Stockholms universitet

Inter-calibration of Laboratory Measurements Using Lasso

Patrick Dinan

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2024:12
Matematisk statistik
Augusti 2024

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2024:12**
http://www.math.su.se

Inter-calibration of Laboratory Measurements Using Lasso

Patrick Dinan*

August 2024

Abstract

The least absolute shrinkage selector operator (lasso) has been popularized for its ability to perform both variable selection and shrinkage. This paper examines the potential of lasso to consider how a change in sampling protocol could affect long-term monitoring programs. By modifying the lasso and implementing a complementary Bayesian approach, simulations studies are conducted to asses various performance metrics relating to point estimation and statistical testing. These evolutions examine how this penalized regression method compares against previous approaches on this topic.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: patrickdinan@icloud.com. Supervisor: Martin Sköld.

Inter-calibration of Laboratory Measurements Using Lasso

Patrick Dinan

August 26, 2024

Abstract

The least absolute shrinkage selector operator (lasso) has been popularized for its ability to perform both variable selection and shrinkage. This paper examines the potential of lasso to consider how a change in sampling protocol could affect long-term monitoring programs. By modifying the lasso and implementing a complementary Bayesian approach, simulations studies are conducted to asses various performance metrics relating to point estimation and statistical testing. These evolutions examine how this penalized regression method compares against previous approaches on this topic.

Acknowledgements

I would like to acknowledge my supervisor Martin Sköld at Stockholm University for the guidance and help throughout this thesis. No Chat GPT or other AI-tools has been used during the writing of this thesis.

Contents

1 Background **3**
1.1 The Lasso 4
1.2 Statistical Testing for the Lasso 5
1.3 Bayesian Lasso 6
1.4 Modifying the Lasso 7

2 Methods **10**
2.1 Aim 10
2.2 Data Generation 10
2.3 Simulation settings 10
2.4 Performance metrics 10

3 Results **12**
3.1 Bias 12
3.2 Root Mean Squared Error 14
3.3 Type I Error 14
3.4 Power 14

4 Discussion **17**
4.1 Summary 17
4.2 Further Extensions 18

1 Background

This problem originates from an unpublished document by Sköld (2024); consider a monitoring program that covers a single-time series of contaminant concentration and has a change of lab from A to B. We would like to infer the trend component of data over some time T, considering the lab change. For this model we'll make some simplifying assumptions about the data. Let the measured log-concentration before the lab change be equal to $c+\delta_A$ then the concentration after the lab change will be equal to $\gamma(c+\delta_B)$, where δ_A and δ_B are i.i.d.

Furthermore, we'll assume that the log-concentrations of the measured concentration is generated from a linear regression process, with equal trends both before and after the change of laboratory. Letting y_t denote the log-concentration at time t and if the lab change happens at some time T_c, then the log-concentration before lab change is given by $y_t = \alpha + \beta t + \epsilon_t$ for $t = 1, ..., T_c$ and $y_t = \alpha + \gamma + \beta t + \epsilon_t$ for $t = T_c + 1, ..., T$, where α, $\alpha + \gamma$ are the respective models intercepts and ϵ_t is iid normal with mean 0 and some variance σ^2. By coding u_t as a dummy for the lab change our model can be written as:

$$\mathbf{y} = \mathbf{X}\theta + \epsilon, \tag{1}$$

x where the vector $\mathbf{y} = (y_1, \ldots, y_T)'$, the matrix $\mathbf{X}$ is a $(T \times 3)$ matrix with columns: $X_1 = (1, 1, \ldots, 1)', X_2 = (1, 2, \ldots, T)', X_3 = (u_1, u_2, \ldots, u_T)'$, the parameter $\theta = (\alpha, \beta, \gamma)'$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_T)$ and $\epsilon_t \sim N(0, \sigma^2)$.

Three different methods for dealing with the lab change in assessing the trend β have been proposed. Firstly, by ignoring the lab factor altogether and estimating the trend β through ols (ordinary least squares) by fitting the data to the restricted model. Secondly, by fully taking the lab change into account, resulting in the ols estimate for β from using the full linear regression model. Lastly, a selection method for using the full or restricted model similar to stepwise regression. Begin with fitting the data to the full model; if the p-value for the estimated $\hat{\gamma}^{full}$ is less than 5%; estimate the trend according to the second method; else use the first method. Note that the critical cut-off point of 5% in our example is chosen arbitrarily.

To compare how these methods estimate the trend β, the Mean Squared Error (MSE) defined as $MSE = E((\hat{\beta} - \beta)^2)$, of the resulting estimators from each of the three methods was analyzed. The results showed that for relatively low signal of γ to noise ratios the restricted model will produce point estimates of β with lower Mean Squared Error (MSE) compared to the full model. This, due to the inflated variance resulting from an extra parameter. But as the signal of γ increases, in relation to the noise, the added bias of the restricted model will produce higher MSE compared to the full model. Because the distribution of the stepwise model's estimate can be

seen as a mixture of the distribution of the estimates resulting from the full respective restricted model, the point estimates will be a weighted average of the two models where the weights depend on the t-test for the estimate $\hat{\gamma}^{full}$. Consequently, as gamma increases relative to noise, the probability of selecting the full relative to the restricted model increases, which will tend to the MSE of the full model.

These effect showed similar behavior in type I error (which in this case would be the probability of falsely rejecting the null-hypothesis that there exist no time trend); where the full model will offer nominal significance the same as the actual, the restricted models actual significance of β will tend to one as γ increases and the stepwise model's estimate nominal significance of β will resemble that resulting from the restricted model for relatively small signals of γ and tend to the actual significance as γ increases in relation to the noise. However, the cost of these unbiased properties of using the estimates of the full model came at a cost of loss of power.

Berk, Brown and Zhao (2009) showed that when estimates are based on a mixture of sampling distributions, arising from a rule such as in our case above, the finite sampling distribution of that estimate can behave in a way that departs from assumptions of how statistic are assumed to behave. They argued that selection procedure such as ours, can lead to invalid statistical test and confidence intervals unless the selection procedure selects a single model with a probability near unity and referred to a possible solution of sample-splitting, although this procedure is restricted for relatively small sample sizes. As an alternative to the estimator resulting from stepwise regression that is offered above, we will examine the use of a penalized regression method known as the lasso and how this can be applied to incorporate the lab change for inferring the trend of our measured contaminant concentration as formalized above. We will compare these approaches to see which and when is more viable.

1.1 The Lasso

As an alternative to regular linear regression, Hoerl and Kennard (1970) proposed ridge regression which minimizes residual sum of squares while constraining the coefficients L_2 norm values to some non-negative value. Tibishirani (1996) introduced the least absolute shrinkage selection operator (lasso), which preforms both variable selection and estimation by penalizing the L_1 norm. Working in the gaussian linear regression setting with N samples and p predictors the lasso solutions are given by

$$\hat{\beta} = \arg\min_{\alpha,\beta} \sum_{i=1}^{N} (y - \alpha - \sum_{i=1}^{p} x_i \beta_j)^2 \quad \text{subject to} \sum_{i=1}^{p} |\beta| \leq t. \tag{2}$$

The solution to (2) can also be given by its Langrarian form;

$$\hat{\beta} = \arg\min_{\alpha,\beta} \sum_{i=1}^{N} (y - \alpha - \sum_{i=1}^{N} x_i \beta_j)^2 + \lambda \sum_{i=1}^{p} |\beta_j|. \tag{3}$$

Similarly to ridge regression, lasso regression has the effect of shrinking some estimates towards zero, but because of penalization on the L_1 norm lasso has the potential of shrinking some coefficients to exactly zero for large enough values of λ. Thus the lasso has the ability of preforming variable selection as-well as estimation.

The value of λ is generally chosen through K-fold cross validation. A process where the sample data is divided randomly into K groups, the K-th group being the test set and the remaining $K-1$ groups the training set. Based on the training set, the lasso solutions are then calculated across a grid of λ values. Then, for each value of λ the mean squared prediction error (MSEP) is calculated based on the test set. This is done for all of the K different groups, after the resulting MSEP values are averaged for each λ in the grid. Generally the λ that produced the smallest MSEP is chosen. However, other values of λ have been proposed, Hastie et al. (2008), chapter 3 for suggest selection on the "1se-rule" which involves the selecting the most sparse model within one standard error of "the best" ("the best" meaning the λ with the smallest MSEP). Furthermore, in the special case of K-fold cross-validation where $K = N$ (where N is the number samples) this becomes what's known as leave-one-out cross validation. Unlike ridge regression, the lasso generally does not offer analytical solutions so other methods are needed. Efron et al (2004) constructed a modified LARS algorithm that produces solutions to the entire path of lasso solutions, which becomes useful when estimating λ through cross validation.

1.2 Statistical Testing for the Lasso

Working within the framework of (1), in the case of ordinary least squares, the solutions are given by

$$\hat{\theta} = (\mathbf{X^T X})^{-1} \mathbf{X^T y}. \tag{4}$$

From which it easily be shown, that the covariance matrix for the estimates is given by

$$Var(\hat{\theta}) = (\mathbf{X^T X})^{-1} \sigma^2, \tag{5}$$

and the standard erors for predictor j could be easily gotten by taking the square root of j:th diagonal element of the covariance matrix multiplied by the standard deviation σ. From there confidence intervals and significance test for predictors can be constructed. In case of the lasso there arrises a problem, since no closed form solution generally exists. Tibishirani (1996) proposed the bootstrap (Efron (1979)) for retrieving the standard errors of the lasso estimates by way of resampling the residuals, which can be used for calculating z-scores and it's related p-values, like in classical hypothesis testing. Another approach for statistical testing is taken by moving away from the frequentist view of hypothesis testing and viewing the problem in the Bayesian context which we will explain in the next section.

1.3 Bayesian Lasso

Tibshiraini (1996) showed that the lasso solutions to β could be seen as the Bayesian posterior mode for independently identically distributed Laplace (double exponential) priors for β_j with density

$$f(\beta) = \prod_{j=1}^{p} \frac{1}{2\tau} e^{-|\beta_j|/\tau}, \text{where } 1/\tau = \lambda.$$

To see this consider the regular linear regression setting (without an intercept) with N samples and p predictors. Then each of the responses are assumed iid with $y_i \sim N(\sum_{j=1}^{p} x_{ij}\beta_j, \sigma^2)$ and therefore, the likelihood of y is given by

$$f(y|\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2}.$$

By Bayes theorem, the posterior is given by is proportional to the likelihood multiplied by the prior, that is:

$$f(\beta|y, \sigma^2) \propto f(y|\beta, \sigma^2) f(\beta).$$

The β that maximizes the posterior is equivalent to the β that minimizes the negative log posterior. Which when using the Laplace distribution above, gives the following negative log posterior;

$$\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|.$$

Thus for a given value σ^2 the posterior mode of β is equivalent to the solutions that minimizes (3) (with $\alpha = 0$). If we were to include an intercept - an independent flat prior for α could be added and would also yield the solution to (3).

Park and Cassella (2008) conditions on σ^2 in the prior distribution of β as to avoid possible problems in the sampling distribution. Furthermore they make use of the Laplace distribution as a scale mixture of normal, makes for easy implementation of Gibbs sampler within this hierarchical framework for the Bayesian model. They used a non-informative prior on σ^2 and suggested an empirical Bayes approach for estimating the penalty parameter λ, and for a full Bayesian they suggested placing a hyper prior on λ thats conjugate. The priors for a full Bayesian is illustrated below.

$$f(\sigma^2) = \sigma^{-2}$$

$$f(\beta|\sigma^2, \lambda^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sigma} e^{-\lambda|\beta_j|/\sigma}$$

$$f(\lambda^2) = \frac{\delta^r}{\Gamma(r)} (\lambda^2)^{r-1} e^{-\delta\lambda^2}.$$

These could be implemented into some MCMC algorithm like Gibbs or Metropolis-Hastings in order to sample from the posterior. Although the posterior mode provides the analogy to the frequentist lasso, Park and Casella (2008), suggest that the posterior mean or median provide a more naturally Bayesian estimate. Furthermore, in contrast to the frequentist lasso that solely provides us points estimates and offered no obvious procedure for dealing with standard error for hypothesis testing, the Bayesian lasso approximates an entire sampling distribution, which can be used for statistical hypothesis testing by constructing credible intervals, say.

1.4 Modifying the Lasso

Many generalizations of the lasso have been offered, Zhu and Hastie (2005) proposed the elastic net, which is a type mixture of ridge and lasso penalization. Zou (2006) introduced the adaptive lasso which penalizes on the L_1 norm but by a different magnitude for each of the coefficients based on some initial estimates (usually ols or ridge). Returning to our model problem in Section 1.1, we wish to modify the lasso as to only penalize on the lab parameter γ, this ensures us that our β coefficient will always be included in the model and leads us to minimize the following function:

$$\arg\min_{\alpha,\beta,\gamma} \sum_{i=1}^{N} (y_i - \alpha - \beta x_i - \gamma u_i)^2 + \lambda|\gamma| \tag{6}$$

Assuming that all the covariates in $\mathbf{X}$ are centered, then for all lasso solutions $\alpha = \bar{y}$. Further, we will assume without loss of generality, that $\bar{y} = 0$. Thus our problem reduces down to minimizing $\sum_{i=1}^{N}(y_i - \beta x_i - \gamma u_i)^2$ subject to $\gamma \leq t$, for some non-negative t. Consider the cost function, this

corresponds to the function of an ellipse centered at the ols solution. That is $RSS = ||y|| + ||x||\beta^2 + \langle x, u\rangle\beta\gamma + ||u||\gamma^2 - 2\langle y, x\rangle\beta - 2\langle y, u\rangle\gamma$. Figure 1 plots this situation for a fixed value of t, where $\hat{\gamma}^m$ and $\hat{\gamma}^s$ corresponds to the ols solution regressing on the full (m) respective restricted (s) model. The point where contour tangents the shaded area corresponds to our lasso solution. This will differ from the full ols estimates as long as the area does not cover this point. Moreover, for all $t \leq |\hat{\gamma}^m|$ the lasso estimate $\hat{\gamma}^l$ will equal $sign(\hat{\gamma}^m) \cdot t$.

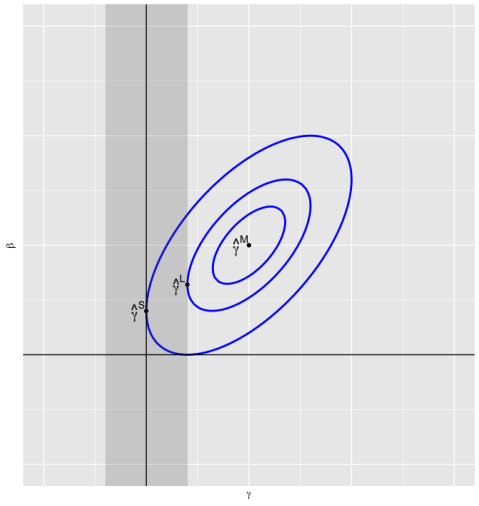

Figure 1: Contours corresponding to the residual sum of squares (blue), shaded area represents the constraint for the modified lasso. The highlighted points represents the various solutions, with $\hat{\gamma}^m$ corresponding to the ols solution regressing on the full model. $\hat{\gamma}^s$ represents the ols solution regressing on the simple mode. $\hat{\gamma}^l$ being the lasso solution.

Assume first that the estimate of the lab parameter resulting from the full ols solution, γ^m is positive. Then we know that lasso solution for the lab parameter $\gamma^l \geq 0$ and the equation can be written as

$$Q(\beta, \gamma) = \sum_{i=1}^{N}(y_i - \beta x_i - \gamma u_i)^2 + \lambda\gamma.$$

Let θ be a (1×2) vector with entries $\theta_1 = \beta$ and $\theta_2 = \gamma$. Then, solving for the θ that minimizes the equation above yields,

$$\frac{\partial Q}{\partial \theta} = \frac{\partial}{\partial \theta} RSS(\theta) + \lambda \frac{\partial \theta_2}{\partial \theta}$$

$$\frac{\partial Q}{\partial \theta} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\theta) + \lambda \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\hat{\theta} = (\mathbf{X^T X})^{-1}(\mathbf{X^T y} - \lambda/2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}).$$

Let the matrix $(\mathbf{X^T X})^{-1}$ be denoted by $\mathbf{C}$, with entries c_{ii} for $i = 1, 2$. Then the lasso-solution for γ is given by $\hat{\gamma}^l = (\hat{\gamma}^m - \lambda c_{22}/2)^+$. Similar calculations show that if $\gamma^m < 0$, then $\hat{\gamma}^l = (\hat{\gamma}^m + \lambda c_{22}/2)^-$. The solution corresponds to the soft thresholding operator $S_{\lambda c_{22}/2}(\gamma^m)$ and thus $\hat{\gamma}^l = sgn(\hat{\gamma}^m)(|\hat{\gamma}^m| - \lambda c_{22}/2)^+$. This differs from the stepwise procedure of estimating γ and is illustrated in Figure 2 for a fixed λ and known variance case.

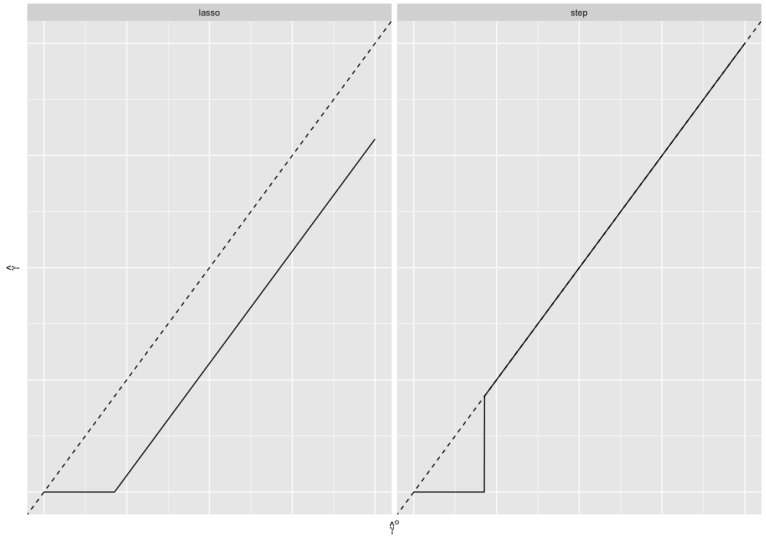

Figure 2: Estimates of $\hat{\gamma}$ using the modified lasso (left) and stepwise regression (right) plotted against the full ols estimate with the 45 degree line (dashed) for reference

In our Bayesian approach the modification will differ from the ordinary Bayesian lasso in that an independent flat prior for both α and β will be used and a double-exponential prior for γ. This way, the posterior mode will correspond to the solution in (3). Furthermore, instead of using a vague gamma prior as in Park and Casella (2008), we will be using a standard half-Cauchy hyper prior on λ as in van Erp et. al (2019). For σ^2 an improper independent prior of the form in section 2.3 is still used. The priors that will be used in this modified Bayesian lasso are summarized below:

$$f(\alpha, \beta, \sigma^2) \propto \sigma^{-2}$$

$$f(\gamma|\lambda, \sigma) = \frac{\lambda}{2\sigma} e^{\frac{\lambda}{\sigma}|\beta|}$$

$$f(\lambda) = \frac{1}{\pi(1 + \lambda^2)}$$

These priors could then be used to sample from the posterior through some Markov Chain Monte Carlo (MCMC) method; see Held and Bove (2020), chapter 8 for example.

2 Methods

2.1 Aim

To evaluate how the lasso preforms in our single time-series monitoring program, a small Monte Carlo (MC) simulation study is preformed. We asses several performance metrics across different parameter values and compare how inference and estimation differ across models and parameter values. These include the frequentist lasso; where the λ is chosen by leave-one-out cross validation, the Bayesian lasso; using a standard half-Cauchy prior for λ and step-wise regression model as described previously; where the lab parameter is omitted if the related p-value falls under 0.05. We will also for comparison add the ordinary least square solutions from the restricted and full model.

2.2 Data Generation

The response values was generated according $\mathbf{y} = \mathbf{X}\theta + \epsilon I_T$ as defined as in (1). For all simulations of data, we will let the intercept $\alpha = 0$, the timeframe equal ten years ($T = 10$). The lab parameter will be constructed such that the first five years corresponds to measurements from lab A and latter five years from lab B. The parameters β, γ and σ will be varied over a range of values.

2.3 Simulation settings

The simple, multiple and stepwise models were all constructed using base R *lm* function. For the frequentist lasso we implemented the *glmnet* package by Friedman et al (2010), using $penalty.facor$ argument to penalize only the time coefficient and setting $nfolds = 10$ for leave-one-out cross-validation. The Bayesian lasso was implemented using *Rstan* by Stan development team (2020); the number of Markov chains was set to 4, with each chain having 5000 iterations and where the first half were used as burn-ins. Across all simulation we set $n = 500$.

2.4 Performance metrics

To evaluate the methods performance of point estimation and statistical hypothesis testing we will be considering the follow metrics, along with their MC approximation:

- *Bias*: This assesses how how far away and in what direction our estimate of the trend parameter $\hat{\beta}$ is compared to true parameter β and is defined as $E[\hat{\beta}] - \beta$.. In case of the Bayesian lasso the point estimate is taken to be the mean of our posterior distribution of β. If $\hat{\beta}_i$ is the estimated coefficient for simulation i then the MC approximation of bias is given by:

$$\widehat{Bias}(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_i - \beta$$

- *Root square mean error(RMSE)*: This is the root of the squared errors defined as $\sqrt{E[(\hat{\beta} - \beta)^2]}$. This can be decomposed into the sum of the estimators squared bias and variance, thus this metric incorporates the standard errors of our estimator as-well as its bias. As in the bias case, we will use the posterior mean for the Bayesian lasso and the Monte Carlo approximation of RMSE is given by

$$\widehat{RMSE}(\hat{\beta}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_i - \beta)^2}$$

- *Type I error*: Type I error defines the probability of a statistical test falsely rejecting the null-hypothesis for given significance level. An appropriate test should have type I close to the target significance level. In our case the null corresponds to the hypothesis that no trend exists ($\beta = 0$). For the frequentist methods we will base our test on the two-sided p-value from the t-statistic given by the $\hat{\beta}/SE(\hat{\beta})$. We will reject the null for p-values less than 0.05. Because of the problem concerning the lasso estimator's standard errors, as discussed in Section 2.2, we will from the suggestion of a referee(M. Sköld), be using the standard errors from the full regression model. If p_i is our p-value for simulation i then the type I errors are approximated by

$$\hat{P}(p < 0.05|\beta = 0) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(p_i < 0.05)$$

  In contrast to frequentist case, which assigns probability to data given that the null is true, in the Bayesian framework we directly test whether the null is true given the data. Therefore, we will define type I error for the Bayesian method based on whether the credible intervals of β covers our null. If the 95 % credible intervals for the posterior of β for simulation i is given by $CI(\beta_i)$, then the Monte Carlo approximation is given by:

$$\hat{P}(0 \notin CI(\beta)) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(0 \notin CI(\beta_i))$$

- *Power*: While the type I error denotes a test false positive the power measures the probability of a true positive, i.e. correctly rejecting the null. The ability to correctly detect a trend ($\beta \neq 0$) in our case. The p-value is derived the same as in the type I above, but with the exception of being one sided corresponding to the true effect's sign and the MC is given by:

$$\hat{P}(p < 0.05 | \beta \neq 0) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(p_i < 0.05)$$

By similar argument as above, we will define the Bayseian power as $P(\beta < 0)$ or $P(\beta > 0)$ and approximate it by:

$$\hat{P}(\beta > 0) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(0 < CI_l(\beta_i)),$$
$$\hat{P}(\beta < 0) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(0 > CI_h(\beta_i))$$

where CI_l and CI_h is the lower and upper bound of a 90 % credible interval of the posterior of β for simulation i.

3 Results

3.1 Bias

Figure 3 illustrates the different methods bias as a function of the lab parameter γ across different noise variances σ. The performance of the frequentist lasso shows a similar pattern as that of step-wise regression, in that it increases with the effect of γ relative to the noise, up to a point where it decreases towards zero. Meaning that the cross validation procedure chooses larger penalties λ up to a certain point, where λ then tends to zero. An advantage of the lasso relative to the stepwise in this case, is the ability of estimating values of γ between zero and the estimate resulting from the full OLS model. As this means that β can take on values in between the values from the full and restricted model, which could explain the lower bias of the lasso compared to the stepwise. However, we could decrease this difference on average, by selecting a higher p-value for the γ estimate in the stepwise procedure. Though this could lead to issues of bi-modality for large γ that wouldn't be captured by inspecting bias. We can further note, that there exists a relatively large discrepancy between the estimator from the frequentist and Bayesian lasso.

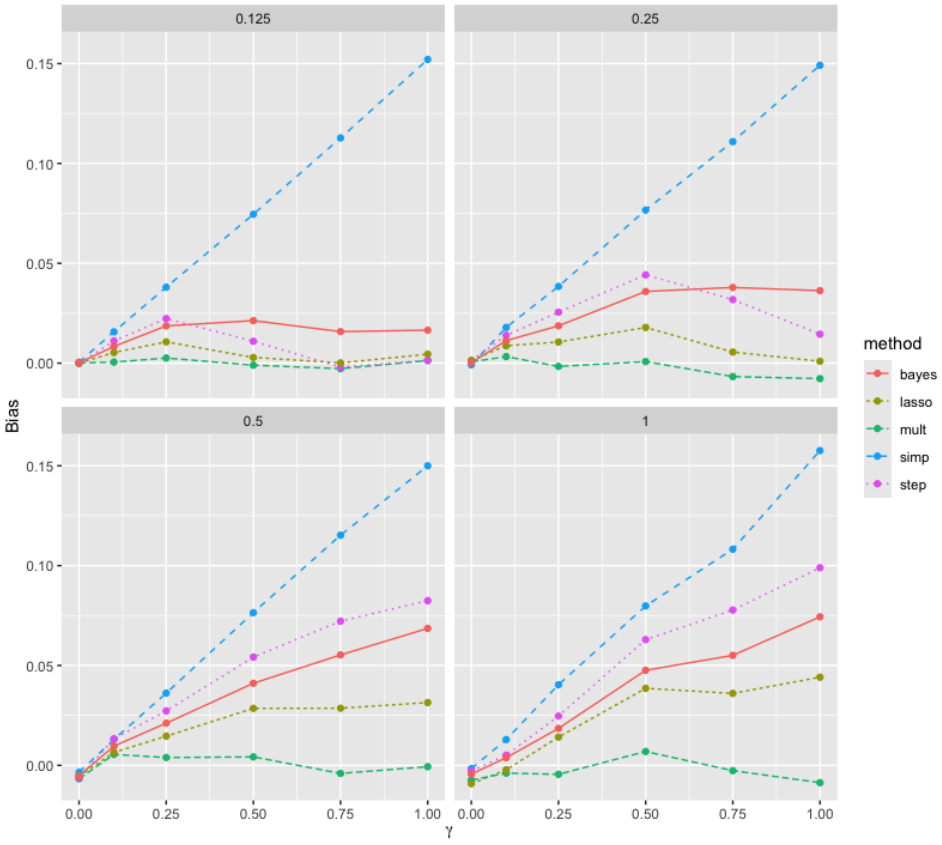

Figure 3: Simulated bias for the different methods as a function of γ at different values of σ when $\beta = 0$

3.2 Root Mean Squared Error

Figure 4 plots the root mean squared error as a function of γ for different values of σ when $\beta = 0$. Because the lasso and step methods all have some degree of penalization or variable selection we can observe a lower RMSE when σ is relatively high compared to the effect of γ. For these values we can also see that the frequentist lasso estimates has a higher RMSE relative to the RMSE estimated from the step-wise method. This, again resulting in the fact that when noise is relatively high the stepwise procedure seems to set the γ coefficient to zero while the lasso solution has the ability to take on nonzero values, resulting in relatively higher standard errors of β because of the extra parameter, but lower than the full fit. As γ increases relative to noise, it would appear that the higher bias of step-wise compared to lasso we observed in the previous figure dominates the decreased variance. It can be noted that the Bayesian lasso outperforms the frequentist lasso for low signal to noise ratios. Because of the discrepancy in bias for the two versions of lasso, illustrated in the previous figure, this would imply that the standard errors of the Bayesian mean estimator is considerably lower than the frequentist one.

3.3 Type I Error

The type I errors for the same value as above are shown in Figure 5. The frequentist and lasso type I errors showed similar behavior as the step ones, it increases for as the relative effect size of γ grows, up till a point where it then begins to decrease. It can be noted that the increase in type I error, that occurs when the effect of γ is relatively low, is smaller for the frequentist lasso in relation to the step-wise procedure and even more so for the Bayesian lasso. Although for relatively highest values of γ, the stepwise enjoys a smaller type I error. Even though the Bayesian lasso provided higher biases on average, the type I error is lower compared to the frequentist. This suggesting that the credible intervals constructed in the Bayesian case produces a comparatively more conservative test.

3.4 Power

In Figure 6 the power of the different methods respective tests are plotted as a function of σ when the lab parameter $\gamma = 0$. We can observe that the power of the lasso method is somewhat lower than the step-wise method. This is expected since the test statistic of the lasso is based on the standard errors of the full fit, compared to the step-wise which also considers using the standard errors of the restricted model (depending on the p-value related to the estimate of γ). Furthermore, we can observe that the power of the Bayesian lasso is even lower than the frequentist method. This again suggests a more conservative credible intervals compare to frequentist methods.

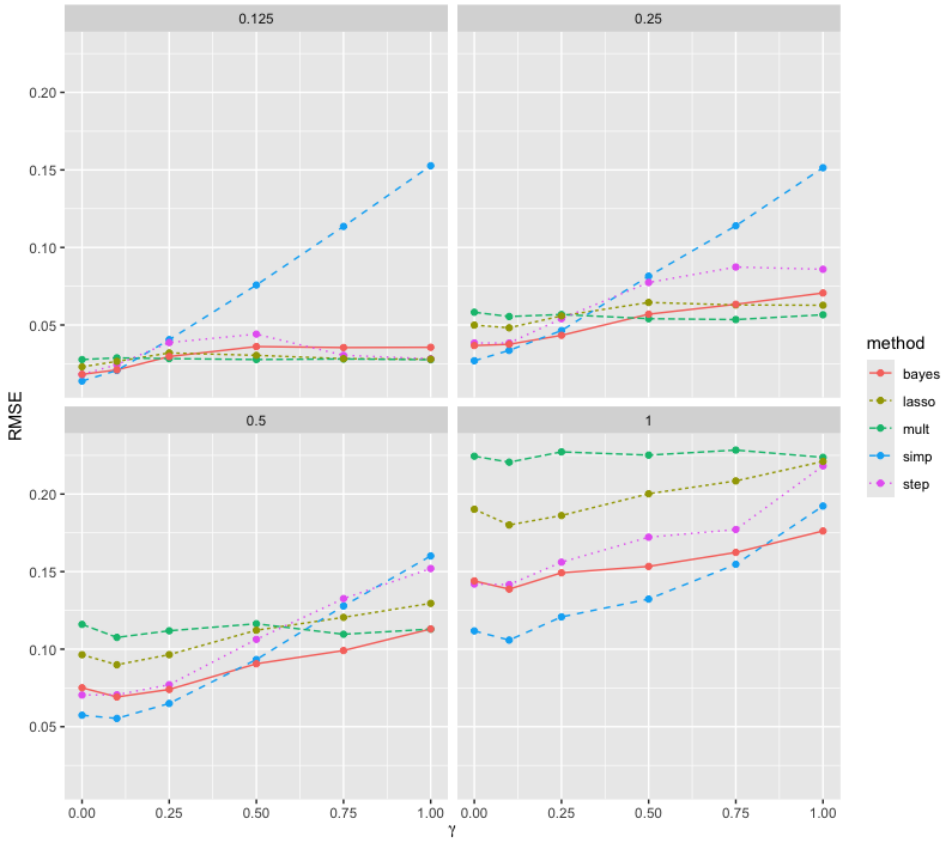

Figure 4: Simulated root mean squared error (RMSE) for the different methods

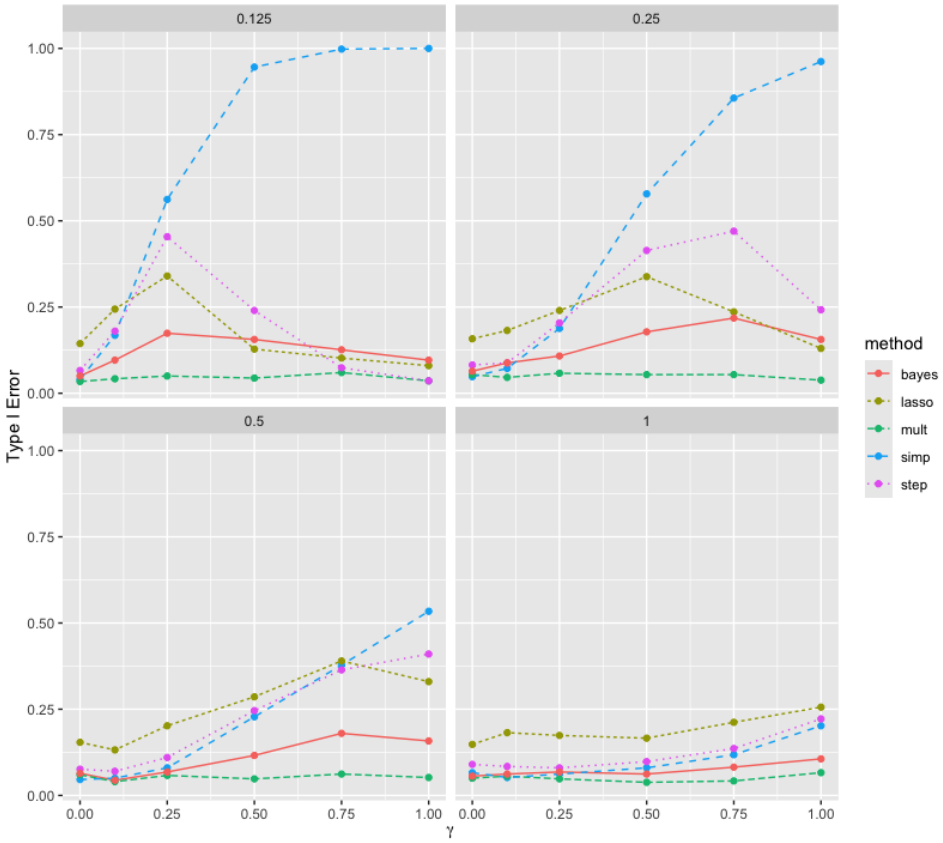

Figure 5: Simulated type I errors the different methods

As γ increases the test would increasingly suffer from false power due to the point estimates biases.

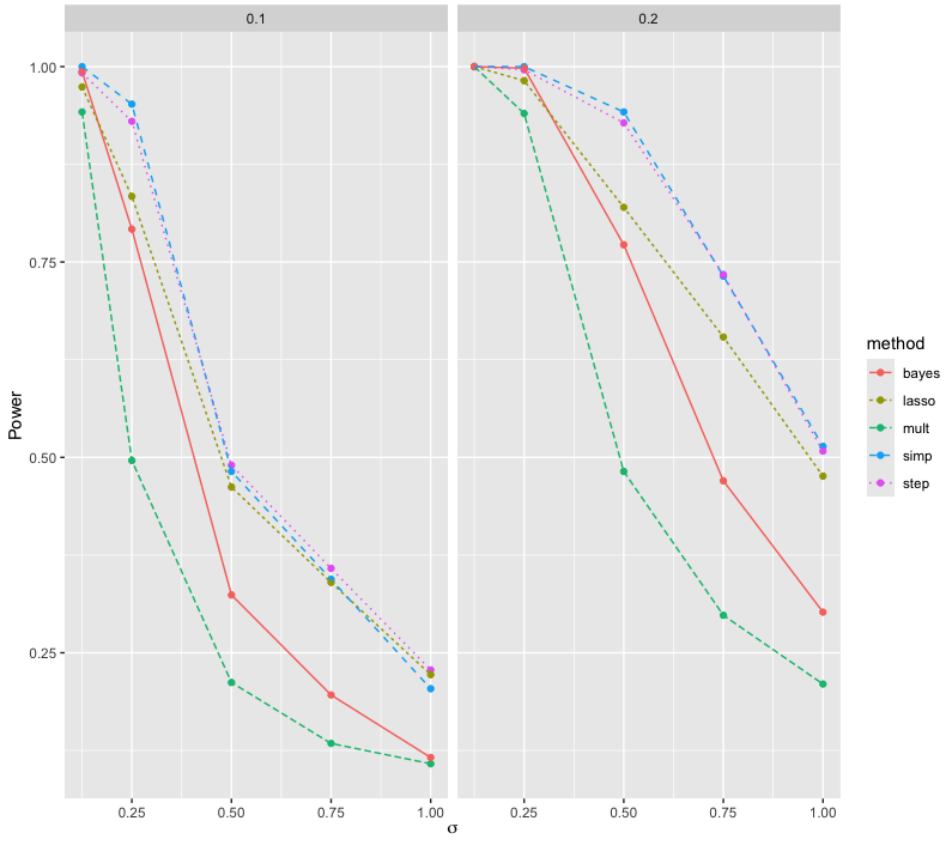

Figure 6: Simulated power of the different methods as a function of the noise standard deviation for $\beta = 0.10$(left) and $\beta = 0.20$(right) when $\gamma = 0$.

4 Discussion

4.1 Summary

In this paper we have investigated the use of the lasso in the normal linear regression setting relating to a problem arising from estimating the trend of a single-time series that has undergone a lab change and how it compares to previously suggested methods. We modified the lasso to fit our problem task and investigated how a Bayesian analog of the lasso-type estimator behaved when preforming estimation and statistical hypothesis test. We did this by conducting a simulation study and evaluated four different methods across a defined set of performance metrics. Concerning point estimation, the results showed that the frequentist lasso showed similar behavior in bias to that of

the stepwise method, but differed significantly from the Bayesian lasso in that regard. The RMSE was reduced for all three relevant methods when noise became significant, in this case though, the frequentist lasso performed worse than both the step-wise procedure and the Bayesian lasso. In the case of statistical hypothesis testing, the type I error of the frequentist lasso also mimicked the behavior of the step-wise method - it increased up to a point where it then began to decrease towards the actual significance. Power analysis showed that the frequentist lasso inhibited lower power compared to the step-wise. The Bayesian lasso showed lower type I than both the previous methods but the cost of significantly lower power.

4.2 Further Extensions

- *Bootstrapping standard errors*: In the frequentist lasso we used the standard errors begotten from the full least squares fit. Briefly mentioned in Section 2.2, an alternative approach would have been method of bootstrapping residuals. Since data is assumed normal a parametrical bootstrapping such as described in Tibshirani et al. (2015), chapter 6 could instead been applied. Based on Efron (2012) that states how the a parametric bootstrap can be used for computation of Bayes posterior sampling distribution, one would expect this parametric bootstrap lasso to show similar performance to the Bayesian lasso and could be worth investigating in future works.

- *Inter-calibration*: In the unpublished work, Sköld (2024) also mentions how an inter-calibration procedure, where the same concentration is measured in both lab A and B yielding a correction factor $\hat{\gamma}^{cf}$ could be used. If the log concentrations in lab A and B is defined by y_A and y_B then the correction factor is given by $\hat{\gamma}^{cf} = y_B - y_A$ and can be used to improve estimating the lab coefficient γ whenever $|\gamma| > |\gamma - \hat{\gamma}^{cf}|$. In case of the lasso this could be used by penalizing on the distance from the correction factor $\hat{\gamma}^{cf}$. Then our lasso-type solution would minimize the residual sum of squares subject to $|\gamma - \hat{\gamma}^{cf}| < t$, for some $t \geq 0$. Geometrically would result in a shifting the constraint area in Figure 1 to be centered at $\hat{\gamma}^{cf}$. In the context of the Bayesian lasso this would similarly shift the location of the Laplace distributed prior on γ from 0 to the estimated correction factor $\hat{\gamma}^{cf}$.

- *Multiple time-series*: A common situation is that the monitoring programs of contaminant concentration has multiple time series which are all location-specific. In this case we could model the log-concentration with location-specific intercepts and annual trends but all having a common lab parameter γ to take into account the lab change. Extending this to the lasso framework could mean just penalizing on the parameter γ, but it could also be extended to penalize on the location

specific time trends. Relevant could even be using the generalization of the lasso known as Group lasso, Yuan and Lin (2005) which would allow all the location-specific trend parameters to be entered or removed from the model simultaneously, thus encouraging sparsity if there is no difference among the locations

References

Richard Berk, Lawrence Brown, and Linda Zhao. Statistical inference after model selection. *Journal of Quantitative Criminology*, 26:217–236, 2010.

B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26, 1979. doi: 10.1214/aos/1176344552. URL `https://doi.org/10.1214/aos/1176344552`.

Bradley Efron. Bayesian inference and the parametric bootstrap. *The annals of applied statistics*, 6(4):1971, 2012.

Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. 2004.

Jerome Friedman, Robert Tibshirani, and Trevor Hastie. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010. doi: 10.18637/jss.v033.i01.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC, 2015. ISBN 1498712169.

L. Held and D.S. Bové. *Likelihood and Bayesian Inference: With Applications in Biology and Medicine*. Statistics for biology and health. Springer, 2020. ISBN 9783662607930. URL `https://books.google.se/books?id=Cx9FzgEACAAJ`.

Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

Trevor Park and George Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103:681–686, 02 2008. doi: 10.2307/27640090.

Martin Sköld. Unpublished work. 2024.

Stan Development Team. RStan: the R interface to Stan, 2020. URL `http://mc-stan.org/`.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 00359246. URL http://www.jstor.org/stable/2346178.

Sara van Erp, Daniel Oberski, and Joris Mulder. Shrinkage priors for bayesian penalized regression. *Journal of Mathematical Psychology*, 89: 31–50, 04 2019. doi: 10.1016/j.jmp.2018.12.004.

Ming Yuan and Yi Lin. Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(1):49–67, 12 2005. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2005.00532.x. URL https://doi.org/10.1111/j.1467-9868.2005.00532.x.

Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006. doi: 10.1198/016214506000000735. URL https://doi.org/10.1198/016214506000000735.

Hui Zou and Trevor Hastie. Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320, 03 2005. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2005.00503.x. URL https://doi.org/10.1111/j.1467-9868.2005.00503.x.