**Report**

Peng Wang

1. **Model**

We start from the linear regression model without considering the error, which is just least square estimator by minimizing the least square error to predict the slop a and intercept b. Actually, if we assume the noise is independent Gaussian distribution, we will end up with the same parameters for the estimator. This is the special property for independent Gaussian noise which can be proved.

However, if the noise is non-Gaussian, such as heavy-tailed distribution, the least square estimator no longer works. Usually the noise distribution is assumed to have zero mean, thus it gives a strong bias to the regression slop towards zero, which is known as regression dilution. To solve this non Gaussian noise problem, we perform probability likelihood maximization to generate regression parameters.

Without loss of generality, we start by assuming student t distribution for the noise that is , where . Since the noise has conditional mean zero given *x*, we exclude the skewed generalized t distribution. We denote the conditional probability density function of for each *x* as . Given any dataset , , ….. , , we can write the probability density under our model as

In multiplying together the probabilities like this, we are using the independence of the noise given in the problem. This is actually the likelihood and we take log of it

which is a function of four unknown parameters. In the method of maximum likelihood, we maximize the likelihood with respect to those parameters. The parameters correspond to likelihood maximum are also the regression estimator parameters.

Besides student t distribution, we have also explored the stable distribution, which is commonly used in financial models. We apply the general α-Stable distributions for the noise and the probability density function is written as six parameters. Due to constraint of zero mean of the noise, for simplicity, we set and . Then the log likelihood function can be written in the same way

and this set of parameters can also be obtained by minimizing the log likelihood with respect to those four parameters.

For further study, mixed tempered stable distributions can be applied to avoid infinite moment of stable distribution, which is left out here.

1. **R programming**

We start from lm() function in R to perform least square minimization and calculate the sum of square residuals (SSR) to evaluate error for the predictor. The we use optim() function to minimize the log likelihood for both cases. We choose the intercept and slop from least square estimator as initial starting point. To avoid singularity during minimization, we use L-BFGS-B algorithm for nonlinear optimization, followed by checking convergence and symmetric positive definiteness for the Hessian matrix to exclude saddle point.

1. **Comparison**

The original dataset for data\_1\_1 and three fitting curves are plotted in Fig. 1. The t distribution and stable distribution are very close; both have larger slop than the least square predictor. The regression dilution effect is significantly removed by the t distribution and stable distribution, which show a better agreement with the overall trend of the original dataset. The values for the slop and intercept, as well as the SSR are presented in Table 1. **Stable distribution model has the largest slop value and smallest intercept, which is also applied on other four datasets to generate final results for those parameters.** However, the least square estimator has the lowest SSR since it is designed to minimize the sum of square errors, while the other two models are designed to maximize the likelihood.

1. **Model Selection**

For statistical inference, such as establish the relation between two variables based on their regression coefficients, the t distribution model and stable distribution model are appropriate, since they correct the regression dilution and reveal the true latent relation between the independent variables and dependent variables.

For prediction modelling, the least square estimator is appropriate since we are only considering the subsequent prediction of dependent variables. The regression model acts as a black box taking every input and output the corresponding value. All the errors and biases are inside the black box as a whole to provide the output.

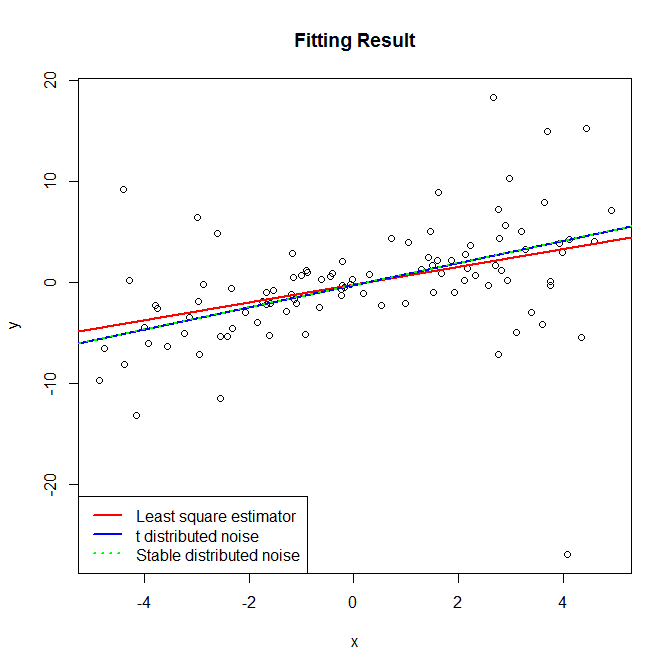
Figure 1. Original data and three fitting curves (For data\_1\_1).

Table 1. Comparison between different models (For data\_1\_1).

|  |  |  |  |
| --- | --- | --- | --- |
| Model | a | b | SSR |
| lse | 0.8808088 | -0.2695265 | 2803.671 |
| t distribution | 1.0916802 | -0.2941103 | 2835.208 |
| stable distribution | 1.0980066 | -0.3011909 | 2837.116 |