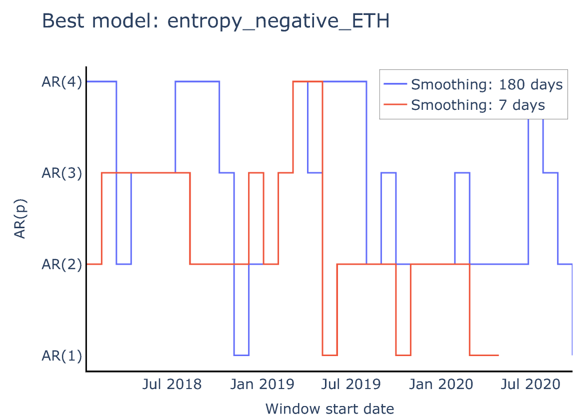
(b) ETH - positive sentiment - $\geq 90\%$ coef. significance



(c) ETH - negative sentiment



(d) ETH - negative sentiment - $\geq 90\%$ coef. significance



C.2 Stage II

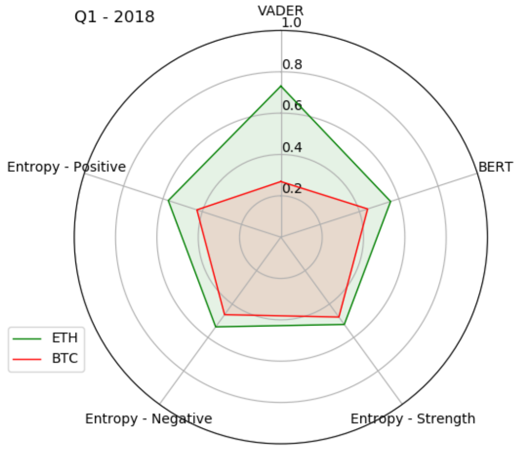
The second stage comprises the use of instrumental variables regression to remove the correlation between the error term and Y_{t-1}. To construct the instrumental variable we have a number of possible regression models which we discuss in detail in Section 6.1.1 of the main manuscript. After trying a range of regression models for the instrumental variable and the residuals of Eq. 6, E_t^y, we will choose the combination that provides both an instrumental variable that is independent from the error term, and a statistically significant fit to E_t^y.

C.2.1 Koyck transform Stage II: Qualitative evidence for the difference between the entropy index and the BERT and VADER indices using the robust median summary

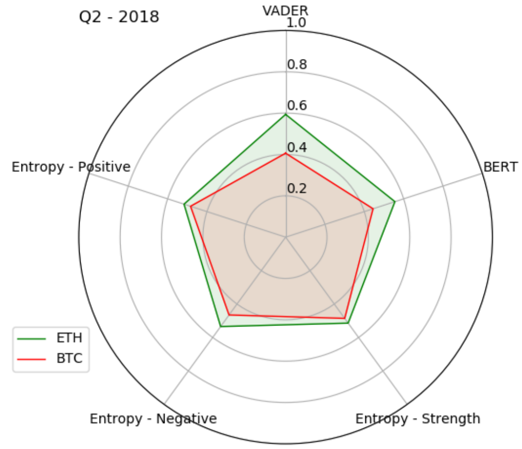
In addition to the IQR volatility-based summary radar plots of the main manuscript, to further illustrate the difference between the signals captured by the examined sentiment constructions we construct the plots of Figures 6 - 9 where we have summarised the sentiment content of the indices in yearly quarters. Figures 6 - 9 we have used the median as a robust summary of the sentiment signal for the corresponding quarters. Looking at Figure 7 (b) we can see that the positive entropy sentiment index exhibits a clear peak in its tendency, showing that it reacted to the sentiment change. However, from Figure 4a of the main manuscript, we observe that BERT captured negative sentiment at that time, whereas VADER captured only a low positive sentiment and mainly fluctuated around neutral.

Figure 6: Sentiment index median per quarter for each of the sentiment indices.

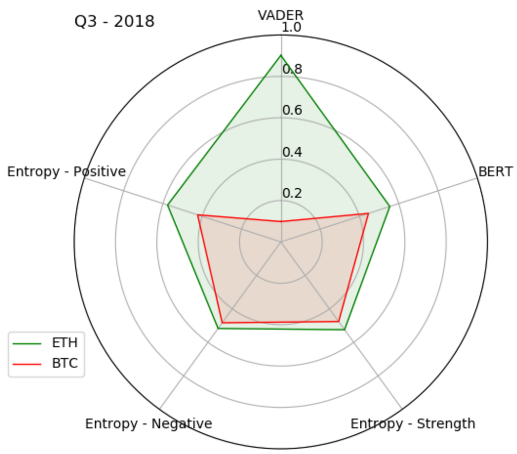
(a) Q1 - 2018



(b) Q2 - 2018



(c) Q3 - 2018



(d) Q4 - 2018

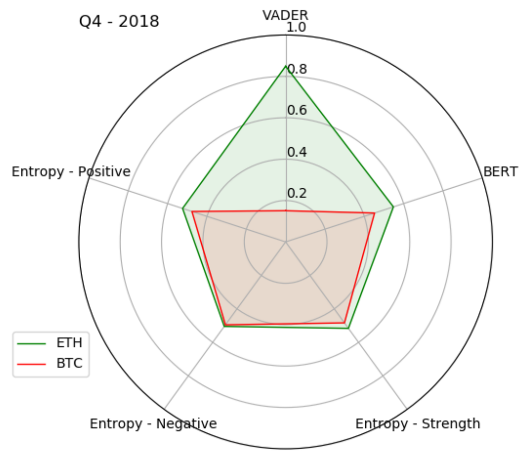
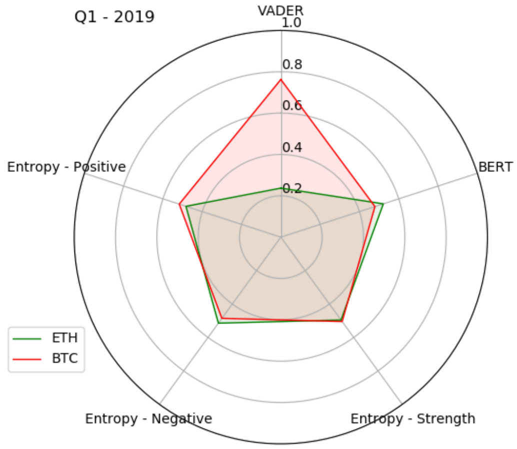
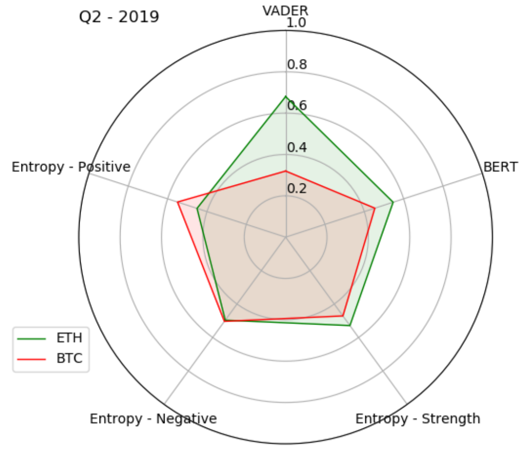


Figure 7: Sentiment index median per quarter for each of the sentiment indices.

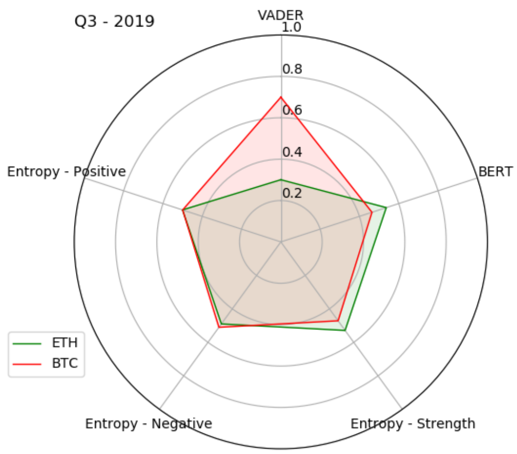
(a) Q1 - 2019



(b) Q2 - 2019



(c) Q3 - 2019



(d) Q4 - 2019

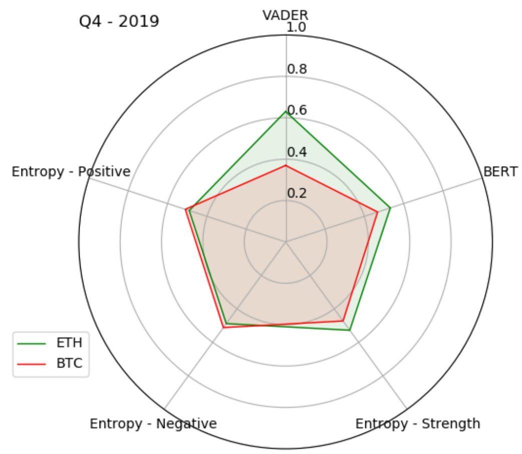
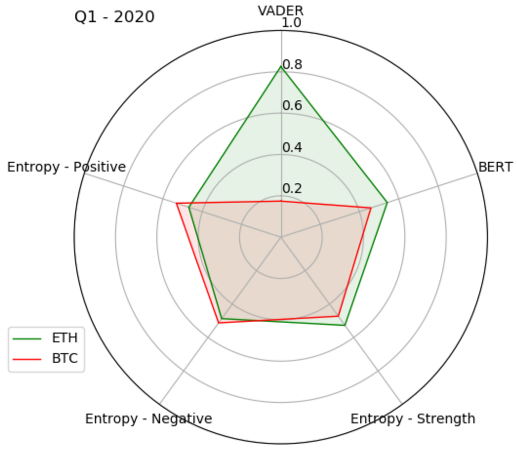
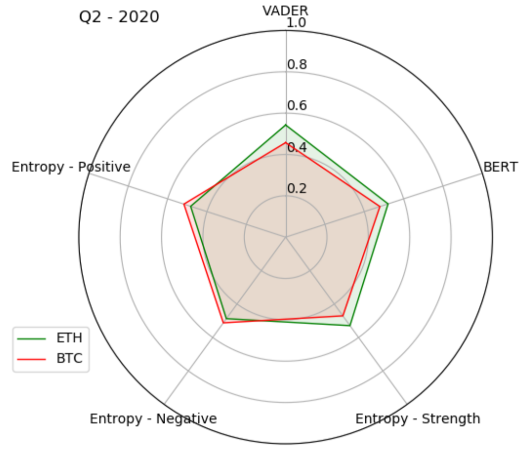


Figure 8: Sentiment index median per quarter for each of the sentiment indices.

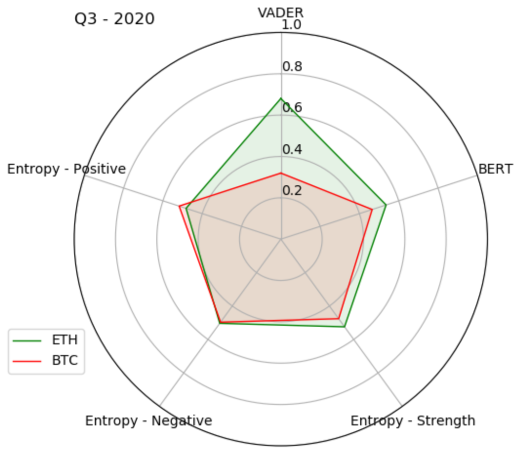
(a) Q1 - 2020



(b) Q2 - 2020



(c) Q3 - 2020



(d) Q4 - 2020

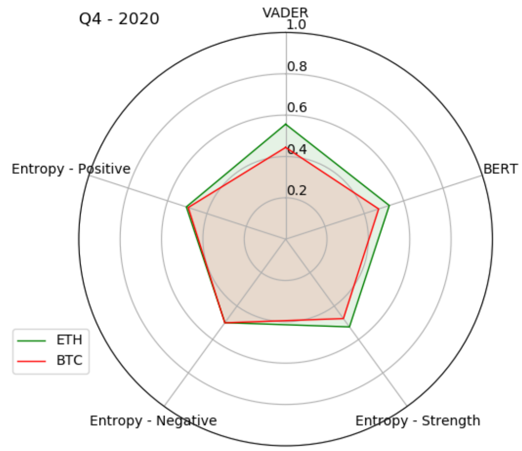
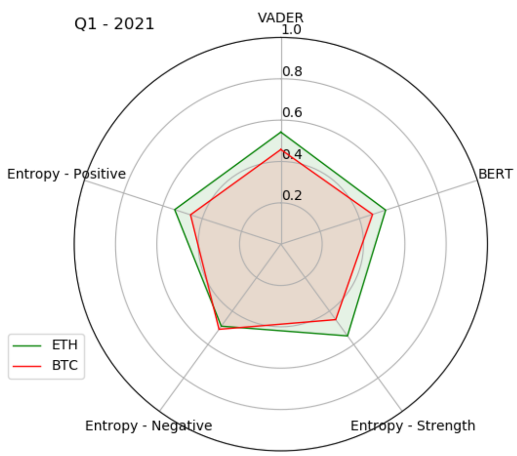
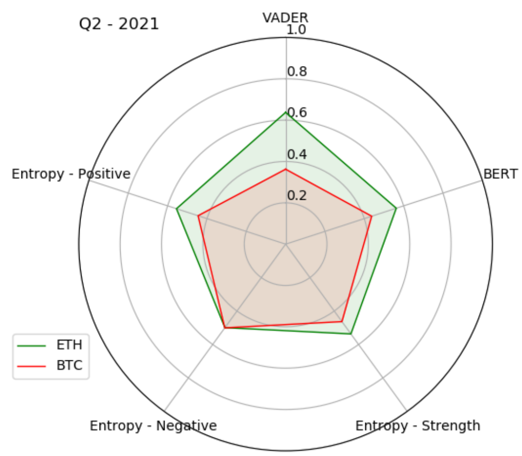


Figure 9: Sentiment index median per quarter for each of the sentiment indices.

(a) Q1 - 2021



(b) Q2 - 2021



D Assessment of regression model over time

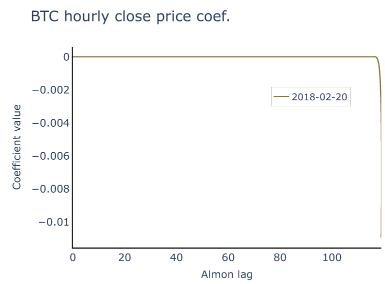
In this section, we present additional plots (Figures 10 - 11) for models M1 and M2 that illustrate the model evolution over time, via the coefficients of the Almon lags for the Bitcoin hourly price and the Hash Rate.

Figure 10: M2: BTC hourly close price - Almon coefficients structure over time. The black trace illustrates the coefficients for the model fit on the complete dataset. The red and green lines demonstrate the Almon coefficients from a rolling window fit. For clarity, we show the fits in separate panels. The highlighted entries in the legend show the start of the 1080-day window fit for the traces that are included in the plot. As the windows go into the past, the trace graduates from red to green colour. No Box-Cox transform was applied.

(a)

(b)

(c)

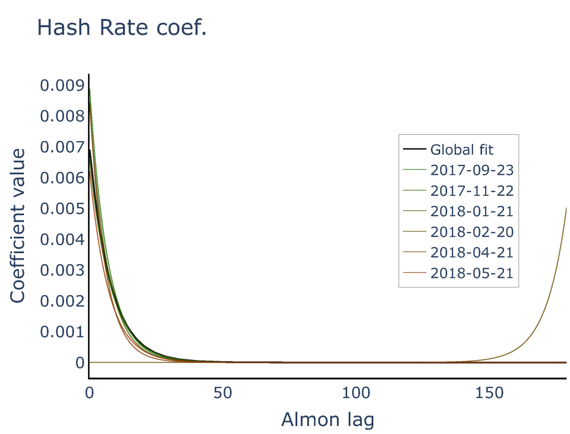


(d)

(e)

Figure 11: M2: Hash Rate - Almon coefficients structure over time. The black trace illustrates the coefficients for the model fit on the complete dataset. The red and green lines demonstrate the Almon coefficients from a rolling window fit. For clarity, we show the fits in separate panels. The highlighted entries in the legend show the start of the 1080-day window fit for the traces that are included in the plot. As the windows go into the past, the trace graduates from red to green colour. No Box-Cox transform was applied.

(a)

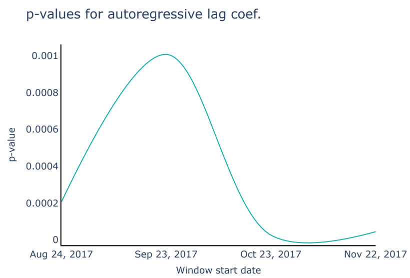


D.1 Statistical significance of interactions between data at mixed frequencies

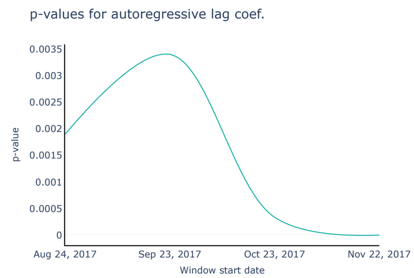
In this section we present further details to explore the statistical significance of the coefficients of the autoregressive covariate, as well as the low- and high-frequency covariates. In Figures 12 - 17 we plot the p-values only when they were statistically significant.

Figure 12: M1: Autoregressive covariate - statistically significant lags. No Box-Cox transform was applied.

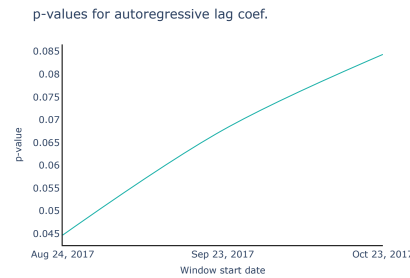
(a) lag 1



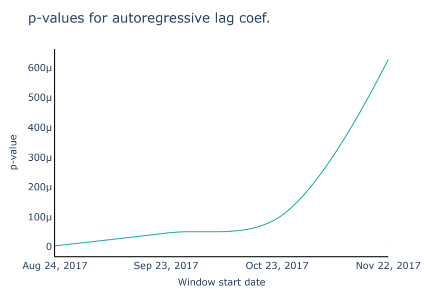
(b) lag 2



(c) lag 3



(d) lag 4



(e) lag 5

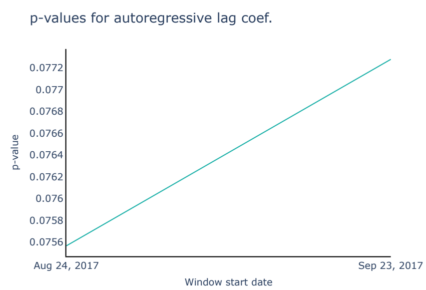
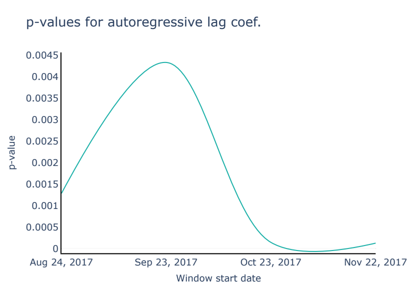
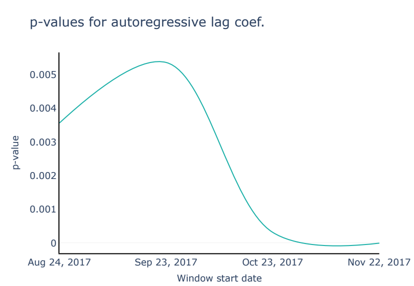


Figure 13: M1: Autoregressive covariate - statistically significant lags. The Box-Cox transform with $\lambda = 0$ was applied.

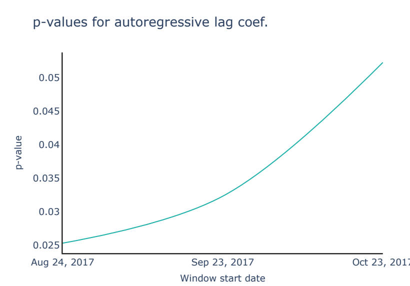
(a) lag 1



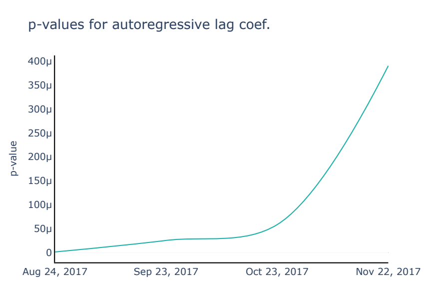
(b) lag 2



(c) lag 3



(d) lag 4



(e) lag 5

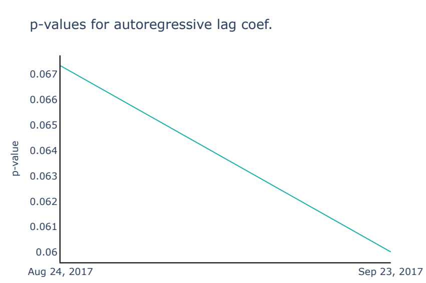
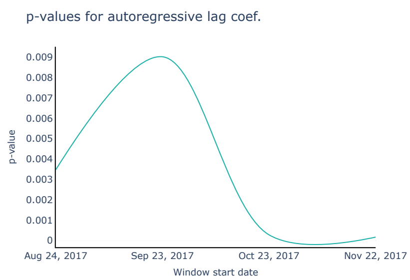
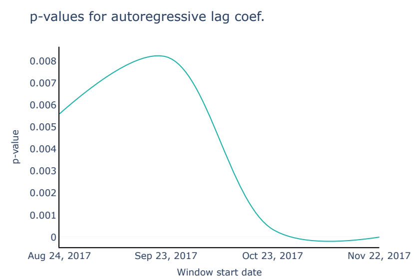


Figure 14: M1: Autoregressive covariate - statistically significant lags. The Box-Cox transform with $\lambda = -0.5$ was applied.

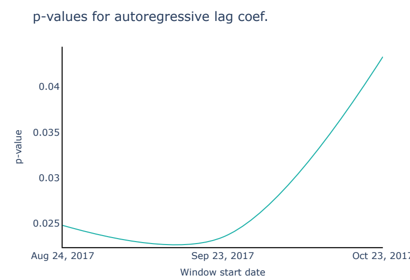
(a) lag 1



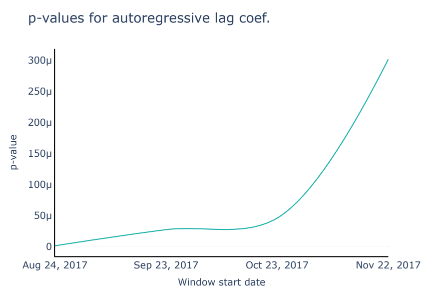
(b) lag 2



(c) lag 3



(d) lag 4



(e) lag 5

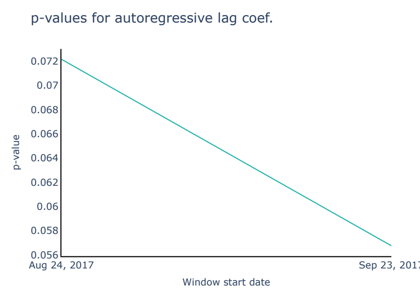
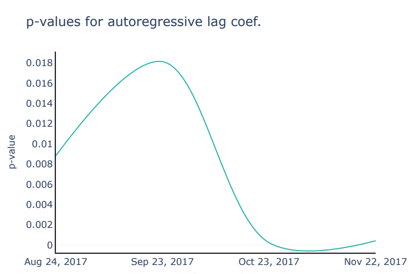
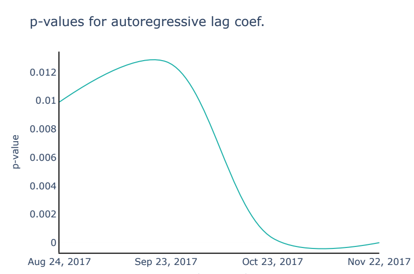


Figure 15: M1: Autoregressive covariate - statistically significant lags. The Box-Cox transform with $\lambda = -1$ was applied.

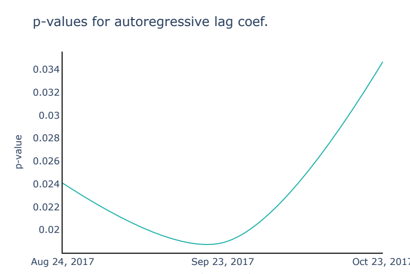
(a) lag 1



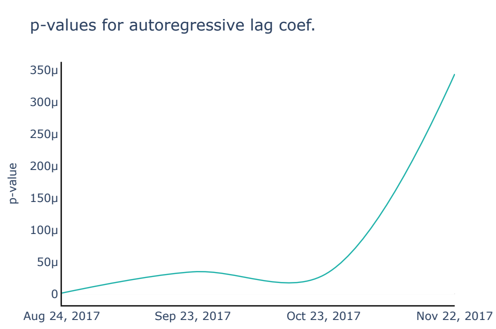
(b) lag 2



(c) lag 3



(d) lag 4



(e) lag 5

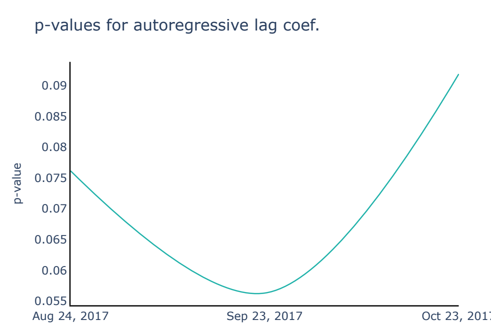


Figure 16: M2: Hash Rate low-freq. covariate - statistically significant Exponential Almon parameters (θ_1, ψ). No Box-Cox transform was applied.

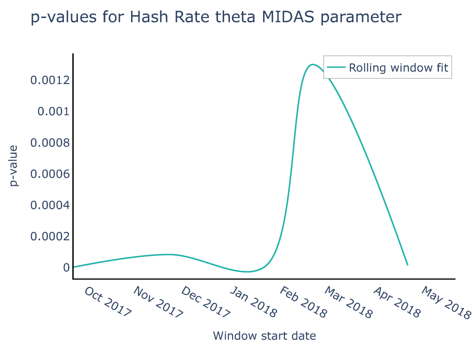
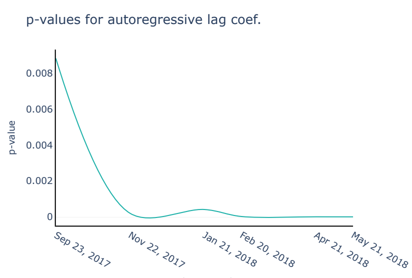
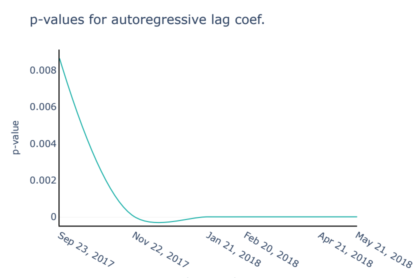


Figure 17: M2: Autoregressive covariate - statistically significant lags. No Box-Cox transform was applied.

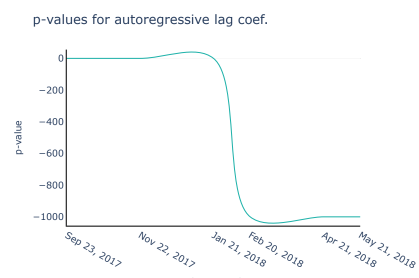
(a) lag 1



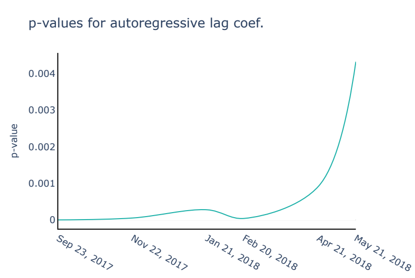
(b) lag 2



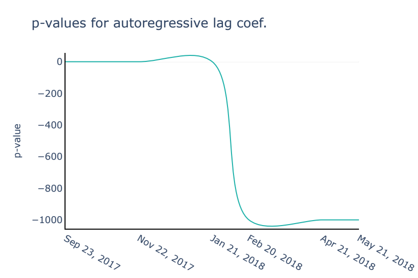
(c) lag 3



(d) lag 4



(e) lag 5

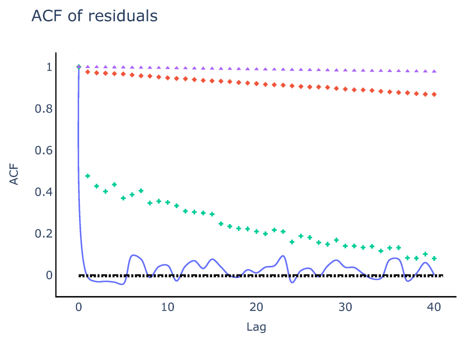


D.2 Mixed-data long memory structures - Autocorrelation functions

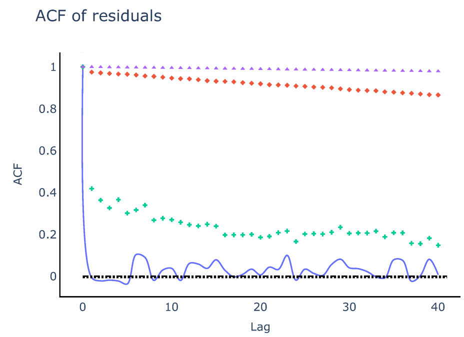
In the case of the long memory Gegenbauer polynomial MIDAS weights, we could estimate the d and u parameters from the residuals using the estimator of Section 5.3 in the main manuscript. The cyclic frequency u can be estimated by observing the autocorrelation plots of the residuals per window: note that the ACF exhibits almost no oscillation around the x axis, which means that $u = 0$, i.e. the long memory is coming from an ARFIMA-type process. We present the ACF plots for models M1 and M2 in Figures 18 - 22.

Figure 18: M1: ACF of residuals per fitting window. The Box-Cox transform with $\lambda = 0$ was applied.

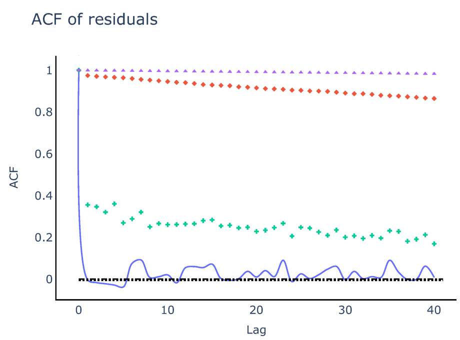
(a)



(b)



(c)



(d)

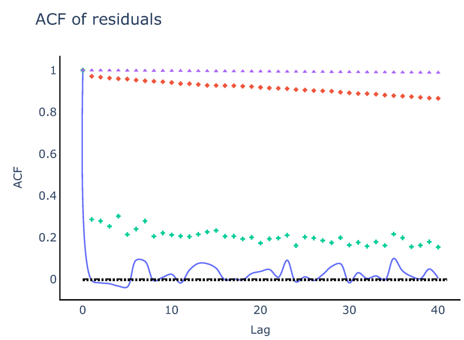
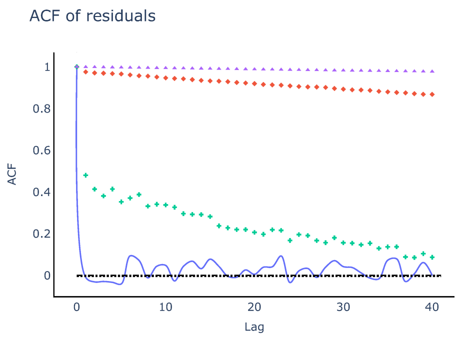
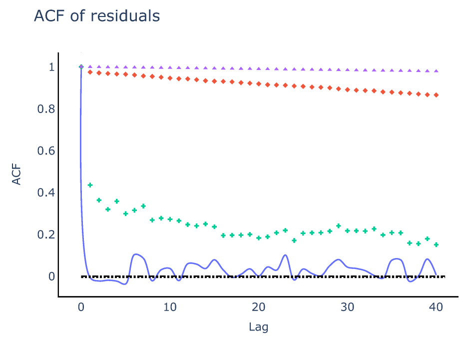


Figure 19: M1: ACF of residuals per fitting window. The Box-Cox transform with $\lambda = -0.5$ was applied.

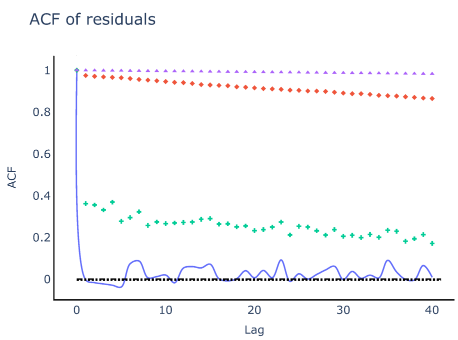
(a)



(b)



(c)



(d)

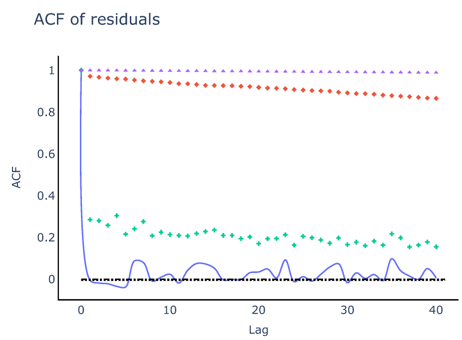
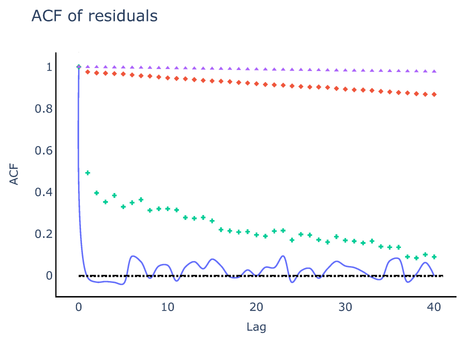
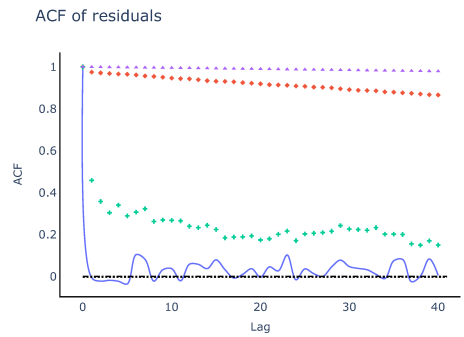


Figure 20: M1: ACF of residuals per fitting window. The Box-Cox transform with $\lambda = -1$ was applied.

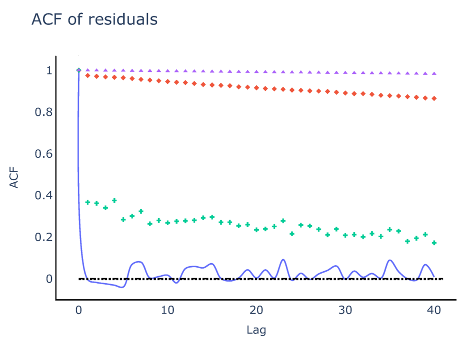
(a)



(b)



(c)



(d)

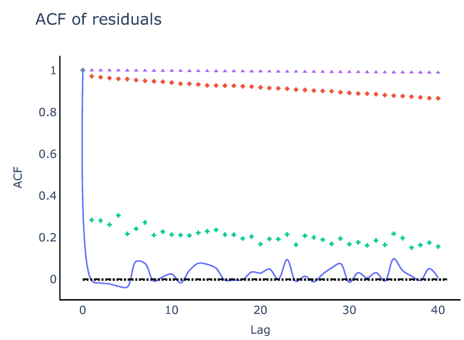
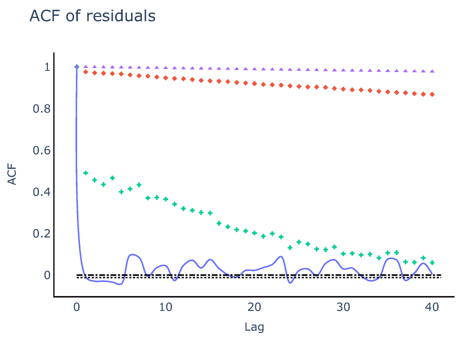
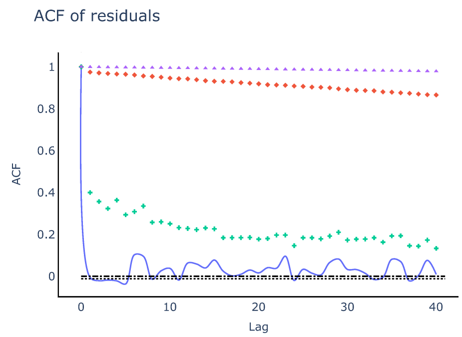


Figure 21: M1: ACF of residuals per fitting window. No Box-Cox transform was applied.

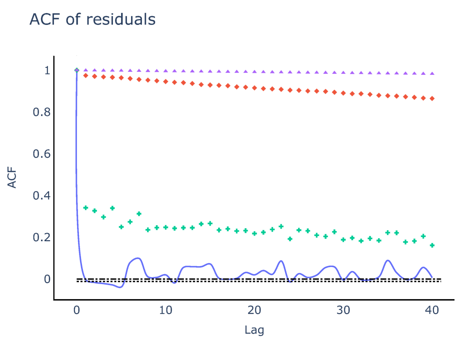
(a)



(b)



(c)



(d)

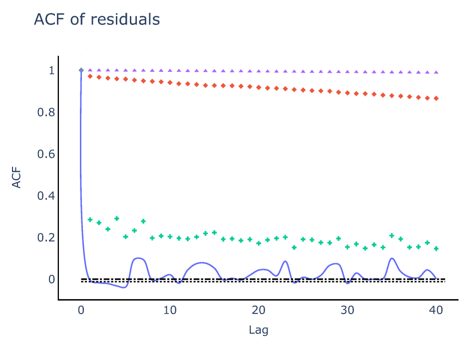
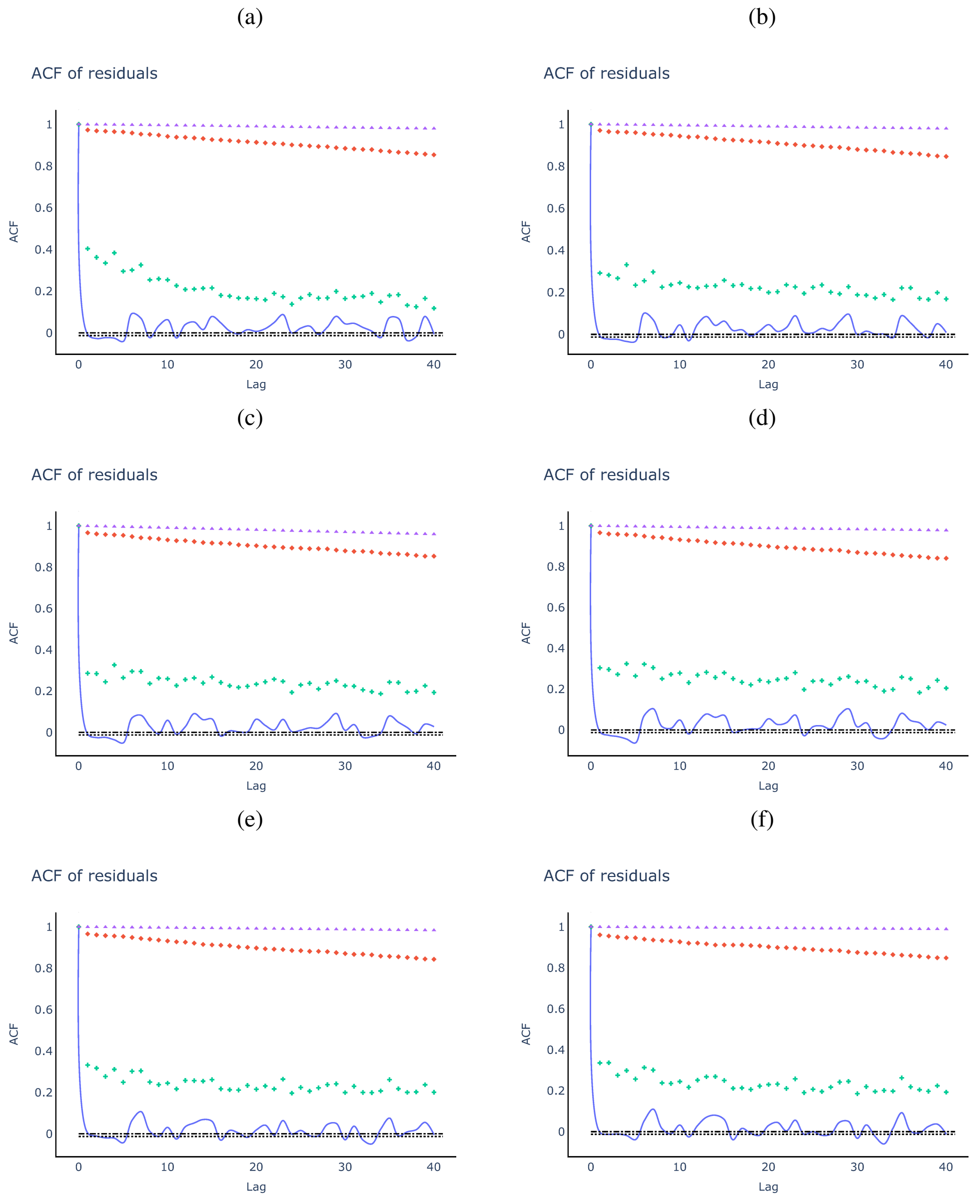


Figure 22: M2: ACF of residuals per fitting window. No Box-Cox transform was applied.



References

[1] H. Akaike, A new look at the statistical model identification, IEEE transactions on automatic control 19 (6) (1974) 716–723.
URL https://doi.org/10.1109/TAC.1974.1100705