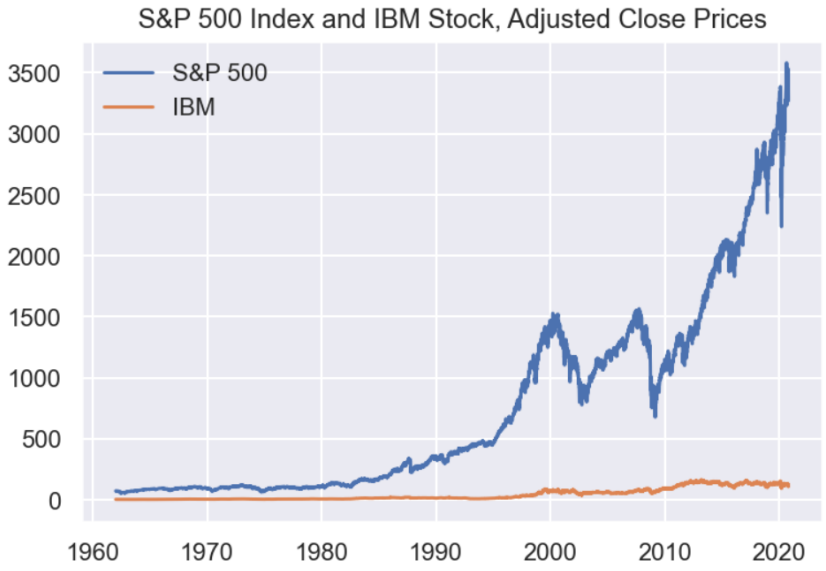
Imperfect Models for Stock Trading

Can imperfect models be profitable? Can a humble stock beat the market? When's the best time to buy? These basic questions are at play as we seek to find answers. The data set comes from Yahoo finance for IBM stock and the S&P 500 index.

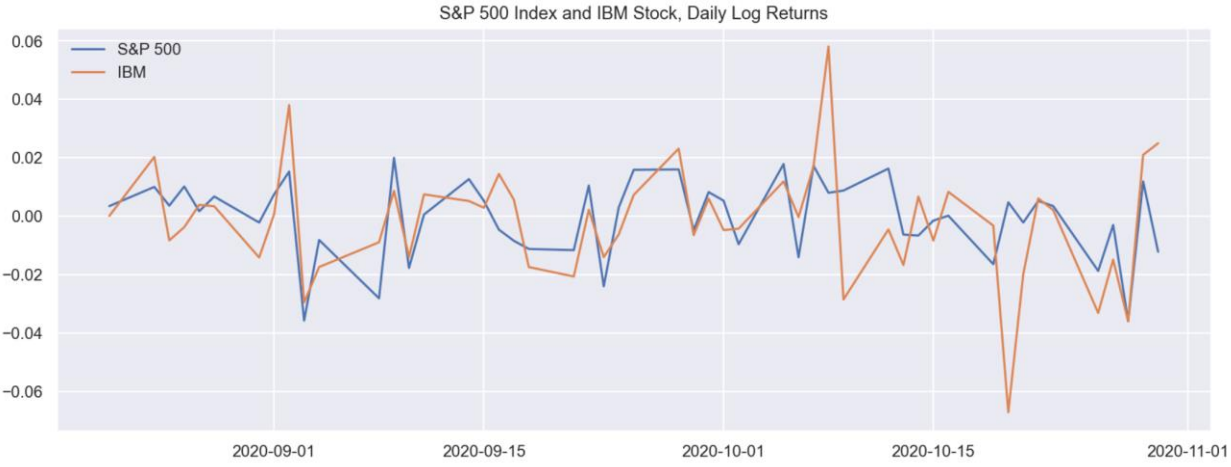
Humble Stock vs. Amazing Index

Actually, IBM is quite awesome to most of us. However, in the shadow of the S&P 500 price history it seems humble. The index dwarfs the stock. We refer to IBM as humble here, relatively speaking. Big stocks like IBM are liquid meaning we can get in and out of the market quickly as needed.



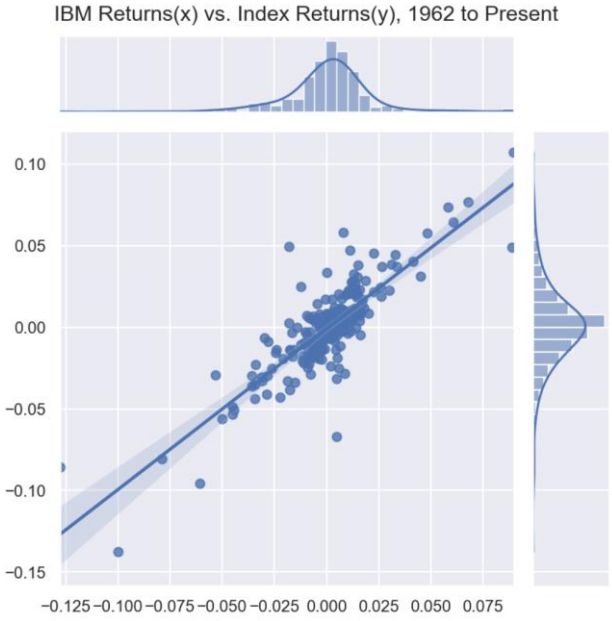
Is there hope?

When we look at a snapshot of log returns, there seems some moments in time when the stock is up happy and the market is down sad. If we could time it right maybe there's some chance to beat the market. It seems within the realm of feasibility. There is hope for this stretch goal.



The Tale of Two Returns

For closer inspection we use a scatter plot of log returns, it seems IBM stock and S&P index are positively correlated with a unit slope regression line. Here the points below the line suggest opportunities of beating the market, relatively speaking. If the market is negative but the stock is less negative, that's helpful in beating it.



Historical Price Features

Maybe price patterns repeat with time in a way that shares relation to next week's price behavior, or not. Here for this model around 250 look-back features going back in time 45 days tie into look-forward historical pricing behavior going forward in time 7 days. Maybe sometimes it works and sometimes it doesn't and that's okay. When the model predicts a target related to high prices next week, this is our cue to buy in chance a past-future relationship holds.

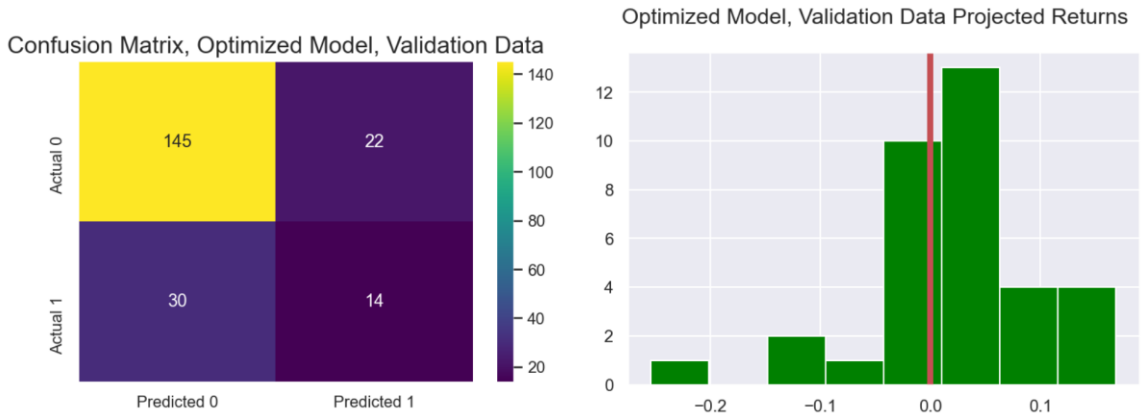
When's the best time to buy?

We tune our model on stock data (and compare performance to index data). But what are the model targets we're aiming for? The best time to buy is when the model predicts a "1". This occurs when feature data indicates high probability that the next week return is in the upper septile or in the high return zone. The septiles divide the next week price return distribution in seven equal portions. We pick the top septile as highest return level, as target 1's zone. All other zones are 0's, non-buy zones.

But the distribution of targets is imbalanced, most are "0's" or non-buy days. We use SMOTE to balance the train set with similar numbers of 1's and 0's. This is fortified with the intent that miss-classified opportunities error on the positive return side and this promotes a better model overall.

Can Imperfect models by profitable?

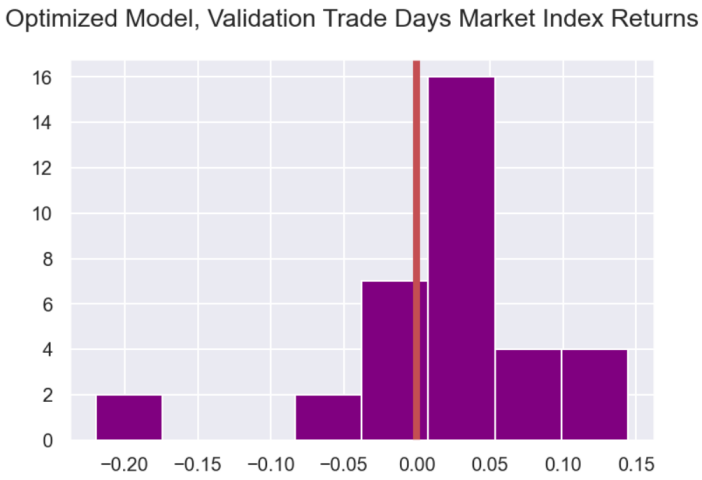
Yes, they can. Here we have a confusion matrix indicating many missed predictions. The model is grossly imperfect, its mean accuracy is around 80%. Yet the corresponding trades are overwhelmingly positive! This demonstrates it's okay to be confused and profitable at the same time. In fact, this explains the stock market in a nutshell for some of us. If most of the trades are positive, they over compensate for the minority that are negative. Note, this system isn't intended to be 100% accurate for any one given trade opportunity. It works to bias most trades as potentially positive. For example, if a losing trade comes around, don't quit – it's seems unlikely to get two losing trades in a row.



Can a humble stock beat the market?

It seems improbable looking at the historical price charts. Using the same trade days that we found for IBM using our model, we also look at what the S&P Index did – guess, what? It was also largely profitable, just not as much. The market earned 78% return those days while IBM did 120%. Okay if we modeled on the index data then maybe the opposite happens, but we wanted to favor the stock.

We're looking for positive expected returns, maybe we got lucky with 120.5% in 2020. We were looking to beat the market in similar trading days, we would have beat the market by 42.5%. We found a humble stock can prevail.



The volatility seems way too high however this is expected for a handful of trades. We like Sharpe ratios about one, we get 1.15, but this may be misleading because the distribution's non-normal. If that seems confusing, more data may be a remedy.

Summary

Yes, Yes, and Yes. An imperfect model can be profitable, a humble stock can beat the market, and using models - we know when to buy. The same is likely true for other stocks. It's a powerful idea, as using a set of humble stocks to beat the market every day. Imperfect models here mean ones that are carefully tuned with parameters that afford error margins that compensate for their misbehavior.

Such models are built on trading rules. Here, we buy the day after the model predicts a "1", and sell exactly a week later without any second guessing it. Why might that be? Because the target returns the model is based upon are derived from such a simple trading rule. A more sophisticated set of rules might train for the sell points as well, and use a fleet of models for a portfolio of humble stocks. The models don't work without using some trading discipline that complements the models.

Try It!

Learning these ideas might inspire some back testing, there's ~50 years of IBM data available, not just the 7-year sample used here. There's likely 90 years of index data making for an ocean of confused models; to be human is to be confused.

The world is full of money and savvy trading may alleviate our monetary disparities. There's a chance people find great joy in weighing trades more positive than negative. As human beings, we share our monetary system and have permission to participate in this ongoing saga.

Disclaimer: These ideas are presented as an educational exercise. Please trade at your own risk. Please don't consider this a basis for real trading where rigorous testing factors in. Happy data to all!

11-01-2020 Patrick Parker

https://github.com/patronical/humble_stock_model.git