What matters to get the best mutual fund risk-adjusted returns?

Gerardo Zamora
UBC #: 25330168

1. Abstract

I will examine certain characteristics of mutual fund managers and determine if they produce a change in risk-adjusted excess returns. The main characteristics are experience and education. Experience will be measured by utilizing the tenure and age of the manager, while education will be analyzed by the average SAT scores and whether they have an MBA. After producing the result this characteristics do influence on the change in risk-adjusted excess returns, however it may not be as significant as it was expected.

2. Introduction

We can say without almost any doubt that the returns in mutual funds is contingent on who manages them. Still we may ask, what is what makes a mutual fund manager produce better risk-adjusted returns? This question is of grand importance for mutual funds as it could help them maximize their returns by choosing their managers based on empirical evidence. Not only it is important for mutual funds but also for the academic field of finance and economics. Not for nothing it has spanned multiple research papers that tackle this same question, each with diverse strategies. This paper will emphasize mainly on performing a regression analysis with four factors for each manager, which are their average SAT scores, whether they have an MBA or not, their age, and how many years they have been managers. This to measure their education and experience. It is predicted that all this variables count positively in the risk-adjusted excess returns of the manager. To validate our regression model, robust checks will be performed. After the model is validated results will be analyzed and conclusions will be drawn. The major contribution to this research study comes from the paper of Judith Chevalier and Glenn Allison [1].

3. Data Description

The data I will be using in this analysis consists of 2029 different observations. Each observation is mapped to different managers, which include five different entries. Each of these entries are used as variables for the model of this study. The variables are as follow:

- ***Return*** is a negative and positive continuous variable that represents the risk-adjusted excess returns derived by each mutual fund manager. Values range [-30.51,32.09] with an average risk-adjusted excess returns of -0.554485 and a standard deviation of 8.358881.
- ***SAT*** is a positive continuous variable that indicates average SAT score of each mutual fund manager. Values range [657,1662] with an average score of 1142.003 and a standard deviation of 143.9482.
- ***MBA*** is a categorical variable that either takes the value of 0 and 1, which describes whether each mutual fund manager has not or has an MBA respectively. Values are fixed as stated previously with an average of 0.5963529 and a standard deviation of 0.4907492.
- ***Age*** is a positive continuous variable that constitute the age of each mutual fund manager. Values range [27,59] with an average age of 42.3312 and a standard deviation of 4.841952.
- ***Tenure*** is a positive continuous variable that state the amount of time measured by years each mutual fund manager has been appointed as manager. Values range [0,8] with an average year-span of 4.248891 and a standard deviation of 1.170249.

It is worth noting that ***SAT*** is used mainly to describe the quality of education received, since a higher ***SAT*** generally leads to a better education, and it is easier to model it rather than trying to rank each university. Further information regarding the data to keep in mind is that the distribution of the continuous variables appear to be normally distributed. Nevertheless, I performed a normality check at section ***5.1 Normality Analysis*** that confirms this in detail. Moreover, in view to the fact that the variables follow a normal distribution, they do not need to be transformed to fit into the model as long as we prove linearity as I did in ***5.4 Linearity Analysis***. Finally, last point to be noted is that certain variables may be dropped depending on the model results.

4. Methodology

The methodology incorporated in this study is a liner regression model constructed with a response variable, in this case ***Return***, while ***SAT***, ***MBA***, ***Age,*** and ***Tenure*** are the explanatory variables. The predicted regression model can be observed as the following:

$$\text{Return} = \beta_0 + \beta_1\text{SAT} + \beta_2\text{MBA} + \beta_3\text{Age} + \beta_4\text{Tenure} \ (1)$$

First, it is imperative to confirm that all explanatory variables present statistically significant results to employ them in the use of the model. Throughout the study a significance level ($\alpha = 0.05$) will be used. To test for significance I regressed the predicted model in STATA, and obtained the following results:

Table 1: Regression Model I

Return Regression Model					
Variables	**Coefficient**	**Standard Error**	**t**	**p > | t |**	**R^2**
β_0	-1.147635	2.195444	-0.52	0.601	
SAT	0.0050736	0.0012814	3.96	0.000	
MBA	0.6744004	0.3759601	1.79	0.073	
Age	-0.1405739	0.0424218	-3.31	0.001	
Tenure	0.0818061	0.1755348	0.47	0.641	
					0.0151

For now we are only interested in one result, and that is p > | t |. Taking a look at the table it is clear that ***MBA***, and ***Tenure*** p-value exceeds the significance level of 5%. This implies we can't reject the null hypothesis at the 95% confidence level, making these variables statistically insignificant for this particular model. This suggest dropping these distinct variables from the model altogether. The new predicted model by dropping the aforementioned variables is as follows:

$$\text{Return} = \beta_0 + \beta_1\text{SAT} + \beta_2\text{Age} \ (2)$$

Now, we must confirm the model has not changed drastically. I ran a regression in STATA on the new predicted model, and the results are as follows:

Table 2: Regression Model II

Return Regression Model					
Variables	**Coefficient**	**Standard Error**	**t**	**p > |t|**	**R^2**
β_0	-0.8592848	2.187331	-0.39	0.694	
SAT	0.005098	0.0012814	3.98	0.000	
Age	-0.1303327	0.0380954	-3.42	0.001	
					0.0134

First, we must make sure the the variables in the model did not undergo drastic changes, which clearly they did not. Glancing over the table it is confirmed the p-value of the variables is still statistically significant and therefore no more adjustments are needed in this predicted model. The final linear regression model that will be used throughout the paper is as follows:

$$\text{Return} = -0.8592848 + (0.005098)*\text{SAT} + (-0.1303327)*\text{Age} \quad (3)$$

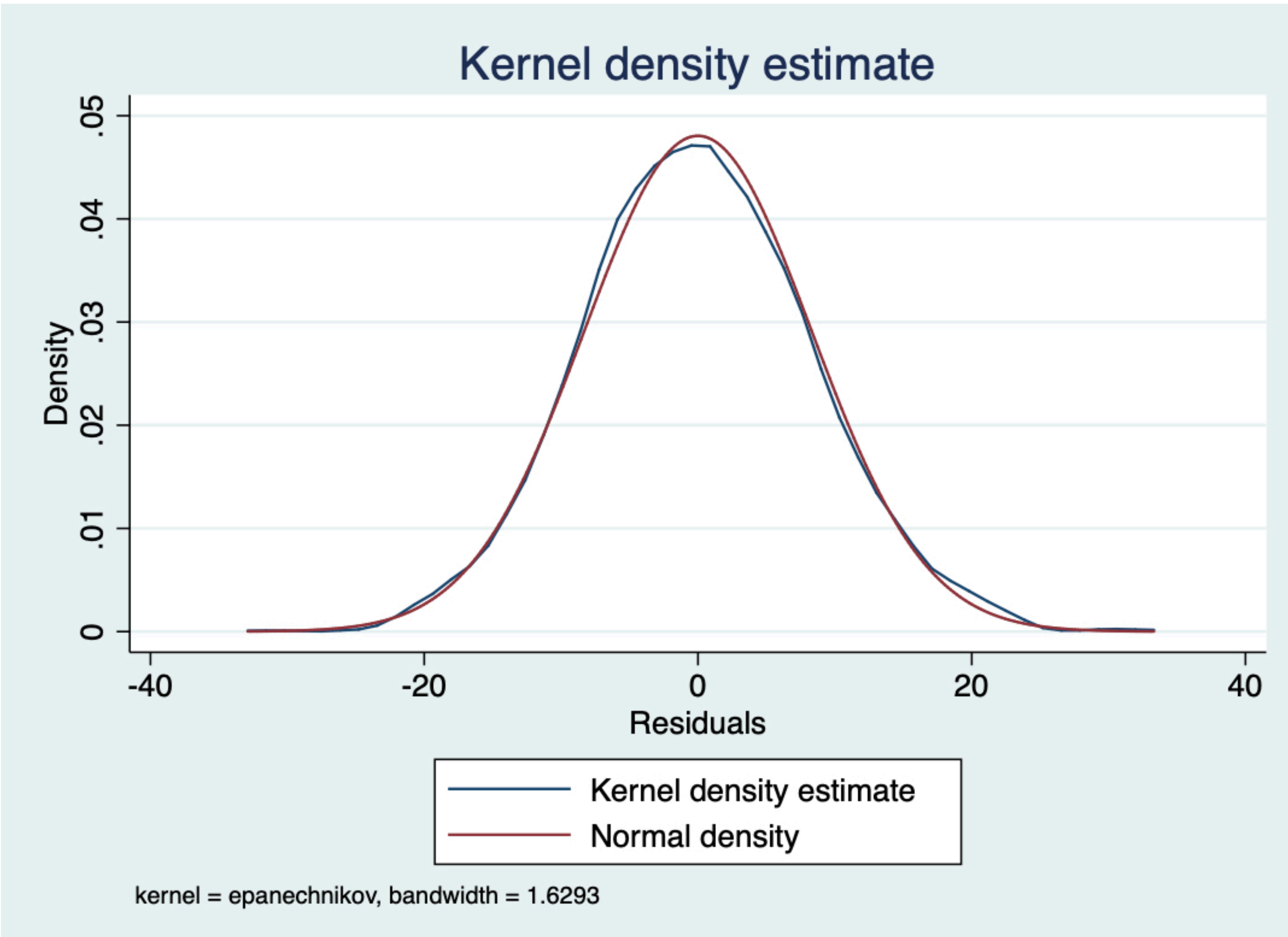
5. Robustness checks

To prove the validity of our model we must first perform diverse tests to prove its correctness. The tests that will be performed to confirm the correctness of our model are the following: normality analysis, homoskedasticity analysis, multicollinearity analysis, and linearity analysis.

5.1. Normality Analysis

To ensure our data fits our current model we need to confirm the residuals follow a normal distribution. To do this we have to predict the residuals using STATA and then perform a kernel density estimate. If the residuals follow a normal distribution, then our data fits our current model, which would imply it exists a zero-conditional mean in our model. The kernel density can bee seen in the following figure:

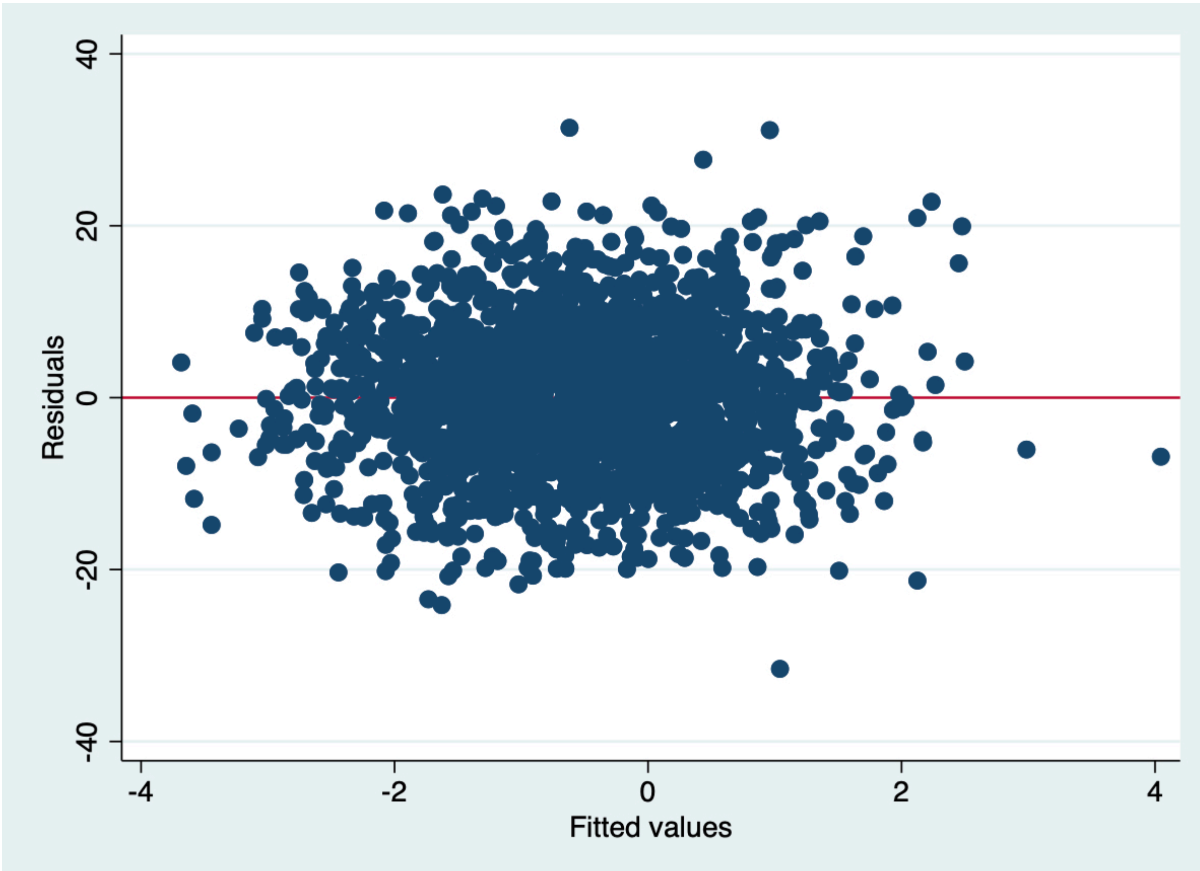
Figure 1: Kernel Density Estimate



As it can be seen in the table above the distribution of the residuals is extremely close as the normal distribution, so it can be implied the residuals follow a normal distribution.

To ensure we are not overestimating the goodness of a fit we need to confirm there exists homogeneity in the variance of the residuals. The first test we can perform is plotting the residuals v.s. fitted values and find the best fitted line. This is done entirely by STATA and it produces the following graph:

Figure 2: Residuals v.s. Fitted Values



Analyzing this graph we can fairly say there exists no correlation at all, however we can further perform a second test to confirm this. The Breusch-Pagan test assumes homoskedasticity as the null-hypothesis. To confirm this we must square the regression of the residuals on the explanatory variables. Doing as previously mentioned in STATA produces a p-value of 0.1642. As this p-value is greater than our significance value of 0.05, we do not reject the null hypothesis. Therefore, we can deduce there is not enough statistical evidence to presume the variance of the residuals is not homogenous.

5.3. Multicollinearity Analysis

To ensure we are not measuring the same effect with two variables we must confirm there exist no collinearity between the explanatory variables. We can examine this by producing a pairwise correlation in STATA as it follows:

Table 3: Pairwise Correlation

Variable	SAT	Age
SAT	1.0000	
Age	-0.0018	1.0000

As it can be observed the correlation between the two variables is merely an extremely low number in absolute terms. Usually it may be worrisome when the absolute value is greater than 0.5, in this case it is 0.0018 in absolute value. We can further expand on multicollinearity by analyzing the Variance Inflation Factor (VIF) that could tell us if a variable is redundant. This can be computed by STATA and the output is as it follows:

Table 4: Variance Inflation Factor

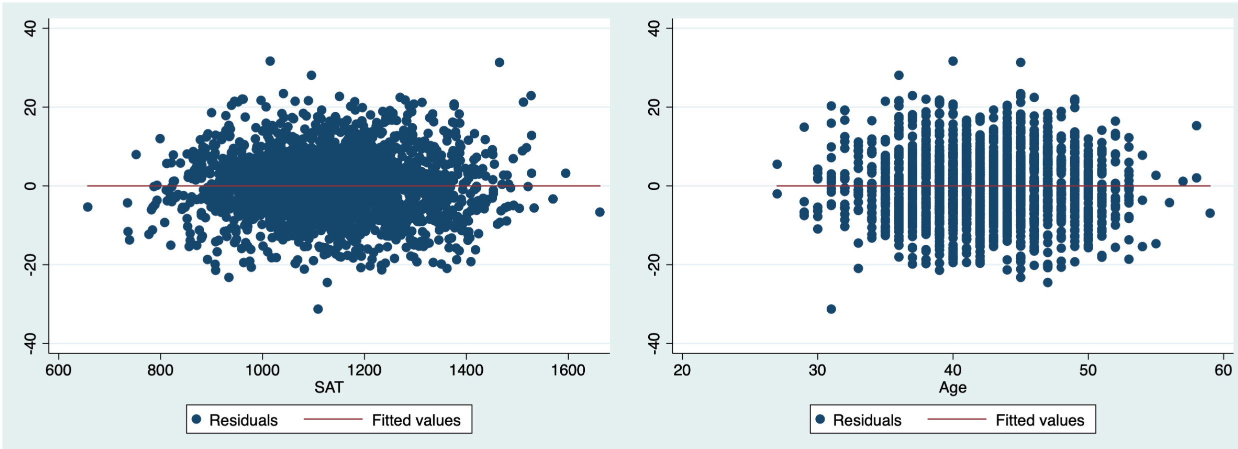
Variable	VIF	1/VIF
SAT	1.00	0.999997
Age	1.00	0.999997
Mean VIF	**1.00**	

Generally a VIF greater than 10 suggest collinearity could exist, however in this case it is just 1. This implies no variable is redundant in our model.

5.4. Linearity Analysis

The last analysis involves examining whether there is linearity and exogeneity holds. For this we must plot the residuals and certify that the residuals have no relationship with the explanatory variables. For this we must plot the residuals and each of the the explanatory and find the fitted values. This graphs can be created in STATA and are the following:

Figure 3 : Residuals v.s. SAT and Age



By analyzing the graphs we notice there is no apparent trend, therefore, we can confirm linearity and exogeneity in our model.

6. Results

After the robustness checks to prove the correctness of our model, we can finally analyze the results of this paper. First, we discovered that our first model (1) in part ***4. Methodology*** contained variables that were statistically insignificant. This variables were ***Tenure*** and ***MBA*** both of which had p-values exceeding our significance level of 0.05. This results can be seen in Table 1 in part ***4. Methodology***. This implies those variables most probably do not explain variation in risk-adjusted excess returns. Next, we simplified the model by removing the aforementioned variables. The second model (2) in part ***4. Methodology*** used only the variables ***SAT*** and ***Age***. The results of this model can be seen in Table 2 in part ***4. Methodology***. This results help us model our final equation (3) in part ***4. Methodology*** to describe the change in ***Return*** based on the changes in ***SAT*** and ***Age***. The coefficient of ***SAT*** is 0.005098, this means keeping everything else constant for each increase of 1 unit in the average SAT score resulted in 0.005098 unit increase in ***Return***. Meanwhile, the coefficient of Age is -0.1303327, this means keeping everything else constant for each increase of 1 year in Age resulted in a 0.1303327 unit decrease in ***Return***. Also, the constant has a coefficient of 0.8592848, this means that if ***SAT*** and ***Age*** were 0, which would be impossible to find, then we would find ***Return*** being equal to 0.8592848. Then, we could analyze the value of R^2 in the aforementioned table and notice it is significantly small. The value is 0.0134, this implies that our final model only explains 1.34% percent in the variation in ***Return***. Finally, our results in part ***5. Robust Checks*** imply that our model has normally distributed residuals, homogenous variance in residuals, its explanatory variable do not convey collinearity, and the data does fit linearly. All of this makes our final model statistically correct and therefore accurate and let us validate our results previously mentioned.

7. Conclusion

We can draw several conclusions by analyzing the results of our final model. The model clearly indicates that the age and the average SAT scores of the mutual fund manager generate a change in the risk-adjusted excess returns they produce. Average SAT scores affect the change positively, which is expected as a manager with higher SAT score probably received a better education. What was not expected is that actually age decreases the performance of the manger; one will expect that with more experience the manger will manage the fund better. This may be because managers close to retirement care less about their performance as they do not have to worry about job security. Another conclusion we can draw is by analyzing the results of our original model before we dropped out variables. We could imply that tenure and age of the manager are correlated and thus can be explained only by one variable same goes with the average SAT and MBA as both represent education received. Coming back to our final model, while it does explain some variation in the risk-adjusted excess returns , it is fairly small, so even with extreme cases the change in risk-adjusted excess returns is almost negligible. I have various hypotheses on why this can be the case. Mainly my best educated guess is the great unpredictability of the market is what offsets the characteristics a certain manager may have. This would mean that most of the variation in the risk-adjusted excess returns actually comes from “luck”. Nevertheless, this may be not the case. We could further improve this study by selecting better categorical variables that measure the skill of the manager even further. My suggestion would be to include the frequency a manager is informed about the current economic and business news, their average GPA in investment and macroeconomics courses, and finally financial designations such as CFP, CFA, CIM, PFP or R.F.P etc.

8. Bibliography

[1] Chevalier, Judith, et al. Are Some Mutual Funds Managers Better Than Others? Cross-Sectional Patterns in Behavior and Performance. National Bureau of Economic Research, 1996.

9. Appendix

```
-------------------------------------------------------------------------------
      name:  <Gerardo Zamora>
       log:  /Users/gerardo/Documents/ECON326A2/Paper.log
  log type:  text
 opened on:  28 Mar 2019, 17:59:29

.
. use "/Users/gerardo/Documents/ECON326A2/MutualFundManagers.DTA",

.
. summ

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      Return |      2,029    -.554485    8.358881     -30.51      32.09
         SAT |      2,029    1142.003    143.9482        657       1662
         MBA |      2,029    .5963529    .4907492          0          1
         Age |      2,029     42.3312    4.841952         27         59
      Tenure |      2,029    4.248891    1.170249          0          8

.
. regress Return SAT MBA Age Tenure

      Source |       SS           df       MS      Number of obs   =     2,029
-------------+----------------------------------   F(4, 2024)      =      7.75
       Model |  2137.16895         4  534.292237   Prob > F        =    0.0000
    Residual |   139560.99     2,024  68.9530582   R-squared       =    0.0151
-------------+----------------------------------   Adj R-squared   =    0.0131
       Total |  141698.159     2,028   69.870887   Root MSE        =    8.3038

------------------------------------------------------------------------------
      Return |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         SAT |   .0050736   .0012814     3.96   0.000     .0025606    .0075866
         MBA |   .6744004   .3759601     1.79   0.073    -.0629087     1.41171
         Age |  -.1405739   .0424218    -3.31   0.001    -.2237689   -.0573789
      Tenure |   .0818061   .1755348     0.47   0.641    -.2624417    .4260539
       _cons |  -1.147635   2.195444    -0.52   0.601    -5.453201    3.157931
------------------------------------------------------------------------------

.
. regress Return SAT Age

      Source |       SS           df       MS      Number of obs   =     2,029
-------------+----------------------------------   F(2, 2026)      =     13.79
       Model |  1903.12753         2  951.563766   Prob > F        =    0.0000
    Residual |  139795.031     2,026   69.000509   R-squared       =    0.0134
-------------+----------------------------------   Adj R-squared   =    0.0125
       Total |  141698.159     2,028   69.870887   Root MSE        =    8.3067

------------------------------------------------------------------------------
      Return |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         SAT |    .005098   .0012814     3.98   0.000      .002585     .007611
         Age |  -.1303327   .0380954    -3.42   0.001    -.2050429   -.0556225
       _cons |  -.8592848   2.187331    -0.39   0.694    -5.148937    3.430367
------------------------------------------------------------------------------

.
. predict r, resid
```

```
.
. kdensity r, normal name(kerneldensity)

.
. rvfplot, yline(0) name(homo)

.
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of Return

         chi2(1)      =     1.94
         Prob > chi2  =   0.1642

.
. correlate SAT Age
(obs=2,029)

             |      SAT      Age
-------------+------------------
         SAT |   1.0000
         Age |  -0.0018   1.0000


.
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
         Age |      1.00    0.999997
         SAT |      1.00    0.999997
-------------+----------------------
    Mean VIF |      1.00

.
. graph twoway (scatter r SAT) (lfit r SAT), name(rSAT)

.
. graph twoway  (scatter r Age) (lfit r Age), name(rAge)

.
.
.
.
.
.
.
.
.
.  log close
      name:  <Gerardo Zamora>
       log:  /Users/gerardo/Documents/ECON326A2/Paper.log
  log type:  text
 closed on:  28 Mar 2019, 17:59:35
----------------------------------------------------------------------------------
```