University of Duisburg-Essen
Faculty of Business Administration and Economics
Chair of Econometrics

Analysing FIFA Data with the Bayesian LASSO

Seminar in Econometrics

Term Paper

Submitted to the Faculty of
Business Administration and Economics
at the
University of Duisburg-Essen

from:

Jens Klenke

Reviewer:	Christoph Hanck
Deadline:	Jan. 17th 2020

Name:	Jens Klenke
Matriculation Number:	3071594
E-Mail:	jens.klenke@stud.uni-due.de
Study Path:	M.Sc. Economics
Semester:	5^{th}
Graduation (est.):	Winter Term 2020

Contents

List of Figures

List of Tables

List of Abbreviations

1 Introduction

In recent years, the LASSO method by Tibshirani (1996) has emerged as an alternative to ordinary least squares estimation. The success of the method is mainly due to its ability to perform both variable selection and estimation. As Tibshirani already pointed out in his original paper the standard LASSO model can be interpreted as a linear regression with a Laplace prior. Park and Casella (2008) were the first to implement the Bayesian LASSO »using a conditional Laplace prior specification«.

Our goal is to compare the results of the Bayesian LASSO with the normal LASSO method and an ordinary least square estimation. The focus is particularly on the number of non-significant parameters in the linear model or, in case of the LASSOs the parameters equal to zero. To compare the different methods we will use the FIFA data sets from 2019 and 2020.

Bayesian statistics has grown very strongly in recent years. This is due to the improvement of computers which have made the calculation possible in the first place. Now they make it possible to calculate these methods in increasing numbers and in even shorter time. In view of these changes, it seems logical that the Bayesian LASSO should be used more and more often, since this method can select variables and also estimate variables by means of distributions. This leads to the fact that the advantages of LASSO and Bayesian statistics come together. Nevertheless, the Bayesian methods still take much longer to calculate than the frequentist methods. Therefore, in this paper we want to analyze whether the Bayesian LASSO has an advantage over the frequentist LASSO.

In the first chapter we will present the general approach of Bayesian statistics. Afterwards, we will describe the data basis and preparation. The next step is to introduce the methods used and to present the results. Before we will draw a conclusion, the models will be analyzed with respect to their residuals.

2 Theory of Bayesian inference

The Bayesian (inference) statistics based on the Bayes' theorem for events.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.1}$$

For Bayesian statistics the event theorem (2.1) gets rewritten to apply it to densities.

Where $\pi(\theta)$ is the prior distribution - which could be gained from prior research or knowledge, $f(y|\theta)$ is the likelihood function, and $\pi(\theta|y)$ is the posterior distribution, which then yields in the following equation:

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f(y)} \tag{2.2}$$

From equation (2.2) the advantages and disadvantages of Bayesian statistics compared to frequentist statistics can directly be retrieved. One major advantage is that the Bayesian approach can account for prior knowledge and points out a philosophical difference to the frequentist approach - that the obtained data stands not alone.

Another key difference and advantage is that in the Bayesian statistic the computations are made with distributions and this leads to a better information level than just the computation of the first and second moment.

The computation with distributions is also the greatest disadvantage or - more neutral - the biggest problem of the Bayesian approach because in high dimensional problems the computation takes a lot of time and in some cases it is even impossible. A solution to that is that with newer and better computers it is possible to simulate the integrals with a Markov chain Monte Carlo (MCMC) method. (Ghosh et al., 2006, p. 35 ff.)

3 Data description

We collected the data from the online database platform *kaggel.* The dataset includes 6 years of data for all players who were included in the soccer simulation game *FIFA* from *EA Sports*. We decided to keep the data for 2019 and 2020, only. The Data for 2019 contains 17538 datapoints which will be used for the estimation of the different models whereas the 2020 data with 18028 will be used to compare the quality of the models with an out of sample RMSE. Both datasets consist of 104 variables which will not all be included in the estimations. Some Variables are just an ID or different length of names and URLs. (Leone, 2020)

A fundamental problem of the dataset consists as goalkeepers are systematically rated differently than field players. Therefore, in the subcategories of the variable *overall* all field player categories were assigned NAs for goalkeepers. Conversely, all field players have NAs in all goalkeeper categories. Because the algorithm of LASSO in R cannot handle NAs they have been set to zero for all models.

It is not very realistic that a fielder has no values in the goalkeeper categories and vice versa. However, it can be argued, at least for outfield players, that goalkeeper attributes play no role in determining market values. This argumentation does not seem to hold for goalkeepers, at least passing can be assumed to be an influential variable for the market value, because is an essential asset for the passing game if the goalkeeper has possession of the ball. Nevertheless, due to the lack of alternatives, all NAs have been replaced by Zero.

Table 1: Summary of some important variables for the 2019 FIFA edition

	year	N	mean	sd
value_eur	2019	17 538	2 473 043.68	5 674 963.22
	2020	18 028	2 518 484.58	5 616 359.21
wage_eur	2019	17 538	10 085.87	22 448.70
	2020	18 028	9 584.81	21 470.29
overall	2019	17 538	66.23	7.01
	2020	18 028	66.21	6.95
age	2019	17 538	25.17	4.64
	2020	18 028	25.23	4.63
potential	2019	17 538	71.40	6.15
	2020	18 028	71.56	6.14

As one can see in Table 1 the differences between the editions for the most

important variables are considerably small.

For example, from 2019 to 2020 the mean player *value* (response variable) increased by 4.54e+04 which is about 1.8 per cent or 0.01 standard deviations. Similar results are observable for the probably most important righthand variables *wage* and *overall* with a difference in the means of -0.02 and -0.003 standard deviations between 2019 and 2020.

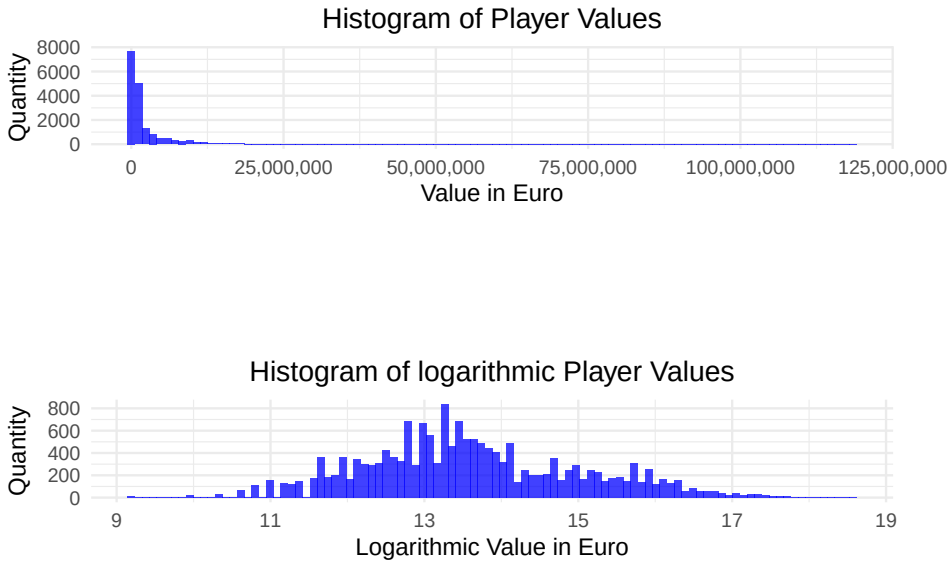

Figure 1: Histograms of player values and log player values

As can be seen for the variable *value* in Figure 1, this relatively strong right-skew is distributed, a similar pattern can be observed for the variable wage. Since we also have estimate a linear model, and this often leads to non-normally distributed residuals, these were logarithmized.

4 Used Models

To compare the Bayesian LASSO we will also analyze the data with a linear multivariate model, and the frequentist LASSO. We will start with the linear model and then gradually modify the model equations respectively the condition for estimating the parameters. So that the coherences between the individual methods become clear.

All three methods have the common assumption that the relationship is linear, at least in the parameters. Inatially, the assumption seems stricter than it is, because the data can be manipulated in such a way that the relationship is linear after all. In our data this was done by logarithmization.

4.1 Linear Model

The frequentist multivariate regression model has the following model equation.

$$\boldsymbol{Y} = \boldsymbol{\beta_0} + \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{4.1}$$

Where $\boldsymbol{y}$ is the $n \times 1$ response vector, $\boldsymbol{X}$ is the $n \times p$ matrix of regressors and, $\boldsymbol{\epsilon}$ is the $n \times 1$ vector of independent and identically distributed (i.i.d.) errors with mean 0 and unknown but constant variance σ^2.

The coefficient will be estimated by the ordinary least square method, which means that $\boldsymbol{\beta}$ should be chosen so that the *Euclidean norm* ($||\mathbf{y} - \mathbf{X}\beta||_2$) is minimal. This yields in the condition for the estimation of coefficients:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{\beta_0} - \boldsymbol{X\beta})^T (\boldsymbol{y} - \boldsymbol{\beta_0} - \boldsymbol{X\beta}) \tag{4.2}$$

4.2 Least Absolute Shrinkage and Selection Operator (LASSO)

In the LASSO method the model equation is the same as the equation for the multivariate but the condition for the optimization of the estimators in equation (4.2) has an additional punishment term. Which leads to the following optimization.

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda \sum_{i=1}^{p} |\beta_j| \tag{4.3}$$

for some $\lambda \geq 0$. This method is also often referred to as L_1 -penalized least squares estimation.

Already in his original paper Tibshirani (1996) has pointed out the possibility that his methods can also be interpreted in a Bayesian way. The LASSO estimates can be considered as posterior mode estimates with a double-exponential Laplace prior.

4.3 Bayesian Lasso

Park and Casella (2008) considered a fully Bayesian approach using a conditional Laplace prior of the form

$$\pi\left(\boldsymbol{\beta}|\sigma^2\right) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^e}} e^{\frac{-\lambda|\beta_j|}{\sqrt{\sigma^2}}} \tag{4.4}$$

The analysis of the FIFA data set is a multidimensional problem and therefore in Bayesian statistics this is an analysis that only works with a hierarchical model. The solution cannot be calculated directly but has to be solved with the Gibbs sampler.

4.3.1 Gibbs Sampler

The Gibbs Sampler is a special case of an MCMC algorithm, which is useful to approximate the combined distribution of two or more regressors in a multidimensional problem.

The algorithm tries to find the approximate joint distribution and therefore the algorithm runs through the sub-vectors of β and draws each subset conditional on all other values. (Gelman, 2004)

In the **monomvn** package in **R** (Gramacy, 2019) the Gibbs sampler for the Bayesian LASSO samples from the following representation of the Laplace distribution. Andrews and Mallows (1974)

$$\frac{a}{2}e^{-a|z|} = \int_0^\infty \frac{1}{2\sqrt{\sigma^2}} e^{-z^2/(2s)} \, \frac{a^2}{2} e^{-a^2 s/2} ds, \qquad a > 0 \tag{4.5}$$

4.3.2 The full Model specification

The full model has the following hierarchical representation

$$\begin{aligned}
\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2 &\sim N_n(\mu \mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n) \\
\boldsymbol{\beta}|\sigma^2, \tau_1^2, \ldots, \tau_p^2 &\sim N_p(\boldsymbol{A}^{-1}\boldsymbol{X}^T\tilde{\boldsymbol{y}}, \sigma^2\boldsymbol{A}^{-1}) \qquad \text{with } \boldsymbol{A} = \boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{D}_\tau^{-1} \\
\boldsymbol{D}_\tau &= diag(\tau_1^2, \ldots, \tau_p^2) \\
\sigma^2, \tau_1^2, \ldots, \tau_p^2 &\sim \pi\left(\sigma^2\right) d\sigma^2 \prod_{j=1}^{p} \frac{\lambda^2}{2} e^{-\lambda^2\tau_j^2/2} d\tau_j^2 \\
\sigma^2, \tau_1^2, \ldots, \tau_p^2 &> 0
\end{aligned}$$

If $\tau_1^2, \ldots, \tau_p^2$ gets integrated out of the conditional prior on $\boldsymbol{\beta}$, we get the form of (4.4). For σ^2 the inverse-gamma function of the form $\pi\left(\sigma^2\right) = \dfrac{1}{\sigma^2}$ was implemented in the **monomvn** package.

5 Estimation and Results of the Models

To compare the performances of the models all three models got, obviously, estimated with the same regressors. We included as righthand variables: *log_wage* , *age*, *height_cm*, *weight_kg*, *overall*, *potential*, *shooting*, *contract_valid_until*, *pace*, *passing*, *dribbling*, and *defending*, so we have 12 explanatory variables to predict the response variable *log_value*.

The reason we chose relatively few variables is that, on the one hand, we have enough variables, so it is very likely that some variables will not become significant, but on the other hand, we also have a better overview of the variables.

Of course, this can lead to biases in the parameters, but the aim of the work is not to provide a causal interpretation of the explanatory variables, but rather to compare the methods.

5.1 Linear Model

Table 2: Summary of the linear model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-8.5970	2.9222	-2.9420	0.0033
log_wage	0.0679	0.0025	26.8466	0.0000
age	-0.1004	0.0008	-119.6665	0.0000
height_cm	0.0012	0.0004	2.7140	0.0067
weight_kg	0.0001	0.0004	0.3175	0.7508
overall	0.2098	0.0008	266.5231	0.0000
potential	-0.0059	0.0007	-8.1962	0.0000
shooting	0.0049	0.0003	17.7691	0.0000
contract_valid_until	0.0051	0.0014	3.5164	0.0004
pace	0.0008	0.0002	3.6118	0.0003
passing	0.0019	0.0004	4.5131	0.0000
dribbling	-0.0016	0.0005	-3.4678	0.0005
defending	-0.0017	0.0002	-10.6851	0.0000

In Table 2 one can see that only 1 parameter *weight_kg* is not significant to a 5 per cent level. The variable *overall* has, naturally, the biggest (positive) impact on the *log_value (value)*, whereas *age* has the biggest negative effect.

$$t = \frac{\beta_i - 0}{se\,(\beta_i)} = \frac{\beta_i - 0}{\frac{s}{\sqrt{n}}} \tag{5.1}$$

Table 2 also shows that some coefficients are relatively small but still significant. However, a general problem with ordinary least squares (OLS) estimation is that with increasing sample size, many "coefficients" become significant, as one can in equation 5.1 . This is because the standard errors become smaller with increasing N, the t-statistic becomes larger, and the p-value smaller. (Royall, 1986)

These coefficients (e.g.: *pace*, or *passing*) could be zero in the LASSO or Bayesian LASSO estimation because of the punishment term.

5.2 Least Absolute Shrinkage and Selection Operator (LASSO)

Table 3: Summary of the LASSO

	Estimate
(Intercept)	1.72337
log_wage	0.066023
age	-0.089616
height_cm	-
weight_kg	-
overall	0.200689
potential	0.001541
shooting	0.004659
contract_valid_until	-
pace	-
passing	-
dribbling	-
defending	-0.000213

For the frequentists LASSO we used the **cv.glmnet** cross-validation function from the **glmnet** package with 100 folds to gain an estimate for λ. A λ of 0.00261 minimized the mean cross-validated error. However, we used a lambda of 0.01156 which is the largest λ such that the λ is in still within one standard error of the minimum. Hastle (2019)

As one can see in Table 3 there are considerable differences to the OLS

results. The LASSO method has shrunk 6 parameters so much that they are no longer included in the model equation.

It may be particularly noticeable, because it seems contra intuitive and the parameter had the biggest impact of all 6 excluded parameters in the linear model, that the variable *contract_valid_until* is also not included in the model.

Since LASSO does not only estimate regressors, but also selects them, no significance tests are needed.

5.3 Bayesian Lasso

With the **blasso** function of the **R** package **monomvn** it is possible to set the hyperparameters λ, for the penalty term, and α and β, which are the shape and rate parameter for the prior. The λ is in our case an empirical parameter which will be approximate through an updating Gibbs sampler. The algorithm uses the parameter of the previous sample. So iteration k uses the Gibbs sampler with hyperparmeter $k-1$. For the frequentists LASSO the λ-parameter was 0.01156, so we decided to set $\lambda = 10$, since the first 25% of the MCMC are not used for the estimation and the sampler convergence rather quickly. (Gramacy, 2019)

$$\lambda^k = \sqrt{\frac{2p}{\sum_{j=1}^{p} E_{\lambda^{(k-1)}}[\tau_j^2|\boldsymbol{y}]}}$$

The expectations are replaced with averages from the previous Gibbs sampler. As Park and Casella (2008) has shown any non-extreme starting value for λ can be used. In the first setting we did not pass any parameters for α or β.

As one can see in Table 4 the *median* for all regressors except, *weight_kg*, are unequal to zero, whereas for the frequentist LASSO we had 6 coefficients which are directly excluded from the model, e.g. zero.

However, it is unlikely that for multidimensional Bayesian model the median for a parameter is zero, since the computation depends on a Gibbs sampler. The Gibbs sampler itself is a special algorithm in the class of MCMC algorithms, which tries to solve the problem of integral formation with the help of random numbers. Therefore it is very unlikely to shrink parameters directly to 0.

Table 4: Summary of the Bayesian LASSO

	median	2.5%	97.5%
log_wage	0.067689	0.062588	0.072707
age	-0.100512	-0.102199	-0.098799
height_cm	0.001249	0.000000	0.001954
weight_kg	0.000000	0.000000	0.000408
overall	0.209842	0.208264	0.211380
potential	-0.005887	-0.007289	-0.004448
shooting	0.004830	0.004277	0.005405
contract_valid_until	0.004853	0.000000	0.007913
pace	0.000651	0.000000	0.001137
passing	0.001660	0.000000	0.002577
dribbling	-0.001395	-0.002481	0.000000
defending	-0.001608	-0.001923	-0.001189
variance	0.057685	0.056509	0.058909
lambda.square	0.000124	0.000037	0.000334

If we instead look at the 95 % credible interval, which is the Bayesian equivalent to a confidence interval and the Bayesian equvivalent to a significant test, we find that 6 of these intervals include the zero.

6 Residual Analysis, Root Mean Squared Error (RMSE) and "Sensitive Analysis"

The next step is to compare the quality of the models. First, we will take a look at the (distribution of the) residuals and after that we will calculate the out-of-sample RMSE for the 2020 *FIFA* data set.

6.1 Residual Analysis

Residuals are defined as the difference between the actual value and the predicted value of the model. As you can see from equation (6.1), negative residuals mean that the model overestimates the value and positive residuals mean that the model underestimates the value. (Hayashi, 2000, p. 16)

$$\epsilon = y_i - \hat{y}_i = y_i - (\beta_0 + \boldsymbol{\beta}_i \boldsymbol{X}) \tag{6.1}$$

One crucial assumption of the linear regression is that the residuals are normally distributed with mean 0 and constant variance σ^2.

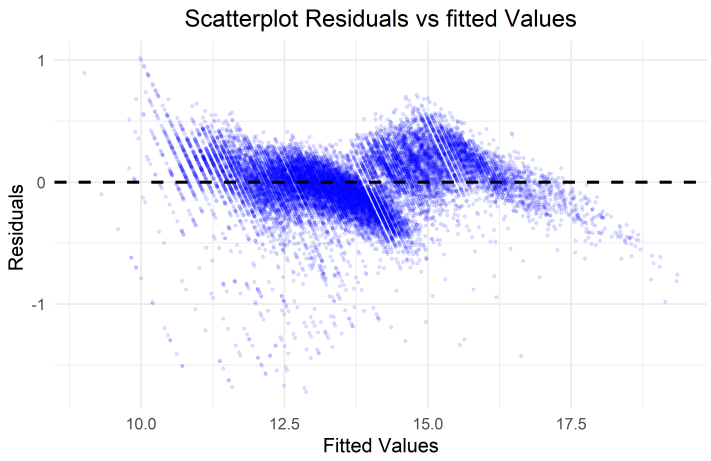

Figure 2: Plot of the Residuals vs Fitted Values for the Bayesian LASSO

In Figure 2 the residuals versus the fitted values were plotted and it appears that several assumptions are violated, on a first glance. On one hand it seems to be that there is a relationship between fitted values and residuals.

On the other hand, there also seem to be clusters with different variances. The variance in the range between 10 and 13 seems to be larger than the variance between 15 and 18, which could be a sign for heteroscedasticity.

Furthermore, the model seems to have a systematic estimation error for high values, all estimated values above 16 have a naegative residuum, i.e. the model overestimates the value of the players. Generally, it can be said that a pattern can be recognized and the residuals do not appear distributed independently of each other.

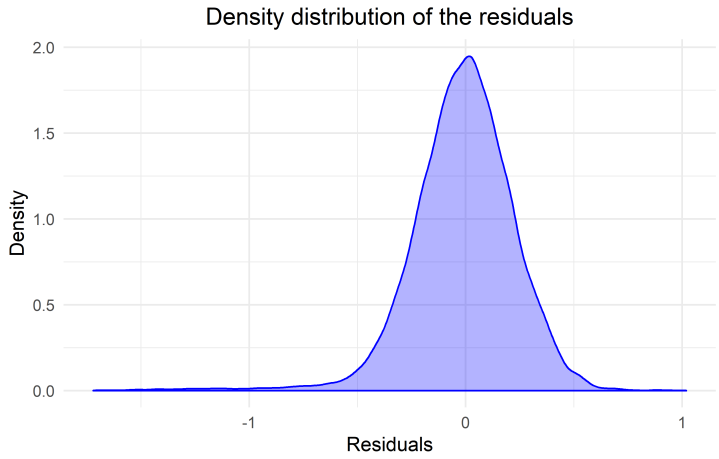

Figure 3: Distribution of the Residuals of the Bayesian LASSO

The distribution of the residuals also does not seem to be normally distributed with a mean value of 0. In Figure 3 it seems that the left tail is much longer and wider than the right tail.

The empirical mean of the residuals is -0.0138247, which is significant different to zero at a one percent level with a t-static of -7.63 and a p-value of 2.50e-14. For the residuals of the OLS and the LASSO model, the observations are the same, as shown in Figures 4 and 5 for the OLS model and Figure 6 and 7 for the LASSO model in the Appendix.

Since the ordinary least squares (OLS) method is our base method for the comparison and we are mainly interested in comparing the number of parameters included (or excluded) in (from) the model. Therefore, the OLS model was estimated again with corrected standard errors for heteroscedasticity.

First of all, we performed a Breusch-Pagan test for heteroscedasticity which was significant to a 5% level with a BP score of 2640.89 and a p-value of $\approx$

0.00e+00 and the non-constant variance score test of Breusch-Pagan test was also significant to a 5% level test with a χ^2 score of 42.1247 and a p-value of $8.5636553 \times 10^{-11}$.

With the corrected standard errors, 2 coefficients are no longer significant, which is 1 coefficient more as in the first OLS model.

6.2 Root Mean Squared Error (RMSE) of the Models

As already mentioned in the introduction, we have taken the *FIFA* data from *2020* as test data. At least briefly, it should be noted here that the leap from the 2019 edition to the 2020 edition has added a further dimension. Between the two years there could be a trend, which could catch other effects like inflation, that increases player values without being explained by any of the variables in the model.

In a quick check using a linear OLS model containing the data for both years and a dummy for the different versions, we get a significant estimator of 0.0322. This estimator is significant with corrected standard errors at a significance level of 5 percent with a t-value of 13.967 and a p-value of 3.37e-44.

The significant dummy for the *versions* does not automatically mean that there is an annual trend. The main thing to keep in mind is that we excluded a number of variables at the beginning and that these can now be included in the dummy as a bias.

However, all three models have the same problem and therefore this should not play a major role in the RMSE's assessment of the models. The RMSE should therefore increase equally for all three models. The Root Mean Squared Error (RMSE) is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}}$$

The RMSE is a relative comparison measure that calculates the performance of different models using the same data set. If the RMSE is determined with other data, as in this case with the 2020 data, it is called an out-of-sample RMSE.

The RMSE for the frequentist linear model is 0.218308, which in itself is difficult to interpret. For the (frequentist) LASSO, the RMSE is 0.217305, which is a reduction of -0.46 percent compared to the base linear regression.

For the Bayesian LASSO, which was estimated with 12.500 MCMC iterations, 2.500 of which were cut off at the beginning and were not used for the calculation. We get a RMSE of 0.216686, which is a change to OLS of -0.74 percent. Compared to the frequentist LASSO we get a reduction of the RMSE of -0.28 percent.

The improvement in the RMSE of the Bayesian model compared to the linear model is relatively small and is even smaller compared to the LASSO. However, the computational effort for the small improvement is relatively high. The linear model (with corrected standard errors) was calculated within a few seconds. Whereas the frequentist LASSO took only slightly longer, despite the cross-validation for the punishment parameter λ, the Bayesian LASSO took almost 18 hours due to the 12.500 MCMC iterations.

6.3 Changing other Hyperparameter

As Park and Casella (2008) have described in their paper, instead of selecting λ directly, one can also use a diffuse prior and only specify the parameters of the gamma distribution. We took the proposed parameters from Park and Casella (2008) with $r = 1$ and $\delta = 1.78$.

Table 5: Summary of the Bayessian LASSO with hyperpriors

	median	2.5%	97.5%
log_wage	0.067611	0.062648	0.072639
age	-0.100563	-0.102264	-0.098894
height_cm	0.001243	0.000000	0.001957
weight_kg	0.000000	0.000000	0.000479
overall	0.209864	0.208337	0.211426
potential	-0.005909	-0.007363	-0.004507
shooting	0.004825	0.004257	0.005428
contract_valid_until	0.004740	0.000000	0.007845
pace	0.000623	0.000000	0.001137
passing	0.001612	0.000000	0.002577
dribbling	-0.001330	-0.002459	0.000000
defending	-0.001601	-0.001924	-0.001163
variance	0.057674	0.056483	0.058891
lambda.square	0.000089	0.000023	0.000261

As Table 5 shows, the results have not really changed. There are still 6 parameters outside the 95 percent credibility interval and, also the size of the effects have hardly changed, which is mainly due to the many iterations of the MCMC algorithm. Also, the λ-parameter differs only slightly , from 1.24e-04 to 8.94e-05. The respective 95 percent credibility intervals overlap, which means that they are not differ significantly.

7 Conclusion

As we have shown in sections 5 and 6, the results of the Bayesian LASSO are very similar to the results of the frequentist LASSO. However, the Bayesian LASSO performs a bit better, measured by the RMSE, which does not seem to be substantially different. Our results are similar to those of Park and Casella (2008), who also found no great difference between Bayesian LASSO and the normal one. However, Hans (2009) found a reduction of the average prediction error from 16 to 36 percent with different dataets.

The small difference in the methods can - of course - also be due to the underlying data, since we estimate and calculate the RMSE with only one data set. As we have seen from the analyses of the residuals and the distribution of the response variable, the data are not optimal either. A possibility for further analysis would be a Box-Cox transformation of the data to see if the Bayesian LASSO has greater advantages. Another possibility would be to generate data with a Monte Carlo simulation and then let the different methods do the estimation again. In this way, certain problems - such as heteroskedasticity - can be controlled.

8 Appendix

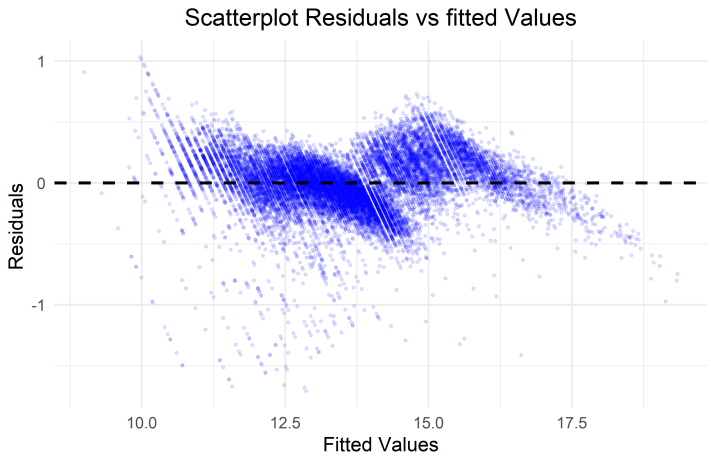

Figure 4: Plot of the Residuals vs Fitted Values for the Linear Model

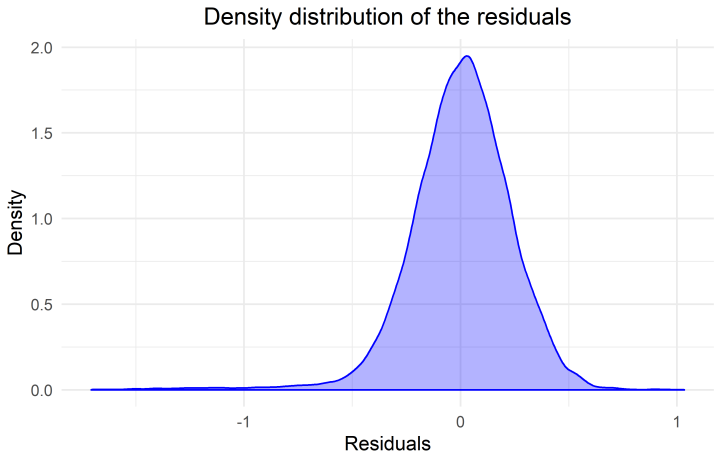

Figure 5: Density Plot of the Residuals of the Linear Model

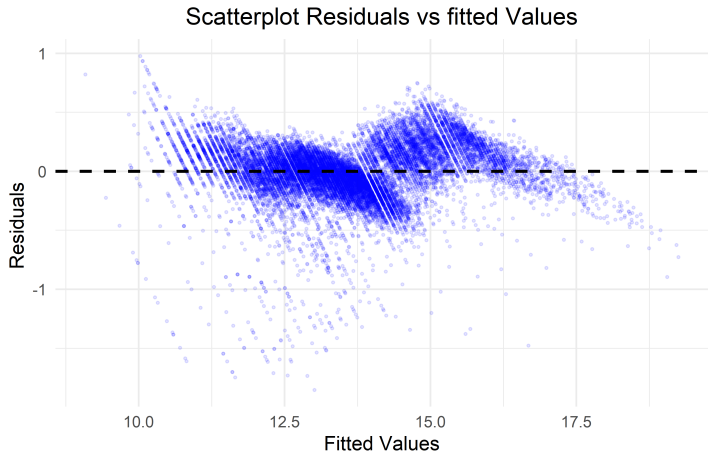

Figure 6: Plot of the Residuals vs Fitted Values for the LASSO Model

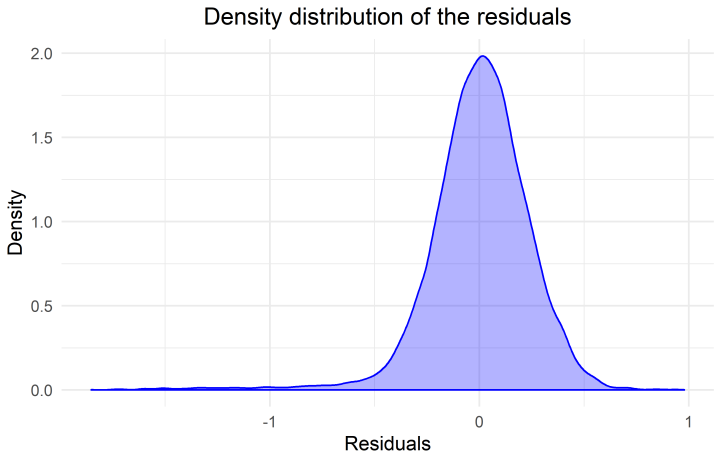

Figure 7: Density Plot of the Residuals of the LASSO Model

References

Andrews, **D. F.**, & **Mallows**, **C. L.** (1974). Scale Mixtures of Normal Distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(1), 99–102. Retrieved January 11, 2020, from https://www.jstor.org/stable/2984774

Gelman, **A.** (2004). *Bayesian data analysis* (2. ed.). Boca Raton [u.a.], Chapman & Hall/CRC.

Ghosh, **J. K.**, **Delampady**, **M.**, & **Samanta**, **T.** (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York, Springer-Verlag. https://doi.org/10.1007/978-0-387-35433-0

Gramacy, **R. B.** (2019). Monomvn: Estimation for MVN and Student-t Data with Monotone Missingness. Retrieved January 12, 2020, from https://CRAN.R-project.org/package=monomvn

Hans, **C.** (2009). Bayesian lasso regression. *Biometrika*, *96*(4), 835–845. Retrieved January 16, 2020, from https://www.jstor.org/stable/27798870

Hastle, **T.** (2019). Glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. Retrieved January 12, 2020, from https://CRAN.R-project.org/package=monomvn

Hayashi, **F.** (2000). *Econometrics*. Princeton [u.a.], Princeton UnivPress.

Leone, **S.** (2020). FIFA 20 complete player dataset. Retrieved January 7, 2020, from https://kaggle.com/stefanoleone992/fifa-20-complete-player-dataset

Park, **T.**, & **Casella**, **G.** (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, *103*(482), 681–686. https://doi.org/10.1198/016214508000000337

Royall, **R. M.** (1986). The Effect of Sample Size on the Meaning of Significance Tests. *The American Statistician*, *40*(4), 313–315. https://doi.org/10.2307/2684616

Tibshirani, **R.** (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. Retrieved January 6, 2020, from https://www.jstor.org/stable/2346178

Eidesstattliche Versicherung

Ich versichere an Eides statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe. Die Arbeit hat in dieser oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Essen, den _____ _____

Jens Klenke