Volatility Modelling and Trading

Artur Sepp

Julius Baer

artur.sepp@gmail.com

Global Derivatives Workshop
Global Derivatives Trading & Risk Management 2016
Budapest
May 13, 2016

Contents

Electronic copy available at: http://ssrn.com/abstract=2810768

1 Introduction: Goals and Scope

Goals of the workshop:

- To provide a practical and technical overview of **volatility trading strategies**:
- Focus on **systematic and rule-based trading strategies** that can be marketed as an invest-able index or a proprietary strategy:
  1. Delta-hedged strategies for capturing the volatility and skew risk-premiums
  2. Without delta-hedge: CBOE and customized options buy-write indices
- Overview important implementation aspects:
  1. Measuring the historic realized volatility
  2. Forecasting the expected realized volatility
  3. Measuring and forecasting implied and realized skew
  4. Computing option delta consistently with empirical dynamics
  5. Analysis of transaction costs
  6. Managing the tail-risk of short volatility strategies

Target Know-how:

- Investment community and asset allocators are increasingly focusing on factor-based approaches for asset allocation
- Volatility-related strategies are typically identified as alternative risk-premium strategies which enjoy growing interest from institutional and high net worth clients
- Increased offering of volatility based strategies:
  1. CBOE expanded its toolkit of buy-write indices to about 25 (these indices are not invest-able yet)
  2. Sell-side investment banks are offering standardizes and customized trading strategies for clients
- The workshop intends to provide:
  1. **The insight for the design and back-testing of systematic volatility strategies**
  2. **Understanding of risk-reward trade-off and potential pitfalls of volatility strategies**
- One is the key topics is how to measure and forecast the realized volatility: a proper method can substantially improve the performance of volatility strategies

Key characteristics of volatility strategies discussed in this workshop:

- The focus is on systematic strategies managed quantitatively with a minimum of discretionary decisions
- Trading frequency: medium to low with typical re-balancing of one month
- Cash management:
  1. Buy-write indices: cash funded, the size of position is set so that the fund value is sufficient to cover exercisers of traded positions
  2. Delta-hedged strategies: leverage is possible, delta risk may be unlimited
- Trading universe:
  1. Options on major stock indices and bonds using futures options
  2. Options on major ETFs and large cap stocks
- Key trade-offs:
  1. Liquidity of the underlying asset (limited apart from major ETFs: SPY, QQQ, IWM and sector ETFs, and some large cap stocks)
  2. Hedging delta risk versus transaction costs which a very important trade-off: take more delta risk and save on transaction costs or take less delta risk and spend on hedging

Remarks for the workshop presentation

- The workshop notes are sometimes technical and reasonably self-contained so they can used as a reference in your own work
- During the workshop, the emphasis will be on illustrations and discussion of key concept
- A lot of the presented material is both current and "perpetual" work-in-progress
- Your feedback and comments is welcome during the workshop and afterwords

2 Realized Volatility

2.1 Return Calculations

We work with the time series data of asset prices observed at the end of a period:

$$[S_0, S_1, S_2, .., S_t, .., S_T] \tag{2.1}$$

where t is an integer indicating the t-th element of the sample, $t = 0, .., T$, $T + 1$ is the size of the sample.

Typically, we use daily observations

Given the sample of $T + 1$ prices corresponding to dates $(d_0, d_1, ..., , d_T)$, we compute the one-period return at time t as follows:

- Absolute return as the spread:

$$r_t^{(s)} = S_t - S_{t-1} \ , \ t = 1, ..., T \tag{2.2}$$

- Logarithmic return:

$$r_t^{(l)} = \ln\left(\frac{S_t}{S_{t-1}}\right) \ , \ t = 1, ..., T \tag{2.3}$$

- Arithmetic return:

$$r_t^{(a)} = \frac{S_t}{S_{t-1}} - 1 \ , \ t = 1, ..., T \tag{2.4}$$

Special Considerations:

- Portfolio returns: when using the arithmetic returns, the portfolio return equals the weighted sum of returns on its components
- Heavy tailed distributions: the exponential moments (necessary for the existence of finite moments of log-returns) are not defined (infinite) for heavy-tailed t-distribution
- My convention is to use arithmetic returns for volatility estimations

2.1.1 Dividends

When a stock or an ETF go ex-dividend at the beginning of period t_k, the arithmetic return is computed by:

$$r_{t_k}^{(a)} = \frac{S_{t_k} + D_{t_k}}{S_{t_k - 1}} - 1 \tag{2.5}$$

The logarithmic and absolute returns are computed by analogy

Special Considerations:

1. Dividends are an important component of total returns
2. All trading and investment strategies should account for dividends and specify whether they are
    - total return strategies with accrued dividends being re-invested
    - dividend distribution strategies with accrued dividends being distributed

2.1.2 Stock Splits

It is not uncommon that the shares of stocks and ETFs have splits

Especially so in pre crisis of 2008 - the emerging market ETF (EEM) had two 3 : 1 splits in 2005 and early 2008

Apple had 7 : 1 split in 2014

The stock split with the ratio of $x : y$ (most common is 2 : 1) means that the holder gets x new shares for y existing shares

When the split is applied at the beginning of period t_s, the return is computed by considering the change in the total position

$$\text{at the end of time } t_s\text{: } xS_{t_s} \text{ vs at the end of time } t_{s-1}\ yS_{t_s-1} \tag{2.6}$$

The arithmetic return is computed by:

$$r_{t_s}^{(a)} = \frac{xS_{t_s}}{yS_{t_s-1}} - 1 \tag{2.7}$$

The logarithmic and absolute returns are computed by analogy.

To deal with splits we can apply the following options

- Create event handler which is provided with split dates and which applies equation (2.7) for return computation: this is more demanding computationally
- Adjust time series post the split date d_{t_s} by dividing by the ratio $x : y$

Special Considerations:

- Data providers also publish adjusted closing prices taking into account dividends and splits.

- Adjusted prices are not aligned with the intra-day prices and are not actual prices for trade executions in the back-test.
- The only possible use of adjusted prices is to verify the internal computations for return computations which accounts for all possible corporate actions.
- Some data providers report dividends on post-split share count (for an example, \$1 dividend paid pre-2 : 1 split is reported as \$0.5)
- The best is to handle stock splits and spin-offs case by case in your internal database

2.2 Sample partitioning

For back-testing of trading strategies and analysis of historical data, we need to partition the existing time series data into sub-samples:

1. Specify dates $\{d_r\}$ when the strategy is rebalanced
2. Specify period prior to each date d_r when input parameters and signals (volatilities) for this strategy are computed
3. Incorporate Markovian property for the trading strategy by using only the information up to date d_r to evaluate inputs for the strategy

Given the time series of prices with the sample of size T, we partition the sample into N sub-samples as follows:

$$\begin{aligned}
&[S_0, S_1, S_2, ..., S_T] \longrightarrow \\
&\left[S_0, S_1, ...S_{T_{K_1}}\right] \\
&\left[S_{T_{K_1}}, S_{T_{K_1}+1}, ...S_{T_{K_2}}\right] \\
&... \\
&\left[S_{T_{K_{N-1}}}, S_{T_{K_{N-1}}+1}, ...S_{T_{K_N}}\right]
\end{aligned} \tag{2.8}$$

Here:

- T_{K_n}, $n = 1..N$ is the strategy re-balancing times, $T_{K_{n-1}}$ is the start time of time series for estimation of the inputs for re-balancing at time T_{K_n}
- The strategy is using data from the previous re-balancing $T_{K_{n-1}}$ up to current re-balancing at T_{K_n}
- The strategy may use longer or shorter periods for estimation of the inputs so that the start date of the estimation period $T_{K_{n-1}}$ needs to be adjusted accordingly

Table 1: Example of Start and End date for strategy with re-balancing at every 3rd Friday of the month and with generation of signal using past month of data

Start	End
17-Jul-15	21-Aug-15
21-Aug-15	18-Sep-15
18-Sep-15	16-Oct-15
16-Oct-15	20-Nov-15
20-Nov-15	18-Dec-15
18-Dec-15	15-Jan-16
15-Jan-16	19-Feb-16
19-Feb-16	18-Mar-16

2.3 Realized Volatility

Given sample of time series of end of day stock prices

$$[S_0, S_1, S_2, ..., S_T] \tag{2.9}$$

and dividends, we obtain the time series of price returns:

$$[r_1, r_2, ..., r_T] \tag{2.10}$$

Fixing the current date by t, we form the sample of prior observations up to time $t - T$ by:

$$[r_{t-T}, ..., r_{t-2}, r_{t-1}] \tag{2.11}$$

In this interpretation:

- r_{t-1} is the most recent return for prior day $t-1$ observed at the beginning of trading day t
- r_{t-T} is the most distant return
- the sample contains T prior returns.

Simple close-to-close estimator of realized variance :

$$\sigma_t^2(T) = \frac{1}{T-1}\sum_{k=1}^{T}\left(r_{t-k} - \overline{r}_t(T)\right)^2 \tag{2.12}$$

where $\overline{r}_t(T)$ is the average sample return

By convention, $\sigma_t(T)$ and $\overline{r}_t(T)$ denote the sample estimates of volatility and average return computed using the returns data at the beginning of period t with the sample containing returns from period $t-1$ to $t-T$

The factor $\frac{1}{T-1}$ is the correction with adjustment for the mean

2.3.1 Annualization

The volatility computed using daily returns and normalized as in Eq (2.12) represents the daily volatility of returns

We annualize the volatility by applying the annualization factor:

$$\widetilde{\sigma}_t^2(T) = AF \times \sigma_t^2(T) \tag{2.13}$$

where $AF = 252$

Special Considerations:

- High frequency (using minute- or hourly returns) and lower frequency (using weekly or monthly returns) volatility estimates can be applied

- It is always robust to report the volatility in the same annualized units to avoid confusion

2.3.2 Volatility vs Variance

It is always more intuitive and clear to report computations in terms of annualized volatilities. It also helps to spot inaccuracies in data and computations

Reporting in terms of variances or (worst) daily variances is confusing - an example is given in the table

Table 2: Volatility vs Variance

Annualized Volatility	5%	10%	15%
Annualized Variance	0.25%	1.00%	2.25%
Daily Variance	0.0010%	0.0040%	0.0089%

2.4 Discrete vs Continuous Volatility and Estimation Methods

In the econometric literature, we following concepts are studied:

- Instantaneous variance: the stochastic driver, denoted by $V(t)$, of instantaneous variance of price returns

$$\mathbb{V}ar[r_t] = V(t)dt \tag{2.14}$$

- Quadratic Variance: the integrated instantaneous variance

$$QV(t) = \int_0^t V(t')dt' \tag{2.15}$$

- Realized Variance: the discrete approximation of the quadratic variance computed over discrete time grid $\{t_k\}$

$$DV(t) = \sum_{t_k \in [0,t]} r_{t_k}^2 \tag{2.16}$$

Volatility estimation and forecasting methods:

1. **Estimation Methods using Sample State**: simple close-to-close estimator, estimators using intra-day prices:
   - Apply the concept of the realized variance using the returns observed in the sample
   - Forecasting is based on flat extrapolation of the most recent volatility into the future - no account for the volatility mean-reversion
2. **Estimation Methods using State Inference (GARCH type methods)**:
   - Work with the concept of the instantaneous variance which is inferred using a state equation

- Apply recursive maximum likelihood method for estimation of model parameters
- Forecasting is based on extrapolation of the most recent volatility into the future accounting for the volatility mean-reversion

3. **Continuous time SV models**:
   - Work with the concept of the instantaneous variance which can be inferred using a state equation or proxied using a short-term realized variance
   - Apply recursive maximum likelihood method for estimation of model parameters
   - Forecasting is based on extrapolation of the most recent volatility into the future accounting for the volatility mean-reversion

3 Volatility Estimation using Sample State Methods

These methods assume the price process driven by a diffusion process and provide inference using the discrete variance

Given the observed sample state of asset prices, they provide an estimate of the volatility

The first two sets of estimators do not account for auto-correlation of volatility

3.1 Close-to-Close Estimator

$$\sigma_t^2(T) = \frac{1}{T-1} \sum_{k=1}^{T} \left(r_{t-k} - \overline{r}_t(T)\right)^2 \tag{3.1}$$

where $\overline{r}_t(T)$ is the average sample return.

3.2 Open-High-Low-Close Estimators

These methods employ intra-day observations and need the time series of daily observations for:

1. open price O_t
2. high price H_t
3. low price L_t
4. close price C_t

3.2.1 Parkinson estimator

Uses the intra-day high and low prices:

$$\sigma_t^2(T) = \frac{1}{4\ln(2)} \sum_{k=1}^{T} \left(\ln\left(\frac{H_k}{L_k}\right)\right)^2 \tag{3.2}$$

3.2.2 Garman-Klass estimator

Extension of Parkinson estimator including opening and closing prices:

$$\sigma_t^2(T) = \sum_{k=1}^{T} \left[\frac{1}{2}\left(\ln\left(\frac{H_k}{L_k}\right)\right)^2 - (2\ln(2)-1)\left(\ln\left(\frac{C_k}{O_k}\right)\right)^2\right] \tag{3.3}$$

3.2.3 Rogers-Satchell estimator

No assumption about zero-mean:

$$\sigma_t^2(T) = \sum_{k=1}^{T} \left[\ln\left(\frac{H_k}{C_k}\right)\ln\left(\frac{H_k}{O_k}\right) + \ln\left(\frac{L_k}{C_k}\right)\ln\left(\frac{L_k}{O_k}\right)\right] \tag{3.4}$$

3.2.4 Yang-Zhang estimator

Is compatible with the drift and overnight jumps:

$$\sigma_t^2(T) = \sigma^2_{\text{overnight}}(T) + c\sigma^2_{\text{open-to-close}}(T) + (1-c)\sigma^2_{\text{Rogers-Satchell}}(T) \tag{3.5}$$

where

$$\sigma^2_{\text{overnight}}(T) = \sum_{k=1}^{T} \left(\ln\left(\frac{O_k}{C_{k-1}}\right) - \overline{\ln\left(\frac{O_k}{C_{k-1}}\right)} \right)^2$$

$$\sigma^2_{\text{open-to-close}}(T) = \sum_{k=1}^{T} \left(\ln\left(\frac{C_k}{O_k}\right) - \overline{\ln\left(\frac{C_k}{O_k}\right)} \right)^2$$

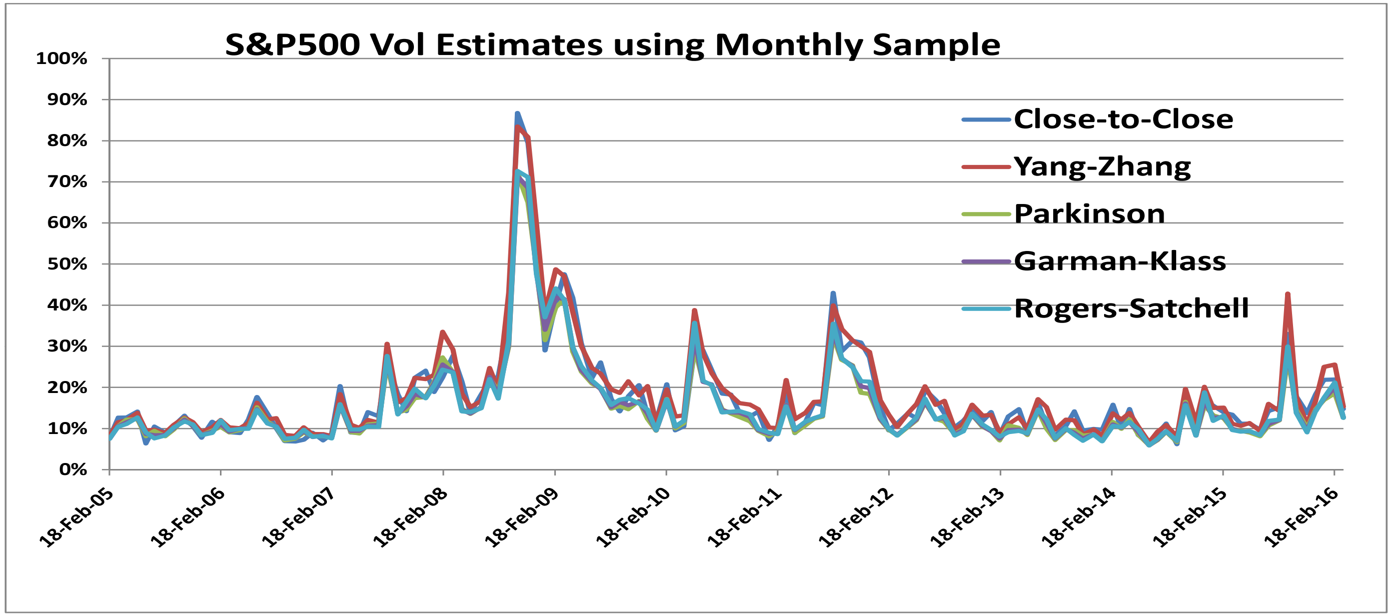
$$c = \frac{0.34}{1.34 + \frac{T+1}{T-1}}$$

Illustrations: Different Estimators

1) Time series of volatility estimates using monthly sample for S&P500 index ETF (SPY) from 2005 to 2016

Key observations:

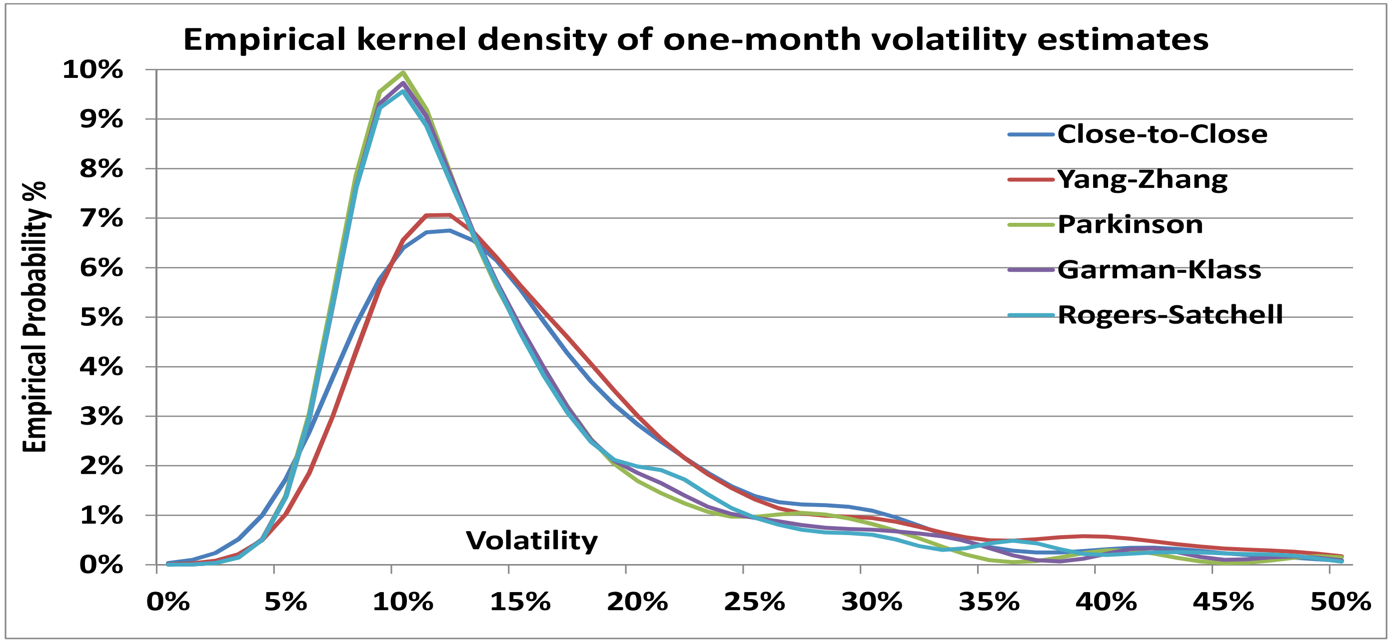
- Close-to-Close and Yang-Zhang employ overnight returns and so provide higher estimates of volatility when the overall level of volatility is high (GFC in late 2008, EU stress in 2010 and 2011)



2) Empirical kernel density of one-month volatility estimates using monthly sample for S&P500 index ETF (SPY) from 2005 to 2016

Key observations:

- Parkinson, Garman-Klass, and Rogers-Satchell do not employ overnight returns so their estimates of the volatility are smaller and less volatility
- For asset allocation programs, Parkinson, Garman-Klass, and Rogers-Satchell provide a more robust way to estimate the volatility and compute allocation weights which are more stable between re-balancings

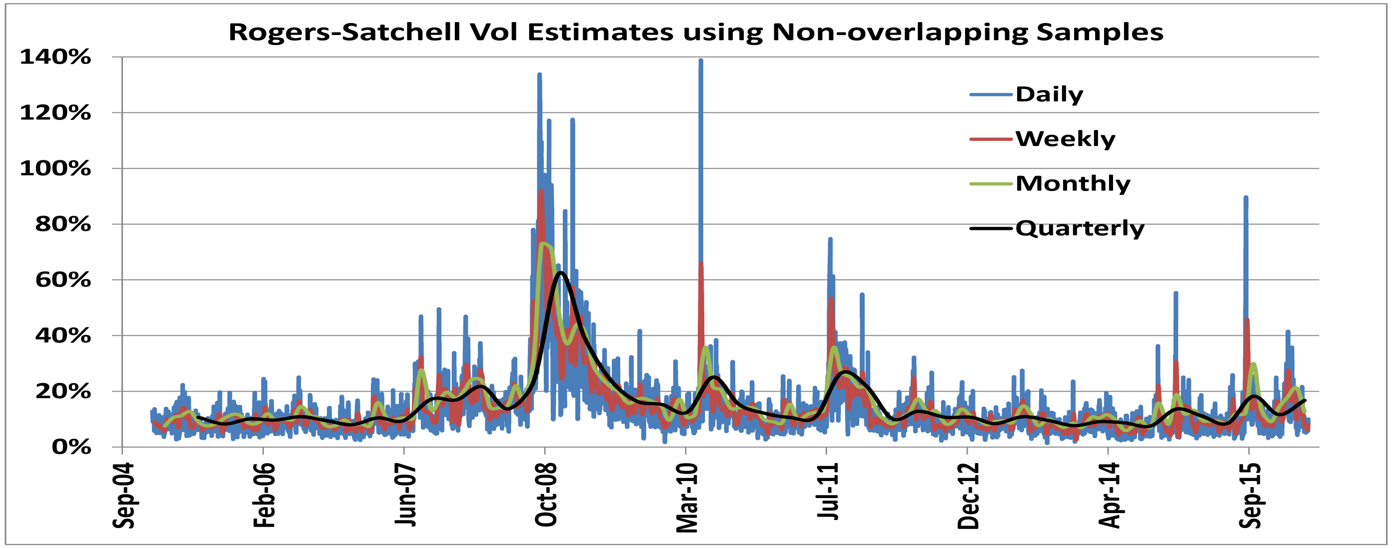


3.3 Specification of Estimation Window

1) Time series of Rogers-Satchell volatility estimates using S&P500 index ETF (SPY) from 2005 to 2016

Key observations:

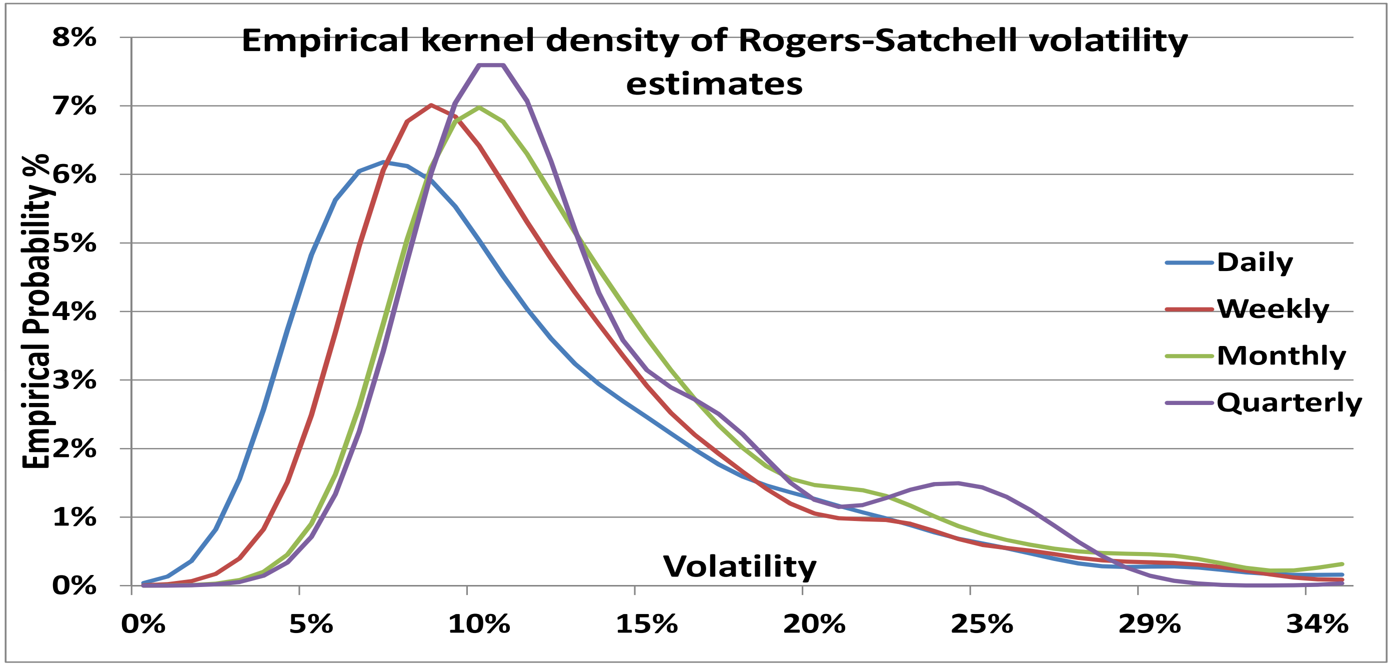
- Typically, the estimation window is based on total returns observed within daily, weekly, monthly, or quarterly period
- I strongly recommend to use non-overlapping intervals for volatility estimation: in this case the statistical estimation errors are uncorrelated and uniform (not prone to heteroscedasticity)
- The estimation window should be related to the frequency of strategy rebalancing: for monthly re-balancing use monthly observation window, etc



2) Empirical kernel density Rogers-Satchell volatility estimates based on total returns observed within daily, weekly, monthly, or quarterly period

Key observations:

- Daily volatility estimate is too dispersed
- Monthly and quarterly volatilities provide sharper estimates



3.4 Bias Correction

The sample standard deviation is computed by:

$$\sigma_t^2(T) = \frac{1}{T-1}\sum_{k=1}^{T}\left(r_{t-k} - \overline{r}_t(T)\right)^2 \tag{3.6}$$

where $T-1$ Bessel's correction to correct for the bias in the estimation of the population variance.

In most of applications we are interested in the estimate of the standard deviation (the volatility) of the population distribution of the returns. The bias correction depends on the particular distribution.

For normal distribution:

$$\begin{aligned}\widehat{\sigma}_t(T) &= \frac{1}{C(T)}\sqrt{\sigma_t^2(T)} \\ C(T) &= \sqrt{\frac{2}{T-1}}\frac{\Gamma\left(\frac{T}{2}\right)}{\Gamma\left(\frac{T-1}{2}\right)} \\ &= 1 - \frac{1}{4}\frac{1}{T} - \frac{7}{32}\frac{1}{T^2} - \frac{19}{128}\frac{1}{T^3} + O\left(T^{-4}\right)\end{aligned}$$

where $\Gamma(x)$ is the gamma function

For weekly estimator with $T = 5$, $C(5) = 0.94$, and the correction increases the estimated volatility by factor of 1.06

For non-normal heavy-tailed distributions an approximation (up to $O(T^{-1})$)

$$C(T) = \sqrt{\frac{1}{T - \frac{3}{2} - \frac{1}{4}m_4}}$$

where m_4 is the population excess kurtosis. The excess kurtosis may be estimated from data.

Where returns are serially correlated (as for volatility and rates), we need also to adjust for the auto-correlation.

3.5 Forecasting Volatility

- Given an estimate of the recent volatility, next step is to make a forecast over the next rebalancing/sample period
- As we observe in figures above, the volatility has mean-reverting features: periods of high volatility are followed by periods of low volatility and vice versa
- For a robust forecasting, we should incorporate the mean-reversion feature

However, the estimators using the sample state are based on the assumption of independent sample and observations within the sample

As a result, the estimate of the volatility computed using the data at the beginning of period t is applied to forecast the volatility at the end of the period t or equivalently at the start of the period $t+1$:

$$\hat{\sigma}^2_{t+1} = \sigma^2_t(T) \tag{3.7}$$

3.6 Volatility Estimators using Exponentially Weighted Moving Averages (EWMA)

Since we observe empirically that the volatility is indeed auto-correlated, now we consider the estimators with the auto-correlation of the volatility

The estimator using exponentially weighted moving averages (EWMA) incorporates the volatility from the previous period to estimate and forecast the volatility for the current and future periods

By the convention, the returns are not mean-adjusted

Estimator employing the **finite EWMA** only uses returns in the sample:

$$\begin{aligned}
\sigma_t^2(T) &= \Lambda(T)\left[r_{t-1}^2 + \lambda r_{t-2}^2 + \lambda^2 r_{t-3}^2 + ... + \lambda^{T-1} r_{t-T}^2\right] \\
&= \Lambda(T) \sum_{k=1}^{T} \lambda^{k-1} r_{t-k}^2 \\
\Lambda(T) &= \frac{1}{1 + \lambda + \lambda^2 + ... + \lambda^{T-1}}
\end{aligned} \tag{3.8}$$

Estimator employing the **infinite EWMA** Infinite EWMA uses returns in the sample plus previous volatility:

$$\begin{aligned}
\sigma_t^2(\infty) &= \frac{1}{1 + \lambda + \lambda^2 + ...}\left[r_{t-1}^2 + \lambda r_{t-2}^2 + \lambda^2 r_{t-3}^2 + ...\right] \\
&= (1 - \lambda) \sum_{k=1}^{\infty} \lambda^{k-1} r_{t-k}^2
\end{aligned} \tag{3.9}$$

Note that:

$$\begin{aligned}\sum_{k=1}^{\infty}\lambda^{k-1}r_{t-k}^2 &= \sum_{k=1}^{T}\lambda^{k-1}r_{t-k}^2+\sum_{k=T+1}^{\infty}\lambda^{k-1}r_{t-k}^2\\ &= \sum_{k=1}^{T}\lambda^{k-1}r_{t-k}^2+\lambda^T\sum_{k=T+1}^{\infty}\lambda^{k-T-1}r_{t-k}^2\\ &= \sum_{k=1}^{T}\lambda^{k-1}r_{t-k}^2+\lambda^T\sum_{k'=1}^{\infty}\lambda^{k'-1}r_{t-k'-T}^2\\ &= \frac{1}{\Lambda(T)}\sigma_t^2(T)+\frac{\lambda^T}{(1-\lambda)}\sigma_{t-T}^2(\infty)\end{aligned} \tag{3.10}$$

using $k'=k-T$

Thus:

$$\sigma_t^2(\infty)=\frac{(1-\lambda)}{\Lambda(T)}\sigma_t^2(T)+\lambda^T\sigma_{t-T}^2(\infty) \tag{3.11}$$

Note that:

$$\begin{aligned}\frac{(1-\lambda)}{\Lambda(T)}+\lambda^T &= (1-\lambda)\left(1+\lambda+\lambda^2+...+\lambda^{T-1}\right)+\lambda^T\\ &= 1\end{aligned} \tag{3.12}$$

Thus, the infinite EWMA is a weighted average of the finite EWMA with window T and the infinite EWMA up to the period T

For $T=1$ we obtain:

$$\sigma_t^2(\infty)=(1-\lambda)r_{t-1}^2+\lambda\sigma_{t-1}^2(\infty) \tag{3.13}$$

using by convention $\sigma_t^2(1) = r_{t-1}^2$

As a result, the infinite EWMA is computed by the recursion starting from the EWMA volatility at the start of the sample $t - T$ (estimated from prior periods) and then applying the recursion (3.13)

3.6.1 Adjusting for the mean

For robust estimation, we need adjust for the mean return

For the finite EWMA, the mean return $\overline{r}_t(T)$ is computed as the EWMA mean:

$$\begin{aligned}\overline{r}_t(T) &= \Lambda(T)\left[r_{t-1} + \lambda r_{t-2} + \lambda^2 r_{t-3} + ... + \lambda^{T-1} r_{t-T}\right] \\ &= \Lambda(T)\sum_{k=1}^{T}\lambda^{k-1} r_{t-k}\end{aligned} \tag{3.14}$$

Then the estimate of EWMA volatility is computed as EWMA of the sum of squared deviations from the mean:

$$\sigma_t^2(T) = \Lambda(T)\sum_{k=1}^{T}\lambda^{k-1}\left(r_{t-k} - \overline{r}_t(T)\right)^2 \tag{3.15}$$

For the infinite EWMA, we can show similarly to Eq (3.11) for $\sigma_t^2(\infty)$ that the infinite EWMA return is the sum of finite EWMA return:

$$\overline{r}_t(\infty) = \frac{(1-\lambda)}{\Lambda(T)}\overline{r}_t(T) + \lambda^T \overline{r}_{t-T}(\infty) \tag{3.16}$$

For practical purposes we can only use returns in the current sample

3.6.2 Volatility Forecast

At the start of trading period at time t we have the information from time $t-1$, $\mathcal{F}_{t-1}$, which we use to compute volatility σ_{t+1}^2 for up to the start on the next trading period $t+1$

The key assumption is that conditional on the available information set $\mathcal{F}_{t-1}$ the time t return is normal

$$r_t \mid \mathcal{F}_{t-1} = \mathbf{N}\left(0, \sigma_t^2(\infty)\right) \tag{3.17}$$

where $\mathbf{N}$ is the PDF of standard normal random variable (it is also possible to use heavy tailed Student t-distribution etc)

For the infinite EWMA, we apply Eq (3.13):

$$\widehat{\sigma}_{t+1}^2(\infty) = (1-\lambda)\mathbb{E}\left[r_t^2 \mid \mathcal{F}_{t-1}\right] + \lambda\sigma_t^2(\infty) \tag{3.18}$$

Applying equation (3.17), we obtain that

$$\mathbb{E}\left[r_t^2 \mid \mathcal{F}_{t-1}\right] = \sigma_t^2(\infty) \tag{3.19}$$

As a result, the volatility forecast from the EWMA estimator is equal to the most recent estimate

$$\widehat{\sigma}_{t+1}^2(\infty) = \sigma_t^2(\infty) \tag{3.20}$$

and for k-th period:

$$\widehat{\sigma}_{t+k}^2(\infty) = \sigma_t^2(\infty) \tag{3.21}$$

Key Observations:

- The forecast conditional volatility is flat and equals to the most recent estimate even though the model assume auto-correlation for the volatility estimate
- Technically, the estimator using EWMA assumes the same type of diffusive Brownian motion-based model for price returns as the model for the equally weighted variance estimator
- The observations that the volatility estimates are time-varying is because of the exponential weighting of data

For the estimator with finite EWMA, we obtain:

$$\begin{aligned} \widehat{\sigma}^2_{t+1} &= \Lambda(T)\left[\mathbb{E}\left[r_t^2 \mid \mathcal{F}_{t-1}\right] + \lambda r_{t-1}^2 + \lambda^2 r_{t-2}^2 + ... + \lambda^{T-1} r_{t-T+1}^2\right] \\ &= \Lambda(T)\left[\sigma_t^2(T) + \lambda r_{t-1}^2 + \lambda^2 r_{t-2}^2 + ... + \lambda^{T-1} r_{t-T+1}^2\right] \end{aligned} \tag{3.22}$$

$$\begin{aligned} \widehat{\sigma}^2_{t+2} &= \Lambda(T)\left[\mathbb{E}\left[r_{t+1}^2 \mid \mathcal{F}_{t-1}\right] + \lambda \mathbb{E}\left[r_t^2 \mid \mathcal{F}_{t-1}\right] + \lambda^2 r_{t-1}^2 + ... + \lambda^{T-1} r_{t-T+2}^2\right] \\ &= \Lambda(T)\left[\widehat{\sigma}^2_{t+1} + \lambda \sigma_t^2(T) + \lambda^2 r_{t-1}^2 + ... + \lambda^{T-1} r_{t-T+2}^2\right] \end{aligned} \tag{3.23}$$

For $q < T$:

$$\widehat{\sigma}^2_{t+q} = \Lambda(T)\left[\widehat{\sigma}^2_{t+q-1} + \lambda \widehat{\sigma}^2_{t+q-2} + ... + \lambda^2 r_{t-1}^2 + ... + \lambda^{T-1} r_{t-T+q}^2\right] \tag{3.24}$$

We observe some recursive dependency, but it is very small and the forecast is not much different from the current estimate

3.7 Empirical Analysis of Sample State Methods

3.7.1 Estimation of Exponential Smoothing and Mean-reversion

For EWMA estimator, we need to estimate the exponential smoothing parameter λ which measures the degree of persistence or mean-reversion in the volatility times series:

- Low value of the smoothing parameter λ: the estimate of the volatility is largely impacted by the most recent returns while the weight of volatility in prior periods is small
- High value of the smoothing parameter λ: the estimate of the volatility is only modestlyy impacted by the most recent returns while the weight of volatility in prior periods is high

Key Observations:

1. The value of the exponential smoothing parameter λ depends on the sampling frequency of the volatility: λ is high for short frequency (daily, weekly), λ is smaller for monthly and quarterly
2. In continuous time SV models, the degree of exponential smoothing is related to the mean-reversion parameter κ proportional to infinitesimal time change dt (in linear one-factor models)
3. On the one hand, the smoothing parameter λ can be estimated using a discrete GARCH-type model (integrated GARCH) for the price returns on daily scale, but in this model the time series of the volatilities is inferred and it is different from the time series of realized volatilities
4. On the other hand, the smoothing parameter λ can be inferred using the estimate of the mean-reversion in a continuous-time SV model, however in this model, the estimate of the mean-reversion is closely linked to the volatility-of-volatility parameter
5. From my own practical experience, by making either λ or κ to be free parameters which are estimated in the process of calibration, we do not improve the explanatory power of an SV model
6. It is robust to estimate either λ or κ beforehand and use this estimated value in other calibrations

For one factor continuous time SV models the auto-correlation between the observed volatilities:

$$\rho\left(V_t, V_0\right) \propto e^{-\kappa t} \tag{3.25}$$

We estimate the empirical value of the auto-correlation for N-lags $\{\rho_n\}$, $n = 1, ..., N$ and fit κ using the following regression:

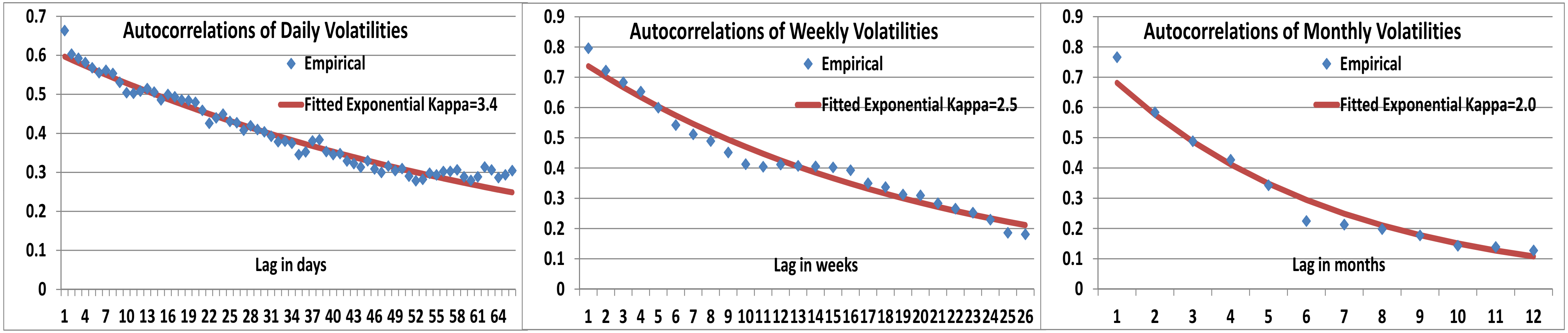
$$-\ln\left(\rho_n\right) = \widehat{\kappa} \times (n \times dt) + \widehat{\alpha}\ ,\ \ n = 1, ..., N \tag{3.26}$$

where $\widehat{\kappa}$ is the estimate for mean-reversion parameter

The value of the estimate depends on the sample window for volatility estimation:

- Daily volatilities exhibit stronger auto-correlation so the smoothing is high
- Weekly and monthly volatilities exhibit weaker auto-correlation so the smoothing is modest

Figure: Auto-correlations for S&P500 index ETF using daily, weekly, and monthly Rogers-Satchell volatilities

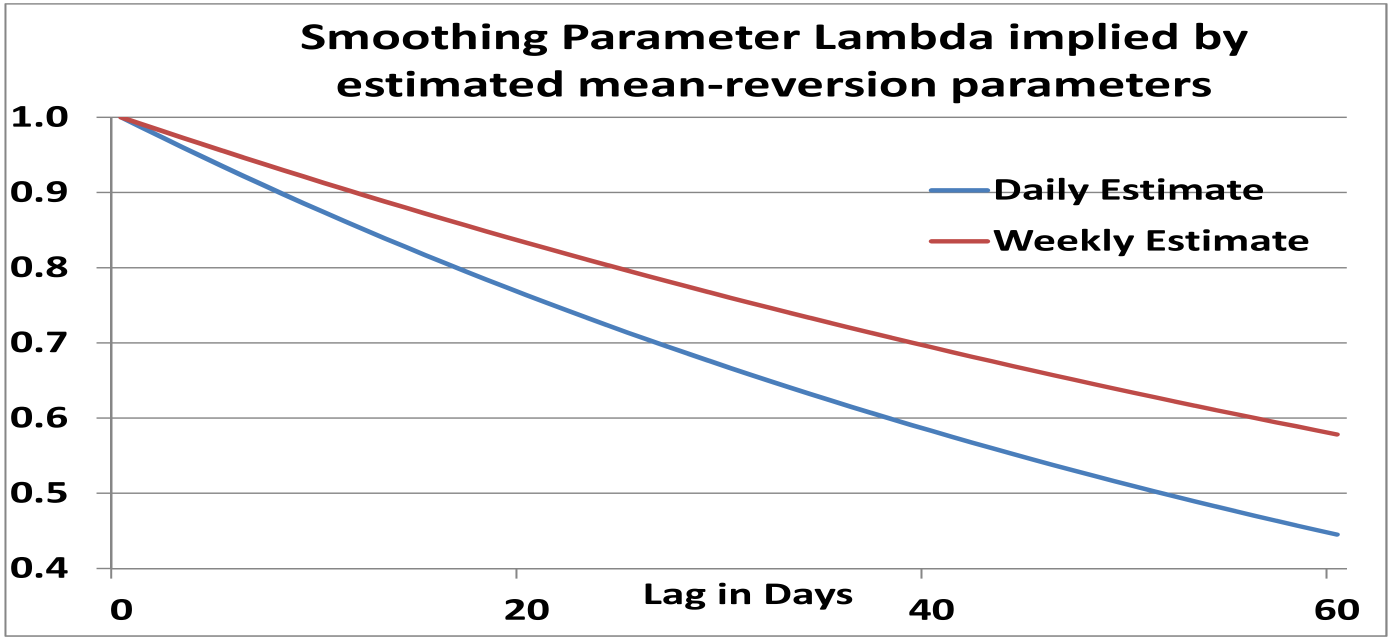


Given the estimated value of mean-reversion parameter $\widehat{\kappa}$, the smoothing parameter lambda for the volatility estimate with sample window of N daily observation is computed by:

$$\lambda_n = \exp\left(-\widehat{\kappa} \times (n \times dt)\right) \tag{3.27}$$

Key practical considerations:

- For medium to low frequency trading strategies with weekly and monthly re-balancing, it is recommended to use the mean-reversion estimated to weekly estimates of volatility
- The estimated values of the mean-reversion are applied to estimate the EWMA parameter λ and are updated infrequently
- For back-test purposes the data for estimation of the mean-reversion should be prior to the date when a signal is generated



3.8 Empirical Tests of Forecasting Power of Sample Space Estimators

To test the predictive power of the volatility forecast for each of the sample space estimators, I design the back-test as follows:

1. Compute the realized volatility $\widehat{\sigma}_t(T)$ for period t
2. Apply the estimate for time $t+1$: $\widehat{\sigma}_t(T) = \widehat{\sigma}_{t+1}(T)$ for period $t+1$
3. Compute the actual realized volatility $\widehat{\sigma}_{t+1}(T)$ for period $t+1$
4. Make the statistical test about the quality of the forecast volatility by apply the following regression model:

$$\widehat{\sigma}_{t+1}(T) = \alpha + \beta\widehat{\sigma}_t(T) + \epsilon_{t+1} \tag{3.28}$$

Key practical considerations:

- The estimator producing a volatility forecast is indicated by high explanatory power R^2, small intercept α and beta coefficient close to one $\beta \approx 1.0$
- Each volatility estimator is applied to gauge the predictive power for:
    1. its own volatility forecast
    2. close-to-close volatility
- It is important that volatility is estimated and forecast over non-overlapping intervals so that forecast errors ϵ_{t+1} are iid (in theory)

Set-up of the back-test:

- 11+ years of time series data from 1-January-2005 up to 1-April-2016
- 40 ETFs and large caps representing major stock index ETFs, sector ETFs, fixed-income asset class ETFs

ETFs and stocks for empirical back-test

Table 3: Sample

Ticker	Underlying Index		Ticker	Underlying Index
Major Market Indices			**Sector ETFs**	
SPY	S&P 500 index		XLP	Staples
QQQ	Nasdaq		XLU	Utilities
VTI	Whole US market		XLV	Health Care
IWM	Small Caps		XLY	Discretionary
IJH	Mid Caps		XLK	Technology
IWF	Russell growth		XLI	Industrials
IWD	Russell value		XLE	Energy
EFA	Developed world ex US & Canada		XLB	Materials
EEM	Emegring markets		XLF	Financials
VGK	Europe		IYR	Real Estate REIT
Large Cap Stocks			**Fixed Income ETFs**	
PG	Staples		SHY	1-3 Year Treasury Bond
T	Telco		EIE	3-7 Year Treasury Bond
JNJ	Health Care		IEF	7-10 Year Treasury Bond
PFE	Health Care		TLT	20+ Year Treasury Bond
JPM	Banks		TIP	Treasury Inflation Protected Bond
XOM	Energy		MBB	Mortgage Bond
GOOGL	Tech		LQD	Investment Grade Corporate Bond
AAPL	Tech		HYG	High Yield Corporate Bond
FB	Tech		EMB	USD Emerging Markets Bond
MSFT	Tech		AGG	U.S. Aggregate Bond

Design of the back-test and performance measurement

Compute volatility estimates using 4 sample windows:

1. Daily
2. Weekly (every Friday)
3. Monthly (every 3rd Friday of the month)
4. Quarterly (every quarter on 3rd Friday of the month)

For each of 40 assets:

1. Compute 7 time series of the realized volatility for each of the estimators
2. For each estimator apply equation (3.28) to compute the predictive power R^2 of each of 7 estimators
3. Rank 7 estimators by their predictive power
4. The predictor in the top place with the highest R^2 gets the score of 3 points
5. The predictor in the second place gets the score of 2 points
6. The predictor in the third place gets the score of 1 points

The score points of each estimator for one asset are aggregated for all 40 assets

First, I start by illustrating the methodology for S&P 500 index ETF

3.8.1 Test for S&P 500 index ETF (SPY)

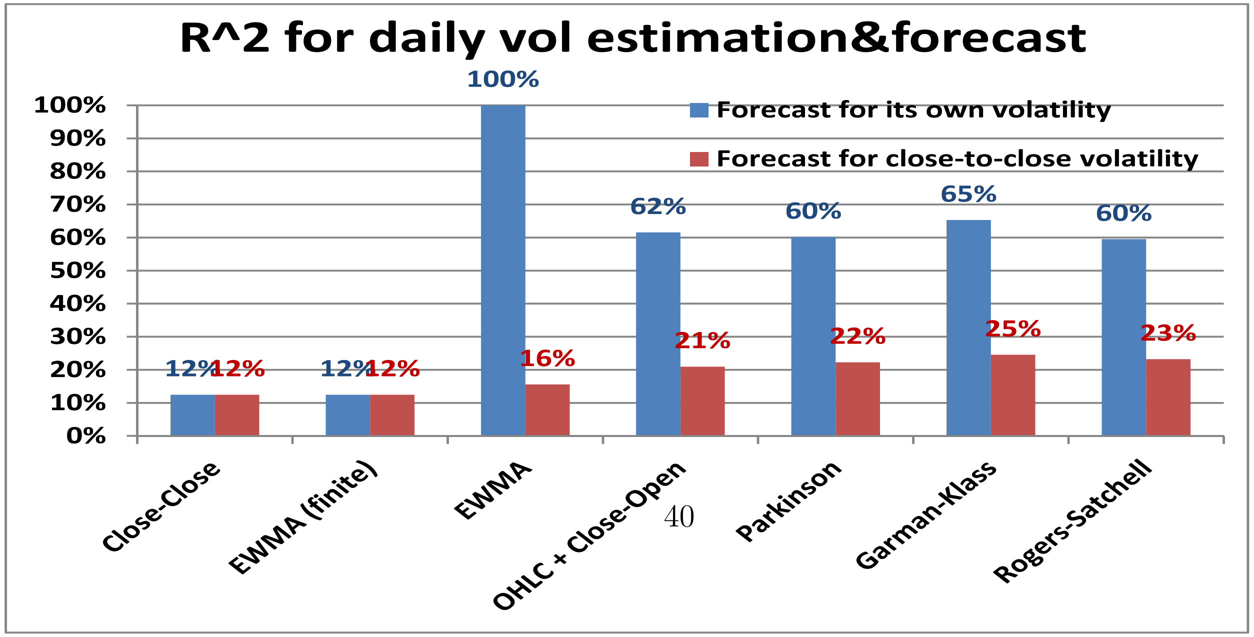
Daily Estimation and Forecast:

- EWMA provides the best forecast for its own volatility but not for close-to-close volatility
- Estimators employing intra-day information provide reasonable explain for their own volatilities and modest explain for close-to-close volatility

Table 4: Left: forecast for its own volatility; Right: forecast for close-to-close volatility

	R^2	α	β
Close-Close	12%	8.5%	0.09
EWMA (finite)	12%	8.5%	0.09
EWMA	100%	0.0%	1.00
OHLC + Close-Open	62%	5.6%	0.52
Parkinson	60%	4.7%	0.52
Garman-Klass	65%	4.4%	0.55
Rogers-Satchell	60%	5.0%	0.49

	R^2	α	β
Close-Close	12%	8.5%	0.09
EWMA (finite)	12%	8.5%	0.09
EWMA	16%	4.8%	0.28
OHLC + Close-Open	21%	5.9%	0.26
Parkinson	22%	5.5%	0.36
Garman-Klass	25%	5.1%	0.39
Rogers-Satchell	23%	5.6%	0.35



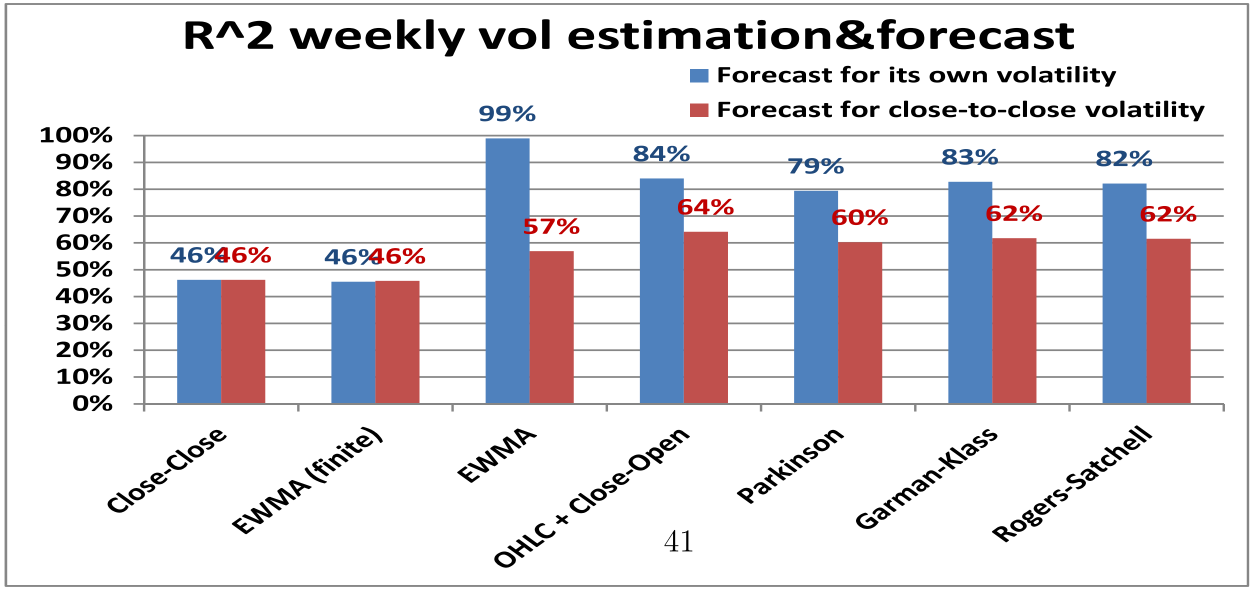
Weekly Estimation and Forecast:

- EWMA still provides the best forecast for its own volatility but not for close-to-close volatility
- Estimators employing intra-day information provide very good explain for both their own volatilities and for close-to-close volatility

Table 5: Left: forecast for its own volatility; Right: forecast for close-to-close volatility

	R^2	α	β
Close-Close	46%	5.8%	0.48
EWMA (finite)	46%	6.1%	0.45
EWMA	99%	0.5%	0.94
OHLC + Close-Open	84%	3.3%	0.74
Parkinson	79%	3.4%	0.68
Garman-Klass	83%	3.1%	0.71
Rogers-Satchell	82%	3.1%	0.71

	R^2	α	β
Close-Close	46%	5.8%	0.48
EWMA (finite)	46%	6.0%	0.46
EWMA	57%	2.4%	0.68
OHLC + Close-Open	64%	3.0%	0.61
Parkinson	60%	3.2%	0.72
Garman-Klass	62%	3.2%	0.72
Rogers-Satchell	62%	3.4%	0.69



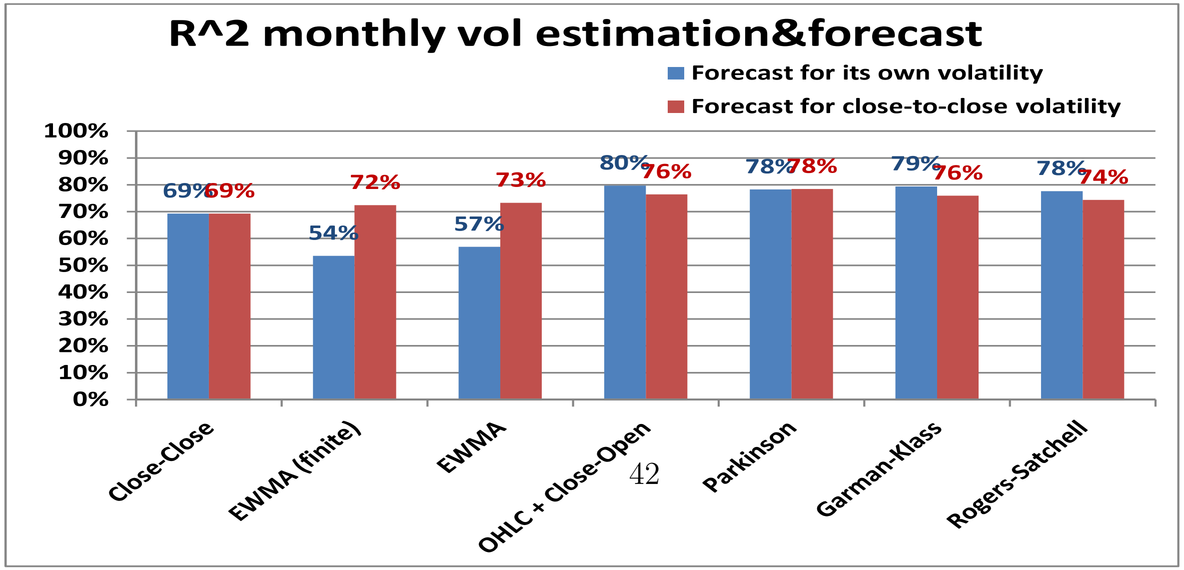
Monthly Estimation and Forecast:

- The predictive power of EWMA estimator becomes the worst for both its own volatility and for close-to-close volatility - the volatility of recent returns becomes more important for the forecast volatility
- Close-to-close estimators has a reasonable explanatory power - the volatility level at medium time scale is persistent
- Estimators employing intra-day information provide very good and consistent explain

Table 6: Left: forecast for its own volatility; Right: forecast for close-to-close volatility

	R^2	α	β
Close-Close	69%	5.1%	0.63
EWMA (finite)	54%	7.3%	0.47
EWMA	57%	7.1%	0.46
OHLC + Close-Open	80%	4.4%	0.67
Parkinson	78%	4.2%	0.62
Garman-Klass	79%	4.2%	0.63
Rogers-Satchell	78%	4.3%	0.62

	R^2	α	β
Close-Close	69%	5.1%	0.63
EWMA (finite)	72%	5.4%	0.68
EWMA	73%	5.2%	0.65
OHLC + Close-Open	76%	3.7%	0.68
Parkinson	78%	2.7%	0.92
Garman-Klass	76%	3.7%	0.83
Rogers-Satchell	74%	4.2%	0.78



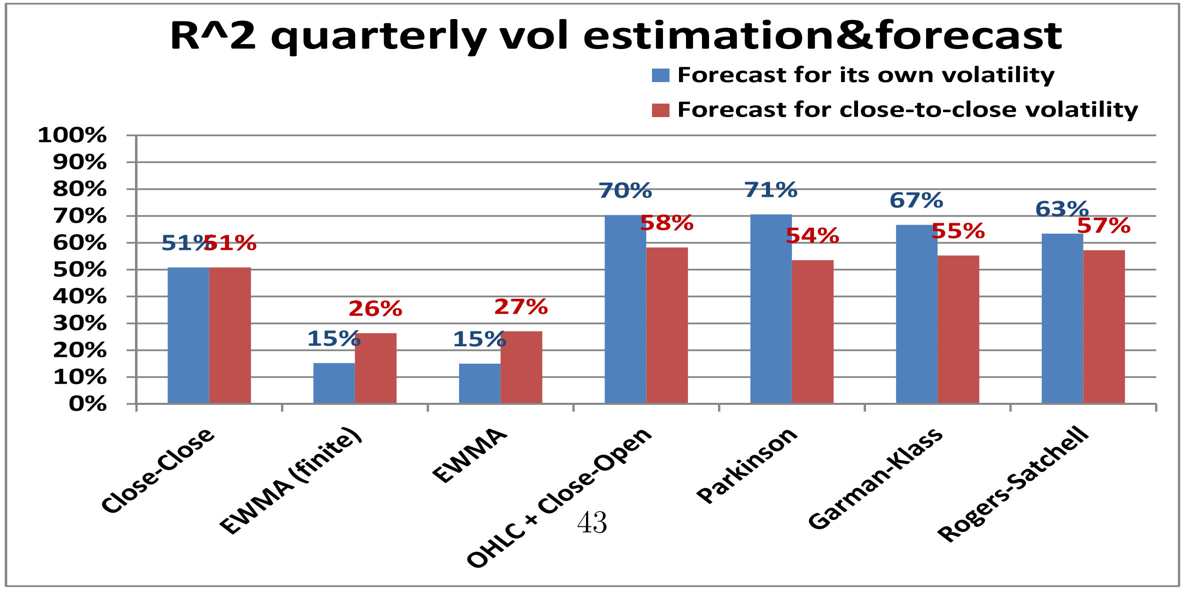
Quarterly Estimation and Forecast:

- As for monthly sampling, the predictive power of EWMA estimator becomes the worst for both its own volatility and for close-to-close volatility
- Close-to-close estimators has a reasonable explanatory power - the volatility level at medium time scale is persistent
- Estimators employing intra-day information provide very good and consistent explain

Table 7: Left: forecast for its own volatility; Right: forecast for close-to-close volatility

	R^2	α	β
Close-Close	51%	7.8%	0.45
EWMA (finite)	15%	8.3%	0.19
EWMA	15%	8.4%	0.20
OHLC + Close-Open	70%	5.8%	0.50
Parkinson	71%	5.1%	0.50
Garman-Klass	67%	5.3%	0.50
Rogers-Satchell	63%	5.4%	0.50

	R^2	α	β
Close-Close	51%	7.8%	0.45
EWMA (finite)	26%	10.8%	0.34
EWMA	27%	10.8%	0.38
OHLC + Close-Open	58%	7.0%	0.45
Parkinson	54%	7.4%	0.55
Garman-Klass	55%	7.4%	0.54
Rogers-Satchell	57%	7.3%	0.52



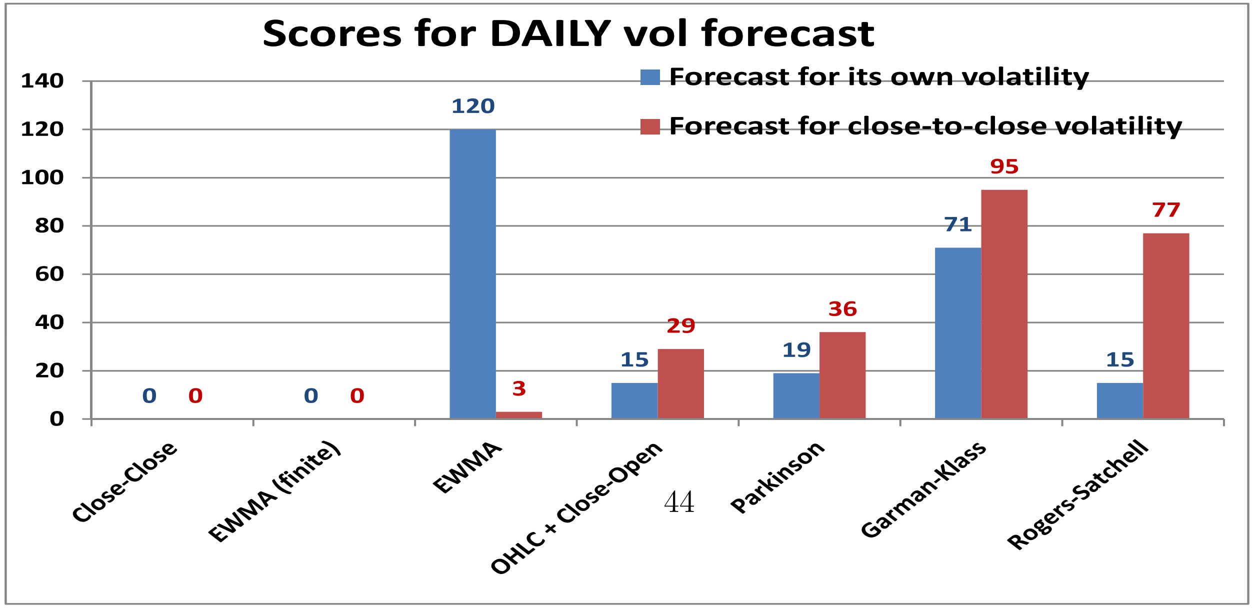
3.8.2 Aggregate Test for 40 assets: total scores

Daily Estimation and Forecast:

- EWMA provides the best forecast for its own volatility but not for close-to-close volatility
- Estimators employing intra-day information provide reasonable explain for their own volatilities and modest explain for close-to-close volatility

Table 8: Left: forecast for its own volatility; Right: forecast for close-to-close volatility

	Scores	Median R^2	Scores	Median R^2
Close-Close	0	11%	0	11%
EWMA (finite)	0	11%	0	11%
EWMA	120	100%	3	13%
OHLC + Close-Open	15	55%	29	18%
Parkinson	19	56%	36	19%
Garman-Klass	71	62%	95	21%
Rogers-Satchell	15	56%	77	21%

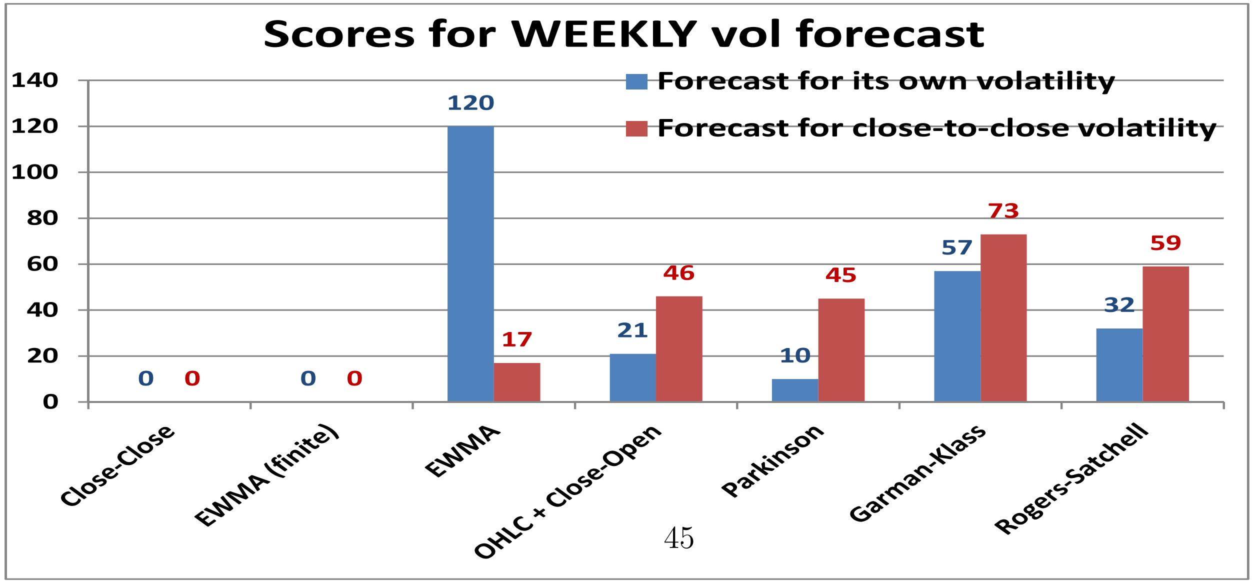


Weekly Estimation and Forecast:

- EWMA still provides the best forecast for its own volatility but not for close-to-close volatility
- Estimators employing intra-day information provide very good explain for both their own volatilities and for close-to-close volatility

Table 9: Left: forecast for its own volatility; Right: forecast for close-to-close volatility

	Scores	Median R^2	Scores	Median R^2
Close-Close	0	40%	0	40%
EWMA (finite)	0	40%	0	40%
EWMA	120	99%	17	52%
OHLC + Close-Open	21	82%	46	58%
Parkinson	10	79%	45	61%
Garman-Klass	57	83%	73	61%
Rogers-Satchell	32	82%	59	61%

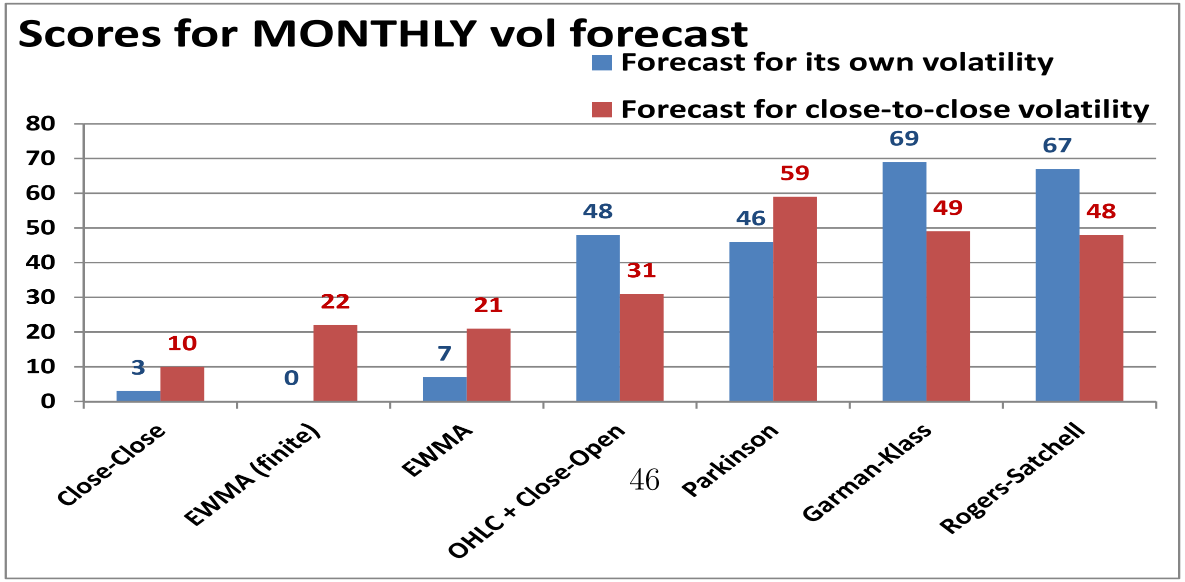


Monthly Estimation and Forecast:

- The predictive power of EWMA estimator becomes the worst for both its own volatility and for close-to-close volatility - the volatility of recent returns becomes more important for the forecast volatility
- Close-to-close estimators has a reasonable explanatory power - the volatility level at medium time scale is persistent
- Estimators employing intra-day information provide very good and consistent explain

Table 10: Left: forecast for its own volatility; Right: forecast for close-to-close volatility

	Scores	Median R^2	Scores	Median R^2
Close-Close	3	64%	10	64%
EWMA (finite)	0	50%	22	64%
EWMA	7	52%	21	65%
OHLC + Close-Open	48	80%	31	68%
Parkinson	46	78%	59	70%
Garman-Klass	69	79%	49	70%
Rogers-Satchell	67	77%	48	69%

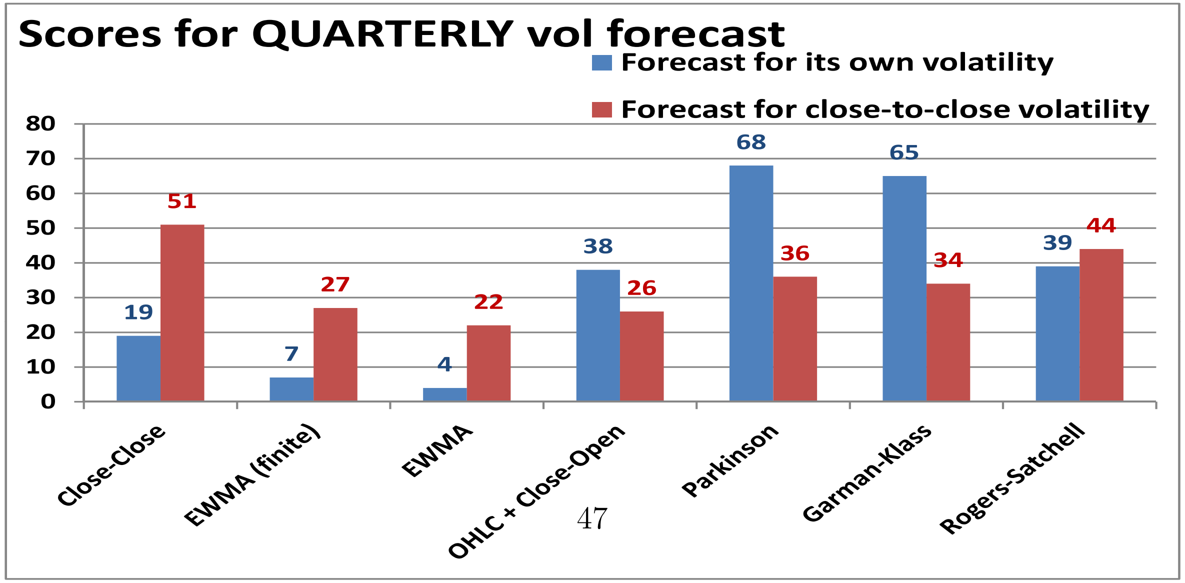


Quarterly Estimation and Forecast:

- As for monthly sampling, the predictive power of EWMA estimator becomes the worst for both its own volatility and for close-to-close volatility
- Close-to-close estimators has a reasonable explanatory power - the volatility level at medium time scale is persistent
- Estimators employing intra-day information provide very good and consistent explain

Table 11: Left: forecast for its own volatility; Right: forecast for close-to-close volatility

	Scores	Median R^2	Scores	Median R^2
Close-Close	19	52%	51	52%
EWMA (finite)	7	19%	27	34%
EWMA	4	20%	22	33%
OHLC + Close-Open	38	57%	26	52%
Parkinson	68	60%	36	52%
Garman-Klass	65	61%	34	53%
Rogers-Satchell	39	60%	44	52%



3.9 Conclusions

Estimators of the volatility with **sample state methods** have the following properties

- The volatility is estimated using the sample of most recent returns
- The length of the estimation sample (daily, weekly, monthly, quarterly) is selected based on the strategy re-balancing, the sample must be
- The volatility forecast is based on flat extrapolation of the most recent volatility for the future period
- The most serious drawback for forecasting is that these methods do not account for the mean-reverting feature of the volatility
- Sample state methods employing intra-day price information provide a reasonable estimate and forecast for the realized volatility:
    1. They also perform well for trading strategies employing the realized volatility: volatility-targeting, volatility-parity allocations
    2. When choosing a particular estimator to apply for a trading strategy, the predictive power of the estimator should be studied separately
- When intra-day price information is not available or the liquidity is low (for example: European ETFs), the EWMA estimator should be applied
- The forecast is not employing the mean-reverting feature of the volatility

To account for the mean-reversion of the volatility we need to apply estimation methods using **state inference** (GARCH type and continuous time SV models)

4 Volatility Estimation and Forecast using Inference State Methods

Inference state methods assume a particular dynamics for the discrete instantenous variance of price returns and incorporate mean-reverting features for volatility forecast

The inference state methods are based on the following model for the discrete returns process:

$$r_t = \mu + \epsilon_t \sigma_t \tag{4.1}$$

where are iid variables with zero mean and unit variance and pdf $f(x)$

σ_t^2 is returns variance driven by a state equation

Conditional on the information $\mathcal{F}_t$, the probability density function (PDF) of returns is normal (or a heavy tailed t-distribution)

The returns are uncorrelated but not iid

Define the residual return at time t, $a_t = r_t - \mu$:

$$a_t = \epsilon_t \sigma_t \tag{4.2}$$

4.1 ARCH (1) Model

Originally introduced by Engle (1982):

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 \tag{4.3}$$

$\alpha_0 > 0$ and $\alpha_1 > 0$ so that the variance is positive, and $\alpha_1 < 1$ for the stationarity

The unconditional variance of a_t:

$$\begin{aligned}
\mathbb{V}ar\left[a_t\right] &= \mathbb{E}\left[a_t^2\right] - \left(\mathbb{E}\left[a_t\right]\right)^2 \\
&= \mathbb{E}\left[a_t^2\right] \ \text{ } (a_t \text{ has zero mean}) \\
&= \mathbb{E}\left[(\epsilon_t \sigma_t)^2\right] \ \text{ (definition of } a_t) \\
&= \mathbb{E}\left[(\epsilon_t)^2\right] \mathbb{E}\left[(\sigma_t)^2\right] \ \text{ (independence of } \epsilon_t \text{ and } a_t) \\
&= \mathbb{E}\left[\alpha_0 + \alpha_1 a_{t-1}^2\right] \ \left(\text{definition of } \sigma_t^2\right) \\
&= \alpha_0 + \alpha_1 \mathbb{E}\left[a_{t-1}^2\right] \\
&= \alpha_0 + \alpha_1 \mathbb{V}ar\left[a_{t-1}\right] \ \left(\ \mathbb{E}\left[a_{t-1}^2\right] = \mathbb{V}ar\left[a_{t-1}\right] \ \right) \\
&= \alpha_0 + \alpha_1 \mathbb{V}ar\left[a_t\right] \ \text{ (stationary of } a_t)
\end{aligned} \tag{4.4}$$

As a result:

$$\mathbb{V}ar\left[a_t\right] = \frac{\alpha_0}{1 - \alpha_1} \tag{4.5}$$

ARCH(1) model is equivalent to the autoregressive AR(1) model on the squared residuals a_t^2

Consider the conditional forecast error between the squared residual and the conditional expectation:

$$\begin{aligned}
e_t &= a_t^2 - \mathbb{E}\left[a_t^2 \mid \mathcal{F}_{t-1}\right] \\
&= a_t^2 - \sigma_t^2
\end{aligned} \tag{4.6}$$

e_t are uncorrelated zero mean variables.

ARCH(1) model can be presented as:

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 \ \text{ (definition)} \\
a_t^2 - e_t &= \alpha_0 + \alpha_1 a_{t-1}^2 \ \text{ (Eq (4.6))} \\
a_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + e_t \ \text{ (AR(1) model with iid noise } e_t \text{)}
\end{aligned} \tag{4.7}$$

The ARCH(1) model implies that a large value of return is followed by a higher volatility

4.2 Generalized Auto-regressive (GARCH) Model

Bollerslev (1986) developed an extension of ARCH(1) model:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{4.8}$$

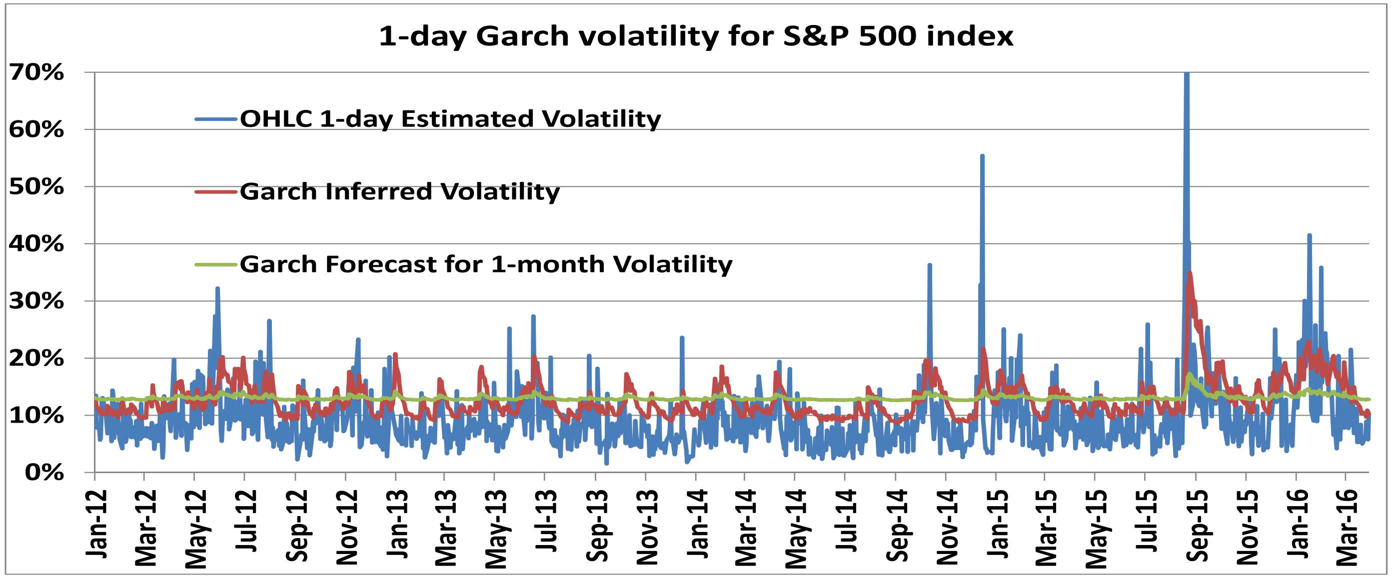
with $\alpha_0 > 0$, $\alpha_1 > 0$, $\beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$

The one step forecast of the variance is a mixture if last period forecast and squared return

The model inferred the state of the volatility given past returns and model parameters

Model parameters are estimated using the maximum likelihood methods

Figure: 1-day Garch volatility for S&P 500 index



GARCH(1,1) model is equivalent to ARMA(1,1) model for squared residuals:

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \text{ (definition)} \\ a_t^2 - e_t &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 (a_{t-1}^2 - e_{t-1}) \text{ (Eq (4.6))} \\ a_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) a_{t-1}^2 + e_t + \beta_1 e_{t-1} \text{ (ARMA(1,1) model with iid noise } e_t \text{)} \end{aligned} \tag{4.9}$$

The unconditional variance of a_t:

$$\begin{aligned} \mathbb{V}ar\left[a_t\right] &= \mathbb{E}\left[a_t^2\right] - \left(\mathbb{E}\left[a_t\right]\right)^2 \\ &= \mathbb{E}\left[a_t^2\right] \text{ } (a_t \text{ has zero mean}) \\ &= \mathbb{E}\left[(\epsilon_t \sigma_t)^2\right] \text{ (definition of } a_t) \\ &= \mathbb{E}\left[(\epsilon_t)^2\right] \mathbb{E}\left[(\sigma_t)^2\right] \text{ (independence of } \epsilon_t \text{ and } a_t) \\ &= \mathbb{E}\left[\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2\right] \text{ } \left(\text{definition of } \sigma_t^2\right) \\ &= \alpha_0 + \alpha_1 \mathbb{E}\left[a_{t-1}^2\right] + \beta_1 \mathbb{E}\left[\sigma_{t-1}^2\right] \\ &= \alpha_0 + \alpha_1 \mathbb{E}\left[a_{t-1}^2\right] + \beta_1 \sigma_{t-1}^2 \text{ } \left(\sigma_{t-1}^2 \text{ is measurable } \right) \\ &= \alpha_0 + (\alpha_1 + \beta_1) \mathbb{E}\left[a_{t-1}^2\right] \text{ } \left(\sigma_{t-1}^2 = \mathbb{E}\left[a_{t-1}^2\right] \right) \\ &= \alpha_0 + (\alpha_1 + \beta_1) \mathbb{V}ar\left[a_{t-1}\right] \text{ } \left(\mathbb{E}\left[a_{t-1}^2\right] = \mathbb{V}ar\left[a_{t-1}\right] \right) \\ &= \alpha_0 + (\alpha_1 + \beta_1) \mathbb{V}ar\left[a_t\right] \text{ (stationary of } a_t) \end{aligned} \tag{4.10}$$

As a result:

$$\mathbb{E}\left[\sigma^2\right] = \mathbb{V}ar\left[a_t\right] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \tag{4.11}$$

where $\mathbb{E}\left[\sigma^2\right]$ is the unconditional (long-run) variance of returns.

Garch model (4.8) can be presented as:

$$\sigma_t^2 = (1 - \alpha_1 - \beta_1) \mathbb{E}\left[\sigma^2\right] + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{4.12}$$

4.3 Volatility Forecast

One-step forecast under the GARCH model:

$$\begin{aligned}
\hat{\sigma}_{t+1}^2 &= \mathbb{E}\left[\sigma_{t+1}^2 \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\alpha_0 + \alpha_1 a_t^2 + \beta_1 \sigma_t^2 \mid \mathcal{F}_{t-1}\right] \\
&= \alpha_0 + \alpha_1 \mathbb{E}\left[a_t^2 \mid \mathcal{F}_{t-1}\right] + \beta_1 \sigma_t^2 \\
&= \alpha_0 + \alpha_1 \sigma_t^2 + \beta_1 \sigma_t^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1)\, \sigma_t^2 \\
&= \mathbb{E}\left[\sigma^2\right](1 - \alpha_1 - \beta_1) + (\alpha_1 + \beta_1)\, \sigma_t^2 \\
&= \mathbb{E}\left[\sigma^2\right] + (\alpha_1 + \beta_1)\left(\sigma_t^2 - \mathbb{E}\left[\sigma^2\right]\right)
\end{aligned} \tag{4.13}$$

Volatility forecast incorporates the long-term variance and its current value

Two-step forecast:

$$\begin{aligned}
\hat{\sigma}_{t+2}^2 &= \mathbb{E}\left[\sigma_{t+2}^2 \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\alpha_0 + \alpha_1 a_{t+1}^2 + \beta_1 \sigma_{t+1}^2 \mid \mathcal{F}_{t-1}\right] \\
&= \alpha_0 + \alpha_1 \mathbb{E}\left[a_{t+1}^2 \mid \mathcal{F}_{t-1}\right] + \beta_1 \hat{\sigma}_{t+1}^2 \\
&= \alpha_0 + \alpha_1 \hat{\sigma}_{t+1}^2 + \beta_1 \hat{\sigma}_{t+1}^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1)\, \hat{\sigma}_{t+1}^2 \\
&= \mathbb{E}\left[\sigma^2\right](1 - \alpha_1 - \beta_1) + (\alpha_1 + \beta_1)\, \hat{\sigma}_{t+1}^2 \\
&= \mathbb{E}\left[\sigma^2\right] + (\alpha_1 + \beta_1)\left(\hat{\sigma}_{t+1}^2 - \mathbb{E}\left[\sigma^2\right]\right) \\
&= \mathbb{E}\left[\sigma^2\right] + (\alpha_1 + \beta_1)^2\left(\sigma_t^2 - \mathbb{E}\left[\sigma^2\right]\right)
\end{aligned} \tag{4.14}$$

k-step forecast:

$$
\begin{aligned}
\hat{\sigma}_{t+k}^2 &= \mathbb{E}\left[\sigma_{t+k}^2 \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\alpha_0 + \alpha_1 a_{t+k-1}^2 + \beta_1 \sigma_{t+k-1}^2 \mid \mathcal{F}_{t-1}\right] \\
&= \alpha_0 + \alpha_1 \mathbb{E}\left[a_{t+k-1}^2 \mid \mathcal{F}_{t-1}\right] + \beta_1 \hat{\sigma}_{t+k-1}^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1)\, \hat{\sigma}_{t+k-1}^2 \\
&= \mathbb{E}\left[\sigma^2\right] + (\alpha_1 + \beta_1)^k \left(\sigma_t^2 - \mathbb{E}\left[\sigma^2\right]\right)
\end{aligned}
\tag{4.15}
$$

Typically, the estimated GARCH model implies that volatility reverts too fast to its mean level

See the figure above for one-month volatility forecast: the one-month forecast is almost flat at all times

4.4 Empirical Tests of Forecasting Power of Inference Space Estimators

To test the predictive power of the volatility forecast for GARCH(1,1), I design the back-test as follows:

- Consider three estimation windows: weekly, monthly, and quarterly
- At the end of estimation window, the parameters of the GARCH model are re-calibrated and the forecast is computed for the next period
- The volatility for the next period is computed using 7 estimators
- Evaluate the GARCH forecated volatility against the actual volatility computed using 7 estimators.
- Make the statistical test about the quality of the forecast volatility by appling the following regression model:

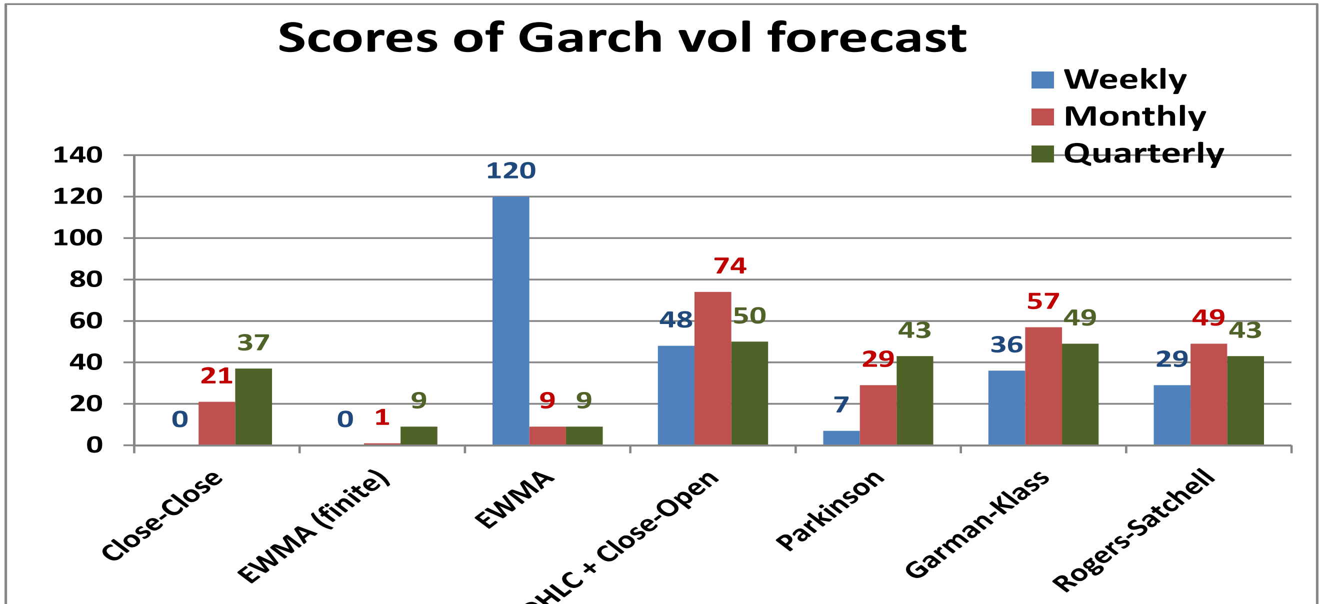
$$\widehat{\sigma}_{t+1}^{\text{actual}}(T) = \alpha + \beta\widehat{\sigma}_t^{GARCH}(T) + \epsilon_{t+1} \tag{4.16}$$

- Use the same data set of 40 assets with 14+ years of time series data from 1-January-2002 (the first 3 years are used for "warm-up" of the calibration) up to 1-April-2016
- Use the same scoring methodology for volatility estimator with the most close forecast to the GARCH model
- Aggregate results for weekly, monthly and quarterly forecast

Total scores of GARCH forecasting power:

- Garch forecast for weekly is the most consistent with EWMA volatility: both estimators are very close to each other in their nature
- For monthly and quarterly forecasts, the GARCH forecasts well the volatility from estimators employing the intra-day data

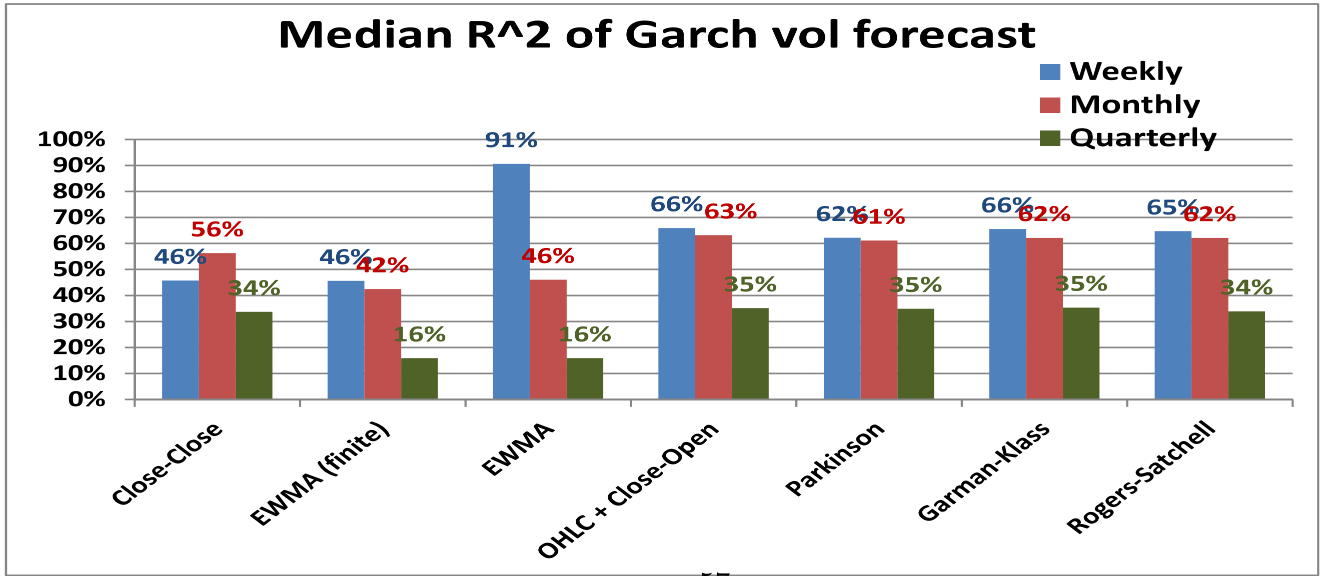
	Weekly	Monthly	Quarterly
Close-Close	0	21	37
EWMA (finite)	0	1	9
EWMA	120	9	9
OHLC + Close-Open	48	74	50
Parkinson	7	29	43
Garman-Klass	36	57	49
Rogers-Satchell	29	49	43



Median R^2 of GARCH forecasting power

- The GARCH has relatively strong explanatory power for weekly and monthly realized volatilities
- The forecast power for longer quarterly periods is rather weak

	Weekly	Monthly	Quarterly
Close-Close	46%	56%	34%
EWMA (finite)	46%	42%	16%
EWMA	91%	46%	16%
OHLC + Close-Open	66%	63%	35%
Parkinson	62%	61%	35%
Garman-Klass	66%	62%	35%
Rogers-Satchell	65%	62%	34%



4.5 Conclusions

- Inference state models and GARCH incorporate the mean level of the volatility for forecasting, however the forecast volatility is relatively flat
- Model estimation and implementation are relatively straight forward
- My experience with these type of models is mixed:
  - I believe they offer value for portfolio allocation problem when re-balancing frequency is low (monthly, quarterly)
  - They can be applied to model the intra-day seasonality and make short-term (less than a day) volatility forecast using high frequency data
- However, these models are rather limited for trading in implied volatilities:
  - The do not account (or weakly) for the leverage between changes in assets price and its volatility
  - They have econometric flavor with little to none interpretation in terms of the implied volatility surface

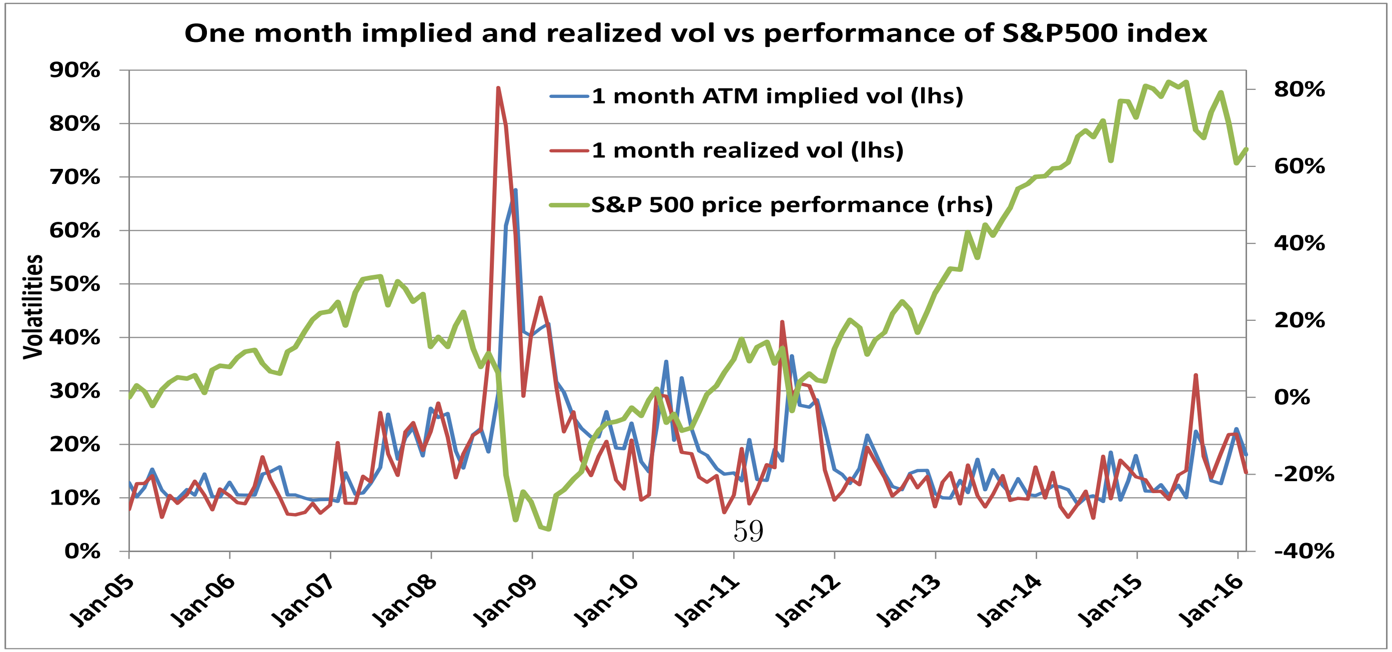
5 Volatility Estimation and Forecast using Continuous-Time Stochastic Volatility Models

For trading implied and realized volatilities the model needs to account for empirical features:

1. Negative leverage between changes in the implied and realized volatilities and price returns
2. Mean-reverting feature of implied and realized volatilities
3. Spread between implied and realized volatilities

From modeling point of view:

1. Consistency with empirical data
2. Intuition behind model parameters and their interpretation



5.1 Beta Stochastic Volatility Model

Why do we need a factor model?

In general, all quant models are applied as factor-based models:

1. For risk computations, we need to project how the changes in primary risk driver, the asset price and delta, translate into secondary driver, the implied volatility and vega
2. For trading strategies we need to explain changes in realized P&L vs changes in risk factors (implied vol vs realized vol)

Given a factor model, we can imply correlations between its factors

But given correlations alone, we cannot specify a factor model

Any dynamic factor model represented in terms of correlation matrix is mis-specified

I present the factor model for volatility dynamics based on my work with Piotr Karasinski (Sepp-Karasinski (2012)): the beta stochastic volatility model

I start by presenting some empirical observations to develop the intuition behind the beta stochastic vol model

The distribution of realized volatility is close to log-normal

The distribution for the volatility is important for stationary model with time-independent parameters

The stationary distribution is typically classified using the scaling of the volatility-of-volatility:

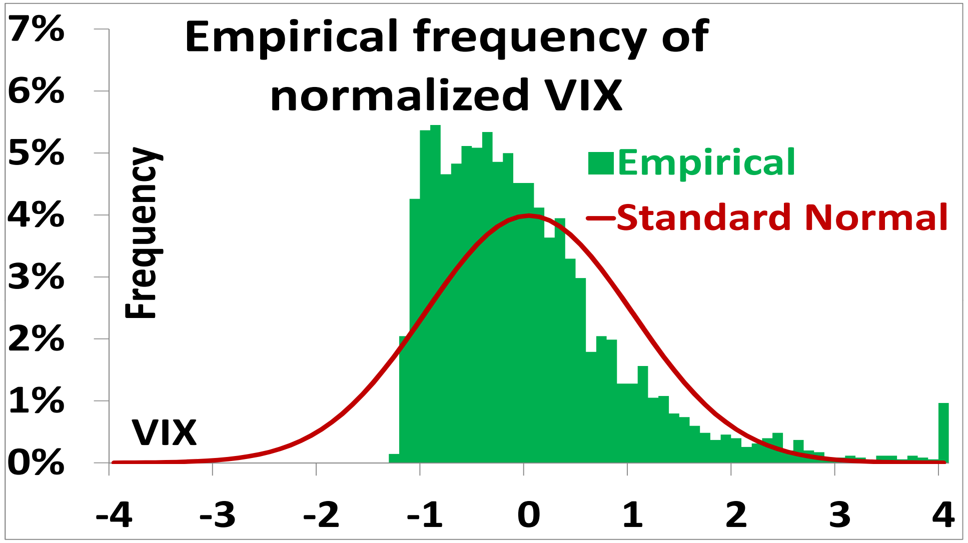
- normal model: the volatility-of-volatility is independent from the level of the volatility
- log-normal model: the volatility-of-volatility is proportional to the level of the volatility
- CEV-type model: the volatility-of-volatility is proportional to the power of the volatility

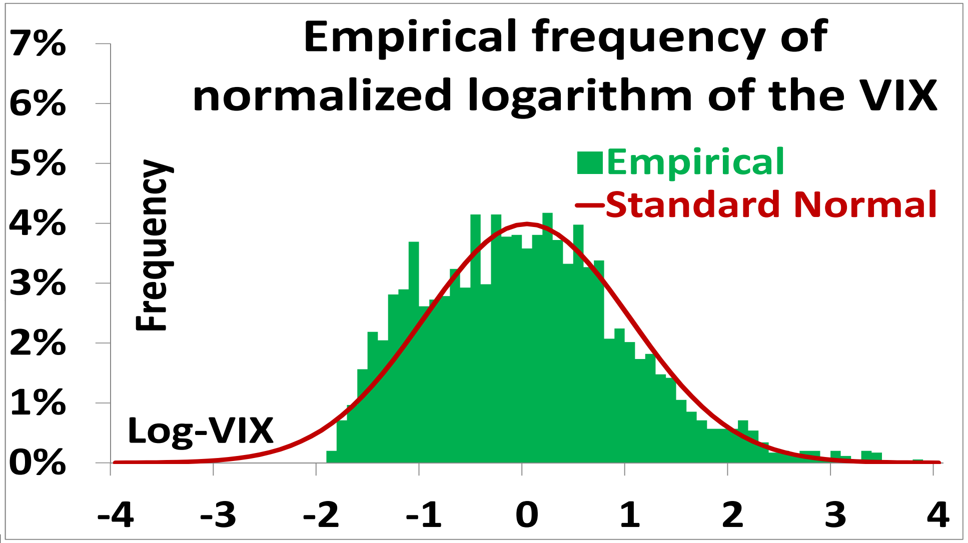
For a simple empirical check for **implied volatilities**: compute the empirical frequency of one-month implied at-the-money (ATM) volatility proxied by the VIX index for last 20 years

Daily observations normalized to have zero mean and unit variance

Left figure: empirical frequency of the VIX: it is definitely not normal

Right figure: frequency of the logarithm of the VIX: it is close to the normal density (especially the right tail)





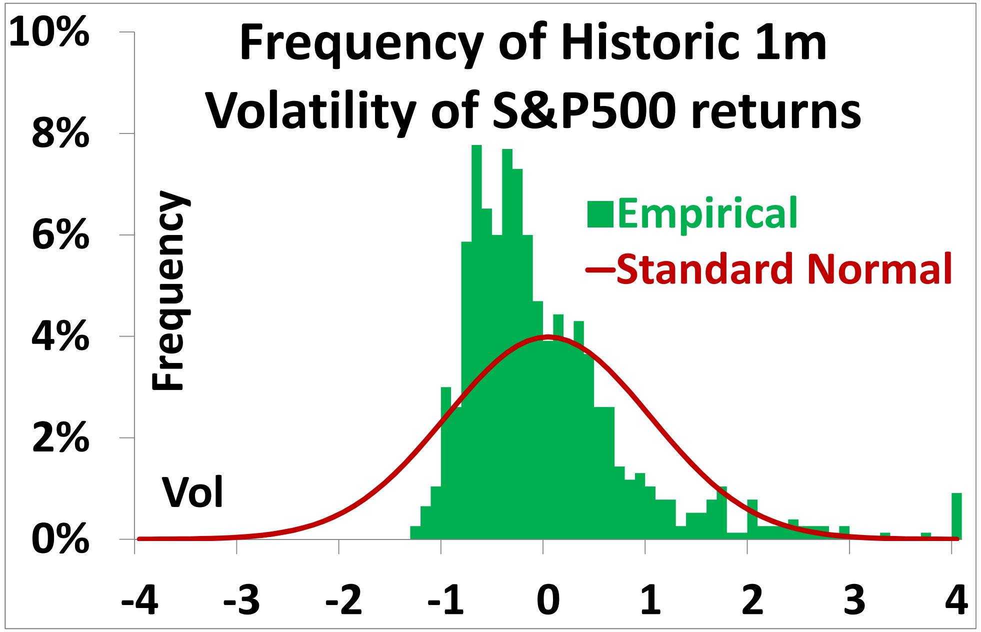
The distribution of implied volatility is also close to log-normal

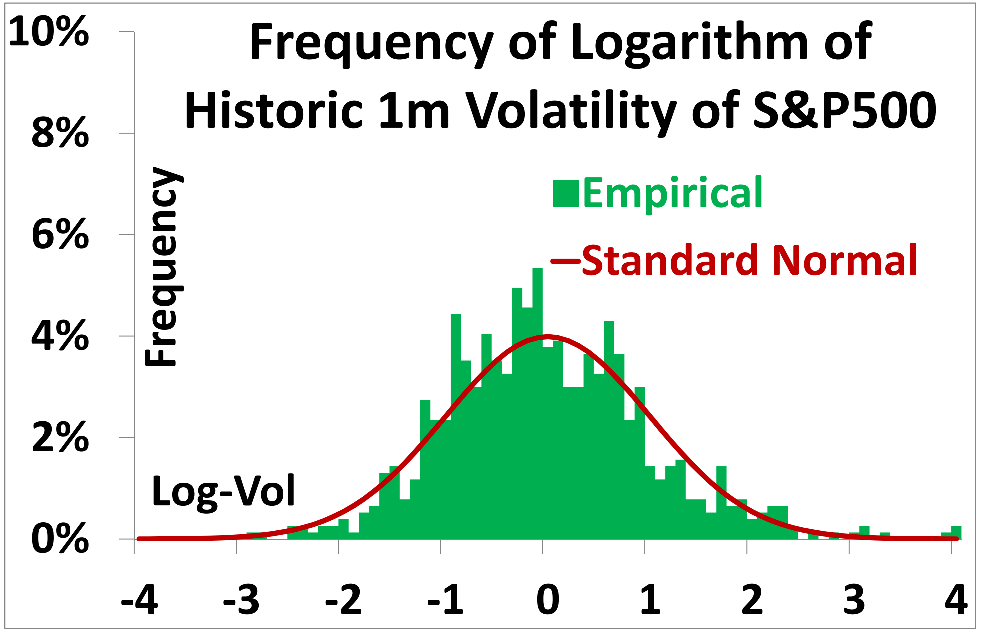
For a simple empirical check for **realized volatilities**: compute one-month realized volatility of daily returns on the S&P 500 index for each month over non-overlapping periods for last 60 years from 1954

Below is the empirical frequency of normalized historical volatility

Left figure: frequency of realized vol - it is definitely not normal

Right figure: frequency of the logarithm of realized vol - again it does look like the normal density (especially for the right tail)





Beta SV model is a factor model for volatility employing the regression model for changes in volatility (realized or implied) $\sigma(t_n)$ **predicted by returns in price** $S(t_n)$:

$$\sigma(t_n) - \sigma(t_{n-1}) = \beta \left[\frac{S(t_n) - S(t_{n-1})}{S(t_{n-1})} \right] + \sigma(t_{n-1})\epsilon_n \tag{5.1}$$

iid normal residuals ϵ_n are scaled by the level of volatility $\sigma(t_{n-1})$ due to log-normality

The regression has a very strong explanatory power explaining up to 80% of variations in volatility

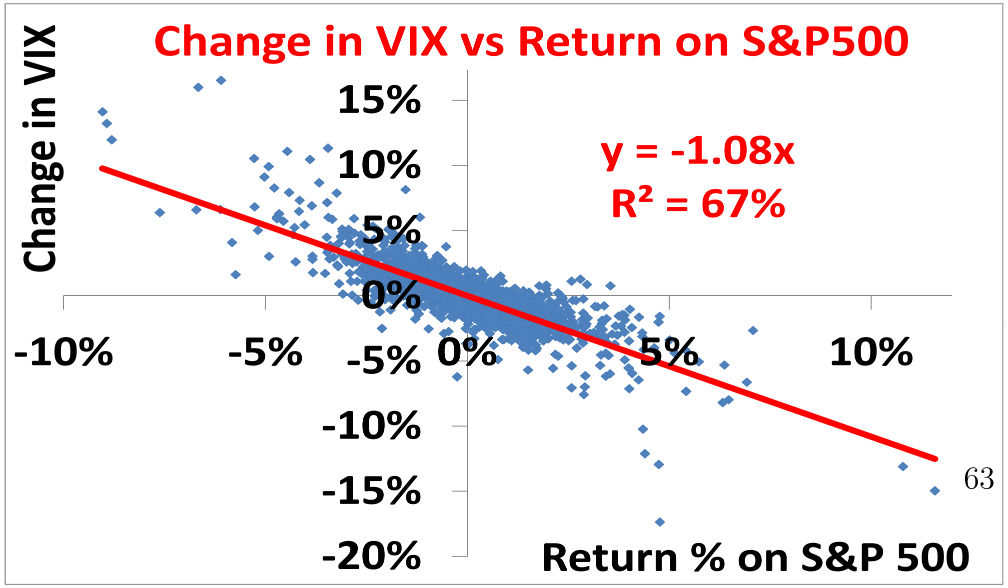
Volatility beta β is negative around -1.0

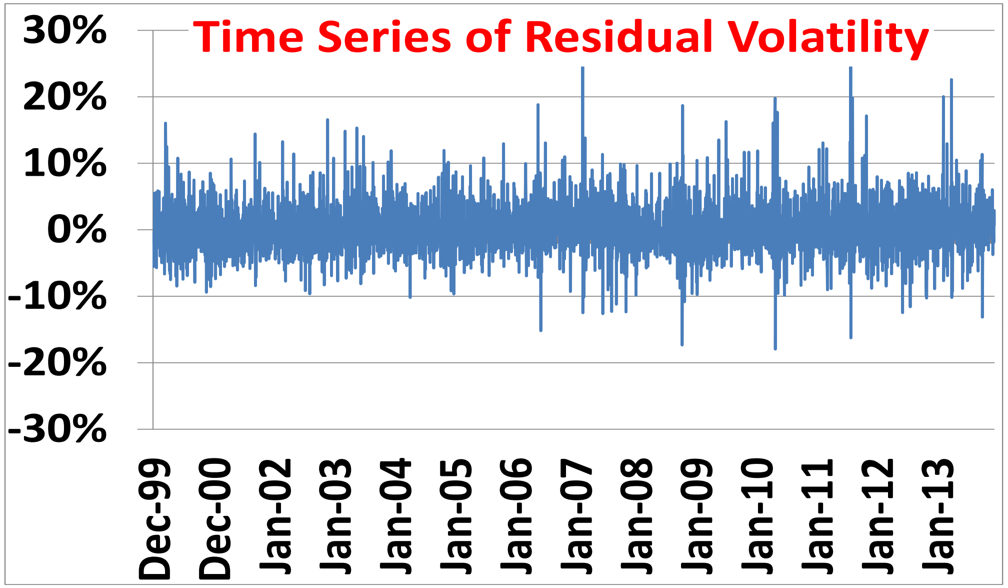
Interpretation of the volatility beta β: expected change in ATM vol predicted by price return: for vol beta of -1.0 and price-return of -2.0%: expected change in vol $= -1.0 \times (-2\%) = 2\%$

Left figure: scatter plot of daily changes in the VIX vs returns on S&P 500 for past 14 years

Right: time series of empirical residuals ϵ_n of regression model (5.1):

- Residual volatility does not exhibit any systemic patterns
- Regression model is stable across different estimation periods





More evidence on log-normal dynamics of volatility: independence of regression parameters on level of ATM volatility

Estimate empirically the elasticity α of volatility by:

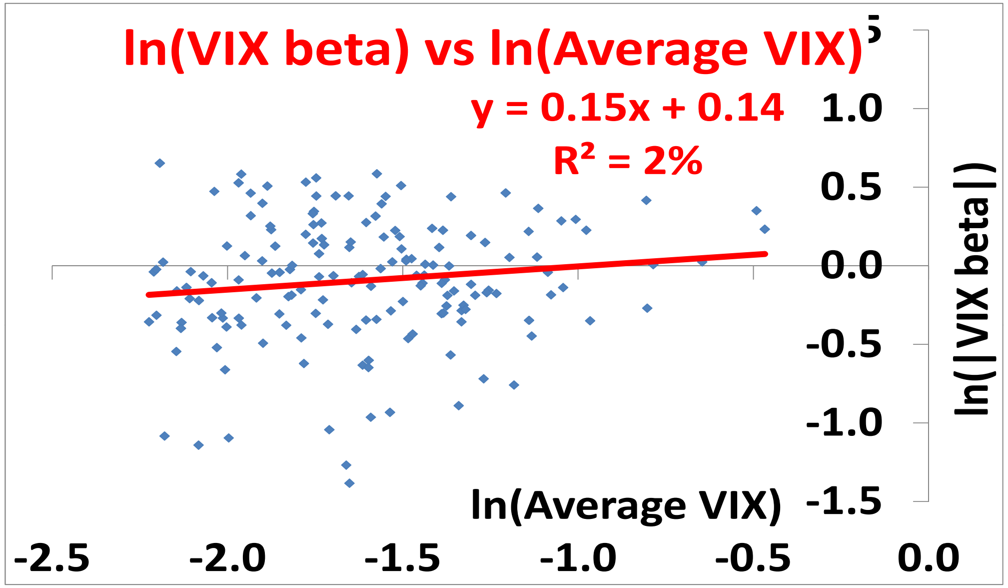
1. Computing the volatility beta and residual volatility-of-volatility for each month using daily returns within this month
2. Test if the logarithm of these variables depends on the log of the VIX in that month using regression model

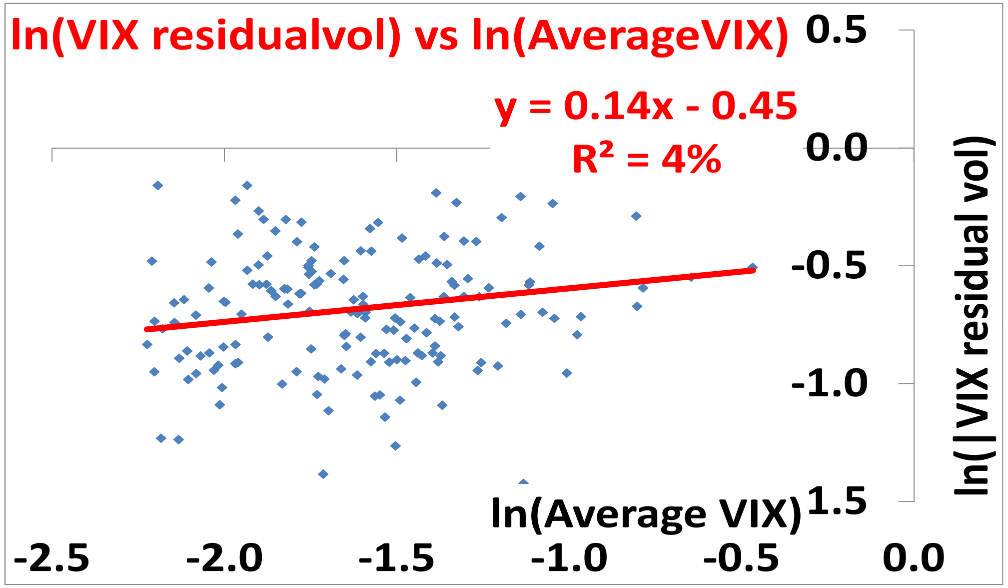
Left figure: test $\hat{\beta}(V) = \beta V^{\alpha}$ by regression model: $\ln\left|\hat{\beta}(V)\right| = \alpha \ln V + c$

Right: test $\hat{\varepsilon}(V) = \varepsilon V^{1+\alpha}$ by regression model: $\ln|\hat{\varepsilon}(V)| = (1+\alpha)\ln V + c$

The estimated value of elasticity α is small and statistically insignificant

Indeed the realized volatility is close to log-normal





Beta stochastic volatility model (Karasinski-Sepp 2012): is obtained by summarizing our empirical findings for dynamics of asset price S_t and volatility σ_t:

$$
\begin{aligned}
dS_t &= \sigma_t S_t dW_t^{(0)} \\
d\sigma_t &= \beta \frac{dS_t}{S_t} + \varepsilon \sigma_t dW_t^{(1)} + \kappa(\theta - \sigma_t)dt
\end{aligned}
\tag{5.2}
$$

σ_t is either returns volatility or short-term ATM implied volatility

$W_t^{(0)}$ and $W_t^{(1)}$ are independent Brownian motions

β is volatility beta - sensitivity of volatility to changes in price

ε is residual volatility-of-volatility - standard deviation of residual changes in vol

Mean-reversion rate κ and volatility mean rate θ are incorporated for the mean-reverting feature and the stationarity of volatility

We arrived to beta SV model (5.2) only by looking at empirical data for realized&implied volatilities and using factor model for the volatility dynamics

Next we consider the econometric estimation of model parameters

5.2 Econometric Estimation

The instantaneous volatility of price returns is not observable directly

To estimate the model parameters for the volatility of price returns, the econometric literature provides the methods:

1. Inference type of methods where the state of the volatility dynamics is inferred using some type of Bayesian or filtering approach: it is relatively computationally expensive
2. Maximum likelihood estimation using probability density function for price returns

Instead, I apply a regression based model

- Compute the time series of the realized volatility using an estimator employing intra-day price data
- The realized volatility is estimated using weekly, monthly or quarterly sampling
- Model parameters are calibrated using the time series of the realized volatility

I apply Euler discretization of the dynamics (5.2) to model change in volatility given price return as follows:

$$\sigma_n - \sigma_{n-1} = (\theta - \sigma_{n-1})\left(1 - e^{-\kappa dt_n}\right) + \beta x_n + o_n \sigma_{n-1}\zeta_n \tag{5.3}$$

where σ_n denotes volatility observed at the estimation times

dt_n, $dt_n = t_n - t_{n-1}$, is time between estimation

ζ_n are iid normals

x_n is realized price return for period t_n: $x_n = \frac{S(t_n)}{S(t_{n-1})} - 1$

o_n is the integrated volatility of volatility defined by:

$$o_n = \varepsilon\sqrt{\frac{1}{2\kappa}\left(1 - e^{-2\kappa dt_n}\right)}$$

Using Eq (5.3), we introduce the normalized variable d_n:

$$d_n = \frac{\sigma_n - \sigma_{n-1} - (\theta - \sigma_{n-1})\left(1 - e^{-\kappa dt_n}\right) - \beta x_n}{o_n \sigma_{n-1}} \tag{5.4}$$

The logarithm of the likelihood function is given by:

$$\begin{aligned} L(\Omega; x, v) &= \frac{1}{N}\sum_{n=1}^{N} l_n(\Omega; x_n, \sigma_n, \sigma_{n-1}) \\ l_n(\Omega; x_n, v_n, v_{n-1}) &= -\frac{1}{2}(d_n)^2 - \ln\left(o_n \sigma_{n-1}\right) - \ln\left(2\pi\right) \end{aligned} \tag{5.5}$$

where Ω is the set of estimation parameters: $\Omega = \{\beta, \epsilon, \kappa, \theta\}$ with estimates are obtained by maximizing the value of ML function (5.5)

Calibration of the mean-reversion parameter κ:

- As I explained when discussing the auto-correlation of realized volatility and smoothing parameter for EWMA model, letting the mean-reversion parameter κ to be a free parameter does not add to the explanatory power of the model
- Fitted values of the mean-reversion parameter κ and residual volatility ϵ are dependent on each other
- The mean-reversion parameter κ is estimated beforehand using the empirical auto-correlations of the volatility and set fixed to ML problem (5.5)

The proposed method does account for jumps in returns because we assume that price returns, $\{x_n\}$, are sampled from empirical distribution and we make no assumption on distribution of $\{x_n\}$

Instead, using Eq (5.3), we make assumption on normality of the distribution of changes in volatility, $\sigma_n - \sigma_{n-1}$, conditional on observed log-returns and previous volatility

5.3 Empirical Tests of Forecasting Power of Beta SV model

To test the predictive power of the volatility forecast for SV beta model, I design the back-test similarly to the GARCH model as follows:

- Consider three estimation windows: weekly, monthly, and quarterly
- At the end of estimation window, the parameters of the model are re-calibrated using past 3 years of data and volatility time series computed using 7 estimators
- The forecast for e estimators is computed for the next period
- Evaluate the forecast volatility against the actual volatility computed using 7 estimators.
- Make the statistical test about the quality of the forecast volatility by applying the following regression model:

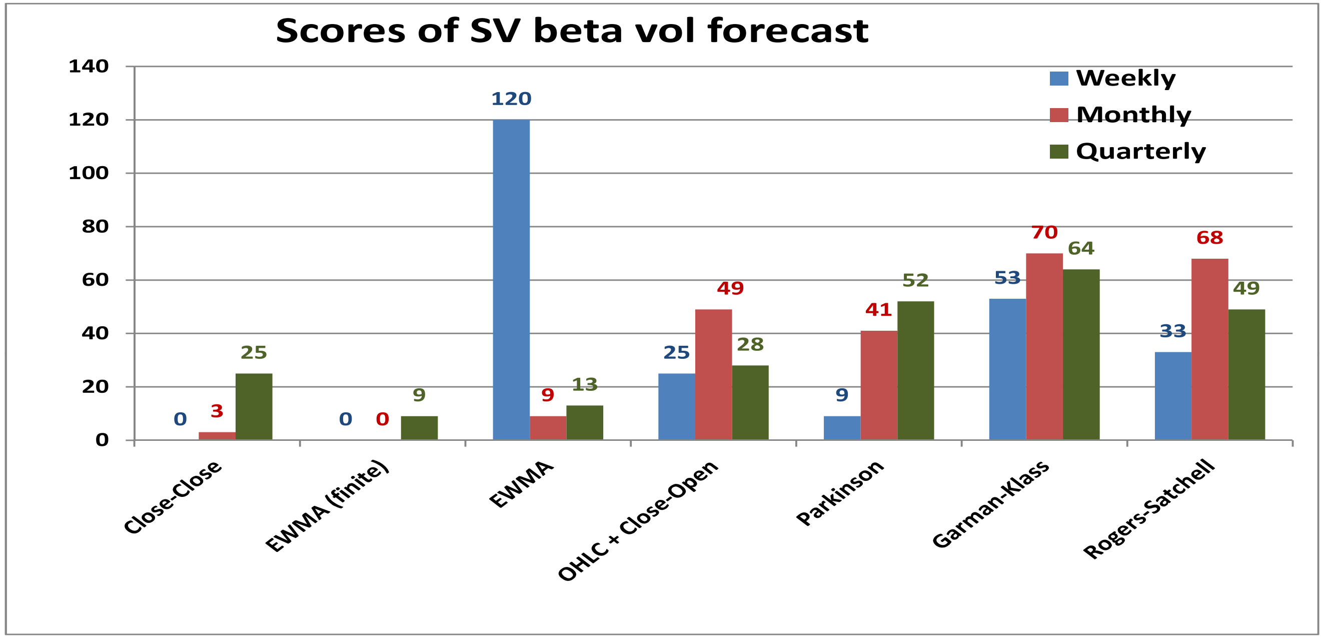
$$\widehat{\sigma}_{t+1}^{\text{actual}}(T) = \alpha + \beta \widehat{\sigma}_{t}^{SV\,beta}(T) + \epsilon_{t+1}$$

- Use the same data set of 40 assets with 14+ years of time series data from 1-January-2002 (the first 3 years are used for "warm-up" of the calibration) up to 1-April-2016
- Use the same scoring methodology for volatility estimator with the most close forecast to the beta SV model
- Aggregate results for weekly, monthly and quarterly forecast

Total scores of SV beta forecasting power:

- Model forecast for weekly volatilities is the most consistent with EWMA volatility
- For monthly and quarterly forecasts, SV beta produces consistent forecast for estimators employing the intra-day data

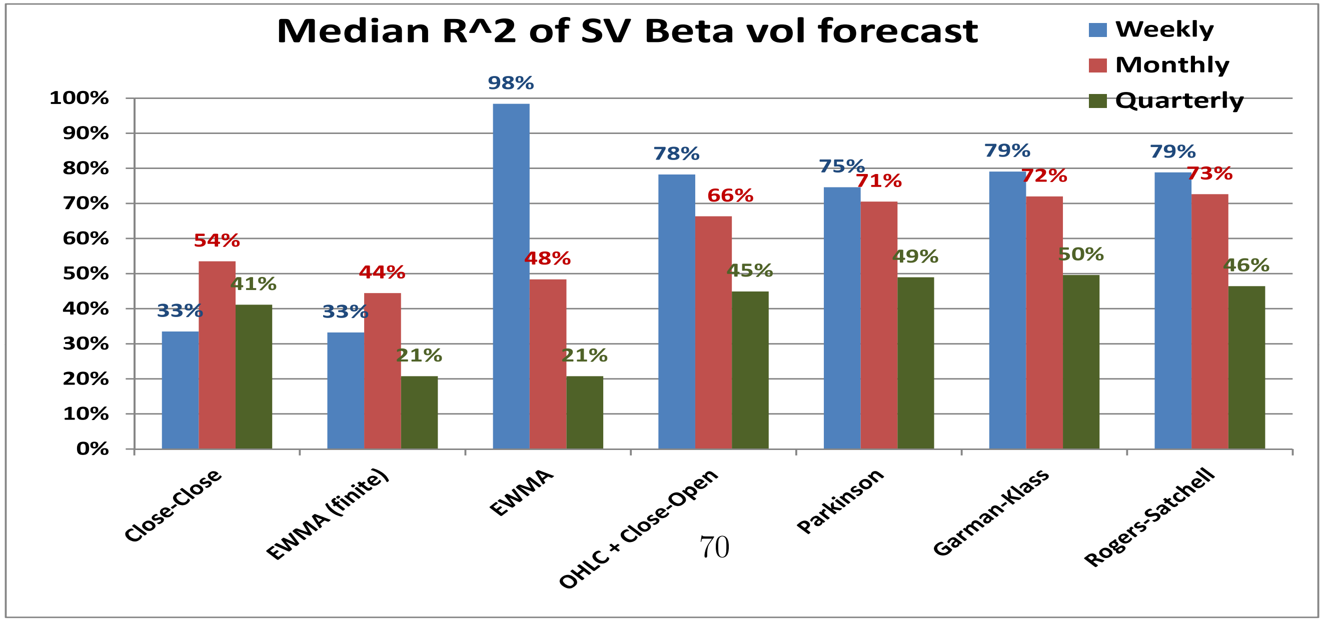
	Weekly	Monthly	Quarterly
Close-Close	0	3	25
EWMA (finite)	0	0	9
EWMA	120	9	13
OHLC + Close-Open	25	49	28
Parkinson	9	41	52
Garman-Klass	53	70	64
Rogers-Satchell	33	68	49



Median R^2 of SV beta forecasting power

- The SV beta has strong explanatory power for weekly and monthly realized volatilities
- The forecast power for longer quarterly periods is modest
- In comparison to the GARCH model, beta SV implies the explanatory power R^2 higher by $10\% - 15\%$

	Weekly	Monthly	Quarterly
Close-Close	33%	54%	41%
EWMA (finite)	33%	44%	21%
EWMA	98%	48%	21%
OHLC + Close-Open	78%	66%	45%
Parkinson	75%	71%	49%
Garman-Klass	79%	72%	50%
Rogers-Satchell	79%	73%	46%



5.4 Conclusions

- The beta SV model incorporates the mean-reversion feature and the leverage between volatility and price returns
- Model can be also estimated using:
  1. The econometric data of short-term implied volatilities
  2. Options data (implied calibration)
- Model estimation and implementation are relatively straight forward

6 Trading Volatility Risk-Premium

Once we identified how to measure and forecast realized volatility, we can make comparison between realized and implied volatilities

For volatility trading:

- Realized volatility is a measure of price-returns volatility to replicate option pay-offs by delta-hedging
- Implied volatility is a measure of volatility at which options are quoted/traded in the market (Black-Scholes volatility)
- The volatility premium is broadly defined as the spread between implied and realized volatilities:

$$\text{Volatility Risk-premium} = \text{Implied volatility} - \text{Realized volatility}$$

Volatility risk-premium can be realized by selling options and delta-hedging

- Selling options at implied volatility indicates the expected costs to replicate options: the premium is received at trade inception and the price of option premium is primarily derived from the implied volatility
- Replication costs are the cost incurred by delta-hedging options to maturity: the higher is the realized volatility - the higher are the replication costs so the total value of replication costs is primarily derived from the realized volatility
- The realized P&L from selling and delta-hedging options is

$$\begin{aligned}\text{Short Volatility P\&L} &= \text{Implied volatility from Selling} - \text{Realized volatility from Delta-Hedging}\\ &= \text{Volatility Risk-premium}\end{aligned}$$

If market were efficient and investors were risk-neutral, the average value of the volatility risk-premium would be close to zero over long-term periods

However, the average value of the volatility risk-premium is positive over long-term periods and it is relatively large so that it can serve as an investment strategy in itself

Why the volatility risk-premium exists is an interesting question which is also important to understand the drivers behind volatility trading: I cover it later

6.1 Quantifying Realized and Expected Volatility Risk-Premium

Consider the strategy that sells options with maturity of one month every 3rd Friday of the month and delta-hedges options during the month up to their expiry, and then rolls into new delta hedged position and so on

The realized P&L on this strategy every month is closely connected to the realized volatility premium:

$$\begin{aligned}\text{Realized Volatility Risk-Premium} &= \text{Implied volatility from Selling at 3rd Friday}\\ &- \text{Realized volatility during month up to the next 3rd Friday}\end{aligned}$$

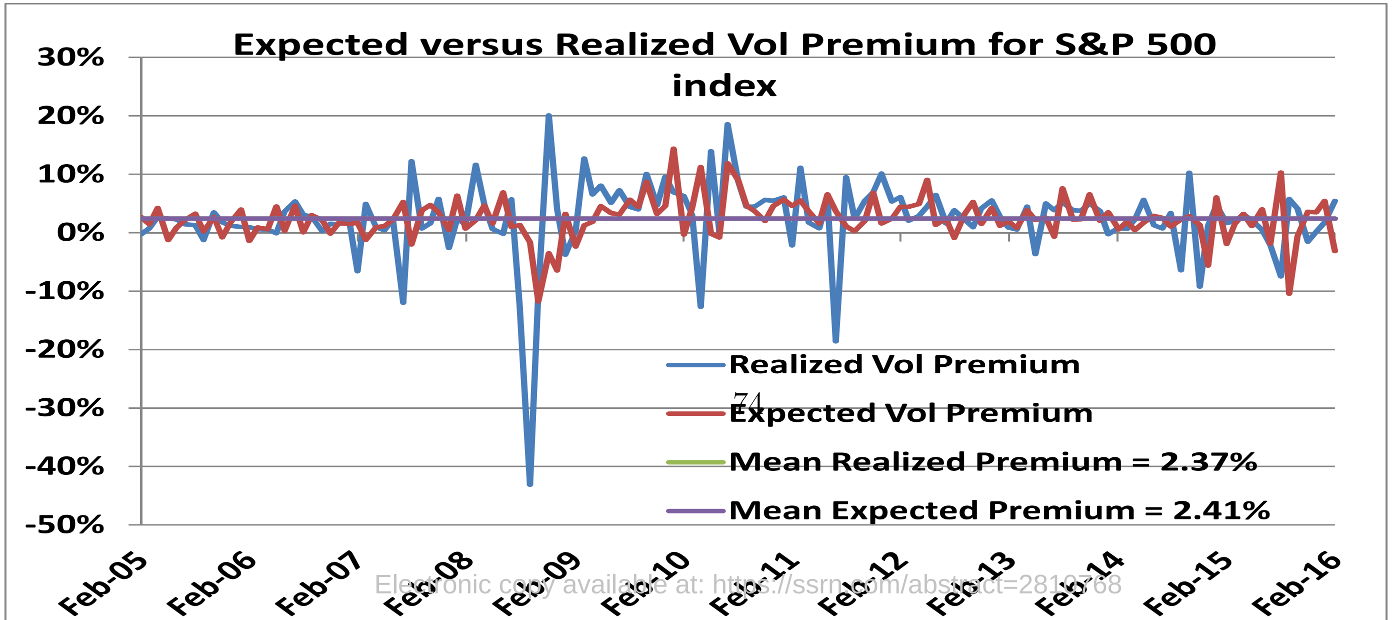
where the realized volatility can be measured by any of the sample space estimators

Also, as a trading signal we consider the expected volatility risk-premium at the beginning of the 3rd Fridy right before the trade inception:

$$\begin{aligned}\text{Expected Volatility Risk-Premium} &= \text{Implied volatility from Selling at 3rd Friday}\\ &- \text{Expected volatility for coming month up to the next 3rd Friday}\end{aligned}$$

where the forecast for realized volatility can be made using sample space estimators, GARCH type models, or the beta SV model

For the S&P 500 index on average both realized and expected vol premiums are positive



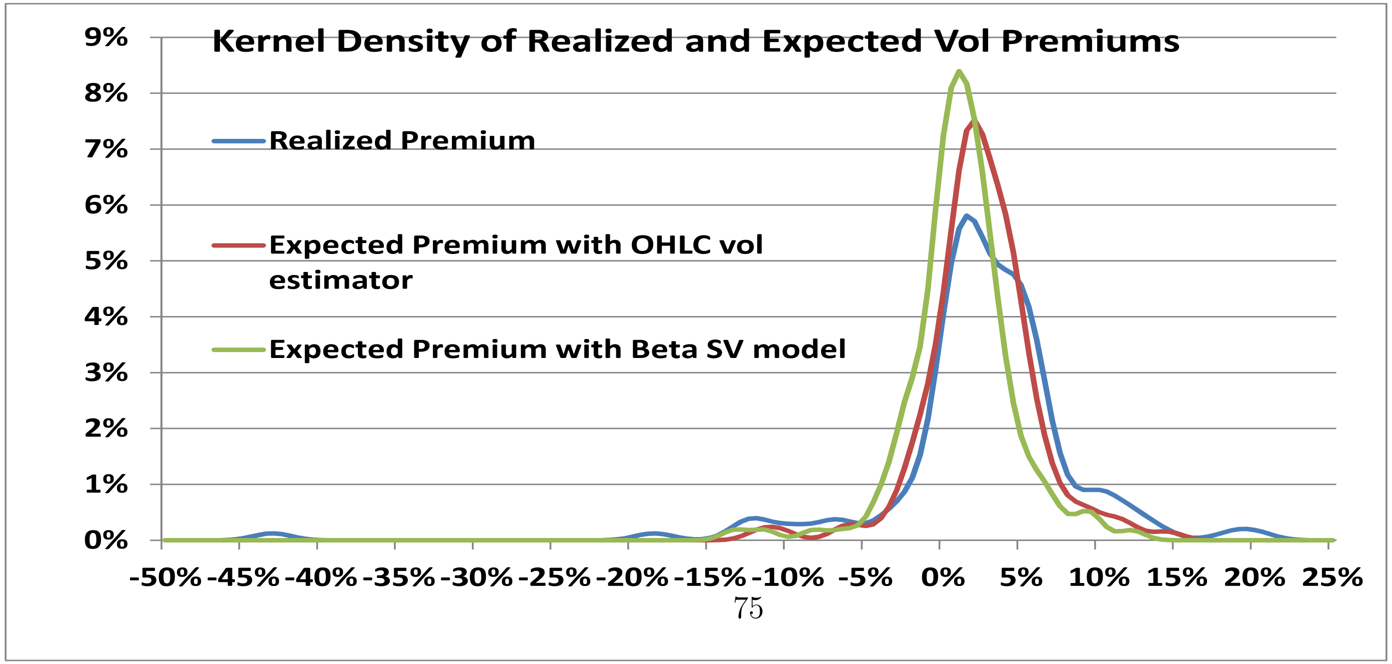
For computing the estimate of the expected volatility risk-premium, we use the forecast volatility computed using a selected method or model

As a result, it is important that the selected model for volatility forecast has a high explanatory power

A robust estimator should produce the empirical probability density of the expected risk premium with a higher peak and less skewed to the right to be more conservative

Illustration of the kernel empirical density (using past 11 years) for

- The realized volatility premium
- The expected volatility premium computed using OHLC estimator
- The expected volatility premium computed using beta SV model



6.1.1 Back-test for Monetizing Volatility Risk Premium

Straddle is a typical strategy for monetizing the volatility risk-premium

Short Straddle = short at-the-money put and call options

Left Figure: Fund value shorting 1 month straddle on S&P500 index ETF (SPY) rolled every month and delta-hedged at end of day from 2005 to 2016

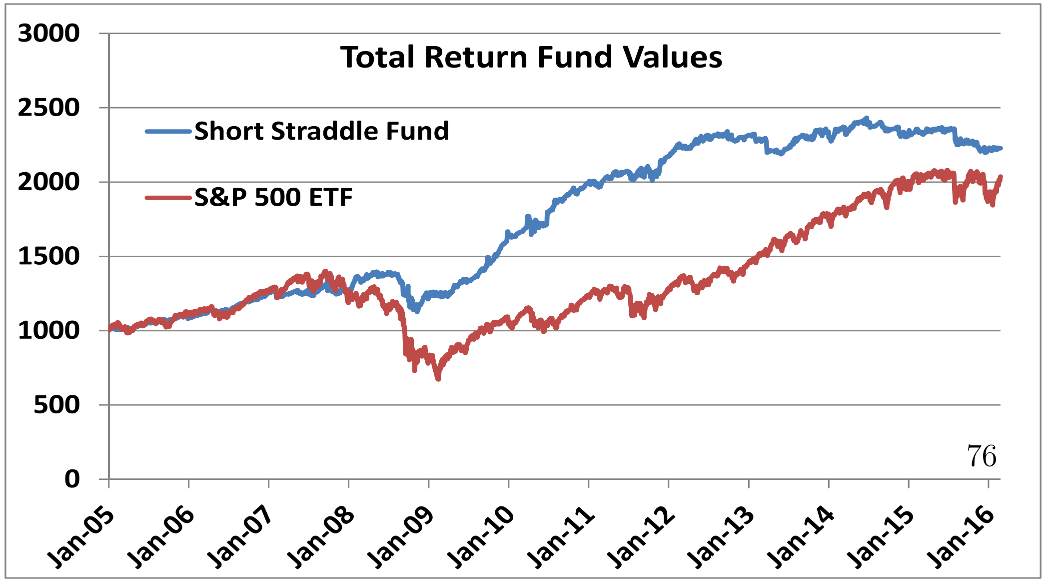
Right Figure: fund running % drawdowns compared to S&P 500 index ETF

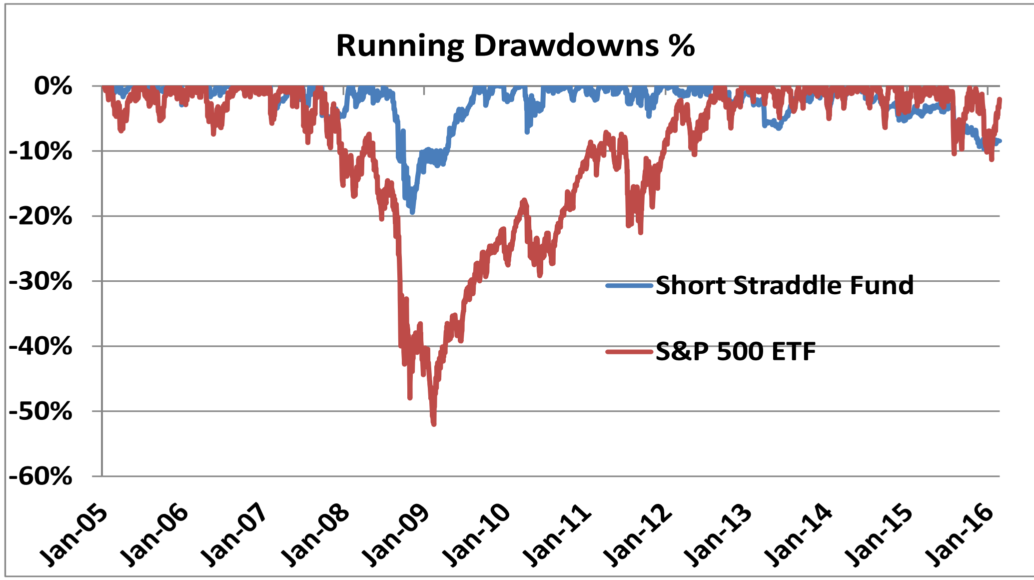
Number of sold option contracts at monthly rolls are set to

$$\text{Number contracts} = 2\frac{\text{Fund Value}}{\text{Spot Price}}$$

Performance vs S&P 500 index ETF

	An. Return	Stdev	Sharpe	Max DrawDown
Short Straddle	7.43%	8%	0.88	-19%
S&P 500	6.58%	18%	0.36	-52%





6.2 Quantifying Implied Skew Risk-Premium

The volatility risk-premium broadly refers to the spread between at-the-money implied volatility and realized statistical volatility

The volatility skew measures the degree of how expensive are out-of-the-money options relative to ATM options measured in terms of implied volatilities

For stock index options:

- Put options are relatively more expensive than ATM options
- Call options are relatively less expensive than ATM options

To measure the volatility skew for puts we can apply the following formulas:

1. Constant %-strike difference:

$$\text{Put 95\% skew} = \text{ATM Vol} - \text{ Put Volatility(Strike=95\% of Spot)}$$

2. Constant Delta difference:

$$\text{Put 25-Delta skew} = \text{ATM Vol} - \text{ Put Volatility(-25 Delta Strike)}$$

3. One standard deviation from the Spot:

$$\text{Put 5\% skew} = \text{ATM Vol} - \text{ Put Volatility(Strike=Spot} \times e^{-\sigma_{ATM}\sqrt{T}})$$

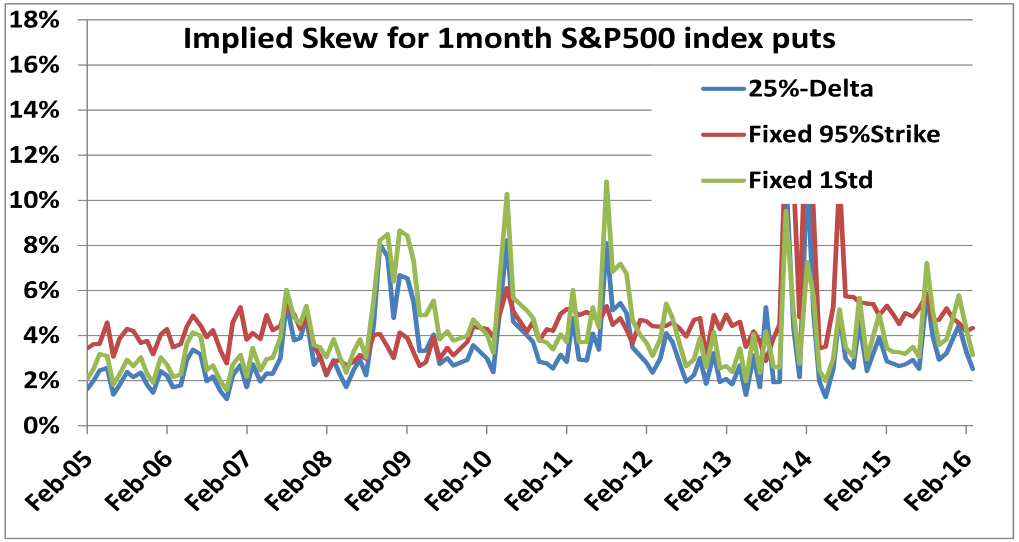
   where T is time to maturity and σ_{ATM} is ATM volatility for this maturity

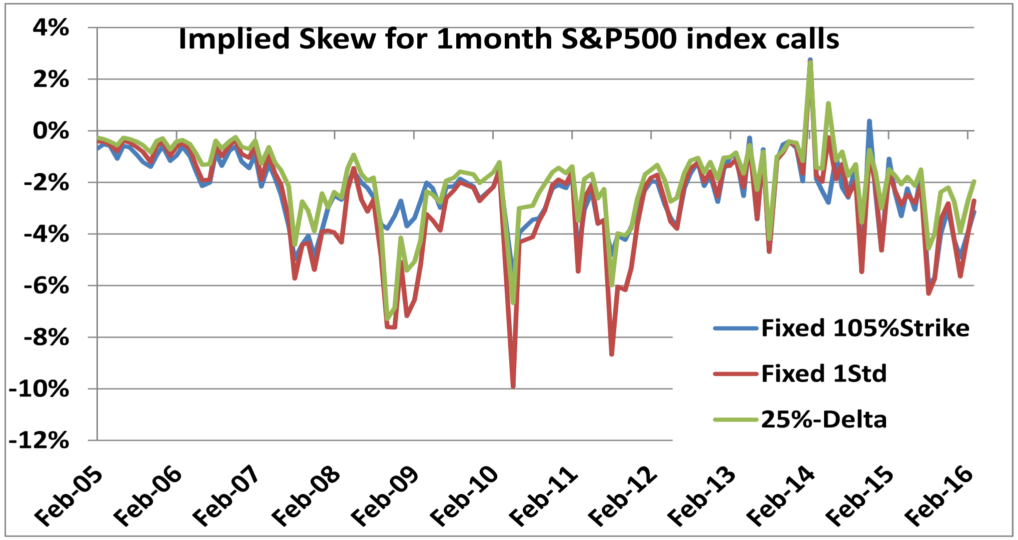
Similarly we define the skew for OTM call options

Put skew for S&P500 index ETF is persistent

The average values for past 11 years:

	Put	Call
25%-Delta	3.20%	-1.86%
Fixed 95%/105% Strike	4.38%	-2.30%
Fixed 1Std	3.96%	-2.58%





6.3 Quantifying Realized and Expected Skew Risk-Premium

The realized implied volatility skew measures how implied at-the-money changes when the asset changes which can be estimated using the following regression model:

$$\sigma_{ATM}(t_n) - \sigma_{ATM}(t_{n-1}) = \beta \times \left(\frac{S(t_n)}{S(t_{n-1})} - 1 \right) + \epsilon_n \tag{6.1}$$

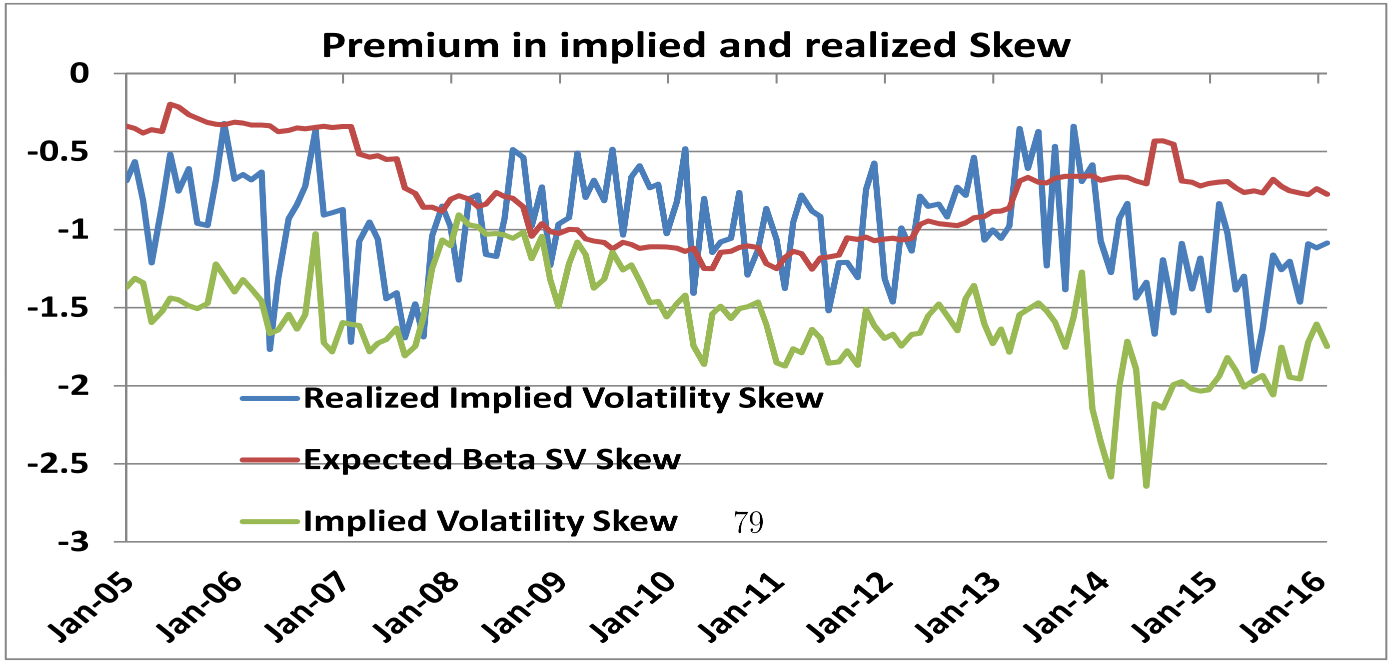
where estimated parameter β measures the realized implied volatility skew

When instead of implied skew we measure change in realized volatility, we apply the estimate of the volatility beta in the beta SV model which serves as a measure of expected skew

Finally, the implied volatility skew is implied from put skew:

$$\text{Implied Volatility skew} = 2 \times \text{Put } 95\% \text{ skew}/5\%$$

The implied volatility skew is consistently and significantly larger than statistical measures of the realized skew



6.3.1 Back-test for Monetizing Skew Risk Premium

Strangle is a typical strategy for monetizing the skew risk-premium

Short Strangle = short out-the-money put and call options

Left Figure: Fund value shorting 1 month Strangle on S&P500 index ETF (SPY) rolled every month and delta-hedged at end of day from 2005 to 2016
Right Figure: fund running % drawdowns compared to S&P 500 index ETF

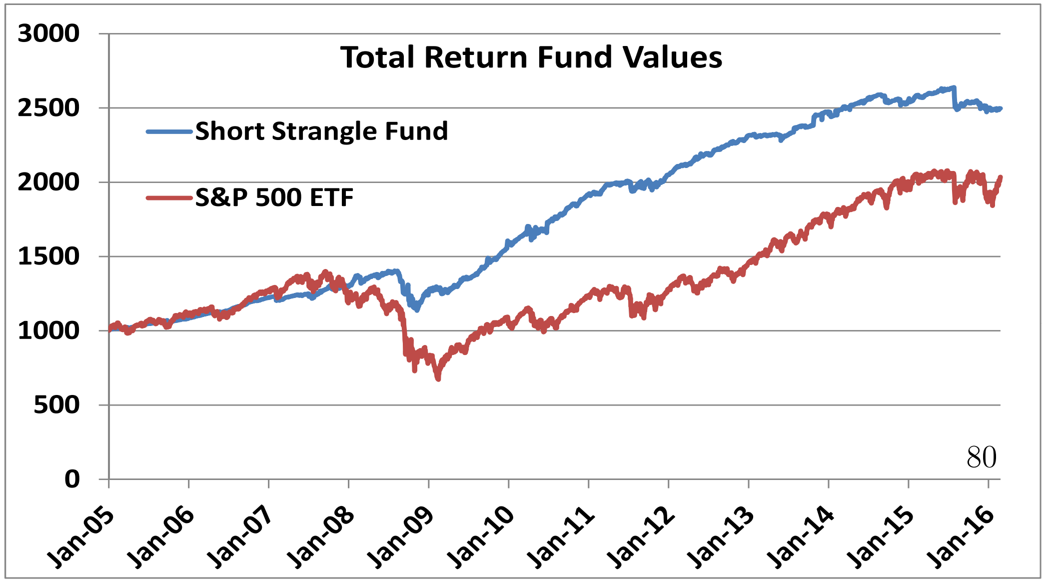
Option strikes at monthly rolls are set to

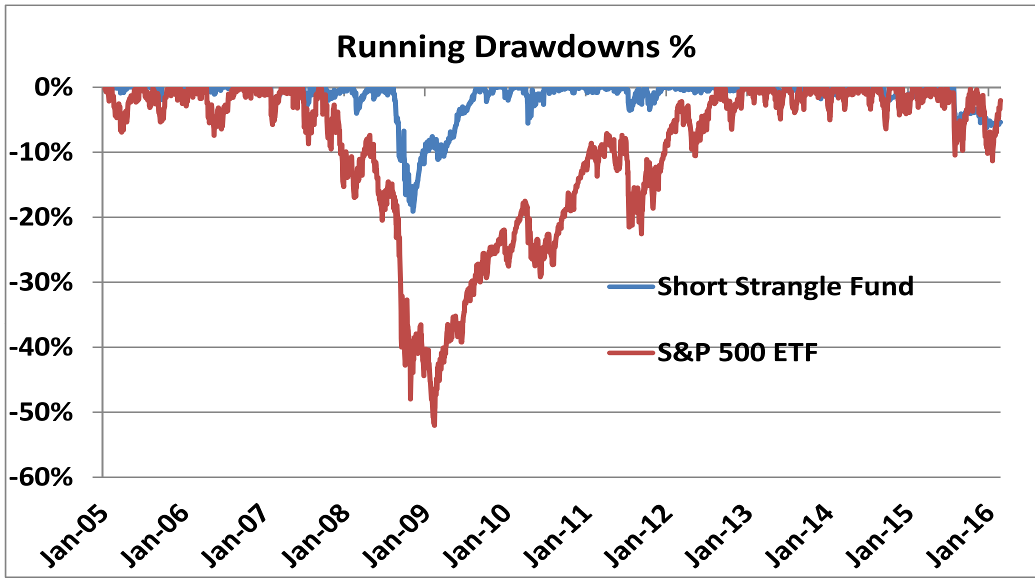
$$PutStrike = Spot \times e^{-\sigma_{ATM}\sqrt{T}}\ ,\ CallStrike = Spot \times e^{+\sigma_{ATM}\sqrt{T}}$$

The number of sold option contracts is set in the same way as for straddle fund

Performance vs S&P 500 index ETF

	An. Return	Stdev	Sharpe	Max DrawDown
Short Straddle	8.55%	7%	1.21	-19%
S&P 500	6.58%	18%	0.36	-52%





6.4 Systematic Volatility Trading Strategies

Typical strategies to monetize volatility premiums:

1. Delta-Hedged strategies:
   - Typically are run by large institutionals
   - Require ability to trade cheaply in the underlying asset
   - Typically lack systematic approach: the flow is client driven, often involves some discretionary decisions about timing, sizing, delta adjustments
2. Naked option selling strategies:
   - Researched and promoted by CBOE
   - Marketed and offered by some investment banks for buy side investors
   - Follow systematic approach: regular re-balancing, rule-based position sizing

6.5 Analysis of Delta-Hedging Cost vs Risk

While delta-hedging reduces risks it also spends a lot on transaction costs

Provided are backtests of four strategies using S&P 500 index ETF from 2005 to 2016:

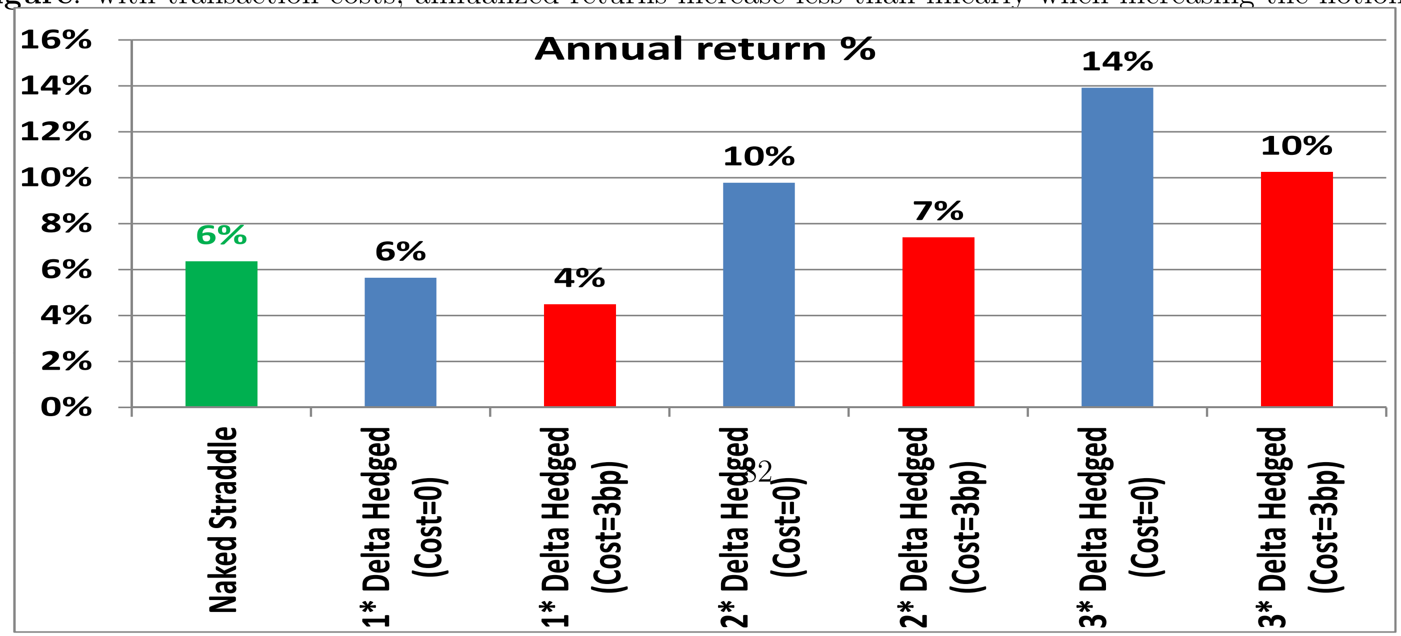
1. Selling naked ATM straddles every month with number of sold contracts set at every roll to

$$\text{Number contracts} = \frac{\text{Fund Value}}{\text{Spot Price}}$$

2. Selling delta-hedged ATM straddles every month with number of sold contracts set at every roll as for 1st strategy
3. Selling delta-hedged ATM straddles with double of the notional
4. Selling delta-hedged ATM straddles with triple of the notional

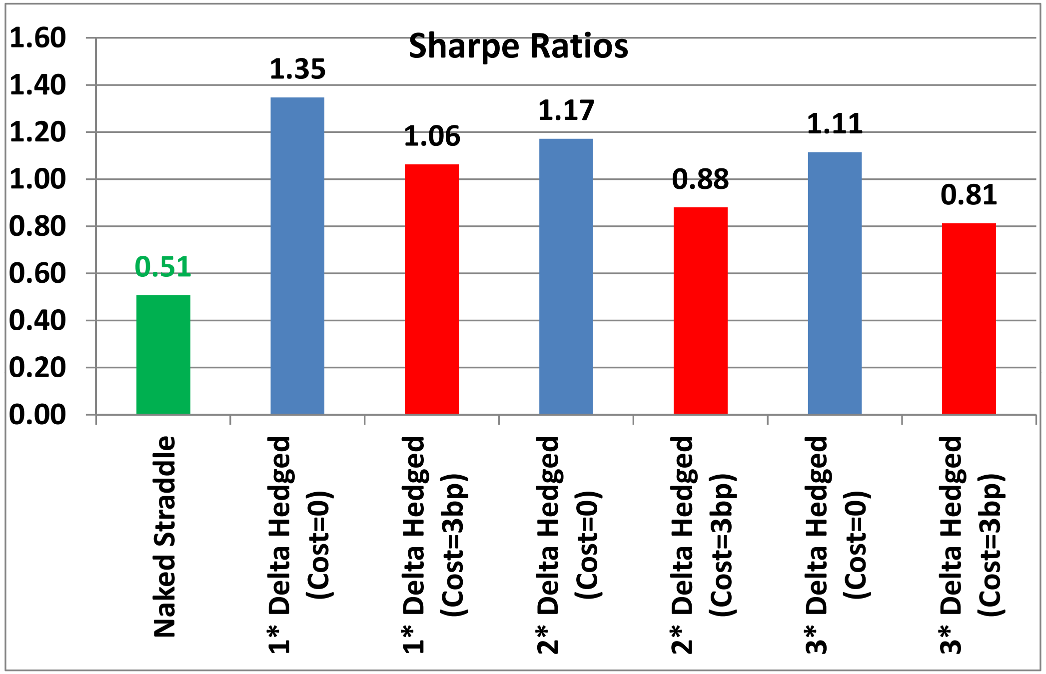
For delta-hedge strategies, the delta-hedging takes place by the end of each trading day and transaction costs are $3bp$ (typical for large cap ETFs)

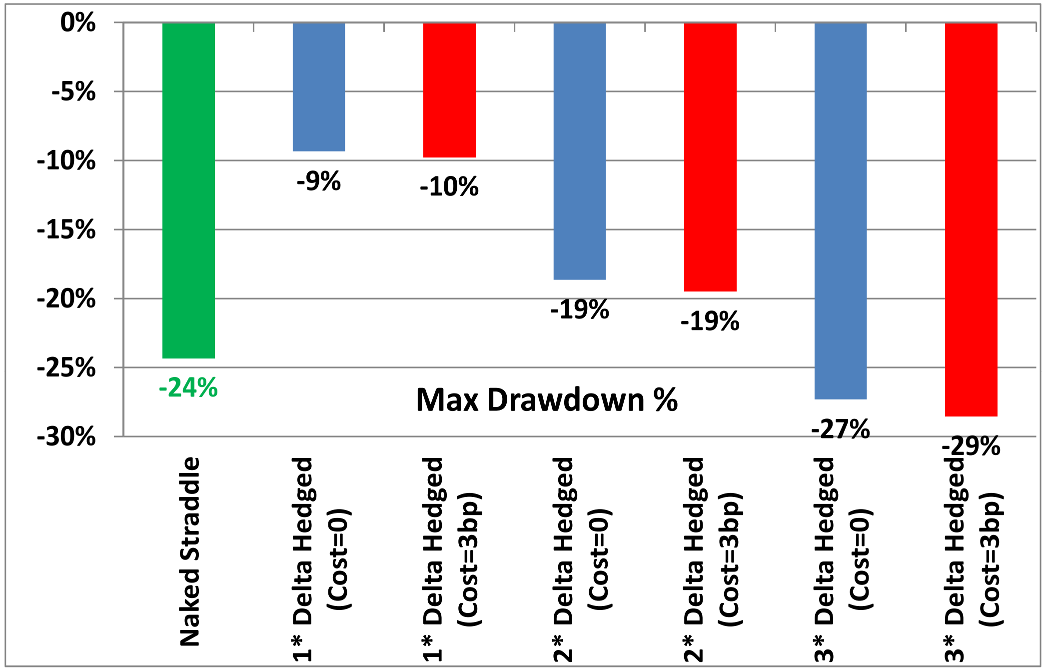
Figure: with transaction costs, annualized returns increase less than linearly when increasing the notional



However when increasing the trade notional, the risk is increasing linearly:

- The Sharpe ratio starts declining
- The realized maximum drawdown increases
- With higher notional, the drawdown risk of delta-hedged strategies is similar to less leveraged naked short option strategies





6.6 Conclusions for Design of Systematic Volatility Trading Strategies

Both delta-hedged and naked short strategies can be implemented in an automated rule-based way and offered to clients

The decision whether implement delta-hedging or not should be based on the risk - reward considerations with regard to transaction costs and tolerance to delta and vega risk

For full automation of delta-hedged strategies the following analytics needs research and implementation:

- Computation of option delta in an optimal way to partially offset vega risks
- Rule-based timing of re-hedging times in order to minimize transaction costs without increasing delta risk too much

Next I consider these two topics

7 Computing Option Delta Consistently with Empirical Dynamics

To compute option delta consistently with observed dynamics of the volatility surface we need to model the dynamics of the volatility surface

Dimensionality: how many factors needed for modeling vol surface ?

Left table: PCA for covariance matrix of weekly changes in ATM vols (1m,2,3m,6m,1y) and returns with 6 factors (variables have scalable vols) - two factors explain more than 90% of changes in returns and ATM vols

Right table: PCA for correlation matrix of changes in ATM vols, skews, convexities, and returns with total of 16 factors - need about four factors to explain 85% of variability - need extra factors to model changes in skew and convexity

Factor	SPX	FTSE	Nikkei	Stoxx
1	93%	89%	56%	86%
2	**97%**	**94%**	**74%**	**94%**
3	100%	97%	85%	98%
4	100%	99%	94%	99%
5	100%	100%	99%	100%
6	100%	100%	100%	100%

Factor	SPX	FTSE	Nikkei	Stoxx
1	42%	34%	34%	39%
2	63%	59%	50%	69%
3	77%	75%	64%	85%
4	**83%**	**85%**	**75%**	**90%**
5	88%	92%	82%	93%
6	91%	94%	87%	96%
7	94%	96%	91%	97%
8	95%	98%	94%	98%
9	97%	99%	96%	99%
10	98%	99%	97%	100%
11	99%	100%	98%	100%
12	99%	100%	99%	100%
13	100%	100%	100%	100%
14	100%	100%	100%	100%
15	100%	100%	100%	100%
16	100%	100%	100%	100%

7.1 Beta SV model for Parametrization of Implied Volatilities

Apply the beta SV model without mean reversion (equivalent to a modified SABR model):

$$
\begin{aligned}
dS_t &= V_t S_t \left[\frac{S_t}{\overline{S}}\right]^{\beta_{LV}} dW_t^{(0)} \\
dV_t &= \beta_{SV}\frac{dS_t}{S_t} + \varepsilon V_t \left[\frac{S_t}{\overline{S}}\right]^{\beta_{LV}} dW_t^{(1)} \\
&= \overline{\beta}_{SV} V_t dW_t^{(0)} + \overline{\varepsilon} V_t dW_t^{(1)}
\end{aligned}
\tag{7.1}
$$

where $\overline{S}$ is a reference spot, typically $\overline{S} = S_0$

To produce volatility skew and price-vol dependence:

1. β_{LV} is **the local vol or backbone beta**, $\beta_{LV} \leq 0$
2. β_{SV} is **the vol beta**, $\beta_{SV} \leq 0$, which measures the sensitivity of instantaneous vol to price returns independent on assumption about the local vol of the spot (it is just a regression of vol changes explained by returns)

β_{SV} and ε **are vol beta and residual vol in terms of the log-normal basis** for the instantaneous volatility of returns; $\overline{\beta}_{SV}$ and $\overline{\varepsilon}$ **are re-scaled for the local volatility basis**:

$$
\overline{\beta}_{SV} = \beta_{SV}\left[\frac{S}{\overline{S}}\right]^{\beta_{LV}} \ , \ \overline{\varepsilon} = \varepsilon\left[\frac{S}{\overline{S}}\right]^{\beta_{LV}}
$$

Connection to SABR model (notations from Hagan *et al* (2002)):

$$
V_0 = \hat{\alpha} \ , \beta_{LV} = \beta - 1 \ , \ \overline{\beta}_{SV} = \nu\rho \ , \ \overline{\varepsilon} = \nu\sqrt{1-\rho^2}
$$

Approximation to Black-Scholes-Merton (BSM) implied vol $\sigma_{BSM}(S;K)$ is obtained:

$$
\sigma_{BSM}(K;S) = \frac{\ln(S/K)}{f(y)} \ , \ y = \frac{1}{V_0}\frac{\left[\frac{S}{\overline{S}}\right]^{-\beta_{LV}} - \left[\frac{K}{\overline{S}}\right]^{-\beta_{LV}}}{-\beta_{LV}}
\tag{7.2}
$$

$$f(y) = \frac{1}{\sqrt{(\overline{\beta}_{SV})^2 + \epsilon^2}} \ln \left(\frac{J(y)\sqrt{(\overline{\beta}_{SV})^2 + \epsilon^2} + ((\overline{\beta}_{SV})^2 + \epsilon^2)y - \overline{\beta}_{SV}}{\sqrt{(\overline{\beta}_{SV})^2 + \epsilon^2} - \overline{\beta}_{SV}} \right)$$

$$J(y) = \sqrt{\left(1 + \left((\overline{\beta}_{SV})^2 + \epsilon^2\right) y^2 - 2\overline{\beta}_{SV} y\right)}$$

Quadratic expansion for BSM implied vols is obtained by assuming $\beta_{LV} \to 0$ we have $y \to \ln(K/S)$, and limit in log-moneyness $k = \ln(K/S)$ becomes:

$$\sigma_{BSM}(K;S) = \left[\frac{S}{\overline{\overline{S}}}\right]^{\beta_{LV}} \left[V_0 + \frac{1}{2}\left(V_0\beta_{LV} + \overline{\beta}_{SV}\right) k + \frac{1}{12}\left(V_0\beta_{LV}^2 - \frac{(\overline{\beta}_{SV})^2}{V_0} + 2\frac{\overline{\varepsilon}^2}{V_0}\right) k^2\right]$$

Empirical testing for skew components: the local β_{LV} and stoch vol β_{SV} are linked by relationship:

$$\sigma_{ATM}(S)\beta_{LV} + \beta_{SV} = 2 \times \text{SKEW}(S)$$

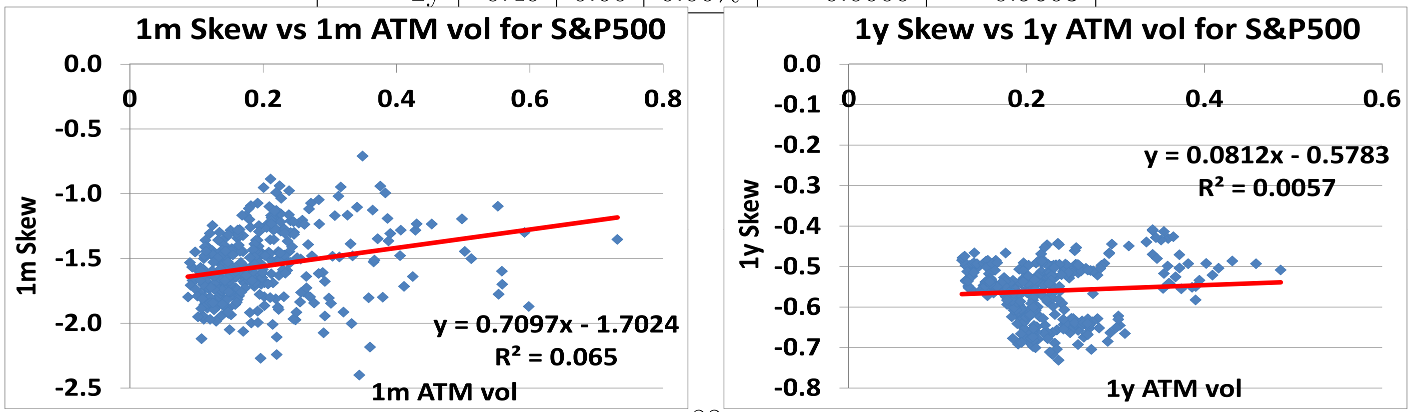
Apply the regression analysis to back-test the empirical relationship between the two using time series of market data

$$2 \times \text{SKEW}_{t_n}(S) = a_0 + a_1 \times \sigma_{ATM,t_n}(S)$$

The significance (small p-values) of a_0 (a_1) means significance of stoch vol (local vol) component to explain the level of the skew

Table&Figure: Estimates for S&P index from 2007 to 2014 - **the empirical link between skew and atm volatility is very weak**

S&P500	$\hat{a}_0$	$\hat{a}_1$	R^2	p-value a_0	p-value a_1
1m	-1.72	0.86	8.79%	0.0000	0.0000
3m	-1.11	0.34	4.31%	0.0000	0.0000
6m	-0.80	0.17	1.56%	0.0000	0.0112
1y	-0.58	0.09	0.52%	0.0000	0.1321
2y	-0.40	0.00	0.00%	0.0000	0.9603

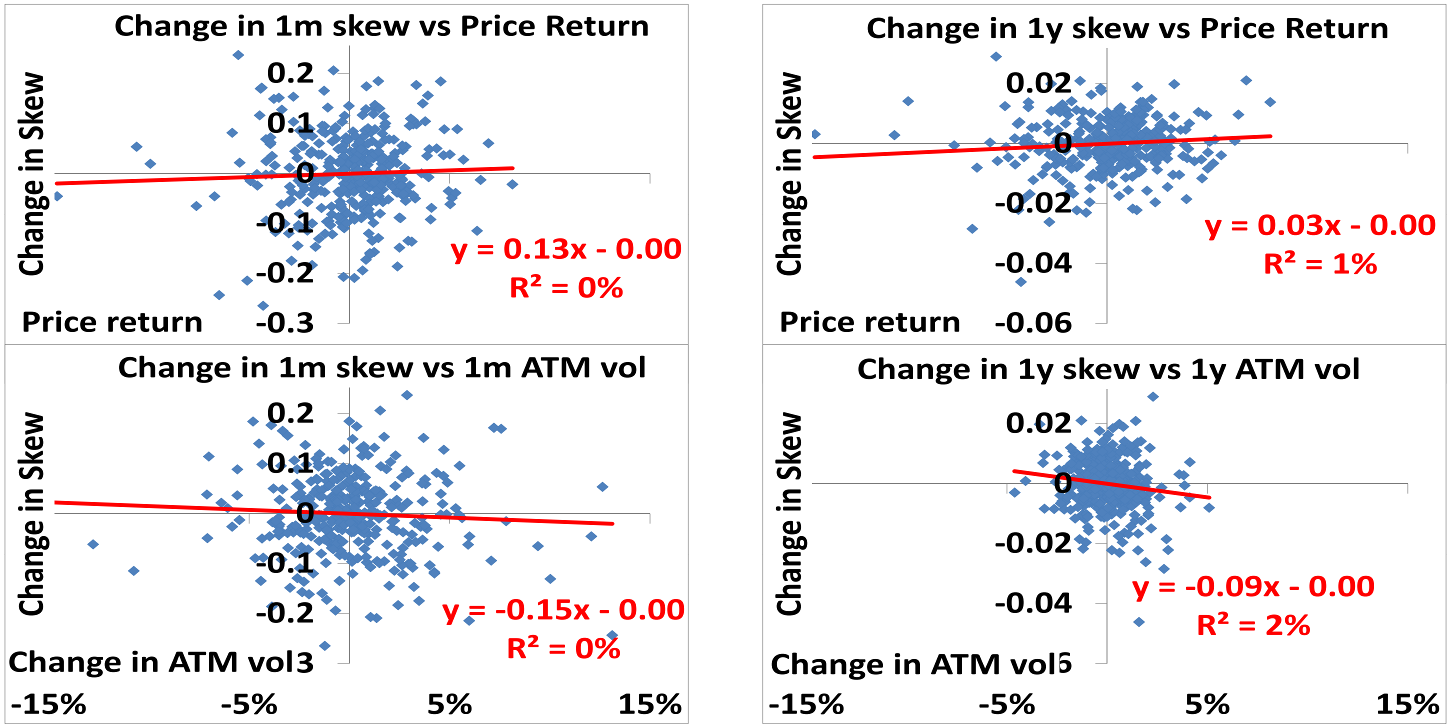


Changes in skew are not correlated to changes in price and ATM vols - important for correct predict of vol and skew P&L

Empirical observations yet again confirm log-normality dynamics - Using S&P500 data from January 2007 to December 2013

Top Figures: weekly changes in $100\% - 95\%$ **skew vs price returns** for maturity of one month (left) and one year (right): Regression slope = 0.13 (1m) & 0.03(1y); $R^2 = 0\%$ (1m) & 1% (1y)

Bottom Figures: weekly changes in $100\% - 95\%$ **skew vs changes in ATM vols** for maturity of one month (left) and one year (right): Regression slope = -0.15 (1m) & -0.06 (1y); $R^2 = 0\%$ (1m) & 2% (1y)



7.2 Minimum Variance Delta

P&L of vanilla options marked with BSM pricer for option price $U\left(S_0, \sigma_{BSM}\left(K; S_0\right)\right)$ with implied vol $\sigma_{BSM}\left(K; S_0\right)$

Delta-hedging portfolio:

$$\Pi(S, \sigma_{BSM}\left(K; S\right)) = U(S, \sigma_{BSM}\left(K; S\right)) - \Delta \times S$$

P&L from change in spot $S \to S + dS$

$$\begin{aligned} \text{P\&L} &= \Pi\left(S + dS, \sigma_{BSM}\left(K; S + dS\right)\right) - \Pi\left(S_0, \sigma_{BSM}\left(K; S_0\right)\right) \\ &\approx \frac{\partial}{\partial S} U\left(S, \sigma_{BSM}\left(K; S\right)\right) dS + \frac{\partial}{\partial \sigma} U\left(S, \sigma_{BSM}\left(K; S\right)\right) d\sigma_{BSM}(K; S + dS) - \Delta \times dS \end{aligned}$$

where $d\sigma_{BSM}(K; S + dS)$ is vol P&L:

$$d\sigma_{BSM}(K; S + dS) = \sigma_{BSM}(K; S + dS) - \sigma_{BSM}(K; S)$$

Minimum-variance delta $\Delta_{\text{min-var}}$ is applied to hedge both delta and delta-induced vega risk:

$$\begin{aligned} \Delta_{\text{min-var}} &= \frac{1}{dS}\left[U\left(S + dS, \sigma_{BSM}\left(K; S + dS\right)\right) - U\left(S_0, \sigma_{BSM}\left(K; S\right)\right)\right] \\ &= \frac{\partial}{\partial S} U\left(S, \sigma_{BSM}\left(K; S\right)\right) + \frac{\partial}{\partial \sigma} U\left(S, \sigma_{BSM}\left(K; S\right)\right) \frac{\sigma_{BSM}(K; S + dS) - \sigma_{BSM}(K; S)}{dS} \end{aligned}$$

BSM vol P&L - as I showed for modified SABR, **any generic local stoch vol can be approximated by quadratic vol for** $k = \ln(S/K)$:

$$\sigma_{BSM}(K; S) = \sigma_{ATM}(S) + \text{SKEW} \times k + \frac{1}{2}\text{CONV} \times k^2$$

Using it, vol P&L is given by:

$$\begin{aligned}\sigma_{BSM}(K;S+dS)-\sigma_{BSM}(K;S) &= d\sigma_{ATM}(S+dS)+\text{SKEW}\times dk+O(k\times dk)+O(dk^2)\\ &\approx d\sigma_{ATM}(S+dS)+\text{SKEW}\times dk\end{aligned}$$

where dk is the change in the money-ness:

$$dk=\log\left(\frac{K}{S+dS}\right)-\log\left(\frac{K}{S}\right)=-\frac{dS}{S}+O\left(\frac{dS}{S}\right)^2\approx-\frac{dS}{S}$$

and $d\sigma_{ATM}(S+dS)$ is the change in the ATM vol:

$$d\sigma_{ATM}(S+dS)=\sigma_{ATM}(S+dS)-\sigma_{ATM}(S)$$

So that, given bump in spot dS, the change in BSM vols is given by:

$$\frac{\sigma_{BSM}(K;S+dS)-\sigma_{BSM}(K;S)}{dS}\approx\frac{1}{dS}\left[\sigma_{ATM}(S+dS)-\sigma_{ATM}(S)\right]-\text{SKEW}\times\frac{1}{S}$$

Minimum-variance delta for a generic model becomes:

$$\begin{aligned}\Delta_{\text{min-var}} = &\frac{\partial}{\partial S}U\left(S,\sigma_{BSM}\left(K;S\right)\right)+\frac{\partial}{\partial\sigma}U\left(S,\sigma_{BSM}\left(K;S\right)\right)\\ &\times\left[\frac{1}{dS}\left\{\sigma_{ATM}(S+dS)-\sigma_{ATM}(S)\right\}-\text{SKEW}\times\frac{1}{S}\right]\end{aligned}$$

Everything here depends only on the implied skew apart from change in the ATM volatility, which is a model implied quantity

So all model calibrated to the same skew produce delta only if their implied changes in the ATM volatility are different

Applying the minimum-var delta, the projected change in vol level V_0 is:

$$V_0(S+dS) - V_0(S) = \beta_{SV} \times \frac{dS}{S}$$

As a result, **for modified SABR the projected change in ATM vol is**:

$$\begin{aligned}\sigma_{ATM}(S+dS) - \sigma_{ATM}(S) &= \beta_{LV} \times \left[\frac{S^{\beta_{LV}-1}}{\overline{S}^{\beta_{LV}}}\right] V_0 \times dS + \beta_{SV}\left[\frac{S}{\overline{S}}\right]^{\beta_{LV}} \times \frac{dS}{S} \\ &= \beta_{LV} \times \sigma_{ATM}(S) \times \frac{dS}{S} + \beta_{SV} \times \frac{dS}{S}\end{aligned}$$

Minimum-variance delta for modified SABR becomes:

$$\begin{aligned}\Delta_{min-var} &= \frac{\partial}{\partial S} U\left(S, \sigma_{BSM}\left(K;S\right)\right) \\ &+ \frac{\partial}{\partial \sigma} U\left(S, \sigma_{BSM}\left(K;S\right)\right) \times \left[\beta_{LV} \times \sigma_{ATM}(S) + \beta_{SV} - \text{SKEW}\right] \frac{1}{S}\end{aligned}$$

The key point of this derivation is that I showed that β_{LV} and β_{SV} are linked to market implied skew by:

$$\sigma_{ATM}(S)\beta_{LV} + \beta_{SV} = 2 \times \text{SKEW}$$

So different values of β_{LV} and β_{SV} always must produce the same skew when fitted to the given market skew

Delta $\Delta_{\text{min-var}}$ becomes:

$$\Delta_{\text{min-var}} = \frac{\partial}{\partial S} U\left(S, \sigma_{BSM}\left(K;S\right)\right) + \frac{\partial}{\partial \sigma} U\left(S, \sigma_{BSM}\left(K;S\right)\right) \times \text{SKEW} \times \frac{1}{S}$$

The minimum-variance delta does not depend on a particular choice of SABR parameters for skew β_{LV} and β_{SV}, as long as they are calibrated to the same skew (!)

Illustration of minimum-variance delta for modified SABR

Given market SKEW and CONV, we set the percentage of skew explained by the local volatility, w_{LV}:

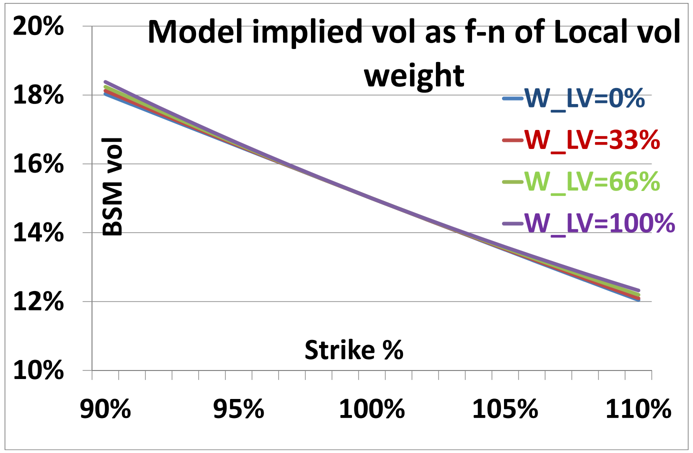
$$\sigma_{ATM}\beta_{LV} = w_{LV} \times [2 \times \text{SKEW}] \Rightarrow \beta_{LV} = w_{LV} \times \frac{2 \times \text{SKEW}}{\sigma_{ATM}}$$
$$\beta_{SV} = 2 \times \text{SKEW} \times (1 - w_{LV})$$

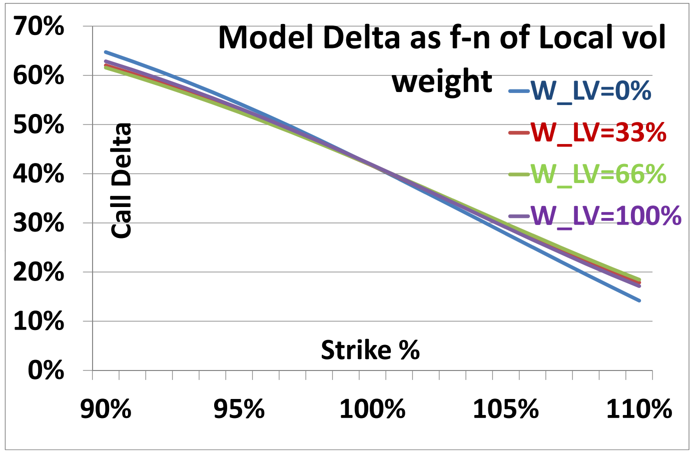
and solve for implied value of residual vol ε to match convexity CONV:

$$\varepsilon^2 = 3\sigma_{ATM} \times \text{CONV} - 2\,[2w_{LV} - 1]\,(\text{SKEW})^2$$

Figure: SABR implied vol (left) and model delta for 1y call (right) using $\sigma_{ATM} = 0.15$, SKEW $= -0.3$, CONV $= 0.2$, $w_{LV} = [0\%, 33\%, 66\%, 100\%]$

Different specifications of local and stoch vol produce the same implied vols and model deltas





Minimum variance delta is important for hedging:

We hedge against changes in risk factors (spot, ATM vol, etc)

The hedging is understood in the least-squares sense

For delta computation, choice of any particular model for stoch or local vol is irrelevant if models are calibrated to same implied vols

Risk-neutral volatility models take market prices as inputs and extrapolate the implied dynamics - they produce wrong hedges if the implied dynamics is different from the realized

In presence of the risk-aversion, implied distributions may (and do) significantly differ from realized especially for fat-tailed distributions

Risk-neutrals model are poor tools for inferences about realized dynamics

Modeling challenge is to have a model that predicts the empirically observed dynamics of the underlying price and its implied vols and yet is consistent with the implied skew observed in the market

7.3 Empirical volatility skew-beta

Empirical volatility skew-beta is important for computing correct option delta consistently with empirical dynamics

Figure 1) Apply regression model (5.1) for time series of ATM vols for maturities $T = \{1m, 3m, 6m, 12m, 24m\}$ (m=month) to estimate regression volatility beta $\beta_{REGRES}(T)$ using S&P500 returns: $\delta\sigma_{ATM}(T) = \beta_{REGRES}(T) \times \delta S$

Volatility beta for SV dynamics is instantaneous beta for very small T

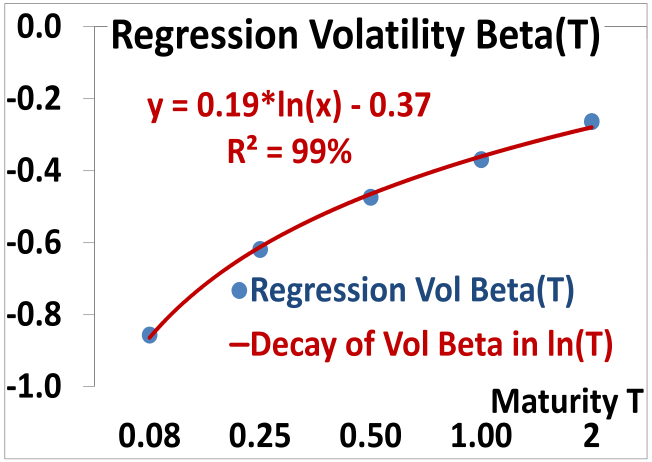
Regression vol beta decays in log-T due to mean-reversion: long-dated ATM vols are less sensitive in absolute values to price-returns

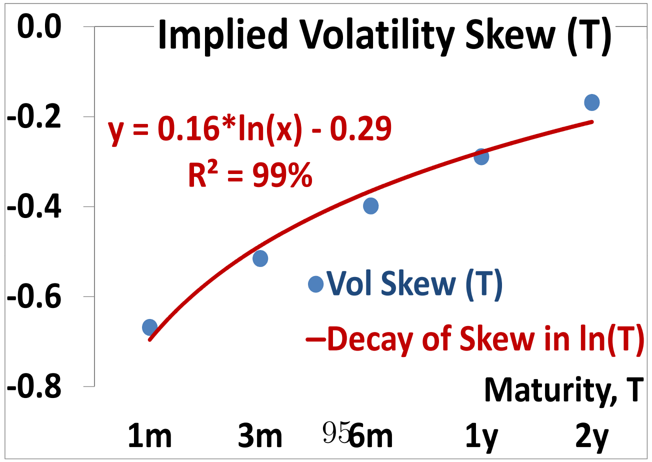
Figure 2) Implied vol skew for maturity T has similar decay in log-T

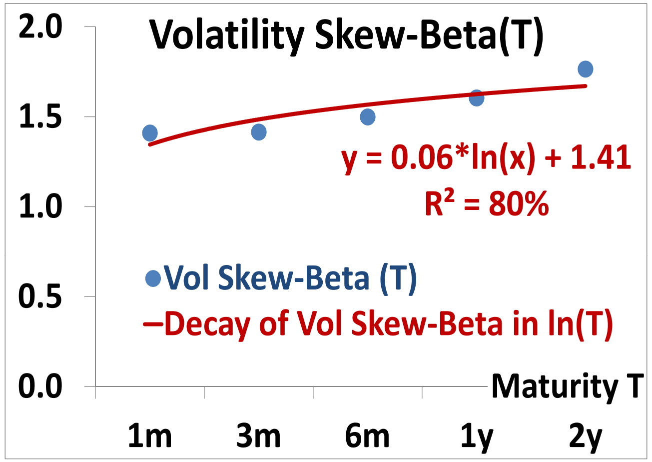
Figure 3) Volatility skew-beta is regression beta divided by skew

$$\text{Skew-Beta}(T) \propto \beta_{REGRES}(T)/\text{SKEW}(T)$$

It is nearly maturity-homogeneous







Volatility skew-beta combines the skew and volatility P&L together

Given price return δS:

$$S \to S\{1 + \delta S\}$$

Volatility P&L is computed by:

1. **For strikes re-based to new ATM level**

   Log-moneyness does not change, $\delta Z(K; S) = 0$

   P&L follows change in ATM vol predicted by regression beta and vol skew-beta:

$$\begin{aligned} \delta\sigma_{BSM}(K) \equiv \delta\sigma_{ATM}(S) &= \beta_{REGRESS} \times \delta S \\ &= \text{SKEWBETA} \times \text{SKEW} \times \delta S \end{aligned}$$

2. **For strikes fixed at old ATM level**

   Log-moneyness changes by $\delta Z(K; S) \approx -\delta S$

   P&L is change in ATM vol adjusted for skew P&L:

$$\begin{aligned} \delta\sigma_{BSM}(K) &\equiv \delta\sigma_{ATM}(S) - \text{SKEW} \times \delta S \\ &= [\text{SKEWBETA} - 1] \times \text{SKEW} \times \delta S \end{aligned}$$

   Positive change in ATM vol from negative return is reduced by skew

Volatility skew-beta under minimum-variance approach is applied to compute min-var delta Δ for hedging against changes in price and price-induced changes in implied vol:

1. We adjust option delta for change in implied vol at fixed strikes
2. The adjustment is proportional to option vega at this strike:

$$\Delta(K,T) = \Delta_{BSM}(K,T) + [\text{SKEWBETA}(T) - 1] \times \text{SKEW}(T) \times \mathcal{V}_{BSM}(K,T)/S$$

$\Delta_{BSM}(K,T)$ is BSM delta for strike K and maturity T

$\mathcal{V}_{BSM}(K,T)$ is BSM vega, both evaluated at volatility skew

I classify volatility regimes (using definitions of Derman 1999) using vol skew-beta for delta-adjustments:

$$\Delta(K,T) = \begin{cases} \Delta_{BSM}(K,T) + \text{SKEW}(T) \times \mathcal{V}_{BSM}(K,T)/S, & \text{Sticky local} \\ \Delta_{BSM}(K,T), & \text{Sticky strike} \\ \Delta_{BSM}(K,T) - \text{SKEW}(T) \times \mathcal{V}_{BSM}(K,T)/S, & \text{Sticky delta} \\ \Delta_{BSM}(K,T) + \frac{1}{2}\text{SKEW}(T) \times \mathcal{V}_{BSM}(K,T)/S, & \text{Empirical S\&P500} \end{cases}$$

"Shadow" delta is obtained using ratio O (may be different from 1/2):

$$\Delta(K,T) = \Delta_{BSM}(K,T) + O \times \text{SKEW}(T) \times \mathcal{V}_{BSM}(K,T)/S$$

which is often an ad-hoc adjustment of option delta made by traders

Instead, we can apply the regression analysis to estimate the adjustment empirically

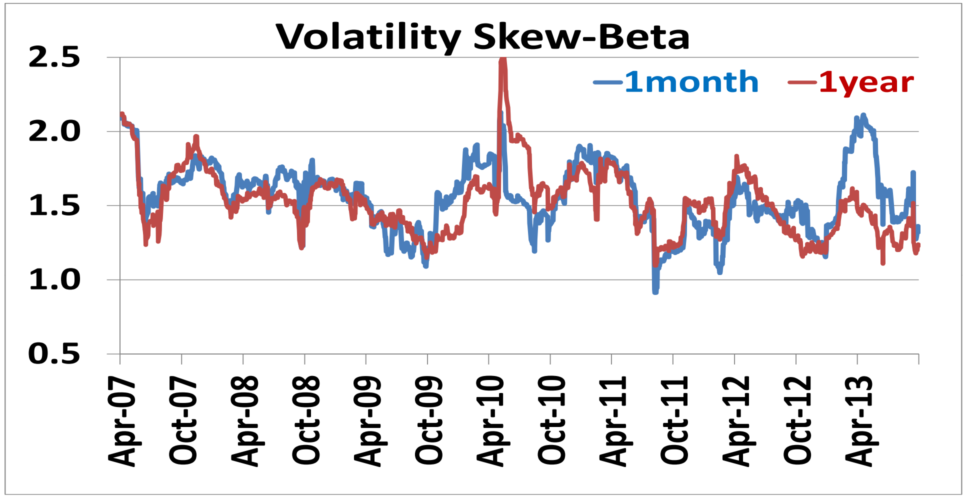
Estimation and stationarity of skew-beta: skew-beta $\beta^{(T)}$ is estimated empirically (for each vol maturity) by regression of changes in ATM vols $\sigma^{(T)}_{\text{ATM}}(t_n)$ predicted by price-return times skew $\text{Skew}^{(T)}$:

$$\sigma^{(T)}_{\text{ATM}}(t_n) - \sigma^{(T)}_{\text{ATM}}(t_{n-1}) = \beta^{(T)} \times \left[\frac{S(t_n) - S(t_{n-1})}{S(t_{n-1})} \times \text{Skew}^{(T)}(t_{n-1}) \right]$$

The skew-beta is range-bound between 1 **and** 2 **with average of about** 1.5, with weak dependence on maturity time

The regression has explain of about 70% **for S&P 500 vols**

Figure: skew-beta for 1m and 1y S&P 500 vol using rolling window of three month



7.4 Empirical Back-test For Volatility Trading Strategies

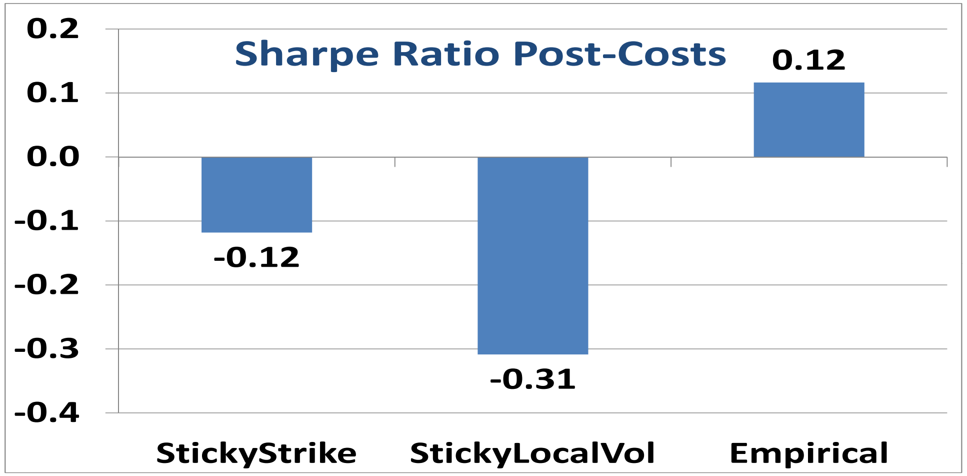
Empirical skew beta applied to vol trading strategies can significantly increase performance - P&L for delta-hedging short 6m straddle on EuroStoxx50

Back-test of monthly rolls into new straddle with maturity of 6m from 2007 to 2015: account for delta-hedge P&L and vol P&L

Delta-hedge daily using specific rules for option delta:

1. StickyStrike - hedging at sticky strike vol (BSM delta)
2. StickyLocalVol - using minimum variance hedge
3. Empirical - using skew-beta with estimation using past data

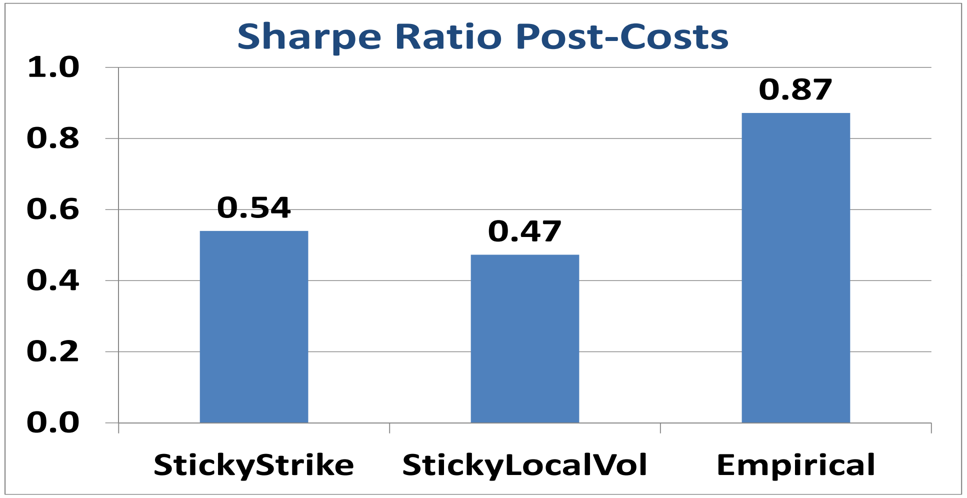
Only empirical hedge produces positive Sharpe ratio post costs



Incorporating trading signal + empirical hedging produces superior P&L for delta-hedging short 6m straddle on EuroStoxx50

Back-test of monthly rolls into new vols with maturity of 6m from 2007 to 2015 with trading signals using inference for market regimes

Delta-hedging using empirical skew-beta outperforms significantly



7.5 Conclusions

To explain reduction in volatility of delta-hedging P&L for given hedging portfolio Π for option U on S:

$$\Pi(t, S, V) = U(t, S, V) - \Delta \times S$$

Minimum-var delta Δ is computed by minimizing variance of Π using SV beta dynamics under risk-neutral pricing kernel with implied vol beta $\beta^{[I]}$:

$$\Delta = U_S + \beta^{[I]} \times U_V / S$$

where U_S and U_V model delta and vega

Given realized return δS, we apply beta SV for change in vol δV under statistical measure:

$$\delta V = \beta^{[R]} \times \delta S + \varepsilon^{[R]} \times V \sqrt{\delta t}$$

By Taylor expansion of realized P&L:

$$\delta \Pi(t, S, V) = \left[\beta^{[R]} - \beta^{[I]}\right] \times U_V \times \delta S + \varepsilon^{[R]} \times U_V \times V \sqrt{\delta t} + O(dt)$$

$\varepsilon^{(R)}$ is random non-hedgable part from residual vol-of-vol

$O(dt)$ part includes quadratic terms $(\delta S)^2$, $(\delta V)^2$, $(\delta S)(\delta V)$

Dependence on return times spread between implied vol beta $\beta^{[I]}$ and realized $\beta^{[R]}$: $\beta^{[R]} - \beta^{[I]}$ - translates into higher P&L vol

It is expensive to hedge vega

Volatility strategy with hedging model consistent with empirical dynamics can reduce costs and P&L volatility

8 Dealing with Transaction Costs

The delta-hedging in practice is applied in the discrete time setting

As a result, to optimise the delta-hedging for the practical implementation, we need to consider the discrete time framework

I will introduce my approach under the discrete time setting and the transaction costs to optimize the delta-hedging

I assume that the option position is related to a single expiry and the position is held to maturity, so that the primary risk to the terminal P&L depends only on the realized path of price returns and not on the realized path of the implied volatility

To implement a quantitative model for signal generation, we need to consider the following three aspects:

1. Given a spread between expected realized volatility and current implied volatility, say of 2%, what is the expected P&L that can be generated at the option maturity by the delta-hedging strategy under the transaction costs?
2. What is the expected volatility of the profit-and-loss (P&L) of this delta-hedging strategy?
3. How to optimize the Sharpe ratio for the delta-hedging strategy using algorithmic and deterministic rules for the re-balancing of the delta-hedge?

We need a quantitative model model for the delta-hedging P&L

8.1 Mathematical model for the Delta-hedging P&L

For an example consider a short position in an at-the-money (ATM) straddle

Given the initial option price, we compute the change in the option price by applying the Taylor expansion in the time passed from the last hedge rebalancing, TimePassed, and in the price change, PriceChange:

$$\text{Option Price change} = -Theta \times TimePassed + Delta \times PriceChange + \frac{1}{2}Gamma \times PriceChange^2$$

Here:

- The first term is the time decay measured by the option Theta
- The second term is the delta term related to the change in price measured by the option Delta
- The third term is the option Gamma term multiplied by the price change squared

For the delta-hedged position, the delta term cancels out as our stock position for the delta-hedge will generate the offsetting P&L

As a result, the P&L of the short delta-hedged position is a quadratic function of the realized return squared:

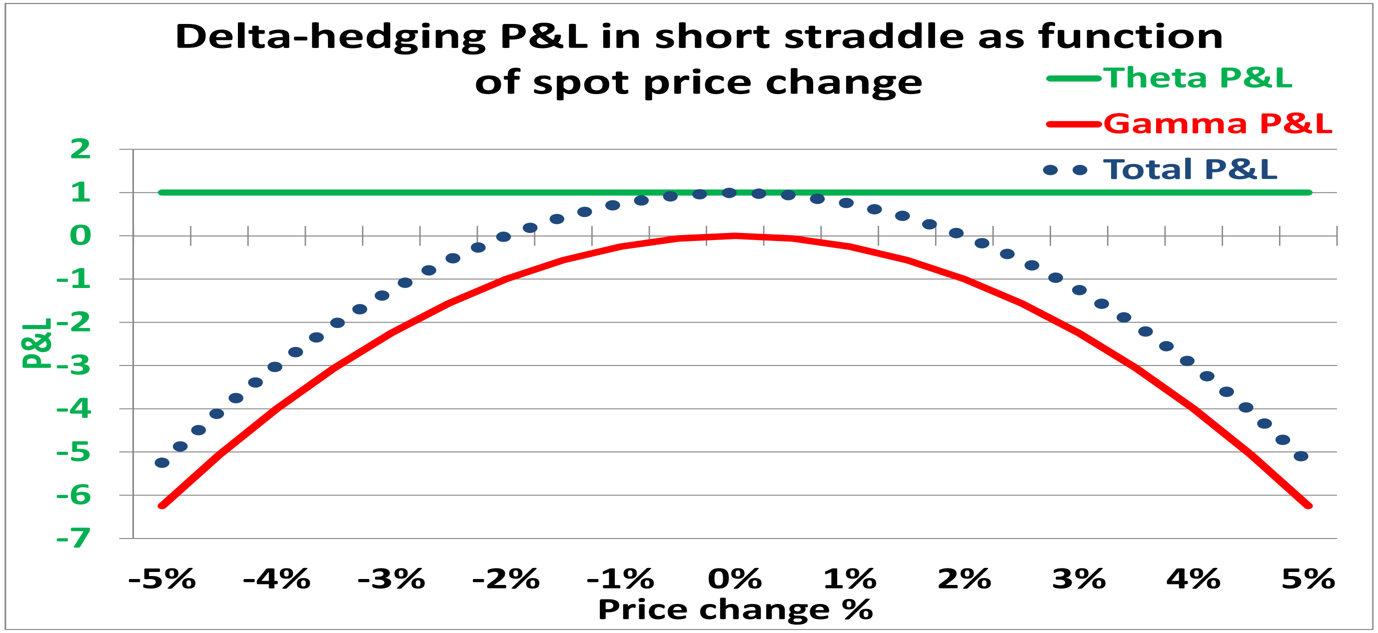
$$P\&L = Theta \times TimePassed - \frac{1}{2}Gamma \times PriceChange^2$$

Figure 1: the profile of the P&L generated by the short delta-hedged position

The green line corresponds to the straddle Theta, which is the gain we receive independently from price return

The red line represents the loss from change in the option price arising from option gamma times the return squared, which has the shape of a parabola

The total P&L is obtained as the sum of the theta and the short gamma P&Ls



The market price of the option value is computed using the Black-Scholes-Merton (BSM) formula with implied volatility denoted by ImpliedVol

Apply the BSM equation to represent the option Theta using the option Gamma as follows:

$$Theta = -0.5 \times Gamma \times SpotPrice^2 \times ImpliedVol^2$$

As a result, the P&L of the short delta-hedged position in the straddle becomes:

$$P\&L(n) = \frac{1}{2}(ImpliedVol^2 - RealizedVol^2) \times \big(Gamma \times SpotPrice^2 \times TimePassed\big)$$

RealizedVol is the realized volatility computed by:

$$RealizedVol^2 = \frac{PriceChange^2}{TimePassed \times SpotPrice^2} \tag{8.1}$$

Therefore the break-even return at which the P&L is zero is:

$$BreakEvenReturn = ImpliedVol \times \sqrt{TimePassed}$$

Expected Total P&L

The total realized P&L realized at option maturity is the sum of single period P&Ls

After some slight simplifications, the expected P&L is the spread between the squares of the implied and realized volatilities multiplied by the option gamma (at the inception of the trade) and maturity time:

$$ExpectedP\&L = \frac{1}{2}(ImpliedVol^2 - RealizedVol^2) \times (Gamma \times SpotPrice(0)^2) \times MaturityTime$$

The remarkable observation is that this result is independent from the hedging frequency

Volatility of P&L

The primary risk for the P&L volatility comes from the volatility of squared returns and the residual volatility arising from the spread between implied and realized volatilities and the stochastic volatility and jumps in price returns

I derive the following formula for the volatility of the P&L that has the hedgable term related to the returns volatility, denoted by ReturnsVolatility, and the residual term, denoted by ResidualVolatility:

$$P\&Lvolatility(HedgingFrequency) = \sqrt{\frac{ReturnsVolatility^2}{HedgingFrequency} + ResidualVolatility^2}$$

The more frequently we delta-hedge, the more P&L volatility we can eliminate

The residual volatility, which cannot be eliminated by delta-hedging, arises from the two sources:

1. The spread between implied and realized volatilities
2. Un-hedgeable delta risk due to jumps and gaps in the price

When the hedging frequency is small the P&L volatility is high

By increasing the hedging frequency the P&L volatility declines but the reduction becomes smaller and smaller as the hedging frequency increases

Transaction costs

Transaction costs occur when we buy or sell the stock to cover our delta

I assume that the strategy is rebalanced using the delta-hedge computed with quoted price S:

- buy orders are executed at price $S * (1 + 0.5 * BidAskSpread)$
- sell orders are executed at price S*(1-0.5*BidAskSpread)

where BidAskSpread measures the bid-ask spread

Delta-hedging with transaction costs is first considered by Leland (1985)

I show that the total expected transaction costs are proportional to the square root of the hedging frequency times the BidAskSpread and some constant C related to gamma:

$$ExpectedCosts(HedgingFrequency) = BidAskSpread \times \sqrt{HedgingFrequency} \times C$$

The total expected P&L is then obtained as the expected P&L minus the expected transaction costs:

$$TotalP\&L(HedgingFrequency) = ExpectedP\&L - ExpectedCosts(HedgingFrequency)$$

Given the obtained formulas for expected P&L and its volatility, I compute the Sharpe ratio of the position as follows:

$$Sharperatio(HedgingFrequency) = \frac{TotalP\&L(HedgingFrequency)}{P\&Lvolatility(HedgingFrequency)} \tag{8.2}$$

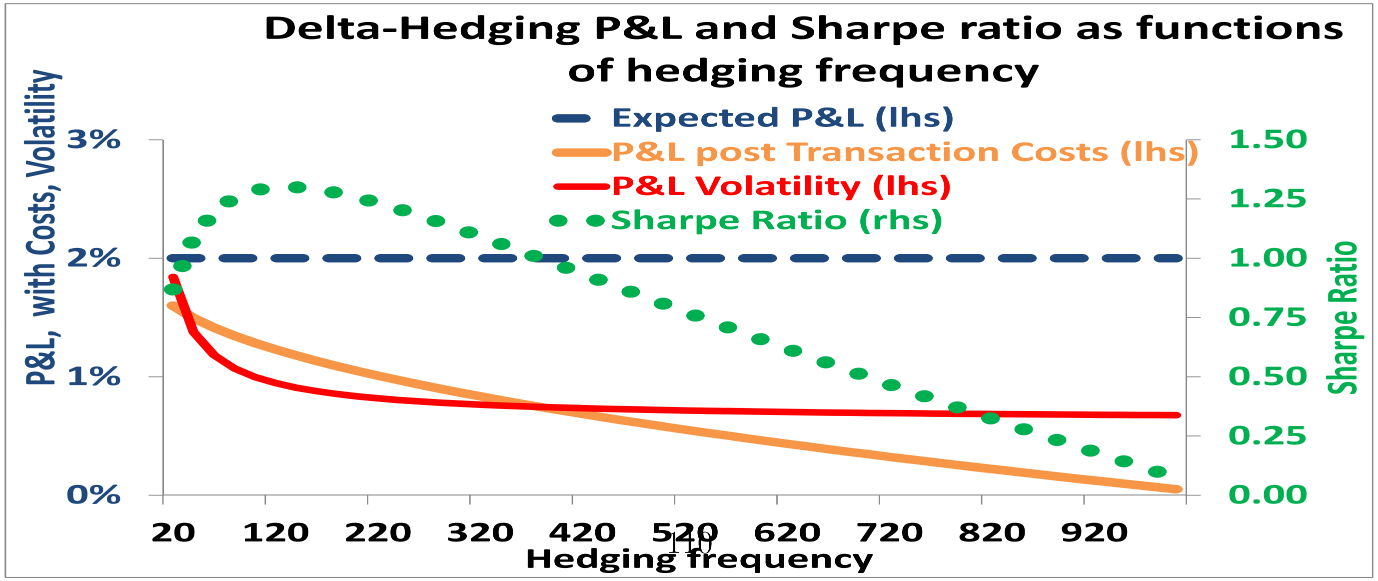
Illustration:

- The expected P&L does not depend on the hedging frequency
- The P&L volatility
- The expected P&L declines as we increase the hedging frequency
- The Sharpe ratio

As we increase the hedging frequency, the P&L volatility declines increasing the Sharpe ratio

However, as we keep increasing the hedging frequency, the transaction costs increase while the P&L volatility drops as a slower rate, so that the Sharpe ratio begins to decline

The analytic solution for the optimal hedging frequency that maximizes the expected Sharpe ratio can be found in my paper (Sepp (2013))



8.2 Re-balancing using price and delta bands

To minimize the transaction costs, I consider the two cases:

1. the re-balancing is based on bands for the price change

2. the re-balancing is triggered by changes in option delta

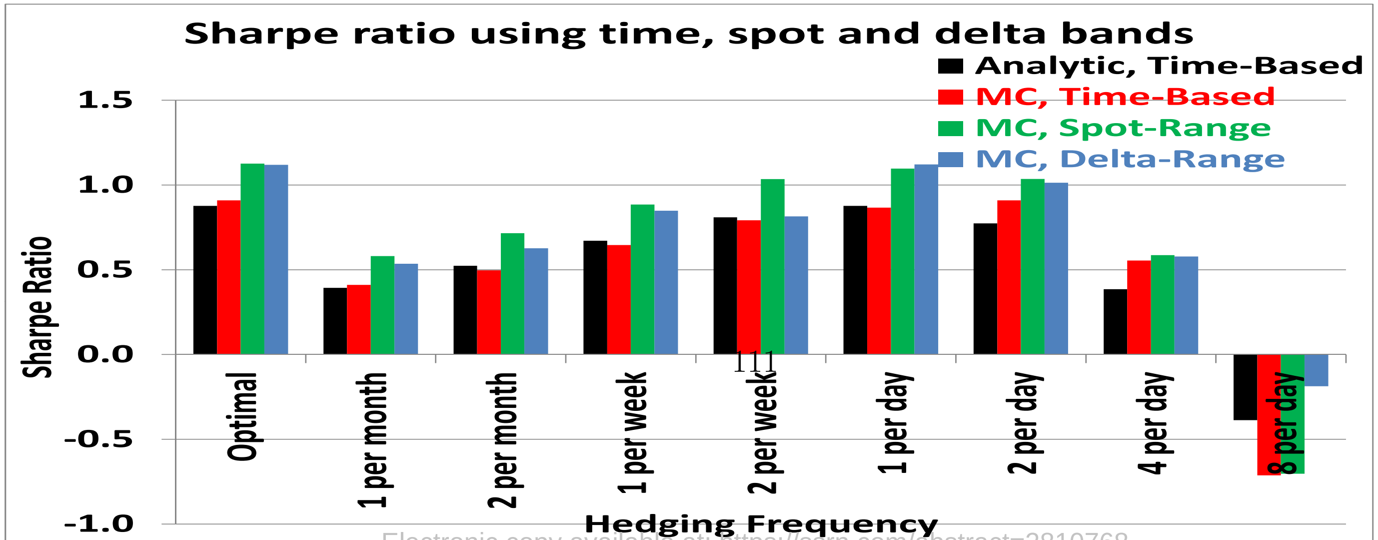
Figure illustrates the Sharpe ratio obtained by using both the analytical formula as well as Monte-Carlo simulations of the delta-hedged P&L using fixed-time and delta and price bands:

- The method provides a very good approximation to the actual Sharpe ratio obtained by Monte Carlo simulations under the time-based re-hedging

- Proposed approximations to convert the time based rule into the price and delta bands also provide a reliable estimate for the Sharpe ratio

- Using the approximations for the price and delta bands, we can first find the optimal hedging frequency for the time-based re-hedging and then apply these approximations to get the equivalent price and delta bands

- Price and delta bands help to increase the realized Sharpe ratio by minimizing realized transaction costs



9 Explaining the volatility risk-premium using econometric model with risk-aversion

It is important to understand and explain the volatility risk-premium in a quantitative way and understand the fundamental origin of P&L of these strategies

These type of questions can be analyzed using Asset Pricing theory (Cochrane (2005), Asset Pricing) which studies:

1. Empirical risk-reward dynamics of different assets and strategies
2. Theoretical models to explain empirical observations and model investment&consumption choices using valuation kernels with risk preferences

Assets include everything that can be considered as an investment:

- Stocks, stock indices, credit bonds,...
- Derivative securities
- Quantitative trading strategies and, in particular, volatility trading strategies

Asset pricing incorporates the risk-neutral valuation theory but not the other way around

The toolkit of asset pricing is developed to describe how risk factors evolve econometrically and price-formation in terms of risk factors

Risk-neutral models with implied calibration assume a given set of market prices and extrapolate the implied dynamics of risk factors, which may be not (typically it is not) consistent with their realized dynamics

Asset pricing using valuation kernels states that the price of any asset p_t equals to the expectation under stochastic discount factor (statistical measure):

$$p_t = \mathbb{E}_t\left[m_{t+1}x_{t+1}\right]$$

x_{t+1} is risky pay-off at time $t+1$ (one-period model)

m_{t+1} is the stochastic discount factor to translate future payoffs in present value terms

$\mathbb{E}$ is expectation under the statistical measure

Introduce gross return $R_{t+1} = (x_{t+1} - p_t)/p_t$ and after re-arrangement:

$$\mathbb{E}_t\left[R_{t+1}\right] - \left(\frac{1}{\mathbb{E}_t\left[m_{t+1}\right]} - 1\right) = -\frac{1}{\mathbb{E}_t\left[m_{t+1}\right]} \times \mathbb{C}ov_t\left[R_{t+1}, m_{t+1}\right]$$

where $\mathbb{C}ov_t$ is the covariance between the return and discount factor

The return R^f on risk-free bond is uncorrelated with the discount factor:

$$R^f_{t+1} = \frac{1}{\mathbb{E}_t\left[m_{t+1}\right]} - 1 \;\Rightarrow\; \mathbb{E}_t\left[m_{t+1}\right] = \frac{1}{1 + R^f_{t+1}}$$

Finally, the expected return is risk-free rate plus risk-adjustment:

$$\mathbb{E}_t\left[R_{t+1}\right] = R^f_{t+1} - (1 + R^f_{t+1}) \times \mathbb{C}ov_t\left[R_{t+1}, m_{t+1}\right]$$

Asset pricing under stochastic discount rate

Assume that the valuation kernel can be presented as:

$$m_{t+1} = e^{-D_{t+1}}$$

where D_{t+1} is market discount rate (think of rate of return on the S&P 500 index or BBB-bond index)

Then covariance between risky return and market rate becomes:

$$\mathbb{C}ov_t\left[R_{t+1}, m_{t+1}\right] \approx -\mathbb{C}ov_t\left[R_{t+1}, D_{t+1}\right]$$

Define the expected return beta to market discount rate:

$$\mathbb{B}eta_t\left[R_{t+1}, D_{t+1}\right] = \frac{\mathbb{C}ov_t\left[R_{t+1}, D_{t+1}\right]}{\mathbb{V}ar_t\left[D_{t+1}\right]}$$

The expected return on risky asset can be decomposed by:

$$\mathbb{E}_t\left[R_{t+1}\right] - R^f_{t+1} = \mathbb{B}eta_t\left[R_{t+1}, D_{t+1}\right] \times \lambda$$

where λ is **the unit price of risk** (generic specification)

$$\lambda = \mathbb{V}ar_t\left[D_{t+1}\right](1 + R^f_{t+1})$$

In the CAPM model, the unit price of risk is specified by spread between expected market return and risk-free rate:

$$\lambda = \mathbb{E}\left[D_{t+1}\right] - R^f_{t+1}$$

Asset pricing with risk-aversion is applied for valuation of securities by introducing risk-aversion and utility functions for discounting risky pay-offs

"Individuals are risk averse if they always prefer to receive a fixed payment to a random payment of equal expected value" (Dumas-Allaz (1996))

For exponential utility with risk-aversion parameter $\gamma > 0$, apply exponential relaxation:

$$q[R, D] \propto e^{-\gamma(R-\overline{R})} \times p[R, D]$$

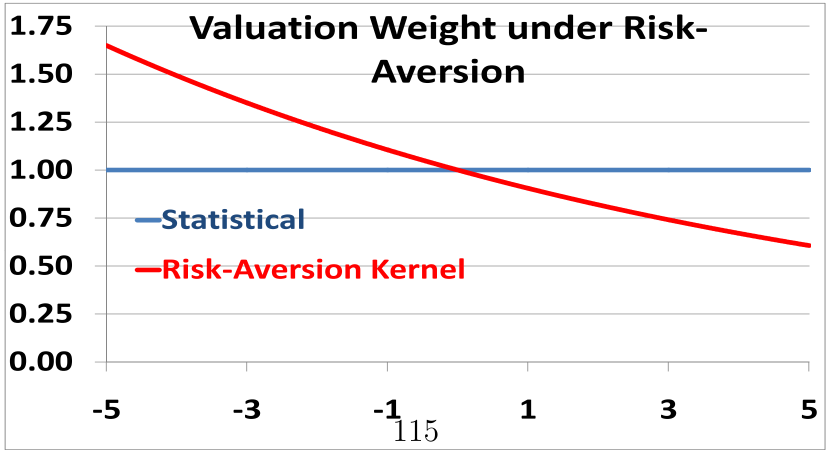
$p[R, D]$ is the statistical kernel (to model econometric dynamics)

$q[R, D]$ is the valuation kernel (to compute expected returns)

$\overline{R}$ is the mean under the statistical kernel p

Figure: Valuation kernel assigns more weight for negative returns

The risk-aversion may change the moments significantly if statistical kernel features fat tails

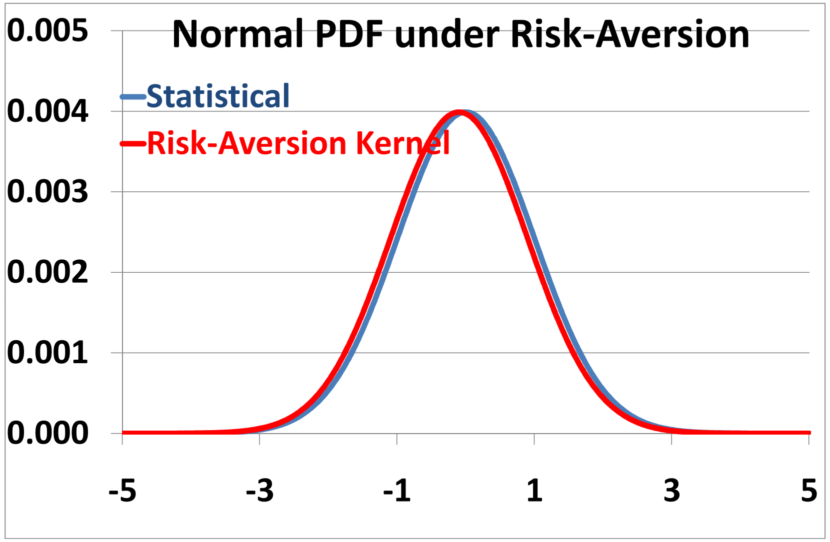


Table&Figure: Risk-Aversion for Gaussian statistical kernel (normalized to volatility of 1.00) with relaxation parameter $\gamma = -0.10$ shifts the valuation kernel to the left

Only the mean under the valuation kernel is changed by $\gamma = -0.10$

Higher moments of valuation kernel are unchanged

	statistical	Risk-Aversion Kernel
Mean	**0.00**	**-0.10**
Volatility	1.00	1.00
Skew	0.00	0.00
Kurtosis	0.00	0.00



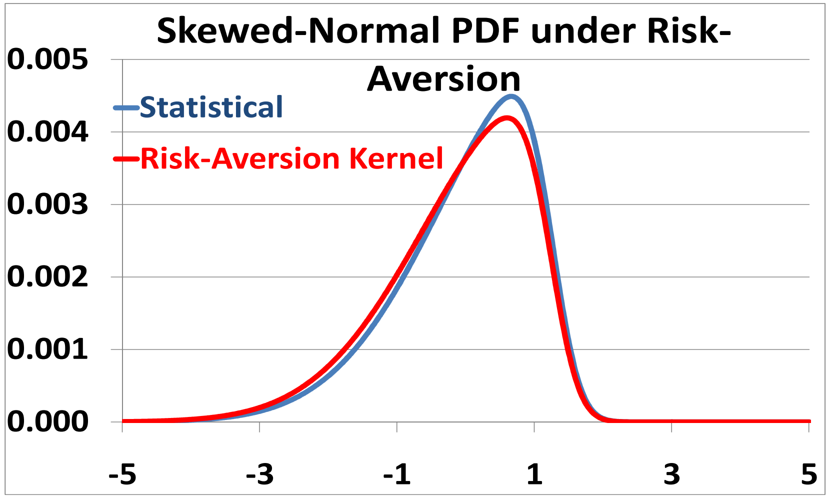
Table&Figure: Risk-Aversion under Skew-Gaussian PDF (normalized to volatility of 1.00, large negative skew of -0.85 and small kurtosis) with relaxation parameter $\gamma = -0.10$

The mean under the valuation kernel is changed by $\gamma = -0.10$

Also, **the implied volatility under the valuation kernel is changed** by 4.5% - the impact from the statistical skew

Higher moments of valuation kernel are slightly changed

	Statistical	Risk-Aversion Kernel
Mean	0.00	**-0.10**
Volatility	**1.00**	**1.04**
Skew	**-0.85**	**-0.81**
Kurtosis	0.71	0.58

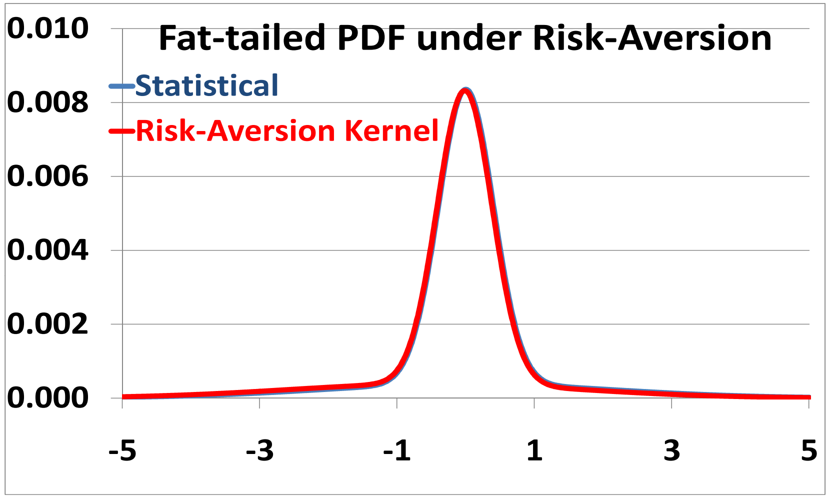


Table&Figure: Risk-Aversion under Fat-Tailed PDF using mixture of normals (normalized to volatility of 1.00, zero skew and large excess kurtosis of 8.50) with relaxation parameter $\gamma = -0.10$

The implied volatility under the valuation kernel is changed by 3.5% - the impact from the statistical kurtosis

The implied skew under the valuation kernel becomes large and negative: the impact from the fat tails of the statistical distribution (statistical kurtosis) is significant

	Statistical	Risk-Aversion Kernel
Mean	0.00	-0.10
Volatility	1.00	1.03
Skew	**0.00**	**-0.82**
Kurtosis	**8.50**	8.30



9.1 Explanation of Volatility Risk-premium using risk-aversion

If asset returns under statistical kernel exhibit skews and fat tails and investors are risk-averse to negative returns:

1. The implied distribution under the valuation kernel has higher volatility and, especially, the skewness
2. The difference between statistical and implied moments (like the volatility and skew) is termed the risk-premium
3. The valuation of derivatives securities should incorporate the risk-premium under risk-aversion

9.2 Empirical Evidence

I show empirical statistics of returns on major risk assets:

1. **S&P500 total return**: weekly returns from 1-Jul-86
2. **Call-S&P500**: CBOE BuyWrite index - long in the S&P 500 index, sell one-month OTM calls
3. **Put-S&P500**: CBOE PutWrite index - sell one-month ATM puts

S&P500 returns have moderate skew and kurtosis: large positive and negative returns are equally likely (-24.70% was realized in crash of 1987)

Selling either call or put options increases considerably skew and kurtosis

Fitting of tail parameter ν of student t-distribution: four moments exists for S&P500 returns; PutWrite index returns may not have finite variance

Realized return on PutWrite is 2% higher and Sharpe ratio of 0.92 vs 0.51 for S&P 500

	S&P500	Call-S&P500	Put-S&P500
Return (annual)	8.76%	9.52%	10.61%
Stdev (annual)	17.25%	12.08%	11.50%
Sharpe	0.51	0.79	0.92
Skew	**-1.08**	**-2.62**	**-4.75**
Excess kurtosis	**10.57**	**32.68**	**76.44**
Max return	13.17%	9.07%	8.28%
Min return	-24.70%	-23.42%	-28.29%
Nu, t-distribution	4.45	2.14	1.61
Finite moment	**4**	**2**	**1**

BofA ML US Corporate Total Return Credit Indices: weekly returns from 3-Jan-94 to 20-Oct-2014 1) AAA: highest quality investment grade credit bonds
2) BBB: credit bonds
3) CCC: high yield and speculative bonds

Returns on all AAA and BBB indices have moderate skew and with large kurtosis: large positive and negative returns are equally likely

Returns on CCC high yield are not skewed but rather large kurtosis: large negative returns are more likely

Fitting of tail parameter ν of student t-distribution: five moments exists for AAA and BBB returns; CCC index returns have finite variance

Large Sharpe ratios for AAA and BBB of (1.07 and 1.30) compared to that for S&P 500 index (0.50)

	AAA	BBB	CCC
Return (annual)	5.82%	6.92%	7.85%
Stdev (annual)	5.44%	5.32%	11.04%
Sharpe	1.07	1.30	0.71
Skew	**-0.37**	**-0.69**	**-0.35**
Excess kurtosis	**5.18**	**5.12**	**9.79**
Max return	5.00%	2.64%	10.84%
Min return	-5.15%	-5.94%	-8.85%
Nu, t-distribution	5.10	5.51	2.07
Finite moment	**5**	**5**	**2**

iShares US Treasury bonds ETFs: weekly returns from 30-Jul-02 to 21-Oct-14

1. 1-3y US Treasury bonds
2. 7-10y US Treasury bonds
3. 20+y US Treasury bonds

Long-term treasuries have very small skew and kurtosis - benefit from safe-haven status

Unlike anything else, returns on 3y T-s have positive skew and with rather large kurtosis with only second moment being finite

	3y	10y	20y
Return (annual)	2.41%	5.94%	8.07%
Stdev (annual)	1.38%	6.64%	12.85%
Sharpe	1.74	0.89	0.63
Skew	**0.53**	**-0.18**	**0.00**
Excess kurtosis	**3.19**	**1.45**	**1.40**
Max return	1.03%	2.94%	7.71%
Min return	-0.78%	-4.52%	-7.95%
Nu, t-distribution	2.92	7.10	7.37
Finite moment	**2**	**7**	**7**

9.3 Portfolio of delta-hedged derivatives under the risk-aversion

This is based on my original framework

Using exponential relaxation γ for valuation of asset with return R_{t+1}, we compute beta of return to the market discount rate, $\mathbb{B}^{\gamma}eta_t\left[R_{t+1}, D_{t+1}\right]$, under the valuation kernel:

$$\mathbb{B}^{(\gamma)}eta_t\left[R_{t+1}, D_{t+1}\right] = \mathbb{B}eta_t\left[R_{t+1}, D_{t+1}\right] - \gamma\mathbb{B}eta_t\left[(R_{t+1} - \overline{R}_{t+1})^2, D_{t+1}\right] + o(\gamma^2)$$

- $\mathbb{B}eta_t\left[R_{t+1}, D_{t+1}\right]$ is **the return beta** to the market discount rate under the statistical kernel
- $\mathbb{B}eta_t\left[(R_{t+1} - \overline{R}_{t+1})^2, D_{t+1}\right]$ is **the variance beta** to the market discount rate under the statistical kernel

Valuation of the expected return under the kernel with risk-aversion:

$$\begin{aligned}\mathbb{E}_t\left[R_{t+1}\right] - R^f_{t+1} &= \mathbb{B}eta_t\left[R_{t+1}, D_{t+1}\right] \times \lambda^{(1)} \\ &\quad - \gamma \times \mathbb{B}eta_t\left[(R_{t+1} - \overline{R}_{t+1})^2, D_{t+1}\right] \times \lambda^{(2)}\end{aligned}$$

$\lambda^{(1)}$ and $\lambda^{(2)}$ are units of risk for return and variance betas

If variance of asset or strategy returns is negatively correlated to discount rate, expected return on this strategy should be higher

For example when selling put options - if market goes down, the variance of short puts increases, so by selling puts we need to charge extra premium for negative co-variance between discount rate and variance of puts

Delta-hedging P&L from trading in derivative security, can also be considered as an asset with return O_{t+1}

Under perfect replication, the beta of return to market discount is zero:

$$\mathbb{B}eta_t\left[O_{t+1}, D_{t+1}\right] = 0$$

Therefore with perfect replication and without the risk-aversion:

$$\mathbb{E}_t\left[O_{t+1}\right] = R^f_{t+1}$$

which is in accordance with the risk-neutral pricing
If the perfect replication is impossible (like hedging is discrete or there are residual risks such as stochastic volatility of returns and implied vols), **the variance of O_{t+1} is not zero**
As a result, even if the first-order risk can be removed:

$$\mathbb{B}eta_t\left[O_{t+1}, D_{t+1}\right] = 0\ ,\ \mathbb{B}eta_t\left[(O_{t+1} - \overline{O}_{t+1})^2, D_{t+1}\right] \neq 0$$

We obtain for expected return:

$$\mathbb{E}_t\left[O_{t+1}\right] = R^f_{t+1} - \gamma \times \mathbb{B}eta_t\left[(O_{t+1} - \overline{O}_{t+1})^2, D_{t+1}\right] \times \lambda^{(2)}$$

As a result, a derivative security with non-zero variance of delta-hedged returns and negative covariance with the market discount rate will demand extra risk-premium
For example, a short position in delta-hedged index puts

Empirical evidence - estimation of variance beta $\beta_{R^2,D}$ by applying regression for monthly returns for assets considered before:

$$Y_{t+1} = \alpha + \beta_{R^2,D} D_{t+1}$$

D_{t+1} is monthly return on the S&P500 index

Top table: $Y_{t+1} = (P_{t+1} - \overline{P}_{t+1})^2$ is realized second moment of monthly asset return

Bottom: $Y_{t+1} = \mathbb{V}ol[P_{t+1}]$ is monthly realized volatility of daily returns

All estimated regression parameters are significant

Variance beta is negative for all assets so that, given a negative return on the S&P500 index, the realized second moment and vols increase

Under the valuation kernel with risk-aversion, expected return on a risky asset is higher&proportional to risk-aversion γ **for this asset**

Factor	S&P500	Call	Put	AAA	BBB	CCC	3y	10y	20y
$\hat{\alpha}$	2.38%	1.32%	1.21%	0.28%	0.37%	2.32%	0.03%	0.49%	1.87%
$\hat{\beta}_{R^2,D}$	-0.24	-0.21	-0.22	-0.03	-0.04	-0.22	0.00	-0.05	-0.26
R^2 adj	6%	10%	10%	4%	3%	2%	10%	7%	8%

Factor	S&P500	Call	Put	AAA	BBB	CCC	3y	10y	20y
$\hat{\alpha}$	16.62%	11.07%	10.00%	5.01%	4.98%	5.43%	1.40%	6.79%	13.06%
$\hat{\beta}_{R^2,D}$	-0.92	-1.04	-1.14	-0.02	-0.09	-0.29	-0.05	-0.11	-0.22
R^2 adj	17%	28%	29%	0%	5%	9%	8%	3%	3%

Empirical evidence - explaining realized returns based on my derived equation for theoretical expected return:

$$\mathbb{E}_t\left[R_{t+1}\right] - R^f_{t+1} = \mathbb{B}_t\left[R_{t+1}, D_{t+1}\right] \times \lambda^{(1)} - \gamma \times \mathbb{B}_t\left[(R_{t+1} - \overline{R}_{t+1})^2, D_{t+1}\right] \times \lambda^{(2)}$$

For empirical estimation, unit prices of risks $\lambda^{(1)}$ and $\lambda^{(2)}$ are specified:

- $\lambda^{(1)}$ - return on the S&P 500 index (expected carry on long risk)
- $\lambda^{(2)}$ - spread between average VIX in given month and realized monthly volatility of S&P500 daily returns (proxy for expected carry on short vol)

Estimation using monthly returns:

$$R_{t+1} = \alpha + \beta_{R,D^2}\lambda^{(1)}_{t+1} + \beta_{R,D^2}\lambda^{(2)}_{t+1}$$

For assets Call and Puts shorting S&P500 volatility and for credit, realized variance beta β_{R,D^2} is positive, in line with the theoretical argument

Treasuries have negative variance betas - long-term realized and implied returns on safety assets are lower

Factor		Call	Put	AAA	BBB	CCC	3y	10y	20y
alpha	$\hat{\alpha}$	-0.02%	0.05%	0.33%	0.07%	-0.52%	0.28%	0.60%	0.93%
Ret	$\hat{\beta}_{R,D}$	0.60	0.52	0.00	0.08	0.50	-0.02	-0.14	-0.28
VIX	$\hat{\beta}_{R,D^2}$	0.08	0.10	0.04	0.11	0.25	-0.02	-0.01	-0.04
Expl	R^2 adj	81%	72%	0%	17%	43%	12%	8%	9%

9.4 Extension to portfolio of volatility strategies

Consider portfolio of short positions in delta-hedged options, O_n, each with weight w_n:

$$\Pi = \sum_n w_n O_n$$

Realized returns $\widetilde{O}_{n,t+1}$ on O_n can be presented using a factor model:

$$\widetilde{O}_{n,t+1} - R^f_{t+1} = \widetilde{\beta}_n \widetilde{D}_{t+1} + \epsilon_{n,t+1}$$

- $\widetilde{D}_{t+1}$ is realized change in the market discount rate
- $\widetilde{\beta}_n$ is realized beta to the market discount rate
- $\epsilon_{n,t+1}$ is realization of idiosyncratic risk factor that balances the left side

Realized return on portfolio $\widetilde{\Pi}_{n,t+1}$ is:

$$\widetilde{\Pi}_{n,t+1} - R^f_{t+1} = \widetilde{\beta}_\Pi \widetilde{D}_{t+1} + \epsilon_{\Pi,t+1}$$

where $\widetilde{\beta}_\Pi$ is **realized portfolio beta** and $\epsilon_{\Pi,t+1}$ is **excess return**

$$\widetilde{\beta}_{\Pi,t+1} = \sum_n w_n \widetilde{\beta}_{n,t+1} \ , \ \epsilon_{\Pi,t+1} = \sum_n w_n \epsilon_{n,t+1}$$

Incorporating the spread between realized and implied returns

Implied excepted return under the valuation kernel with risk-aversion:

$$\mathbb{E}_t^{(\gamma)}[O_{t+1}] - R_{t+1}^f = \beta_{\Pi,D} \times \mathbb{E}[D_{t+1}] - \gamma \times \beta_{\Pi^2,D} \times \lambda$$

$\widetilde{\beta}_{\Pi,D}$ and $\widetilde{\beta}_{\Pi^2,D}$ are portfolio implied betas and $\epsilon_{\Pi,t+1}$ is excess return:

$$\beta_{\Pi,D} = \sum_n w_n \mathbb{B}eta_{n,t}\left[R_{t+1}, D_{t+1}\right]$$
$$\beta_{\Pi^2,D} = \sum_n w_n \mathbb{B}eta_{n,t}\left[(O_{n,t+1} - \overline{O}_{n,t+1})^2, D_{t+1}\right]$$

The difference between implied and realized returns:

$$\begin{aligned} H &= \mathbb{E}_t^{(\gamma)}[O_{t+1}] - \widetilde{\Pi}_{n,t+1} \\ &= \left(\beta_\Pi \times \mathbb{E}^{(\gamma)}[D_{t+1}] - \widetilde{\beta}_\Pi \widetilde{D}_{t+1}\right) - \epsilon_{\Pi,t+1} - \gamma \times \beta_{\Pi^2,D} \times \lambda \end{aligned}$$

We would expect that for delta-hedged portfolio both implied and realized betas to the market discount rate are negligible, so that:

$$H = -\epsilon_{\Pi,t+1} - \gamma \times \beta_{\Pi^2,D} \times \lambda$$

As a result, the portfolio of delta-hedged short-options makes money if the risk-aversion is present:

- The implied beta of portfolio return variance (short options) is positively correlated to market discount rate so $\beta_{\Pi^2,D} > 0$

- For a diverse portfolio, residual risks $\epsilon_{\Pi,t+1}$ should be negligible (but see next slide!)

9.5 Portfolio P&L and systemic gap risks: LTCM and MFG

Idiosyncratic risk $\epsilon_{\Pi,t+1}$ arises from being short risk factors such as implied vols or credit bonds

For a diverse portfolio, the sum of idiosyncratic risks $\epsilon_{\Pi,t+1}$ should be negligible

However, this is not true if shocks are systemic and portfolio idiosyncratic losses do not cancel each other

For example, if all position produce the same gap losses over short-period, idiosyncratic risk become systemic:

$$\epsilon_{n,t+1} = J \;\Rightarrow\; \epsilon_{\Pi,t+1} = -J$$

and the portfolio P&L over short-period becomes:

$$\text{P\&L} = -J - \gamma \times \beta_{\Pi^2,D} \times \lambda$$

As a result, for a portfolio shorting market implied risk factors (vols and credit), we expect **the cyclical pattern of P&L with long periods of positive returns followed by sudden and large systemic losses**

This is very representative of derivatives business at large: I give one example using Long Term Capital Management (LTCM)

Position mis-sizing: blow-up of Long-Term Capital Management

Was successful in the first tree years from founding in 1994

Subsequent compression in risk-premiums lead to massive (about $\times 25$) increase in leverage in many quantitative strategies shorting the risk-premium across different markets (in implied vols, credit, term premium in sovereign bonds, etc) - strategies were believed to be uncorrelated

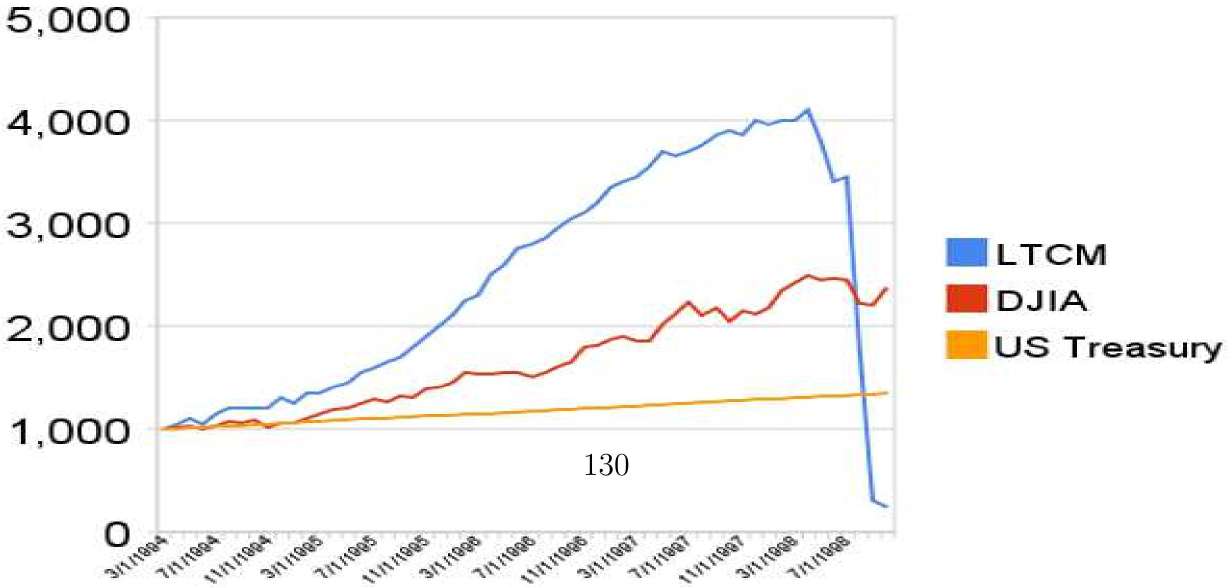
In line with the equation I derive for portfolio in short delta-hedged options with risk-premiums:

$$\text{P\&L} = -\gamma \times \beta_{\Pi^2,D} \times \lambda - J \times \mathbf{1}_{\{\text{Systemic Loss}\}}$$

LTCM made money in good times by shorting risk premium betting on $\beta_{\Pi^2,D} < 0$

However in 1998 LTCM suffered the systemic loss J across different strategies and the firm was liquidated

Figure: Value of $1,000$\$ invested in LTCM, DJIA, T-s (source: Lowenstein (2000))



Another example of misfortune in sovereign risk-premium: position mis-timing relative to market cycle by MF Global

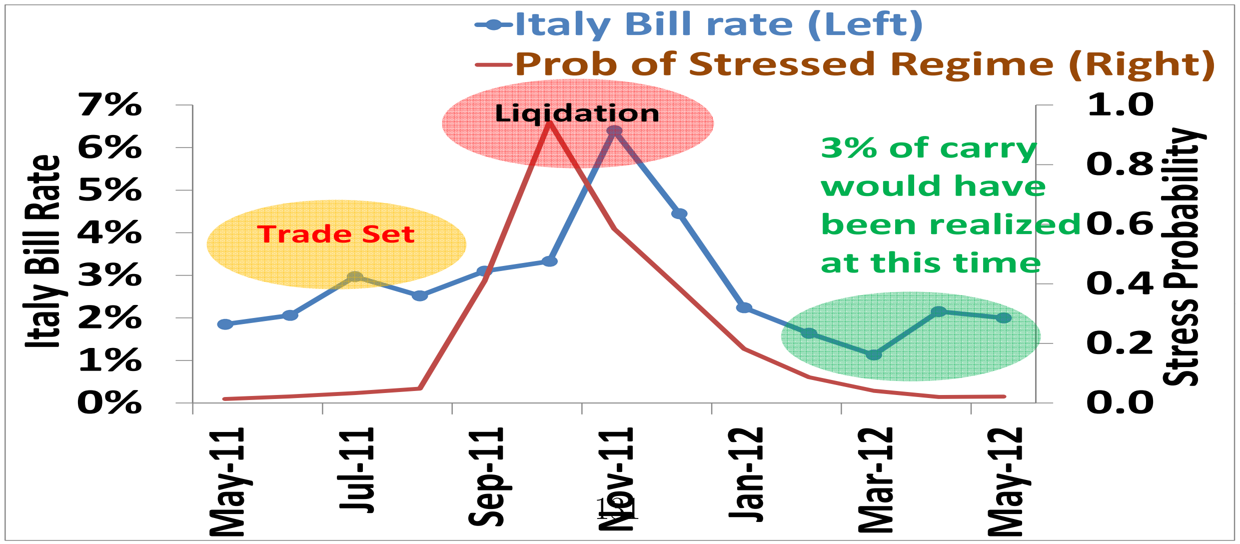
During the summer of 2011, MFG accumulated about 6.3$ billion position (about $\times 4$ of MGF equity) in Italy bills

The trade was robust fundamental view on Italy bills with about one year to maturity at yield of 3% (carry of 0.2$ billions or 12% on the equity)

But timing and position sizing were poor relative to the market cycle: the position was accumulated right in the middle of EU crisis with high probability of being in stressed regime

In October 2011, MF Global experienced a big loss on its positions and was liquidated: just before Italy bills started to rally and markets recovered (in 2016, Italy 3y bond traded close to zero and at negative rates)

Figure: Italy bills rate (left) and probability of stressed regime (right) for the S&P 500



10 Managing market cycles and tail risk of volatility strategies

To generate alpha from selling volatility, the strategy need to model market cycles and size positions accordingly

In good times:

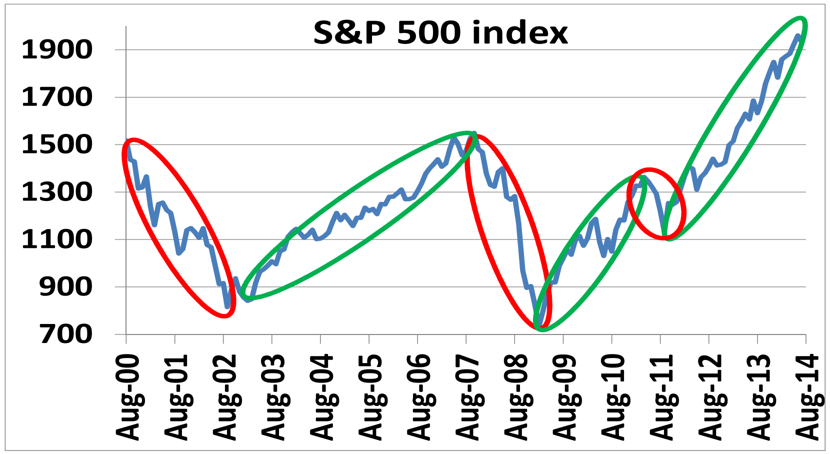
- Exposure to index returns is removed by delta-hedging
- Residual risks of volatility premiums (gamma, vega) are idiosyncratic so for diversified portfolio these are negligible

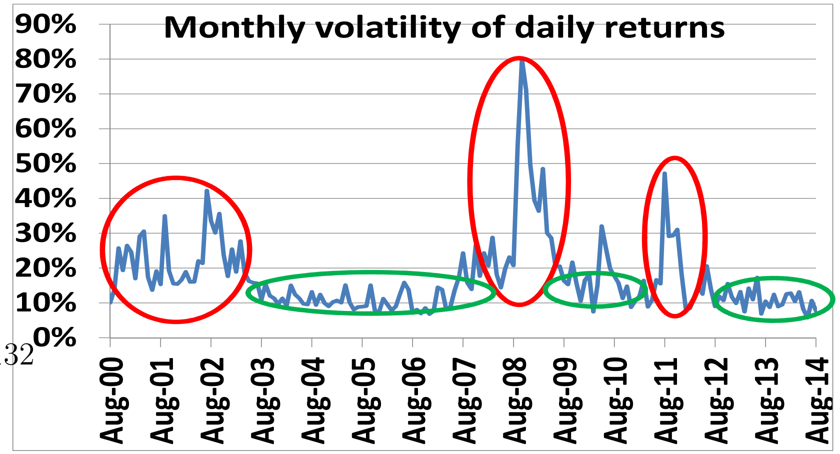
In bad times: Residual risks become common with risk-premiums spiking down

Volatility risk-premiums is interpreted as compensation to bear losses in bad market cycles

Market cycles are evident in empirical data - We can clearly spot the pattern of periods with positive drift and low volatility followed by periods of large negative returns and high volatility:

- **Right:** Time series of the S&P 500 index from Jan 2000
- **Left:** Time series of realized historical volatility of daily returns within each month





Market cycles are statistically significant with fat tails of returns explained by regime changes in realized volatility

Right: QQ-plot of monthly returns on the S&P 500 index

Left: QQ-plot of monthly returns normalized by the realized historic volatility of daily returns within given month

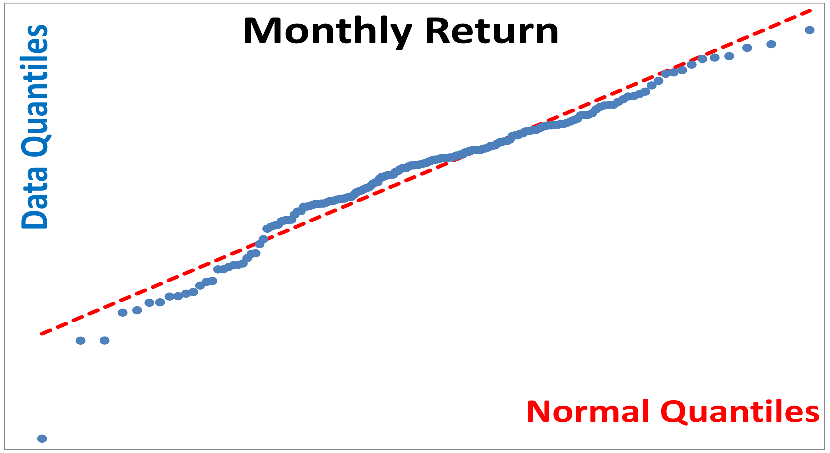
Anderson-Darling test for normality of returns (H0 hypothesis)

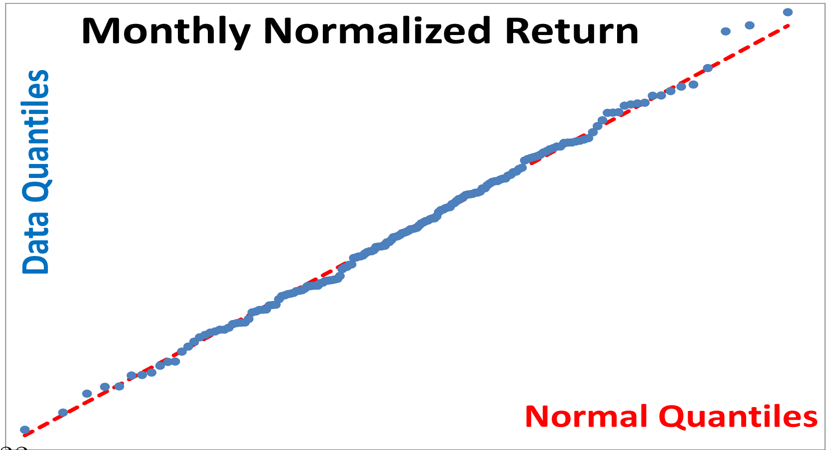
	Returns	Normalized Returns
Reject H0	Yes	No
p-value	0.0005	0.9515

We have strong presumption against normality of returns

However we cannot reject the hypothesis that volatility-normalized returns

Non-normality of returns can be explained by regimes in the volatility



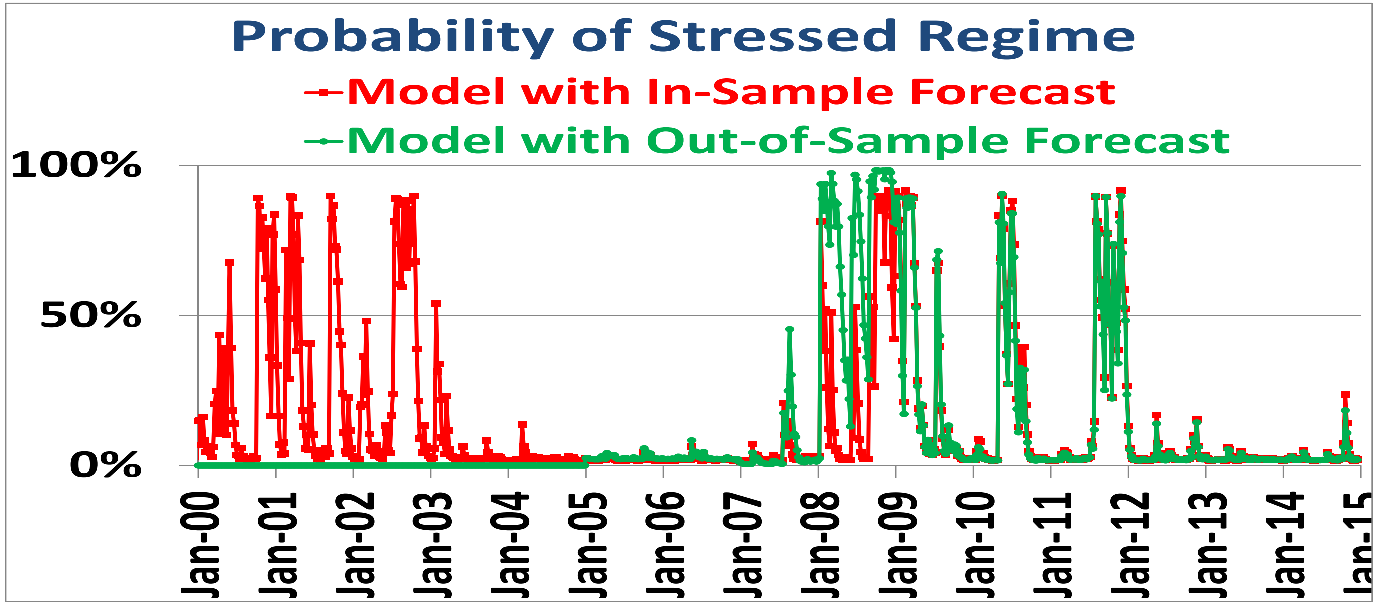


Kalman filter of Hamilton (1988) is applied to infer the probabilities of the stressed regimes out-of-sample

Apply **regime-switching model with normal and stressed market cycles**

Model statistical inference:

1. Distributions of returns in normal and stressed regimes
2. In-sample inferred probability of stressed regime from observed data
3. Out-of-sample forecast for regime probabilities using Bayesian-type update with new market data



10.1 Examples of regime identification for optimal position sizing

Maximize the log of investment value to obtain the Kelly betting rule conditional on market regimes:

$$\text{Position In Normal Regime} = \frac{\text{Expected Return In Normal Regime}}{\text{Variance of Return In Normal Regime}}$$

$$\text{Position In Stressed Regime} = \frac{\text{Expected Return In Stressed Regime}}{\text{Variance of Return In Stressed Regime}}$$

Expected Return In Normal Regime is positive so be long risk-premiums

Expected Return In Stressed Regime is negative so be short risk-premiums

Volatility selling using 1m delta-hedged short Straddle on the S&P500 index - the strategy with optimal position sizing outperforms

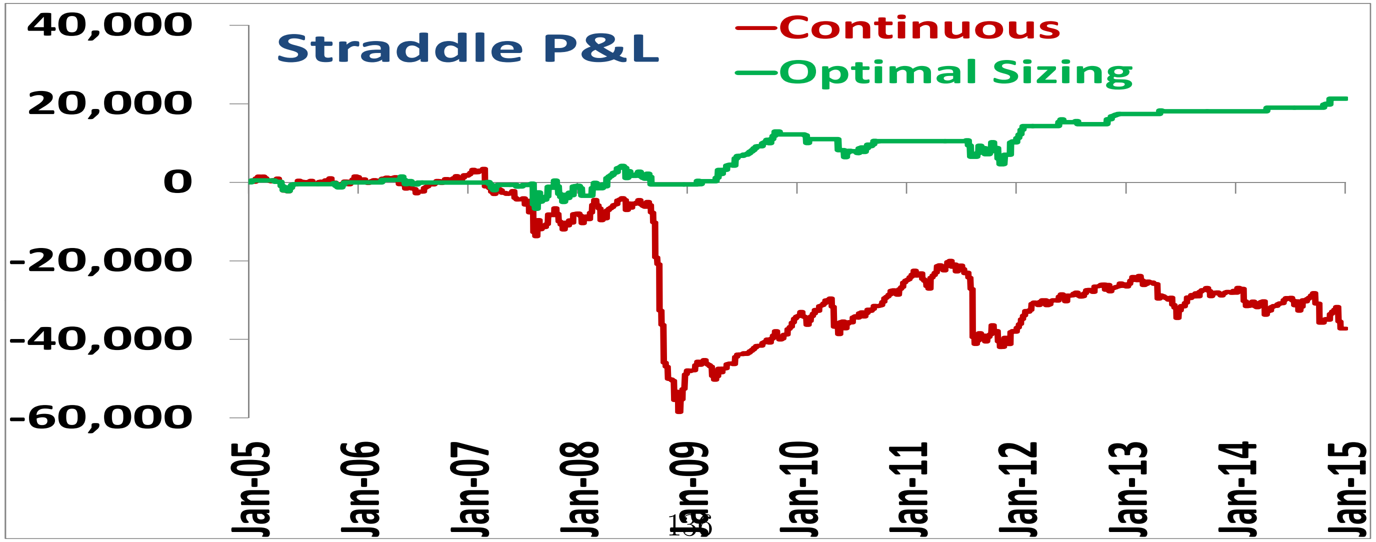
Daily re-hedging and weekly re-balancing

Transaction costs: 5bp per spot, 1% per implied vol

- **Continuous:** roll into new straddle every week
- **Optimal:** size position using out-of-sample forecast of regime probabilities:

No trading when forecast probability of stressed regimes is high

	Continuous	**Optimal**	%
Trades #	519	173	33%
Sortino	-0.39	**0.98**	
Sharpe	-0.34	**0.74**	

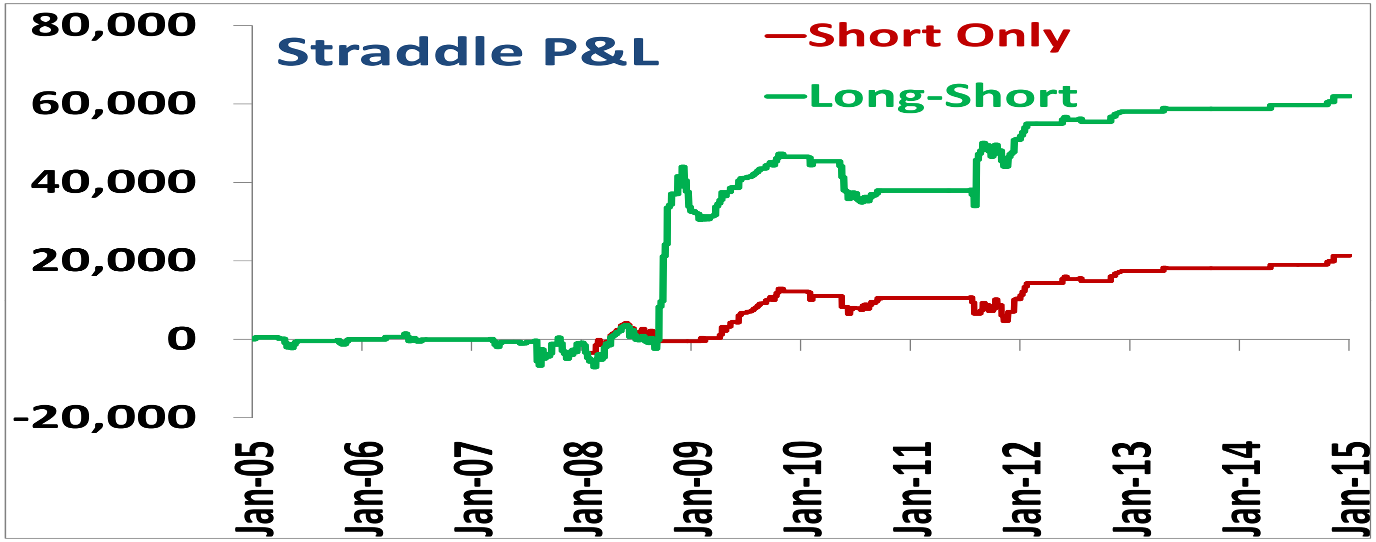


Volatility trading with 1m straddles on the S&P500 index - the strategy with optimal long/short positions

Long/short am Straddle with daily re-hedging and weekly re-balancing

- **Short Only:** sell using out-of-sample forecast of regime probabilities
- **Long-Short:** buy and sell straddles using regime probabilities

	Short Only	Long-Short	%
Trades #	173	218	26%
Sortino	0.98	**1.86**	**90%**
Sharpe	0.74	**1.04**	**40%**

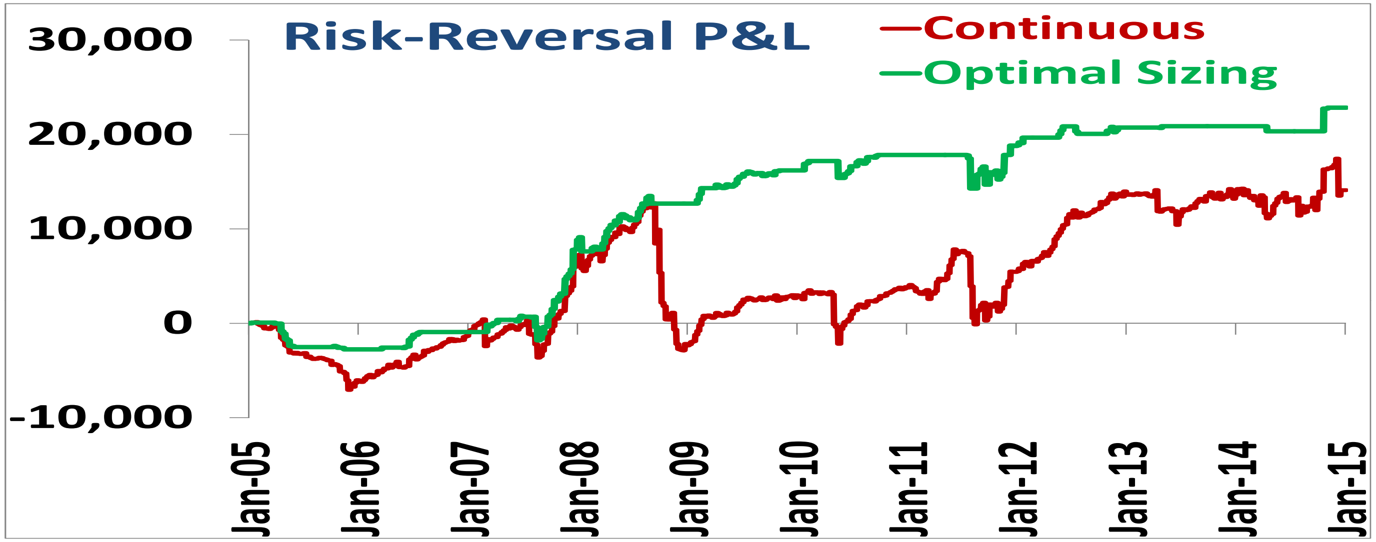


Skew selling using risk-reversals on the S&P500 index - the strategy with optimal position sizing outperforms

1m delta-hedged short 97.5%-102.5% Risk-reversal with daily re-hedging and weekly re-balancing

- **Continuous:** roll into new risk-reversal every week
- **Optimal:** size position using out-of-sample forecast of regime probabilities

	Continuous	**Optimal**	%
Trades #	519	173	33%
Sortino	0.34	**2.09**	$\times$**6**
Sharpe	0.29	**1.56**	$\times$**5**



10.2 Conclusions

Fundamental ideas: connection to Minsky theory - in 1986, American economist Minsky proposed the finantial stability theory:

1. In normal cycles of the economy, the corporate and household cash-flows rise above what is needed to pay off debt
2. A speculative euphoria emerges, soon followed by short-fall between income and debt
3. Banks and lenders tighten credit supply, followed by declines in stock markets and economic contraction
4. The Federal Reserve is a lender of last resort

"**A fundamental characteristic of our economy is that the financial system swings between robustness and fragility and these swings are an integral part of the process that generates business cycles**"

Risk-premiums can be interpreted as costs to protect against systemic shocks

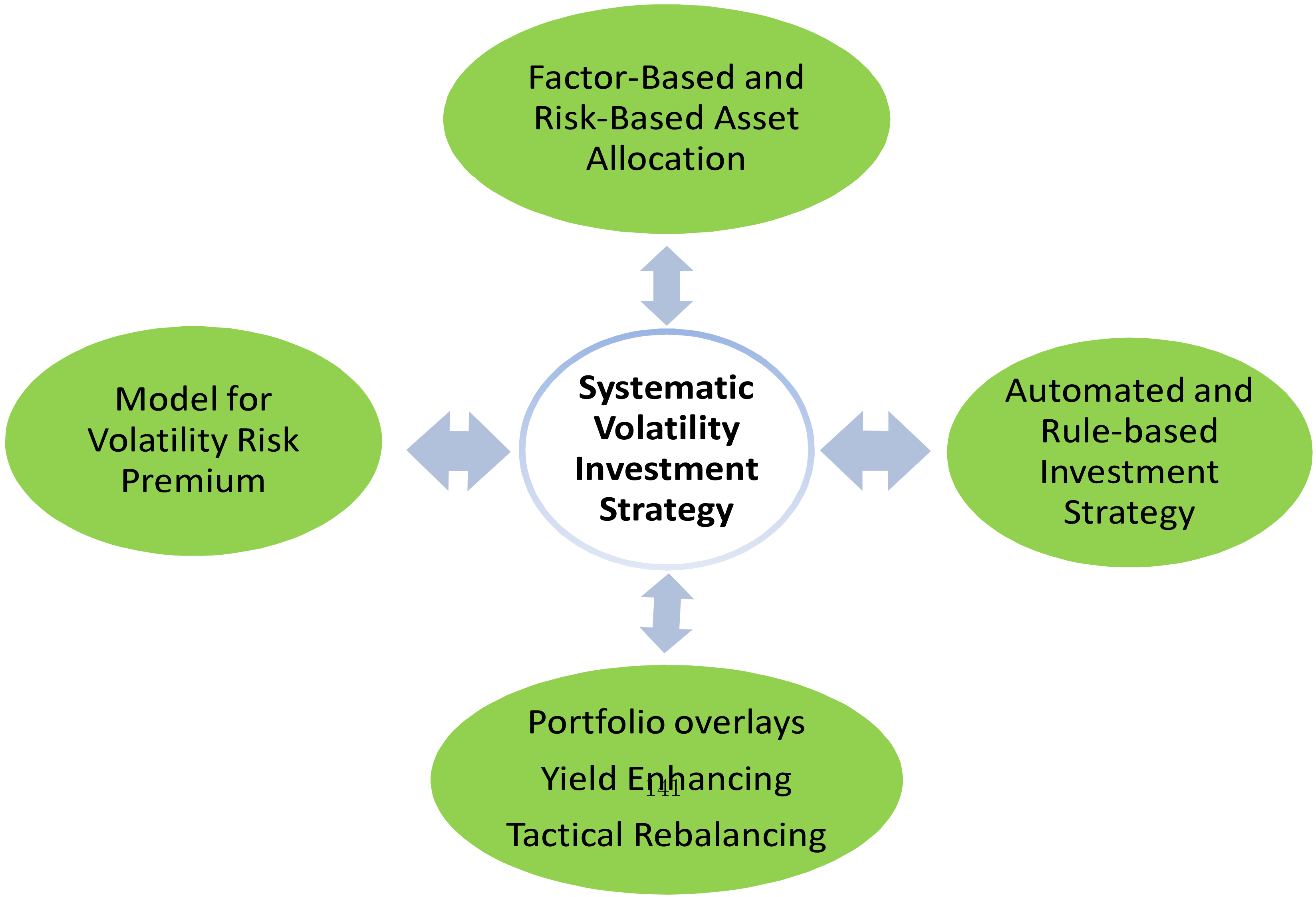
The regime-switching model with state-dependent risk-premiums fits very well into this picture

The model can help manage the exposure to market cycles

The same approach can be applied to another risk-premium strategis (credit, high yield, etc)

11 Using Volatility As A Systematic Investment Strategy

Motivation



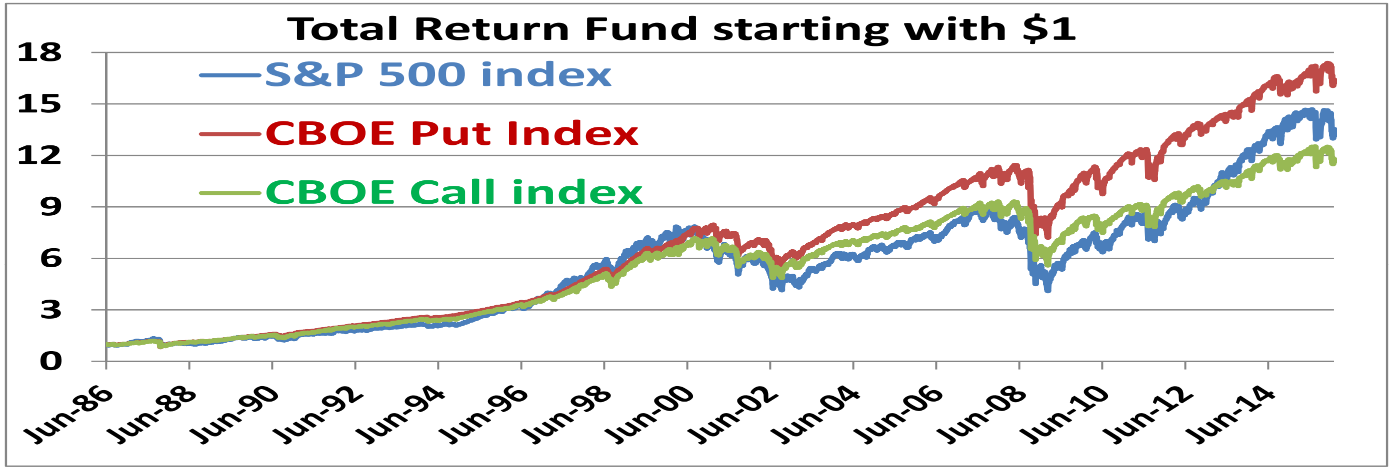
Volatility Investing is Implemented by Trading in Options

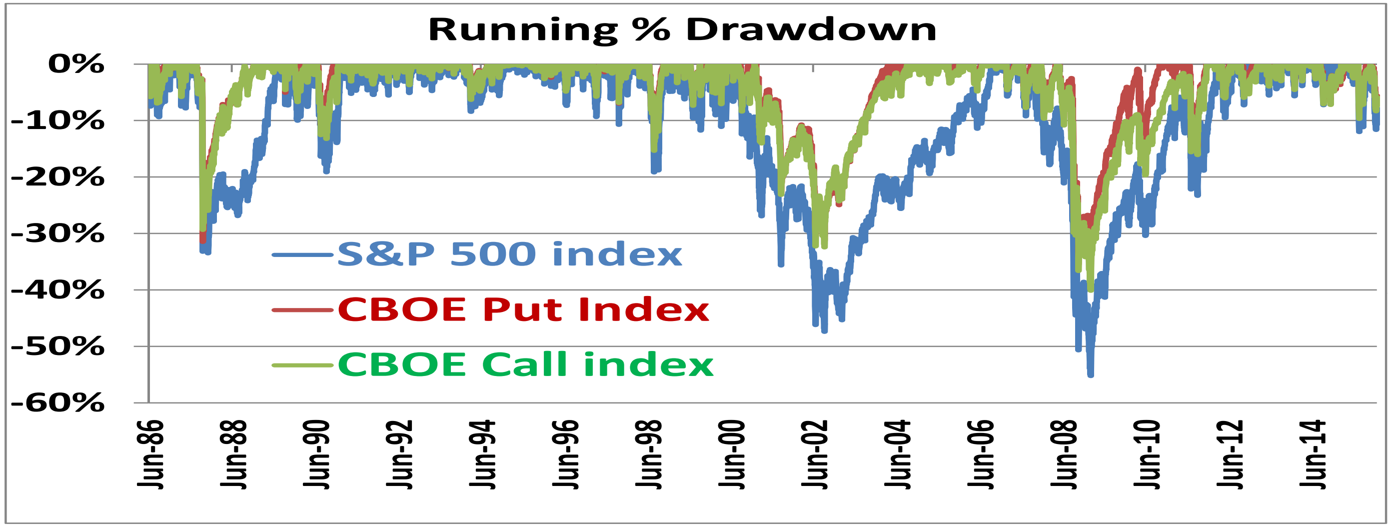
Selling volatility:

- "**Profit**" is the **traded option premium** derived from the **implied volatility**
- "**Loss**" is the **realized option premium** (through delta-hedging or without it) derived from the **realized volatility**

Realized **Profit&Loss** equals to the spread between implied and realized volatilities

Empirical evidence: systematic short volatility strategies out-perform the benchmark with smaller risk



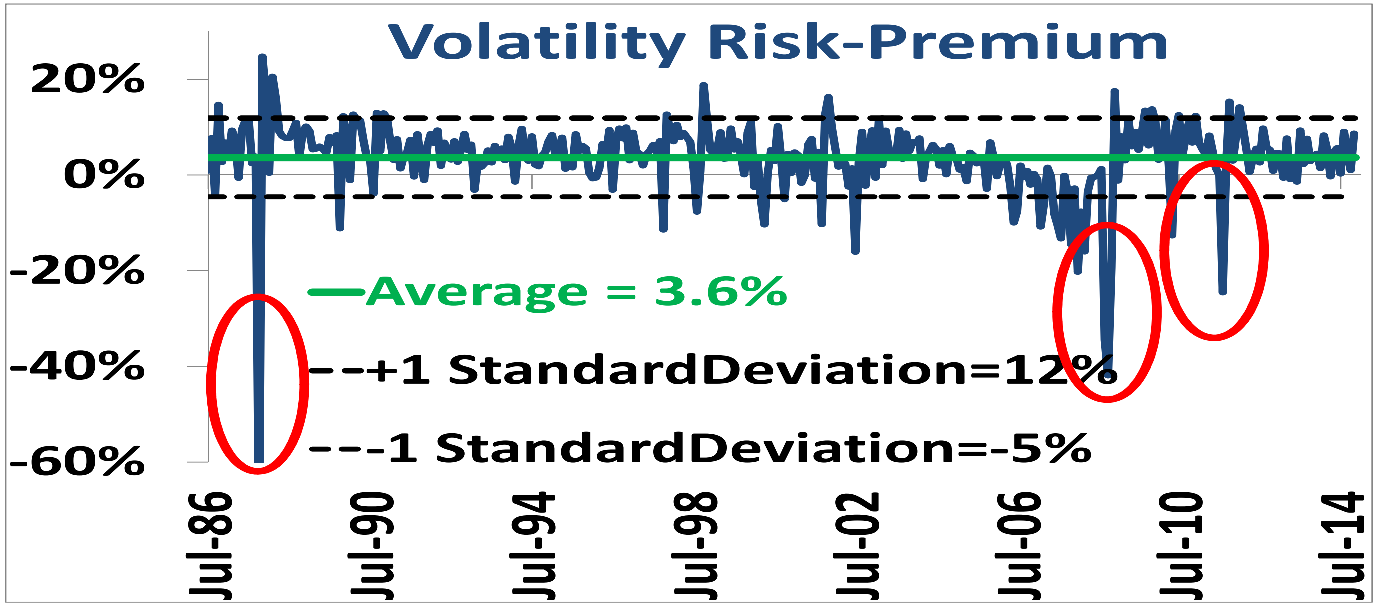


Out-performance of volatility strategies is related to volatility risk-premium

$$\text{Volatility Risk-premium} = \text{Implied volatility} - \text{Realized volatility}$$

Figure:

$$\begin{aligned}\text{Proxy Risk-premium} = {} & \text{VIX at month start} \\ & - \text{Realized volatility of S\&P500 in this month}\end{aligned}$$



Design of generic rule-based volatility indices I

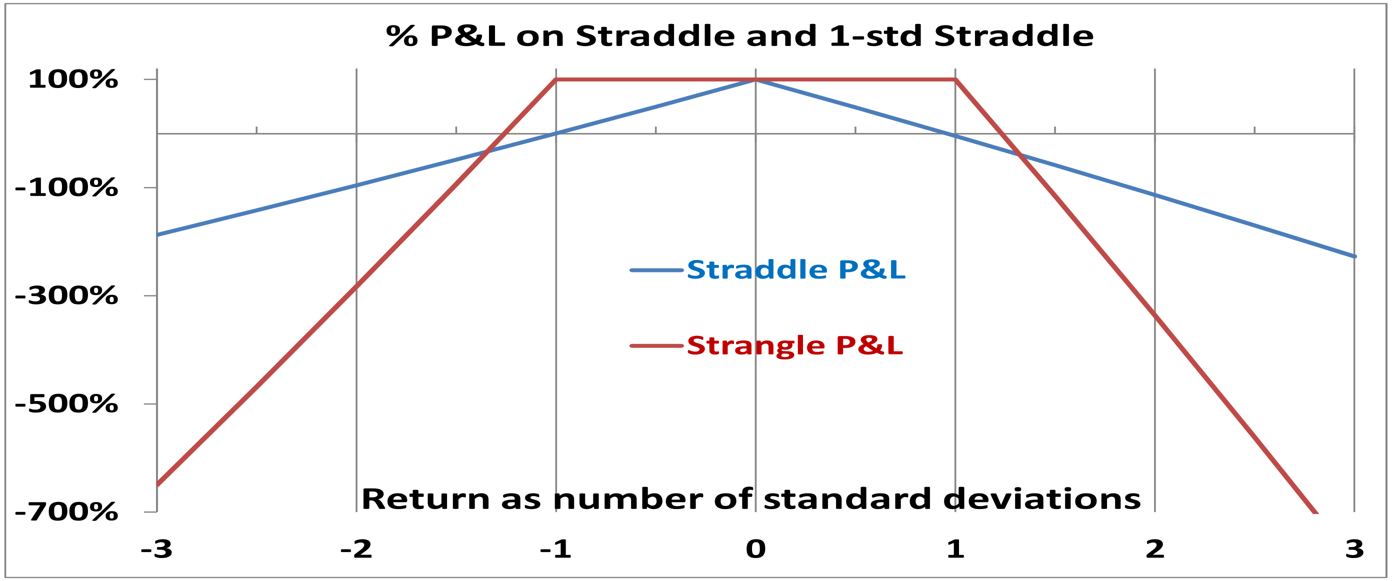
1. Liquidity:
    - No OTC
    - Large cap ETFs and stocks with liquid option markets: SPY, QQQ, IWM, TLT, XLF, AAPL, ...
2. Option Maturity:
    - Short-dated options: 1 week, 1 month, 1 quarter
    - More frequent re-balancing and capturing of volatility premium
    - Volatility skew-premium is higher for short-dated options
    - Position is held up to the expiry date: no account for vega risk

Typical Option strategies Strategies

Express returns and option log money-ness in terms of number of:

$$\text{Standard deviation} = \text{Volatility} \times \sqrt{\text{Maturity Time}}$$

Straddle: short at-the-money call and put options
Straddle P&L is positive if implied deviation is higher than realized
Strangle: short out-of-the money call and put options with strikes of one implied standard deviation away from spot price
Strangle premium picks up volatility, skew and tail premiums

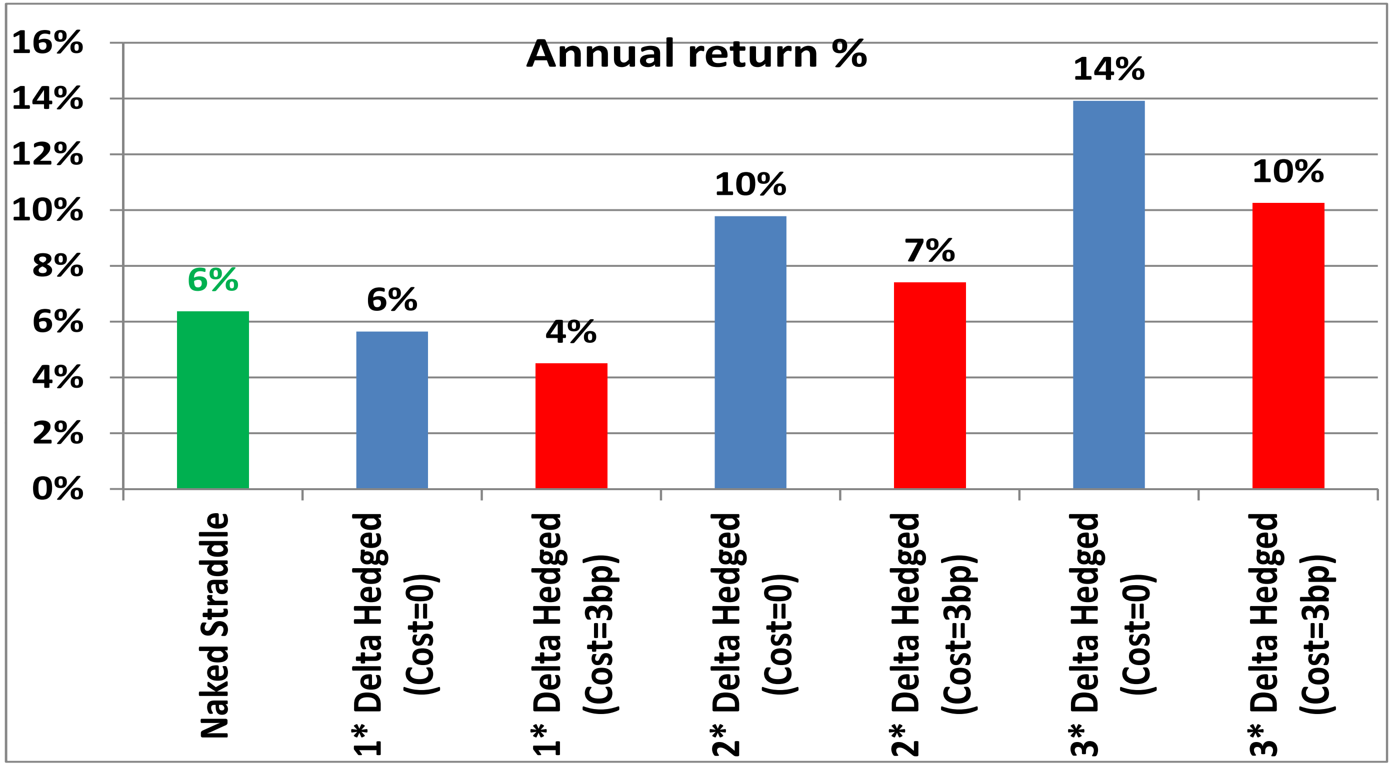


Design of generic rule-based volatility indices II

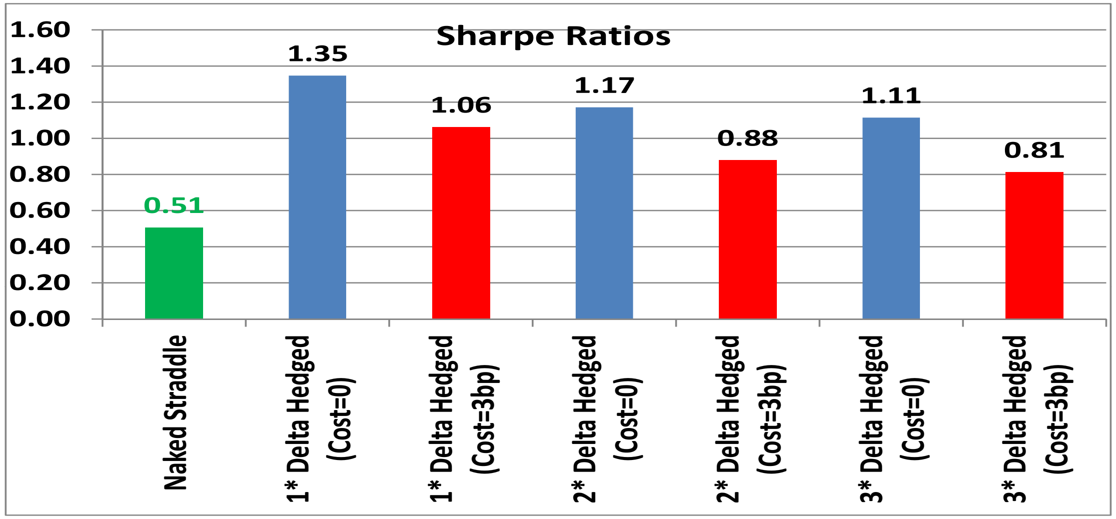
1. Position sizing:
    - Naked short: cash covers strike exposure at expiry
    - Delta-hedged: leverage is determined
2. Delta-hedged or naked:
    - Transaction costs for delta-hedge
    - Computation of option delta

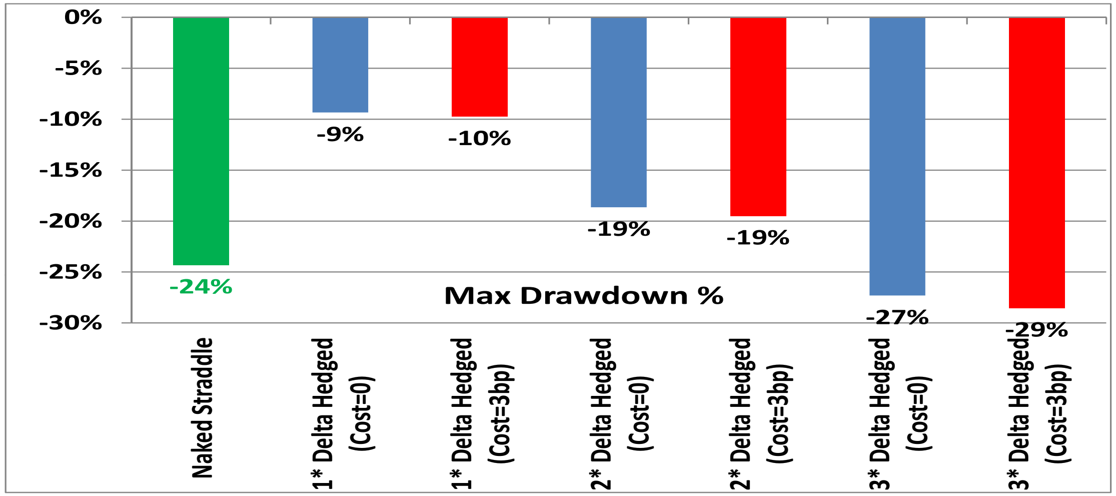
Delta-Hedge reduces reduces realized returns in presence of Transaction Costs

Monthly rolls for S&P 500 index etf (SPY) from 2005 to 2016



Increased exposure of delta-hedged position leads to higher tail risk





Covered Calls

Hybrid version of short volatility strategy

$$\begin{aligned}\text{Covered at-the-money call} &= \text{Long Stock} + \text{Short Call}\\ &= \text{Short Put}\\ &= 50\%\ \text{Long Stock} + 50\%\ \text{Short Straddle}\end{aligned}$$

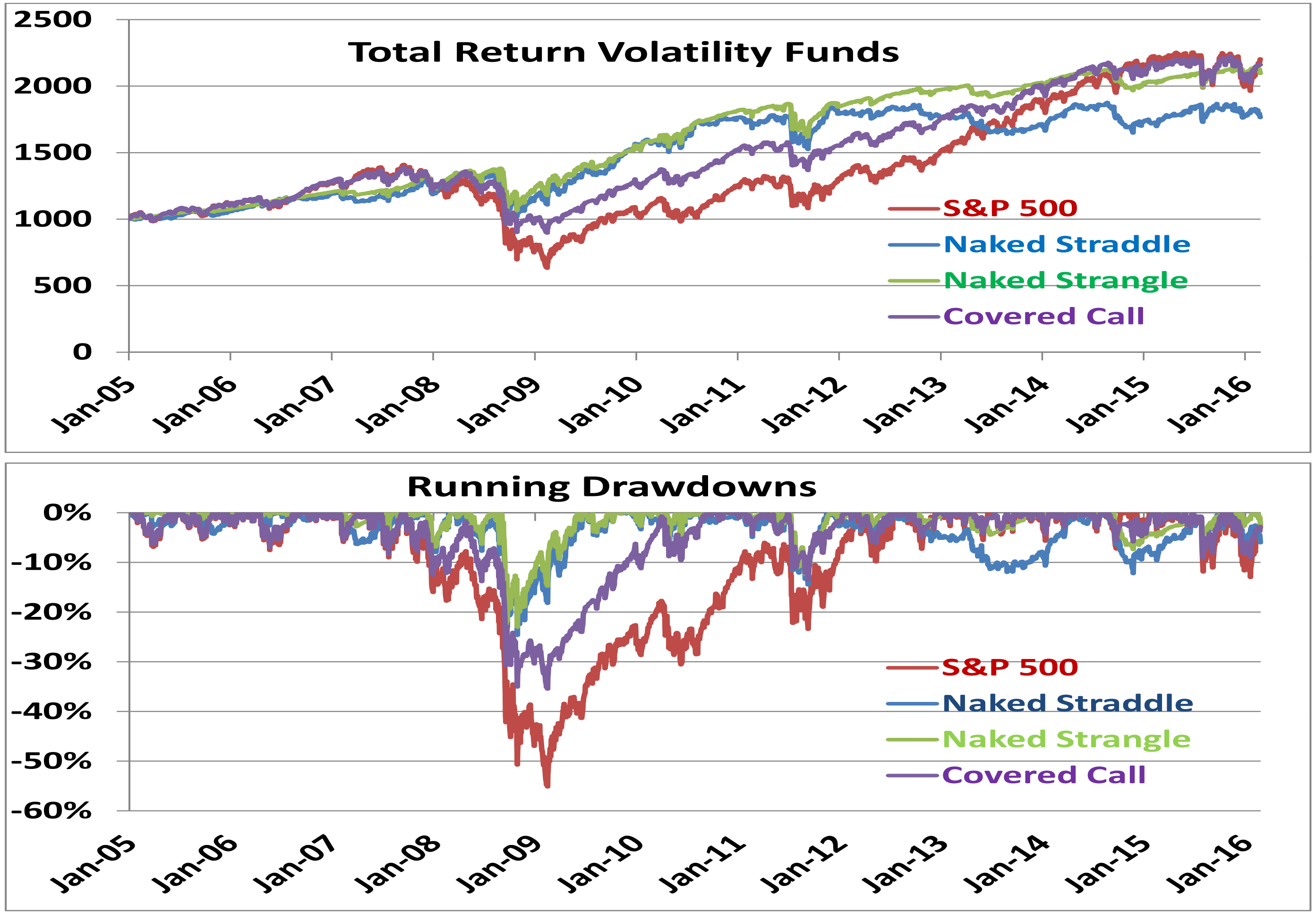
+ Implied vs realized volatility carry but long-term income is generated only if implied volatility is higher than realized

+ 50% exposure to Stock price (near at-the-money)

+ 100% Dividends

- Cost of funding for long stock

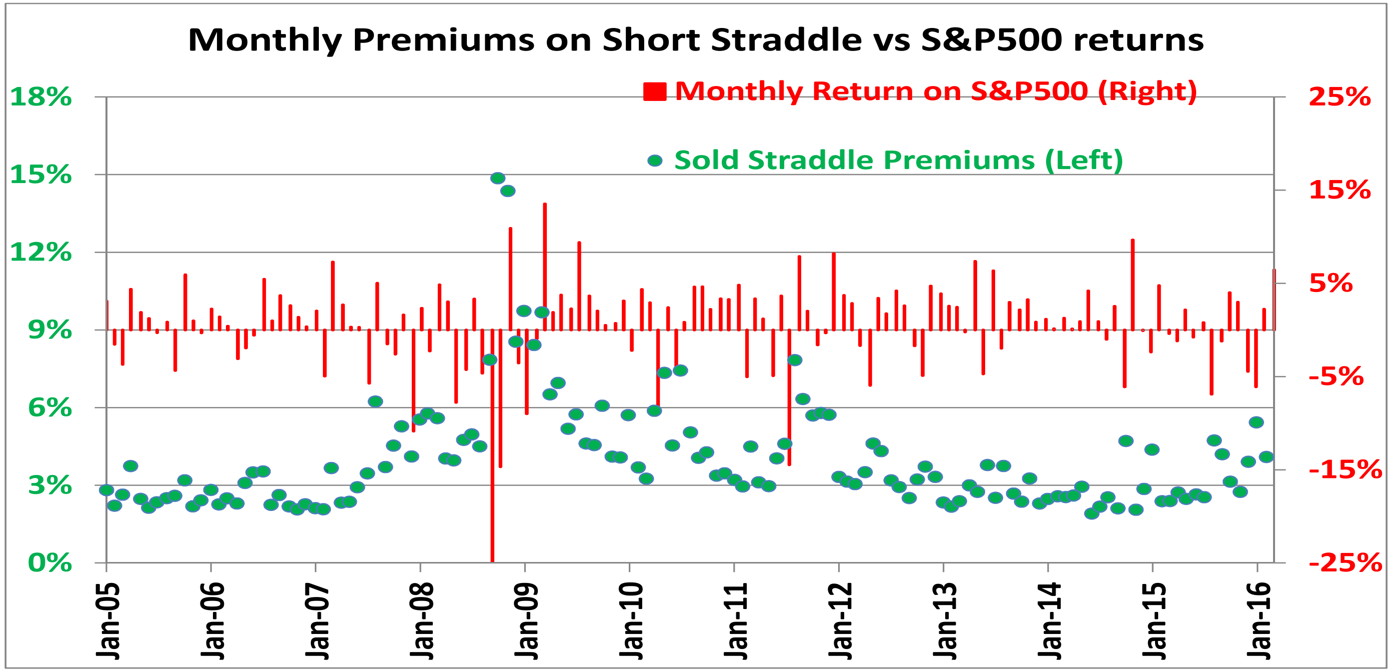
Generic volatility indices for S&P 500 index ETF (SPY) produce robust risk-adjusted performance since 2005



Systematic rule-based investment policy is crucial for long-term performance

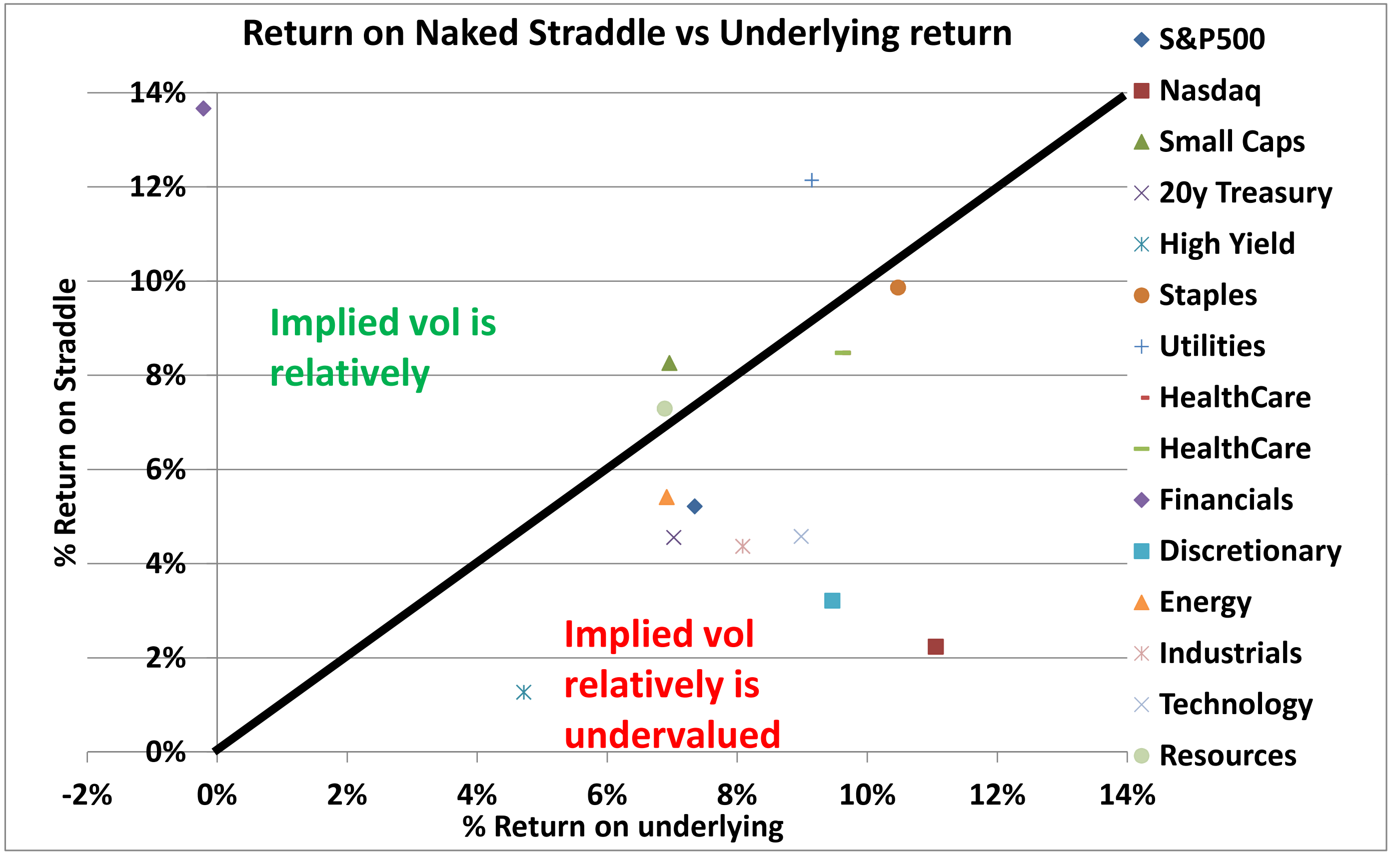
High premiums are realized following periods of market corrections
To smoothed the cyclical volatility:

- Manage macro risk with a proper model
- Portfolio overlay with allocation to "safe-haven" assets



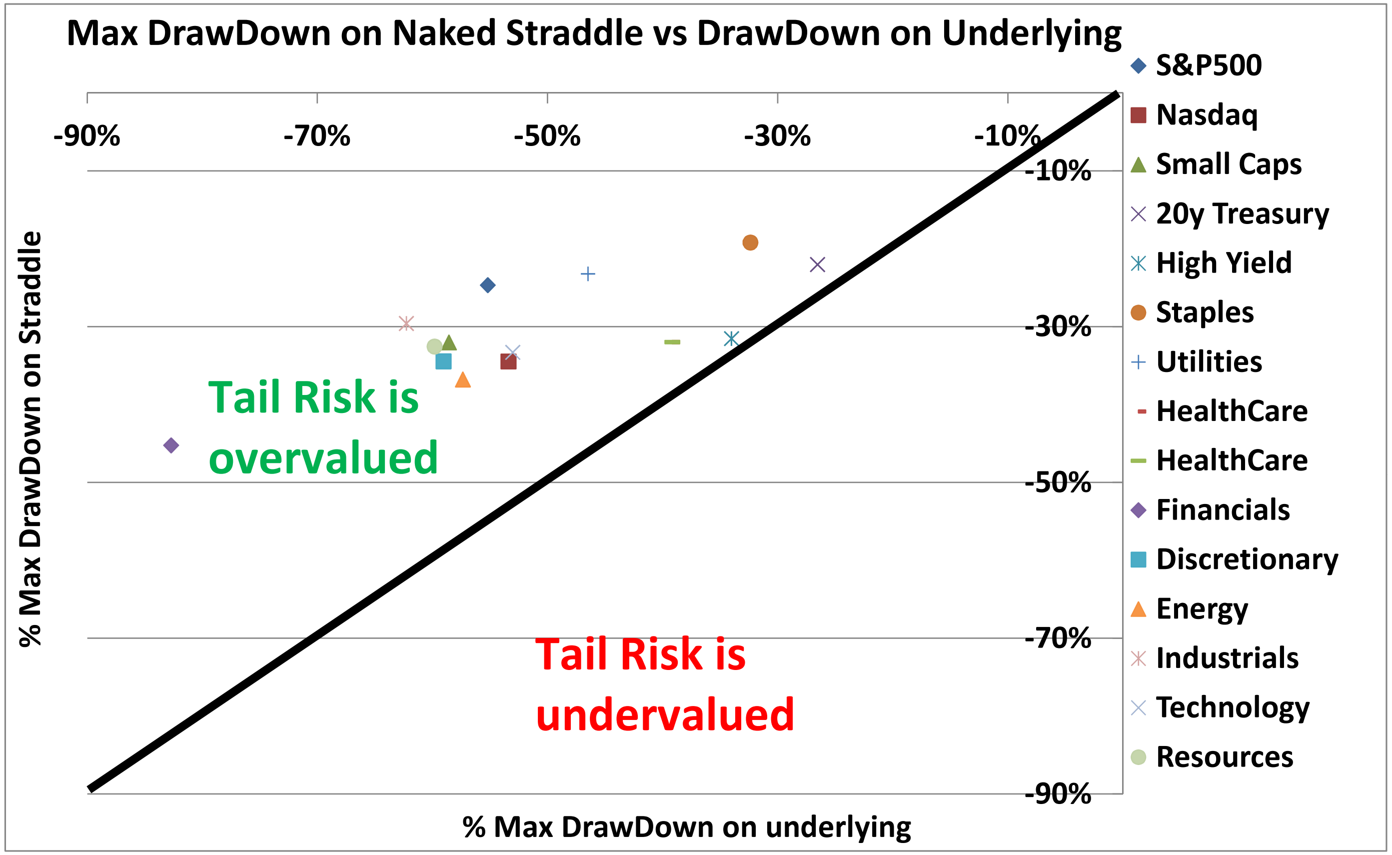
Return on Naked Short Straddle vs Return on Underlying for Broad Indices and Sector ETFs

Outliers +: Financials (XLF), Utilities (XLU), Small Caps (IWM)
Outliers -: Nasdaq (QQQ), Discretionary (XLY), High Yield (HYG)



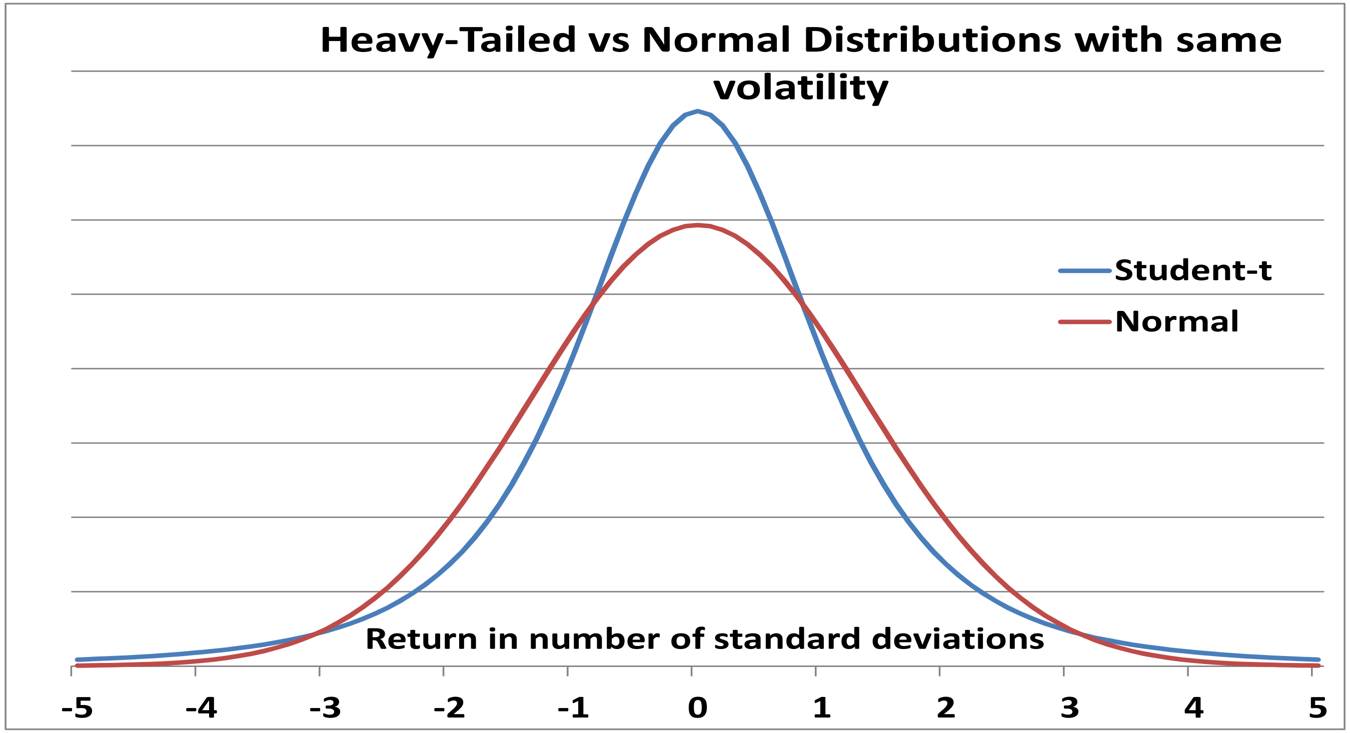
Max DrawDown on Naked Short Straddle vs Max DrawDown on Underlying for Broad Indices and Sector ETFs

Tail is risk is significantly and consistently overvalued



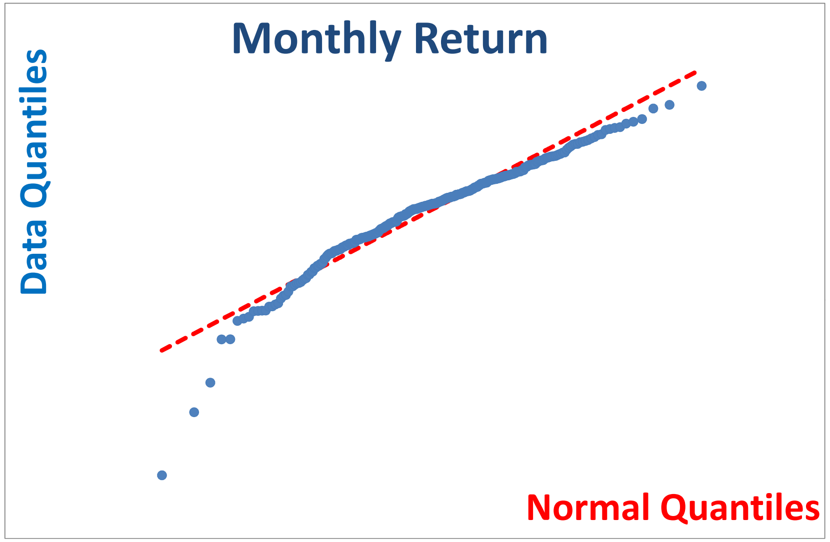
Challenge on quantitative side is how to explain the persistence of the risk-premium

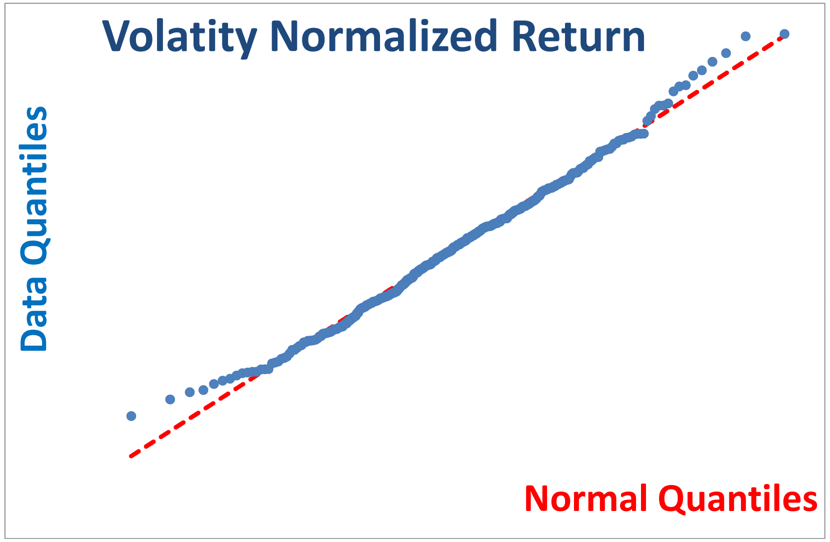
"Normal" volatility vs volatility of heavy-tailed distributions



Fat tails can be explained by market cycles with regime changes in realized volatility

Left figure: QQ-plot of monthly returns on the S&P 500 from 1986
Right: QQ-plot of monthly volatility-normalized returns
Volatility risk-premium is derived from fat tails





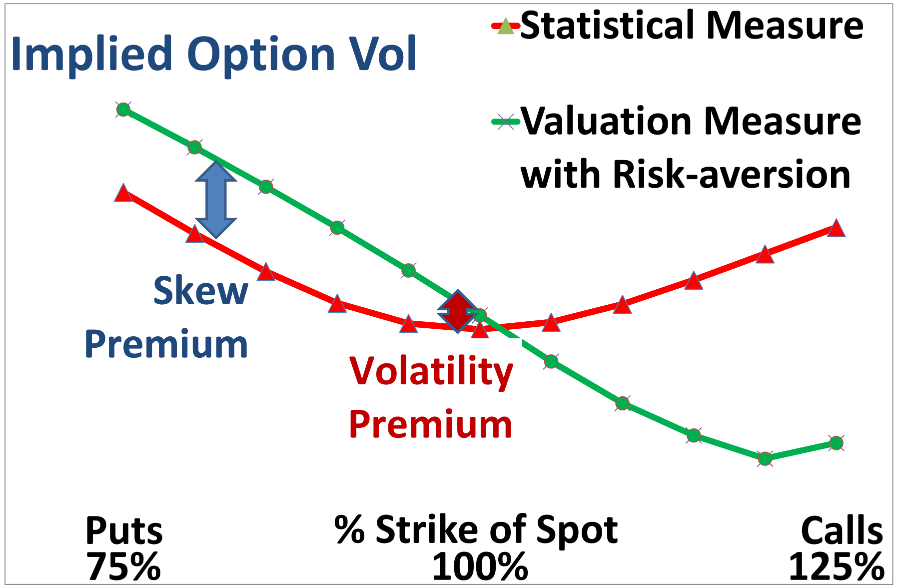
Quantitative model implies that the volatility risk-premium arises from:

- **Fat tails of the statistical distribution of returns**
- **Investors' risk-aversion**

As driver for P&L:

$$\text{Risk-premium} = \text{Implied risk} - \text{Realized risk}$$

risk = volatility, skew etc



Volatility carry is the most significant factor for a volatility strategy

$$\begin{aligned}\text{Short Option P\&L} &= \text{Implied Costs} - \text{Realized Costs} \\ &= \text{Alpha} \\ &+ \text{Beta To Asset} \times \text{Realized Return} \\ &+ \text{Beta To Vol Premium} \times \text{Realized Vol Premium}\end{aligned}$$

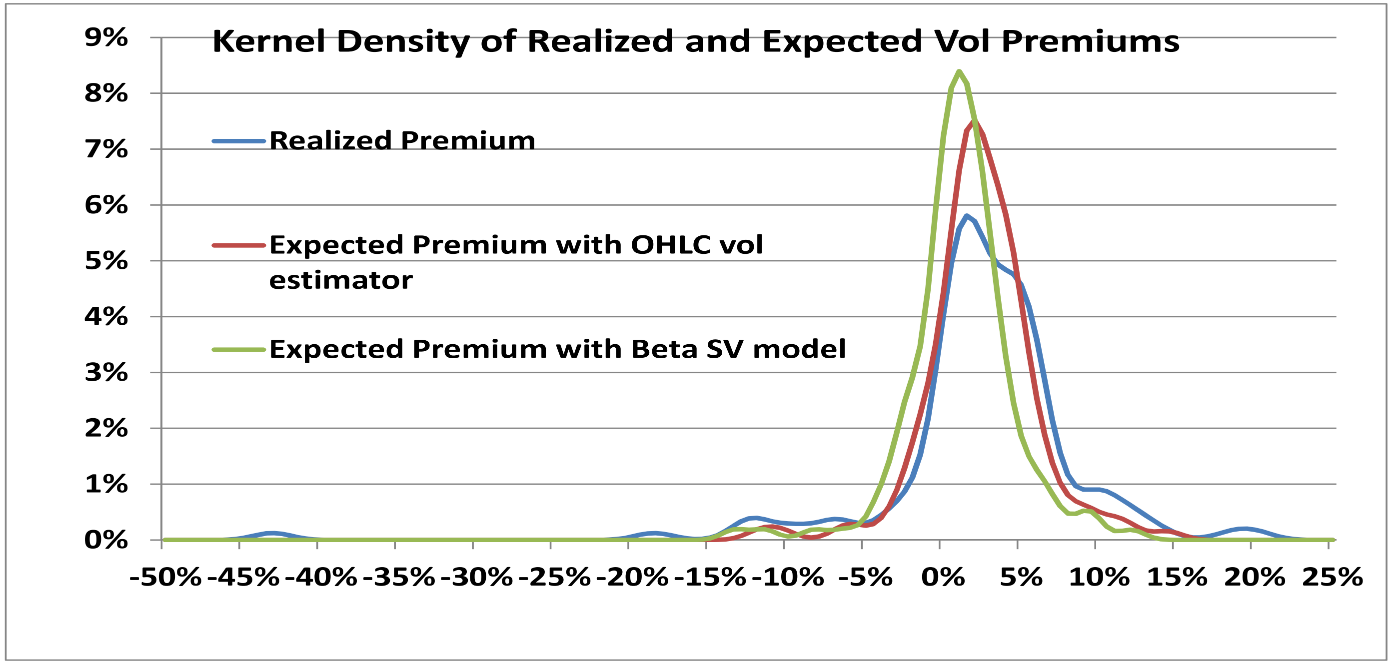
Empirical regression with only significant coefficients for vol strategies on S&P500 ETF

Strategy	Alpha	S&P500	Vol Premium	R^2
Naked Straddle			0.26	31%
Naked Strangle			0.26	45%
Delta-Hedged Straddle (×2)	3%		0.22	52%
Delta-Hedged Strangle (×2)	4%		0.20	59%
Covered Call		0.54	0.09	88%

Modeling expected volatility risk-premium

Necessary to deliver "alpha"

Apply the expected risk-premium to size exposures

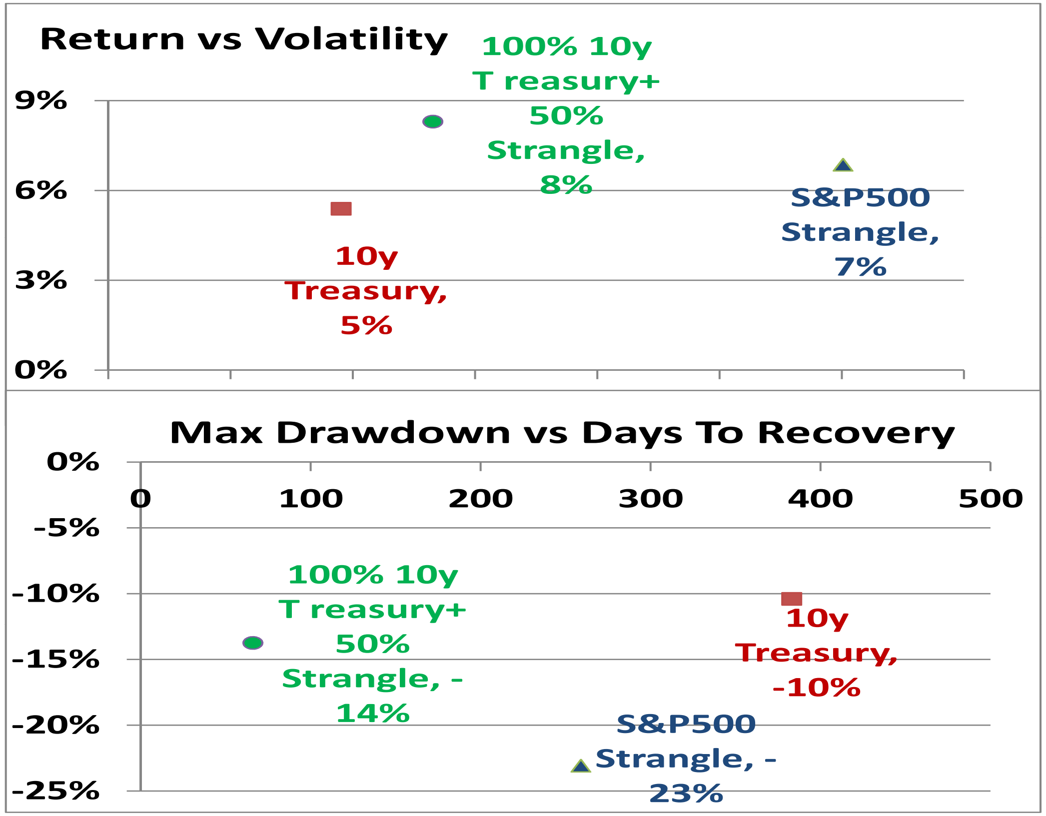


Design of generic rule-based volatility indices on multi-assets

1. Allocation type for individual asset:
    - Long only
    - Option overlay
2. Reduce tail risk of strategy by being long "safety" assets and selling options on "risky" assets
3. Factor-based risk management for long position and option exposures:
    - Static "strategic" weight
    - Risk-based weight

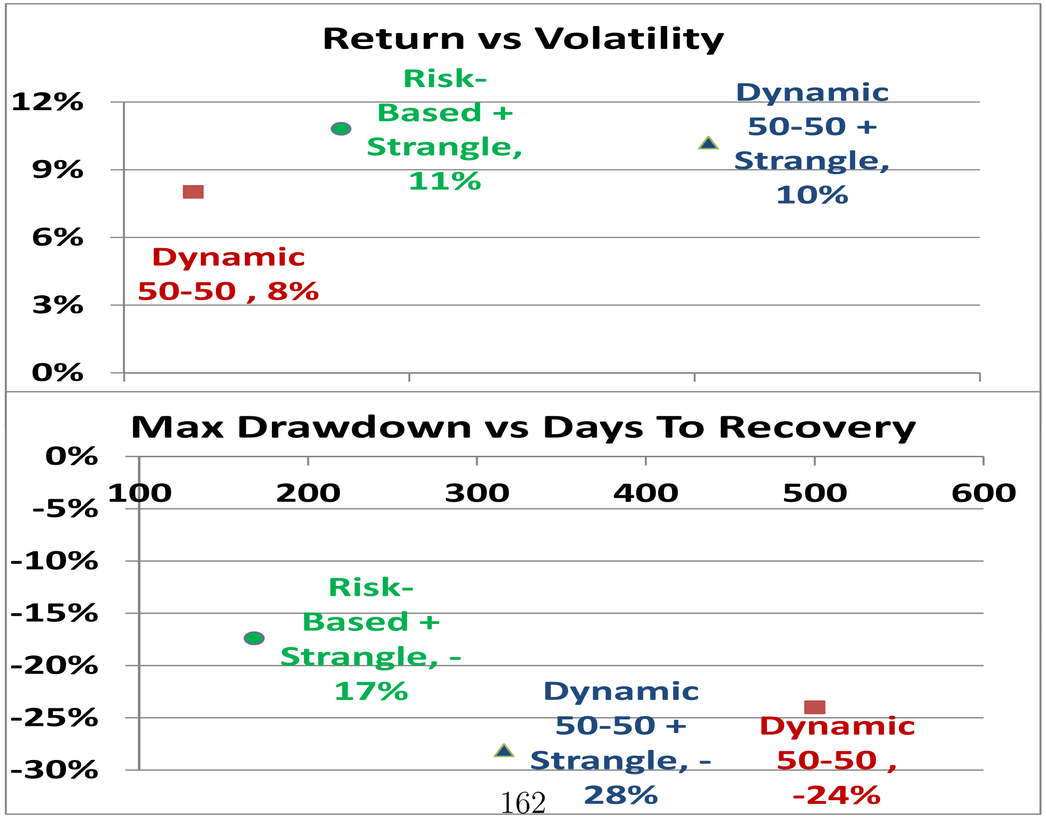
Application I: bond portfolio yield-enhancing strategy in 10y treasury ETF (IEF) with 50% strangle on S&P 500 index ETF (back-test from 2005)

1) 100% allocation to 10y Treasury ETF
2) cash + 100% S&P500 Strangle
3) 100% allocation to 10y Treasury ETF + 50% Strangle



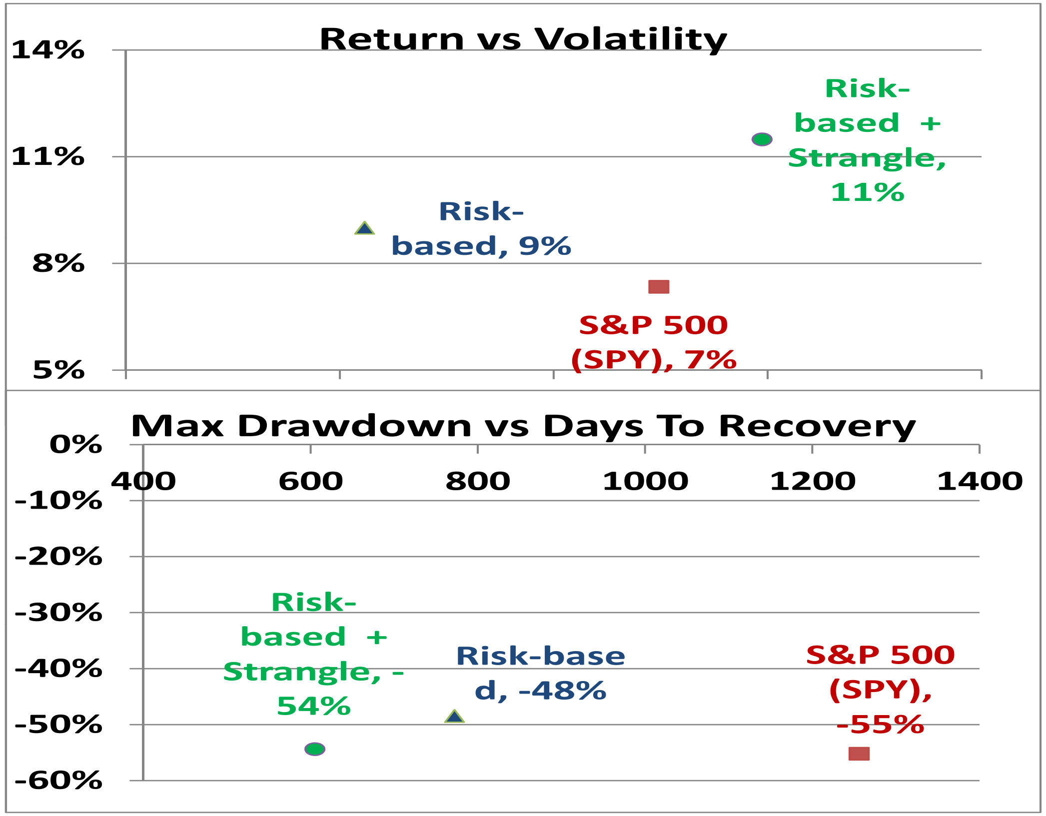
Application II: tactical re-balancing of strategic asset allocation with 50%-50% in 20y treasury bond ETF (TLT) and S&P500 index ETF (SPY) using strangles in both ETFs

1) 50%-50% allocation
2) 50%-50% + Strangles
3) Risk-based allocation + Strangles



Application III: full equity portfolio with risk-based allocation to 9 sector ETFs (XLE, XLF,...) with Strangle overlays

1) S&P 500 index ETF (SPY) as the benchmark
2) Risk-based allocation to 9 sector ETFs
3) Risk-based allocation + Strangles overlay



12 References

Beta stochastic volatility model:

Karasinski, P., Sepp, A., (2012), "Beta stochastic volatility model," *Risk, October*, 67-73
http://ssrn.com/abstract=2150614

Sepp, A., (2016), "Log-Normal Stochastic Volatility Model: Affine Decomposition of Moment Generating Function and Pricing of Vanilla Options," *Working Paper*
http://ssrn.com/abstract=2522425

Computing option delta consistently with empirical dynamics:

Sepp, A., (2014), "Empirical Calibration and Minimum-Variance Delta Under Log-Normal Stochastic Volatility Dynamics"
http://ssrn.com/abstract=2387845

Optimal delta-hedging strategies for discrete hedging with transaction costs:

Sepp, A., (2013), "When You Hedge Discretely: Optimization of Sharpe Ratio for Delta-Hedging Strategy under Discrete Hedging and Transaction Costs," Journal of Investment Strategies 3(1), 19-59
http://ssrn.com/abstract=1865998