Improving Smart Beta ETFs with Smart Beta.
Chrilly Donninger
Chief Scientist, Sibyl-Project
Sibyl-Working-Paper, August 2016
Revision 1: 2016-08-15

http://www.godotfinance.com/

Smart beta portfolios have been the object of considerable marketing hype. They are more a testament to smart marketing rather than smart investing.
(Burton G. Malkiel [1])

Abstract:
In a previous working paper [2] I analyzed the performance of pairs-trading from a universe of ETF stocks. This paper uses a similar approach but considers long-only strategies which belong to the general group of smart beta strategies: Alternative weighting schemes of the stocks and subset selection according to volatility, beta or momentum. One selects a smart sub-ETF. If the original ETF is already smart the subset is a sort of smart ETF of the second order. The working paper shows that the question of different weighting-schemes is over-weighted. It answers also the FAQ about the impact of different re-balance periods. The risk-adjusted performance can be improved by either low-volatility or small-beta subsets. Momentum and Mean-Reversal works for some but not all ETFs. Revision 1 adds the results for idiosyncratic Volatility.

Introduction:
There is no universally accepted definition of smart better strategies. What most people who use the term have in mind is that it may be possible to achieve greater-than market returns using a variety of relatively passive investment strategies that involve no more risk than would be assumed by investing in a low-cost total stock market index [1]. D. Glushkov has done in [3] a comprehensive analysis of smart-beta ETFs from 2003-2014. His results are mixed. Most of the figures would be even worse in the last two years. A lot of smart-beta ETFs have a small-size and value tilt. The large-cap Russel 1000 ETF IWB has clearly outperformed the small-cap Russel-2000 IWM since the end of the study. Only low-volatility ETFs consistently improved the risk-adjusted – but not the total – performance. The SPLV and the USMV are in the last 2 years even in absolute terms ahead of the crowd whereas an ETF like the SPHB (S&P-500 High-Beta) is in this period a disaster. Some smart ETFs like the PKW (PowerShares Buyback Achievers) worked till 2014 quite fine but under perform the SPY and especially the low-volatility indexes SPLV and USMV since then. This is also a general result of Glushkov. Every style has its good and bad times. But one knows this only in hindsight. It is difficult to find a ETF style selection strategy.

The current working paper employs only smart-beta rules which can be directly calculated from the price data. No other information like value is used. This task is delegated to the ETF providers. It is assumed that they have done their job already properly. The challenge is to improve this selection further.
The data:
I downloaded the recent holdings of industry sector ETFs like IYE, ITB, IBB or IDU and general style ETFs like OEF, QQQ, DIA, USMV, SPLV, PKW from the ETF providers. No attempt was made to reconstruct the historical holdings of these ETFs (retrieving and processing the current holdings is cumbersome enough). The historic simulation is from 2011-01-01 to the current date 2016-08-04. Some of the stocks were during these period not a member of the ETF. The historic simulation has no notion of this. It is not in the strict sense a simulation of e.g. the IYE-stocks. Some of the stocks were even not traded from the beginning. The universe is in most cases at the beginning of the simulation period smaller than the current one. Some stocks which were initially in the ETF are not selected any more. One has a survival bias. This bias is difficult to quantify. The changing weights of the stocks has no effect. The strategies ignore this information. All data are downloaded from yahoo. The calculation uses the adjusted closing prices.

Equal Weighting:
The most trivial smart-beta approach is Equal-Weighting. As already noted simulated trading is done from 2011-01-01 till 2016-08-04. One starts with an initial cash of 500,000$. This is a convention I use in all my trading papers. The position is re-balanced each year (every 252 trading days).

Graphic-1 shows the performance of the Nasdaq-100 QQQ. The red line is the ETF itself. The yellow line the performance of the equal weighted portfolio. Green is the S&P-100 OEF and blue the equal-weighted OEF. The P&L of the QQQ is 123%, with a Sharpe ratio of 0.84, and a max. relative drawdown of 16.1%. The equal weight is with a P&L of 196.8% a Sharpe ratio of 0.99 and a drawdown of 20.2% clearly superior. The values for the OEF are a P&L of 88.6%, a Sharpe ratio of 0.81 and a drawdown of 16.9%. The equal weight is again with a P&L of 126.3%, a Sharpe ratio of 0.95 and a drawdown of 16.7% much better. This result holds for almost all ETFs. The performance difference can certainly not only attributed to the different weighting scheme. There is also a survival/selection bias. The superior performance of equal-weighting is usually explained by a small value tilt but additionally by exploiting mean-reversion by the re-balance.

Graphic-1: Performance of QQQ (red, yellow) and OEF (green, blue)
But the re-balance period has only a minor effect on the performance. It does not really matter if one re-balances every 21, 63, 126 or 252 trading days. The performance drops for longer periods. But the effect is even for 1024 trading days (4 years) not very dramatic. This holds for all considered ETFs.

The PKW (Powershares Buyback Archivers) is the only ETF where the equal-weight is worse than the ETF. The PKW is from a universe of 1000 Nasdaq stocks. *A stock is selected if that have effected a net reduction in shares outstanding of 5% or more in the trailing 12 months.* The ETF holds currently 228 stocks. The PKW outperformed the QQQ till the beginning of 2015 and clearly under performs the QQQ since then. The PKW tries to combine small-size with quality. If a company can buyback shares it has obviously enough money to do so. Small-size under-performed in the last time. This explains the relation between PKW and QQQ. Equal weight has a small-size tilt. This explains the widening gap between the red and yellow line. The general under performance of the equal weight can be explained by a negative selection bias. A company is in the index if it has reduced the shares within the last year. Usually it has not done this in the past. Reducing shares is a way to improve the short-term performance of the stock (the effect on the dividend-yield is permanent, but this effect is priced in).

---

**Graphic-2:** Performance of PWK (red) and Equal-Weight (yellow).

**Graphic-3:** Performance of USMV Equal-Weight with trading costs.
Graphic-3 shows the performance of the USMV (iShares Minimum Volatility US). The P&L is 152.5% with a Sharpe ratio of 1.12 and a max. relative drawdown of 15.6%. The equal-weight portfolio of the QQQ has a higher P&L, but USMV is risk-adjusted ahead. The USMV worked also fine in Pairs-Trading. The ETF and the equal-weight portfolio can not be compared in the considered time range. The USMV was introduced in autumn 2011. The equal-weighted portfolio is for the common time range superior. The USMV has from his inception the highest Sharpe ratio of all considered ETFs. There is as expected a high correlation between the ETFs and the equal-weighted portfolio. So far the P&L was calculated without trading costs. The yellow line in Graphic-3 assumes 10 cents per trade and share. The green line 50 cents. The impact on the performance is only minor. This holds for all equal-weighted portfolios with yearly re-balance. As the period has only an insignificant effect re-balancing once a year is the best choice. The situation is quite different to Pairs Trading. The break-even point was 16 cents per share and trade. The trading costs are for the yearly re-balance for all considered strategies a minor issue.

**Inverse Beta- and Volatility-Weighting:**
A little bit smarter is inverse Beta- and Volatility-Weighting. The weight is in the first case inverse to beta. Beta is calculated in relation to the ETF. If there are – at the beginning of the considered time-range - no ETF data available the SPY is used instead. Beta is calculated with a 1 year time-window. Inverse volatility-weighting is in principle the same. But instead of beta one uses the inverse-volatility. The volatility is calculated from the last 63 trading days. Graphic-4 shows the performance of the Equal-Weight (red), inverse Volatility-Weight (yellow) and inverse Beta-Weight (green) strategy for USMV. The differences are minor. This is a general result. Only PKW is again slightly different. The inverse Beta-Weight is clearly best (but by far worse than for USMV). The inverse Volatility- and Equal-Weight are on par.

![Graphic-4: Performance of USMV Equal-, inverse Beta- and Volatility-Weight.](image)

Although the inverse beta and volatility weights should have a different factor loading the overall effect is small. The stock-selection (which is done by the provider) is obviously much more important. The effect of the re-balance period is also minor.
Graphic-5 shows the performance of the SPY and the Equal-Weighted S&P-500 ETF RSP. The P&L is in the considered time range practically the same. The RSP is just a little bit more noisy. **The question of different weighting is clearly over-weighted.**

Graphic-6: SPY (red) and RSP(yellow)

**Low-Volatility-Subsets:**
The strategies so far have done no stock-picking. The whole universe was traded. Only the weights were altered. The following strategies pick a subset according a given criterion. For this strategy one selects at each re-balance 20 (7) stocks with the lowest volatility. The re-balance is done every 252 trading days. The re-balance period has this time a larger impact than in the strategies above. But the effect is not uniform. Sometimes a period of 126 trading days, sometimes of 252 is best. Volatility is calculated from the last 63 trading days. The picked stocks are equal-weighted. Each pick gets a weight of 5 (14.3) %. The selection logic is similar to the SPLV. The SPLV selects 100 low-volatility stocks from the S&P-500. The USMV 170 stocks from a larger universe. But these ETFs do not use equal-weighting for the selected stocks. The weights are in-between equal- and caps-weighting (maybe they are inverse volatility-weighted). As the weight-question has overall only a minor influence I used the simple equal-weight. Graphic-6 shows the performance of the USMV (red), QQQ (yellow), OEF (green), PKW (blue) and ICF (dark-blue) subsets. The ICF (iShares Cohen&Steers REIT) holds only 30 stocks. So the selection was restricted to 7. For all the other ETFs 20 stocks are picked. The QQQ subset has with a P&L of 189.4% and a Sharpe ratio of 1.10 and drawdown of 15.9% the best performance. The P&L is a little bit lower than the equal-weight portfolio, but the Sharpe ratio is considerable improved. The OEF has a much smaller P&L, but the Sharpe-ratio is close to the QQQ. Interesting is also the behavior of the PKW subset. The performance does not drop off in 2015. The low-volatility filter removes the junk. The P&L of the ICF is quite good (much better than the overall Equal-Weight), but the small subset introduces a relative high volatility.
One can also select the 20 stocks with the highest volatility. Once upon a time it was assumed that investors get rewarded with higher returns for the additional risk. This is partly true for QQQ. The return is till 2014 much higher. But then it deteriorates. For the other ETFs the P&L is not higher (usually worse) and the Sharpe ratio is uniformly much lower.

**Low-Beta-Subsets:**
One can also select by beta. One picks the 20 (7) stocks with the lowest beta. Graphic-7 shows the performance of the Low-Beta subset for the same ETFs (and colors) as before.

The best ETF is now the PKW with a P&L of 177.4% a Sharpe ratio of 1.07 and a drawdown of 17.0. But the OEF has the highest Sharpe ratio of 1.09. The ICF has with a P&L of 218.4% a stellar win but the up and downs are considerable. The Sharpe ratio is hence only 1.01.

The performance of the USMV is smooth, but the P&L is much smaller. The USMV is already a low-volatility selection. Selecting from this universe in a second step low-beta stocks is too much of the same recipe.
Graphic-8 shows the performance industry-sector ETFs ITB (iShares US Home Construction) with 7 stocks selected in red and IYH (iShares US Healthcare) with a 20 (yellow) and 7 stocks (green) subset. ITB has with a P&L of 313.9% a Sharpe ratio of 0.99, and a drawdown of 31.8% a very high profit. But due to the small subset (there are only 42 stocks in ITB) there are considerable up-and downs.

IYH has with 7 selected stocks a P&L of 290.0%, a Sharpe ratio of 1.08 and a drawdown of 20.2%. For 20 stocks the P&L is 176.9 with a Sharpe ratio of 0.99 and a drawdown of 19.9. The sector ETFs have also small-size stocks in their portfolio. These stocks perform bad in recent time. The low-beta filter does a good job to remove this junk and to leave nevertheless interesting stocks in the portfolio. The up and downs are more an effect of the small subset size.

**High-Beta-Subsets:**
Graphic-9 shows the performance is one selects the 20 (7) high-beta stocks.

This improves for the USMV the P&L to 193.2%. The Sharpe-ratio is with 1.01 similar to the Low-Beta selection. The QQQ has first a stellar performance but the strategy does
not work anymore since 2015 (yellow). For the other ETFs the P&L and the Sharpe ratio is worse than for the Low-Beta selection. The trading costs are for all the subset strategies a minor issue. There is little exchange of stocks in the subsets. It is essentially also an Equal-Weight re-balance.

**Momentum:**
According to Asness et al. momentum is everywhere [4]. This claim has been questioned by other authors. One calculates the return within the last year (252 trading days) but excludes the last month (21 trading days). One selects the stocks with the highest momentum. The overall result was similar to Pairs Trading (see [2]). It works reasonable for high-quality ETFs, but if fails for junk.

Graphic-10 shows the performance for USMV (red), QQQ (yellow), ITB (green) and IYH (blue). USMV has with a P&L of 157.5%, a Sharpe ratio of 1.14 and a drawdown of 15.6%. ITB and IYH have with 7 selected stocks again a stellar performance, but the up- and downs are not for the faint-hearted. The trading costs are somewhat higher than in the previous selections. A low-beta stock is usually also in the next period within the selection. Momentum stocks have a higher turn-over rate. For trading-costs of 10 cent per share and trade the P&L of UMSV drops to 150.4% with a Sharpe ratio of 1.12 and a relative drawdown of 15.6%. A re-balance period of 21 tradings days improves the performance without trading costs. But this does not compensate the considerable higher trading cost drag.

![Graphic-10: Performance of Momentum-Subsets](image)

**Mean-Reversal:**
One excludes for the momentum calculation the last month (21 trading days) because there is usually a mean-reversion effect. It is therefore sensible to select the stocks which have performed worst within the last month.

This works indeed similar to momentum. Graphic-11 shows USMV (red), QQQ (yellow), OEF(green) and ITB(blue). It is argued in the literature that the mean-reversal lasts only for a short term. The performance is indeed for a re-balance period of 21 trading-days better. But the yearly re-balance has after trading costs a clear edge.
The USMV performs with a P&L of 215.4%, a Sharpe ratio of 1.16 and a relative drawdown of 13.1% quite nice. The QQQ has a similar P&L but higher volatility. The ITB has a P&L of 308% but the Sharpe ratio is due to the large swings relative low. The trading costs are similar to the momentum strategy. Mean-Reversal is a close relative of momentum. It works and fails for the same group of ETFs.

**Strategy-Comparison:**
Graphic-11: Performance of Mean-Reversal-Subsets

Graphic-12 shows the performance of different strategies for the SPMO (S&P-500 Momentum). “The Index tracks the performance of stocks in the S&P 500 Index that have a high "momentum score". The Fund and Index are reconstituted and rebalanced twice a year on the third Fridays of March and September. Constituents are weighted by their market capitalization and their momentum score” (from the fact-sheet).

The ETF was introduced at 2015-10-22. It was so far no success story. Liquidity is practically zero. The ETF has very high weights for few selected stocks. The weight for MSFT is currently 9.19, Google (GOOG and GOOGL) has a weight of 9.22. But one can use the universe to calculate the performance of the different strategies.

Graphic-12: Strategies-Performance for SPMO.
The equal-weighted portfolio (red) has a P&L of 219.0%, a Sharpe ratio of 1.15 and a relative drawdown of 17%. The minimum volatility subset with 20 stocks (yellow) has P&L of 187.7%, a Sharpe ratio of 1.17 and a drawdown of 9.8%. The minimum-beta subset (green) a P&L of 196.7%, a Sharpe ratio of 1.18 and a drawdown of 10.9%. The momentum subset (light-blue) a P&L of 175.6% a Sharpe ratio of 1.13 and a drawdown of 16.1%. Mean-Reversion (dark-blue) has with 282.9% the by far highest P&L, but there are considerable up- and downs. The Sharpe ratio is 1.15 and the drawdown 15.7%. The ETF seems to be – due to the concentration on a few stocks – no good idea. But the selected stock universe is an interesting starting point for the strategies.

Graphic-13 shows the performance of the USMV for the different strategies. The meaning of the colors are the same than for the SPMO. The detailed results have already been presented above and can be found also in table-1 in the Appendix.

Idiosyncratic Volatility: (Revision 1).
There is an own strand of literature about the effects of idiosyncratic volatility. Conceptually the meaning of the term is clear. But there are different approaches how to measure it (see [5]). I used the most trivial approach.

\[ IR_t = R_t - IND_t \]

The idiosyncratic Return is the deviation of the stock from the index. The index is for this application the ETF. If there are no data for the ETF available the SPY is used. This is the same procedure than for the beta calculation. Another method is to estimate first beta and to calculate the deviation from beta*IND. I preferred the more stable and robust approach of above. According to [5] the results are similar.

Graphic 14 shows the performance for the USMV and the SPMO universe. Idiosyncratic volatility is calculated – like plain volatility – over the last 63 trading days. Re-balance is done once a year. The red chart selects from USMV the stocks with the lowest, the yellow line with the 20 stocks with the highest volatility.
The performance is – surprisingly – similar. The low-volatility selection has a P&L of 172.7%, a Sharpe ratio of 1.08 and a drawdown of 18.2%. The high-volatility subset (yellow) a P&L 158.2%, a Sharpe ratio of 1.11 and a relative drawdown of 16.6%. These values are better than the plain volatility strategy. The low- and high-volatility strategy differs for the SPMO significantly. The low-volatility subset (green) has a P&L of 202.8%, a Sharpe ratio of 1.08 and a drawdown of 18.7%. The high-volatility subset (blue) a P&L of 283.6%, a Sharpe ratio of 1.12 and a drawdown of 21.1%.

Graphic-15 shows the results for OEF and QQQ. The low-volatility OEF subset (red) has a P&L of 113.1%, a Sharpe ratio of 0.92 and a drawdown of 15.8%. The high-volatility OEF subset (yellow) a P&L of 155.8% a Sharpe ratio of 0.86 and a drawdown of 23.5%. The low-volatility QQQ subset (green) has a P&L of 139.1%, a Sharpe ratio of 0.94 and a drawdown of 19.1%. The high-volatility QQQ subset (blue) has a P&L of 293.9% a Sharpe ratio of 0.91 and a relative drawdown of 28.0%.

The results confirm – besides for USMV – the literature. High idiosyncratic volatility stocks have a higher return. But there is no free lunch. The Sharpe ratios are not superior. The often used term “anomaly” is in my view not appropriate.
Conclusion:
The question if different weighting schemes beat a cap-weighted portfolio can not be answered with this study. There is a survival/selection bias. The bias is negative in case of PKW and very likely positive for the other ETFs. But the selection bias should have only a small influence when one compares the strategies with each other. The effect of different weighting schemes is minor. If one selects the whole universe also the re-balance period has within a wide range a minor effect. A yearly re-balance is the best compromise. Selecting a subset usually improves the performance. Selecting low-volatility stocks seems to be a safe bet. One does not improve P&L, but reduces the risk. Momentum and – to my surprise – Mean-Reversion works reasonable if one selects from a high-quality universe like USMV. The results for idiosyncratic volatility are inline with the literature. High idiosyncratic volatility boosts return. But there is no free lunch. The Sharpe ratios are about the same than in the other strategies. The idiosyncratic volatility results confirm once again: Universe selection is more important than the strategy. This was also the main conclusion of the Pairs-Trading strategy. But there is the usual disclaimer: Past performance is not a guarantee of future results; current performance may be higher or lower than performance quoted. This can be best seen for some of the smart-beta ETFs with a small-size tilt. They worked – like PKW – fine for some time but deteriorated considerable in the last two years. The same can of course happen with the considered selection rules.

References:
## Appendix: ETF-Performance-Table:

<table>
<thead>
<tr>
<th>ETF</th>
<th>Equal-Weight</th>
<th>Min.-Volatility</th>
<th>Min.-Beta</th>
<th>Momentum</th>
<th>Mean-Reversion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P&amp;L</td>
<td>ShR</td>
<td>DD</td>
<td>P&amp;L</td>
<td>ShR</td>
</tr>
<tr>
<td>OEF</td>
<td>126.3</td>
<td>0.94</td>
<td>16.7</td>
<td>118.7</td>
<td>1.06</td>
</tr>
<tr>
<td>QQQ</td>
<td>196.8</td>
<td>0.99</td>
<td>20.1</td>
<td>189.3</td>
<td>1.09</td>
</tr>
<tr>
<td>PKW</td>
<td>81.3</td>
<td>0.64</td>
<td>27.6</td>
<td>128.2</td>
<td>1.02</td>
</tr>
<tr>
<td>SPLV</td>
<td>141.2</td>
<td>1.08</td>
<td>15.0</td>
<td>129.3</td>
<td>1.04</td>
</tr>
<tr>
<td>USMV</td>
<td>152.5</td>
<td>1.12</td>
<td>15.6</td>
<td>131.2</td>
<td>1.05</td>
</tr>
<tr>
<td>SPMO</td>
<td>219.0</td>
<td>1.15</td>
<td>17.0</td>
<td>187.7</td>
<td>1.17</td>
</tr>
<tr>
<td>SPVU</td>
<td>79.3</td>
<td>0.60</td>
<td>29.3</td>
<td>79.8</td>
<td>0.81</td>
</tr>
<tr>
<td>SPHB</td>
<td>80.1</td>
<td>0.46</td>
<td>35.5</td>
<td>55.4</td>
<td>0.40</td>
</tr>
<tr>
<td>SPHD*</td>
<td>69.6</td>
<td>0.76</td>
<td>17.8</td>
<td>99.0</td>
<td>0.93</td>
</tr>
<tr>
<td>DSY</td>
<td>99.2</td>
<td>0.88</td>
<td>17.7</td>
<td>110.7</td>
<td>0.98</td>
</tr>
<tr>
<td>FTS*</td>
<td>157.1</td>
<td>1.07</td>
<td>17.1</td>
<td>144.4</td>
<td>1.07</td>
</tr>
<tr>
<td>HDV</td>
<td>100.6</td>
<td>0.90</td>
<td>16.7</td>
<td>107.7</td>
<td>0.95</td>
</tr>
<tr>
<td>IBB</td>
<td>229.9</td>
<td>0.66</td>
<td>47.9</td>
<td>159.1</td>
<td>0.75</td>
</tr>
<tr>
<td>ICE*</td>
<td>142.0</td>
<td>0.80</td>
<td>22.4</td>
<td>161.5</td>
<td>0.87</td>
</tr>
<tr>
<td>IDU*</td>
<td>117.7</td>
<td>0.86</td>
<td>16.6</td>
<td>125.0</td>
<td>0.89</td>
</tr>
<tr>
<td>ITB*</td>
<td>173.5</td>
<td>0.68</td>
<td>42.3</td>
<td>192.7</td>
<td>0.94</td>
</tr>
<tr>
<td>IYC</td>
<td>146.4</td>
<td>0.92</td>
<td>21.5</td>
<td>154.8</td>
<td>1.05</td>
</tr>
<tr>
<td>IYE</td>
<td>11.9</td>
<td>0.00</td>
<td>61.3</td>
<td>-0.6</td>
<td>0.00</td>
</tr>
<tr>
<td>IYH</td>
<td>251.2</td>
<td>0.95</td>
<td>29.6</td>
<td>153.4</td>
<td>1.04</td>
</tr>
<tr>
<td>IYM*</td>
<td>82.4</td>
<td>0.46</td>
<td>36.9</td>
<td>82.3</td>
<td>0.67</td>
</tr>
<tr>
<td>QUAL</td>
<td>139.7</td>
<td>0.94</td>
<td>19.3</td>
<td>133.3</td>
<td>1.08</td>
</tr>
<tr>
<td>XRLV</td>
<td>135.7</td>
<td>1.00</td>
<td>18.9</td>
<td>131.6</td>
<td>1.05</td>
</tr>
</tbody>
</table>

ETFs marked with a "**" select 7 stocks, all others 20.
ShR = Sharpe Ratio
DD = max. relative Drawdown
Entries in red-contour mark the highest P&L of a strategy.
Entries in green-contour mark the highest Sharpe-Ratio of a strategy.
The numbers do not consider trading costs.