AST 5731: Project 2

Aritra Banerjee, Alex Granados, Kiet Pham, Sarah Taft

November 24, 2022

1 Introduction

The Stefan-Boltzmann law is a well-defined astrophysical descriptor of blackbody radiation that relates stellar luminosity to stellar temperature and radius. Formally, the expression takes the form:

$$L = 4\pi R^2 \sigma T^4 \tag{1}$$

By taking the log of both sides and assuming stellar radii are constant, Equation 1 can be re-expressed as:

$$log(L) \propto 4log(T) \tag{2}$$

This is a linear relationship, which makes sense as Equation 1's original form is a power law. As such, the purpose of this project is to use observational data of stellar systems to model the linear relationship of the log-transformed Stefan-Boltzmann law via Bayesian Normal Regression. Section 2 will detail the data and filtering techniques employed for this analysis. Sections 3 and 4 will discuss the results of the Normal Regression, excluding and including the associated parameter errors in the analysis respectively, and assuming the proportionality given in Equation 2. Section 5 will discuss some of the limitations of these analyses, and Sections 6 and 8 will explore model improvements and associated results to address these limitations.

2 Data & Motivation

Data were obtained from DEBCat, a catalog summarizing the physical properties of well-studied eclipsing binary star systems [1]. Numerous physical properties are summarized in this catalog, however the only ones utilized in our analysis are the log of effective stellar temperature, the log of solar normalized stellar luminosity, stellar radius, and associated errors. We note that errors for stellar temperature and luminosity were given in log space, so transforming those data was not necessary.

While DEBCat includes physical properties for both stars in each binary, we chose to only use data from one star in each system. The left panel of Figure 1 depicts the data after systems missing luminosity or temperature information were filtered out, but before any other filtering mechanisms were applied. Two clear regions of the data are present, a smaller upper region and a larger lower region, corresponding to red giant stars and main sequence stars respectively. For our preliminary analysis in Sections 3 and 4, given that we want to fit a linear model to our data, we decided to analyze main sequence stars only and filter out red giants. All red giant points above the dashed line in the central panel of Figure 1 were filtered out, leaving only main sequence points, as shown in the right panel of Figure 1. Sections 6 and 8 include the red giant data in their analyses, corresponding to the left panel of Figure 1.

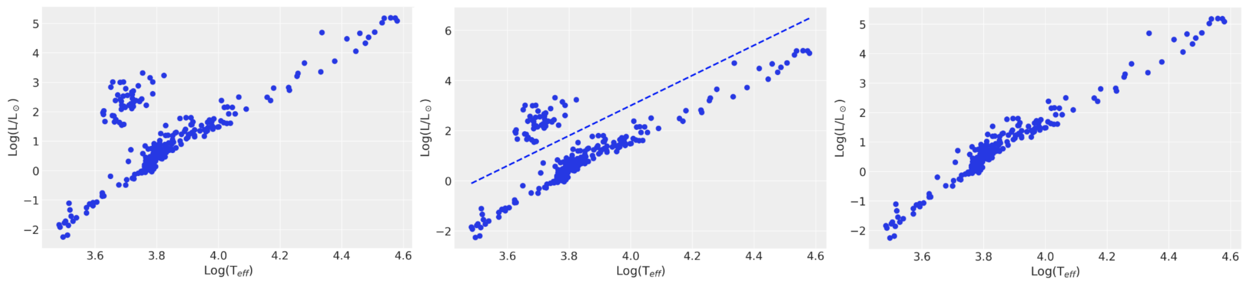

Figure 1: Log of solar-normalized stellar luminosity vs. log of effective stellar temperature at different filtering points. Left panel: Data before red giants are filtered out. Center panel: Identical to left but including the red giant filter cutoff. Right panel: Filtered data excluding red giants.

.

3 Regression: Without Errors

We employed a linear regression model to test the relationship between the logarithms of luminosity and temperature described in Equation 2 as seen in Equation 3

$$\mu = \beta x \tag{3}$$

Where $x = (1, log(T)$ and $\beta = (\beta_0, \beta_1)$ are the regression coefficients that correspond to the intercept and coefficient of $log(T)$ respectfully.

3.1 Statistical Model

For the components of the statistical model, β is the parameter we are interested in, specifically β_1 as it indicates by which power L and T are related. $\beta \in \mathbb{R}$ as there are no restrictions on the slope or intercept besides the fact they must be real values. The logarithm of the temperature, $X_i = log(T_i)$, of the ith star, where $i = 1, 2, ..., n$ and $n = 212$, has a sample space $\mathcal{X}$ of all possible vectors of n real-valued numbers. As mentioned in Section 2, the catalog is a summary of properties from a well studied system. Therefore we assume our likelihood is normally distributed:

$$p(y|\mu) \sim N(\mu, \tau^{-1}) \tag{4}$$

Where μ is Equation 3 and τ is the precision. For the priors, we assume that $\beta \sim N(a, b^{-1}I_2)$ and $\tau \sim Gamma(c, d)$. We assume the standard reference prior for both β and τ, $(a = 0, b = 1e-3, c = d = 0.5)$ in order to leave as many values open as possible.

3.2 Results

We ran a Markov Chain Monte Carlo simulation for 10000 iterations and summarize the results below:

	Mean	sd	hdi 2.5%	hdi 97.5%
β_0	-23.615	0.358	-24.333	-22.911
β_1	6.338	0.092	6.151	6.516
τ	11.354	1.071	9.342	13.535

In Figure 2, the posterior distributions for our parameters β_0, $\beta 1$, τ, our mean of 6.3 for β_1 with a confidence interval between (6.1,6.5) is not what we were expecting. This result would imply that approximately L is proportional to T^6 which directly contradicts Equation 1.

In Figure 3, the posterior predictive check is displayed to check the validity of our model. 100 draws were used for this test. The model clearly represents a smoother realization than what was actually observed, as there are visible divergences between the posterior predictive mean and observed lines. This supports the idea that our results for β_1 may not be valid.

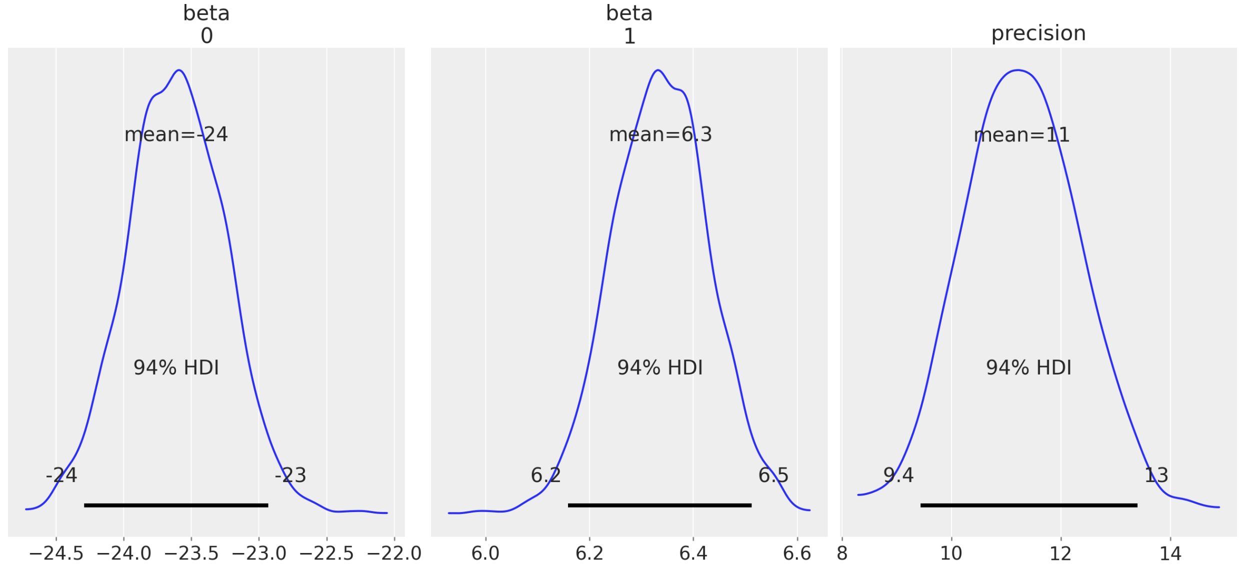

Figure 2: Posterior distribution of β and τ for a normal linear regression model where we assumed no errors in L nor T. Left panel: Posterior distribution of β_0 with mean at -24 and 94% HDI of (-24,-23). Center panel: Posterior distribution of β_1 with mean at 6.3 and 94% HDI of (6.2,6.5). Right panel: Posterior distribution of τ with mean at 11 and 94% HDI of (9.4,13).

.

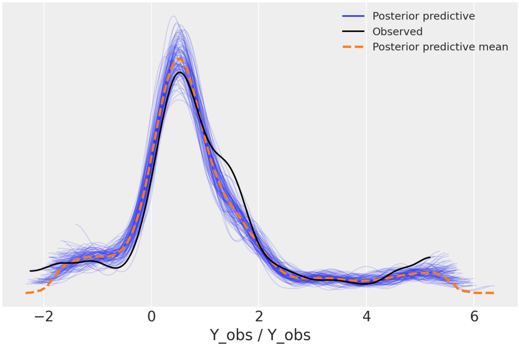

Figure 3: Posterior Predictive Check Displaying 100 draws (blue), the mean of all the draws (orange), and the observed $log(L)$ data (black).

.

3.3 Sensitivity Analysis

In order to test the assumptions made in this model, we performed a sensitivity analysis, changing the constants assumed in the prior and observing the change in the posterior distributions for β and τ. Below are the test priors used in the sensitivity analysis.

	a	b	c	d
Original Prior	0	0.001	0.5	0.5
Test Prior 1	0	0.1	0.5	0.5
Test Prior 2	5	0.001	0.5	0.5
Test Prior 3	0	0.001	1	1
Test Prior 4	5	0.1	1	1

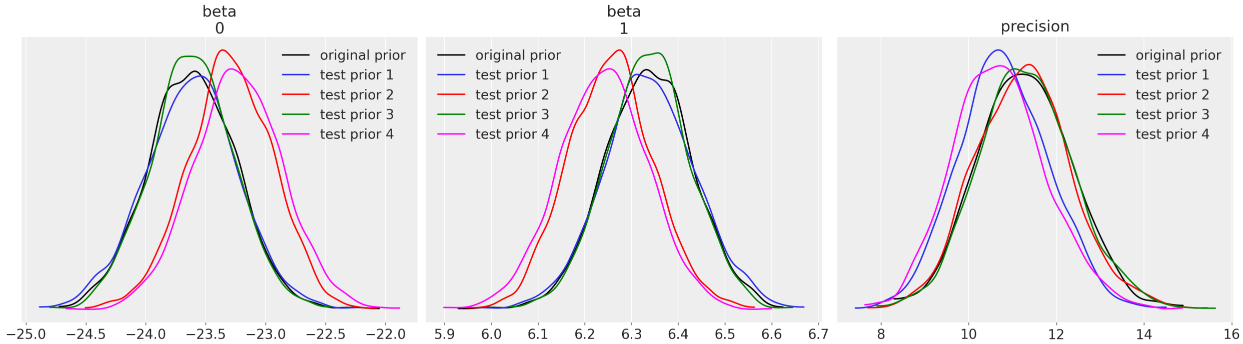

Figure 4: Sensitivity analysis displaying posteriors of β and τ given different priors: Original prior used in Bayesian analysis above (black), test prior 1 (blue), test prior 2 (red), test prior 3 (green), test prior 4 (magenta). Left panel: Posterior distributions for β_0. Center panel: Posterior distributions for β_1. Right panel: Posterior distributions for τ.

.

In testing these various priors, Figure 4, displaying the posterior distributions of β and τ, shows the stability of the posterior. The largest variations are tied to the priors that do not include the mean value displayed in the posterior, and even those variations are minimal.

4 Regression: With Errors

This section employs the same methodology used in Section 3, this time including the errors of both $log(L/L_\odot)$ and $log(T_{eff})$ in our analysis. As with the previous section, we denote $\beta = (\beta_0, \beta_1)$ where β_0 corresponds to the intercept coefficient and β_1 corresponds to the slope coefficient. The space in which β must exist, the sample space $\mathcal{X}$ of $X_i = log(T_i)$, and the assumed normal likelihood distribution are the same as described in Section 3.

4.1 Statistical Model

The statistical model for this section's analysis takes the following form:

$$\begin{aligned}
Y_i^* &\sim N(y_i, a^{-1}) \\
Y_i &\sim N(x_i\beta, \tau^{-1}) \\
X_i^* &\sim N(x_i, b^{-1}) \\
X_i &\sim N(0, c^{-1}), iid \\
\beta &\sim N_2(0, d^{-1}I_2) \\
\tau &\sim Gamma(e, f)
\end{aligned}$$

With error information on Y_i^* and X_i^* represented as a and b respectively, and $c = d = 1e-3$ and $e = f = 0.5$ serving as our original constant values.

4.2 Results

We ran a Markov Chain Monte Carlo simulation for 10000 iterations and summarize the results below:

	Mean	sd	hdi 2.5%	hdi 97.5%
β_0	-23.318	3.315	-29.934	-16.890
β_1	6.261	0.854	4.608	7.976
τ	0.132	0.009	0.114	0.151

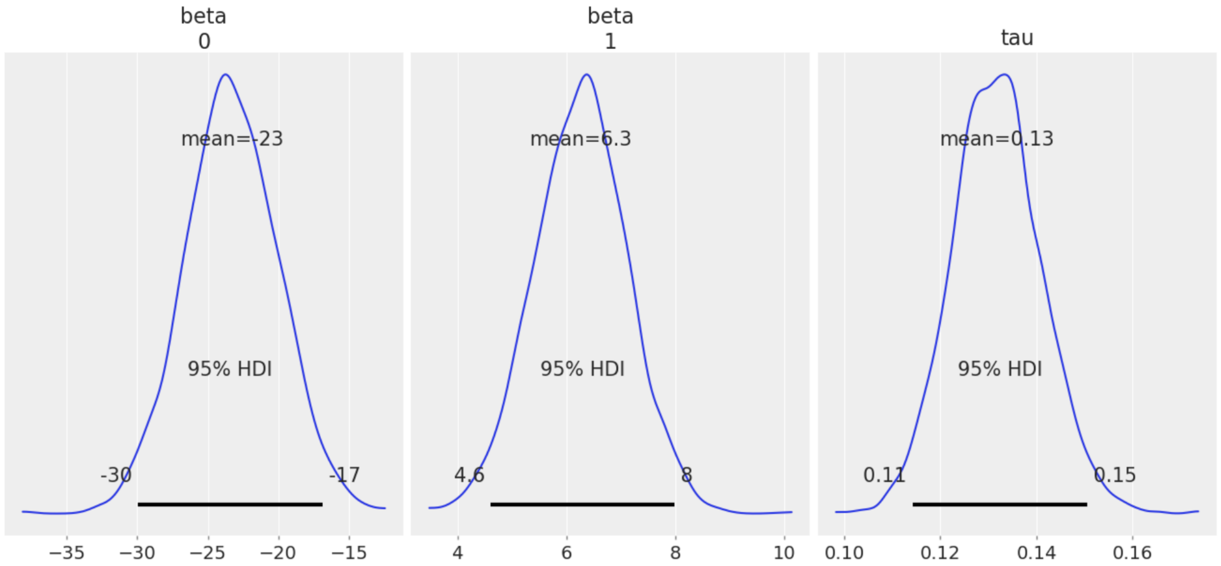

Figure 5: Posterior distribution of β and τ for a normal linear regression model, including errors on $log(L)$ and $log(T)$. Left panel: Posterior distribution of β_0 with mean at -23 and 95% HDI of (-30,-17). Center panel: Posterior distribution of β_1 with mean at 6.3 and 95% HDI of (4.6,8.0). Right panel: Posterior distribution of τ with mean at 0.13 and 95% HDI of (0.11,0.15)

.

In accordance with Section 3, from Figure 5, the mean β_1 value of 6.2 from the posterior distribution does not well agree with the theoretical value of 4 as expected from the Stefan-Boltzmann law. An interesting note, however, is compared to Section 3 the standard deviations on both β_0 and β_1 are larger by nearly a factor of 10, resulting in a broader range in confidence intervals. This broader range yields a lower level value for β_1 of 4.608 which, while not exactly, is closer to agreeing with the theoretical value of 4 than any value in the confidence interval in Section 3. Additionally, the standard deviation on τ is nearly 100 times smaller than the standard deviation on τ from Section 3, which makes sense given this analysis includes errors on $log(L)$ and $log(T)$.

Figure 6 depicts the posterior predictive check to test the validity of our model. 100 draws were used for this test, and qualitatively the posterior predictive, observed, and posterior predictive mean values on both Y^* and X^* are in better agreement with one another than they were in Section 3.

4.3 Sensitivity Analysis

Following the same logic and methodology as discussed in Section 3, we used the following table of constant values to vary prior parameters and observe changes in the resulting posterior distributions:

	c	d	e	f
Original Prior	1e-3	1e-3	0.5	0.5
Test Prior 1	1e-3	1e-3	2	2
Test Prior 2	5e-3	5e-3	3	5
Test Prior 3	1e-4	1e-4	4	9
Test Prior 4	1e-5	1e-5	5	10

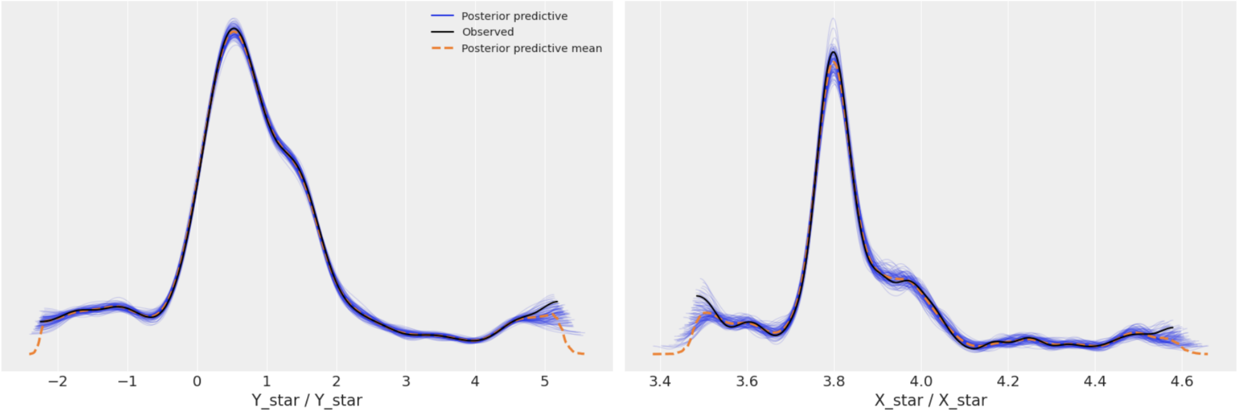

Figure 6: Posterior Predictive Check of 100 draws (blue), the mean of all the draws (orange), and the observed data (black) for Y^* ($Log(L)$, left panel) and X^* ($Log(T)$, right panel).

.

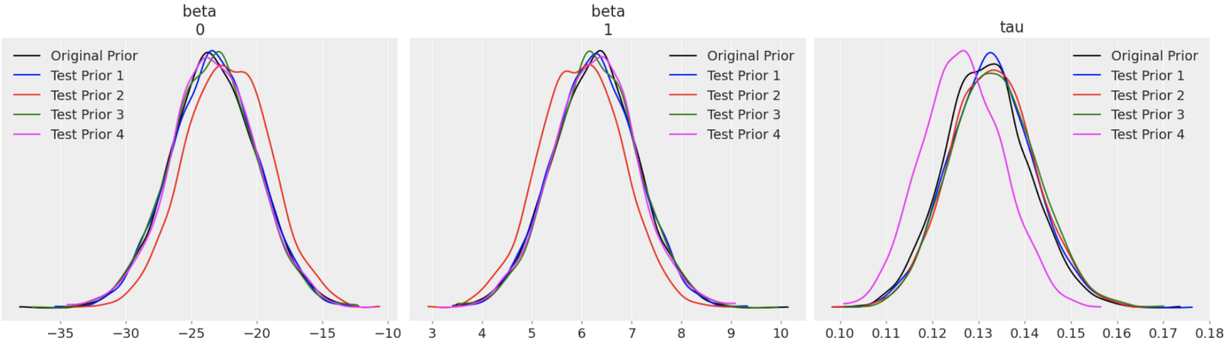

Figure 7: Sensitivity analysis displaying posteriors (determined using errors) of β and τ given different priors: Original prior used in Bayesian analysis above (black), test prior 1 (blue), test prior 2 (red), test prior 3 (green), test prior 4 (magenta). Left panel: Posterior distributions for β_0. Center panel: Posterior distributions for β_1. Right panel: Posterior distributions for τ.

.

Figure 7 displays these resulting posterior distributions of β and τ. Even the largest variations in constant values do not yield much change, showing that the priors are robust to influence, even more so than those explored in the sensitivity analysis in Section 3.

5 Issues & Discussion

There were issues and features that arose during the analysis shown for the two models. The posteriors' samples of β_0 and β_1 for the linear regression model with and without errors are shown in the right and left figures of Figure 8, respectively.

One features that were common for both models were the tight correlation between β_0 and β_1. We understood this correlation as a feature of our data, due to the small determinant of the X^TX matrix. As shown in the posterior densities of the two models, as well as the histograms of β_0, and β_1 posterior samples, there is a significant different in the width of the distribution between the two models. We plotted the posterior densities of β_1 inferred from the two models together and show them in Figure 9, along with our theoretical expectation value of 4 for β_1 according to equation (2). As shown in figure 9, the width of the density of the "without error" model is larger than its counter part. In addition, neither models agree well with the theoretical prediction, since 4 is outside both of the 95% confidence intervals for β_1s from both

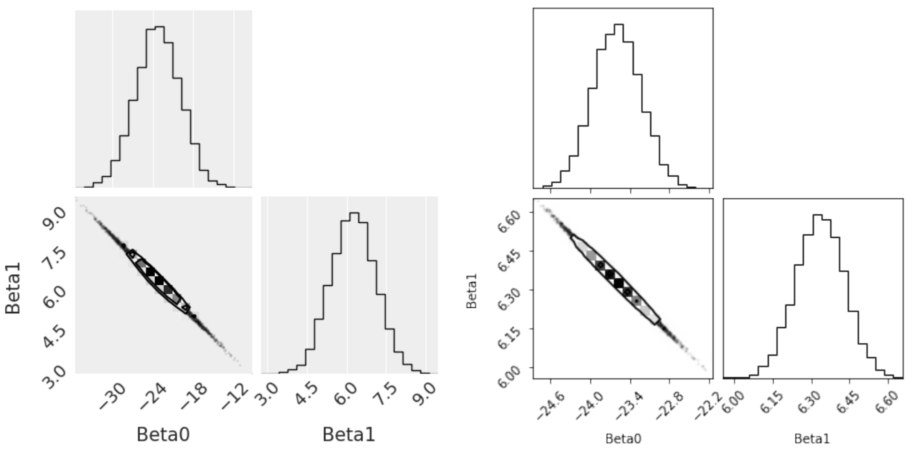

Figure 8: *Left:* Corner plot showing the histogram of MCMC β_0 and β_1 posteriors samples for the linear regression model presented in section 3: Linear regression model without error. *Right:* Corner plot showing the histogram of MCMC simulated β_0 and β_1 for the linear regression model presented in section 4: Linear regression model with error.

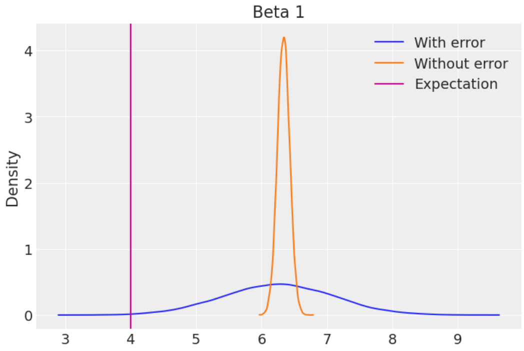

Figure 9: Posteriors densities of β_1 for the two linear regression with and without error models plotting on the same plot.

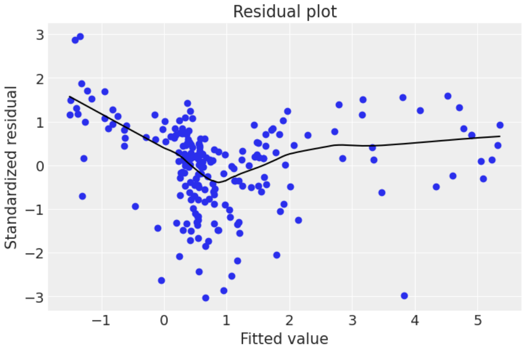

Figure 10: Standardized residual plot using the mean posterior densities of the fit values and the LOWESS fitted line for the linear regression model that include errors.

models, especially the "without error" model since it does not include the prediction value at all in the distribution.

Lastly, figure 10 shown the standardized residual plot using the posterior mean values for β_0 and β_1 of the linear regression models that includes error in section 4. We expected the shape of the residual plots for models in section 3 and 4 would be the same since the posterior means of β_0 and β_1 from the two models agreed well with each other even though the width of the distribution were different. As seen in figure 10, there were noticeable trend within the residual plots suggesting further improvement were needed. In the following sections, we expanded our works on the linear regression model to improve the residuals as well as explaining the discrepancy between our inferred β_1 and its prediction.

6 Regression: Without errors, including red giants

In Figure 1 we saw that there are quite a few data points corresponding to the red giants that were discarded in the earlier analysis. But we can take care of them if we use 2 separate lines for the main sequence stars by adding a dummy variable that takes 1 for the red giants and 0 for the main sequence ones. This new regression line looks like

$$logL_i = \beta_0 + \beta_1 logT_i + \beta_2 \mathbf{1}_{RG_i} + \epsilon_i$$

And the fitted lines look like as shown in Figure 11 For the Bayesian analysis without the errors, we assume the reference prior, i.e.

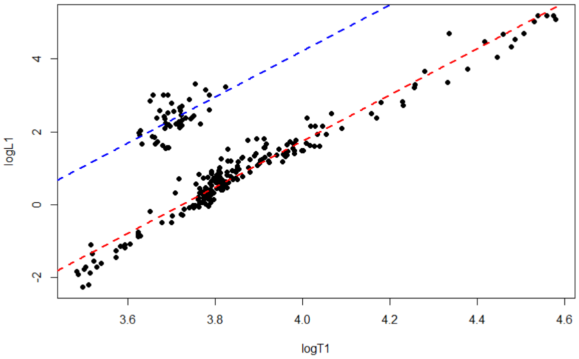

Figure 11: Pair of parallel regression lines for two groups

$$Y_i \overset{ind}{\sim} \mathcal{N}(x_i^T \beta, \sigma^2) \;\; i = 1, 2, \cdots, n$$
$$\beta, \sigma^2 \sim \nu(\beta, \sigma^2) = \frac{1}{\sigma^2}$$

Based on this model, the ACF of the the posterior samples drawn using the Linchpin sampling is found to have no significant correlation beyond lag 0 (Figure 12) And the distributions themselves look like as shown in Figure 15 and as tabulated in Table 6

	Mean	sd	hdi 2.5%	hdi 97.5%
τ	10.323604	0.89840112	8.640190	12.183833
β_0	-23.57657	0.37635811	-24.318149	-22.836829
β_1	6.327940	0.09699357	6.137456	6.517898
β_2	2.485393	0.05191732	2.383657	2.587194

The joint distribution of the regression coefficients can be found in Figure 13.

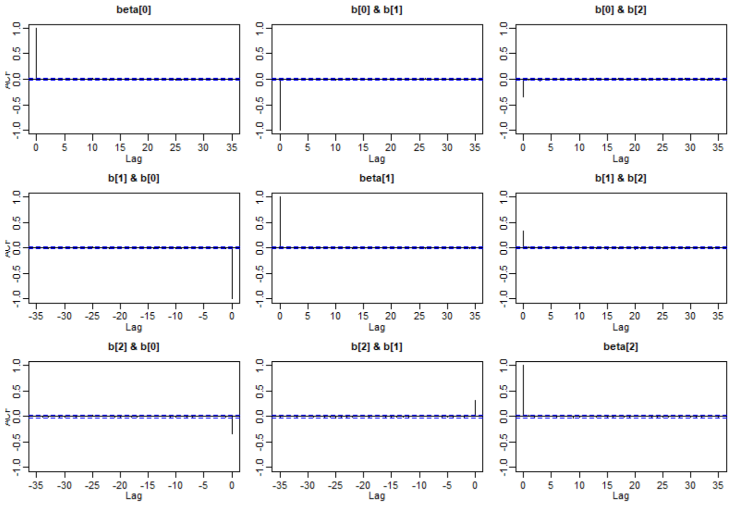

Figure 12: ACF of the posterior samples

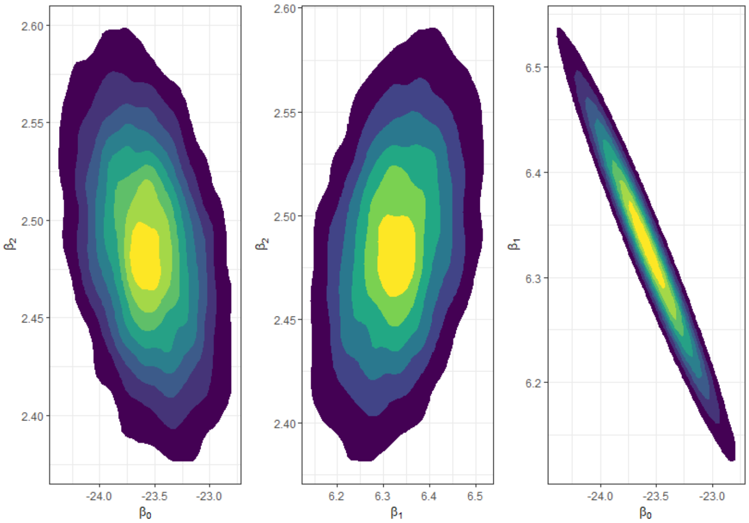

Figure 13: Joint distribution of the regression coefficients

We computed the same using independence prior as well with

$$\begin{aligned} Y_i &\overset{ind}{\sim} \mathcal{N}(x_i^T\beta, 1/\tau) \\ \beta &\sim \mathcal{N}(0, (b_1, b_1, \cdots, b_1)) \\ \tau &\sim Gamma(c, d) \end{aligned}$$

here we set $c = d = 0.5$ and $b_1 = 1000$. This yields similar posterior distributions (Figure 16). And setting $c = d = 10$ and $b_1 = 1$ yields the posterior distributions as described in Figure 17. The outcome is very similar to the one we obtained from the reference prior case. This sensitivity analysis of the data shows that the choice of prior does significantly change the results. Again, we see similar joint distributions of the coefficients as well.

Figure 14: Caption

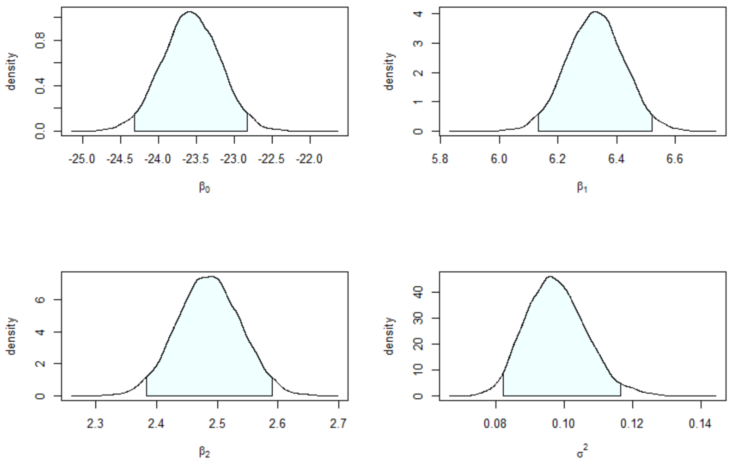

Figure 15: Posterior Distribution of the Regression Parameters

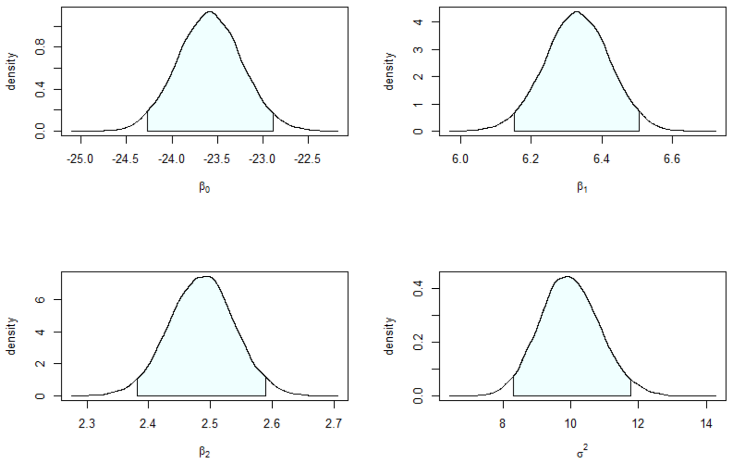

Figure 16: Posterior Distribution of the Regression Parameters

7 Regression: Without Errors, Including Giants and Radii

As we saw in the earlier analysis, the slope came out to be around 6.3 which is far from the theoretical value of 4. The reason being we did not account for the radius. As stated in equation 1

$$L = 4\pi R^2 \sigma T^4 \tag{5}$$

Through out the previous regression models, fitting on $log(L)$ and $log(R)$ means that the radii are assumed to be constant through out all the data points (or stars), which is clearly not the case.

So, now we consider the following model

$$logL_i = \beta_0 + \beta_1 logT_i + \beta_2 R_i + \mathbf{1}_{RG_i}\beta_3 + \epsilon_i$$

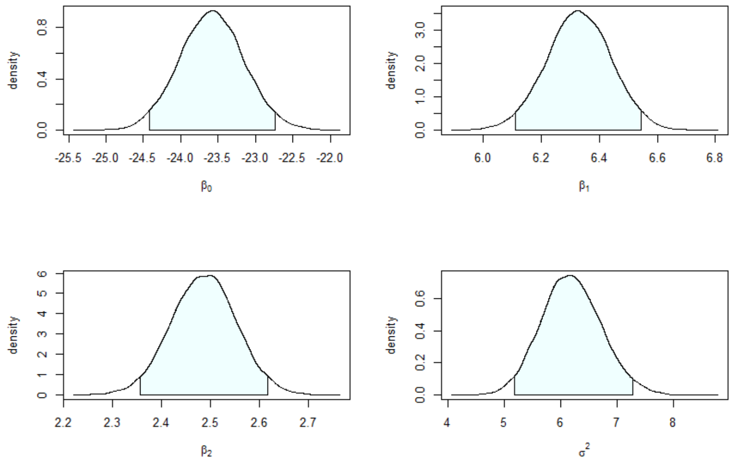

Figure 17: Posterior Distribution of the Regression Parameters

For this analysis, we again use the Linchpin variable sampling to get the posterior samples. The posterior distribution looks like as shown in Figure 18. The outcomes are tabulated in Table 7. Here we can clearly see that including the radius has made the bayes' estimates very close to the theoretical values. Another interesting thing to note here is that the coefficient for the dummy variable turns out to be insignificant as 0 is present in the 95% credible interval which means we don't need 2 separate planes to explain the association for the main sequence stars and red giants. This is because a 3d visualisation of the data (Figure 19) shows all the data points essentially lie on the same plane

	Mean	sd	hdi 2.5%	hdi 97.5%
τ	5273.11	467.707	4396.271	6220.998
β_0	-15.08289	0.029	-15.139	-15.027
β_1	4.009881	0.008	3.995	4.025
β_2	1.994967	0.005	1.984	2.006
β_3	0.007.14534	0.007	-0.007	0.021

8 Regression: With Errors, Including Giants & Radii

In this model, we were attempting to include the log radii data ($log(R)$) as well as their errors in the models. In this case, all the data points of the red giants were also included since their radii are larger in comparison to the main-sequence stars' ones. It would play a similar role as the indicator variables distinguishing the two different types of stars.

8.1 Statistical model

Equation (1) could be written in its logarithmic form as:

$$log(L) \propto 4log(T) + 2log(R) \tag{6}$$

Where the intercept constant is not included. We denote $\beta = (\beta_0, \beta_1, \beta_2)$ are the regression coefficients that correspond to the intercept, coefficients of $log(T)$ and $log(R)$, respectively

Since the errors of $log(R)$ were also included, here we presented our statistical model:

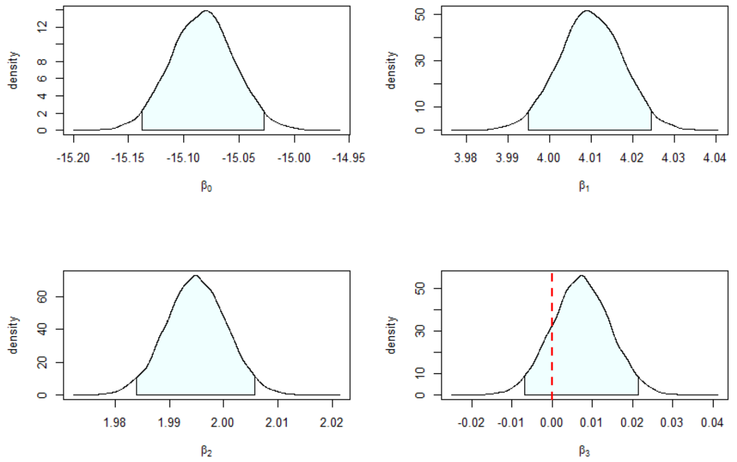

Figure 18: Posterior Distribution of the Regression Parameters

$$\begin{aligned}
Y_i^*|Y_i &\overset{ind}{\sim} N(y_i, a^{-1}) \\
Y_i|\beta, X_i, \tau_y &\overset{ind}{\sim} N(x_i\beta, \tau_y^{-1}) \\
X_i^*|X_i, \tau_x &\overset{ind}{\sim} N(x_i, \tau_x^{-1}) \\
X_i &\overset{ind}{\sim} N_2(b_i, c_i^{-1}) \\
\beta &\sim N_3(0, d^{-1}I_3) \\
\tau_x &\sim Gamma(e_x, f_x) \\
\tau_y &\sim Gamma(e_y, f_y)
\end{aligned}$$

Standard reference priors incorporate the error in X^* with $b_i = 0$ and the c_i given in the data, then d^{-1} is taken to be at least 1e3. Standard reference priors incorporate the error in Y^* with a_i given in the data. Tau_x here include both $\tau_{x1} \sim Gamma(e_{x1}, f_{x1})$ and $\tau_{x2} \sim Gamma(e_{x2}, f_{x2})$. The standard recommendation is to choose $e_{x1} = e_{x2} = e_y = f_{x1} = f_{x2}$ =$f_y = 0.5$ which essentially mimics a unit information prior.

8.2 Results

We ran our MCMC over 10,000 iterations and summarized our results below

	Mean	sd	hdi 2.5%	hdi 97.5%
τ	0.196	0.010	0.177	0.217
β_0	-14.940	2.559	-19.908	-9.927
β_1	3.973	0.670	2.671	5.290
β_2	2.001	0.286	1.440	2.565

Our posterior densities are shown in the Figure 20. Our expectation values of (4,2) for (β_1, β_2) fall well within the 95% confidence intervals and lie approximately at the mean of the distribution. Therefore, our linear regression results for this model agree well with the theoretical prediction.

To test the validity of the model, a 100 posterior predictive sample was also produced and shown in figure 21. The simulated data overlapped well with the observed data and there were no visible distinguishable

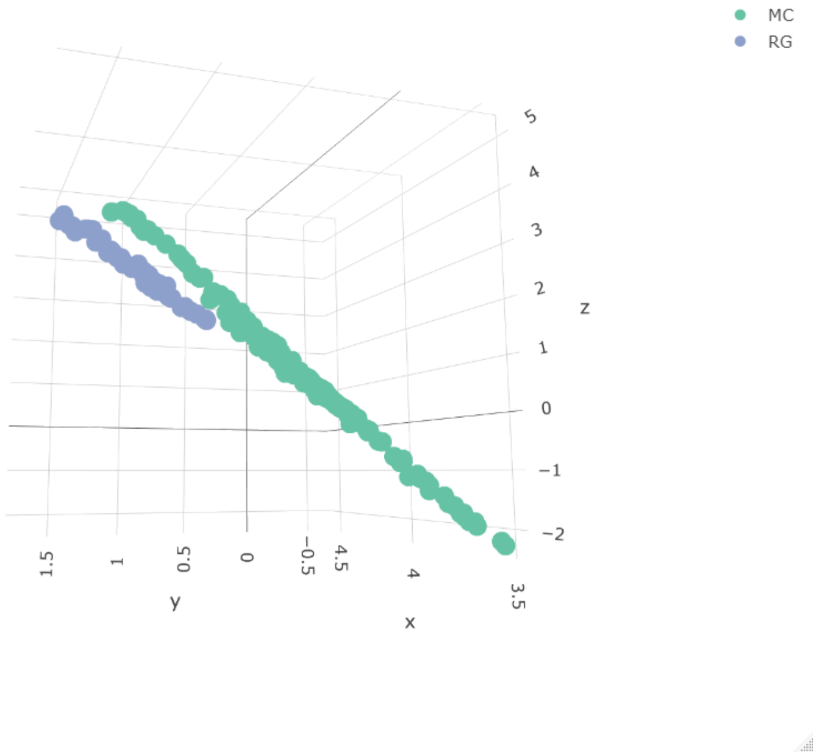

Figure 19: logL vs (logR, logT)

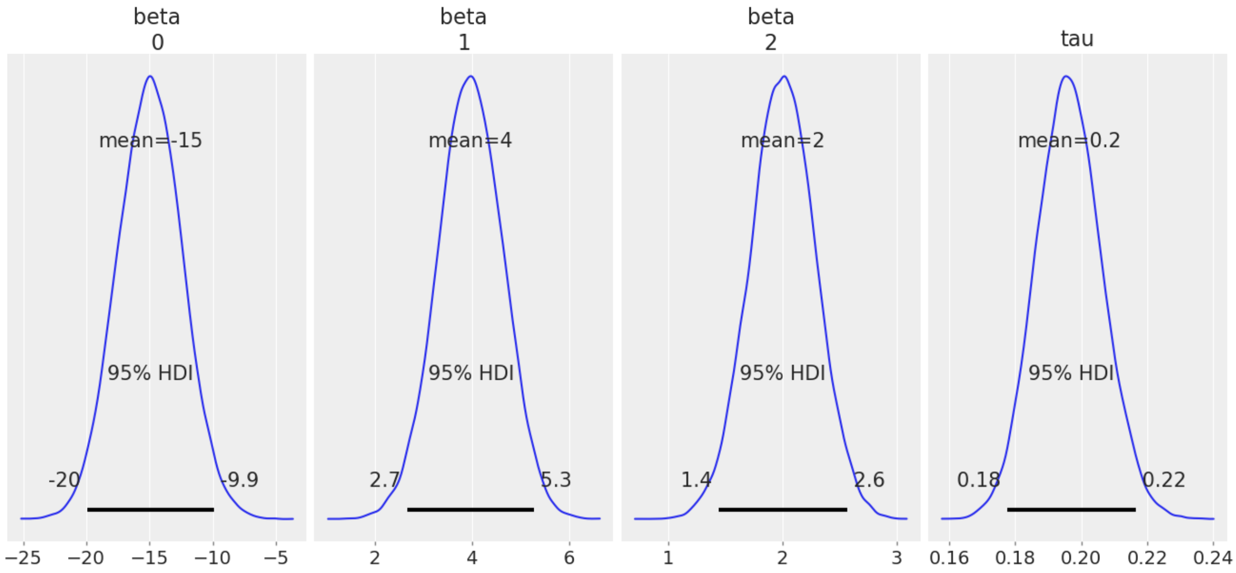

Figure 20: Posteriors density of our regressions coefficients (β_0, β_1, β_2, τ) and their 95% confidence interval. Our expectations for (β_1, β_2) of (4,2) fall well within the distributions as shown in the two middle distributions.

trends between the two for all three variables. The standardized residual plot was also produced and shown in figure 22. The residual plots improved significantly as there was no clear pattern except near the high

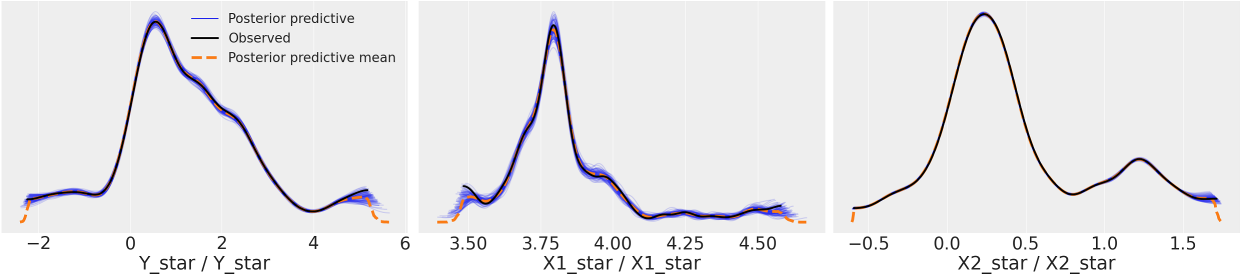

Figure 21: Posteriors predictive of 100 realizations and their means for Y^*, X_1^* and X_2^*, which corresponds to the simulated $Log(L)$, $Log(T)$ and $Log(R)$, respectively.

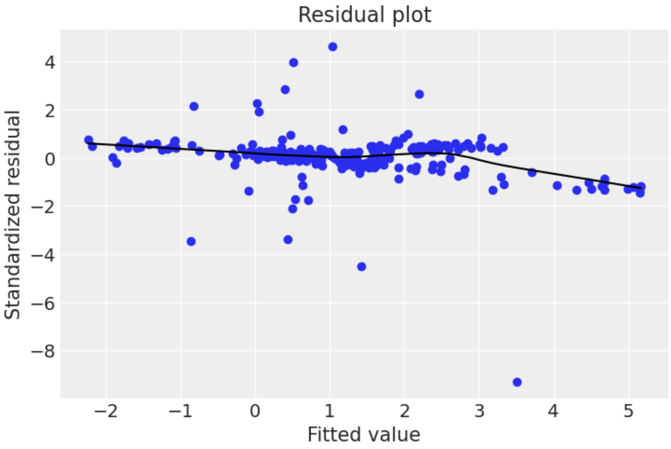

Figure 22: Standardized residual plot using the mean posterior densities of the fit values and the LOWESS fitted line for the linear regression model that included Log(R) data as well as the associated errors

fitted values, and the data were randomly distributed for the most part of the plot. Therefore, our linear regression model successfully validating the Stefan-Boltzmann law shown in equation 1 using the data.

8.3 Sensitivity Analysis

In this section we investigate the sensitity of our model that include radii and errors data by varying the prior parameters, in particular the parameters in the gamma distribution of $\tau_{x1} \sim Gamma(e_{x1}, f_{x1})$, $\tau_{x2} \sim Gamma(e_{x2}, f_{x2})$, and $\tau_y \sim Gamma(e_y, f_y)$.

Figure 23 showed the posterior densities for β_1 and β_2 inferred from different choices of prior. The specific numerical values for the parameters' sets are reported in the legends of the figures in the format of $(e_{x1}, f_{x1}, e_{x2}, f_{x2}, e_y, f_y)$. As the parameters varied, the peaks position and means of the distributions stayed approximately the same for both β_1 and β_2, and they all agreed with the theoretical expectation value of $(\beta_1, \beta_2) = (4, 2)$. However, as shown in the plot, the width of the distribution was sensitive to the choice of prior as it varied with different sets of priors' parameters.

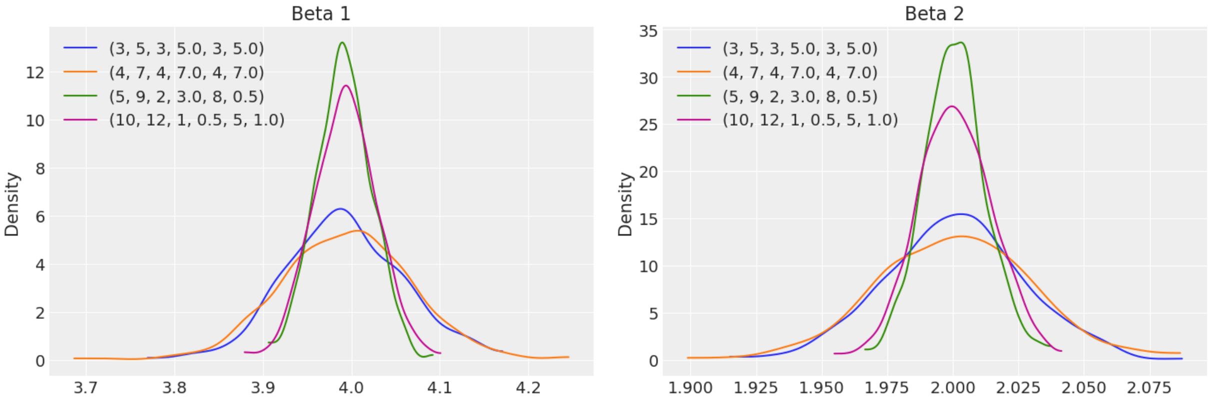

Figure 23: *Left:* Posteriors densities of β_1 of different MCMC runs with different parameters sets for the gamma distribution for τs. The legends are in the format of $(e_{x1}, f_{x1}, e_{x2}, f_{x2}, e_y, f_y)$. *Right:* Posteriors densities of β_1 of different MCMC runs with different parameters sets for the gamma distribution for τs. The legends are in the format of $(e_{x1}, f_{x1}, e_{x2}, f_{x2}, e_y, f_y)$.

References

[1] J. Andersen, “Accurate masses and radii of normal stars,” , vol. 3, no. 2, pp. 91–126, Jan. 1991.