Abstract:
This working paper uses as a starting point the filtered historical simulation (FHS) approach developed by Barone-Adesi et al. ([1],[2]). One builds a GJR-GARCH model and generates Monte-Carlo return/price paths with normalized returns. This introduces a severe drift-bias. The Stochastic Volatility Regime Simulation (SVRS) avoids the bias by sampling from the same volatility regime. As an alternative to GJR-GARCH an asymmetric HAR and a GARCH-VIX model is used. Path sampling is done in the same way. As a model free alternative a VIX based approach is additionally investigated. This alternative clearly beats the models during the pre- and post-Brexit market turmoil. Barone-Adesi et al. transform the real-world into the risk-neutral measure. The current model stays in the real-measure. One simulates a realistic trading behavior by hedging the options along the Monte-Carlo paths. One can calibrate the model by adding external noise.

The GARCH-Model:
The purpose of GARCH is to model volatility. Volatility is mean-reverting and persistent. Everybody in finance knows what volatility is. But nobody knows how to measure it. The traditional proxy is the squared daily return. But the squared-returns are not persistent. Even in a high-volatility regime there can be days with a low squared return. On the other hand there can be spikes in low volatility phases. One can interpret volatility-models as smoothers or filters of the proxy. As the true value of volatility is not known, there is no clear-cut criterion for selecting a model. This is a very fortunate state for the academic world. Thousands of papers have been written about the virtues of dozens of different GARCH variants. Following [1] I implemented the GJR-GARCH model of [4]. GJR is a shortcut of the authors (Glosten, Jagannathan, Runkle). This is one of the more popular GARCH models.

\[
\sigma^2(t) = a_0 + a_1 \sigma^2(t-1) + a_2 R^2(t) + a_3 I(t) R^2(t) \\
I(t) = (R(t)<0)
\]

\(a_0\) is a sort of ground-level. In combination with the other parameters it is also responsible for the mean-reverting behavior. \(a_1\) is the persistence or smoothing factor. The value is typically between 0.8 and 0.95. \(a_2\) models the impact of the squared daily return. This factor is typically negative. \(a_3\) is the impact of negative returns. The indicator \(I\) is zero for positive returns and 1 for negative. \(a_3\) is positive and larger than the absolute
value of a2. a2 and a3 model the different impact of negative and positive returns. The plain GARCH model has only the parameters a0, a1, a2. Negative and positive returns have the same effect on volatility. This is reasonable for FX, but not for stocks and stock-indexes like the S&P-500. There are several other approaches to model the asymmetric behavior of stock-indexes. GJR-GARCH is the simplest one.

The main drawback of GJR is: If a2 is negative (and a2 is in most cases negative), the expression on the left side can become negative. This is clearly impossible. As a simple fix the implementation clamps the value to the minimum of 0.5*VIX in the historic time window. The time window was for all reported results 1500 trading-days. The start value is set to 0.5*VIX on the corresponding day. But the first 30 days are not used for the GARCH parameter estimation. The factor 0.5 was chosen, because GARCH-volatility is usually somewhat smaller than the VIX. The VIX contains a risk-premium.

Note: To convert from VIX to \( \sigma \) one has to scale by \( \frac{1.0}{\sqrt{252.0} \times 100.0} \)

If there is no implied volatility-measure like the VIX available, one could instead use plain realized volatility. The purpose is to have a reasonable starting value and a lower-bound.

It's well known, that estimating GARCH parameters is a nasty numerical problem [5]. There are many combinations of the parameters which result in a similar filter behavior. Hence the ML-function has many local minima and is in the solution area very flat. Although the Amoeba optimization method of Press et al. [6] is rather robust, I had in a previous study [3] convergence problems (Amoeba is generally known as the downhill simplex method due to Nelder and Mead). Winker&Maringer solve in [5] the problem with their favorite method: Threshold-Accepting. My favorite is Differential-Evolution (DE). The number of individuals was set to 16. The number of generations to 200. DE generated always slightly better results than Amoeba and has no convergence problems. DE is – like every meta-heuristic – slower than Amoeba. But optimizing 4 parameters with a meta-heuristic is on modern processors no issue.

Graphic-1 shows the GJR-GARCH volatility and the VIX in the last 5 years from 2011-06-24 till 2016-06-29. To get the same scale \( \sigma \) is multiplied by \( \sqrt{252.0} \times 100.0 \). The general level of \( \sigma \) is lower than the VIX. But the GARCH volatility reacts faster and stronger. This can be best seen in the August 2011 crash (yellow line in Graphic-1). On 2011-08-08 \( \sigma \) was 66.54, the VIX at 48.0. The last values on the right of the chart are the effect of the Brexit. The GJR-GARCH volatility was on Thursday 23. June, before the result of the vote was known, with 9.13 rather low. It exploded to 30.04 at the close of Friday. The VIX closed on Thursday with 17.25 and went up to 25.76. During the day the VIX was higher, but the model captures only close to close movements. The S&P declined on Monday 27th by -1.8%. The VIX decreased nevertheless down to 23.85. The GJR-GARCH volatility increased to 31.55.

On the days before the Brexit the VIX and the SPX did not move according their usual pattern. The Realized Volatility was rather low, the SPX moved even up. But the VIX contained already a risk-premium and was higher. Generally markets and also betting shops did not anticipate the result of the vote. They expected a Bremain. The Brexit shows the problems of a realized volatility model. If the incertitude is not reflected in the
market prices of the underlying the model prices will be far too low. The correlation between the VIX and the GJR-GARCH volatility is over the whole time range 0.85.

Graphic-1: GJR-GARCH (yellow) and VIX (red) from 2011-06-24 till 2016-06-29

The asymmetric HAR-Model:
F. Corsi proposed in [7] an additive cascade model of realized volatility aggregated at different time horizons. This cascade of heterogeneous volatility components leads to a simple ARCH type model that considers volatility realized over different time horizons and is thus called *Heterogeneous Auto Regressive* (HAR) [8]. The model is much simpler to estimate and to interpret than the GARCH model. There are many variations on this model (see [8]). I used for this work an asymmetric extension.

\[
V_{t+1} = b_0 + b_1 V_{t}^{(1)} + b_2 I_t V_{t}^{(1)} + b_3 V_{t}^{(5)} + b_4 V_{t}^{(21)} \quad [1]
\]

- \( V_{t}^{(1)} \) is the volatility at time \( t \).
- \( I_t \) is 1 if the return at \( t \) is negative, otherwise 0. This is the asymmetric leverage effect.
- \( V_{t}^{(5)} \) is the mean of the volatility over the last 5 trading days.
- \( V_{t}^{(21)} \) is the mean of the volatility over the last 21 trading days.

Formula [1] is often also specified with logarithmic values. The current implementation avoids this transformation. Although the model is in the strict sense no long-memory model, it captures the long memory behavior of volatility reasonable well. The parameter \( b_1 \) is usually slightly negative and \( b_2 \) much larger and positive. A positive current return dampens volatility. A large negative return has a strong increasing effect. The HAR \( \sigma \) is usually closer to the VIX. There is not such a strong explosion during crashes. The volatility is in quiet times usually above the GJR-GARCH value.
The GARCH-VIX-Model:
This model is probably my own contribution to the GARCH zoo. The equation is the same than for the GJR-GARCH model but the parameters are tuned to minimize the absolute deviation of the model $\sigma$ to the current VIX. The persistence parameter $a_1$ is in the plain model around 0.86. It is for this variant around 0.96. The behavior of the model is much smoother. The peaks are never above the VIX, but the model volatility decreases also more slowly than the VIX. The correlation with the VIX is 0.92 (to 0.84 of the plain model).

Graphic-4 shows the different behavior of the 3 model sigmas. GJR is the most reactive, HAR is somewhat smoother. Especially it is more limited on the downside. The GJR-VIX approach is much smoother and – during quiet times – considerable higher than the other two measures. I tried also a HAR-VIX approach. But this did not give reasonable results in the following calculation stages. The behavior of the model sigma is too smooth. It does not differentiate well between the volatility regimes.
Filtered Historical Simulation:

Historical simulation is the simple idea: One samples from the past $N$ returns randomly and generates in this way Monte-Carlo paths. The method is – besides the window-length – parameter free. One avoids the nasty question about the statistical distribution of the returns. But this naive form destroys the volatility-clustering. One stacks returns from different volatility-regimes together. Filtered historical simulation addresses this point. One first normalizes the returns by dividing with the model-volatility. Normalized returns are also much closer to a normal-distribution than the returns itself. One could interpret the market behavior as a brownian motion with a time-varying temperature of the liquid. The invisible hand turns the Bunsen-burner on and off.

$$nR(t) = R(t)/\sigma(t)$$

For path generation one picks randomly a normalized return and multiplies with the current volatility. The generated return is plugged in the GJR-GARCH formula and a new $\sigma$ is calculated.

$$R(T) = nR(t) * \sigma(T)$$

There is one problem with this approach. The mean of the returns aka the drift is dependent on the volatility regime. If one picks in a high-volatility phase a normalized return from a low-volatility-regime, one has a systematic wrong drift. FHS generates in high-volatility-regimes in the mean a strong bull market. One can argue that the drift is anyway hedged away. This is – in a realistic setting – only partly true. But another shortcoming of FHS creates even in a complete market a significant bias. The frequency of historic positive returns is almost always higher than the negative ones. The market moves up in many small steps and goes down in a few larger ones. Hence FHS samples in high-volatility-regimes the wrong ratio of up- and down-moves. GJR-GARCH reacts by construction different to negative and positive returns. The combination of FHS and GJR-GARCH is in contrast to the claims in [1] and [2] not well suited. The situation is in principle the same for an asymmetric HAR model.
Stochastic Volatility Regime Simulation

Stochastic Volatility Regime Simulation (SVRS) tackles the problems of FHS by sampling from the same historic regime. One defines 3 regimes. One simply sorts the entries according the GARCH, HAR or GARCH-VIX volatility. The sampling process is independent from the model generation process. The lowest 45% form the quiet-market regime. The entries from 45-90% form the mid-volatility regime and the top 10% the times of troubles.

One compares the current volatility along a simulation path with the regime boundaries and samples randomly the returns in the corresponding regime-bucket. One calculates from the return the next model volatility and repeats the cycle till the simulation horizon is reached. One gets alongside each path the model-volatility and the summed up historic returns.

There is one problem with this approach. If e.g. the current volatility corresponds to the 44% entry the simulated return will be likely either small or positive. The transition probability to the mid-regime is rather low. If the current value is a little bit above the threshold the generated path will look in most cases quite different. The starting conditions will have a significant impact on the final distribution. The same happens of course also along a simulation path. But the effect is most pronounced at the root of the simulation tree. Besides this the behavior is also in real market live (strongly) influenced by external shocks like political turmoil, natural disasters, FOMC decisions....

One adds therefore to the current volatility a noise term and looks up this value as described above. This noise term could also be described as a regime-transition probability. But the transition probability depends also on the current distance to the boundary. If the 45% threshold is at 17.0 and the current value is 16.5% a relative small external shock can trigger the regime transition. If it is at 12.0%, the external shock must be quite large.

The external shock is modeled as an exponential distribution. The choice of this distribution was purely pragmatic. The noise-term is always positive. This is also for real markets a more realistic behavior than a shock that reduces current volatility (although sometimes the “nice” words of Mrs. Yellen have a calmative effect). The exponential distribution has fat tails and last but not least it is trivial to generate an exponential distributed variable with the parameter lambda > 0 in Java (or any other programming language) with.

\[ E = \frac{-\text{Math.log}(1.0 - \text{Random.nextDouble()})}{\text{VOLA_FAC*lambda}}. \]

Note: One has to use the expression \((1.0 - \text{Random.nextDouble()})\) and NOT nextDouble(). The method nextDouble() returns a value in the half-open interval \([0,1)\). Subtracting the random value from 1.0 avoids a NaN exception for Math.log(0).

The added noise has no effect as long as the volatility stays below the regime-threshold. One samples in the same way from the same regime. It is an indirect effect if the regime-boundary is crossed. The return distribution differs between regimes. But it is only the sampled return which influences the further path. The noise E has no direct effect on the model volatility. The effect of the noise E is persistent if the model volatility crosses due to a large (negative) return the regime boundary. The effect is larger for the GARCH model. GARCH is more sensitive to large negative returns than the HAR or GARCH-
VIX model (see Graphic-4). Lambda can be interpreted as a calibration parameter. The standard deviation of the return-distribution at the horizon increases with a smaller lambda. The distribution gets also more left-skewed. One could calibrate with lambda the model value to the market price of the option. A small lambda means the option is expensive, a large one the option is cheap. But it should be noted that the calibration works only within a limited range.

At of this writing (2016-06-30) the ESU6 Future is at 2060, the VIX at 16.4. The market is in the recovery phase after the Brexit turmoil. Graphic 5 shows the final distribution of the S&P ESU6 Future at 2016-07-15 (The expiry of the N6 options). The GARCH distribution is quite different from the other two models. The GARCH $\sigma$ has exploded in the Brexit crash. The value is still high and the simulation starts in the high-volatility regime. Some of the simulation paths move down in the mid- or low-regime. But a considerable amount stays also in the high-volatility regime. The other two models are already back in the mid-regime and the distribution is much more centered. The expected return is about zero, whereas the GARCH model expects – in the mean – a serious market drop. Graphic-6 shows the situation for the Q6 expiry (2016-08-19). The GARCH model is now clearly bimodal. The HAR and GARCH-VIX model are in agreement with the market which has already digested the Brexit.
Options Price Calculation:
The SVRS generates paths with the volatility and return at each daily time-step. The options values are calculated by delta-hedging along the path. Delta is calculated with the model volatility. The final value is the payoff at the horizon/expiry plus the P&L of the hedging activity. The model price is the mean of the sampled values. But one gets additionally the full payoff distribution. No attempt is made to transfer this distribution into a risk-neutral measure. After all trading is also done in the real measure and one (or the trading-algo) has to decide if there is enough fun for the risk.

Graphic-7 shows the market (red) and model-prices of the EW3N6 P1950 Option (expiry 2016-07) from 2016-05-12 till 2016-07-01. The calculation is done with the prices at 20:00 UTC (15:00 CST). Lambda is 0.75 for the GJR and 0.5 for the HAR and GJR-VIX model. This gives the best overall fit. The P1950 is an – most of the time – OTM Put. Especially GJR is up to 2016-06-12 quite close to the market price. But then all the model prices are considerable lower. At this time the first pro-Brexit opinion polls were published. This was reflected in an increasing VIX. But it was not reflected in the realized volatility of the S&P. The market mood was “Brexit danger ahead”. The models were in a quiet regime. After the surprising result of the vote the realized volatility become very high. The fast upwards moving GJR σ was much higher than the VIX. The model values are therefore far off. The effect is less pronounced for the GJR-VIX model. The HAR model is during the Brexit crash and in the following recovery in relative good agreement with the market prices.

Graphic-7: EW3N6 P1950 Market (red), GJR (yellow), HAR(blue), GJR-VIX (green)

Graphic-8 shows the performance for the EW3N6 P2000 option. The overall picture is similar. The GJR and GJR-VIX model the market-prices till mid June quite well. The HAR price is somewhat too high. All the models underestimate like for P1950 the prices in the days before the Brexit and overestimate them after wards. HAR does in this phase a better job, but the agreement is not as good as for the P1950 option. The P2000 moves in this phase from OTM to (almost) ATM.

Graphic-8 shows the performance for the EW3N6 P2000 option. The overall picture is similar. The GJR and GJR-VIX model the market-prices till mid June quite well. The HAR price is somewhat too high. All the models underestimate like for P1950 the prices in the days before the Brexit and overestimate them after wards. HAR does in this phase a better job, but the agreement is not as good as for the P1950 option. The P2000 moves in this phase from OTM to (almost) ATM.
Graphic-8: EW3N6 P2000 Market (red), GJR (yellow), HAR (blue), GJR-VIX (green)

Graphic-9 shows the behavior of the EW3Q6 P1950 (expiry 2016-08-19) option. The lambdas are 0.5 for GJR and HAR, and 0.4 for GJR-VIX. The overall pattern is similar to the EW3N6 case. The evaluation is till mid June quite well, underestimates the pre-Brexit phase. The GJR and GJR-VIX model strongly overestimate the post-Brexit position. The HAR model handles this phase quite well. The same pattern holds for all Puts at the two expires.

Graphic-9: EW3Q6 P1950 Market (red), GJR (yellow), HAR (blue), GJR-VIX (green)

Graphic-10 shows the calculation of the EW3N6 C2100 Call. Lambda was set for GJR and GJR-VIX to 1.0, for HAR to 10.0 (this almost switches the external noise off). The evaluation is for GJR- and GJR-VIX for the first half quite well. The larger lambda is consistent with the fact that Calls have a lower IV than Puts. The evaluation is again for the pre-Brexit phase too low. In the post Brexit phase the GJR value is again much too high, the GJR-VIX evaluation is somewhat better, but still too high. There is no way to tune the HAR evaluation in the first half down to the market value. The evaluation is simply too high. The pre-Brexit phase is for the same reason okay. The HAR is in the post-Brexit market regime also too high.

Graphic-10: EW3N6 C2100 Market (red), GJR (yellow), HAR (blue), GJR-VIX (green)
Graphic-10: EW3N6 C2100 Market (red), GJR (yellow), HAR(blue), GJR-VIX (green)

Graphic-11 shows the evaluation for the EW3Q6 C2100 option. Lambda was set to 0.75 for GJR and GJR-VIX. The smaller lambda (larger noise) is consistent with the result above for the EW3Q6 Puts. For the HAR-model lambda was set to 10.0. The result is similar to the EW3N6 Call.

Graphic-11: EW3Q6 C2100 Market (red), GJR (yellow), HAR(blue), GJR-VIX (green)

**A Model-Free VIX-Model:**
The GJR- and GJR-VIX model generate reasonable prices for different moneyness and different expiry under normal market conditions. There is no chance for a model based on the realized volatility of the underlying to model the fear for an upcoming event if this fear is not directly reflected in the market prices. As an alternative I tried a model free VIX model. The market regime is defined by the Implied Volatility Term Structure (IVTS). The IVTS was used in previous papers as a market-timing signal [9]. It is also implemented and used with good success in my fully autonomous trading system CashBot. The IVTS is the 1-month VIX divided by the 3-month VXV.

**IVTS= VIX/VXV.**
Both volatility indexes are calculated during the regular trading hours by the CBOE in real time (VIX is now even available from 2:15 CST). It was shown in [9] and in [10] that the IVTS is well suited for regime-classification. It is superior to the stand alone VIX. After a crash the VIX is usually still high, but the IVTS signals already a recovery. The IVTS usually signals also market-danger ahead of the VIX. The problems are related. If one wants to capture the large wins in the recovery phase, one has to set the threshold relative high. But this means that one closes at the formation period of a crash the position too late. If one sets the threshold lower to avoid this problem, one misses the recovery wins. The IVTS solves this problem much better. Graphic-12 shows the IVTS (red) and the VIX (yellow) in the last 5 years. The values were scaled to 100 for direct comparisons. The correlation between the two series is only 0.60.

Following earlier IVTS models and papers there four regimes are defined:
Low with IVTS < 0.91, mid with IVTS < 0.97, high with IVTS < 1.03 and very high with IVTS >= 1.03.

One classifies the entries in the historic window according these thresholds. But the classification is overlapping with a margin of 0.01. All the entries in the low-regime have an IVTS up to 0.91+0.01. The entries in the mid-regime have an IVTS between 0.91-0.01 and 0.97+0.01 and so on. This avoids sharp regime-boundary effects. One samples along the simulation path from the appropriate regime. The return on the next day is used as the simulation value. One calculates the new IVTS from the closing values of VIX and VXV at t+1. The simulation path jumps from one IVTS value of the historic window to the next. As before one adds an exponential noise to the current IVTS

\[ E = -\text{Math.log}(1.0-\text{Random.nextDouble()})/(100.0*\text{lambda}). \]

The options calculation is done like before. The current VIX is used as the volatility parameter for calculating the hedging delta.

Graphic-13 compares the performance of the IVTS model with a lambda of 1.5 for the EW3N6 P1950 Option with the best HAR model of Graphic-7. The model estimates are in the pre-Brexit phase still somewhat too low. But the behavior after the Brexit is quite well. The simulation starts in the correct market regime. The overall fit is clearly superior.
The same holds for the EW3N6 P2000 option. Lambda is set to 2.0. The fit is especially during the Brexit turmoil and the following recovery remarkable good.
The fit is for the August 2016 Option EW3Q6 P1950 during the Brexit phase not so well. But the VIX model is overall superior to the HAR model (Graphic-15).

For the EW3N6 C2100 Option the VIX evaluation is in the first half too high. But the fit is in the pre-Brexit and the post-Brexit phase almost perfect. Lambda is set to 1000.0. This means the random noise is practically switched off. The HAR model has a similar problem. It also evaluates the (ATM-) Calls too high. The GJR-VIX model does a better job in the first half, but is by far worse in the pre- and post-Brexit phase.

Conclusion:
The GJR- and GJR-VIX model generate reasonable prices for different moneyness and different expiry under normal market conditions. There is no chance for a model based on the realized volatility of the underlying to model the fear for an upcoming event if this
fear is not directly reflected in the market prices. There is considerable potential for improvement in the recovery phase of the crash. The model values are plain wrong.

The model free VIX/IVTS model is overall the clear winner. One could argue that this model is against the spirit of the general approach: A benchmark for answering the question if options are cheap or expensive. After all the model starts with the current VIX. But the VIX model answers additional questions: One gets the full final distribution and can hence also calculate the risk. Another application is the evaluation of more sophisticated trading strategies. E.g. evaluate the value of a calendar spread or other strategies with stop-loss rules. The VIX model has according the current results the most realistic and especially the most robust behavior of all models. The model can – with some modifications – extended to other assets than the SPX. The CBOE calculates nowadays for a host of volatility indexes. One can also handcraft an own index from the option prices.

Further Work:
The most promising direction for further work seems to be the VIX/IVTS model. Currently the VIX is used as the input for calculating the hedge-delta. OTM calls are overhedged, far OTM Puts underhedged. One can incorporate the smile into the model. Another interesting direction is the evaluation of more sophisticated trading strategies. As already noted are calendar spreads an interesting application. For this kind of applications it is critical to have besides a realistic price distribution of underlying also a reasonable model value for the implied volatility at hand.

References: