Mykyta Medvediev
Professor Lu Zheng
Mgmt 190
February 17, 2022

Barbell Portfolio

The purpose of this paper is to summarize and expand on the midterm project I did for this class. The topic is Nassim Taleb's idea of a barbell portfolio - a set of hyper-secure and hyper-risky investments that make the overall portfolio antifragile (robust to extreme negative returns yet exposed to extreme positive returns).

This paper is structured in the following way: I will shortly recap the midterm paper (background, asset allocation and sample selection, and hypothesis tests) and expand on it with new empirical findings (regression analysis and Monte Carlo simulations). The last section will be the economic conclusions one can draw from this project.

First things first, I'd like to summarize what I already have at this moment. I am a fan of Nassim Taleb, and one of the most useful financial concepts I learned from his books was the idea of antifragility. In his *Skin in the Game,* the author also talks about a particular financial implementation of this concept - namely, barbell portfolio. The idea is simple: cover the worst case scenario (bankruptcy) and stay exposed to the unlimited upside. The way to do this is to invest most of one's funds into paranoidly secure assets (I chose 85% for TIPS, short-term bonds, and gold) and a small portion into ridiculously risky assets (I chose 15% for emerging-market ETFs, high-yield bonds, meme stocks, and cryptocurrency). After some calculations, I observed that both sides had big excess kurtosis, and the risky side of the portfolio was also positively skewed, which indicated that I was moving in the correct direction - covering the worst case scenario with safe securities, while being open for upside. I tested whether there was any correlation between the sides' returns and variances with hypothesis testing and proved that there wasn't any. It means that the sides of the portfolio will not move in the same direction in case of destructive crises.

Now, I also conducted regression analysis to see the extent to which market movement affects the portfolio, as well as how portfolio sides affect each other. But before I reveal the findings, I'd like to take a second to discuss the sample limitation. All of the following calculations for anything that is related to portfolio and the right (safe) side will have a sample size of 643 - just shy of 2 years period. The reason for it is, unfortunately, very trivial - the funds I used to analyze cryptocurrencies started working in 2019, and the way Python Pandas's dataframe works is that it has to have all funds use the same length time-series. Therefore, if some samples are bigger than others, all of them will be cut short to the length of the shortest one. It is unfortunate because the best insight from data would have come from the crises - events that the portfolio is designed to protect from. In principle, if cryptocurrencies were removed from asset selection, I am confident I could at least use 2008 crisis data, although it is just one single crisis and I am not sure whether it would've given me any solid insight. Regardless, all data that are somehow related to the right side will only have a 2-year sample. The same is true for the left side, and although its sample size is higher - 3017, almost 11 years of data - it still doesn't capture any of the most serious crises that it should. This limitation will seriously affect both economic and statistical conclusions.

To the regression analysis we go. First things first, I regressed each side's stocks individually against S&P 500 (to which I will refer to as "the market") returns and found individual beta and intercept (Tables 1 and 2). From observing the results, it is clear that overall, the safe side's betas are much lower than the risky side's ones. It confirms our economic intuition and proves that the safe side is actually safe, and the risky side is actually risky (that is, compared to the safe one).

Next, I conducted four univariate regressions: left side vs the market, gold vs the market, left side vs right side, and portfolio vs the market. (Figures 1-4 respectively.) Let's go through each regression result one by one.

First, I wanted to see how market returns affect the returns of my safe side (Figure 1). Judging by the p-value of 0, I concluded that the coefficient for market return is statistically significant and worthy of interpretation. The coefficient itself is -0.0132, which means that for each 1% that the market returns increase, the returns for the left side of my portfolio decreases by 1.32 basis points. This goes along with my economic intuition because when the market returns go up, market securities become more appealing to investors compared to bonds and commodities, and the latter usually drops in price slightly (since investors give it less value compared to equity securities). The coefficient of determination is 0.004, which means that the market explains 0.4% of the left side's returns (which is what we would expect: in the end, government bonds and equity securities shouldn't depend on each other much), and this can be confirmed by looking at the F-statistic that shows us the statistical significance of the whole model.

From that, I wanted to test regress gold against the market because I know empirically (just by looking on the graph of gold prices) that gold increases in value during market crashes. The results are in Figure 2. First second that I saw a negative coefficient, I was happy that my intuition was more or less correct, but the p-value of 46.6% means that I can't reject the null hypothesis that the coefficient is actually 0, so, unfortunately, I shouldn't overinterpret this finding. Moreso, R^2 indicates that the market explains 0% of variation of gold returns, and F-statistic shows that the model's statistical significance is likely to be non-existent.

Next, I wanted to see how the left side of the portfolio reacts to gains from the right side. The results are in Figure 3. Again, the p-value of 11.8% makes results inconclusive since I can't reject the null hypothesis that the coefficient is 0. R^2 is almost 0, which means that the right side doesn't explain variation of returns of the left side (which is good), and F-stat, yet again, shows that the model's results are likely to be statistically insignificant.

Finally, I wanted to test the overall portfolio against the market (Figure 4). The p-value of the coefficient is 0, which means that it's statistically significant and that we can reject the null hypothesis that the coefficient itself is 0. In fact, the model indicates that it is 0.3357, which means that for each 1% increase in market's returns, portfolio returns increase by 33.57 basis points. The model itself is statistically significant, which is proven by the F-statistic, and the coefficient of determination tells us that the market explains 9% of the portfolio returns variation. It is good because the purpose of the portfolio is to not follow the market up and down but to be prepared when it crashes.

After making conclusions about the regression, I wanted to statistically simulate returns from each side and the overall portfolio, for which I used Monte Carlo simulation. Now, there is a big problem with this kind of simulating and the ensuing analysis: I know as a fact from my midterm calculations that both sides, as well as the portfolio as a whole, have a non-normal underlying distribution - it is thick-tailed with big excess kurtosis and positive skewness; and yet, Monte Carlo simulation is done with normal distribution as the underlying one. I researched the whole internet on how to set up this simulation as if it was thick-tailed, but the only answer I found was in R language and involved so much unnecessary math (frankly, I didn't understand any of it nor how to even blindly implement it) that the mental cost-benefit analysis proved it was definitely not worth it. Besides, we don't really need to know how the barbell portfolio performs statistically as long as we know what it does conceptually.

With that said, Figures 5-8 represent the following, respectively: simulation of the left side returns, right side returns, and portfolio returns in graph and histogram forms. The only robust conclusion to infer from these results is that the right side's returns are much more volatile than that of the left, which is exactly what allows the upside exposure. Also, by observing the histogram, one can conclude that the returns are positively skewed, even without the explicitly embedded skewness.

In conclusion, small sample size seriously affected the results of both regression analysis and Monte Carlo simulation. Barbell portfolio is designed to be antifragile to any kind of event - especially market crushes - and it is unfortunate that my security selection limits the possibility of testing the portfolio under such extreme conditions. Regression analysis has shown that our economic intuitions are mostly right: the safe side's returns are slightly negatively affected with the market returns, while for the overall portfolio, having a bigger market coefficient, the market explains only 9% of the variation of returns. We concluded that it's good because the portfolio shouldn't be affected by market returns much. Monte Carlo simulation, on the other hand, in spite of the fact that its usage in this case is conceptually flawed, still showed that our economic intuition is right and that the overall portfolio has a bigger impact from the risky side as well as a positive skewness, in other words, exposition to the extreme upside, which is one of its two important features.

Tables and figures

Table 1

	beta	intercept
EMSH	0.038466	6.940044e-07
EEM	-0.307605	7.677202e-04
VWOB	-0.060362	2.328074e-04
SPHY	-0.050233	3.338234e-04
BSJO	-0.033589	2.219482e-04
HYG	-0.069050	2.864583e-04
AMC	-0.168981	8.208237e-03
GME	0.070533	1.075110e-02
GBTC	-0.276120	3.528564e-03
ETHE	-0.357526	1.196926e-02

Table 2

	beta	intercept
TIP	0.005288	0.000161
SHV	0.000664	0.000020
BIL	0.000193	0.000015
LTPZ	0.014706	0.000315
VGSH	0.002921	0.000044
VCSH	-0.003790	0.000118
SHM	0.011186	0.000055
BSV	0.002618	0.000075
BWZ	0.031208	-0.000045
GOLD	-0.005142	0.000217

Figure 1

Dep. Variable:	Gold	R-squared:	0.000
Model:	OLS	Adj. R-squared:	-0.000
Method:	Least Squares	F-statistic:	0.5305
Date:	Sun, 13 Mar 2022	Prob (F-statistic):	0.466
Time:	21:06:44	Log-Likelihood:	16228.
No. Observations:	5265	AIC:	-3.245e+04
Df Residuals:	5263	BIC:	-3.244e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0004	0.000	2.716	0.007	0.000	0.001
SPY	-0.0091	0.012	-0.728	0.466	-0.033	0.015

Omnibus:	651.438	Durbin-Watson:	2.015
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6153.269
Skew:	-0.223	Prob(JB):	0.00
Kurtosis:	8.277	Cond. No.	81.4

Figure 2

Dep. Variable:	Left_mean	R-squared:	0.004
Model:	OLS	Adj. R-squared:	0.004
Method:	Least Squares	F-statistic:	13.37
Date:	Sun, 13 Mar 2022	Prob (F-statistic):	0.000260
Time:	21:06:44	Log-Likelihood:	14252.
No. Observations:	3017	AIC:	-2.850e+04
Df Residuals:	3015	BIC:	-2.849e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0001	3.92e-05	2.753	0.006	3.1e-05	0.000
SPY	-0.0132	0.004	-3.657	0.000	-0.020	-0.006

Omnibus:	627.370	Durbin-Watson:	1.751
Prob(Omnibus):	0.000	Jarque-Bera (JB):	18800.623
Skew:	0.247	Prob(JB):	0.00
Kurtosis:	15.219	Cond. No.	92.4

Figure 3

Dep. Variable:	Left	**R-squared:**	0.004
Model:	OLS	**Adj. R-squared:**	0.002
Method:	Least Squares	**F-statistic:**	2.457
Date:	Sun, 13 Mar 2022	**Prob (F-statistic):**	0.118
Time:	21:06:44	**Log-Likelihood:**	2890.3
No. Observations:	643	**AIC:**	-5777.
Df Residuals:	641	**BIC:**	-5768.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.0002	0.000	1.561	0.119	-4.32e-05	0.000
Right	0.0050	0.003	1.567	0.118	-0.001	0.011

Omnibus:	275.685	**Durbin-Watson:**	1.388
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	11453.749
Skew:	1.176	**Prob(JB):**	0.00
Kurtosis:	23.542	**Cond. No.**	30.2

Figure 4

Dep. Variable:	Portfolio	**R-squared:**	0.090
Model:	OLS	**Adj. R-squared:**	0.089
Method:	Least Squares	**F-statistic:**	63.64
Date:	Sun, 13 Mar 2022	**Prob (F-statistic):**	6.90e-15
Time:	21:06:44	**Log-Likelihood:**	1749.4
No. Observations:	643	**AIC:**	-3495.
Df Residuals:	641	**BIC:**	-3486.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.0016	0.001	2.465	0.014	0.000	0.003
SPY	0.3357	0.042	7.977	0.000	0.253	0.418

Omnibus:	914.581	**Durbin-Watson:**	1.928
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	249324.165
Skew:	7.540	**Prob(JB):**	0.00
Kurtosis:	98.282	**Cond. No.**	66.9

Figure 5

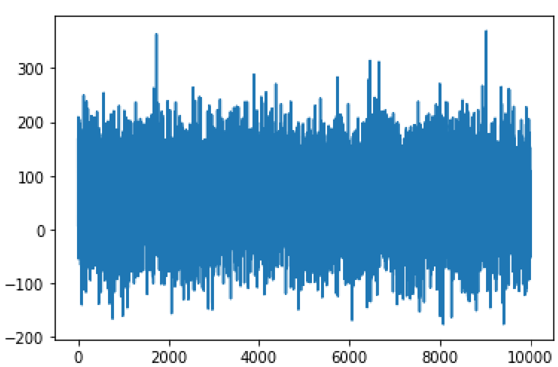

Figure 6

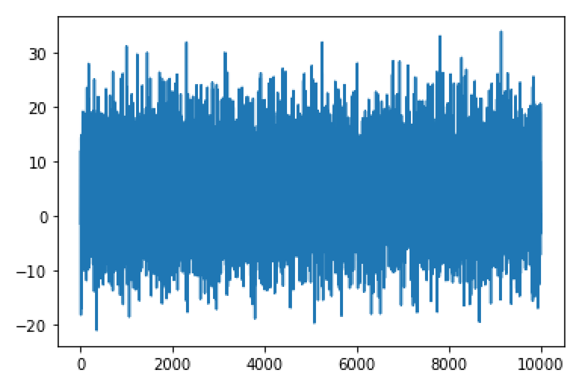

Figure 7

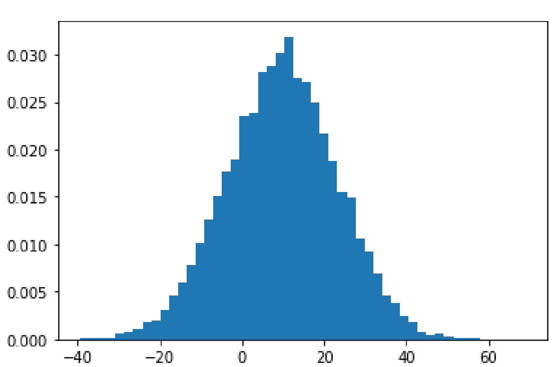

Figure 8