Signal Correlation Test Brief Notes for Factor Model

Niyu Jia

Introduction

Based upon the perspective that correlation can show the predictive power of the signal, I use the local linear regression to find the non-linear correlation between the signal and SPX.

1. Data Cleaning

First, I lag the signal data to make sure that we do not use future data in prediction. Then I used quantile to deal with the outliers.

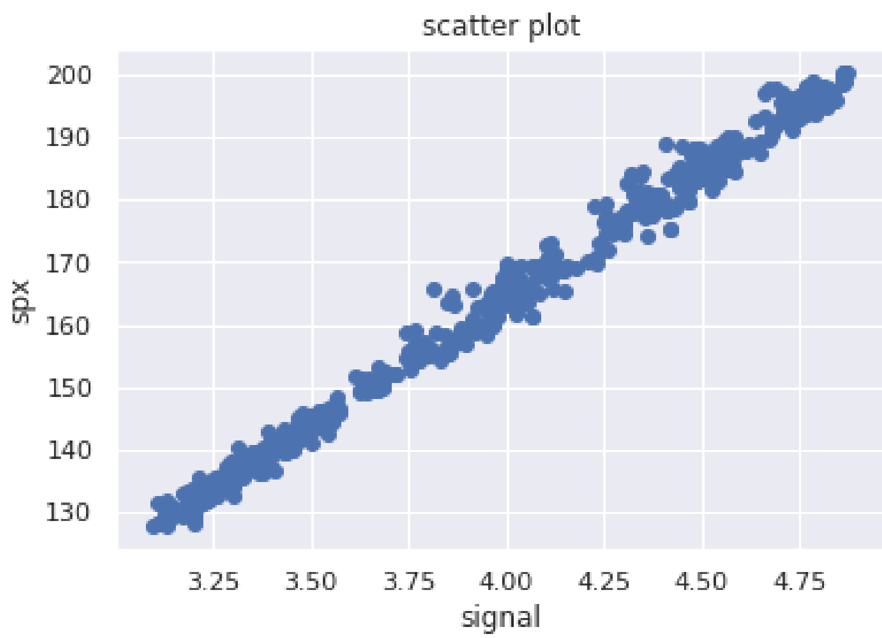
Here I present the outliers and their index:

Index	Signal
304	-3.8
460	430
461	433
571	0.00456
572	0.455
573	0.456
662	4.87
666	4.88

Index	ClosePrice
1	128
421	619
422	619
423	710
666	201

Note that we may also delete some normal data due to the quantile we choose. But overall it will not affect our correlation analysis.After being detected as outliers, the observations will be deleted (the whole row). Then we keep total 654 observations for next step.

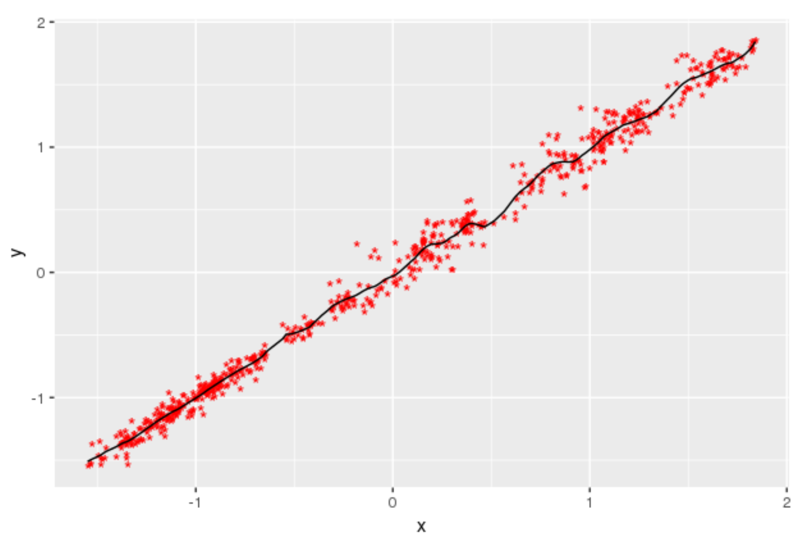
Here is a simple scatter plot of cleaned spx against signal. It seems to be some linear correlation between the two columns. We will do a non-parametric local linear regression to check the correlation.



2. Local Linear Regression

Of course, the first step I scaled the signal and spx data. I denote the signal data as X and the spx close price as Y for notation.

For the regression, I choose the bandwidths by cross validation. Here is the plot of the local linear regression. The black line is the true value and the red star is the estimated value by regression. All data is scaled.



Local correlation formula:

$$\rho(x) = \frac{\sigma_1 \beta(x)}{\left(\sigma_1^2 \beta^2(x) + \sigma^2(x)\right)^{\frac{1}{2}}}$$

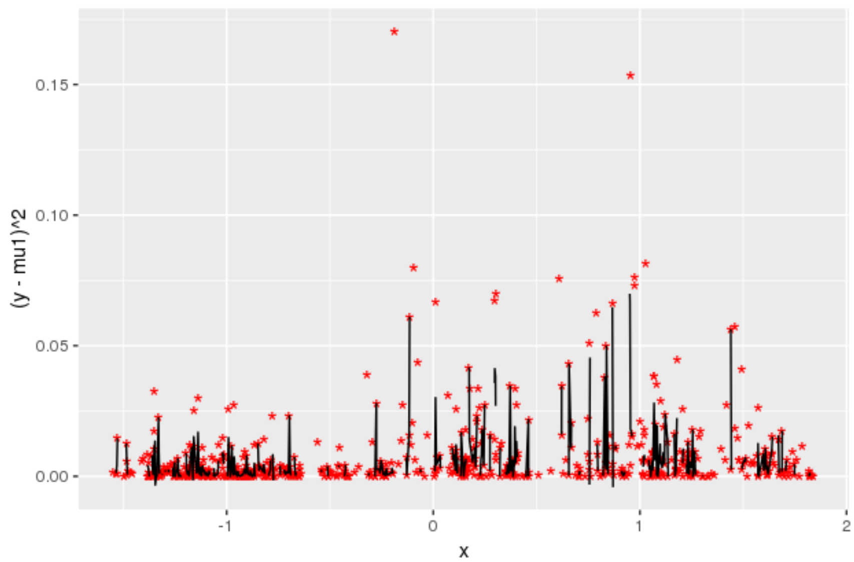
where:

$$\sigma_1^2 = var(X)$$

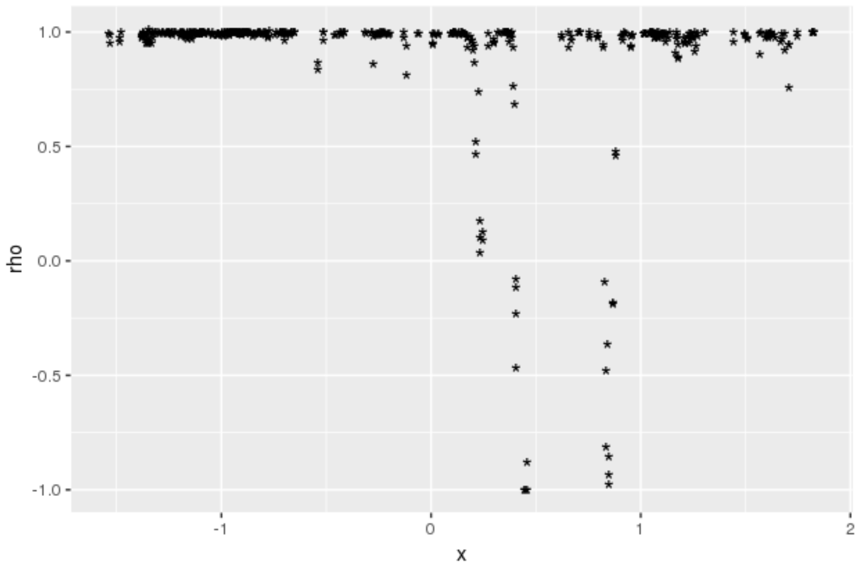
$$\beta(x) = \frac{d\,E(Y|X=x)}{dx}$$

$$\sigma^2(x) = var(y|X=x)$$

In order to calculate the local correlation, the biggest problem is to estimate the conditional variance term. And for that, we use another local linear regression. We regress the residual square against the X variable and get following regression plot:



Again, the black line is the real residual squares and the red stars are the estimated residual squre, which is also the $\sigma^2(x)$ we want in the local correlation formula. Now we can easily compute the local correlation. Following plot shows the correlations and its corresponding X:



Conclusions:

When scaled x is in [0,1], the are little correlation between the signal and spx closing price. But in other region, there appears strong correlation (approximately perfectly correlated!) between the signal and spx.

Therefore my suggestions for portfolio managers would be:

1. If your standardized signal locates between 0 and 1, please do not rely on this signal. It shows no predication power of the spx
2. If your standardized signal locates outside of [0,1], you can assume there is large positive correlation between the signal and spx.
3. Single signal is not reliable enough for complex real market. Other techniques should also take into consideration

3. Viability and Shortcomings

- Viability: this method is asymptotic unbiased and can take care of out-of-sample performance.
- Shortcomings:

  1. Ill-performed for small data set

  2. Bandwidth selection is always sensitive in regression step

4. Reference

[1] Fan, Jianqing, and Qiwei Yao. *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media, 2008.

[2] Doksum, Kjell, et al. "Correlation curves as local measures of variance explained by regression." *Journal of the American Statistical Association* 89.426 (1994): 571-582.