
NE
E
SE
S
SW
W

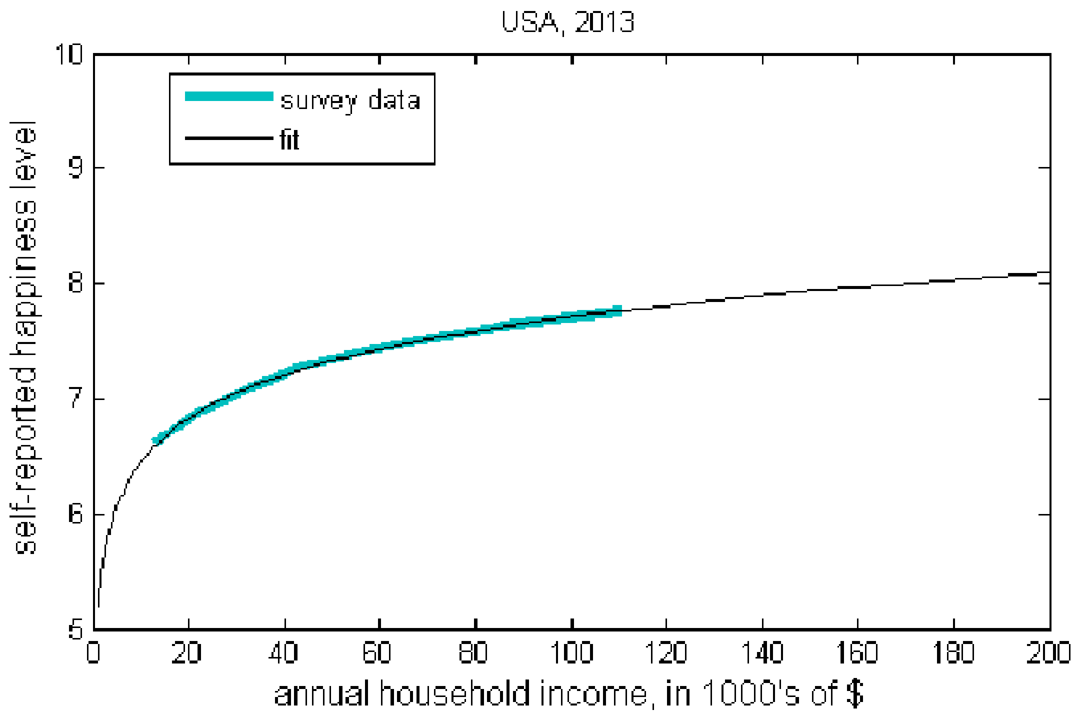
Quantitative Finance

Why quantitative finance is so hard

This is not financial advice.

Quantitative finance (QF) is the art of using mathematics to extract money from a securities market. A **security** is a fungible financial asset. Securities include stocks, bonds, futures, currencies, cryptocurrencies and so on. People often use the techniques of QF to extract money from prediction markets too, particularly sports betting pools.

Expected return is future outcomes weighted by probability. A trade has **edge** if its expected return is positive. You should never make a trade with negative expected return. It is not enough just to use expected return. Most peoples' value functions curve downward. The marginal value of money decreases the more you have. Most people have approximately logarithmic value functions.



A logarithmic curve is approximately linear when you zoom in. Losing 1% of your net worth hurts you slightly more than earning 1% of your net worth helps you. But the difference is usually small enough to ignore. The difference between earning 99% of your net worth and losing 99% of your net worth is not ignorable.

When you gain or lose 1% of your net worth, the expected change to the logarithm of your wealth is a tiny -0.01%. When you gain or lose 99% of your net worth the expected change to the logarithm of your wealth is -400%.

$$\log(1.01) + \log(0.99) \approx -0.0001$$

$$\log(1.99) + \log(0.01) \approx -4$$

This is called a **risk premium**. For every positive edge you can use the Kelly criterion to calculate a bet small enough such that the you edge exceeds your risk premium. In practice traders tend to use fractional Kelly.

Minimum transaction costs are often constant. It is not sufficient for your edge to merely exceed your risk premium. It must exceed your risk premium plus the transaction cost. Risk premium is defined as a fraction of your net worth but transaction costs are often constant. If you have lots of money then you can place larger bets while keeping your risk premium constant. This is one of the reasons hedge funds like having large war chests. Larger funds can harvest risk-adjusted returns from smaller edges.

Getting an Edge

The only free lunch in finance is **diversification**. If you invest in two uncorrelated assets with equal edge then your risk goes down. This is the principle behind index funds. If you know you're going to pick stocks with the skill of a monkey then you might as well maximize diversification by picking all the stocks. As world markets become more interconnected they become more correlated too. The more people invest in index funds, the less risk-adjusted return diversification buys you. Nevertheless, standard investment advice for most[1] people is to invest in bonds and index funds. FEMA recommends you add food and water.

All of the above is baseline. Baseline rents you can extract by mindlessly owning the means of production is called **beta** β. Earning money in excess of beta by beating the market is called **alpha** α.

> There are three ways to make a living in this business: be first, be smarter or cheat.
>
> —John Tuld in *Margin Call*

You can be first by being fast or using alternative data. Spread Networks laid a $300 million fiber optic cable in close to a straight line from New York City to Chicago. Being fast is expensive. If you use your own satellites to predict crop prices then you can beat the market. Alternative data is expensive too.

If you want to cheat go listen to *Darknet Diaries*. Prison is expensive.

Being smart is cheap.

Science will not save you

Science [ideal] applies Occam's Razor to distinguish good theories from bad. Science [experimental] is the process of shooting a firehose of facts at hypotheses until only the most robust survive. Science [human institution] works when you have lots of new data coming in. If the data dries up then science [human institution] stops working. Lee Smolin asserts this has happened to theoretical physics.

If you have two competing hypotheses with equal prior probability then you need one bit of entropy to determine which one is true. If you have four competing hypotheses with equal prior probability then you need two bits of entropy to determine which one is true. I call your prior probability weighted set of competing hypotheses a **hypothesis space**. To determine which hypothesis in the hypothesis space is true you need training data. The entropy of your training data must exceed the entropy of your hypothesis space.

The entropy of n competing hypotheses with equal prior probability is $\log n$. Suppose your training dataset has entropy T. The number of competing hypotheses you can handle grows exponentially as a function of T.

$$\log n = T$$
$$n = e^T$$

The above equation only works if all the variables in each hypothesis are hard-coded. A hypothesis $y = 2.2x + 3.1$ counts as a separate hypothesis from $y = 2.1x + 3.1$.

A hypothesis can instead use tunable parameters. Tunable parameters eat up the entropy of our training data fast. You can measure the entropy of a hypothesis by counting how many tunable parameters it has. A one-dimensional linear model $y = ax + b$ has two tunable parameters. A one-dimensional quadratic $y = ax^2 + bx + c$ model has three tunable parameters. A one-dimensional cubic model $y = ax^3 + bx^2 + cx + d$ has four tunable parameters. Suppose each tunable parameter has e bits of entropy. The total entropy needed to collapse a hypothesis space with m tunable parameters equals m. The entropy of a hypothesis space with m tunable parameters equals m.

We can combine these equations. Suppose your hypothesis space has n separate hypotheses each with m tunable parameters. The total entropy J equals the entropy necessary to distinguish hypotheses from each other plus the entropy necessary to tune a hypothesis's parameters.

$$J = m + \log n$$

Logarithmic functions grow slower than linear functions. The number of hypotheses n is inside the logarithm. The number of tunable parameters m is outside of it. The entropy of our hypothesis space is dominated by m. The number of competing hypotheses we can distinguish grows exponentially slower than the entropy of our training data. You can distinguish competing hypotheses from each other by throwing training data at a problem if they have few tunable parameters. If you have tunable parameters then the entropy required to collapse your hypothesis space goes up fast.

If you have lots of entropy in your training data then you can train a high-parameter model. Silicon Valley gets away with using high-parameter models to run its self-driving cars and image classifiers because it is easy to create new data. There is so much data available that Silicon Valley data scientists focus their attention on compute efficiency.

Wall Street is the opposite. Quants are bottlenecked by training data entropy.

Past performance is not indicative of future results

If you are testing a drug, training a self driving car or classifying images then past performance tends to be indicative of future results. If you are examining financial data then past performance is not indicative of future results. Consider a financial bubble. The price of tulips goes up. It goes up some more. It keeps going up. Past performance indicates the price ought to keep going up. Yet buying into a bubble has negative expected return.

Wikipedia lists 25 economic crises in the 20th century plus 20 in the 21st century to date for a total of 45. Financial crises matter. Hedge funds tend to be highly leveraged. A single crisis can wipe out a firm. If a strategy cannot ride out financial crises then it is unviable. Learning from your mistakes does not work if you do not survive your mistakes.

When Tesla needs more training data to train its self-training cars they can drive more cars around. If a hedge fund needs 45 more financial crises to train its model then they have to wait a century. World conditions change. Competing actors respond to the historical data. New variables appear faster than new training data. You cannot predict financial crises just by waiting for more training data because the entropy of your hypothesis space outraces the entropy of your training data.

You cannot predict a once-in-history event by applying a high-parameter model to historical data alone.

---

1. If your government subsidizes mortgages or another kind of investment then you may be able to beat the market. ↩

Hypothesis Space Entropy

In Why quantitative finance is so hard I explained why the entropy of your dataset must exceed the entropy of your hypothesis space. I used a simple hypothesis space with n equally likely hypotheses each with m tunable parameters. Real life is not usually so homogeneous.

No Tunable Parameters

Consider an inhomogeneous hypothesis space with zero tunable parameters. Instead of $H = \log n$ which works for homogeneous hypothesis spaces, we must use more complicated entropy equation.

$$H = -\sum_{i=1}^{n} \rho_i \ln \rho_i$$

This equation makes intuitive sense. It vanishes when one $\rho_{i=j}$ equals 1 and all other $\rho_{i \neq j}$ equal 0. Our equation is extremized when all ρ_i are equal at $\frac{1}{n}$. $H = \log n$ is the maximal case when $\rho_i = \frac{1}{n} \forall i \in \{1, \ldots, n\}$[1].

With Tunable Parameters

Suppose each hypothesis i has m_i tunable parameters. We can plug m_i into our entropy equation.

$$H = \sum_{i=1}^{n} \rho_i (m_i - \ln \rho_i)$$

Our old equation $H = m + \log n$ is just the special case where all ρ_i are homogenous and m_i are homogeneous too.

We have so far treated m_i as representative of each hypothesis's tunable parameters. More generally, m_i represents each hypothesis's internal entropy. If we think of hypotheses as a weighted tree, m_i is what you get when you iterate one level down the tree. Our variable H identifies the root of the tree. Suppose ith branch of the next level down is called H_i.

$$H = \sum_{i=1}^{n} \rho_i (H_i - \ln \rho_i)$$

We can define the entropy of the rest of the tree with a recursive equation.

$$H_\mu = \sum_{i=1}^{n} \rho_i (H_{\mu,i} - \ln \rho_i)$$

$$= \sum_{i=1}^{n} (\rho_i H_{\mu,i} - \rho_i \ln \rho_i)$$

There are two parts to this equation: the **recursive component** $\rho_i H_{\mu,i}$ and the **branching component** $-\rho_i \ln \rho_i$.

Branching component $-\rho_i \ln \rho_i$

The $-\rho_i \ln \rho_i$ component is maximized when $\rho_i = \frac{1}{e}$.

$$-\rho_i \ln \rho_i = -\frac{1}{e} \ln \frac{1}{e}$$

$$= \frac{1}{e} \ln e$$

$$= \frac{1}{e}$$

The branching component tops out at $\frac{1}{e}$. It can never contribute a massive quantity of entropy to your hypothesis space because it is limited to $\frac{1}{e}$ entropy per level of the tree.

$$0 \leq -\rho_i \ln \rho_i \leq \frac{1}{e}$$

The branching factor is mostly unimportant. The bulk of our entropy comes from the recursive component.

Recursive component $\rho_i H_{\mu,i}$

Fix ρ_i at a positive value. There is no limit to how big $H_{\mu,i}$ can become. You can make it arbitrarily large just by adding parameters. Consequently $\rho_i H_{\mu,i}$ can become arbitrarily large too. In real world situations we should expect the recursive components of our hypothesis space to dominate the branching components.

If ρ_i vanishes then the recursive component disappears. This might explain why human minds like to round "extremely unlikely" $\epsilon > \rho_i > 0$ to "impossible" $\rho_i = 0$ when $H_{\mu,i}$ is large. It removes lots of entropy from our hypothesis space still being right almost all of the time. This may be related to synaptic pruning.

Lessons for Hypothesis Space Design

Once again, we have confirmed that having hypotheses with lots of parameters is a worse problem that having lots of hypotheses to choose between. More generally, one or more hypotheses with exceptionally high entropy dominate the total entropy of your hypothesis space. If you want better priors then the first step of your optimization should be to eliminate these complex subtrees from your hypothesis space.

---

1. Proof:

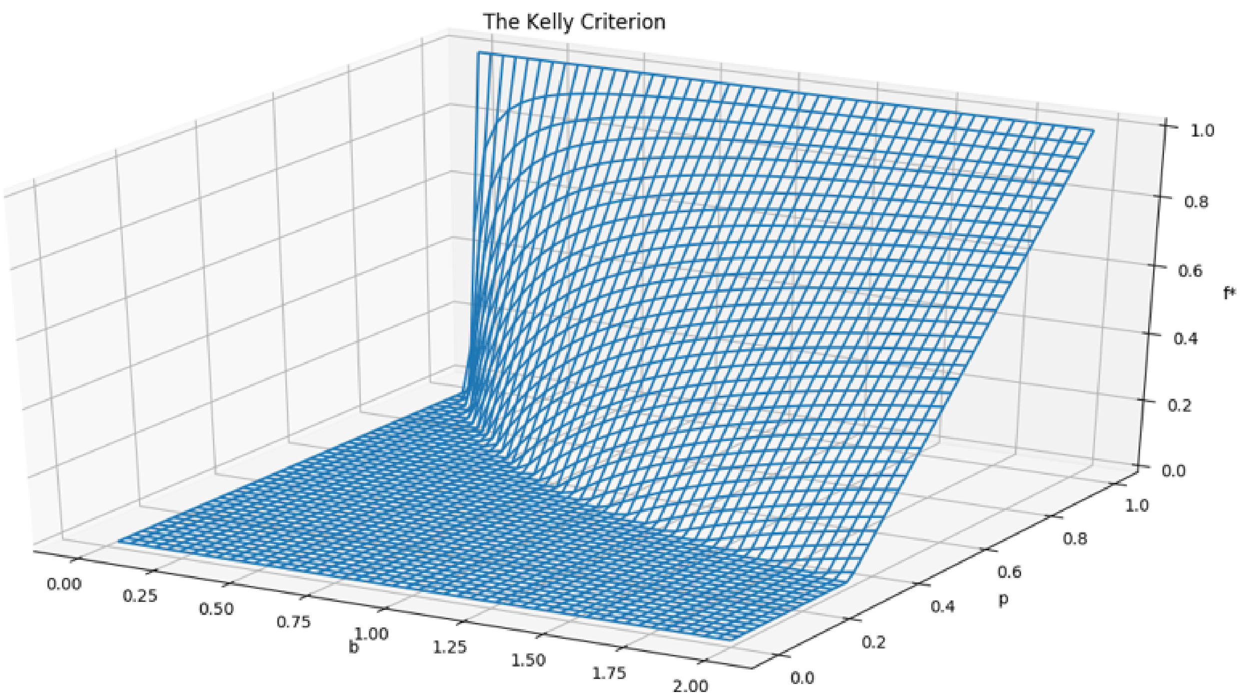
$$\begin{aligned} H &= -\sum_{i=1}^{n} \rho_i \ln \rho_i \\ &= -\sum_{i=1}^{n} \frac{1}{n} \ln \frac{1}{n} \\ &= -\ln \frac{1}{n} \\ &= -\ln(n^{-1}) \\ &= \ln n \end{aligned}$$

↩

The Kelly Criterion in 3D

The Kelly Criterion is a gambling strategy which maximizes the logarithm of your expected wealth. The Kelly Criterion tells you what fraction f^* of your bankroll to wager. It is a function of the net fractional odds[1] received $b > 0$ and the probability of a win $p \in (0, 1)$.

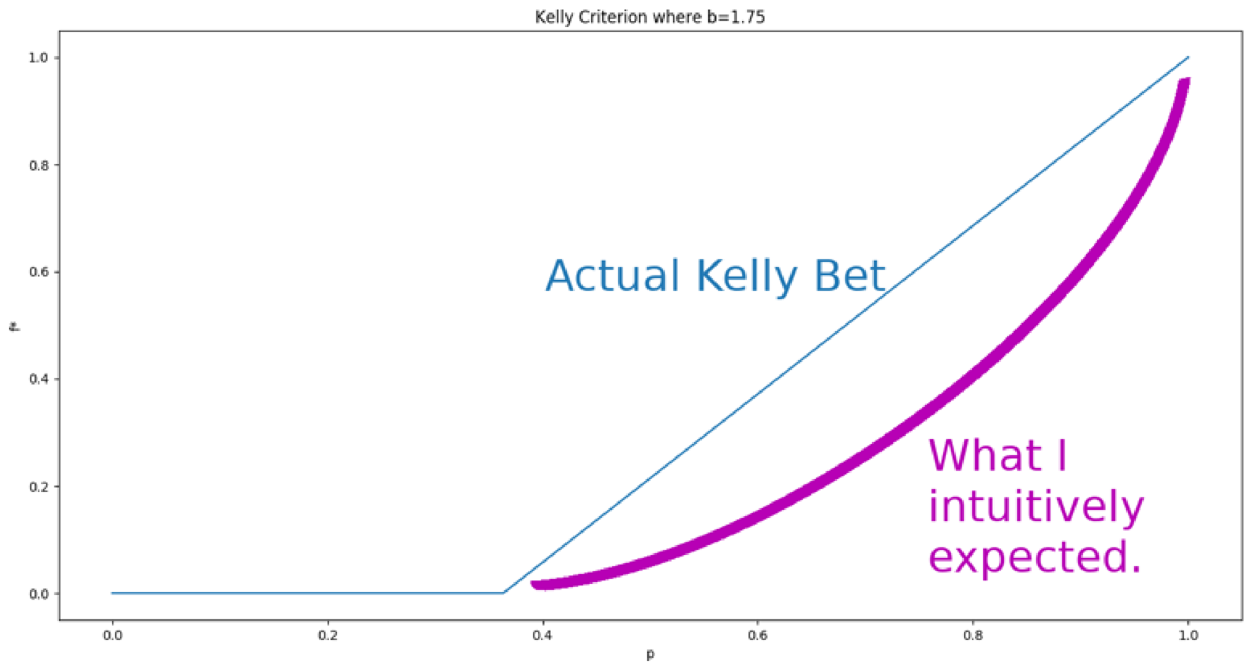
$$f^* = \frac{p(b+1) - 1}{b}$$



Some properties are intuitively easy to understand.

- The Kelly wager is positive iff the expected value $bp - (1 - p)$ is positive. The Kelly wager is zero otherwise.
- The Kelly wager is 1 for all $p = 1$. (Ignore $b = 0$.)
- The Kelly wager is 0 for all $b = 0$. (Ignore $p = 1$.)

What surprised me is that if you fix b and restrain p to the region of positive f^* then f^* is a linear function of p. This was not intuitive to me.



I expected asymptotic behavior with the greatest $\frac{df^*}{dp}$ in a neighborhood of p = 1. In other words, I expected the fractional wager to increase slowly at first and then increase faster as p approached 1. Actually, p is linear.

Kelly wagers tend to be more aggressive then human intuitions. I knew this and I *still* underestimated the Kelly wager. I didn't mess this up in a high stakes situation where fear throws off my calculations. I didn't even mess this up in a real-world situation where uncertainty complicates things. I underestimated the Kelly wager on a purely conceptual level.

I have written before about the utility of [my fear heuristic](). My fear heuristic might be helping to compensate for my Kelly miscalibration.

Recalibrating

I'm good at tolerating risk when it comes the small number of gigantic risky bets guiding my professional career. I'm also good at tolerating risk in the domain of painlessly small bets. (Not that there is much risk to tolerate in this latter case.) Judging by this post's analysis, I am *awful* at calibrating my risk tolerance for wagers between 0.05% and 1% of my net worth. Specifically, I am insufficiently risk tolerant.

What makes this worse is that the region of 0.05% to 1% of my net worth is full of long tails. The wagers I'm skipping could easily repay themselves a thousandfold. If I take a wager like this every day for 2 years and just a single one of them repays itself a thousandfold then I win bigtime.

I need to gamble more.

Optional Practice

I used these problems is to help develop my intuitive grasp of the Kelly criterion.

Q1: If p = 0.01 and b = 1000 then what is the corresponding f^*?

0.9%

The above number is way higher than what my intuition tells me is appropriate.

Q2: If p = 0.1 and b = 20 then what is the corresponding f^*?

5.5%

The above number is *higher* than the answer to Q1. This result was, again, unintuitive to me. I expected it to be smaller because b is smaller in absolute terms. But I didn't pay sufficient attention to bp = 2. The average return is 2× your initial investment.

Q3: If p = 0.51 and b = 1 then what is the corresponding f^*?

2%

Q4: If p = 0.51 and b = 2 then what is the corresponding f^*? (Not that b = 1 means you get back your original wager *plus* double you wager for a total of 3× your wager.)

26.5%

Q5: If p = 0.65 and b = 100 then what is the corresponding f^*? (In practice, opportunities like this are so rare you will usually not get to wager a full Kelly.)

65%

I was a little surprised; I had expected a higher result. The logarithmic value function is doing the work of keeping Kelly down.

Q6: If p = 0.05 and b = 100 then what is the corresponding f^*?

4%

---

1. The "net fractional odds" b indicate how much you win in the case of a win. If you wager x and lose then you lose x. If you wager x and win then you get your x back plus an additional xb. ↩

How to Price a Futures Contract

All investments are risky. The techniques explained in this series are exceptionally risky. I am not a registered investor, attorney, advisor, broker, banker, lawyer, dealer or anything of the sort. All opinions expressed here are for my personal exploration. I present it to you for entertainment. This post may help you understand policies related to financial regulation for the purposes of more informed voting. This post is not investment advice. This post may contain errors.

A **security** is a tradable financial asset. A **future** is the obligation to buy or sell a security at a predetermined strike price k.

Suppose a security has price S_0 right now and the risk-free bond interest rate is r.

What is the market equilibrium strike price k of a security at temporal displacement T into the future? You may assume the existence of an idealized free market.

$$k = S_0 e^{rT}$$

The price is forced to satisfy the above equation due to arbitrage.

What is the market equilibrium price S_0 right now of a future with strike price k?

$$S_0 = k e^{-rT}$$

You can simulate a future by short-selling the underlying security and buying a bond with the revenue. You can simulate short-selling the same future by borrowing money (selling a bond) and using the money to buy the underlying security.

If the strike price of a future is anything other than $k = S_0 e^{rT}$ then you can earn money from it.

How do you extract risk-free profit from an future with strike price $k^- < k$?

The strike price k is too low compared to S_0. We can earn money by longing the future. In the case of a future, "longing" means agreeing to buy the underlying asset at time T.

It is not enough to merely earn money on average. We care about *risk-adjusted* returns. The fact that the future price deviates from $S_0 = k e^{-rT}$ means we can extract money *risk free*. We can hedge our long of the future by shorting the underlying asset. Then we can use the bond market to bridge the temporal distance between shorting the underlying asset *now* and when the future contract expires.

Perform the following three actions simultaneously.

- Long the future. You agree to buy the underlying security at time T for price k^-. This contract costs you nothing upfront.
- Short-sell the underlying security for ke^{-rT} in cash.
- Lend the cash k as a bond with interest rate r.

When the future expires, perform the following three actions simultaneously.

- Collect the principal and interest from your bond for revenue ke^{rT}.
- Fulfill your end of the future contract by purchasing the underlying security at price k^-.
- Immediately hand off that underlying security to fulfill your end of the short-selling contract.

You pocket a profit of $(k - k^-)e^{rT}$.

How do you extract risk-free profit from an underpriced future with strike price $k^+ > k$?

Same as the previous question except you short the future, borrow the cash and buy the underlying security. Your net profit is $(k^+ - k)e^{rT}$.

Of course, there are limitations on this in real life. Real markets have transaction costs. You cannot borrow money at the risk-free rate. The equation ignores margin calls and other limits on your own short-selling.

Despite these limitations, this analysis shows some interesting principles. In particular, the trading strategies outlined here require no underlying equity. They are constructed from pure leverage.

This is great for increasing market efficiency. It also sets the stage for liquidity crises.

Alpha α and Beta β

All investments are risky. The techniques explained in this series are exceptionally risky. I am not a registered investor, attorney, advisor, broker, banker, lawyer, dealer or anything of the sort. All opinions expressed here are for my personal exploration. I present it to you for entertainment. This post may help you understand policies related to financial regulation for the purposes of more informed voting. This post is not investment advice. This post may contain errors.

There are two ways to profit from the stock market: alpha α and beta β. Extracting rent from your ownership of capital is called β. Arbitraging away market inefficiencies is called α.

Arbitrage has two components: inefficiencies and hedging.

You start by spotting a market inefficiency. In the previous post, a futures contract price $S_0 \neq ke^{-rT}$ constituted a market inefficiency. Wherever there is a market inefficiency there is the opportunity to earn money *on average*. But we are not interested in earning money on average. We care about *risk-adjusted* returns.

Once you have discovered a market inefficiency the second thing you do is construct a financial derivative to hedge away as much risk as you can.

In the first example of the previous post, we created an artificial future by buying the underlying security with borrowed money. In the second example of the previous post, we artificially shorted the future by shorting the underlying security and then lending the resultant cash as a bond. These artificial futures hedged our risk from the original mispriced future. We earned a profit of exactly $|S_0e^{rT} - k|$ independent of how the actual price behaved.

In addition to eliminating our risk, our hedge also removed the necessity of tying up capital. In theory, our futures trading algorithm required zero underlying capital. It is constructed from pure leverage. We can continue the strategy over and over again until $|S_0e^{rT} - k|$ drops below our transaction costs. (Actually, we can continue the strategy over and over again until we are issued a margin call.)

In this way, financial derivatives let you use the magic of leverage to gamble assets worth far more than your capital holdings.

As another example, suppose you knew Tesla will go up 1% in value over the next month. If you are a normal human being then you might consider buying Tesla stock. But if you have been following along so far then you understand why "buying Tesla stock" is almost completely wrong. What should you do instead?

As we showed in the previous post, the price of Tesla futures is a function of the present stock price and the bond rate and has nothing to do with the fact Tesla stock is going to increase in value over the next month. The bond rate r is unlikely to exceed $1.01^{12} - 1 = 0.13$ ROI per year. (In the unusual circumstance it does, then just add a negative sign to everything.) Borrow money at the bond rate and buy Tesla futures expiring in one month. Repeat until you run out of credit or until the present Tesla price represents the future-discounted price $1.01e^{-r(1\text{ month})}$.

Our better solution involves no capital investment. It is pure leverage.

As before, this trading strategy is theoretical. There are hidden risks I have not addressed yet.