Portfolio stock return modeling with Bayesian approach

Israel Diego, Isaac Jones, Daichun Li

4/25/2020

1 Introduction

The purpose of our project is to predict daily average stock returns while accounting for the recent COVID-19 outbreak. Our goal is to achieve reasonable prediction accuracy given recent data. The data is daily closing stock prices from Apple and Microsoft from 2010 to present. This results in a total of 2,592 observations for each stock. We compute the return on an equally-weighted portfolio consisting of the two stocks. All our developed Bayesian time-series models are based on this portfolio return time-series. Figure 1, below, gives a depiction of the three time-series. While the three-time series appear similar they are on three different scales (e.g. observe their y-axis values). The portfolio return has a smaller scale, because of the averaging between two relatively riskier stocks. This indicates that the portfolio return is less risky and serves as an illustration on the benefits of portfolio diversification.

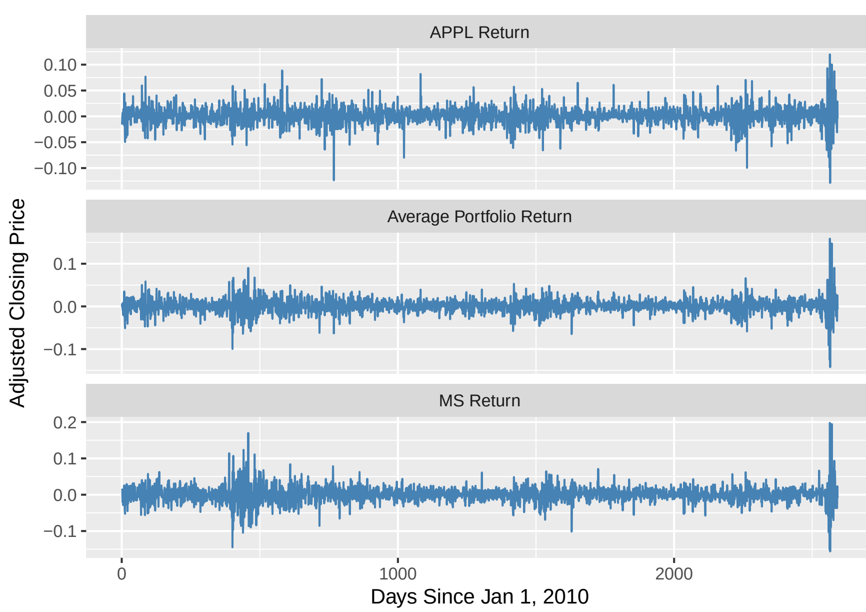

Figure 1: Time Series plot of Apple, Microsoft, and average closing adjusted returns from Jan 1, 2010 to present

2 Methodology

Several models worth considering when fitting the average adjusted returns of Apple and Microsoft: $AR(p), ARMA(p,q)$, and $ARMA(p_a,q_a) - GARCH(p_g,p_g)$. Additionally, we conduct posterior predictive checks and comparisons between models in order to find the best fitting model to the observed data.

2.1 Autoregressive(AR) Model

For stationary time series process $\{X\}_{n=1}^N$, the $AR(p_a)$ model can be expressed by:

$$\phi(B) = 1 - \phi B - \phi_2 B^2 - \cdots - \phi_{p_a} B^{p_a}$$

The ϵ_n is generally the IID Gaussian noise, which means $\epsilon \sim$ i.i.d. $N(0,\sigma^2)$. Similarly, the t-distribution type of white noise is also checked in our report since the heavy-tailed nature of financial return data, which means $\epsilon \sim$ i.i.d. $TDist(\nu, 0, \sigma^2)$ and degree of freedom serves to control the extent of the heavy tail.

2.2 Autoregressive Moving-Average (ARMA) Model

In addition to the AR Model, the error terms can also be correlated and should be modeled in practice. The $ARMA(p_a,q_a)$ model can be expressed by:

$$\phi(B)X_n = \Psi(B)\epsilon_n.$$

Where

$$\psi(B) = 1 - \psi_(B) - \psi_2(B^2) - \cdots - \phi_{p_a} B^{p_a}$$
$$\Psi = 1 + \Psi_1 B + \Psi_2 B^2 + \cdots + \Psi_{q_a} B^{q_a}$$

The error term is generally the IID Gaussian noise $\epsilon_n \sim$ i.i.d. $N(0,\sigma^2)$.

2.3 Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model

To address the self-exciting and local volatility dependence nature of financial data, the GARCH model is necessary. The $GARCH(p_g,q_g)$ model can be expressed by:

$$X_n = \epsilon_n \sqrt{V_n}.$$

Where

$$V_n = \alpha_0 + \sum_{i=1}^{p_g} \alpha_i X_{n-i}^2 + \sum_{j=1}^{q_g} \beta_j X_{n-j}^2$$

Notice the ϵ_n still covers IID Gaussian noise $\epsilon_n \sim N(0,\sigma^2)$ and IID t-distribution noise $\delta_n \sim$ i.i.d. $TDIST(\nu, 0, \sigma^2)$, the self-exciting nature is modeled by the terms in V_n.

2.4 Autoregressive Moving-Average with Generalized Autoregressive Conditional Heteroskedasticity Error(ARMA-GARCH) Model

A more general model is the ARMA model with GARCH error, which captures both the ARMA type trend and local volatility dependence. The $ARMA(p_a, q_a) + GARCH(p_g, q_g)$ model can be expressed by:

$$\phi(B)X_n = \psi(B)\epsilon_n.$$

Where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_{p_a} B^{p_a}$$
$$\psi(B) = 1 + \psi_1 B + \phi_2 B^2 + \cdots + \psi_{q_a} B^{q_a},$$

and

$$\epsilon_n = \sigma_n \delta_n.$$

Where

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{p_g} \alpha_i \epsilon_{n-i}^2 + \sum_{j=1}^{q_g} \beta_j \sigma_{n-j}^2$$

The δ_n is the IID white noise this time. And this report will both cover the Gaussian noise $\delta_n \sim$ i.i.d. $N(0, 2)$ and t-distribution white noise $\delta_n \sim$ i.i.d. $TDIST(\nu, 0, \sigma^2)$ for our ARMA-GARCH model.

3 Posterior Predictive Checks

Sufficient methods and criteria are implemented in posterior predictive checks to ensure the results and prediction of the model we selected is more reliable and accurate than any other models we compared.

3.1 Log-Likelihood (LLH)

The LLH is given by

$$LLH = \log p(y|\hat{\theta}_{MLE})$$

3.2 Akaike Information Criterion (AIC)

The AIC is given by

$$AIC = -2 \log p(y|\hat{\theta}_{MLE}) + 2k,$$

where k is the number of parameters estimated in the model. In addition we generate density plots of posterior prediction simulations and compare them to the distribution of the observed data.

4 Results

In order to illustrate the performance of each model, we will build our models sequentially. We will start with the GARCH(1,1) and compare it to ARMA(1,1)+GARCH(1,1). We will also compare model coefficients and their posterior intervals with results from fitting the models using built-in R functions such as `arima()` and the `fGarch` package. We impose $N(0, 2)$ weakly informative prior on the model coefficients and cauchy$(0, 5)$ on the standard deviation parameters.

4.1 GARCH model

4.1.1 Estimated Coefficients and 95% Credible Intervals

From the trace plots of the coefficients in GARCH(1,1) with Gaussian noise, we can see the coefficients of our GARCH model are convergent after reasonable iterations. Next we look at potential scale reduction $\hat{R}$, which is defined by $\hat{R} = \sqrt{\frac{v\hat{a}r^{+}(\psi|y)}{W}}$, where $v\hat{a}r^{+}(\psi|y)$ is the marginal posterior variance of the estimand and $W = \frac{1}{m}\sum_{j=1}^{m} s_j^2$. We can see that the $\hat{R}$s of the model coefficients are close to 1, which means the potential scale reduction will be low if we proceed with further simulations. Thus, the simulations we have done are enough for inference about the target distribution. The ESSs are also high and agree with our analysis.

Table 1: GARCH(1,1) Stan Fitted Model

	mean	se_mean	sd	2.5%	97.5%	n_eff	Rhat
mu	0.0013	0.0000	0.0003	0.0008	0.0018	4629.606	0.9999
alpha0	0.0000	0.0000	0.0000	0.0000	0.0000	1231.529	1.0011
alpha1	0.1148	0.0003	0.0145	0.0889	0.1454	1790.141	1.0015
beta1	0.8529	0.0005	0.0181	0.8156	0.8845	1395.631	1.0019
lp___	9680.8617	0.0477	1.6812	9676.7635	9683.1020	1243.040	0.9999

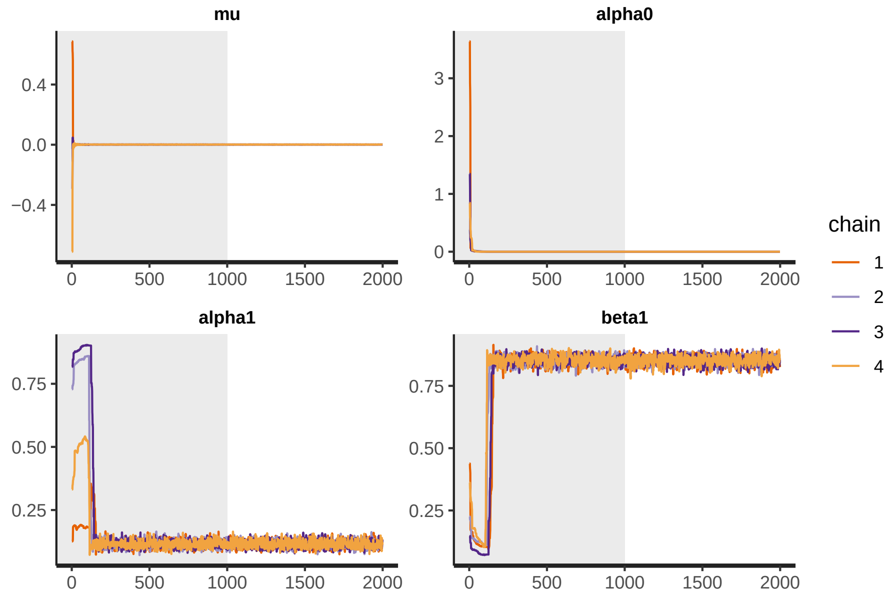

Figure 2: GARCH(1,1) Trace plot

For the GARCH(1,1) we have four parameters to estimate. We notice that the pareto k value, `p_loo`, is about 8.3, which is higher than the the number of parameters we are estimating, and indicates our model is likely to be misspecified.

Table 2: Likelihood of GARCH(1,1) model

	Estimate	SE
elpd_loo	7316.483140	51.2094398673147
p_loo	8.301691	1.07348661926317
looic	-14632.966280	102.418879734629
AIC	-14624.966280	

4.1.2 Maximum Likelihood Estimation Comparison (i.e. garchFit function in R)

Comparing our model results to the maximum likelihood approach, we see that the estimates of the coefficients and standard errors are close to our posterior sampling approach. The log likelihood is a bit higher for MLE approach.

```
##             Estimate   Std. Error   t value     Pr(>|t|)
## mu     1.307548e-03 2.516526e-04  5.195846 2.037904e-07
## omega  7.730011e-06 1.659113e-06  4.659123 3.175596e-06
## alpha1 1.096479e-01 1.377493e-02  7.959960 1.776357e-15
## beta1  8.608806e-01 1.722240e-02 49.986088 0.000000e+00
```

```
## LogLikelihood
##      7323.437
```

4.1.3 Posterior Predictive Checks

Below we show a density overlay plot comparing our posterior samples to the observed distribution. We see that the observed distribution has a higher peak than our posterior samples. This indicates we may want to modify our model to achieve a better fit.

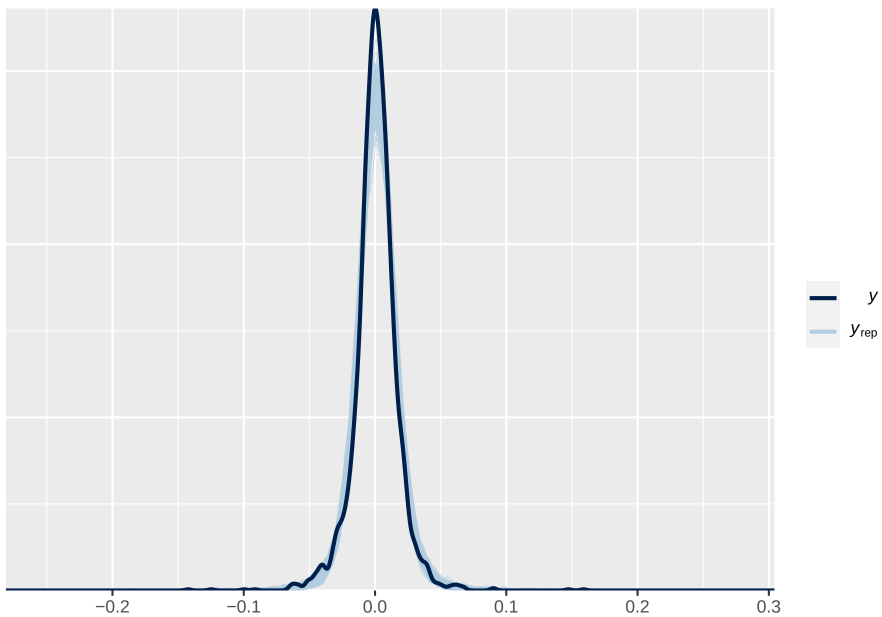

Figure 3: GARCH(1,1) Kernal Density of Data compared to replicates.

4.2 ARMA + GARCH model

We incorporate a time-varying mean component to our model by combining ARMA and GARCH models in order to see whether the fit of the model would improve.

4.2.1 Estimated Coefficients and 95% Credible Intervals

Similarly, for convergence analysis of our ARMA(1,1)-GARCH(1,1) with Gaussian white noise, we can see the trace plots are generally good. However, the variability of β_1 and θ are undesirable and worth reconsidering. By checking $\hat{R}$s, we find they are close to 1, including for β_1 and θ, which suggests that all coefficients converge based on our current simulation iterations. The ESSs are also above 1000 except for β_1 and θ. In summary, most of coefficients converge after reasonable iterations, but there are still some problems with the convergence of β_1 and θ.

Table 3: ARMA(1,1)+GARCH(1,1) Stan Fitted Model

	mean	se_mean	sd	2.5%	97.5%	n_eff	Rhat
alpha	0.0016	0.0000	0.0006	0.0005	0.0028	1261.7069	1.0045
beta[1]	-0.2353	0.0125	0.3808	-0.8116	0.5493	925.2211	1.0039
theta	0.2118	0.0124	0.3801	-0.5738	0.7952	934.6029	1.0039
sigma1	0.0223	0.0002	0.0102	0.0070	0.0462	2215.6235	1.0009
alpha0	0.0000	0.0000	0.0000	0.0000	0.0000	1497.0497	1.0008
alpha1	0.1147	0.0003	0.0143	0.0883	0.1446	1915.1857	1.0030
beta1	0.8533	0.0004	0.0180	0.8163	0.8867	1615.9727	1.0015
lp__	9681.0590	0.0526	1.8024	9676.7039	9683.5578	1174.4846	1.0012

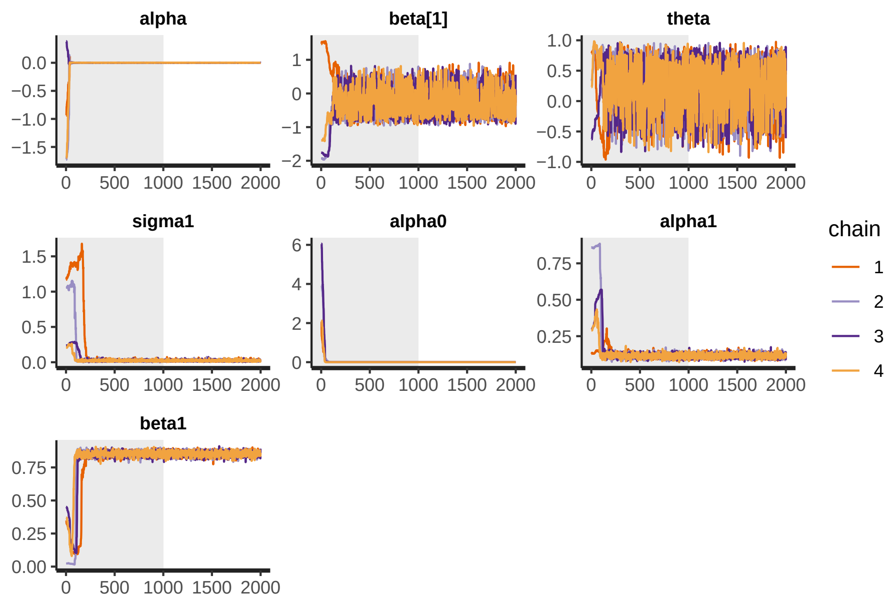

Figure 4: ARMA(1,1)+GARCH(1,1) Trace plot

In the table below, we see that the log-likelihood is 7316, which is the same as the GARCH model. According to the AIC, we would favor the GARCH model since it is a simpler model and has lower AIC than the ARMA+GARCH model.

Table 4: Likelihood of ARMA(1,1)+GARCH(1,1) model

	Estimate	SE
elpd_loo	7316.435411	51.1909903746143
p_loo	9.259267	1.04557166195069
looic	-14632.870821	102.381980749229
AIC	-14620.870821	

4.2.2 Maximum Likelihood Estimation Comparison (i.e. garchFit function in R)

The MLE approach also shows that there is no improvement in the fit by including ARMA component in our model since the log-likelihood barely increased.

```
##             Estimate   Std. Error    t value     Pr(>|t|)
## mu      2.129127e-03 4.924866e-04  4.323217 1.537704e-05
## ar1    -6.231172e-01 2.133781e-01 -2.920249 3.497514e-03
## ma1     5.977725e-01 2.186384e-01  2.734069 6.255688e-03
## omega   7.695875e-06 1.655144e-06  4.649672 3.324638e-06
## alpha1  1.097999e-01 1.380345e-02  7.954528 1.776357e-15
## beta1   8.609031e-01 1.723225e-02 49.958837 0.000000e+00
```

```
## LogLikelihood
##      7324.667
```

4.2.3 Posterior Predictive Checks

The density overlay plot shows that the posterior samples from the ARMA+GARCH model still fall short of the peak of the true distribution.

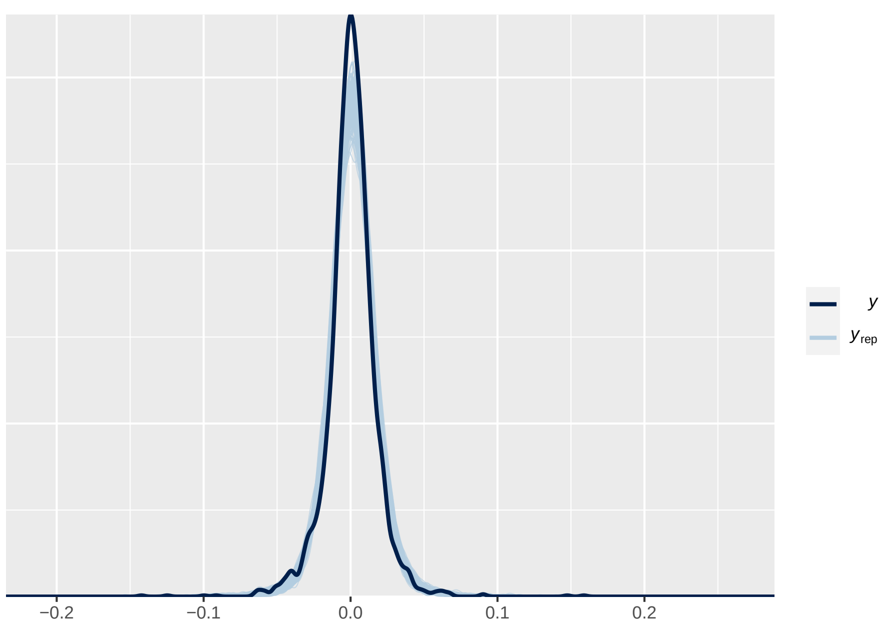

Figure 5: ARMA(1,1)+GARCH(1,1) Kernal Density of Data compared to replicates.

4.3 GARCH model (t-error)

Another approach to improving the GARCH model is to assume a different error distribution. Below we show a QQ plot of the standardized residuals of the GARCH model. We see that the residuals are heavier-tailed than the standard normal. Instead we will fit GARCH model with t-distributed error, where the degrees of freedom for the t-distribution are computed from the kurtosis of the data.

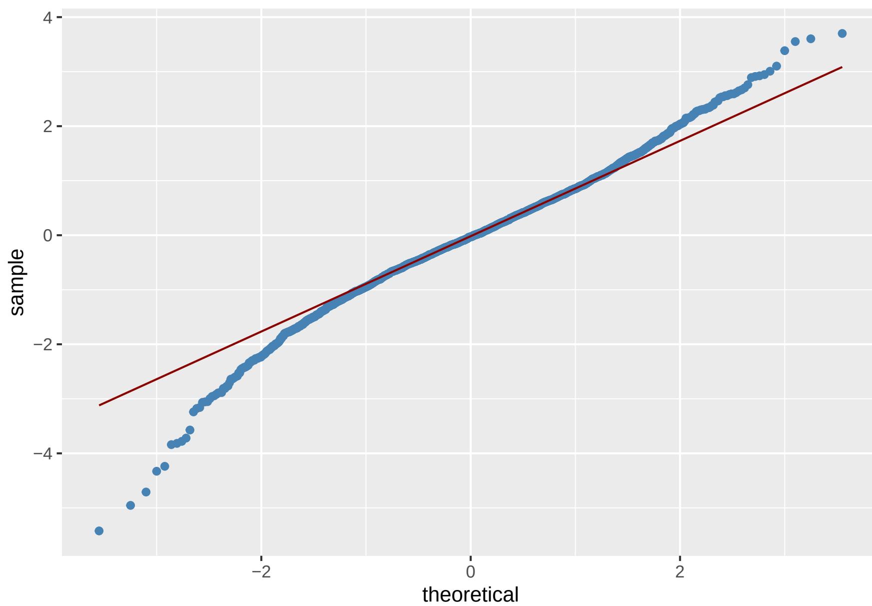

Figure 6: GARCH(1,1) QQ plot of standardized residuals.

4.3.1 Estimated Coefficients and 95% Credible Intervals

For GARCH(1,1) with t-distributed noise, the trace plots shows all coefficients are fine with respect to convergence. Similarly, all $\hat{R}$s are close to 1 and ESSs are high enough for us to draw inference about the target distributions. To summarize, the coefficients converge well under our GARCH(1,1) with t distribution noise model.

Table 5: GARCH(1,1) (t-distributed) Stan Fitted Model

	mean	se_mean	sd	2.5%	97.5%	n_eff	Rhat
mu	0.0013	0.0000	0.0002	0.0009	0.0018	4591.058	1.0002
alpha0	0.0000	0.0000	0.0000	0.0000	0.0000	1156.569	1.0008
alpha1	0.0691	0.0003	0.0109	0.0501	0.0927	1407.963	0.9991
beta1	0.8677	0.0006	0.0206	0.8240	0.9039	1054.996	0.9998
lp__	9867.3885	0.0454	1.6226	9863.3925	9869.5318	1277.943	1.0030

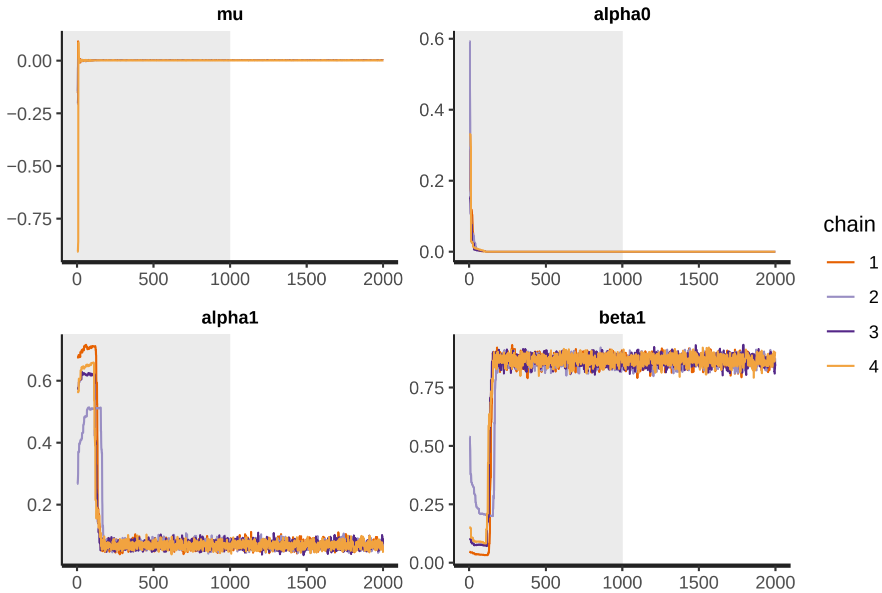

Figure 7: GARCH(1,1) (t-distributed) Trace plot

This model fits the data the best. It has an improved log likelihood of 7371, while the built-in R-estimated model has a slightly higher log-likelihood of 7380.

Table 6: Likelihood of GARCH(1,1) (t-distributed) model

	Estimate	SE
elpd_loo	7371.570745	47.444111229031
p_loo	5.279545	0.451724032473223
looic	-14743.141490	94.888222458062
AIC	-14733.141490	

4.3.2 Maximum Likelihood Estimation Comparison (i.e. garchFit function in R)

```
##              Estimate   Std. Error   t value     Pr(>|t|)
## mu     1.331712e-03 2.376761e-04  5.603055 2.106066e-08
## omega  6.075587e-06 1.746677e-06  3.478369 5.044755e-04
## alpha1 1.040105e-01 1.619819e-02  6.421122 1.352731e-10
## beta1  8.755647e-01 1.909087e-02 45.863009 0.000000e+00
## shape  6.301325e+00 7.694065e-01  8.189851 2.220446e-16
```

```
## LogLikelihood
##      7379.603
```

4.3.3 Posterior Predictive Checks

The density overlay plot below shows that the GARCH(1,1) model with t-distributed errors fits the observed data better than the GARCH(1,1) model with gaussian distributed errors. This largely occurs because stock profolio returns are heavy-tailed.

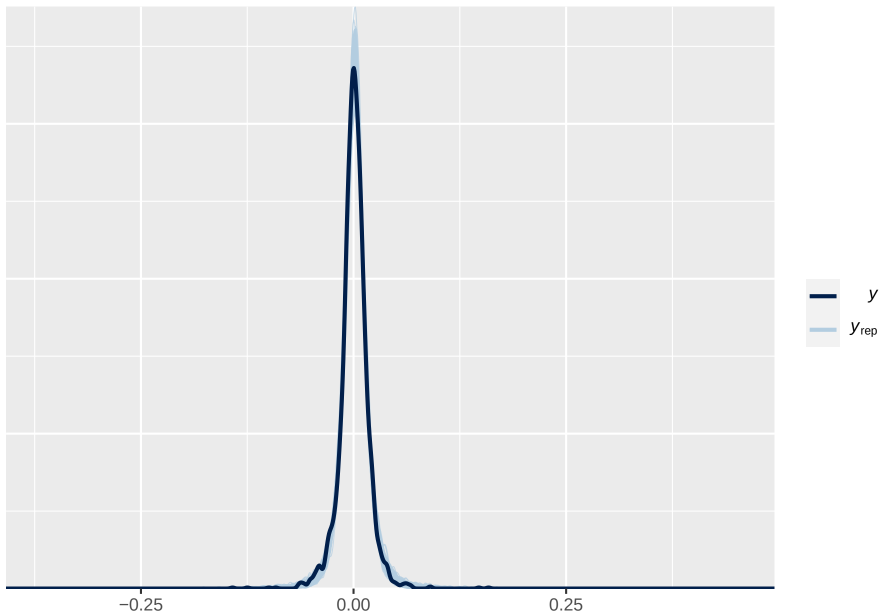

Figure 8: GARCH(1,1) (t-distributed) Kernal Density of Data compared to replicates.

5 Conclusion

We have surveyed several methods for stock return predictions. The GARCH model with t-distributed errors performed best. This implies our portfolio returns have time-varying variance, but not a time-varying mean; if the latter was true the ARIMA(1,1)-GARCH(1,1) model would have performed better. Additionally, we also have indirectly shown some of the limitations of fitting timeseries models in R using maximum likelihood. While, these models perform well when their assumptions are not violated, they are not flexible enough to incorporate additional information. For example, informative priors on parameters or, as we have shown, substituting gaussian distributed errors for t-distributed errors when model fit can be improved.

6 References

[1] Gelman, Andrew, John B. Carlin, Hal Steven. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. Bayesian Data Analysis. Baca Raton: CRC Press, 2015.

[2] Team, Stan Development. "Stan User's Guide." Stan. Accessed April 25, 2020. https://mc-stan.org/docs/2_23/stan-users-guide/time-series-chapter.html.