Analysis of ZB After Hours Price Action

By Alizerin

Introduction

I aimed to analyze the after hours price action of the 30 year bond futures contract ZB. Particularly, how often the current day's opening price was the same or close to the previous day's close. Hypothesis: regardless of how far the price action deviates away from the close in after hours trading, there is a tendency for it to regress back to that previous day's closing price by the time the next day's regular trading hours start. As it turns out, it is quite common for the price to regress back to the close (within 4 ticks). This happens 70% of the time. Perhaps even more interesting, I was able to derive two probability distributions: one pertaining to how far the price will initially deviate from the daily close and the other pertaining to how far back to the close price is likely to regress. With these two distributions I was able to create a model that can calculate an expectation value for a trading strategy. I then backtest this strategy showing confluence with the model and data. The backtested result was an average profit of $25,626 per year trading one contract each trade.

Data and Methods

I used Yahoo Financeas the source for the data, turning their website into an API. I am using a time series data set with 'high', 'low', 'open' and 'close' data fields. My current analysis uses two years worth of hourly data, the most Yahoo would serve. Later, I will concatenate multiple two year groups to get better statistics and look for pattern changes over different time periods.

First, I shifted the data set to reflect Dubai time. In this time zone, after hours trading happens on the same day (vs spanningtwo2 days in USA time zones) which allows me to easily group by date. Then I calculated five basic metrics:

- *max-deviation*: how far the price deviated from the previous close during after hours trading.
- *deviation-direction*: whether price deviated above or below the close.
- *regression*: once reaching *max-deviation*, how far the price retraced back towards the close.
- *regression-ratio*: *regression / max-deviation.*
- *frequency*: how often price regressed back to within 4 ticks of the close.

Next, I fit a distribution function to both the *max_deviation* data and the *regression* data. For the two PDFs (probability density functions) I used the SciPy library for python to test multiple distributions and get the best fit parameters. The fitting algorithm used was Maximum Likelihood Estimation. Then I used Sum of Squared Errors as well as residuals histograms to determine the best distribution out of the following distributions:

- Burr
- Burr12
- Gumbel
- Log Normal
- Weibull Min
- Weibull Max
- Gamma
- Fréchet (Inverse Weibull)

These distributions are common knowledge in statistics and well studied. More information on these distributions and the fitting methods used can be found in the SciPy Docs as well Wikipedia, which is luckily still useful for Mathematics.

For all the data and code, I have shared my Python Notebook. While in this document I only show the best fit curves, in the code you can see all the different fit curves, how they compare and the algorithms I used to calculate them.

Results

max-deviation **(normalized) [%]**	
Mean	0.64
Median	0.56
Standard Deviation	0.38
regression **(normalized) [%]**	
Mean	0.57
Median	0.47
Standard Deviation	0.40
regression-ratio	
Median (expressed as %)	91.30%

frequency	
(expressed as %)	70.11%
deviation-direction	
Above (raw count)	306
Below (raw count)	246

Table 1: *Shows the Mean, Median and Standard Deviation Values for the main metrics.*

After binning the *max_deviation* data and creating a histogram, the data shaped into a positively skewed distribution. The best fit came from the Fréchet distribution. Figure 1 below shows the data histogram, best fit curve and the residuals histogram. The PDF for this distribution is

$$PDF_{Fréchet} = \frac{\alpha}{s}\left(\frac{x-m}{s}\right)^{-1-\alpha} e^{\left(\frac{x-m}{s}\right)^{-\alpha}},$$

where the best fit parameters are:

$$\alpha = 23.87$$
$$s = 6.63$$
$$m = -6.17$$

The median and mean for the Fréchet, using the fit parameters, were calculated as

$$mean_{Fréchet} = 0.63, \quad median_{Fréchet} = 0.56\ .$$

For the data set, the $mean_{data} = 0.64$ and $median_{data} = 0.56$. These numbers are consistent and give me some more confidence that the curve is fitting the data well.

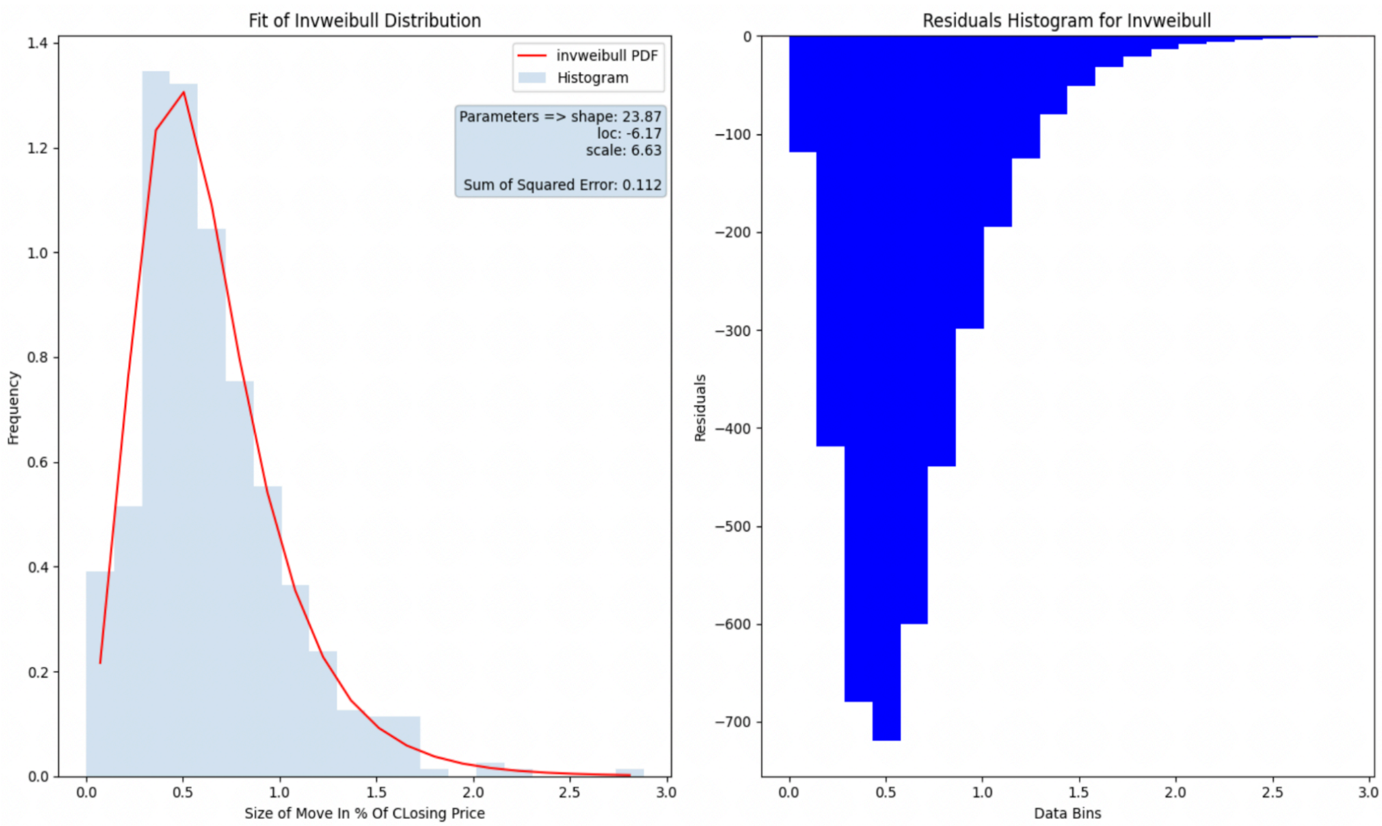


Figure 1: Shows best fit curve for max_deviation data

Similarly, after binning the *regression* data, it also shaped into a positively skewed distribution. The best fit for the *regression* data came from the [Log Normal distribution](). Figure 2 shows the best fit for this data along with the residuals. As is implied in the name, the Log Normal distribution is similar to the Normal Distribution but uses the natural log of the data.

$$PDF_{lognormal} = \frac{1}{\alpha(x-m)\sqrt{2\pi}} e^{-\left(\frac{ln^2\left(\frac{x-m}{s}\right)}{2\alpha^2}\right)}$$

And the best fit parameters are:

$$\alpha = 0.57$$
$$s = 0.54$$
$$m = -0.07$$

The mean and median for the Log Normal were calculate to be

$$mean_{lognorm} = 0.57, \; median_{lognorm} = 0.47.$$

The data set $mean_{data} = 0.57$ and $median_{data} = 0.47$. These numbers are the same within two significant figures of precision, once again giving confidence that the model fits well.

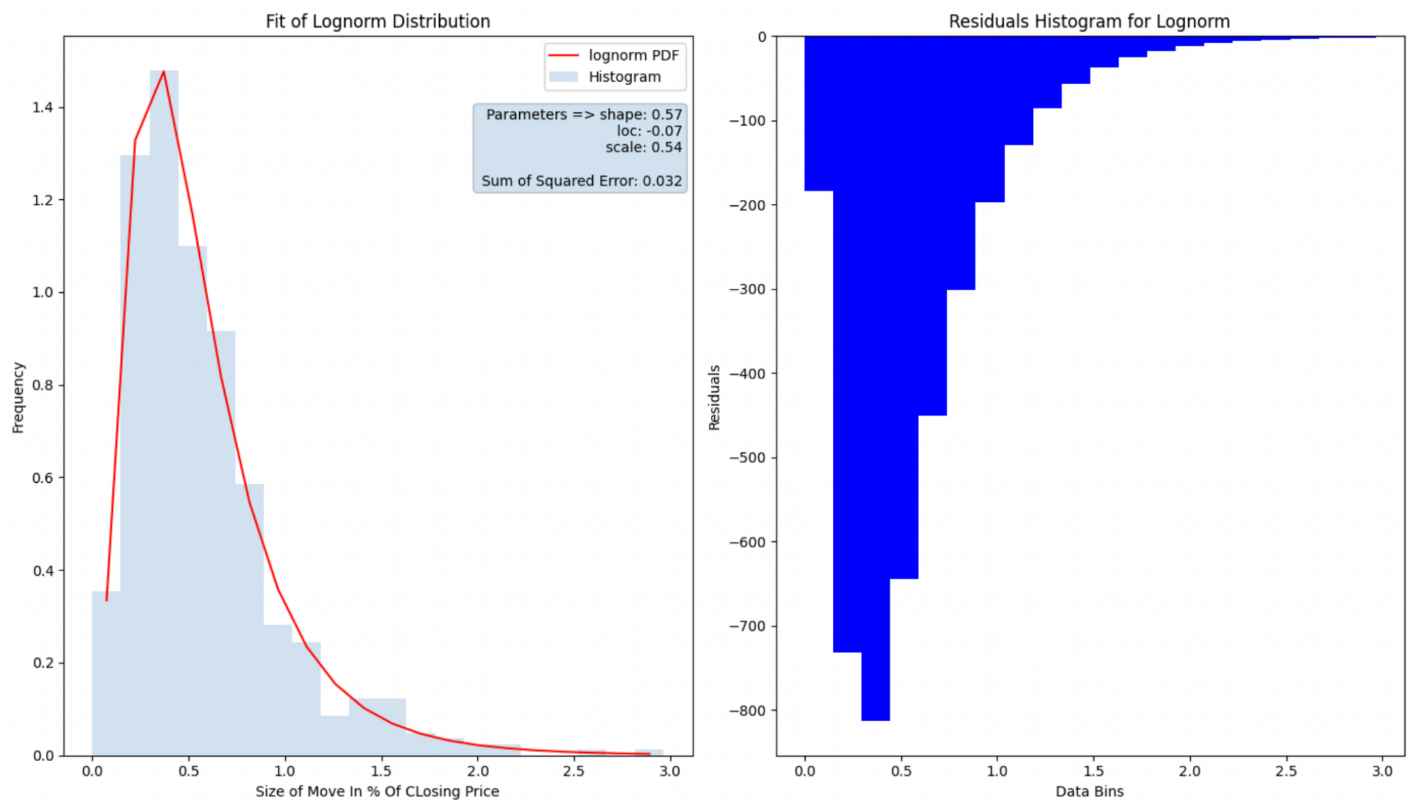


Figure 2: Shows best fit curve for regression data

With these probability density functions, I can now calculate the probability of an entry limit order getting executed, the probability of getting stopped out for a given stop price and the probability of the take profit order getting executed. In the Trading Strategy section below I will show how to create the expectation value for a possible strategy and compare it against a backtest of the data.

Discussion

The mean and median values of the data are quite different, which can be seen from the skew in the histograms above. The chart from Figure 3 confirms this, so I believe the median values are more appropriate to use.

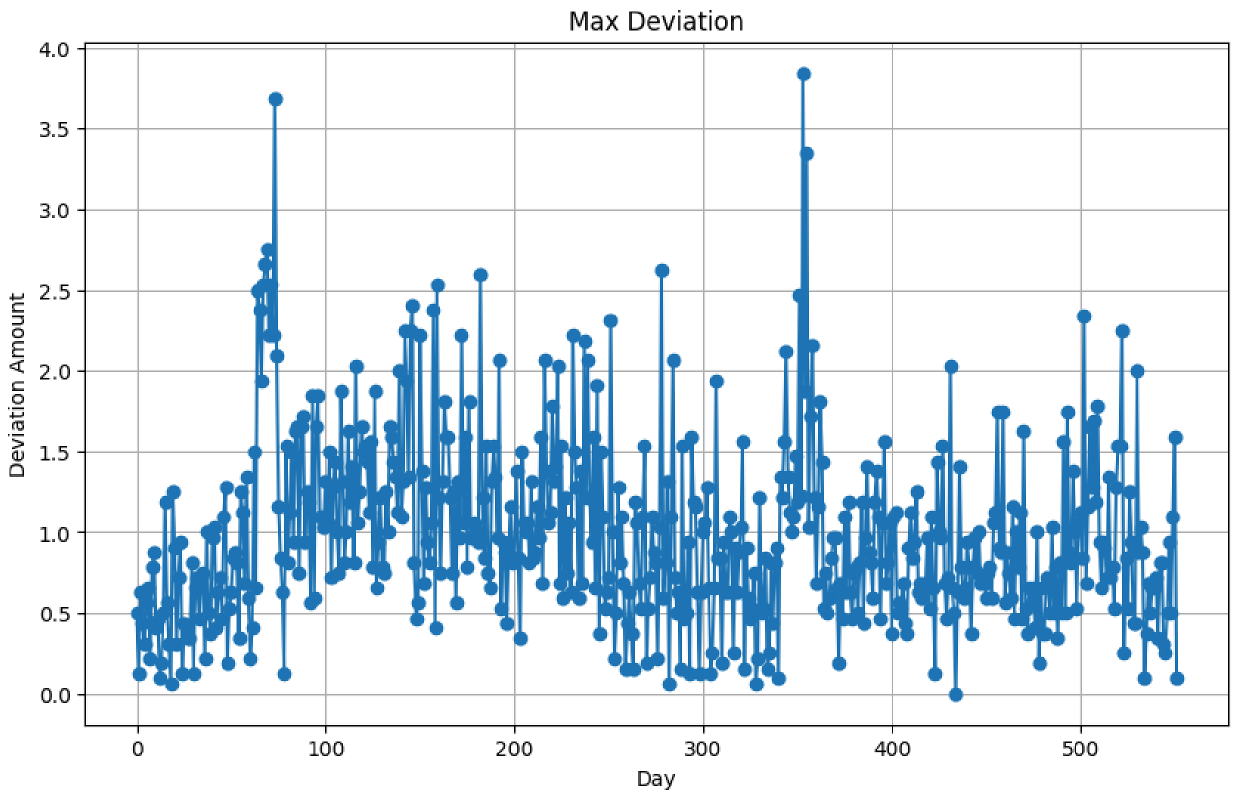


Figure 3: Shows the max-deviation for each after hours trading session.

The median *max-deviation* is 0.56% of the closing price, or about 28 ticks. The median regression is 0.47% of the closing price, or about 22 ticks. At first glance this seems promising as price action is retracing its initial deviation by a substantial amount. In fact, the median regression is 91% of the total deviation. This is certainly a tradable phenomena. However, it is more complicated because of the variance of the data.

The standard deviation is quite large with respect to the median, indicating a wide range of values (assuming a normal distribution). Had the variance been small and the *max-deviations* fell within a tighter range, we could simply have entered a trade at some value close to the median, perhaps the median +/-0.5 standard deviations, and exited when we came close to the median regression value. Unfortunately the variance is large, so despite identifying a tradable pattern, we still cannot predict with good probability what value the *max-deviation* will be using these simple metrics alone. In the trading strategy section that follows I will show how to use the distributions we fit above to find a possible entry.

The proportion of times the price regressed back within 4 ticks of the closing price was 70%. This is also promising for a strategy that trades the regression back to the close, but I do think that a more meaningful metric is not simply how many sessions regressed back to the close, but how many sessions retraced more than a given percentage of the *max-deviation*.

For instance, let's say that the previous close was 110 and that price hit 120 during after hours trading, then regressed back down to 115 for the open. The *max-deviation* would be 10 and the *regression* would be 5, which is a 50% retracement of the deviation from the close. A more meaningful question would be: How many of these sessions had at least a 50% retracement from the *max-deviation*? How many had a 60% retracement? After all, we don't need a full 100% retracement to make a profitable trade, and knowing the probabilities associated with a given retracement can help us properly structure our risk management strategy.

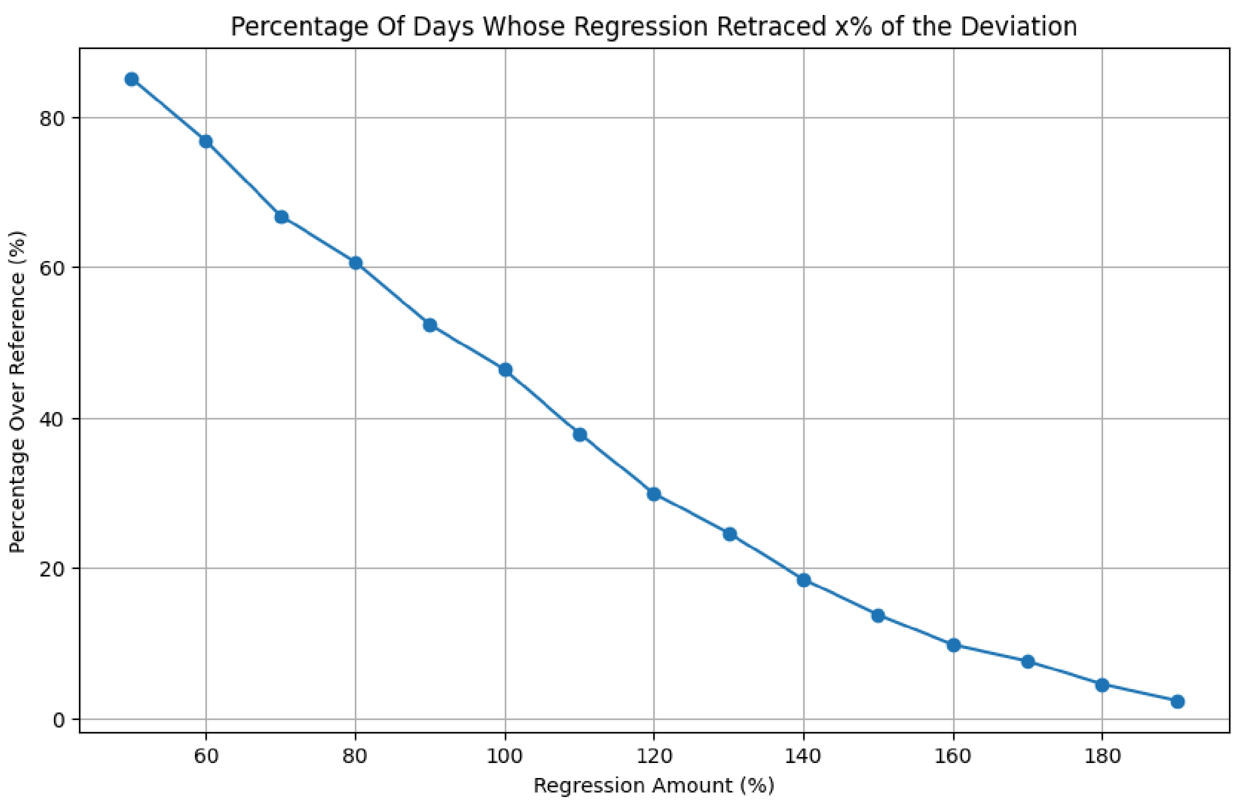
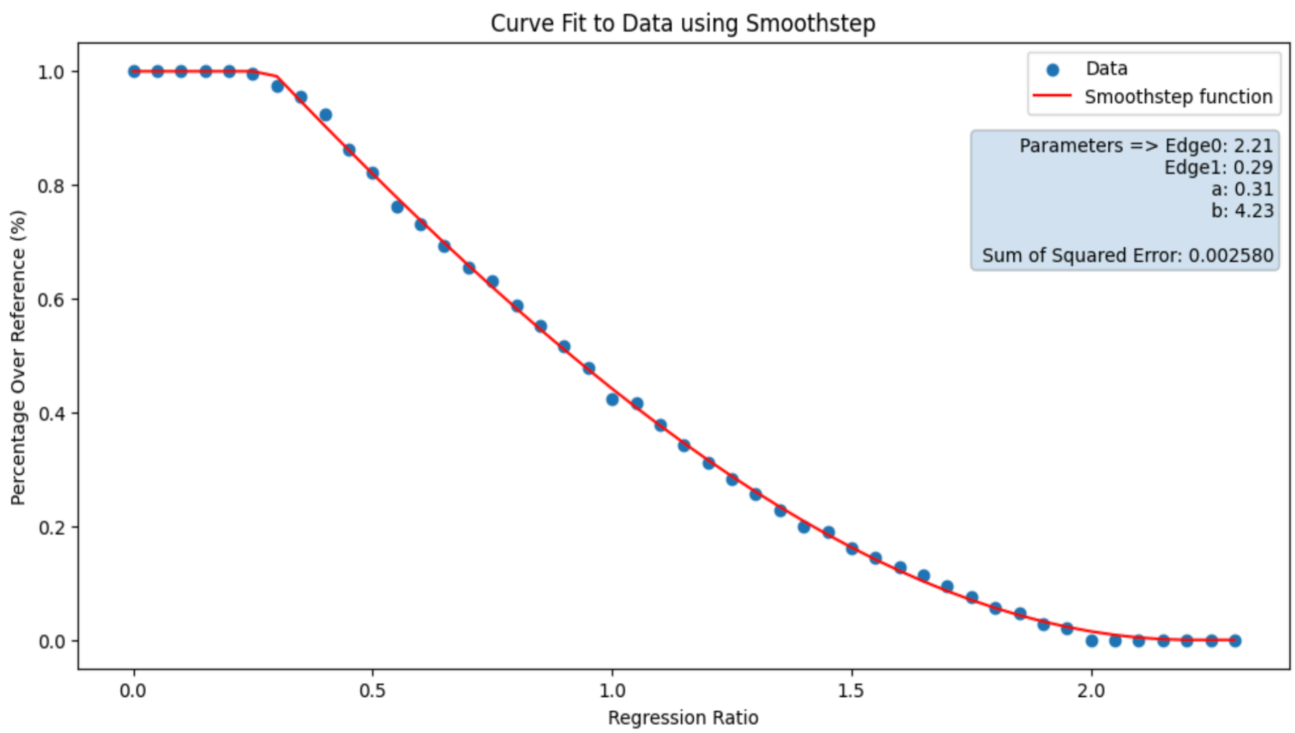


Figure 4: Shows the frequency of sessions that retraced at least a given percentage.

Figure 4 gives insight into how often these after hours sessions retrace by a given percentage. If we look at the first data point, it shows that the proportion of sessions that retraced at least 50% from the *max-deviation* was 85%. This is a very high number and should give confidence that if we find a *max-deviation* inflection point, there is likely to be at least a 50% retrace of that move. Similarly, if we look at the 4th data point, it shows that the proportion of sessions that retraced at least 80% from the *max-deviation* was 61%.

The distribution in Figure 2 essentially gives us the probability that a given exit price will be hit (assuming we entered perfectly at the *max-deviation* value), so should be useful for our exit strategy and refining our risk management. We can model this distribution as well, and the best fit curve was the Smoothstep function. Figure 5 shows the curve, which fits well to the data. We could use this distribution to help with calculating probabilities for our trade exits, but this is a piecewise function, which is hard to deal with,particularly its derivative, as it is not a continuous function. So the better approach will be to stick with the Log Normal distribution from above. You can check the notebook code to see the parameters and fit algorithms for the Smoothstep if interested.



Trading Strategy

This is a hands off strategy that sets the limit orders right at the daily close. There is no need for technical analysis in this instance or to try and track price throughout the night for a potential better entry. Everything is calculated ahead of time. If the trade enters but does not reach the take profit nor get stopped out, the trade is closed at the end of the after hours trading session.

This strategy requires two orders to be implemented at the same time. One if the price deviates downward (a long entry) and one if the price deviates upward (a short entry). Because this cannot be done using futures contracts on one account, either two accounts would need to be used or one could implement a trading bot that enters the correct trade when the entry price is reached.

Based on the day's closing price, we enter in a limit order with an entry, stop and exit price. These values will come from optimizing the expectation value function below. The values are relative to the closing price, so if the optimal entry value is found to be 1%, this means that the entry price should be 1% above that day's closing price for a short entry or 1% below that day's closing price for a long entry. Similarly for the other values. The optimized values are fixed, meaning that the same values are used for every trade. However, because they are percentage values relative to the closing price, the actual *prices* for entry, stop and exit will change each day as the closing price changes each day.

Calculate the Expectation Value

There are 3 expectation values (EV's) to calculate: The first is the for when the trade is stopped out, the second is for when the trade reaches the take profit and the third is when the trade enters, but does not reach the stop or take profit and only exits when the session is over, potentially in some profit or at some loss. Because the model does not consider which direction price has deviated, only the magnitude of the deviation, I will model each trade as a short, assuming that the price traded higher during the after hours session.

For the stopped out scenario, we simply need the probability that the *max_deviation* is larger than the stop value. We can get this value by using $sf(x) = 1 - CDF(x)$ where $CDF(x) = \int PDF(x)$. Then we multiply by the loss we would experience when hitting the stop.

$$EV_{stopped} = sf(stop)(entry - stop)$$

For the take profit scenario, we need a weighted average of the probabilities for each deviation between our entry and stop price, along with a weighted average for each regression that can reach our take profit level, multiplied by the profit at that exit price. We need a similar setup for the scenario where we only exit due to time constraints, but in this case we sum over all regressions that do not trigger our take profit. We can combine these together to get the final EV equation,

$$: \int_{\mathrm{entry}}^{\mathrm{stop}} \left(\int_0^{y-\mathrm{exit}} \left(\mathrm{PDF}_{\mathrm{Fréchet}}(y) \mathrm{PDF}_{\mathrm{lognorm}}(x)(\mathrm{entry} - y + x) dx dy \right) + \int_{y-\mathrm{exit}}^{\infty} \left(\mathrm{PDF}_{\mathrm{Fréchet}}(y) \mathrm{PDF}_{\mathrm{lognorm}}(x)(\mathrm{entry} - \mathrm{exit}) dx dy \right) \right)$$

$$+ \mathbf{sf}_{\mathbf{Fréchet}}(\mathbf{stop})(\mathbf{entry} - \mathbf{stop})$$

Referring to the definitions for these PDFs above, this EV will be too difficult to evaluate analytically, even with a symbolic package like SymPy. Instead, I will evaluate it numerically and use an optimization function, in this case the Nelder-Mead algorithm, to find the values of entry, stop and exit for which the EV is maximized. Once again refer to my python code to see how I implemented the optimization.

The optimized values are

$$entry = 0.48,\ stop = 4.77\ and\ exit = 4.5 \times 10^{-9}.$$

The values represent a percent of the closing price, so for example the entry price should be $1.0048 \times price_{close}$. These values indicate entering near the median *max_deviation* value and setting wide stops and exit points.

The maximized EV value is 0.09, indicating we should average a small profit on each trade. The model predicts around 62% of trades should be entered, and if we use 1 contract for every trade and a median closing price of 130 over a two year period for 342 trades, the gain is

$$gain = 130 \cdot \frac{EV}{100} \cdot 32_{ticks} \cdot \$31.25_{per\ tick} \cdot 342 = \$40,014.$$

So the model is predicting a pretty nice yearly gain when just using one contract per trade.

Next I am going to feed these values into a backtest function and optimize them even further. The reason for creating the model first is that the backtest function is somewhat complex for an optimizer, meaning that if we do not start close to an ideal value, the optimization may not converge. And just like in technical analysis, I want to see confluence from two different approaches.

Backtest

When I feed these initial optimized values into the backtest using actual data, I get slightly better results with the total gain being

$$gain = \$41,589,$$

but still very close to model predictions. If I then run the backtest function through the same optimizer, starting with the initial optimized values, I get a new set of optimized values

$$entry = 0.71,\ stop = 2.23\ and\ exit = 0.095.$$

These values increase the entry price a little bit above the median, being more discriminating for entering a trade while also moving the stop and exit points a little closer. The number of trades entered drops from 342 to 211. Figure 6 below shows the performance of the final optimized parameters over the total two year dataset.

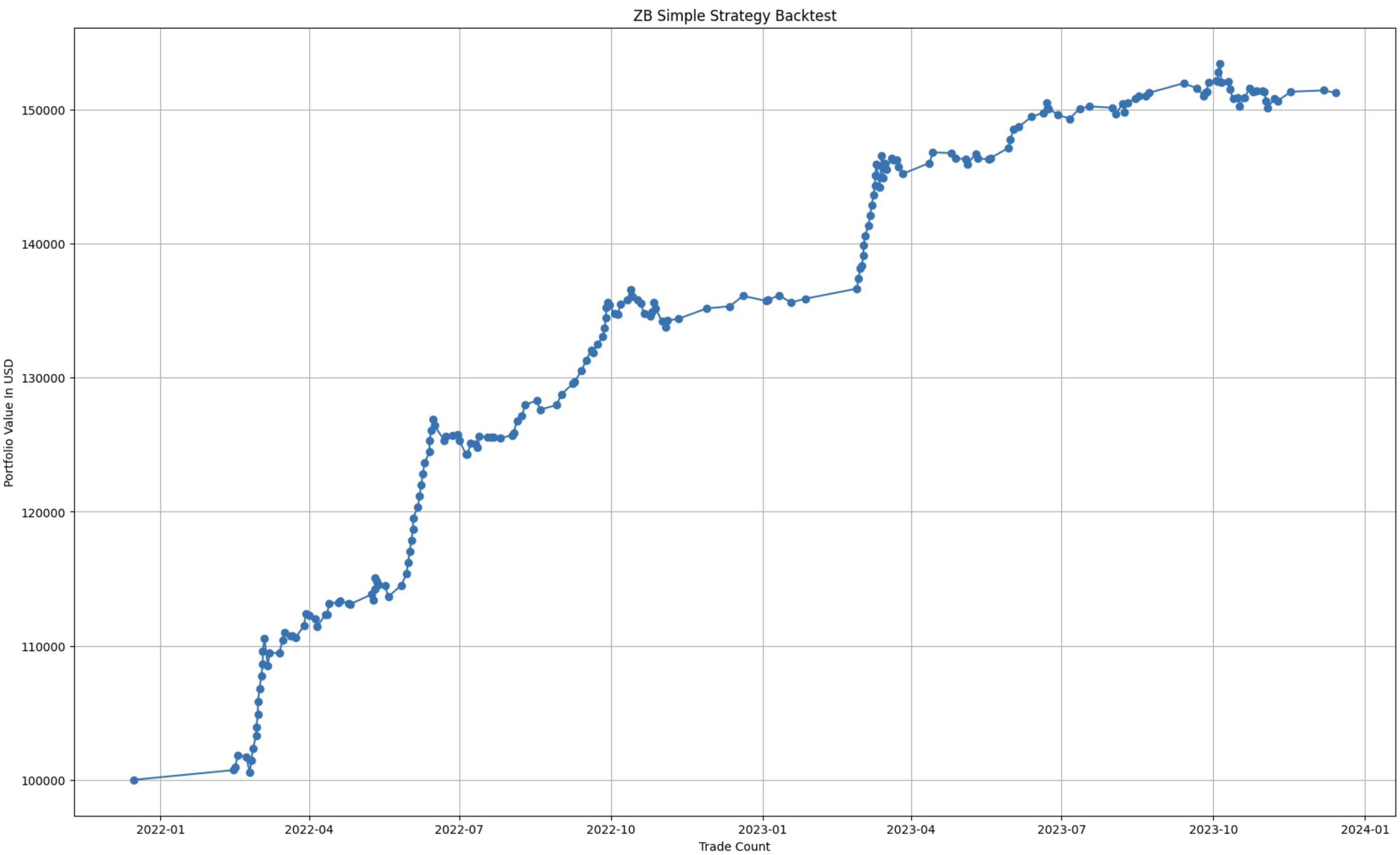


Figure 6: Performance of backtest strategy of a 2 year period.

Using the final optimized values, the backtest showed a total gain of

$$gain = \$51,252.$$

In spite of trading much less, these new values resulted in an increase of about $10,000 for the whole period over the initial values. The strategy then predicts an average performance of

$25,626 per year, per contract.

Assuming sufficient liquidity for 10 contracts, this would be potentially $250,000 per year. Given the simplicity of the strategy, this is a great result. The max drawdown is about $3,000, so with a margin requirement of $6,200 for a single ZB contract, a conservative minimum portfolio starting balance would be around $15,000. To make the strategy profit faster, one could increase the number of contracts traded as the portfolio starts to make gains from the previous trades.

Figure 6 also shows that the strategy goes through “boom” and “consolidation” periods, indicating that the strategy performs better under certain conditions than others. What these conditions are will be left for later work. That being said, the strategy had no long term downtrends, indicating that it is somewhat of an all weather strategy.

Risk Considerations

As was mentioned in the previous section, this strategy saw a max drawdown of about $3000. This is relatively small but should be taken into consideration when deciding on an initial balance. Also, this strategy cannot guarantee that the future price action will adhere to model predictions. We are entering a somewhat new macro landscape and it is possible that the regression pattern we have seen may change enough such that the strategy is no longer profitable long term. Although I think that would be unlikely, anyone using this strategy should keep a close eye on performance and start with minimal investment to make sure that the strategy performs as expected.

Future Work

The data I used only went back 2 years, the maximum Yahoo would give me in one request. We have entered a new macro landscape these past couple years, so it’s very possible this regression pattern was different in the past. It would be better to get data going back to at least 1980. Rather than just average the data over the whole set, it would be good to break it into epochs and see how the price action has changed over the years and correlate this behavior to different macro environments.

Figure 6 shows that the portfolio gains happened in certain “boom” and “consolidation” periods. I need to go through these time periods and find out what may be the driving factors between these different periods so I can refine the strategy even further.

The trading strategy used was very simple. I didn’t track the price throughout the after hours session or use any complex technical analysis tools. This is powerful as it is easy to implement, but it may be leaving out a lot of profit. It would be ideal to explore how I may refine the entry by using more sophisticated tools, or perhaps even look into trading options instead of the futures contract itself.

Conclusion

ZB does exhibit a tradable phenomena during after hours, whereby after deviating a certain amount after close, price is likely to regress back to the closing price. I was able to fit two probability distributions to the data which allowed me to calculate optimized entry, stop and exit values for a simple trading strategy. Backtesting this strategy over a two year period using the optimized values and 1 contract per trade resulted in an average gain of

$25,626 per year.

Whether this is a sufficient return will be up to each individual trader. In the case of EAG Fund, for which this analysis was primarily meant for, this may not be an aggressive enough of a return. However, I do think this return is good and will more than likely experiment with this strategy personally.