A. By construction, we know that the marginal prior distribution $p(\theta)$ is a gamma mixture of normals. Show that this takes the form of a centered, scaled t distribution:

$$p(\theta) \propto \left(1 + \frac{1}{\nu} \cdot \frac{(x-m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}$$

with center m, scale s, and degrees of freedom ν, where you fill in the blank for m, s^2, and ν in terms of the four parameters of the normal-gamma family. Note: you did a problem just like this on a previous exercise! This shouldn't be a lengthy re-derivation.

We can calculate $p(\theta)$ as the integral of the joint prior $p(\theta, \omega)$ with respect to ω:

$$\begin{aligned}
p(\theta) &= \int_\Omega p(\theta, \omega) d\omega \\
&\propto \underbrace{\int_\Omega w^{\frac{d+1}{2}-1} \exp\left\{-\omega\left(\frac{\kappa(\theta-\mu)^2}{2} + \frac{\eta}{2}\right)\right\} d\omega}_{\text{kernel of Gamma}\left(\frac{d+1}{2}, \frac{\kappa(\theta-\mu)^2}{2} + \frac{\eta}{2}\right)} \\
&\propto \left(\frac{\left[\frac{\kappa(\theta-\mu)^2}{2} + \frac{\eta}{2}\right]^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})}\right)^{-1} \\
&\propto \left[\frac{\kappa(\theta-\mu)^2}{2} + \frac{\eta}{2}\right]^{-\frac{d+1}{2}} \\
&\propto \left[\left(\frac{\kappa(\theta-\mu)^2}{2} + \frac{\eta}{2}\right)\frac{2}{\eta}\right]^{-\frac{d+1}{2}} \frac{\eta}{2}^{-\frac{d+1}{2}} \\
&\propto \left[\frac{1}{d} \cdot \frac{(\theta-\mu)^2}{\eta/(\kappa d)} + 1\right]^{-\frac{d+1}{2}},
\end{aligned}$$

which takes the form of a centered, scaled t distribution with parameters

$$\begin{aligned}
m &= \mu \\
s &= \sqrt{\frac{\eta}{\kappa d}} \\
\nu &= d
\end{aligned}$$

(c.f. "Bayesian Inference in Simple Conjugate Families" (F))

B. Assume the normal sampling model in Equation (1) and the normal-gamma prior in Equation (2). Calculate the joint posterior density $p(\theta, \omega \mid \mathbf{y})$, up to constant factors not depending on ω or θ. Show that this is also a normal/gamma prior in the same form as above:

$$p(\theta, \omega \mid \mathbf{y}) \propto \omega^{(d^\star+1)/2-1} \exp\left\{-\omega \cdot \frac{\kappa^\star(\theta-\mu^\star)^2}{2}\right\} \cdot \exp\left\{-\omega \cdot \frac{\eta^\star}{2}\right\} \quad (1)$$

From this form of the posterior, you should able to read off the new updated parameters, by pattern-matching against the functional form in Equation (2):

- $\mu \longrightarrow \mu^\star = ?$
- $\kappa \longrightarrow \kappa^\star = ?$
- $d \longrightarrow d^\star = ?$
- $\eta \longrightarrow \eta^\star = ?$

We can find the joint posterior by taking the product of the likelihood and the joint prior:

$$\begin{aligned}
p(\theta, \omega|\mathbf{y}) &\propto p(\mathbf{y}|\theta, \omega)p(\theta, \omega) \\
&\propto \prod_{i=1}^{n} \sqrt{\omega} \exp\left\{-\frac{\omega}{2}(y_i-\theta)^2\right\} \omega^{\frac{d+1}{2}-1} \exp\left\{-\omega\left(\frac{\kappa(\theta-\mu)^2}{2}+\frac{\eta}{2}\right)\right\} \\
&\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[\sum_{i=1}^{n}\left(y_i^2-2y_i\theta+\theta^2\right)+\kappa(\theta^2-2\theta\mu+\mu^2)+\eta\right]\right\} \\
&\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[\sum_{i=1}^{n} y_i^2-2n\bar{y}\theta+n\theta^2+\kappa\theta^2-2\kappa\theta\mu+\kappa\mu^2+\eta\right]\right\} \\
&\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[\sum_{i=1}^{n} y_i^2+(\kappa+n)\theta^2-2(n\bar{y}+\kappa\mu)\theta+\kappa\mu^2+\eta\right]\right\} \\
&\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[(\kappa+n)\theta^2-2(n\bar{y}+\kappa\mu)\theta+\left(\kappa\mu^2+\eta+\sum_{i=1}^{n} y_i^2\right)\right]\right\} \\
&\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\omega\left[\frac{(\kappa+n)\left(\theta-\frac{n\bar{y}+\kappa\mu}{\kappa+n}\right)^2}{2}\right]\right\} \exp\left\{-\omega\left(\frac{\kappa\mu^2+\eta+\sum_{i=1}^{n} y_i^2-\frac{(n\bar{y}+\kappa\mu)^2}{\kappa+n}}{2}\right)\right\} \\
&\equiv \omega^{\frac{d^*+1}{2}-1} \exp\left\{-\omega\frac{\kappa^*(\theta-\mu^*)^2}{2}\right\} \exp\left\{-\omega\frac{\eta^*}{2}\right\},
\end{aligned}$$

where

$$\begin{aligned}
d^* &= d + n \\
\kappa^* &= \kappa + n \\
\mu^* &= \frac{n\bar{y} + \kappa\mu}{\kappa + n} \\
\eta^* &= \eta + \kappa\mu^2 + \sum_{i=1}^{n} y_i^2 - \frac{(n\bar{y} + \kappa\mu)^2}{\kappa + n} \\
&= \eta + S_y + \frac{n\kappa(\bar{y} - \mu)^2}{n + \kappa}
\end{aligned}$$

C. From the joint posterior you just derived, what is the conditional posterior distribution $p(\theta \mid \mathbf{y}, \omega)$? Note: this should require no calculation—you should just be able to read it off directly from the joint distribution, since you took care to set up things so that the joint posterior was in the same form as Equation (2).

By pattern matching, we see that

$$\begin{aligned} p(\theta|\mathbf{y},\omega) &\propto \exp\left\{-\frac{\omega\kappa^*}{2}(\theta-\mu^*)^2\right\} \\ &\equiv \text{Normal}(\mu^*, (\omega\kappa^*)^{-1}), \end{aligned}$$

where $\mu^* = \frac{n\bar{y}+\kappa\mu}{\kappa+n}$ and $\kappa^* = \kappa + n$.

D. From the joint posterior you calculated in (A), what is the marginal posterior distribution $p(\omega \mid \mathbf{y})$? Unlike the previous question, this one doesn't come 100% for free—you have to integrate over θ. But it shouldn't be too hard, since you can ignore constants not depending on ω in calculating this integral.

Let's integrate $p(\omega, \theta|\mathbf{y})$ with respect to θ to get our marginal posterior for ω:

$$\begin{aligned} p(\omega|\mathbf{y}) &= \int p(\omega,\theta|\mathbf{y})d\theta \\ &\propto w^{\frac{d^*+1}{2}-1}\exp\left\{-\omega\frac{\eta^*}{2}\right\}\underbrace{\int \exp\left\{-\frac{\omega\kappa^*}{2}(\theta-\mu^*)^2\right\}d\theta}_{\text{kernel of Normal}(\mu^*,(\omega\kappa^*)^{-1})} \\ &\propto \omega^{\frac{d^*+1}{2}-1}\exp\left\{-\omega\frac{\eta^*}{2}\right\}(\sqrt{\omega\kappa^*})^{-1} \\ &\propto \omega^{\frac{d^*}{2}-1}\exp\left\{-\omega\frac{\eta^*}{2}\right\}, \end{aligned}$$

so $p(\omega|\mathbf{y}) \equiv \text{Gamma}\left(\frac{d^*}{2}, \frac{\eta^*}{2}\right)$.

E. Show that the marginal posterior $p(\theta \mid \mathbf{y})$ takes the form of a centered, scaled t distribution and express the parameters of this t distribution in terms of the four parameters of the normal-gamma posterior for (θ, ω). Note: since you've set up the normal-gamma family in this careful conjugate form, this should require no extra work. It's just part (A), except for the posterior rather than the prior.

We can simply pattern-match from (A):

$$p(\theta|\mathbf{y}) \propto \left(1 + \frac{1}{\nu} \cdot \frac{(x-m)^2}{s^2}\right)^{\frac{\nu+1}{2}},$$

where $\nu = d^*, m = \mu^*$, and $s^2 = \frac{\eta^*}{d^* \kappa^*}$.

F. True or false: in the limit as the prior parameters κ, d, and η approach zero, the priors $p(\theta)$ and $p(\omega)$ are valid probability distributions. (Remember that a valid probability distribution must integrate to 1 (or something finite, so that it can normalized to integrate to 1) over its domain.)

FALSE; both priors will be undefined if their hyperparameters are 0.

G. True or false: in the limit as the prior parameters κ, d, and η approach zero, the posteriors $p(\theta \mid \mathbf{y})$ and $p(\omega \mid \mathbf{y})$ are valid probability distributions.

TRUE; if $\kappa, d, \eta \to 0$, then $d^*, \kappa^* \to n, \mu^* \to \bar{y}$, and $\eta^* \to S_y$. Therefore, the parameters in both marginal posteriors (θ and ω) are defined and the probability distributions are valid.

H. Your result in (E) implies that a Bayesian credible interval for θ takes the form

$$\theta \in m \pm t^{\star} \cdot s \,,$$

where m and s are the posterior center and scale parameters from (E), and $t^{\star}$ is the appropriate critical value of the t distribution for your coverage level and degrees of freedom (e.g. it would be 1.96 for a 95% interval under the normal distribution).

True or false: In the limit as the prior parameters κ, d, and η approach zero, the Bayesian credible interval for θ becomes identical to the classical (frequentist) confidence interval for θ at the same confidence level.

TRUE; as stated in (G), $d^*, \kappa^* \to n$, $\mu^* \to \bar{y}$, and $\eta^* \to S_y$. So, we have

$$\begin{aligned} s &= \sqrt{\frac{\eta^*}{d^* \kappa^*}} \\ &\to \sqrt{\frac{S_y}{n^2}} \\ &= \frac{\sqrt{S_y}}{n}, \\ m &= \mu^* \\ &\to \bar{y} \end{aligned}$$

Therefore,

$$m \pm t^{\star} \cdot s \to \bar{y} \pm t^* \frac{\sqrt{S_y}}{n},$$

which is identical to the frequentist confidence intervals for θ.

A. Derive the conditional posterior $p(\beta|\mathbf{y},\omega)$.

First, let's obtain the joint posterior distribution:

$$
\begin{aligned}
p(\beta,\omega|\mathbf{y}) &\propto p(\mathbf{y}|\beta,\omega)\cdot p(\beta|\omega)\cdot p(\omega) \\
&\propto |(\omega\Lambda)^{-1}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{y}-X\beta)^T\omega\Lambda(\mathbf{y}-X\beta)\right\} \\
&\quad\cdot |(\omega K)^{-1}|^{-1/2}\exp\left\{-\frac{1}{2}(\beta-m)^T\omega K(\beta-m)\right\} \\
&\quad\cdot \omega^{d/2-1}\exp\left\{-\frac{\omega\eta}{2}\right\} \\
&\propto \omega^{(n+p+d)/2-1} \\
&\quad\cdot \exp\left\{-\frac{1}{2}(\mathbf{y}-X\beta)^T\omega\Lambda(\mathbf{y}-X\beta)\right\} \\
&\quad\cdot \exp\left\{-\frac{1}{2}(\beta-m)^T\omega K(\beta-m)\right\}\exp\left\{-\frac{\omega\eta}{2}\right\}
\end{aligned}
$$

From here, we can obtain the conditional posterior for β by only keeping terms containing β:

$$
\begin{aligned}
p(\beta|\omega,\mathbf{y}) &\propto \exp\left\{-\frac{1}{2}\left[-2\mathbf{y}^T\omega\Lambda X\beta+\beta^T X^T\omega\Lambda X\beta-2m^T\omega K\beta+\beta^T\omega K\beta\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[-2\underbrace{(\mathbf{y}^T\omega\Lambda X+m^T\omega K)}_{\mathbf{b}^T}\beta+\beta^T\underbrace{(X^T\omega\Lambda X+\omega K)}_{A}\beta\right]\right\}
\end{aligned}
$$

Therefore,

$$
p(\beta|\omega,\mathbf{y}) = \text{Normal}\left((X^T\Lambda X+K)^{-1}(X^T\Lambda\mathbf{y}+K^Tm),(X^T\omega\Lambda X+\omega K)^{-1}\right)
$$

B. Derive the marginal posterior $p(\omega|\mathbf{y})$.

To obtain the marginal posterior for ω, we integrate the integrate the joint posterior with respect to β:

$$\begin{aligned}
p(\omega|\mathbf{y}) &\propto \int_\beta p(\omega,\beta|\mathbf{y})d\beta \\
&\propto \omega^{(n+p+d)/2-1}\exp\left\{-\frac{\omega}{2}\left[\mathbf{y}^T\Lambda\mathbf{y}+m^TKm+\eta\right]\right\} \\
&\quad\cdot\int_\beta \underbrace{\exp\left\{-\frac{1}{2}\left[-2(\mathbf{y}^T\omega\Lambda X+m^T\omega K)\beta+\beta^T(X^T\omega\Lambda X+\omega K)\beta\right]\right\}}_{\text{kernel of Normal}(A^{-1}b,A^{-1})}d\beta \\
&\propto \omega^{(n+p+d)/2-1}\exp\left\{-\frac{\omega}{2}\left[\mathbf{y}^T\Lambda\mathbf{y}+m^TKm+\eta\right]\right\} \\
&\quad\cdot|\omega^{-1}(X^T\Lambda X+K)^{-1}|^{1/2}\cdot \underbrace{\exp\left\{\frac{1}{2}\left[b^TAb\right]\right\}}_{\text{term that doesn't contain }\beta\text{ in the above integrand}} \\
&\propto \omega^{(n+d)/2-1}\exp\left\{-\frac{\omega}{2}\left[\mathbf{y}^T\Lambda\mathbf{y}+m^TKm+\eta\right]\right\} \\
&\quad\cdot\exp\left\{\frac{\omega}{2}\left[(\mathbf{y}^T\Lambda X+m^TK)^T(X^T\Lambda X+K)^{-1}(\mathbf{y}^T\Lambda X+m^TK)\right]\right\}
\end{aligned}$$

Therefore, the marginal posterior for ω is

$$\begin{aligned}
p(\omega|\mathbf{y}) &= \text{Gamma}\left(\frac{n+d}{2},\frac{\eta^*}{2}\right), \\
\eta^* &= \eta+\mathbf{y}^T\Lambda\mathbf{y}+m^TKm-(\mathbf{y}^T\Lambda X+m^TK)^T(X^T\Lambda X+K)^{-1}(\mathbf{y}^T\Lambda X+m^TK)
\end{aligned}$$

C. Derive the marginal posterior $p(\beta|\mathbf{y})$.

We can derive the marginal posterior for β using Bayes' rule:

$$\begin{aligned}
p(\beta|\mathbf{y}) &\propto \int_\Omega p(\omega, \beta|\mathbf{y}) d\omega \\
&\propto \int_\Omega p(\beta|\omega, \mathbf{y}) p(\omega|\mathbf{y}) d\omega \\
&\propto \int_\Omega \omega^{(n+p+d)/2-1} \\
&\quad \cdot \exp\left\{-\frac{1}{2}(\mathbf{y} - X\beta)^T \omega\Lambda(\mathbf{y} - X\beta)\right\} \\
&\quad \cdot \exp\left\{-\frac{1}{2}(\beta - m)^T \omega K(\beta - m)\right\} \exp\left\{-\frac{\omega\eta}{2}\right\} d\omega,
\end{aligned}$$

which resembles the kernel of a Gamma distribution. Thus, we see that

$$\begin{aligned}
p(\beta|\mathbf{y}) &\propto \left(\frac{(\beta^*)^{a^*}}{\Gamma(a^*)}\right)^{-1} \\
&\propto \frac{\Gamma\left(\frac{n+p+d}{2}\right)}{\left(\frac{1}{2}\left[(\mathbf{y} - X\beta)^T\Lambda(\mathbf{y} - X\beta) + (\beta - m)^T K(\beta - m) + \eta\right]\right)^{\frac{n+p+d}{2}}} \\
&\propto \left(\frac{1}{2}\left[(\mathbf{y} - X\beta)^T\Lambda(\mathbf{y} - X\beta) + (\beta - m)^T K(\beta - m) + \eta\right]\right)^{-\frac{n+p+d}{2}} \\
&\propto \left(\frac{1}{2}\left[-2(\mathbf{y}^T\Lambda X + m^T K)\beta + \beta^T(X^T\Lambda X + K)\beta\right]\right)^{-\frac{\nu^*+p}{2}} \\
&\propto \left(\frac{1}{2}\left[-2(\mathbf{y}^T\Lambda X + m^T K)\beta + \beta^T\Lambda^*\beta\right]\right)^{-\frac{\nu^*+p}{2}} \\
&\propto \left((\beta - \mu^*)^T\Lambda^*(\beta - \mu^*) + \eta^*\right)^{-\frac{\nu^*+p}{2}} \\
&\propto \left(\frac{1}{n+d}(\beta - \mu^*)^T\frac{n+d}{\eta^*}\Lambda^*(\beta - \mu^*) + 1\right)^{-\frac{\nu^*+p}{2}} \\
&\propto \left(\frac{1}{\nu^*}(\beta - \mu^*)^T\Sigma^*(\beta - \mu^*) + 1\right)^{-\frac{\nu^*+p}{2}}
\end{aligned}$$

where $\nu^* = n + d$, $\Lambda^* = X^T\Lambda X + K$, $\mu^* = (\Lambda^*)^{-1}(X^T\Lambda\mathbf{y} + K^T m)$ and $\Sigma^* = \frac{\nu^*}{\eta^*}\Lambda^*$. From the above, we see that the marginal posterior for β is a t-distribution with ν^* degrees of freedom, location μ^*, and covariance Σ^*.

D. Using the "greenbuildings.csv" data set, what is the 95% Bayesian credible interval for the coefficient on the green rating variable? How does your result compare to the classical 95% confidence interval? What does a histogram of the model residuals reveal? Are you happy with your model?

Using the derivation for the marginal posterior $p(\beta|\mathbf{y})$, we are able to sample β without using a Gibbs sampling scheme. From our derivation in (c), we get the following 95% Bayesian credible intervals for the coefficient on the green rating variable:

$$[-0.03, 1.72] \qquad \text{(using HDI estimation)}$$
$$[-0.04, 1.70] \qquad \text{(using ETI estimation)}$$

In comparison, our 95% confidence interval using *lm()* in *R* for the same covariate is

$$[-0.087, 1.69]$$

It appears that the 95% confidence interval is wider than our 95% credible intervals with a notably lower lower-bound. In Figure 1, we report a histogram of the model residuals. It's important to note that there is a long right tail, implying that our model is not ideal. This is most likely the case because of two reasons: (1) our model should not assume normality and (2) our assumptions on the covariance appear to be flawed. I'm not happy with this model because the histogram of these residuals should not display any skewness.

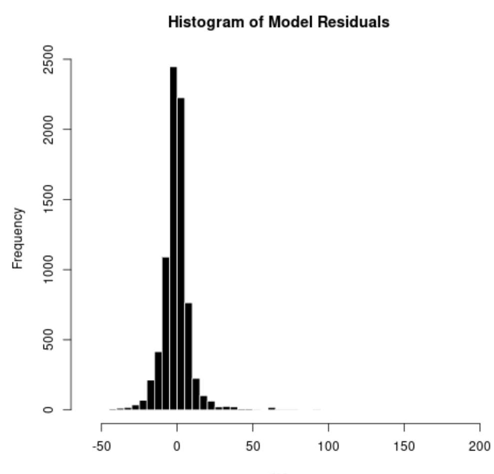

Figure 1: Histogram of Model Residuals

A. Under this model, what is the implied conditional distribution $p(y_i|\beta,\omega)$?

We can obtain the desired conditional distribution by marginalizing out λ_i:

$$
\begin{aligned}
p(y_i|\beta,\omega) &= \int_\Lambda p(y_i|\lambda_i,\beta,\omega)\cdot p(\lambda_i)d\lambda_i \\
&\propto \int_\Lambda \sqrt{\lambda_i}\exp\left\{-\frac{\omega\lambda_i}{2}\left(y_i - X_i^T\beta\right)^2\right\}\cdot \lambda_i^{\frac{h}{2}-1}e^{-\frac{\lambda_i h}{2}}d\lambda_i \\
&\propto \int_\Lambda \underbrace{\lambda_i^{\frac{h+1}{2}-1}\exp\left\{-\lambda_i\left[\frac{\omega}{2}\left(y_i - X_i^T\beta\right)^2 + \frac{h}{2}\right]\right\}}_{\text{kernel of Gamma}\left(\frac{h+1}{2},\frac{\omega}{2}\left(y_i - X_i^T\beta\right)^2+\frac{h}{2}\right)} d\lambda_i \\
&\propto \left[\frac{\omega}{2}(y_i - X_i^T\beta)^2 + \frac{h}{2}\right]^{-\frac{h+1}{2}} \\
&\propto \left[\left(\frac{\omega}{2}(y_i - X_i^T\beta)^2 + \frac{h}{2}\right)\frac{2}{h}\right]^{-\frac{h+1}{2}}\left(\frac{2}{h}\right)^{\frac{h+1}{2}} \\
&\propto \left[\frac{\omega}{h}(y_i - X_i^T\beta)^2 + 1\right]^{-\frac{h+1}{2}},
\end{aligned}
$$

which resembles a t-distribution. In fact,

$$
p(y_i|\beta,\omega) = t\left(\nu = h, m = X_i^T\beta, s^2 = \frac{1}{\omega}\right)
$$

B. What is the conditional posterior distribution $p(\lambda_i|\beta,\omega)$?

This is the integrand of the above problem, so we can easily see that

$$
p(\lambda_i|\beta,\omega) = \text{Gamma}\left(\frac{h+1}{2}, \frac{\omega}{2}\left(y_i - X_i^T\beta\right)^2 + \frac{h}{2}\right)
$$

C. Code a Gibbs sampler using the full-conditional for β (in the previous section), the marginal posterior for ω (in the previous section), and the conditional posterior distribution for λ_i. Apply it to the green buildings data set. Are you happier with the fit? How do the 95% credible intervals on each model term compare to our previous model? Are there certain regions of predictor space that seem to be associated with higher variance residuals?

In Figure 2, we see a side-by-side comparison of the histogram of residuals between our heteroskedastic (blue) and homoskedastic (red) model. It appears that with my choice of $h = 2$, both models demonstrate quite heavy tails; this means that my choice in $h = 2$ does not greatly improve our fit to the data. It seems that the heteroskedastic model has residuals that are more concentrated at 0, so that's at least a bit of improvement!

In Figure 3, we see the plot that Dr. Scott hinted at. Here, we see that for data points with smaller values, the relative variance could be significantly greater than the variance for data points of higher values. We see a somewhat-funnel shape to this plot, implying that greater values in data points have relatively less variance.

In Figure 4, we see the 95% credible intervals from our old (homoskedastic) and new (heteroskedastic) model. They appear to match well for $\beta_1, \beta_3, \beta_4$; however, there are noticeable changes in estimation for $\beta_2, \beta_5, \beta_6$.

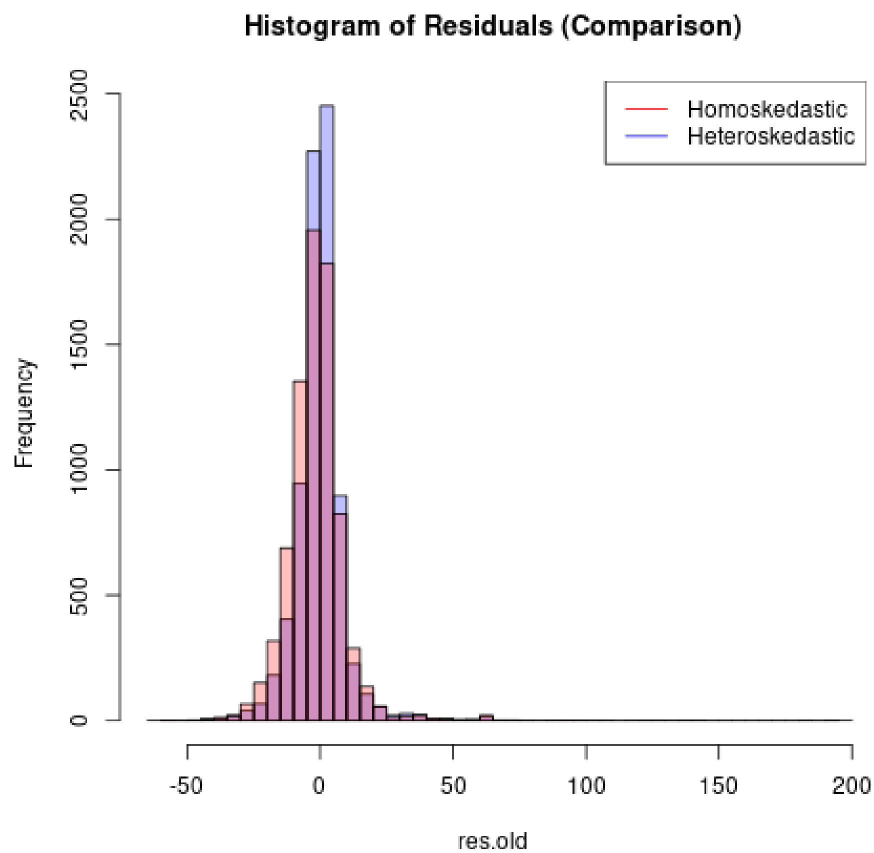

Figure 2: Compare Histograms of Residuals.

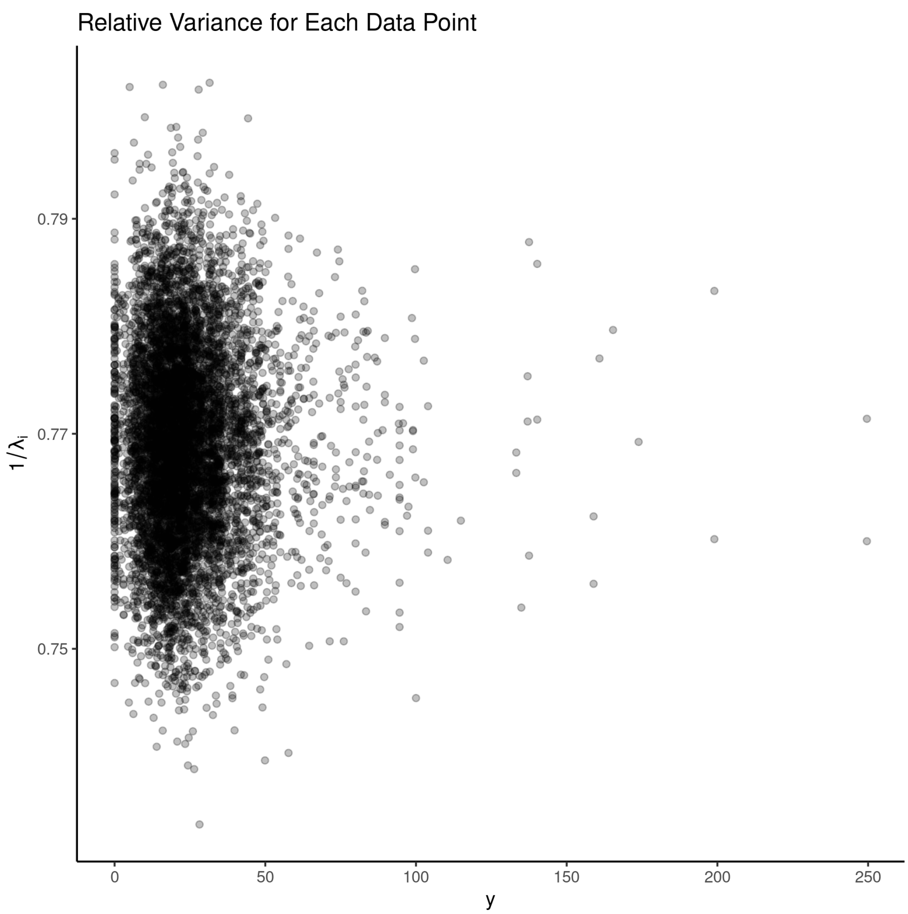

Figure 3: Plot of $\frac{1}{\lambda_i}$ v. y_i.

```
Old Model with  beta 1 : 95% HDI: [-10.00, -7.94]
NewModel with  beta 1 : 95% HDI: [-10.62, -7.29]

Old Model with  beta 2 : 95% HDI: [0.60, 2.32]
NewModel with  beta 2 : 95% HDI: [0.33, 2.40]

Old Model with  beta 3 : 95% HDI: [0.98, 1.02]
NewModel with  beta 3 : 95% HDI: [0.96, 1.05]

Old Model with  beta 4 : 95% HDI: [-0.01, 0.01]
NewModel with  beta 4 : 95% HDI: [-0.01, 0.01]

Old Model with  beta 5 : 95% HDI: [7.63, 9.34]
NewModel with  beta 5 : 95% HDI: [7.28, 9.61]

Old Model with  beta 6 : 95% HDI: [3.28, 4.69]
NewModel with  beta 6 : 95% HDI: [3.02, 5.05]
```

Figure 4: Compare the 95% Credible Intervals Between Models for Each Parameter Term.