Timing the Tail-Risk-Protection of the SPY with VIX-Futures by a Hidden Markov Model.
The Wool-Milk-Sow Strategy.

Chrilly Donninger
Chief Scientist, Sibyl-Project
Sibyl-Working-Paper, April 2017

"In short, by the time we are aware of a crisis it is usually too late to diversify into volatility"[1].

Abstract:
Protecting an equity market portfolio with VIX-Futures eats not only the kurtosis[2] but also the profits of the portfolio. Being constantly VIX-Futures long is too expensive[1]. Therefore one has to find an appropriate timing strategy. This working paper presents a Hidden-Markov-Model which not only has a reasonable tail-risk-protection but even improves the overall return of the SPY. The strategy is – at least in the historic simulation – close to what is called in German an "eierlegende Wollmilchsau" ("egg-laying wool-milk-sow"[3]).

Introduction:
There are papers which show the profitability of protecting an equity market portfolio with VIX-Futures. But they only consider the period of the 2008 crash. It is a trivial result that an insurance is beneficial in case of a damage. There is general consensus that paying constantly the insurance premium is too costly. One has to find a reasonable working timing-mechanism. The starting point of this work was the paper "Effective and Cost-Efficient Volatility Hedging Capital Allocation: Evidence from the CBOE Volatility Derivatives" by Yueh-Neng Lin[4]. The volume of the long VIX-Futures position depends on different market regimes. The regimes are defined by the realized volatility of the SPY and the trend of the VIX. If the realized volatility RV is below 10% and the VIX is additionally trending down no Futures are hold at all. If the RV is above 45% and the VIX is trending up, one holds 65% of the SPY position in VIX-Futures. The other regimes/states are in between. Lin also considers other hedging tools like S&P-Puts and Variance-Futures. But his results are best for the VIX-Futures. The presented strategy is – albeit with other parameters – according my own simulations indeed relative attractive. The tail-risk is reduced considerably, the Sharpe-ratio is slightly improved but there is also a significant drag in performance.
The intuitive notion behind this – and other – approaches is: One should hedge in a risky, volatile regime and one should not pay the insurance premium in a quiet bull-market. A well known statistical tool to address such a question is the Hidden-Markov-Model (HMM). The method was initially developed in the field of temporal pattern recognition such as speech, handwriting or gesture recognition.
The assumptions of the model are: The probability of the observed sequence is identically and independent distribution for a given – non observable – state, but differs significantly between states. The observed variable is usually correlated. The correlation is a result of state transitions. The state-transitions depend only on the last state (this is the Markov-condition) and not on the full state-space history. The Markov-condition greatly simplifies the estimation of the hidden latent states. The number of states must be specified beforehand. The observed sequence in the original applications is a discrete variable. But it is relative straightforward to extend the model to a continuous distribution. It is usual to assume a Multivariate Normal-Distribution. This assumption can be justified by the fact that almost any (reasonable) distribution can be approximated by a mixture of Gaussian's. But this is only a theoretical result. The number of states would usually be too large to be of any practical use. According the motto "All models are wrong, some are useful" one is interested in calculating in a reliable way the

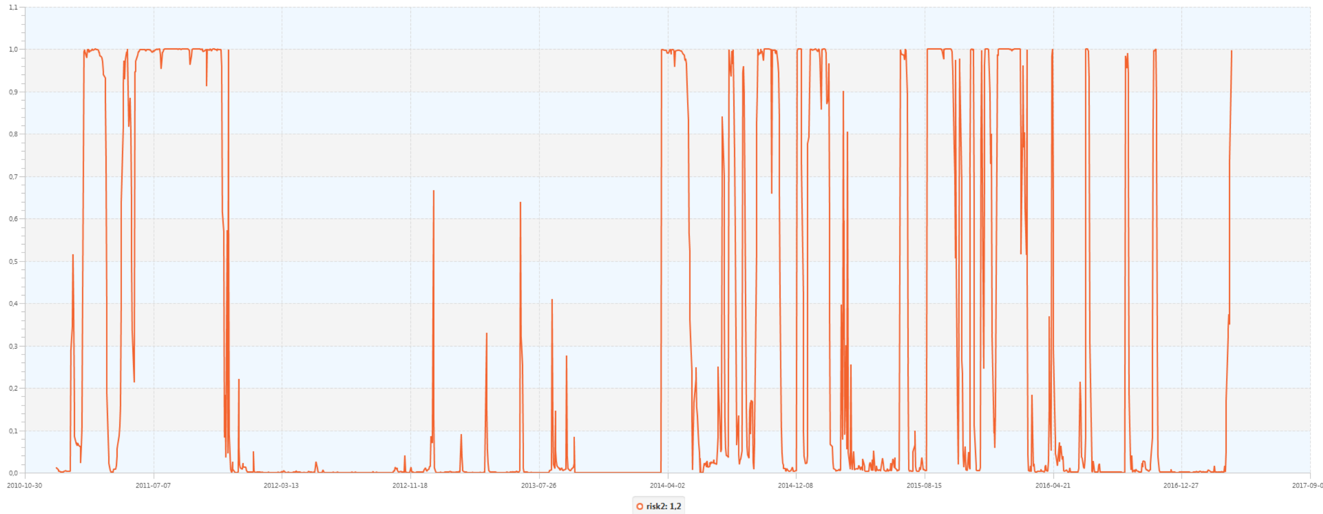
latent states and does not care too much about the Gaussian assumption and the Markov-condition.

Following the mentioned paper of Lin the HMM uses a Multivariate-Normal with 2 Dimensions. For Dimension 1 I tried the daily-return and the daily-absolute-return.
Note: The daily return is scaled by 100.0, the absolute-return by 100*sqrt(252.0) to get easy to interpret numbers. This does not change the mathematical properties of the model.
The daily-absolute-return aka volatility performs generally slightly better. But the effect is not dramatic.
For the second VIX-related dimension I tried the VIX, the Implied-Volatility-Term-Structure (IVTS) and the VIX-Futures Term-Structure.
The IVTS is defined as VIX/VXV. The VXV is the 3-month volatility index. I have used the IVTS in several trading strategies with good success (see [5] and the references herein).
The VIX-Futures Term-Structure is the difference in price between the 2nd and 1st Future. The 1st Future is rolled over 5 trading days before expiry. This avoids the erratic price behavior of the 1st Future immediately before expiry. The hedging-strategy uses the same roll-rule.
Using the VIX does not work at all. It makes a big difference if the VIX is falling from 30 to 20 or increasing from 15 to 20. The Lin-paper addresses this problem by using the VIX-trend. The IVTS does considerable better. The VIX moves much faster than the VXV. The IVTS takes not only the absolute value but also the movement of the VIX into account.
The clear winner is the VIX-Futures Term-Structure. The Term-Structure is usually positive, the Futures are in contango. The price difference is typically 1.0. The Futures are only during a market-turmoil in backwardation. The Future has in backwardation a positive roll-yield and hence good chances to be profitable. Even if the classification of the HMM – *attention, buy insurance, we are entering a risky-state* – is wrong, the decision can be nevertheless profitable. In other words: the VIX-Futures Term-Structure is not necessarily the best regime-classifier. But it is gives clearly the best trading advice.
Note: The IVTS and the Futures-Term-Structure are highly correlated. But the 1st and 2nd Futures react faster and the Term-Structure measures directly the relevant trading information.

Implementation and Trading-Rules:
The parameters of the HMM are calculated with the Baum-Welch Algorithm. The algorithm is conceptually simple, but handling the numeric problems of over- and underflow is not trivial.
I ported the C++ implementation of Press[6] to Java. The implementation of Press assumes a discrete distribution. It is straightforward to replace this term by the density of the 2D-Normal. But the implementation failed when I tried also a plain Normal-Distribution. The algorithm selects for one state a single observation and classifies the other measurements to the remaining state. This single observation has a variance of zero and hence an infinite density. This is a known problem which can be addressed by regularization terms[7]. I did not encounter this problem for the 2-dimensional case.
For trading purposes one recalculates for each day the HMM. One uses always a sliding window of 2 years (504 trading-days). I tried a HMM with 2 and 3 states. The 2 states model is clearly superior.
The Baum-Welch algorithm calculates for the whole time-range the daily state-probabilities. But only the last entry is of interest. The algorithm has of course no notion of a risky state. It classifies the trading states to maximize the overall likelihood. The numbering of the states is arbitrary. It happens that the risky-state is at time T state-0 and at time T+1 state-1. A state k (with k=0,1) is defined risky, if the state-marginal mean of the absolute return of k is greater than of the marginal mean of the other state and the marginal-mean of the VIX-Futures Term-Structure is less than in the alternative state.
There are a few occasions where e.g. state 1 has the greater mean absolute return but also the greater mean Term-Structure. This can happen if there was not a major market-turmoil within the last 2 years. There is in fact only one market regime. But the model always tries to find two and places the

measurements rather arbitrarily into the state-bins. This situation was resolved by setting the risky-state-probability to Zero. Theoretically it could also be a crash lasting for 2 years. But this is an extremely unlikely situation. Using a 2 years window means that the definition of risky, the distribution of the absolute-return and of the Term-Structures differs from time period to time-period. It is a relative measure: Risky in terms of the last 2 years.

Graphic-1 shows the probability of being in the risky-state. The transition is usually very fast from a probability close to Zero to close to One. This holds also in the other direction. One sees on the left the relative long lasting market-turmoil in the summer of 2011. The other risky-regimes are of shorter duration.

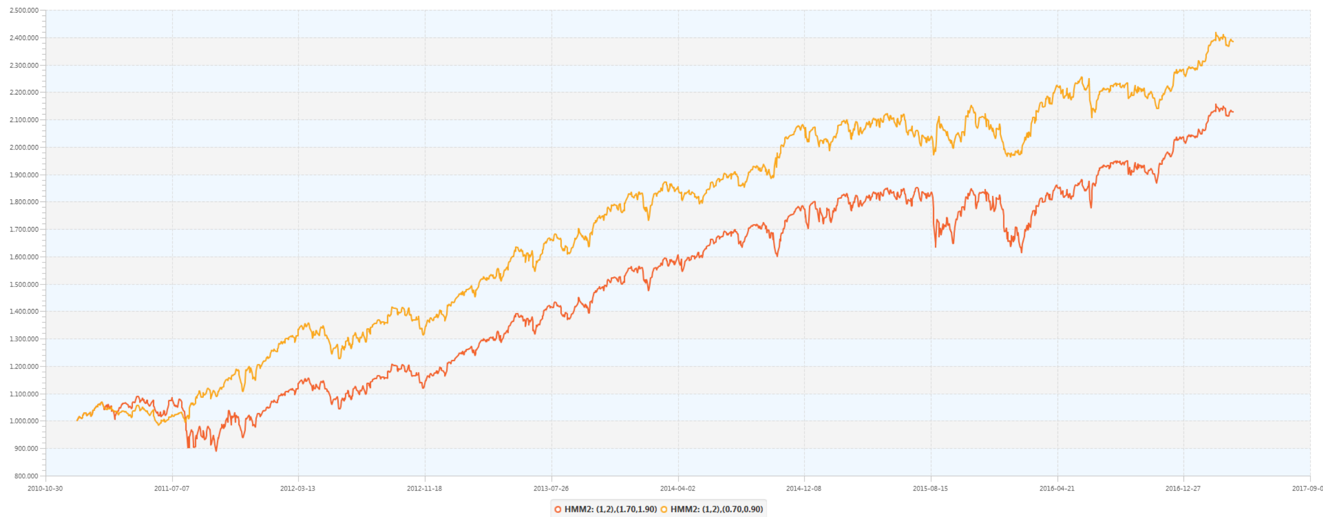


Graphic-1: Risky-State-Probability from 2011-01-03 till 2017-04-04

The Trading-Strategy:

One buys initially at 2011-01-03 for 1.000.000$ the SPY-ETF. The hedge goes the 1st VIX-Future long. The position is rolled over 5 trading-days before maturity. The minimum-maturity of a new VIX-Future must be 15-trading days. This reduces the number of roll-overs. The trading cost is the minimum spread of 0.05 or 50$ per Future and Trade. One holds no Futures if the Risk-Probability is below 0.7. Between 0.7 and 0.9 the Futures positions is 6% and above 0.9 18% of the current value of the SPY-position. If the SPY has currently a value of 1.000.000 one would invest a value of 60.000 or 180.000. If the Futures price is 20.0, one would buy 3 or 9 Futures. One can only trade a full Future. The volume is rounded to the nearest integer-value. One does not adjust the Futures-volume as long as one stays within the same regime. The relative value of the Futures position can thus (considerably) differ from the initial allocation.
The Futures are financed from a cash-account. But the cash-account is cleared at the end of each month. If the Futures made some profit, one increases the SPY allocation accordingly. If the cash balance is negative, one sells the same amount of the SPY. The SPY can also only be hold in full units. All the operations are up to rounding effects. The trading costs of the SPY are 1 Cent per Trade. The overall transaction volume of the SPY is relative low. Different (higher) transactions costs change the results only marginally.

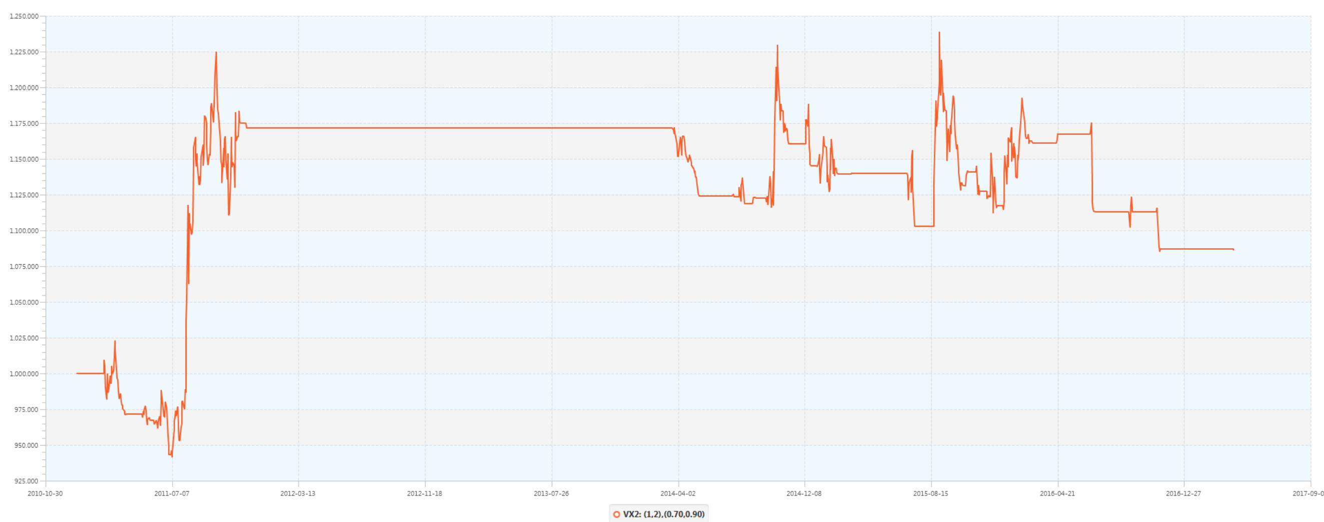
Graphic-2 shows in red the performance of the Buy&Hold strategy. The calculation uses the adjusted close-price. The overall Return is 112.9% with a Sharpe-Ratio of 0.86, and a max. relative Drawdown of 18.6%. The yellow chart is the performance of the HMM-Hedge with an overall Return of 138.5%, a Sharpe-Ratio of 1.0 and a max. relative Drawdown of 9.7%.



Graphic-2: SPY Buy&Hold (red) and HMM-Hedged (yellow) from 2011-01-03 till 2017-04-04

Graphic-3 shows the performance of a strategy where one has initially a cash account of 1.000.000$. One trades only the VIX-Futures according the hedge-rules. One makes an overall profit of 8.6%. The profit is made in the crash of 2011. After the crash there are till March 2014 no trading-activities. The Futures looses in spring 2014 some money, which is recovered afterwards. Overall there are – after the 2011-crash some slight losses. But the spikes hedge quite good the SPY-losses.

Note that the combined effect of the SPY&Futures position is larger. The profit of the VIX-Futures position is invested end of August, September, October and November 2011 in the SPY. One has some money available to buy low. The same happens to a lesser amount in the 2014 and 2015 Drawdowns of the SPY. The losses of the Futures position are usually paid from a rising SPY. This effect is also valid for other hedging-strategies. But the drag of the hedge is usually much larger.



Graphic-3: Performance of Long-Futures from 2011-01-03 till 2017-04-04

Conclusion:

The performance of the strategy is the best I have found so far in the literature. It is almost too good to be true. I have therefore checked with considerable rigor my calculations to find a bug. Especially if the HMM model uses information from the future trading days and is hence a smoother and not a filter. But

I have found – so far – no error. If one replaces the VIX-Futures Term-Structure with the IVTS or with the VIX, the performance is – for an otherwise identical calculation – also much worse.
I have not tuned the parameters of the 2 years estimation-window. This was an educated guess before I started any Risk- and Trading-calculations. The settings of the risk-probability thresholds is not critical. As can be seen in Graphic-1 the probability moves quite fast from very low to very high. It does no really matter if one sets the high-threshold to 0.85, 0.90 or 0.95. The leverage of 6% and 18% was tuned somewhat by hand. But the strategy has within wide bounds a superior return and a much lower Drawdown than the buy&hold strategy.

Further Work:

One can apply the HMM to other allocation problems. The Herorats strategy[8] switches between an equity- and a bond-market ETF. The decision is based on the IVTS. If the IVTS is high, one holds the bonds, if the IVTS is low the stocks. Maybe this strategy can also be improved by an HMM.
There are several papers which use the HMM for modeling high-frequency FX trading[9]. This seems to be the most interesting direction for further investigations.

References:

1. Carol Alexander, Dimitris Korovilas: Effective and Cost-Efficient Volatility Hedging Capital Allocation: Evidence from the CBOE Volatility Derivatives. ICMA Centre Discussion Papers in Finance DP2011-04.
2. Keith Black: Improving hedge fund risk exposures by hedging equity market volatility, or how the VIX ate my Kurtosis, Journal of Trading 1, 6–15.
3. Wikitionary - Eierlegende Wollmilchsau.
4. Yueh-Neng Lin: Effective and Cost-Efficient Volatility Hedging Capital Allocation: Evidence from the CBOE Volatility Derivatives.
5. Chrilly Donninger, Improving the S&P Dynamic VIX Futures Index. The Mojito 3.0 Strategy, Sibyl-Working-Paper, Dec. 2013
6. William Press et. al: Numerical Recipes, 3rd Edition, Chapter 16.3.
7. Andrew Fraser: Hidden Markov Models and Dynamical Systems. Chapter 3.1.2
8. Chrilly Donninger: How to beat the market with the Implied Volatility Term Structure. The HeroRATs Strategy. Sibyl-Working-Paper, Dec. 2013
9. Anton Tenyakov, Rogemar Mamon: Modelling high-frequency FX rate dynamics: A zero-delay multi-dimensional HMM-based approach.