Seminar: Topics in Machine Learning and Applied Econometrics

Seminar Paper

Rhine Water Level Forecasting

Philipp Georg Bößendörfer

Am Hardtberg 9

35423 Lich

Number of enrollment: 4006282

Teacher: Dr. Arne Warnke

Contents

1 Introduction

The general idea of the paper is to use relatively simple methods, i.e. methods with a high degree of clarity regarding the calculations and solving methods and compare them to more sophisticated estimation methods. For this reason and by the inspection of the time series itself, I restrict the estimation to Autoregressive (AR) models, Autoregressive Distributed Lag (ADL) models and Vector Autoregressive (VAR) models, respectively.
I use the Bayes Information criterion for choosing the lag length. To assess the forecasting accuracy among the different models, I use Pseudo-out-of-sample forecasts (Poos) and the corresponding Root Mean Squared Forecasting Error (RMSFE). I demonstrate that VAR models with the additional regressor of rainfall in the catchment area of the Rhine perform best, and even better than the more sophisticated ARIMA models.
The remainder of this paper is structured as follows. In the next section, I describe the test for stationarity and breaks and explain the conclusions I draw from them. In section 3, I compare different models in their forecasting performance. In section 4, I present the forecasting results. Section 5 summarizes and proposes extensions of the estimation.

2 Analysis of the Rhine water level time series

I start the analysis of the Rhine water level by plotting the time series. The plot is given in panel (a) of Figure 5.1. To get a sense of appropriate models, I plot the estimated autocorrelation function (ACF) and the estimated partial autocorrelation function (PACF).[1] The plots are illustrated in Figure 5.2. The ACF plot reveals a strong autocorrelation among the first lags. The partial autocorrelation is much weaker which indicates that a linear model might be a good fit. To further investigate the relationship between the lags, I plot the lagged values up to lag 4. The results are illustrated in Figure 5.3.
As a starting point I estimate Autoregressive models of order p ($AR(p)$)[2]:

$$level_t = \alpha_0 + \alpha_1 level_{t-1} + \alpha_2 level_{t-2} + ... + \alpha_p level_{t-p} + U_t \tag{1}$$

where p is the number of lags included, and $level_t$ describes the Rhine water level at time t, α_0 is defined as $\alpha_0 = (1 - \alpha_1 - \alpha_2 - \ldots - \alpha_p)\mu$. I use Ordinary Least Squares (OLS) to estimate the coefficients in equation (1).
For the decision of the lag length, I use the Bayes Information Criterion (BIC)[3], which is further defined in the Appendix. Table 5.1 summarizes the results for a lag length of up to ten. The model with $p = 4$ lags minimizes the Bayes Information criterion. The coefficients, the adjusted R^2 and the heteroscedastic robust standard errors for the $AR(4)$ model are reported in column 1 of table 5.2. As the inspection of the ACF and PACF already suggested, a simple linear model fits the data fairly good, with an adjusted R^2 of

[1] See for example Shumway and Stoffer (2017).
[2] See for example Stock and Watson (2020).
[3] See Stock and Watson (2020) p.594.

0.986 and a Residual Standard Error (SER) of 12.32.

2.1 Test for Stationarity - Augmented Dickey Fuller Test

To decide whether the assumption of stationarity is reasonable, I perform the Augmented Dickey Fuller (ADF) test for the AR(4) model. To obtain the appropriate test statistic, I transform the general $AR(p)$ model of equation (1) into the following form[4]:

$$\Delta level_t = \alpha_0 + \delta level_{t-1} + \gamma_1 \Delta level_{t-1} + ... + \gamma_{p-1} \Delta level_{t-p+1} + U_t \tag{2}$$

where $level_t$ indicates the Rhine water level at time point t, $level_{t-p}$ is the p-lagged value, U_t is the error term and Δ indicates the change of the Rhine water level, $\delta = \alpha_1 + \alpha_2 + ... + \alpha_p - 1$. Note: this transformation is only done for the computation of the test statistic, it does not change the model itself; for example, the BIC of both models is the same.

The test for stationary corresponds to the following hypothesis: $H_0: \delta = 0$, i.e. the time series is nonstationary against the $H_1: \delta < 0$, i.e. the time series is stationary.[5] The critical values for the test statistic do not follow the normal distribution, but a different distribution, which can be retrieved from Stock and Watson (2020). Table 5.3 summarizes the results for the model with $p = 4$ lags with and without a deterministic (linear) time trend. I reject the hypothesis of nonstationary for the model with and without a deterministic trend at the 1% level and conclude that the time series of the Rhine water level is stationary. Therefore, I do not transform the time series, e.g. by taking the first difference to account for seasonality. I account for the effects of seasonality in a later step by including rainfall as an additional predictor.

[4] See Stock and Watson (2020).

[5] It can be shown that this test refers to the Unit Root test, this relationship is well explained in Stock and Watson (2020).

2.2 Testing for Breaks – Quandt Likelihood (QLR) Test

To test for breaks in the relationship of the predictors and the predicted variable, I perform the Quandt Likelihood Ratio Test (QLR test)[6] for the inner 70% of the dataset. I.e., I test for breaks in the period from September 2016 to October 2019. In a second step, I test the model performance in an informal test via the Root Mean Squared Forecasting Error (RMSFE) for the last 15% of the observations to ensure that there is no break at the very end of the data.

The QLR test corresponds to the following model:

$$level_t = AR(p) + \gamma_0 D_t(\tau) + \gamma_1(D_t(\tau) level_{t-1}) + ... + \gamma_p(D_t(\tau) level_{t-p}) + U_t \qquad (3)$$

where $D_t(\tau) = 1$ if $t > \tau$, and $D_t(\tau) = 0$ otherwise and $AR(p)$ is described as in equation (1). I test for every possible date within the inner 70% of the observations the $H_0 : \gamma_0 = \gamma_1 = ... = \gamma_p = 0$ against the H_1 : at least one of the $\gamma_0, ..., \gamma_p$ is nonzero. For this test I calculate the ordinary F-test, however, I use heteroskedastic robust standard errors.

A time series plot for the examined period is given in panel (a) of Figure 5.4. The maximum value of 2.18 occurs in November 2018 and is smaller than the critical value of 3.26[7] at the 5% level, I thus do not reject the H_0 that there is no break.

To further ensure that there is no break at the very end of the time series, i.e. the part that is of interest in forecasting, I compare the Standard Error of Residuals (SER) for the regression of the first 85% of the observations (i.e., for the period from January 2016 to March 2019) with the RMSFE of the Pseudo-out-of-sample Forecasts[8] of the last 15% of the observations. The results are summarized in column 2 of Table 5.2. The SER is 12.35 and the RMSFE for the period from March 2019 to June 2020[9] is 12.23 (Table 5.4, column 1). I conclude that there is no break at the very end of the dataset.

[6] See Stock and Watson (2020) p.609.
[7] The critical values may be obtained in Stock and Watson (2020).
[8] Described in Stock and Watson (2020) p.613.
[9] i.e. the last 15% of observations.

3 Model Choice

To assess whether there is a significant correlation in the residuals of the AR(4) model, I start with plotting the ACF-function of the residuals up to lag 20. Inspecting the Bartlett bounds[10] in panel (a) of Figure 5.5 does not indicate whether the residuals are independent identical distributed (i.i.d.) and thus White Noise.

To test the hypothesis that the residuals are i.i.d., I perform the Ljung-Box-Test. The Ljung-Box-Test[11] refers to the following test-statistic:

$$Q_{LB} = n \sum_{j=1}^{p} \hat{\rho}_j \frac{n+2}{n-2} \quad \sim \quad \chi_p^2 \tag{4}$$

where n is the number of observations in the time series, $\hat{\rho}_j$ is the estimated autocorrelation up to lag j. The H_0, that the residuals are i.i.d. is rejected if the QLR test statistic is larger than the critical value, which follows a χ_p^2 distribution.

The test statistic for the QLR test is 32.376 (Table 5.2 column 1) which is larger than the critical value at the 5% level (31.41). I thus conclude that the residuals are not i.i.d. and that there are possible improvements in the model perfromance.

3.1 Autoregressive Distributed Lag (ADL) Model

The result from the QLR test is the motivation to include additional regressors in the $AR(4)$ model. As additional predictors, I use the rainfall in the catchment area of the Rhine.[12] The Autoregressive Distributed Lag (ADL) model with p lags in both predictors

[10] i.e. the 5% significance boundary that the autocorrelation for this lag is different from zero.

[11] See Brockwell and Davis (2016) p.31.

[12] Retrieved form the CDC Portal of "Deutsche Wetterdienst". A time series plot of the rainfall is provided in in panel (b) of Figure 5.1. Catchment area is plotted in Figure 5.7, panel (a).I use the

is described by:

$$level_t = \alpha_0 + \alpha_1 level_{t-1} + ... + \alpha_p level_{t-p} + \beta_1 rain_{t-1} + ... + \beta_p rain_{t-p} + U_t \qquad (5)$$

where $rain_t$ is the rainfall within the catchment area of the Rhine.

I, again, use the BIC criterion to decide for the lag length.[13] I restrict the model choice to models with the same lag length in both predictors, i.e. to $ADL(p, p)$ models. The BIC for the ADL model is summarized in Table 5.1. The ADL model with $p = 5$ lags minimizes the BIC criterion.

In panel (b) of Figure 5.5 the ACF for the residuals for the $ADL(5, 5)$ model is plotted; the results of the Ljung-Box-Test are summarized in column 3.1 of Table 5.2. The QLR statistic is 17.43, the critical value is 31.41, I thus do not reject the hypothesis of White Noise. This result is in line with the increase in the in-sample model performance: the adjusted R^2 increases from 0.986($AR(4)$) to 0.9905 ($ADL(5, 5)$). I perform the QLR test for the additional predictors of rainfall, a time series plot for the period from September 2016 to October 2019 is given in panel (b) of Figure 5.4. The maximum value is 1.92, which is smaller than the critical value at the 5% level. Thus, I do not reject the H_0 that there is no break.

3.2 Pseudo-out-of-sample Forecasts (Poos) and Root Mean Forecasting Error

To assess the model performance for out-of-sample values, I use Pseudo-out-of-sample forecasts (Poos) and calculate upon this the Root Mean Squared Forecasting Error (RMSFE). The idea of poos is to re-estimate a given model ever period up to $t = t_1 - 1, ..., n - 1$ and then compute the Pseudo-out-of sample forecast $\tilde{Y}_{t+1}$ at date $t + 1$ and compare this value with the actual value.[14] With the forecasted values the RMSFE can be computed

arithmetic mean of the measurement stations in the catchment area.

[13]For the definition see Appendix.

[14]See Stock and Watson (2020).

as:

$$RMSFE = \sqrt{\frac{1}{n}\sum_{t=t_1-1}^{n-1}(Y_{t+1} - \tilde{Y}_{t+1})^2} \tag{6}$$

I compute the RMSFE for up to ten step ahead forecasts, therefore, it is necessary to also estimate the rainfall. Thus, the ADL models become a Vector Autoregressive (VAR) model, where the rainfall and the water level are estimated simultaneously. The VAR(p) model is described by the following two equations[15]:

$$level_t = \alpha_0 + \alpha_1 level_{t-1} + ... + \alpha_p level_{t-p} + \beta_1 raint_{t-1} + ... + \beta_p rain_{t-p} + U_t \tag{7}$$

$$rain_t = \alpha_0 + \alpha_1 level_{t-1} + ... + \alpha_p level_{t-p} + \beta_1 raint_{t-1} + ... + \beta_p rain_{t-p} + U_t \tag{8}$$

Table 5.4 summarizes the results of the Pseudo-out-of-sample forecasting. According to the RMSFE, the $VAR(5)$ model performs best; the additional predictors of rainfall decrease the RMSFE substantially compared to the $AR(4)$ model. The more sophisticated ARIMA[16] model performs, according to this criterion, worse than the $AR(4)$ and $VAR(5)$ model.

[15]See Stock and Watson (2020)

[16]The model is estimated via the package "forecast" and the function auto.arima.

4 Forecasts

For the 10-step ahead forecasts I use the model that minimizes the RMSFE, i.e. the VAR(5) model. The forecasts are summarized in Table 5.5 and are illustrated in Figure 5.6. The shaded area illustrates the 99%, 95% and 90% forecast intervals. To construct the forecast intervals, I use the estimated RMSFE of section 3. As the increasing RMSFE of section 3 already revealed, the uncertainty surrounding the forecasts increases as the forecasting horizon increases. This effect is, besides the model uncertainty, also driven by the reliance on the forecasted rainfall in the catchment area. Compared to the forecasting of the Rhine water level the ADL(5,5) model for rainfall performs poorly in explaining the variation in rainfall, with an adjusted R^2 of 0.1367.

5 Conclusion and Extensions

I demonstrated that with relatively simple estimation methods it is possible to get an adequate estimation for the Rhine water level. Specifically, the forecasts for the first, second and third days work well, however, the uncertainty increases as the forecasting horizon increases. This is also due to the observation that in the VAR model, the rainfall must be forecasted which increases the forecasting uncertainty.
A natural extension of the model from a statistical point of view is to use a different model class for the forecast of the rainfall in the catchment area. It might further be worthwhile to account for the differing densities in the measurement grid. This might be done by Thiessen-Polygon technique or Isohyeten Method, which give areas with low densities a larger weight.
As a starting point of the analysis of the catchment area, I use different subsets of the default catchment area of the Rhine. The catchment areas are illustrated in Figure 5.7. Panel (a) illustrates the default catchment area for the Rhine , panel (b) is a reduced version of the catchment area, where the measurement stations downstream are excluded by hand and plot (c) is the catchment area without Bavaria. The RMSFE of the different catchment areas are summarizes in Columns 5-7 in Table 5.4. The default catchment area outperforms the other two catchment areas, with the lowest RMSFE for the six steps ahead forecast. This result indicates that it might be necessary to further model the reflow of the Rhine.
In general, it might be worthwhile to model the inflow and outflow of the water level and include further predictors that effect the water level, e.g. the evaporization rate or the phreatic level.

Tables and Figures in the main body

Table 5.1: Bayes Information Criterion for AR(p) and ADL(p,p) models

Number of lags (p)	AR(p) Model	ADL(p,p) Model
1	5.86	5.714
2	5.221	4.83
3	5.045	4.686
4	5.042	4.693
5	5.042	4.679
6	5.046	4.686
7	5.049	4.691
8	5.053	4.697
9	5.054	4.701
10	5.055	4.704

Table 5.2: Model Results

	[1] $level_t$ AR(4)		[2] $level_t$ AR (4) (85%)		[3] VAR(5) [3.1] $level_t$ ADL(5,5)		[3] VAR(5) [3.2] $rain_t$ ADL(5,5)	
	Coefficients	Standard Error	Coefficients	Standard Error	Coefficients	Standard Error	Coefficients	Standard Error
Constant	3.4932	(-0.664)	3.3535	(-0.7089)	-0.2471	(0.5774)	662.7167	(100.3823)
$level_{t-1}$	1.9776	(-0.0657)	1.9775	(-0.067)	1.7389	(0.0538)	-2.5414	(4.2237)
$level_{t-2}$	-1.4855	(0.1139)	-1.4982	(-0.1205)	-1.1252	(0.089)	5.3733	(7.7779)
$level_{t-3}$	0.5799	(0.0872)	0.6005	(-0.0928)	0.5043	(0.0759)	-7.168	(8.7406)
$level_{t-4}$	-0.0896	(0.0337)	-0.0969	(-0.0343)	-0.2147	(0.0544)	5.5781	(8.4092)
$level_{t-5}$					0.0715	(0.0219)	-0.8203	(3.9382)
$rain_{t-1}$					0.0018	(0.0002)	0.3568	(0.0373)
$rain_{t-2}$					0.0024	(0.0003)	0.0349	(0.0323)
$rain_{t-3}$					0.0006	(0.0003)	0.0051	(0.031)
$rain_{t-4}$					-0.0001	(0.0002)	-0.0019	(0.0302)
$rain_{t-5}$					-0.0008	(0.0002)	0.0484	(0.0293)
Adjusted R^2	0.986		0.986		0.9905		0.1367	
Residual Standard Error	12.32		12.35		10.17		1902	
Ljung Box Test-Statistic	32.376				17.432			

Coefficients are estimated via OLS. Column [2] is estimated with the first 85% of the observations (January 2016 – March 2019). LJung Box-Test Statistic with 20 lags for the residuals. Heteroskedastic robust standard errors.

Table 5.3: Augmented Dickey Fuller Test Results

	[1] $\Delta level_t$ ADF Model	[2] $\Delta level_t$ ADF Model (time trend)
$level_{t-1}$	-0.018	-0.018
	(-5.775)	(-5.831)
$\Delta level_{t-1}$	0.995	0.995
	(-40.393)	-40.384
$\Delta level_{t-2}$	-0.49	-0.49
	(-15.036)	(-15.028)
$\Delta level_{t-3}$	0.09	0.09
	(-3.605)	-3.615
t		-0.001
		(-0.816)
Intercept	3.493	3.996
	(-5.127)	(-4.35)
Residual Standard Error	12.32	12.32

Models are estimated via OLS. Ordinary t-test statistic in parentheses (no heteroskedastic standard errors). ADF critical values at 1% level: -3.43; -3.96 (with time trend)

Table 5.4: RMSFE for different forecasting horizons

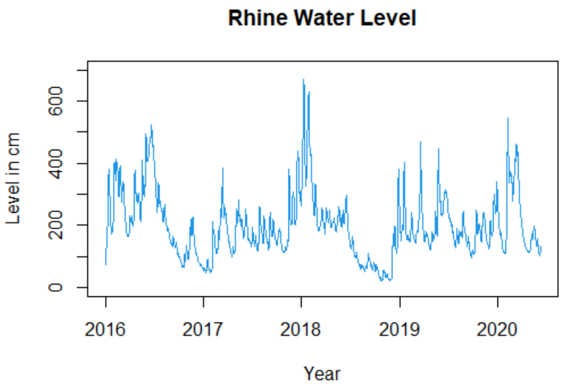
	[1] AR(4) (15%)	[2] AR(4)	[3] VAR(5)	[4] Auto ARIMA	[5] VAR(5), $catchment_{(a)}$	[6] VAR(5), $catchment_{(b)}$	[7] VAR(5), $catchment_{(c)}$
$RMSFE_{h=1}$	12.226	12.889	9.765	13.157	9.765	10.855	10.223
$RMSFE_{h=2}$	-	29.491	19.749	30.568	19.749	23.057	21.389
$RMSFE_{h=3}$	-	44.715	30.066	47.04	30.066	34.369	32.365
$RMSFE_{h=4}$	-	57.076	42.494	60.875	42.494	45.613	44.227
$RMSFE_{h=5}$	-	66.606	54.759	72.009	54.759	56.359	55.861
$RMSFE_{h=6}$	-	73.42	63.821	80.468	63.821	64.378	64.494
$RMSFE_{h=7}$	-	78.528	70.576	87.149	70.576	70.56	70.985
$RMSFE_{h=8}$	-	82.82	75.65	92.832	75.65	75.384	75.9
$RMSFE_{h=9}$	-	86.99	80.281	98.15	80.281	79.851	80.432
$RMSFE_{h=10}$	-	90.916	84.923	103.038	84.923	84.301	84.965

RMSFE calculated via Poos for the period December 2019 – June 2020. Column [1] refers to the last 15% of observation. Catchment area in columns [5] – [7] refers to panels (a) –(c) of Figure 5.7. Results for Column 4 are obtained via the package "forecast", auto.arima function.

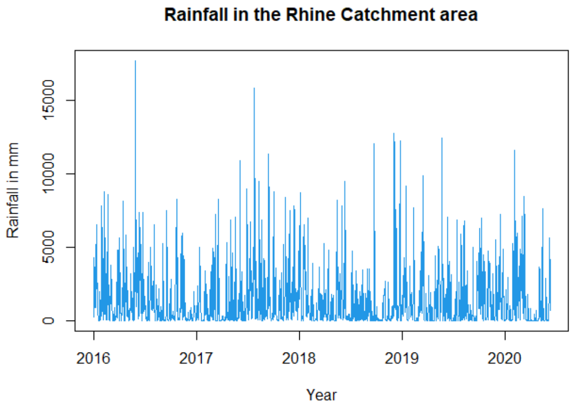
Table 5.5: 10-step ahead forecasts

Date	08.06.2020	09.06.2020	10.06.2020	11.06.2020	12.06.2020	13.06.2020	14.06.2020	15.06.2020	16.06.2020	17.06.2020
Water Level Forecast	146.3277	149.9686	145.7078	142.4284	143.3195	146.008	148.7648	151.2359	153.4262	155.3787

Figure 5.1: Time Series Plots (January 2016 – June 2020)

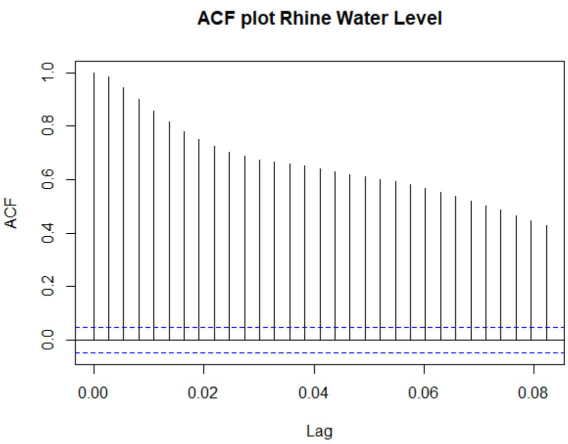


(a) Rhine Water Level

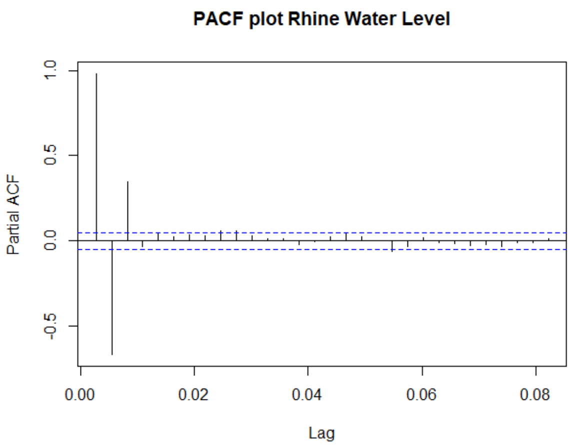


(b) Rainfall in the Catchment Area

Figure 5.2: ACF and PACF plots of the Rhine Water Level



(a) Autocorrelation Function up to lag 30 of the Rhine Water Level



(b) Partial Autocorrelation Function up to lag 30 of the Rhine Water Level

Figure 5.3: Rhine Water Level lag plots up to lag 4

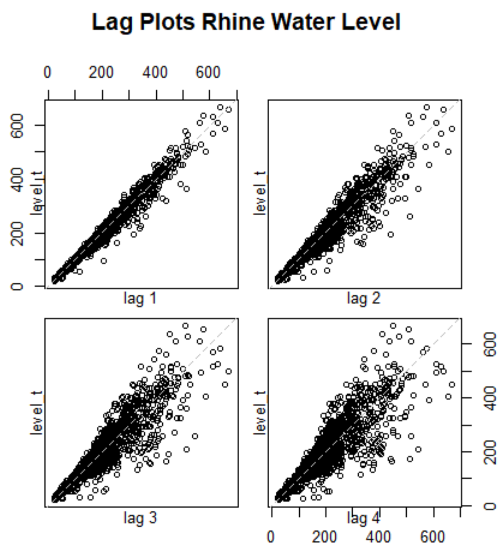
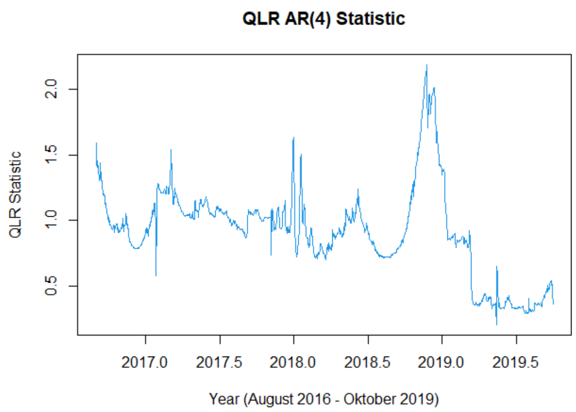
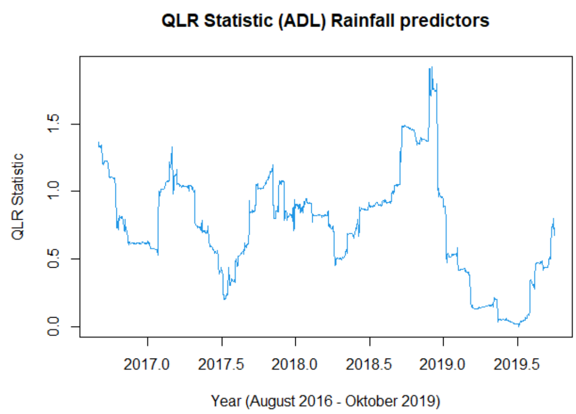


Figure 5.4: Quandt Likelihood Ratio Plots (September 2016 – October 2019)

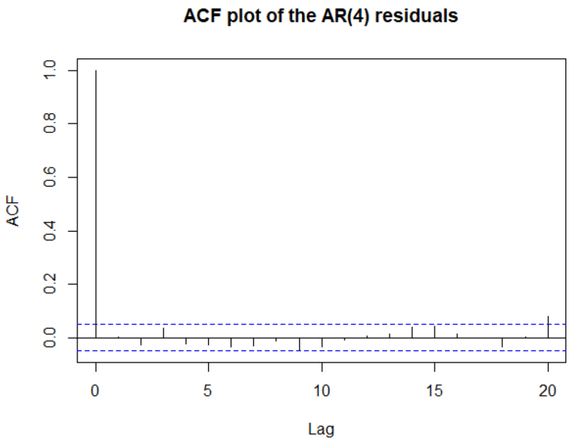


(a) Quandt Likelihood Plot for AR(4) model

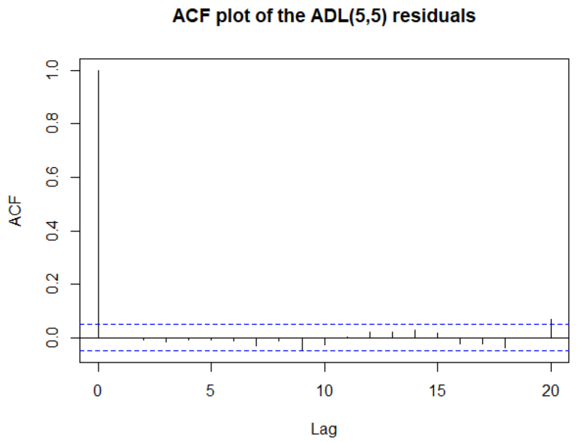


(b) Quandt Likelihood Plot for rainfall predictors in the ADL(5,5) model

Figure 5.5: ACF and PACF of the Residuals of AR(4) and ADL(5,5) models

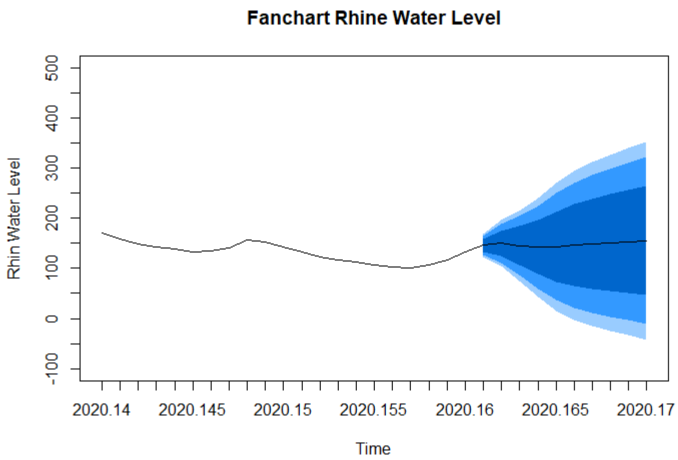


(a) Autocorrelation Function of the AR(4) model residuals up to lag 20



(b) Partial Autocorrelation Function of the ADL(5,5) model residuals up to lag 20

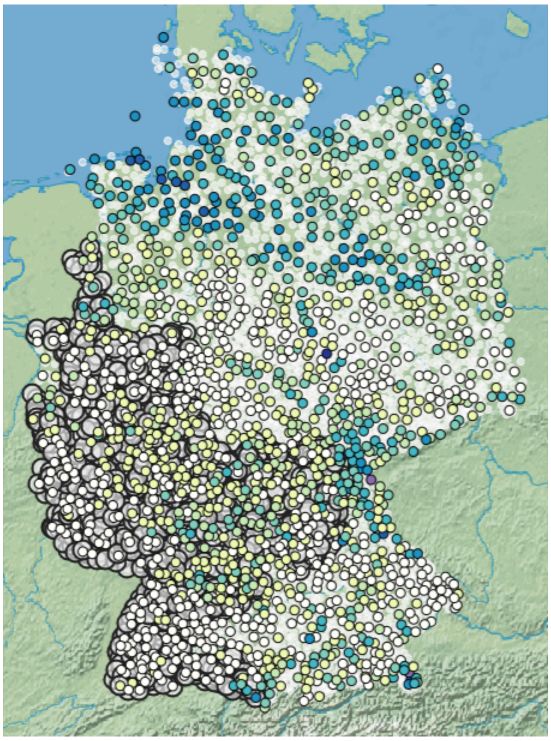
Figure 5.6: Fan Chart of the Rhine Water Level Forecasts



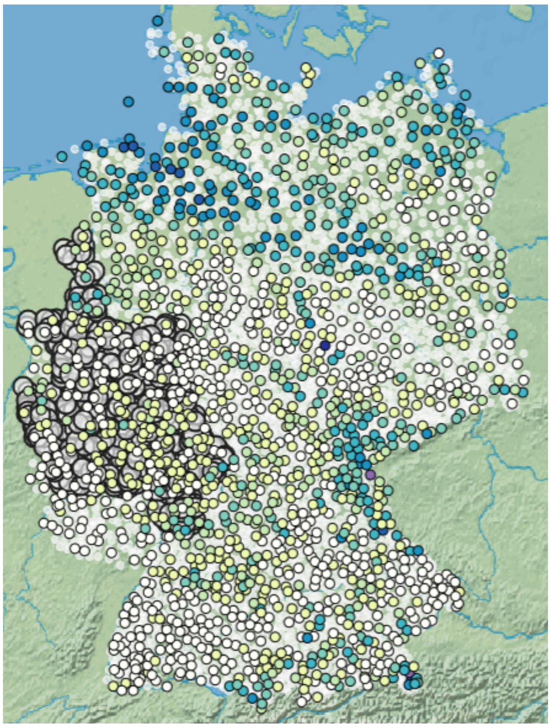
Forecasts are obtained via VAR(5) model. The shaded areas indicate the 90%, 95% and 99% forecasting intervals where the RMSFE of Table 5.4 was used to construct the intervals.

Figure 5.7: Different Catchment Areas

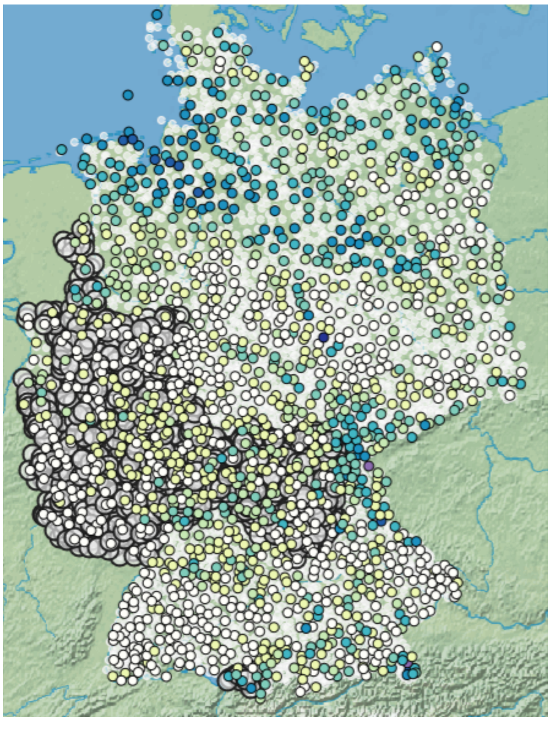
This Figure illustrates the different catchment areas used for the estimation in columns 5 -7 in Table 5.4. The data is retrieved from the cdc portal of the Deutsche Wetterdienst,https://cdc.dwd.de/portal/201912031600/mapview, as of 14.06.2020.



(a) Complete Catchment Area of the Rhine



(b) Catchment Area of the Rhine, excluding downstream stations



(c) Catchment area of the Rhine, excluding measurement stations in Bavaria

A. Appendix

The Bayes Information Criterion for AR(p) models is described by the following equation:

$$BIC(p) = ln(\frac{SRR(p)}{n}) + (p+1)\frac{ln(n)}{n} \tag{9}$$

where $SRR = \sum_{i=1}^{n} \hat{U}_i^2$. The BIC(p) trades of bias and variance to determine a best value of p for the model.[17]

The Bayes Information Criterion for ADL(p) models is described by the following equation:

$$BIC(K) = ln(\frac{SRR(p)}{n}) + K\frac{ln(n)}{n} \tag{10}$$

[17] Stock and Watson (2020).

Bibliography

Brockwell, P. J. and Davis, R. A. (2016). *Introduction to time series and forecasting.* Springer, Cham, Switzerland.

Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications - With Examples in R.* Springer, Cham, Switzerland.

Stock, J. H. and Watson, M. W. (2020). *Introduction to econometrics.* Pearson, Harlow, England.