

Simulation Studies

1 Simulation Studies

This chapter investigates the effect of manipulating some of the `mcmc_ridge()` inputs and thereby sampling parameters in a controlled environment, such that changes in the estimation outcome can be directly linked to respective changes in the model inputs. After an introductory exploration phase, we decided for a subset of all possible model variations that indicated the greatest potential for interesting and relevant findings.

As a result, the simulation studies discussed in this chapter will be conducted on the following components:

- The **data** input (sections 1.1, 1.2 and 1.3), which is captured by the function argument `m` or alternatively the combination of `X`, `Z` and `y`. Here, the sampler's *robustness* as well as the *shrinkage effect* of the Ridge penalty is tested across various scenarios.
- The **sample size** `n` (section 1.4). Here, we are looking for a possible stabilization process with increasing size of the input data hinting at *asymptotic/convergence* properties.
- The **hyperparameters** `a_tau`, `b_tau`, `a_xi` and `b_xi` (section 1.5) of the Inverse Gamma prior distribution of the variance parameters τ^2 and ξ^2 , as specified in the first report. The effect of hyperparameters in a hierarchical Bayesian model can be difficult to predict based on pure logical reasoning. Therefore simulations are a useful tool to either confirm prior assumptions or discover unexpected behaviour.

Since the resulting simulation studies serve different purposes (e.g. diagnostic vs. explorative), they demand for different approaches in the simulation settings, the implementation as well as the analysis and presentation of the results. For that reason, we decided against forcing all of the following sections into one common rigid framework. Instead, each section individually motivates, explains and interprets the methods chosen for its particular use case.

In order to keep the analysis in this chapter compact and succinct, there will be almost no code included. It is worth noting though that the **R Markdown** document itself as well as all **R Scripts** used for the simulations are contained in the `simulation-studies` folder of the `asp21bridge` package. Thus, each figure as well as all numerical results are fully reproducible and can be repeated and extended by the reader.

1.1 Correlated Predictor Variables

Up to this point, we have often illustrated the usage and results of the `mcmc_ridge()` sampler with simulated data from the built-in `toy_data` set. As stated in section ??, each regressor variable is independently sampled from a normal distribution and the outcome variable is simulated based on the correctly specified location-scale regression model introduced in chapter ?. All these conditions lead to an excellent performance of the `mcmc_ridge()` sampler, but might arguably not represent the most challenging task.

Sections 1.1 and 1.2 analyze the sampler's performance on simulated data, which might be closer to data found in the real world. First, we will induce correlation among the predictor variables, whereas in the second part the distributional assumptions are considerably changed. Further, the `mcmc_ridge()` performance is compared to the Maximum Likelihood based `lmls()` estimates and the Markov Chain Monte Carlo `mcmc()` sampler without penalty from the `lmls` package.

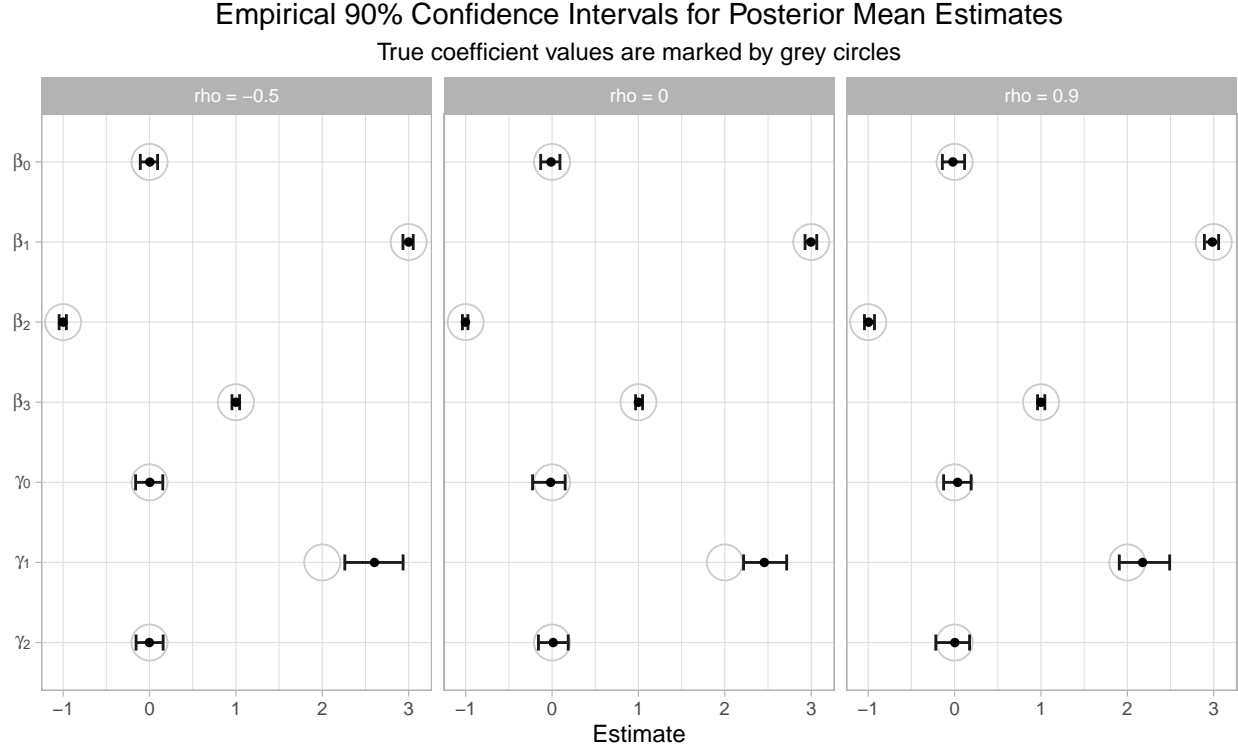


Figure 1: Comparison of Correlation Structures - 50 Simulation Cycles

Simulation Setting

Briefly stated, we simulated data from a three dimensional multivariate normal distribution with three different values for the pairwise correlation ρ . Then, we fitted the three models `mcmc_ridge()`, `mcmc()` and `lmls()`, where the number of simulations was set to 10000 for the Bayesian Samplers. The exact simulation setting is described in full detail in the Appendix.

Moreover, we compared the performance of the classical `mcmc_ridge()` implementation, which draws β from the closed form full conditional (multivariate normal) distribution, with an alternative sampling process that uses a Metropolis-Hastings approach for both, the location parameter β as well as the scale parameter γ .

Simulation Results

All three models perform very similarly for each correlation structure: The estimates for β are almost exactly equal to the true values, whereas there are slight deviations for the γ vector. These errors, however, are consistent in all three sampling scenarios and, thus, do not depend on the strength nor the direction of the correlation between the covariates. Due to this similarity between the three models, the remaining analysis of this simulation study is dedicated to the `mcmc_ridge()` model.

There is one additional aspect to be noted: The performance of the `mcmc_ridge()` function is considerably worse, when β is sampled by means of the Metropolis-Hastings algorithm. Although the acceptance rates of the proposed values are in acceptable ranges for a random walk proposal (between 35% and 40%), the chains are strongly correlated and did not fully converge even after 10000 iterations. For that reason, we limit the Metropolis-Hastings sampling process for β to this one example and will focus on the classical `mcmc_ridge()` implementation in the remaining parts of chapter 1.

The corresponding graphical display for the comparison between the three models and between the two sampling procedures for β is omitted here for brevity reasons, since the findings are not that surprising. The code as well as all plots can be found and fully reproduced in the `regressor-correlation.R` file located

inside of the `simulation-studies` folder of the project directory.

In order to make any conclusions about bias and variance, the above procedure is repeated 50 times with 1000 simulations each. The black points in Figure 1 represent the mean of these 50 Posterior Mean estimates from the penalized Ridge sampler. For a better visual comparison, the true values for each coefficient are indicated by grey circles. Since we cannot rely on distributional theory for the standard errors, the variability of the estimates is displayed by nonparametric ‘confidence’ intervals, which are simply given by the range from the empirical 0.05 quantile to the 0.95 quantile of the 50 estimated values.

These intervals are very narrow and centered around the true value in all cases except for γ_1 , where the estimated are biased upwards for all correlation structures. This result is somewhat surprising, since the Ridge penalty might suggest a bias towards zero. Although the `mcmc_ridge()` function does indeed show the greatest variability across many different scenarios for the γ vector which is sampled via Metropolis-Hastings, this upward bias is not consistent and might be induced by chance in this case.

Further, it is worth noting that the results are partially affected by the standardization of the covariates: Although the behaviour illustrated here is quite similar overall to the performance for unstandardized data (as used for the second report), there are two differences:

- The variability of the Posterior Mean estimates of β_0 is significantly reduced. This indicates a greater sampling stability for the standardized case.
- The bias of the Posterior Mean estimates of γ_1 is not present, when the design matrices are not centered and scaled (referring to the second report). This again suggests that the finding in the standardized case is due to randomness, since there is no convincing explanation why upward bias should be connected to scaled covariates.

1.2 Challenging the Model Assumptions

This simulation study is structured in a very similar way to the study considered in section 1.1. Instead of varying the correlation structure among the regressors in the underlying data set, both the regressors and the outcome variable y are sampled from distributions that are more challenging for estimation than the normal distribution.

Simulation Setting

The covariates of the design matrices \mathbf{X} and \mathbf{Z} are generated by a mix of distributions including the Normal, Exponential, Uniform, Bernoulli and t - distribution. Given these values, the outcome variable \mathbf{y} is drawn from three different distributions (Normal, t and Uniform) corresponding to three separate scenarios, which are compared in the following analysis. As usual, the full simulation design is stated in the Appendix.

Note that the `lmls()`, `mcmc()` and `mcmc_ridge()` models are built upon the assumption of a Gaussian distribution for \mathbf{y} . Hence, we expect all three estimation procedures to perform well under the first outcome specification, which they were designed for. The remaining two cases analyze the performance in presence of a mild (t distribution) and a moderately strong (Uniform distribution) violation of this model assumption.

Simulation Results

Just as in section 1.1 the results of one complete simulation cycle (each of the $3 \cdot 3 = 9$ models was fitted once / each data point represents one estimate) are displayed in Figure 2. Note that the second facet is labeled by $y \sim t$, although it is formally sampled from an affine transformation of a t -distributed random variable, which does not follow an exact t distribution.

The differences within each facet as well as between the facets are significant. All three models seem to estimate the β vector well, when there are no or only mild violations of the normal assumption for y . If y is sampled from a uniform distribution, there are major differences for β_0 , β_2 and β_4 (notice the extended x -scale in the third facet). Interestingly, the γ vector is estimated very well in the latter case with more deviations in the setting, where y is based on the t distribution.

Posterior Means / MLE for (misspecified) Regression Models

True coefficient values are marked by grey circles

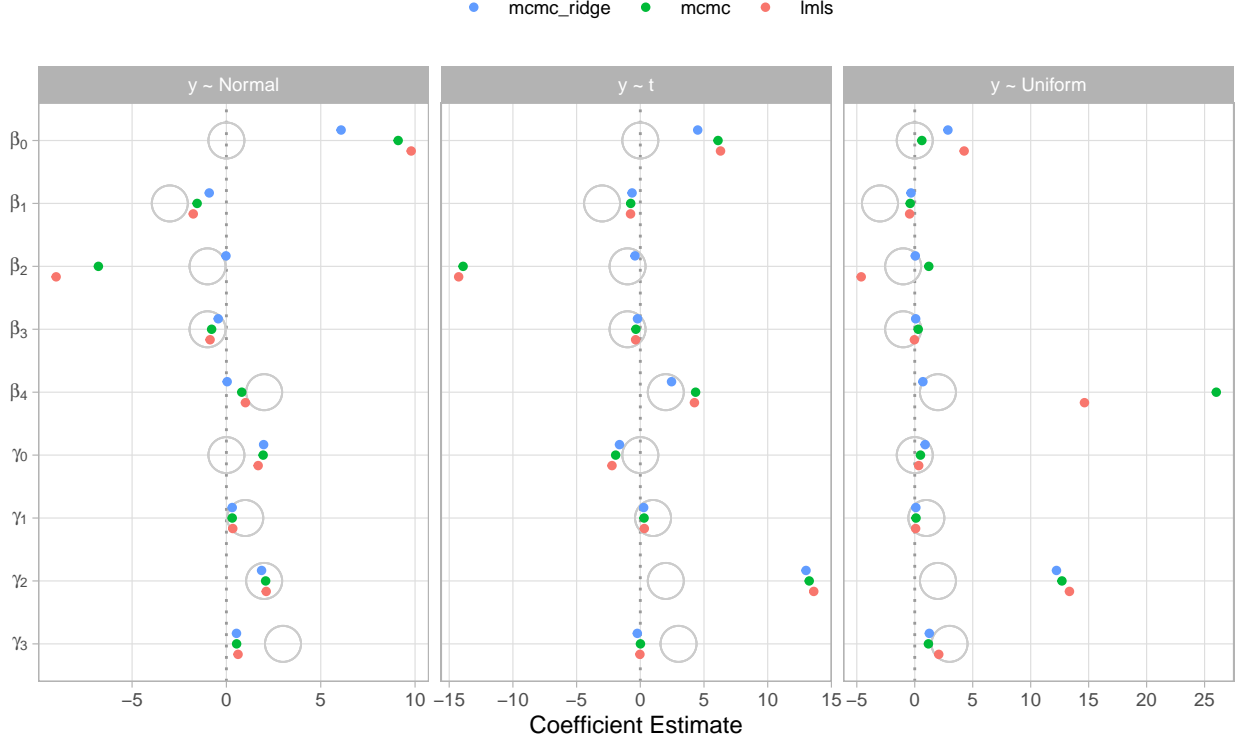


Figure 2: Comparison of Outcome Distributions - One Simulation Cycle

To gain insights beyond this single simulation cycle, which could well be disturbed by random noise, we repeat the sampling process 50 times. The resulting means as well as empirical confidence intervals (analogous to section 1.1) are plotted in Figure 3.

This plot (literally) paints a drastically different picture, emphasizing the necessity of repeating experiments multiple times whenever possible. Across all 50 simulations the deviation of the estimates for β_2 (corresponding to the regressor variable from the exponential distribution) is huge for all three distributional specifications of y .

The small bias induced by all three models is negligible compared to the wide confidence intervals, which is particularly interesting, when y stems from a normal distribution. In this case all models should perform well, however the `lmls()` and the `mcmc()` Posterior Mean / Maximum Likelihood estimates vary wildly across the simulation cycles. A similar effect can be observed for β_0 in case of the uniform distribution. Here, all models overestimate the true value on average, while the `mcmc_ridge()` function again shows the smallest variability. In contrast, the estimates for γ_0 in the right facet are fairly stable across simulation cycles, but also consistently wrong at the same time.

Estimates of the bias and the standard error of the Posterior Mean / Maximum Likelihood estimates can be more distinctly compared by their numerical values provided in Tables 1 and 2. In order to emphasize the interesting/differing entries, both tables only include a subset of the estimated coefficients.

Considering the bias estimates in Table 1 first, there are no obvious patterns that would suggest the superiority of one model. Further, none of the three models tend to only over- or underestimate the true coefficient values. The most interesting entries are the bias estimates for β_0 and γ_0 in the uniform case, where all three models agree to significantly overestimate. However, the intercept coefficients are often of minor interest.

Empirical 90% Confidence Intervals for Posterior Mean Estimates

True coefficient values are marked by grey circles

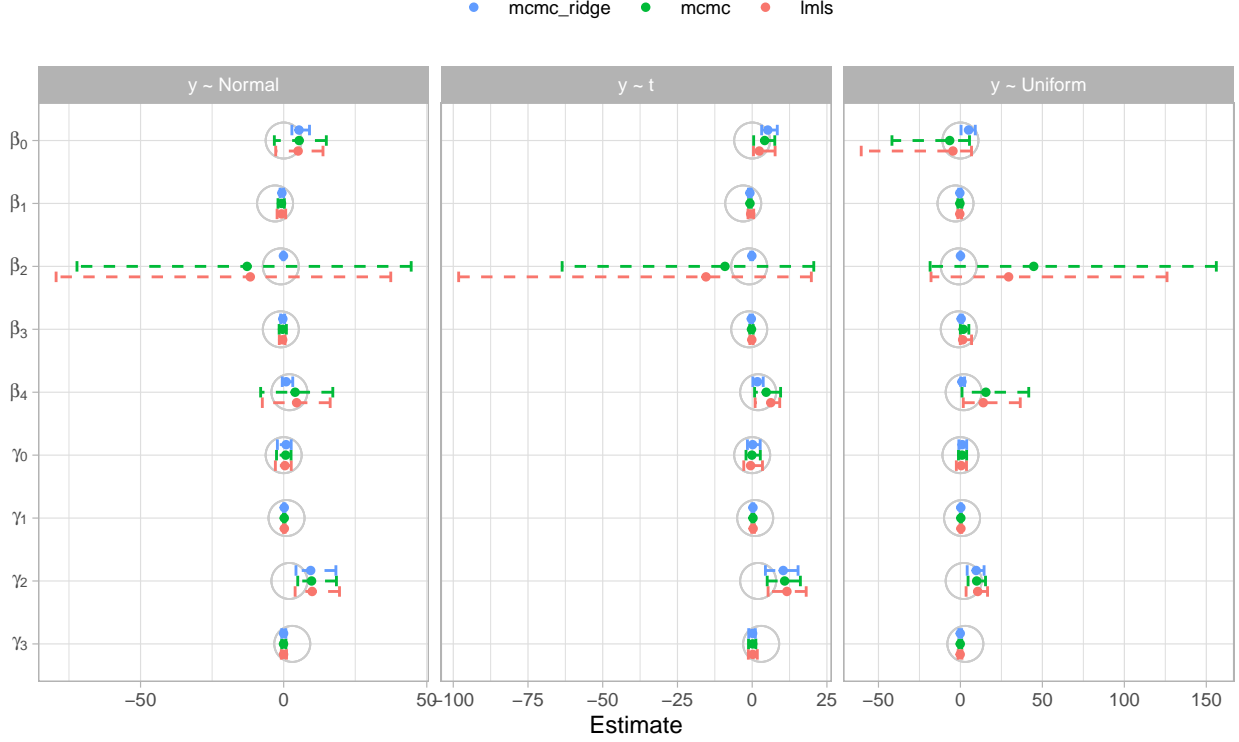


Figure 3: Comparison of Outcome Distributions - 50 Simulation Cycles

The standard error estimates displayed in Table 2 clearly indicate the worst performance of the `lmls()` and the `mcmc()` model for β_2 . In almost all cases (and sometimes very significantly), the `mcmc_ridge()` sampler has the smallest standard error. This finding nicely confirms the underlying mathematical theory: The present prior specifications in the Bayesian setting, which induces the equivalent form of a frequentist Ridge penalty, can lead to biased estimation.

However, this loss in accuracy can be (as it is in this case) dominated by the gain in precision by the shrinkage effect of the penalty. Note that (except for γ_0 in the most right facet in Figure 3) the `mcmc_ridge()` sampler slightly *overestimates* coefficients with true *negative* values and *underestimates* those with true *positive* values. This again is caused by the Ridge penalty leading to estimated coefficients close to zero.

In summary, the following conclusions can be drawn:

Table 1: Bias of Coefficient Estimates

| | Normal | | | t | | | Uniform | | |
|------------|--------|--------|------------|--------|-------|------------|---------|-------|------------|
| | lmls | mcmc | mcmc_ridge | lmls | mcmc | mcmc_ridge | lmls | mcmc | mcmc_ridge |
| β_0 | 5.05 | 5.40 | 5.35 | 2.38 | 4.19 | 5.22 | -4.57 | -6.54 | 5.00 |
| β_2 | -10.66 | -11.79 | 0.93 | -14.46 | -8.11 | 0.88 | 30.40 | 45.71 | 1.05 |
| β_4 | 2.52 | 2.00 | -1.15 | 4.22 | 2.69 | -0.26 | 11.93 | 13.52 | -1.06 |
| γ_0 | 0.43 | 0.70 | 0.84 | -0.53 | -0.10 | 0.14 | 0.48 | 0.98 | 1.10 |
| γ_2 | 7.95 | 7.74 | 7.43 | 9.62 | 8.86 | 8.39 | 8.56 | 7.93 | 7.60 |

Table 2: Standard Errors of Coefficient Estimates

| | Normal | | | t | | | Uniform | | |
|------------|--------|-------|------------|-------|-------|------------|---------|-------|------------|
| | lmls | mcmc | mcmc_ridge | lmls | mcmc | mcmc_ridge | lmls | mcmc | mcmc_ridge |
| β_0 | 6.33 | 5.61 | 2.15 | 12.61 | 2.97 | 1.78 | 20.81 | 18.38 | 3.01 |
| β_2 | 37.98 | 37.45 | 0.18 | 48.35 | 28.81 | 0.27 | 46.32 | 60.91 | 0.12 |
| β_4 | 7.61 | 7.44 | 1.11 | 12.11 | 3.21 | 1.16 | 21.25 | 19.46 | 0.92 |
| γ_0 | 1.65 | 1.49 | 1.38 | 1.95 | 1.49 | 1.37 | 2.01 | 1.64 | 1.53 |
| γ_2 | 5.17 | 4.64 | 4.42 | 4.38 | 3.86 | 3.62 | 4.46 | 3.80 | 3.64 |

1. All three models are affected by changes in the regressor and/or outcome distributions. In the former case the regressor variables sampled from the Exponential and the Bernoulli distributions were the greatest challenge, in the latter case the outcome variable from the Uniform distribution. This is generally not surprising, since these distributions deviate most from the nicely behaved normally distributed case.
2. As expected, the `lmls()` function is affected strongly by violating the model assumptions, since its estimation process is based on the normal (log) Likelihood. Surprisingly, the `mcmc()` sampler without penalty often did not perform much better.
3. While the `mcmc_ridge()` function does not excel at estimation accuracy, it does lead to the most stable estimation with smallest standard errors in the vast majority of cases. As emphasized above, this behaviour nicely agrees with the mathematical theory of Ridge penalization.
4. Had we not conducted repeated experiments, our conclusions would have been quite different. Simulation results are therefore always worth repeating many times to consolidate the correct interpretations.

Technical Aspects

As outlined in the previous paragraph, a total of $50 \cdot 3 \cdot 3 = 450$ models were fitted to analyze the performance differences. In order to speed up the involved computations of this specific and some of the other simulation studies in this report, we used the *parallel computing* capabilities of R.

There are many options from various packages to choose from. We decided to use the `furrr` package, which is built on top of the `future` package specialized on parallel processing. As the name suggests, `furrr` provides a convenient way to use many functions from the popular `purrr` package, while using multiple cores at the same time. This *functional programming* based approach (similar to the `apply()` family in ‘base R’) is particularly well suited for simulation studies and provides some structural as well as minor performance advantages compared to the classical `for`-loop approach.

The following (slightly modified) code snippet provides a brief insight into the implementation:

```
plan(multisession, workers = 8)

full_results <- tibble(id = 1:50) %>%
  mutate(samples = future_map(
    .x = id,
    .f = ~ show_results(n = 50, num_sim = 1000),
    .options = furrr_options(seed = 1)
  ))
```

The `plan()` function borrowed from the `future` package initializes the parallel computing process and determines the number of cores/workers available for computation. The `show_results()` helper function fits all three models `mcmc_ridge()`, `mcmc()` and `lmls()` for each outcome distribution in a single simulation cycle.

This entire procedure is repeated 50 times in parallel using the `future_map()` function from the `furrr` package, where the results of all 450 models are saved in a well organized structure inside of a list column. This new column of the data frame contains complete information about all simulations, such that any required element for the further analysis can be easily extracted and post processed.

Finally, the `.options()` argument allows the specification of a random seed. Random number generation in the context of parallel computing is slightly more involved compared to the sequential approach. This additional complexity is automatically handled by the `future_map()` function, such that all results are sampled in a statistically valid and fully reproducible manner.

1.3 Redundant Covariates

Similar to the simulation study from section 1.1, this section investigates the sampler’s behaviour in presence of strong *pairwise* correlation among the covariates. However, in contrast to all previous simulation studies, we add additional *redundant* regressors to the model, that were not used for generating the outcome variable \mathbf{y} . The statistical theory of (Ridge) penalization suggests, that the coefficient estimates from the `mcmc_ridge()` function are affected by the shrinkage effect induced by the coefficient’s prior distributions in presence of high (compared to the sample size n) number of correlated covariates. Thus, the magnitude of the estimated β vector should be smaller compared to models without regularizing penalty component.

Simulation Setting

The model contains 10 pairs of covariates drawn from a multivariate normal distribution with correlation $\rho = 0.9$. For each pair, the first covariate contributes to generating the outcome variable \mathbf{y} with a true coefficient value of 1. The second covariate has no impact on \mathbf{y} , but is included nonetheless in the model fitting stage. Similar to section 1.2, the `lmls()`, `mcmc()` and `mcmc_ridge()` functions are used to generate estimates.

The exact design of this simulation study can be found in the Appendix, whereas the following paragraph focuses on the analysis of the obtained results.

Simulation Results

This study exclusively focuses on the location parameter β . There are two aspects of interest we wish to investigate further:

1. Can each model differentiate between the relevant and the redundant covariate for each pair of regressor and, thus, isolate the true effect on \mathbf{y} ?
2. Due to the high number of 20 covariates compared to the low sample size of 50 observations and the presence of correlated and partially redundant regressors, the setup of this study is prone to overfitting. Can we therefore observe the hypothesized shrinkage effect on the coefficient estimates for the `mcmc_ridge()` model compared to the models without penalty?

Figure 4 illustrates the results obtained by fitting all three models. We start with the answer to the first question. For each of the three models and each coefficient pair the absolute value of the coefficient corresponding to relevant regressor is larger than the magnitude of the paired quantity. Thus all three models always contributed the major impact on the outcome variable to the correct covariate. Further, the differences between estimations for those coefficients with true value 1 and those with true value 0 do not differ in an *obvious* manner across the models. Since the variance of the estimates seems to be largest for the `lmls()` model and smallest for the `mcmc_ridge()` model, the ratio (instead of the difference) tends to be largest for the regularized model.

The reduced variance of the `mcmc_ridge()` model can be explained by the Ridge penalty: Since all estimates are shrunken towards zero, fewer ‘outlier’ estimates that are far away from zero will be observed. This property is beneficial for the red points, since the estimates clutter around their true value in this case. However, the coefficients corresponding to the green points are penalized as well such that a bias towards

Posterior Means / MLE for pairwise correlated Covariates

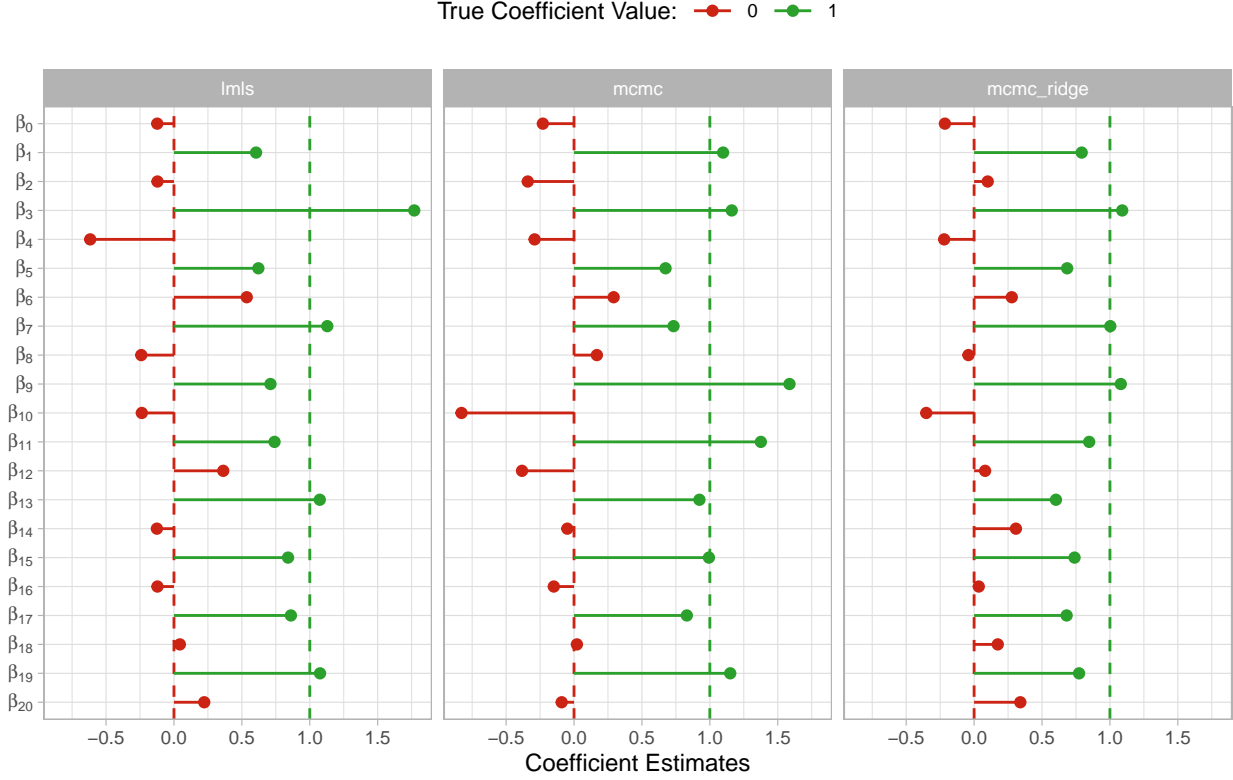


Figure 4: Shrinkage Effect of Ridge Penalty in Presence of Redundant Covariates

zero is induced, which is not present for the two unpenalized models. Hence, the hypothesized shrinkage effect can indeed be observed.

It is insightful to compare the observed effect from the Ridge penalty to the expected effect of the *LASSO* penalty. The LASSO penalty encourages *sparse* solutions: each coefficient is assigned the same weight such that shrinkage of large estimates is not preferred over shrinkage of small estimates. Thus, small estimates are often set to exactly zero, while large estimates, although reduced, can still be significantly different from zero.

In contrast, the Ridge penalty generally does not induce sparsity. In fact, quite the opposite can be the case: Small coefficient estimates such as β_{14} in Figure 4 are sometimes *increased* in magnitude compared to the unpenalized models. Large estimates such as β_4 or β_6 , however, are reduced in size quite heavily. The Ridge penalty tries to put all coefficients on the same scale and, thus, prioritizes shrinkage of large estimates compared to small ones!

In order to quantify the observed shrinkage effect from Figure 4 numerically, we calculate the sum of squared coefficient estimates, i.e. the squared Euclidean norm $\|\beta\|^2$, for each model. Moreover, these magnitudes are compared separately for all coefficients corresponding to relevant regressors (those with an *odd* subscript) and those corresponding to redundant regressors (*even* subscript). Table 3 summarizes the results.

A perfect model that estimates a coefficient value of 0 for all 10 unnecessary covariates and a value of 1 for all relevant covariates would have achieved squared vector norms of 0 and 10, respectively. The `lmls()` model is very close to the perfect value for the second group, indicating that the estimated values are quite precise *on average*. However, as we have seen in Figure 4, this conclusion does not transfer to the individual estimates due to the significant variability among the estimates.

The most interesting and reassuring finding is obtained from the `mcmc_ridge()` vector norms: In agreement

Table 3: Squared Euclidean Norms of Coefficient Estimates

| | True Value = 0 | True Value = 1 |
|------------|----------------|----------------|
| lmls | 1.03 | 9.97 |
| mcmc | 1.23 | 11.80 |
| mcmc_ridge | 0.56 | 7.15 |

with the visual illustration, the Ridge regularization effect can be numerically detected for all coefficient estimates, independent of the true value. While this property induces a bias that is not desired for the second column (underestimating the true values on average), it does indeed prevent overfitting by shrinking the coefficients corresponding to the redundant covariates, which is indicated by the lowest value in the first column. In that sense, the Ridge penalty increases the model *robustness* in presence of a high number of (correlated) regressors and often leads to an overall lower Mean Squared Error due to the reduction in variance.

1.4 Sample Size

This simulation study analyzes the effect of the sample size \mathbf{n} on the means of the posterior distribution for the coefficients of β and γ . There are two main goals of this simulation study: On the one hand, we want to investigate whether the posterior means of large samples are closer to the true values than the posterior means of small samples. On the other hand, we want to analyze whether the `mcmc_ridge()` penalty affects the location of the posterior means.

Simulation Setting

- The design matrix $\mathbf{X} = (\mathbf{1}_n \quad \mathbf{x}_1 \quad \mathbf{x}_2)$ contains two independently sampled regressor variables plus one intercept column:
 - $\mathbf{x}_1 \stackrel{iid}{\sim} \mathcal{N}(1, 1)$,
 - $\mathbf{x}_2 \stackrel{iid}{\sim} \mathcal{N}(2, 1)$.
- The design matrix $\mathbf{Z} = (\mathbf{1}_n \quad \mathbf{z}_1 \quad \mathbf{z}_2)$ is structured in the same way with the regressor variables:
 - $\mathbf{z}_1 \stackrel{iid}{\sim} \mathcal{N}(1, 1)$,
 - $\mathbf{z}_2 \stackrel{iid}{\sim} \mathcal{N}(2, 1)$.
- The true coefficient vectors are given by $\beta = (\beta_0 \quad \beta_1 \quad \beta_2)^T = (1 \quad -1 \quad 4)^T$ and $\gamma = (\gamma_0 \quad \gamma_1 \quad \gamma_2)^T = (0 \quad -0.5 \quad 1)^T$.
- The posterior means are analyzed with respect to 6 different sample sizes: $\mathbf{n} \in \{0, 50, 100, 200, 300, 500\}$.
- In the next step, the outcome vector $y \in \mathbb{R}^n$ is simulated and passed to the `mcmc_ridge()` function with `nsim = 500` simulations.
- To make the results more stable, the above procedure is repeated 100 times. For each coefficient, the mean value of the Posterior Mean estimates of each coefficient is calculated as well as the Mean Absolute Error (MAE) with respect to the true values of β and γ .

Simulation Results

The means of the Posterior Mean estimates are displayed in Figure 5. For larger sample sizes ($\mathbf{n} \geq 200$) none of the six parameters are extremely biased.

Moreover, for $\mathbf{n} = 30$, β_0 and β_2 are significantly biased, which might be caused by the high `mcmc_ridge()` penalty for $\beta_2 = 4$. The significant bias of β_0 might be explained by a counteract of the β_2 bias.

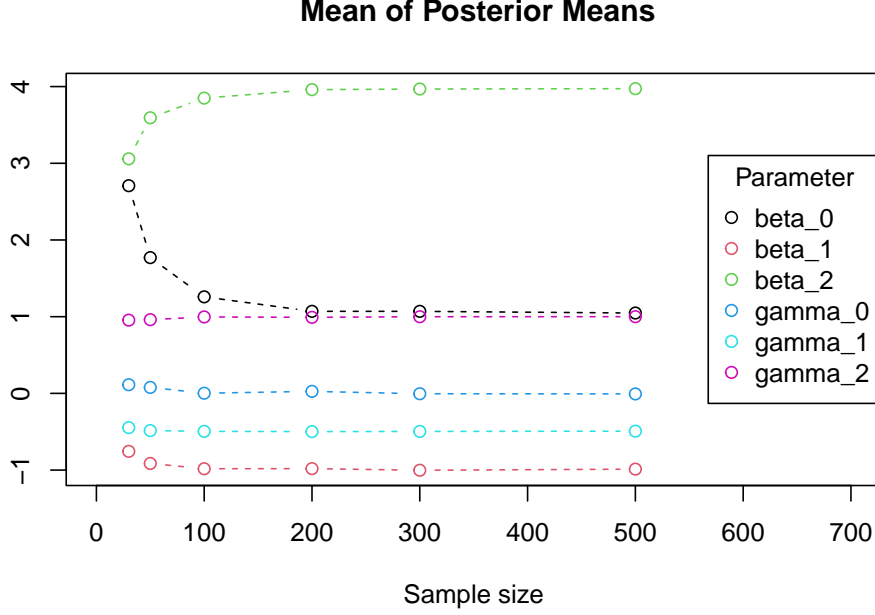


Figure 5: Mean value of 100 Posterior Mean Estimates

After getting an impression about empirical biases of the coefficients, we now focus on the variability of the posterior means of the coefficients, which are measured by the *MAE* based on the results of the 100 repetitions. Figure 6 points out that the posterior means of β_0 have significantly larger errors than the posterior means of β_2 for $n = 30$. However, this might also be explained by the fact that for $n = 30$, β_0 has a greater empirical bias than β_2 as could be observed in Figure 5.

In addition, for increasing sample sizes, the *MAE* of the Posterior Means tend to zero for all coefficients except β_0 . Nevertheless, also the errors of β_0 seem to become smaller with increasing sample size.

1.5 Hyperparameters

In the past, we have been sampling data with the `mcmc_ridge()` function without having a closer look on the effect of the hyperparameters and model inputs `a_tau`, `b_tau`, `a_xi` and `b_xi`. However, they affect the Full Conditional Distributions of τ^2 and ξ^2 , as stated in sections ?? and ?? in chapter ??.

Moreover, the mean vector $\boldsymbol{\mu}_{\beta_{\text{eta}}}$ and covariance matrix $\boldsymbol{\Sigma}_{\beta_{\text{eta}}}$ of the $\boldsymbol{\beta}$ vector both depend on τ^2 and, thus, implicitly on the hyperparameters a_τ and b_τ (see section ??). Analogously, section ?? illustrates the direct effect of the Full Conditional distribution of γ on ξ^2 , which in turn depends on the hyperparameters a_ξ and b_ξ .

Finally, cross effects can be observed, since $f(\boldsymbol{\beta} \mid \cdot)$ depends on γ through the quantities \mathbf{W} and \mathbf{u} as defined in chapter ?? and $f(\gamma \mid \cdot)$ directly depends on $\boldsymbol{\beta}$. These dependencies are reflected in the `mcmc_ridge()` sampler by the iterative sampling procedure which is discussed in great detail in the previous sections ?? and ??.

Thus, the hyperparameter choice of a_τ , b_τ , a_ξ and b_ξ inevitably impacts the result of *all* coefficient estimates contained in the model in a nontrivial way, such that pure analytical reasoning might be misleading. For this reason, this section investigates these effects based on a simulation approach.

Simulation Setting

- The design matrix $\mathbf{X} = (\mathbf{x}_1 \quad \mathbf{x}_2)$ is simulated from a two dimensional normal distribution $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean vector $\boldsymbol{\mu} = (1 \quad 2)^T$ and identity covariance matrix $\boldsymbol{\Sigma} = \mathbf{I}_2$. The same holds true for the

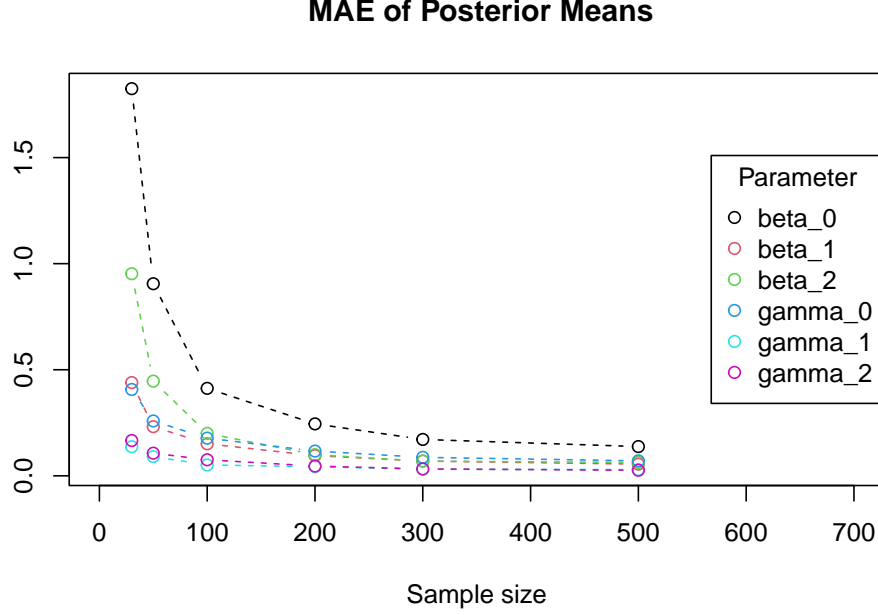


Figure 6: Mean Absolute Error Estimates

design matrix $\mathbf{Z} = (\mathbf{z}_1 \quad \mathbf{z}_2)$ with mean vector $\boldsymbol{\mu} = (5 \quad 3)^T$ and identity covariance matrix.

- In both design matrices intercept columns are added for estimation purposes. The true coefficient vectors are given by $\boldsymbol{\beta} = (\beta_0 \quad \beta_1 \quad \beta_2)^T = (0 \quad -1 \quad 4)^T$ and $\boldsymbol{\gamma} = (\gamma_0 \quad \gamma_1 \quad \gamma_2)^T = (0 \quad -2 \quad 1)^T$.
- For sampling the location parameter, the full conditional multivariate normal distribution of $\boldsymbol{\beta}$ is chosen, i.e. `mcmc_ridge(..., mh_location = FALSE)` is used. Therefore, the location estimate is directly affected by the hyperparameters.
- For simulating the influence of the hyperparameters, nine different values are chosen: $a_\tau, b_\tau, a_\xi, b_\xi \in \{-1, 0, 0.5, 1, 2, 10, 50, 100, 200\}$. Since for statistical properties like the mean of an Inverse Gamma distribution $\frac{b}{a-1}$ the condition $a > 1$ is required, particular attention is given to larger values. However, it is an aim to inspect the performance of the sampler for smaller hyperparameter values than 1 as well.

Simulation Results

The first two plots of Figures 7 and 8 display the absolute deviations of the Posterior Mean estimates from the true parameters with the stated different values for a_τ and b_τ . For each estimate, the Posterior Mean averages over 1000 simulations of the `mcmc_ridge()` sampler. Note, that location and scale parameters are plotted separately, according to the relationship mentioned above. For a better overview, the dotted line displays the linear trend of all estimate deviations.

The x -axis is transformed by a pseudo logarithm in order to clearly visualize the deviations in the range of -1 to 10 , which would not be possible on original scales. Since -1 and 0 are also part of the hyperparameter values, the `pseudo_log_trans()` function of the `scales` package is applied, log-transforming positive values only.

It can be observed, that the intercept estimates in each plot show the largest deviations from their true value. In Figure 7, however, the overall deviations of $\boldsymbol{\beta}$ estimates from their corresponding true value are small in absolute value. In contrast, deviations of the $\boldsymbol{\gamma}$ estimates in Figure 8 are fairly significant, especially for γ_0 .

The functional chain that applies to the estimates of $\boldsymbol{\beta}$ can be described by the effect of the mean of the inverse gamma distribution on τ^2 : A larger value for b_τ leads to larger values of τ^2 , which are again affecting the full posterior parameters of $\boldsymbol{\beta}$ and, thus, potentially increase the absolute deviation of the corresponding

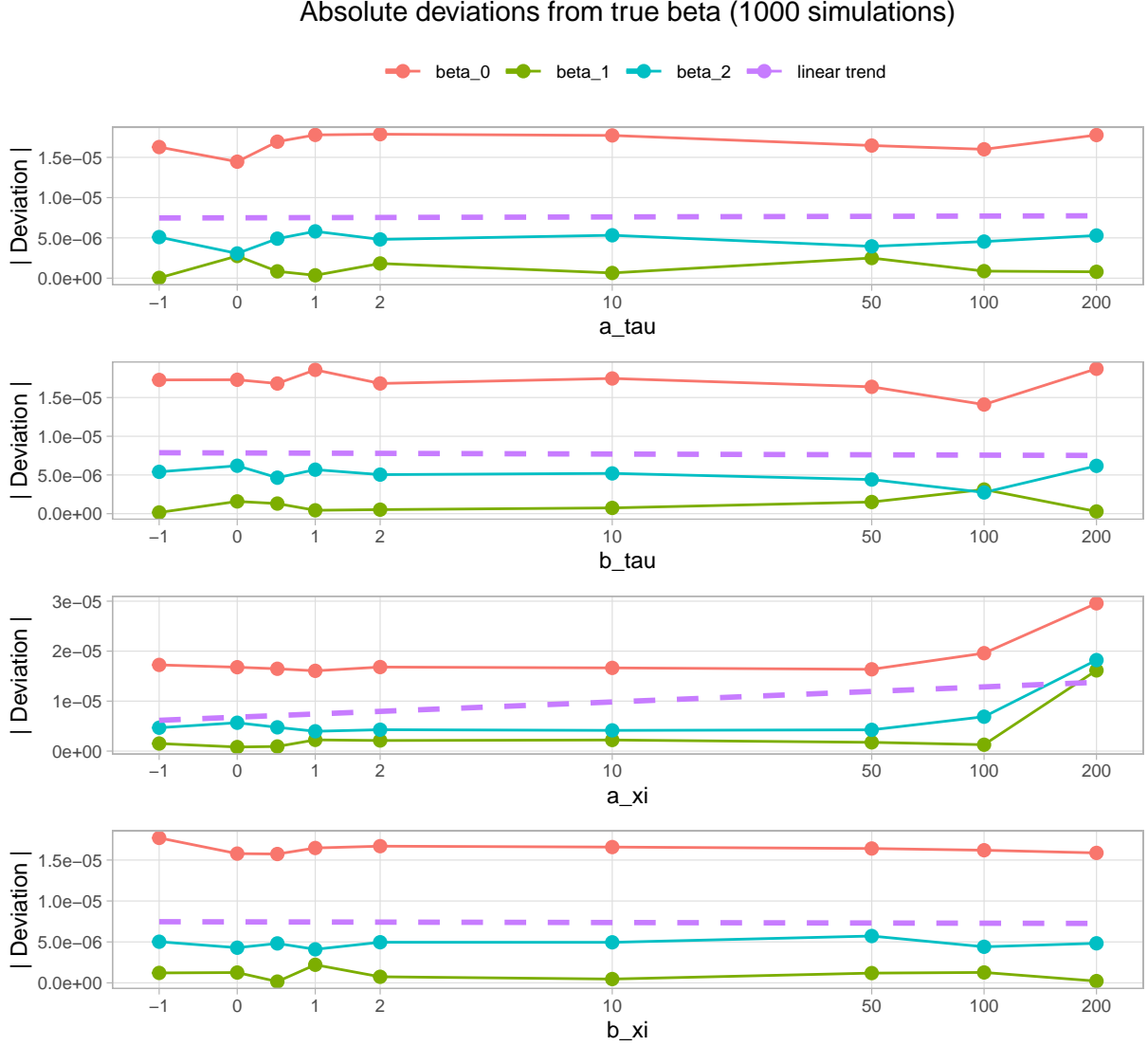


Figure 7: Comparison of the absolute deviations of beta parameters

estimates from their true values. a_τ causes the opposite effect. This numerically observable effect, however, is hidden by the overall small deviation in the first two plots of Figure 7.

It is remarkable, that the deviation of β estimates is smallest when $a_\tau, b_\tau \in \{50, 100\}$. For values of $a_\tau \leq 1$, one obtains wider variances of absolute deviations, since the Posterior Mean requires values larger than one.

In the upper two plots of Figure 8, there is no clear impact of τ^2 and its parameters. Rooted in no direct effect of τ^2 on γ according to our underlying mathematical model, one observes cross-effects through the sampling procedure of the `mcmc_ridge()` sampler, where the full posterior $f(\gamma | \cdot)$ depends on β .

Anyway, our sampler produces the lowest deviation of γ estimates for $a_\tau, b_\tau \in \{0, 0.5, 200\}$, where 0.5 is chosen by coincidence for a_τ here, since wide variations for $a_\tau \leq 1$ of absolute deviations are observable again.

The lower two plots of Figures 7 and 8 are constructed analogously, but showing the impact of a_ξ and b_ξ on the location and scale parameters respectively. Again, the overall absolute deviations for the β estimates from their true values are small, whereas the deviations for the γ estimates are considerably larger. Once again, the intercept estimates display the largest deviations from their true value.

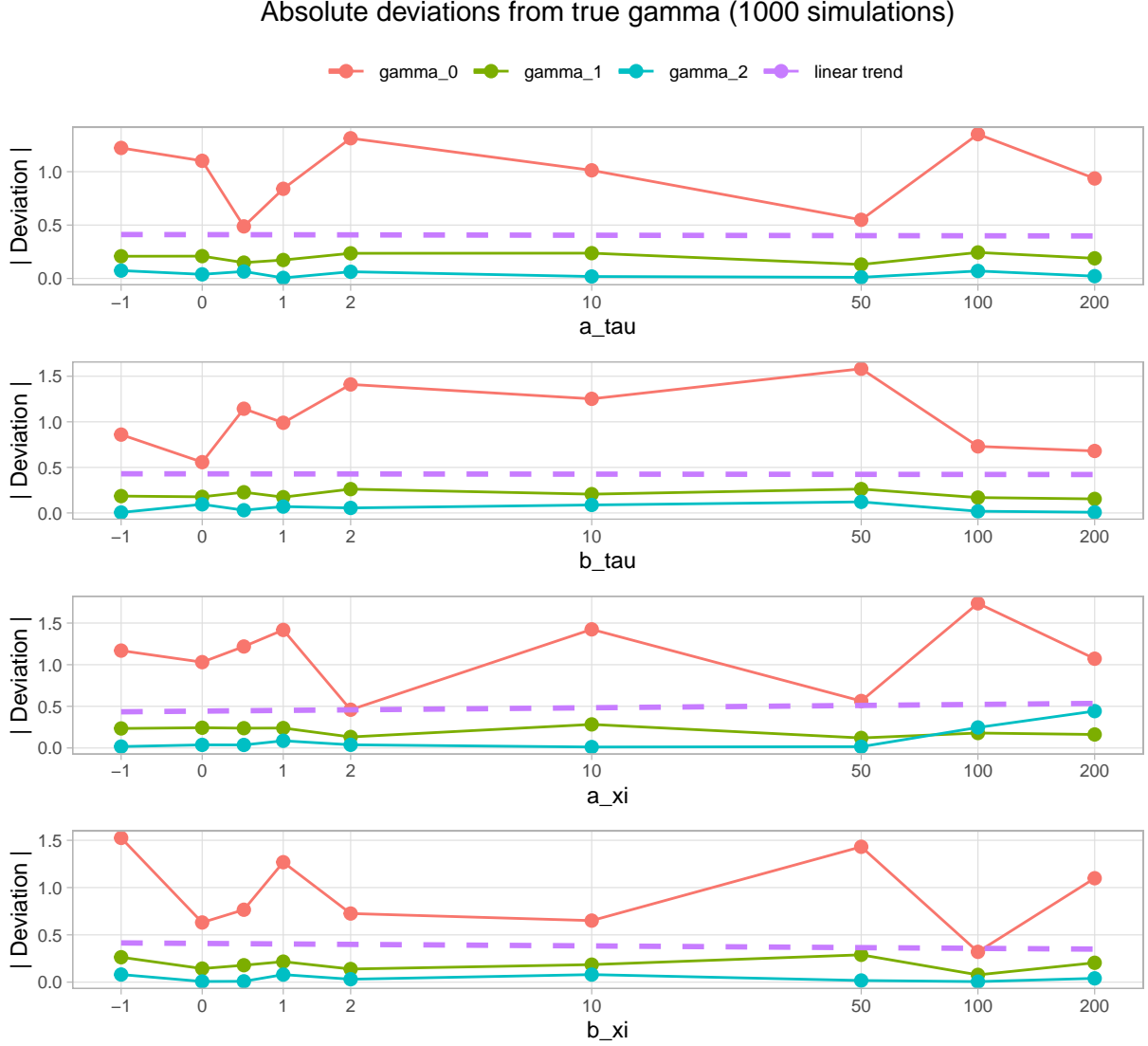


Figure 8: Comparison of the absolute deviations of gamma parameters

Arguing with the mean of the Inverse Gamma distribution of ξ^2 in a similar way, one obtains larger mean values for b_ξ , while a_ξ lowers them. The impact of ξ^2 on γ is assumed to decrease $f(\gamma | \cdot)$ according to our underlying theoretical model. This effect is indicated by the linear trend lines in the second half of Figure 8.

In general, one obtains smaller deviations for larger values of b_ξ and lower ones of a_ξ , where especially lots of randomness occurs in the deviations of γ_0 . Therefore, the impact of a_ξ and b_ξ on the scale intercepts is overshadowed by the randomness induced by the Metropolis Hastings algorithm. The same wide variations exclusively for $a_\xi \leq 1$ cannot be obtained in the same manner as for a_τ .

The sampler exhibits the best results for the *scale* estimates for $a_\xi = 2$ and $b_\xi = 100$. However, due to the wide overall variation, these results must be taken with care.

The effect of a_ξ and b_ξ on β can be explained through the cross-effects of the matrix \mathbf{W} and the vector \mathbf{u} introduced at the beginning of this section, both containing γ . These diminish with increasing values of the γ entries.

The matrix \mathbf{W} affects the variance of the full conditional distribution of β negatively, while the mean is positively affected. Hence, larger values of a_ξ cause higher Posterior Means of the location parameters. The

positive linear trend in the second half of Figure 7 for values of a_ξ is particularly interesting. For values of b_ξ , the trend comes off inferior. The wider variations of deviations for $a_\xi \leq 1$ is again not observable here. Nonetheless, the randomness observable for scale estimates does not show up for location estimates anymore.

The smallest deviations of the *location* estimates can be detected for $a_\xi = 1$ and $b_\xi = 200$.

Shortly noted, the acceptance rates of the Metropolis Hastings algorithm for sampling γ are always between 0.31 and 0.53. For the value range of a_τ , b_τ and a_ξ , no distinct pattern is observable in this regard. With growing values of b_ξ , however, acceptance rates are more likely to grow. Since the acceptance rates are in reasonable ranges enabling statistically valid estimation, these results are not further investigated here.