

# The Identification of Market-Regimes with a Hidden-Markov Model

by

Dr. Chrilly Donniger

Chief Scientist, Sibyl-Project

Sibyl-Working-Paper, June 2012

<http://www.godotfinance.com/>

Financial assets follow according to the classical model a brownian motion. This is also the basic assumption of the Black-Scholes options formula. It is well known, that this model is too simple-minded. There is a clear volatility clustering effect. The temperature of the brownian liquid varies over time. But also the overall trend changes between bull- bear- and sideways markets. Correct classification of the market-regime aka sentiment is of paramount importance for successful trading. Hence a lot of indicators have been developed for this purpose. A simple example is the number of advancing versus declining stocks. This widely used indicator is in my view not very useful. The stocks are highly correlated with the overall market. One can just look at the S&P-500 to get the same information. If the S&P rises significantly, most stocks rise too. If the S&P goes sideways, about half of the stocks move slightly down, the other half up. If the S&P declines sharply, practically all stocks follow (the correlation of assets is generally higher in falling than in rising markets).

A slightly more sophisticated and probably more useful indicator is the comparison of new highs and new lows. The Aroon indicator developed by T. Chande in 1995 is a further refinement of this idea. The Aroon indicators measure the number of periods since price recorded an x-day high or low (see [1]) The indicator has a sound basis in the theory of extreme-values (see [2]). But the technical analysis literature does not mention this connection. (market-technicians are usually not expert's in advanced mathematics).

This working paper follows the approach of Mulvey&Zhao [3]. The market-regimes are modeled as a hidden-markov model. In a markov-model the system has a fixed number of states. In each time step it switches from one state to another (it can of course stay in the same state. In this case the switch is to its previous step). The term "markov" means, that the switching probabilities depend only on the previous state. The history before does not matter at all. Systems have usually memory. But the Markov-assumption simplifies the mathematics and the computations considerable.

In a *hidden* Markov model, the state is not directly visible, but output, dependent on the state, is visible. Each state has a probability distribution over the possible output. One infers from the output the hidden-state. Hidden Markov models are especially known for their application in speech-recognition and bioinformatics. In the original model the output are discrete tokens. E.g. the different DNA nucleotides. One infers from the statistical distribution of the nucleotides the position or the function of the DNA-sequence. In case of assets prices the output are the assets-returns. This is a continuous variable. To infer the hidden states one has to specify a distribution model. The usual assumption is that the returns are drawn from a multivariate normal-distribution. It is well known, that assets-returns are not normal. They have fat tails. But the assumption only states, that the returns of each state follows a multivariate distribution. The observed values are a mixture of the state-distributions. The mixture has also fat-tails. The assumption is hence not too far from reality.

The hidden-markov algorithm solves – surprisingly – two problems at once. It determines for each time-step the state-probabilities and the parameters of the state-dependent multivariate distributions. The values are found iteratively with the Baum-Welch algorithm. One starts with arbitrary parameters for the state and the distribution. The algorithm holds the distribution fixed and calculates state-probabilities which are in better agreement with the data. Once the state-probabilities are calculated, the distribution parameters are improved. The iteration starts again by calculating the new state-probabilities. It can be shown, that the algorithm converges to a model

which explains in a statistical sense the data best. A good description of the algorithm plus efficient C++-code can be found in Press et al. [5]. I have ported this code for the Godot-Server to Java. The original code assumes discrete tokens. I have extended the algorithm to a multivariate distribution. The Baum-Welch algorithm is conceptually simple but it is numerically quite nasty. The algorithm calculates the probability of a given sequence. This probability tends very fast to zero. Avoiding underflow is a considerable implementation challenge.

Mulvey&Zhao use in [3] eight representative economic indicators: The S&P-500 (STK), Zero Coupon Bond Prices (BND), U.S. \$ Index (USD), The CBOE Volatility Index VIX, Equity Dividend Yield (EDY), the U.S. Interbank 3M Interest Rate (UIR), Treasury Yield Spread (TYS) and US. Credit Spread (UYS). They use weekly data. For STK, BND, USD and VIX the state-output is the return. For the other data the difference of the values is used.

Using weekly data is a compromise. If one uses daily values, one has at the time horizon the latest available information. Like in any estimation problem the quality of the Baum-Welch parameter estimation increases also with the number of data points. But daily data are rather noisy. One feeds a lot of noise in the model. Weekly data have a better signal to noise ratio. The weekly returns are calculated by the authors from Wednesday to Wednesday. Wednesday has the least number of public holidays. One avoids also weekend effects.

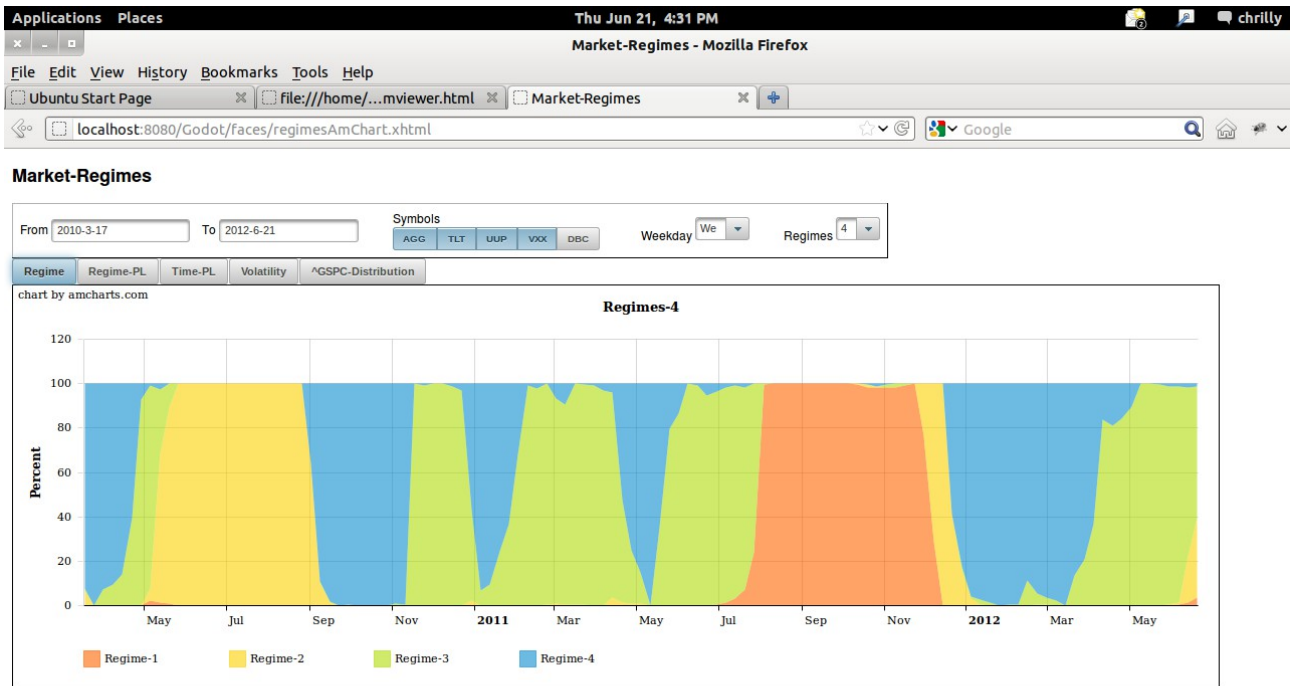
For the Godot-model I have replaced most of the time-series by ETF's. ETF data can be easily downloaded from yahoo-finance (and other sources). The Godot-Model uses ^GSPC (SP&500), AGG (iShares Aggregate Bonds), TLT (iShares 20+ Year Treasury Bonds), UUP (PowerShares DB US Dollar Index), VXX (iPath S&P-500 VIX Short-Term Futures), DBC (PowerShares DB Commodity Index). One uses for all values the weekly-return.

Another important parameter is the number of states/regimes. All models I am aware of use either 3 or 4 regimes.

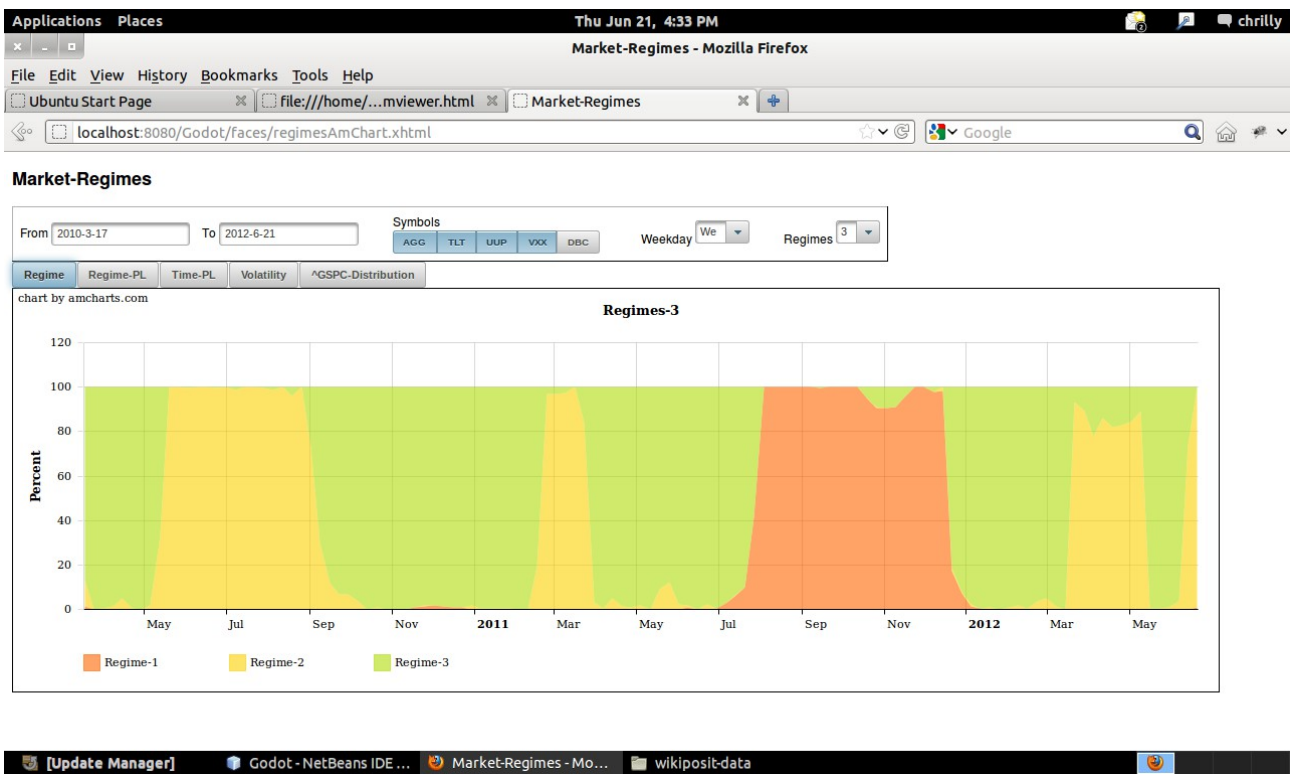
Graphic-1 shows a screenshot from the Godot Market-Regime homepage. The user can select the time-range for the regime-calculation. Per default its from 2010-03-17 till current. The start is an arbitrary date I have used also in other working papers. He can pick any combination of the ETF's listed above. The ^GSPC is included always in the model and can hence not (de-)selected. One can pick also any other weekdays for calculating the weekly return. The last entry sets the number of states/regimes (3 or 4).

The graphic is a stacked area chart. It shows for each time-step (week) the probability of the different states. E.g. in graphic-1 the orange area (regime 1) rises in July 2011 fast to 100%. This is the crash of summer 2011. But note, the regime-numbers are arbitrary labels. In a recalculation with somewhat different data regime-1 can be a rally. The best way to check the meaning of the regimes is to look at the volatility charts (see description of graphic 3) below.

The regime-classification is – of course – somewhat different for the 3-state model (graphic-2). But here again the orange area in summer 2011 marks the crash.

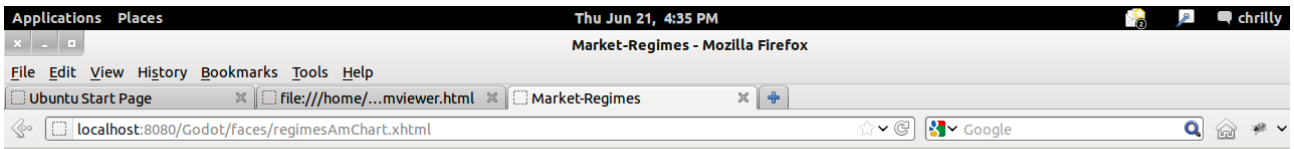


Graphic-1: Market-Regimes with 4 states

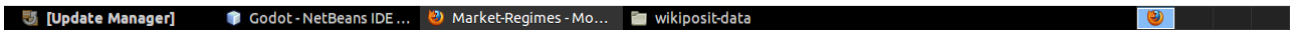
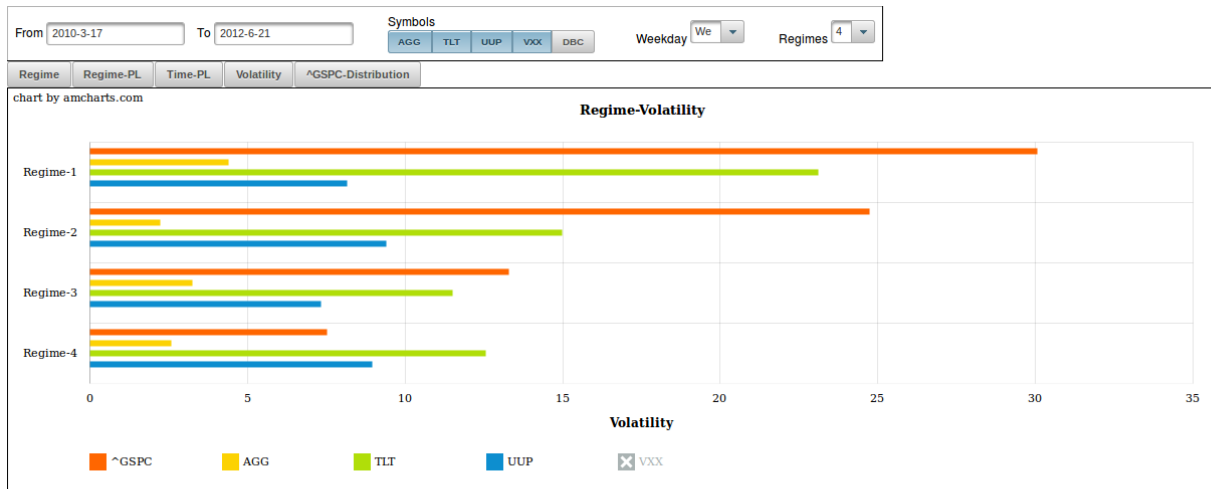


Graphic-2: Market-Regimes with 3 states

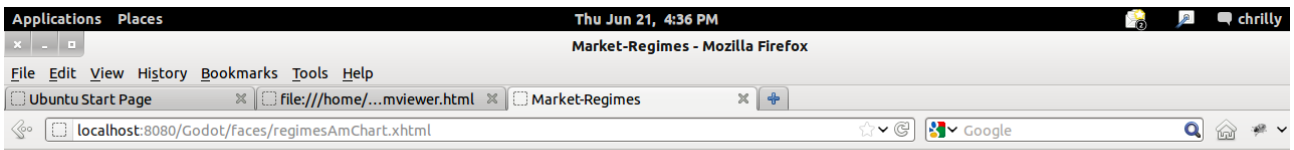
Note: One can (de-) select the different regimes in the charts by pressing at the legends at the bottom.  
 At the time of this writing (2012-06-21) (the market was in a mixture of regime 2 and 3 (4-states) or in regime-2 (3-states).



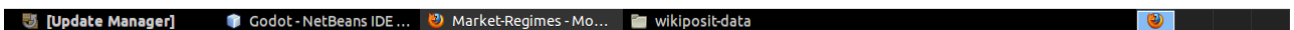
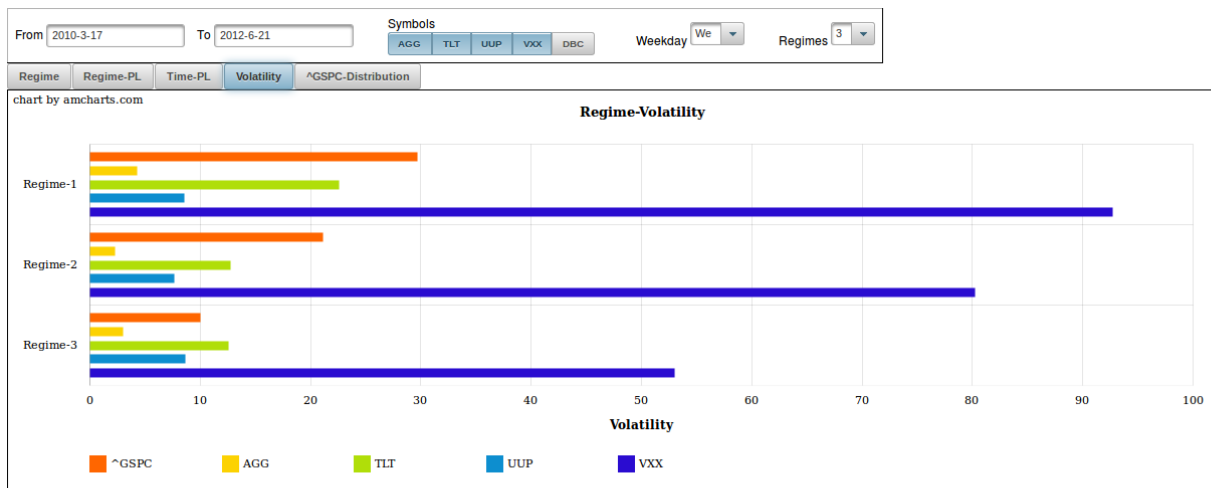
**Market-Regimes**



Graphic-3: Volatility of the ETF's for each regime (4-states).  
VXX was deselected, because the VIX-based volatility is much larger than the rest.



**Market-Regimes**

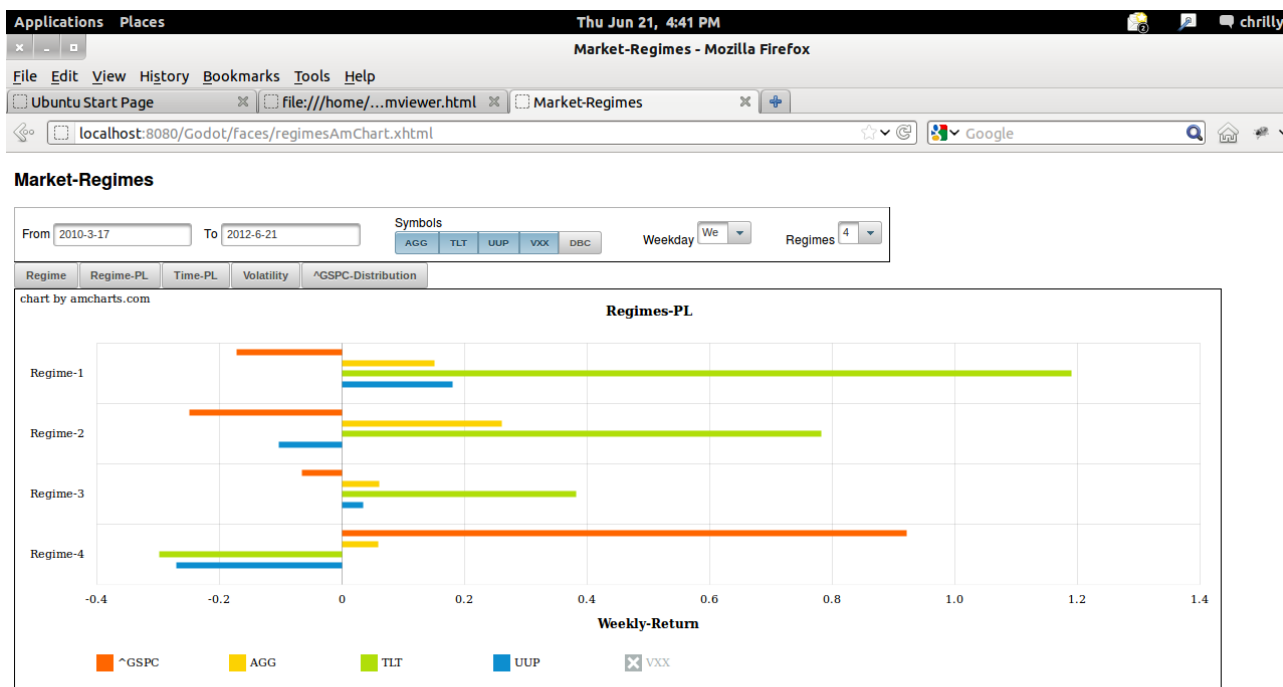


Graphic-4: Volatility of the ETF's for each regime (3-states).  
VXX was deselected, because the VIX-based volatility is much larger than the rest.

As can be seen in graphic-3 the volatility is highest in regime-1 and lowest in regime 4. In regime-1 we are in times of troubles. Volatility jumps up in a crash. Regime-4 is a rally. Regime 2 and 3 are in between. Graphic-4 shows a similar picture. Regime-1 is a crash with high volatility, regime-3 a

rally.

As noted already above, the regime-numbers are arbitrary. In another calculation regime-1 can be a rally, regime-2 the crash ... One always has to check the other graphics to identify the meaning of each regime. One could think about an automatic classification. But currently the meaning of the charts is in the eye of the user.

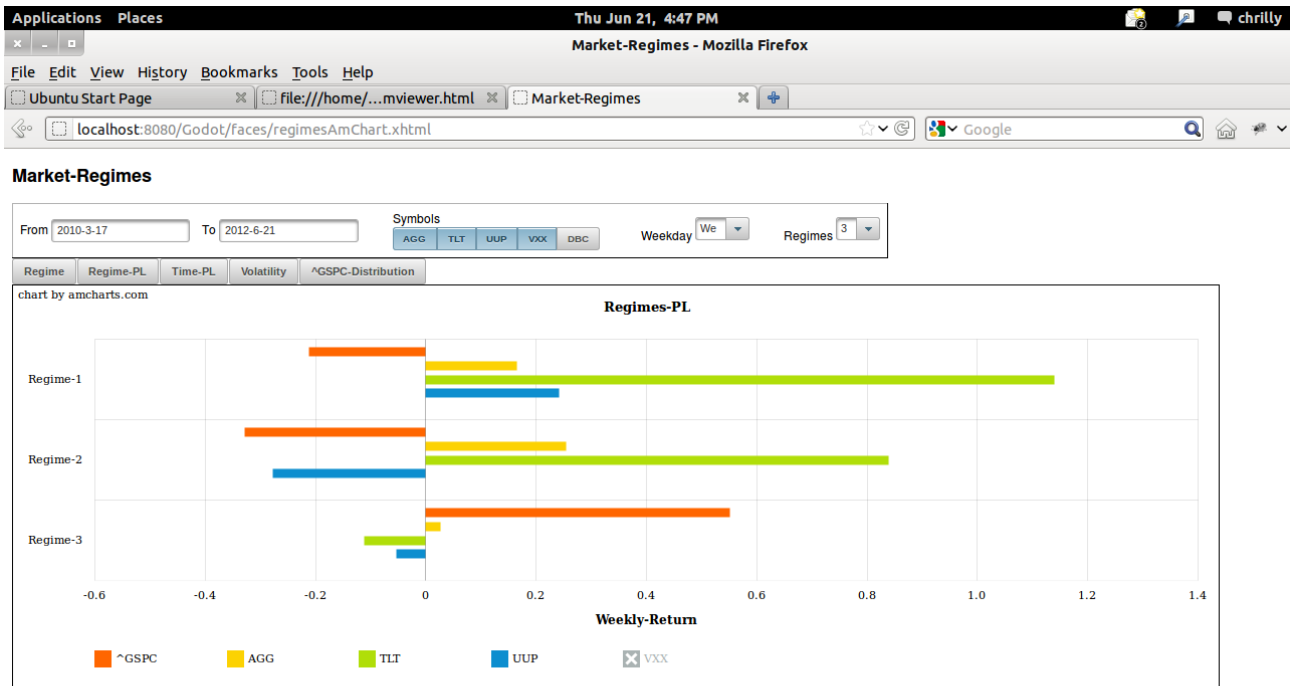


Graphic-5: Profit&Loss charts for the different regimes (VXX deselected). 4-States.

According to graphic-5 the S&P (orange) loses considerable in regime-1 and 2. There is a minor loss in regime-3 and a large expected win in regime-4. TLT (20 years Treasury) has the opposite pattern. TLT wins most in regime-1 and loses only in regime-4. In times of troubles (regime-1) investors consider US-Treasuries as a safe-heaven. AGG (Bonds) win – in the considered time range – in all 4 regimes. But the profit is generally less than for the other assets. UUP (US-\$) wins in regime-1 and 3, but loses in 2 and 4. The VXX (not shown) has exactly the inverse pattern of the S&P.

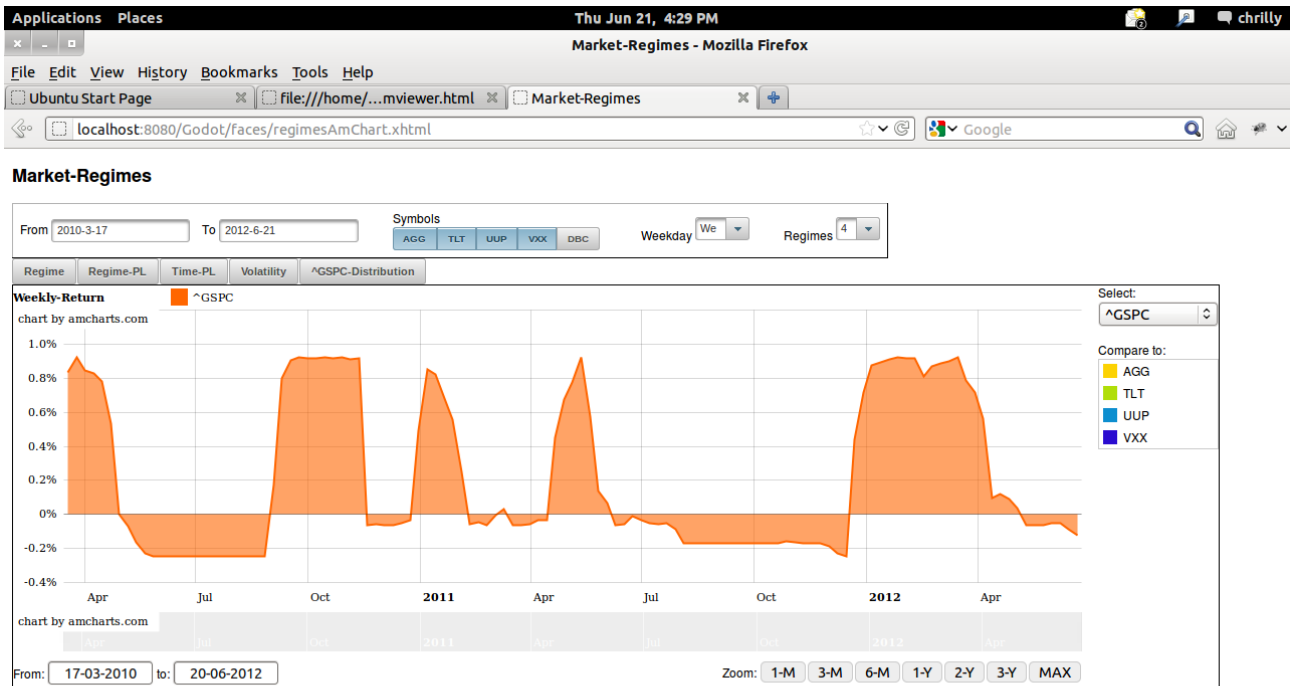
**The different behavior of the assets in different states is the main point of the model.** In regime-1 one would shift the investment into treasuries, in regime-4 in the S&P. At the time of this writing there is a mixture of regime-2 and 3 (see graphic-1). One would therefore choose TLT. In a more sophisticated approach one would form a portfolio of e.g. AGG and TLT. One could calculate also the regime-dependent performance of other assets which are not part of the hidden-markov-model and invest accordingly.

Graphic-6 shows the same chart for the 3-state model. According to graphic-2 we are currently in regime-2. The 3-state model prefers therefore also investments in TLT and AGG.



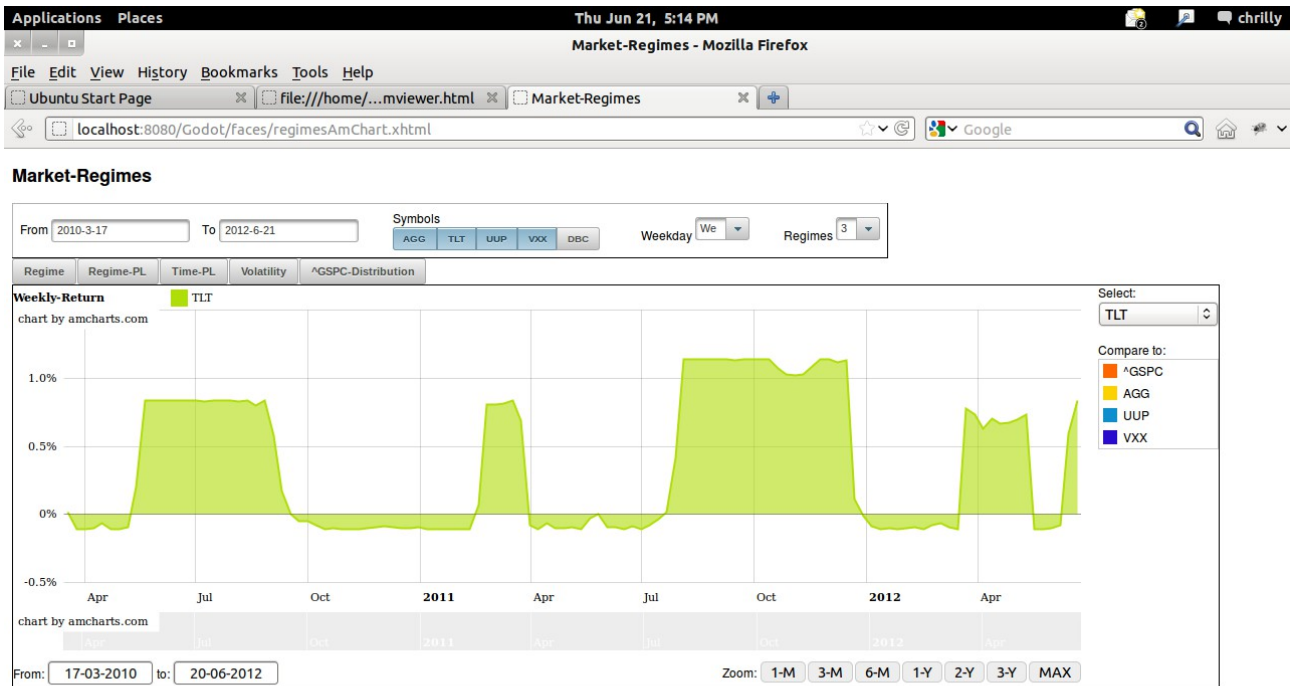
Graphic-6: Profit&Loss Charts for the different regimes (VXX deselected). 3-States.

Graphic-7 shows the expected performance of the ^GSPC along the time-dimension. There are considerable losses in may-july 2010 (flash-crash) and another strong drawdown in summer 2011. In spring 2012 the model shows a strong rally.



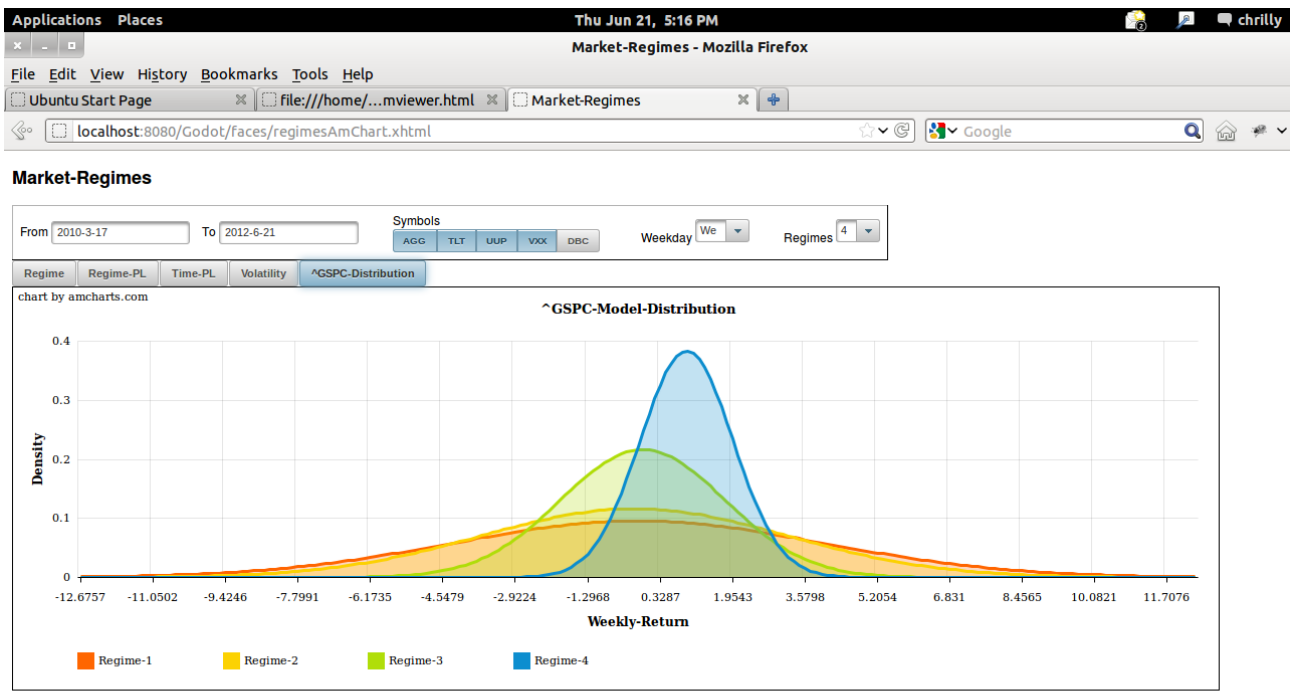
Graphic-7: Expected-Weekly Return of ^GSPC over time.

TLT shows the inverse behavior (graphic-8).



Graphic-8: Expected-Weekly Return of TLT over time.

Graphic-9 shows the distribution of  $\wedge$ GSPC for the different regimes. The blue area is the distribution of weekly-returns during a rally (regime-4). The peak is clearly on the positive side, variance/volatility is relative low. The red-line is the distribution of the  $\wedge$ GSPC regimes during the crash-regime. Note that is is not the actual, empirical distribution, but the theoretical model-distribution. The empirical distribution is not normal. One could replace the normal-distribution with the Student-t.



Graphic-10: Distribution of  $\wedge$ GSPC in different regimes.

This would not address the major shortcoming of the theoretical-distribution. Both distributions are symmetric. But it is well known, that the empirical distribution is skewed towards left (large-losses are more probable than large wins). Estimating the Baum-Welch algorithm with the empirical distribution is not possible, as the states are hidden. One can not directly measure the empirical distribution in a given state. Under the assumption of a normal distribution the problem reduces to the estimation of the parameters of the multivariate Normal-distribution. This problem can be solved.

As already noted the user can select any combination of the ETF's with ^GSPC. Every combination results in a different regime-classification. There is no clear-cut criterion which combination is best. One would have to make extensive simulations of different trading-strategies. The current default combination with ^GSPC, TLT, AGG, VIX and UUP (but without DBC) produced a reasonable looking result.

#### References:

- [1] A description of the Aroon-indicator can be found at [http://stockcharts.com/school/doku.php?id=chart\\_school:technical\\_indicators:aroon](http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:aroon)  
or  
<http://www.investopedia.com/terms/a/aroon.asp#axzz1yPvpwpqp>
- [2] Gumbel E.J.: Statistics of Extremes, Dover Books 2004.
- [3] Mulvey John M., Yonggan Zhao: An Investment Model via Regime-Switching Economic Indicators, Nov. 2010  
<http://www.ccf.org.cn/cicf2011/papers/20110109040345.pdf>
- [4] [http://en.wikipedia.org/wiki/Hidden\\_Markov\\_model](http://en.wikipedia.org/wiki/Hidden_Markov_model)
- [5] Press W.H, Teukolsky S.A., Vetterling W.T., Flannery B.P: Numerical Recipes, The Art of Scientific Computing, 3<sup>rd</sup> Ed. Chap. 16.3: Markov-Models and Hidden Markov Modeling.