Empirical Finance

From Time Series Analysis to Risk Neutral Pricing

Florian ielpo

Centre d’Economie de la Sorbonne (CERMSEM) – Paris, France
and
Lombard Odier Investment Managers – Geneva, Switzerland

January 2015
Questions that probably run in your minds

① Why do I need to learn time series analysis when I’m already drowning into stochastic calculus?
② Why do I need to learn the Black Scholes model and its tedious calculations since it’s been long outdated?
③ What is this mystical "leverage effect" mentioned in so many books and papers and does it really matter?
④ Is there a connection between the "risk neutral" distribution and the time series of returns?
⑤ If the world is really non Gaussian, why care about Gaussian models?
⑥ What is the problem of econometricians with the Heston model? Why don’t they use it instead of their GARCH?
⑦ Can I price options with a GARCH model? Is it worth the trouble?
⑧ ...
Introduction

Basic models in finance:

- The Markowitz's approach to portfolio management.

Let $\mu_i$ and $\sigma_i$ be the expected return and volatility associated to asset $i$. In the single asset case, the investor should invest its money accordingly to the Sharpe ratio multiplied by risk aversion:

$$\omega_i = \frac{\mu_i}{\sigma_i} \times \frac{1}{\gamma - 1},$$

where $\gamma$ is the relative risk aversion of the representative agent.

- The Capital Asset Pricing Model (CAPM):

$$r_i = r_f + \frac{\sigma_{i,m}}{\sigma_m^2} (r_m - r_f)$$
• The **Vasicek** term structure model:

\[ dr_t = \kappa(\theta - r_t)dt + \sigma dW_t \quad (3) \]

• The **Black Scholes** model:

\[ \frac{dS_t}{S_t} = \mu dt + \sigma dW_t \quad (4) \]

• ... and the risk management approach of many banks: VaR computation from Monte Carlo scenario involving both expectation and volatility, the tracking errors essential to fund management...

These models are based on the assumption that *we can measure expected return and volatility*. This is however in general false.
Focusing on volatility

Both expectation and volatility are non measurable quantities.

Most of the academic efforts are directed at volatility modeling, whereas most of the asset management attention is attracted by expected returns.

**Two explanations:**

1. Most of the academic research is directed at risk modeling and little to returns prediction: the second order moment is easier to modellize than the first order one, because of its persistence.
2. Should anybody find a way to predict efficiently the expected returns, then it would be obviously foolish to make publications and to teach classes about it.
Which volatility?

Two kinds of volatility need to be disentangled:

1. the **historical volatility**: the volatility associated to the time series behavior of asset prices.
2. the so-called **risk neutral volatility**, reflecting the willingness of investors to buy and speculate over protection.

The ”risk neutral” world is a theoretical world in which either:

1. the expected returns are equal to the current risk free rate for the simplest models
2. where agents are neutral toward risk taking: it does not only involve the expected return, but also higher order moments.
Which volatility?

**Point 1:** Historical and risk neutral volatility are related but not equal.

**Point 2:** Risk neutral volatility is a function of option prices that is driven by:

- the *hedging demand*: buy a stock and a put will offer you the protection regarding your maximum loss. The market being structurally long, the buying of protection is high.

- *volatility trading*: buy calls or puts out-of-the-money and hedge them against the underlying’s moves and accelerations.

On average, the risk neutral volatility is higher than the historical one: the *volatility premium* is negative.
Organization of the class

1. Modelling the historical volatility
   (a) Volatility’s stylized facts
   (b) The time series approach to financial markets laws of motion
   (c) The use of high-frequency data

2. Modelling the risk neutral volatility
   (a) Stylized facts about implied volatilities
   (b) Models for the risk neutral volatility
   (c) The computation and the information content in the volatility indices

3. Bridging the gap between historical and risk neutral distributions
   (a) The pricing kernel and aggregated market risk aversion
   (b) Estimation strategies to the pricing kernel
   (c) The consistency between the RN and the historical distribution
Readings

Core readings:

Other readings:
- J. Hull. Options, futures and other derivatives. Prentice Hall.
I. Modelling historical volatility

I-1. Volatility’s stylized facts
The historical volatility of financial markets

- Understanding the phenomenon is essential before thinking of building a volatility model.
- Volatility is usually considered as a risk measure.
- Volatility is the squared root of variance:
\[
\sigma(X) = \sqrt{V[X]} \tag{5}
\]
- Variance measures the expected quadratic spread between a random variable and its expectation
\[
V[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[x])^2 f(x) dx \tag{6}
\]
- It measures *dispersion*: the higher the vol and the higher the probability that the random variable reaches a value far from its expectation.
- It says nothing about asymmetry and the rate of decay of the tails of the random variable
Volatility estimation

– When a random variable is stationary and ergodic or iid, volatility can be estimated using the method of moment estimator.
  – A random variable is stationary whenever its expectation and autocorrelation structure do not depend over time.
  – A process is ergodic when its statistical properties can be estimated from a single sample of it.
    More technical definitions are available.

– The method of moment estimator is then:

\[
\mathbb{E} [X^i] = \frac{1}{n} \sum_{t=1}^{n} (x_t)^i
\]  

(7)

– This is a direct application of the Law of Large Numbers.
Stationary Processes

• A process $(X)_{t \in \mathbb{Z}}$ is strictly stationary if for all $t_i$ and $t_i+h$, the sequence $x_{t_i}, x_{t_i+1}, \ldots, x_{t_n}$ has the same distribution as $x_{t_i+h}, x_{t_i+h+1}, \ldots, x_{t_n+h}$.

• A process $(X)_{t \in \mathbb{Z}}$ is second order stationary if
  \[
  \forall t \in \mathbb{Z}, \mathbb{E}[X_t] = \mu < \infty \\
  \forall t \in \mathbb{Z}, \text{Var}[X_t] = \sigma < \infty \\
  \forall t, h \in \mathbb{Z}, \mathbb{E}[(X_t)(X_{t+h})] = \gamma(h)
  \]

• Examples of stationary processes in finance:
  – Equity returns
  – Bond index returns
  – Currency rate variations
  – Interest rates in Germany and USA from 1998 onward

• Examples of non stationary processes in finance:
  – Financial prices: equities, bonds, carbon allowances...
  – Currency rates
DAX Returns

Jan-91  Jan-92  Jan-93  Jan-94  Jan-95  Jan-96  Jan-97  Jan-98  Jan-99  Jan-00  Jan-01  Jan-02  Jan-03  Jan-04  Jan-05  Jan-06  Jan-07  Jan-08  Jan-09  Jan-10

DAX Returns
-0.1
-0.05
0
0.05
0.1
0.15
Scales of volatility

When facing anything new, rough estimates are essential. Volatility is always presented in terms of annualized volatility.

Suppose \((r)_{t \in \mathbb{Z}}\) is the daily returns stationary process, with

\[
\mathbb{V}[r_t] = \sigma^2 \quad \mathbb{E}[r_t r_{t-1}] = 0
\]  

(8)

where \(\sigma\) is the daily volatility. The annualized variance is then

\[
\mathbb{V} \left[ \sum_{t=1}^{T} r_t \right] = \sum_{t=1}^{T} \mathbb{V}[r_t] = T \times \sigma^2
\]  

(9)

Hence, the rule of the thumb: annualized volatility is equal to

- \(\sqrt{252} \times\) the daily volatility
- \(\sqrt{52} \times\) the weekly volatility
- \(\sqrt{12} \times\) the monthly volatility

In general, the correlation is non zero and the annualized vol from daily data is higher than the annualized vol computed from monthly data.

The lower the sample frequency and the more Gaussian the data.
Economic fundamentals to volatility

Volatility arises during periods of fear and uncertainty. Differences in volatility is related to the exposure to risk factors.

When it comes to Bonds and Stocks, the difference in volatility stems from different factors:

- Debt holder and share holder do not bear the same level of risk: during bankruptcy, debt is payed back before shares.
- The situation is the same amongst of bond holder: the degree of seniority leads to different levels of risk.
- The industry type as well as the rating of the firm explains the different in risks, from investment to speculative grade/high yield.
- Finally, the country to which the firm is mostly related can bear different risks, such as a political risk.

In the case of commodities, supply and demand games explain the hierarchy of risks.
Balance Sheet and Risk

**Assets**
- Senior Debt
- Subordinated Debt:
  - Lower Tier 2
  - Upper Tier 2

**Liabilities**
- Senior Debt
- Subordinated Debt:
  - Lower Tier 2
  - Upper Tier 2
  - Tier 1
- Shares

Dropping chances to see your money back during a bankruptcy episode
Stylized facts #1: volatility is time varying

Risk perception changes through time.

It is related to different fundamentals:

- Global economic stance of the world – developed and emerging markets: as profit perspectives drop, bankruptcy risks rise sharply and prices drop.
  These periods are cyclical.
  Volatility cycles match the NBER dating of economic crises.

- The order relation between asset classes is however globally unchanged: the volatility of low volatility asset classes remain below the one of high volatility asset classes.

- Most of the asset classes have a volatility that move together.
Volatility and economic growth

Volatility and economic growth
Stylized fact #2: the clustering effect

The key elements:

1. Volatility happens by clusters.

2. Periods of high volatility have a tendency to last for periods of a lower length than low volatility ones.

Rationale:

- Volatility reflects fear amongst market participants. When the market is shocked, it takes some time for the shock to vanish.

- Volatility is related to the business cycle: crisis last for several months. Periods of expansion are usually longer than periods of expansion.

- Volatility reflects fear toward specific cases. In the 2010 Greek debt case, volatility cannot go down before a institutional solution is proposed.
Stylized fact #2: the clustering effect

In a few textbooks, volatility clustering is disentangled from the Taylor (1986) effect.

Absolute values for returns is the returns’ transformation creating the biggest autocorrelation.

The autocorrelation function:

\[ \gamma(h) = \frac{\text{Cov}(r_t, r_{t+h})}{\text{Var}[r_t]} \]  \hspace{1cm} (10)

A process is said to be persistent whenever \( \gamma(1) \) is statistically different from zero. The method of moment estimator has the following distribution:

\[ \frac{\hat{\gamma}(h)}{\sigma(\hat{\gamma}(h))} \xrightarrow{d} N(0, 1) \]  \hspace{1cm} (11)

Rationale: Absolute returns are meant to represent volatility and volatility is a persistent process.
Stylized effect #3: the leverage effect

The key elements:

1. Volatility is higher during periods of negative returns.

2. Negative returns contribute more to a rise in volatility than positive ones.

Rationale:

- Volatility measures risk. Negative returns lead to a riskier world for investors. However, volatility does not measure only downward risks.

- For stocks, periods of economic expansion leads to an improving access to credit. The temptation to get leveraged (borrowing money to obtain several times the return on equity) is high, increasing in the mean time the economic consequences of a bankruptcy. Hence, the net loss perspectives are higher than the net gains.

- For commodities for which the supply is limited, the effect is converse: periods of high vol coincides with periods of positive returns, indicating an expected shortage. However, as in the stock case, long positions are more common than short ones.
Index returns vs. VIX variations

- SPX
- Small caps
## Leverage effect

Correlation between returns and variations of VIX and VDAX

<table>
<thead>
<tr>
<th></th>
<th>VIX</th>
<th>VDAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP500</td>
<td>-76%</td>
<td>-39%</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>-64%</td>
<td>-31%</td>
</tr>
<tr>
<td>DAX</td>
<td>-42%</td>
<td>-67%</td>
</tr>
<tr>
<td>CAC</td>
<td>-41%</td>
<td>-66%</td>
</tr>
<tr>
<td>SC.US</td>
<td>-68%</td>
<td>-38%</td>
</tr>
<tr>
<td>SC.EU</td>
<td>-42%</td>
<td>-68%</td>
</tr>
<tr>
<td>HY.EU</td>
<td>-1%</td>
<td>-2%</td>
</tr>
<tr>
<td>CORP.EU</td>
<td>6%</td>
<td>12%</td>
</tr>
<tr>
<td>CHINA</td>
<td>-11%</td>
<td>-20%</td>
</tr>
<tr>
<td>EM</td>
<td>-99%</td>
<td>-47%</td>
</tr>
<tr>
<td>EU.GOVT</td>
<td>8%</td>
<td>14%</td>
</tr>
<tr>
<td>COMMODITY</td>
<td>-14%</td>
<td>-15%</td>
</tr>
<tr>
<td>OIL</td>
<td>-9%</td>
<td>-8%</td>
</tr>
<tr>
<td>GOLD</td>
<td>-2%</td>
<td>-3%</td>
</tr>
<tr>
<td>WORLD.GOVT.INDEX</td>
<td>18%</td>
<td>25%</td>
</tr>
</tbody>
</table>
Problems with the method of moment estimator

- How do I choose n to compute the volatility?
  - Usually 20-days rolling volatility is retained, e.g. for the Bollinger bands. Why is this?
  - It gives to volatility a persistency based on no scientific rule (i.e. you cannot check the robustness)

- The contribution of returns to risk may be non-linear.

- Most of the time, we are interested in the conditional volatility.

- It gives no clues regarding the future evolution of risk.

- A degree of uncertainty is associated to the estimates of volatility as

\[
\sqrt{\frac{1}{T} \sum_{t=1}^{T} (r_t)^2} = \frac{1}{T} \left( \frac{\mathbb{E}[r_t^4]}{\text{Kurtosis related}} - \frac{\mathbb{E}^2[r_t^2]}{\text{Squared volatility related}} \right)
\]  

(12)
I. Modelling historical volatility

I-2. How to estimate volatility
Why volatility needs to be estimated

• For most financial asset pricing model, volatility is a key element

• However, historical volatility cannot be observed in the real world of finance

• Its very measure raises questions that are difficult to answer

⇒ An estimation framework is required here.
An introduction to estimation

Two different approaches:

1. Calibration:
   - Use market prices over a small time bucket
   - Try to minimize a given error metric
   - Discriminate between
     - More parameters and a nearly perfect fit
     - and less parameter with more estimation stability

2. Estimation:
   - Use historical data
   - Look at the statistical properties of the datasets
   - Define an estimation strategy for the model
   - Compare the stability of the estimation / forecasting power (when possible) / the statistical fit of the stylized facts.

⇒ Both approaches cannot be compared since they aim at different goals
An introduction to estimation

The estimation story:

\[ X = (x_1 \ x_2 \ x_3 \ \ldots \ x_n)^\top \] is a dataset of \( n \) observations such as:

- returns of stocks
- interest rates daily variations
- any currency
- oil prices

You assume that \( X \) is driven by model involving:

- a certain type of randomness
- and driven by a vector of parameter \( \theta \)

Estimating the model is basically:

- deciding upon a reliable criterion \( h(\theta|X) \)
- whose extremum guaranties that the chosen parameters make the model fit the really in the best way possible
An introduction to estimation

Usually: find $\theta^*$ that solves

$$\min \theta \text{ or } \max \theta \ h(\theta|X)$$

Main difference with calibration here?

- Since the (maybe asymptotic) distribution of $X$ is known
- The distribution of the empirical counterpart to $h$ is known for most of the cases
- The distribution of $\hat{\theta}$ is known $\Rightarrow$ test for $\hat{\theta} = 0$ elementwise

The statistical discussion goes beyond what calibration can tell you about your model, while delivering a poorer fit.
An introduction to estimation

Example: The Capital Asset Pricing Model
W. Sharpe, 1964. Equity returns are assumed to follow:

\[ r^i_t = r^f_t + \beta(r^m_t - r^f_t), \quad t = 1, \ldots, n \]

One estimation strategy:
- Introduce randomness: \( r^i_t = r^f_t + \beta(r^m_t - r^f_t) + \epsilon_t, \quad t = 1, \ldots, n, \)
  \( \epsilon_t \sim N(0, \sigma) \)
- Parameters driving the model: \( \theta = (\beta, \sigma) \)
- Find a criterion:
  \[ h(\theta|X) = \frac{1}{n} \sum_{t=1}^{n} (r^i_t - r^f_t - \beta(r^m_t - r^f_t))^2 \quad (13) \]
  and solve to get the estimates \( \Rightarrow \) Ordinary Least Squares estimator (OLS)

Finally, test for statistical signficicativity is straightforward.
An introduction to estimation

Different estimation strategies exist. However, the most widespread approach is the **(quasi) maximum likelihood**.

**How does it work?**

- Starting point: the joint distribution of the sample: \( f(x_1, x_2, \ldots, x_n | \theta) \).
- We are interested in the probability of occurrence of our sample, given the parameters involved

\[
f(x_1, x_2, \ldots, x_n | \theta) = f(\theta | x_1, x_2, \ldots, x_n)
\]

Key idea here: look for the vector of parameters \( \theta^* \) that makes the probability of occurrence of our sample maximum. Use density instead of probability for a continuous random variable:

\[
\theta_{ML} : \max_{\theta} \log f(\theta | x_1, x_2, \ldots, x_n)
\]
A basic example

The Black and Scholes (1973) model can be discretized as:

\[ r_t = \mu + \sigma \epsilon_t. \]  

(14)

Given that \( \epsilon_t \) is distributed as a \( N(0, 1) \), the likelihood of the sample can be written as:

\[ f(r_1, r_2, \ldots, r_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{1}{2} \frac{(r_t - \mu)^2}{\sigma^2} \right) \]  

(15)

The loglikelihood of the process is then

\[ \log L = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2} \sum_{t=1}^{n} \frac{(r_t - \mu)^2}{\sigma^2} \]  

(16)

Optimizing for \( \sigma \), the ML estimate is

\[ \hat{\sigma}_{\text{ML}} = \frac{1}{n} \sum_{t=1}^{n} (r_t - \bar{r}_t)^2 \]  

(17)
The ARCH approach to measuring volatility

ARCH is AutoRegressive Conditionally Heteroscedastic model, after Engle (1982)’s seminal work.

- Initially created to measure the volatility of inflation in the UK.
- The primary focus was monetary economics. Financial applications came after that.
- There exist a linear relation between variance and squared returns, when looking at the "naive" method of moment estimator:

\[
\frac{\partial}{\partial (r_t)^2} \frac{1}{n} \sum_{t=1}^{n} r_t^2 = \frac{1}{n}
\]  

(18)

Engle focused on finding scaling parameters such that today’s squared returns could be turned into today’s conditional variance:

\[
r_t = \mu + \sigma_t \epsilon_t, \epsilon_t \sim N(0, 1)
\]  

(19)

\[
\sigma_t^2 = \omega + \alpha (r_{t-1} - \mu)^2
\]  

(20)
Problems with the ARCH approach

1. In the ARCH model with 1 lag, a rise in the absolute value of $\epsilon_t$ will have an effect only on volatility at time $t + 1$. With a daily timing, volatility rises simultaneously with large returns.

2. The model becomes more realistic when other lags than $\epsilon_{t-1}$ is included in the variance dynamics:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2$$  \hspace{1cm} (21)

By doing so, the persistence of the process is increases as past shocks are still used to compute time $t$ conditional variance.

The variance process is said to be a Moving Average process of order $p$ and the model is called ARCH(p).
The GARCH model

Including more lags in the variance dynamics is a statistical artifact that can be obtained by considering an autoregressive process of order 1 for variance.

\[ \sigma_t^2 = \omega + \alpha (r_{t-1} - \mu)^2 + \beta \sigma_{t-1}^2 \]  \hspace{1cm} (22)

This process is said to be a process mixing an AutoRegressive process of order 1 and a moving average process of order 1.

However, it can be written as an infinite moving process, with loadings exponentially decreasing for past lags:

\[ \sigma_t^2 = \omega^* + \sum_{i=0}^{\infty} (\alpha^*)^i \epsilon_{t-i}^2 \]  \hspace{1cm} (23)

Hence, Bollerslev and Engle’s GARCH model:

\[ r_t = \mu + \sigma_t \epsilon_t \]  \hspace{1cm} (24)

\[ \sigma_t = \omega + \alpha (r_{t-1} - \mu)^2 + \beta \sigma_{t-1}^2 \]  \hspace{1cm} (25)
The GARCH approach

The key features of the model:
- use volatility clustering
- assume variance is a linear combination of past variance and squared returns

Purpose: "nowcasting" and forecasting of conditional variance/volatility.
With such a model:

$$\mathbb{E}[r_t | r_{t-1}, r_{t2}, ...] = \mathbb{E}[r_t] = \mu$$
$$\mathbb{V}[r_t | r_{t-1}, r_{t2}, ...] = \sigma_t^2$$

Conditional volatility is a function of past returns: no disturbance on its own.
As long as $\beta + \alpha < 1$ the process is stationary. When $=1$ the volatility process is a random walk.
Maximum likelihood applied to GARCH models

Specific issues:

- The sample is not i.i.d.. A conditional approach is required here.
- The loglikelihood writes

\[ f(x_1, x_2, ..., x_n|\theta) = f(x_1|\theta)f(x_2|x_1, \theta)f(x_3|x_1, x_2, \theta)...f(x_n|x_{n-1}, x_{n-2}, ..., x_1, \theta) \]

- \( \epsilon_t \sim N(0, 1) \)
- Stationary whenever \( \omega_1 + \omega_2 < 1 \) and positive parameters

Equity parameters rough estimates:

- \( \alpha \approx 0.2 \): the contribution of new shocks to volatility
- \( \beta \approx 0.8 \): this parameter is related to the persistence of the volatility process, that is the contribution of old shocks to volatility
- \( \omega \) guides the unconditional volatility level, as

\[
\mathbb{E} \left[ \sigma_t^2 \right] = \frac{\omega}{1 - \alpha - \beta}
\] (26)
The riskmetrics approach

The sum of $\alpha$ and $\beta$ is often found to be close to 1: the estimated volatility process is usually found to be strongly persistent. Riskmetrics propose a risk management system based on conditionally Gaussian models that look like GARCH models, with fixed parameters balancing the contribution of news in the volatility computation.

$$\sigma_t^2 = \beta \sigma_{t-1}^2 + (1 - \beta) r_{t-1}^2$$  \hspace{1cm} (27)

Volatility measuring is a difficult task, as:

- when entering the crisis, if the process is too strongly persistent, the VaR calculation will be by far too small;
- when exiting a crisis, again, the VaR will stand a good chance to be too large.

Finding the balance between the two requires to discriminate between crisis and market expansion periods.
The Markov Switching approach

Globally, risky markets rise and fall together, being governed by two different states market participants refer to as bull and bear market:

\[ r_t = \mu_i + \sigma_i \epsilon_t, \text{ with } s_t = i \]  \hspace{1cm} (28)

\[ s_t \text{ is the state variable. The transition from one state to another is given by the transition probabilities matrix} \]

\[ P(s_t = i | s_{t-1} = j) = p_{i,j} \]  \hspace{1cm} (29)

Estimating this model is more complex as both the state and \( \epsilon_t \) are random variables, thus necessitating the use of an optimal filtering method. In the case where \( \epsilon_t \) is Gaussian, the estimation can be performed using the Kalman filter.

When working with equity markets, typical results involve two states \( i = 1, 2 \) such that

\[ \mu_1 > \mu_2 \] \hspace{1cm} (30)

\[ \sigma_1 < \sigma_2 \] \hspace{1cm} (31)

\[ p_{ii} \sim 1 \] \hspace{1cm} (32)

The persistence of volatility is captured by the persistence of the state.
Simulated sample

Scatterplot

Density

mu1=0.05 mu2=-0.1 sigma1=0.05 sigma2=0.05 p11=0.99 p22=0.98
The leverage effect

The impact of negative returns on variance is usually higher than the contribution of positive ones in equity markets. The financial literature call this effect the "leverage effect". 

Rationale behind:

- Financial crises are characterized by negative returns and a sharp rise in risk. The least you can expect from a risk measure is to grow faster during crises periods than during market pauses.

- During expansion periods, a leveraging of the balance sheet happened:

\[
\text{Results} = r \times K + (r - i)D \Rightarrow \frac{\text{Results}}{K} = r \times + (r - i) \frac{D}{K} \quad (33)
\]

This financial leverage leads to overpriced equities and to sharp drops in asset prices, given that the interest rates is to be added to the already negative returns.

- In the oil case, these effects are to be related to political issues leading to larger rises in prices when a short-squeeze is expected.
The leverage effect

Different models account for leverage effects:

1. GARCH GJR (Glosten, Jagannathan and Runkle, 1993):

   \[ r_t = \mu + \sigma_t \epsilon_t \]  
   \[ \sigma^2_t = \omega_0 + \omega_1 (r_{t-1} - \mu)^2 + \omega_2 \sigma^2_{t-1} + \gamma \max(-(r_{t-1} - \mu), 0)^2 \]  


   \[ r_t = \mu + \sigma_t \epsilon_t \]  
   \[ \log \sigma^2_t = \omega_0 + \omega_1 \epsilon_{t-1} + \gamma \left( |\epsilon_{t-1}| - \sqrt{\frac{2}{\pi}} \right) + \beta \log \sigma^2_{t-1} \]

Another possibility: leverage effect is related to conditional skewness:

\[ \text{Cov}_t(r_{t-1} - r_t, \sigma^2_{t+1} - \sigma^2_t) = \omega_1 \mathbb{E}[r_t^3] \]

We label the neg. returns backfire on volatility **conditional skewness** and the unconditional distribution’s skewness the **unconditional skewness**.
The non linearity of volatility

Most of the GARCH models are based on the intuition that the squared volatility is linearly related to squared returns. This is however unsure. Granger and Ding (1996) proposed a model that dwells on a possibly non linear relation such as:

\[ r_t = \mu + \sigma_t \epsilon_t \]  
\[ \sigma_t^\delta = \omega + \alpha \epsilon_t^{\delta - 1} + \beta \sigma_t^{\delta - 1} \]

The model is called ”Asymmetric Power ARCH model”.

When testing whether \( \delta \) is statistically different from 0 we get the answer whether such a linear relation exists or not.

For most markets this relation does not hold.
Conditional and unconditional skewness

\[ r_t = \mu + \sigma_t \epsilon_t \]

Neg. returns to volatility transmission \quad Conditional distribution of the returns

Until now: **implicit assumption that returns are conditionally Gaussian.**

The empirical finance literature shows that other distributions better suits the statistical properties of returns:

- The Student and asymmetric distribution distributions
- The Hyperbolic distribution
- The Normal Inverse Gaussian distribution
- The Generalized Hyperbolic distribution
- The mixture of Gaussian distribution
The Generalized Hyperbolic Distribution

- For \((\lambda, \alpha, \beta, \delta, \mu) \in \mathbb{R}^5\) with \(\delta > 0\) and \(\alpha > |\beta| > 0\), the density is

\[
d_{GH}(x, \lambda, \alpha, \beta, \delta, \mu) = \frac{(\sqrt{\alpha^2 - \beta^2}/\delta)^\lambda}{\sqrt{2\pi K_\lambda(\delta \sqrt{\alpha^2 - \beta^2})}} e^{\beta(x-\mu)} \frac{K_{\lambda-1/2} \left(\alpha \sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{1/2-\lambda}}
\]

where \(K_\lambda\) is the modified Bessel function of the third kind.

- Encompass fat tails (but not too fat) and various degrees of skewness

- Offers a very nice fit for return time series (CAC, DAX, SP500 and FTSE), especially for the lower tail
The Generalized Hyperbolic Distribution

With properly chosen parameters, this distribution reduces to the following distributions:

1. \( \lambda = 1 \): hyperbolic distribution
2. \( \lambda = -1/2 \): NIG distribution
3. \( \lambda = 1 \) and \( \xi \to 0 \): Normal distribution
4. \( \lambda = 1 \) and \( \xi \to 1 \): Symmetric and asymmetric Laplace distribution
5. \( \lambda = 1 \) and \( \chi \to \pm \xi \): Inverse Gaussian distribution
6. \( \lambda = 1 \) and \( |\chi| \to 1 \): Exponential distribution
7. \( -\infty < \lambda < -2 \): Asymmetric Student
8. \( -\infty < \lambda < -2 \) and \( \beta = 0 \): Symmetric Student
9. \( \gamma = 0 \) and \( 0 < \lambda < \infty \): Asymmetric Normal Gamma distribution

with

\[
\zeta = \delta \sqrt{\alpha^2 - \beta^2}, \quad \rho = \beta/\alpha \tag{42}
\]
\[
\xi = (1 + \zeta)^{-1/2}, \quad \chi = \xi \rho \tag{43}
\]
\[
\bar{\alpha} = \alpha \delta, \quad \bar{\beta} = \beta \delta. \tag{44}
\]
The Generalized Hyperbolic Distribution
Disentangling cond. from unconditional skewness

A new issue:
When \( \epsilon_t \) is a mixture of Gaussian distribution and \( \sigma_t \) is chosen to be EGARCH:

\[
    r_t = \mu + \epsilon_t(\theta_{\epsilon})\sigma_t(\theta_{\sigma}),
\]

\[
    \mathbb{E}[\epsilon_t] = 0, \mathbb{V}[\epsilon_t] = 1
\]

(45) (46)

skewness now comes from both ends of \( \epsilon_t\sigma_t \). Statistically speaking, we stand a good chance that the model is no longer unique: it is very likely that different sets of parameters \( \theta_{\epsilon} \) and \( \theta_{\sigma} \) deliver the same distribution for \( r_t \).

This is especially true when it comes to the asymmetry of the distribution. Part of the skewness of \( r_t \) comes from \( \epsilon_t \) (unconditional skewness) and part of it comes from \( \sigma_t \)... but which is which?
It’s all in the estimation strategy

Three kinds of estimation strategies:

1. Maximum likelihood: estimate both $\theta_\epsilon$ and $\theta_\sigma$ in the mean time by maximizing:

$$\max_{\theta_\epsilon, \theta_\sigma} \log f(r_1, r_2, ..., r_T; \theta_\epsilon, \theta_\sigma)$$  \hspace{1cm} (47)

2. Quasi maximum likelihood: estimate first $\theta_\sigma$, assuming that $\epsilon_t$ is Gaussian. Then estimate $\theta_\epsilon$ keeping the previous estimates for $\theta_\sigma$ fixed:

$$\begin{align*}
(1) & \max_{\theta_\epsilon} \log f(r_1, r_2, ..., r_T; \theta_\sigma) \\
(2) & \max_{\theta_\epsilon} \log f(r_1, r_2, ..., r_T; \theta_\epsilon, \theta_\sigma^*)
\end{align*}$$  \hspace{1cm} (48) \hspace{1cm} (49)

3. Recursive likelihood: starting from the QML estimates, run $n$ times the several estimation steps:

(a) Re-estimate $\theta_\sigma$ holding $\theta_\epsilon$ fixed: $\max_{\theta_\sigma} \log f(r_1, r_2, ..., r_T; \theta_\epsilon^*, \theta_\sigma)$.

(b) Re-estimate $\theta_\epsilon$ holding $\theta_\sigma$ fixed: $\max_{\theta_\epsilon} \log f(r_1, r_2, ..., r_T; \theta_\epsilon, \theta_\sigma^*)$
Using actual data

We propose to estimate the various models proposed earlier using a data set of 1506 returns on the SP 500, the US equity index.

The data set starts on January, 2\textsuperscript{nd} 1998 and ends on December, 31\textsuperscript{st} 2003. It includes different market phases, both bull and bear, as it includes the 1998 market rally and the explosion of the technology bubble in 2001.

We estimate the parameters of the models combining the EGARCH and APARCH volatility structures and the MN and GH distributions and the three estimation strategies: ML, QML and REC. The starting values to perform these estimates are the same for each method and are set to be equal to the values used in the Monte Carlo experience.
Test methodology

The focus here is on the **goodness of fit** of the distribution of $r_t$.

We propose to use the joint density of the sample as a score to differentiate the estimation strategies, following the Vuong (1980)'s density test.

Say we deal with a time series model for the log-returns whose conditional density at time $t$ is $f_1(Y_t|Y_{t-1}, \theta_1)$, where $Y_{t-1} = (Y_1, \ldots, Y_{t-1})$ and $\theta_1$ is the vector of parameters describing the shape of this conditional distribution and the volatility structure. We compare this model to another one defined by the conditional density $f_2(Y_t|Y_{t-1}, \theta_2)$, with $\theta_2$ being the parameters associated to this second model.

The null hypothesis of the test is "models 1 and 2 provide a similar fit of the log-return's conditional distribution". The corresponding test statistic used is:

$$t_{1,2} = \frac{1}{n} \sum_{t=1}^{n} (\log f_1(Y_t|Y_{t-1}, \theta_1) - \log f_2(Y_t|Y_{t-1}, \theta_2)),$$

where $n$ is the total number of observations available. Under the null hypothesis

$$\frac{t_{1,2}}{\hat{\sigma}_n} \sqrt{n} \xrightarrow{n \to +\infty} \mathcal{N}(0, 1).$$
Using actual data

<table>
<thead>
<tr>
<th>Models</th>
<th>ML vs. QML</th>
<th>REC vs. QML</th>
<th>REC vs. ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGARCH - GH</td>
<td>40.27</td>
<td>83.13</td>
<td>29.74</td>
</tr>
<tr>
<td>APARCH - GH</td>
<td>19</td>
<td>75.84</td>
<td>15.55</td>
</tr>
<tr>
<td>EGARCH - MN</td>
<td>43.32</td>
<td>86.67</td>
<td>34.22</td>
</tr>
<tr>
<td>APARCH - MN</td>
<td>19.02</td>
<td>91.1</td>
<td>14.92</td>
</tr>
</tbody>
</table>

The data set starts on January, 2\textsuperscript{nd} 1998 and ends on December, 31\textsuperscript{st} 2003. It includes 1506 data points. The returns computed are logarithmic returns. ML stands for maximum likelihood, QML stands for Quasi Maximum Likelihood and REC stands for Recursive Likelihood. When comparing the accuracy of model 1 (with parameters $\theta_1$) vs. model 2 (with parameters $\theta_2$) to fit the joint distribution of a given sample, the test statistic is computed as follows:

$$t_{1,2} = \frac{1}{n} \sum_{t=1}^{n} \left( \log f_1(Y_t|Y_{t-1}, \theta_1) - \log f_2(Y_t|Y_{t-1}, \theta_2) \right),$$

with $f(\cdot)$ the selected conditional density. The test reads as follows: in the EGARCH-GH case, the column "ML vs. QML" uses the estimated parameters by ML as model 1 and the parameters estimated by QML as model 2. The test statistics value is 40.27: this value being outside the $[-1.96 : 1.96]$ 5\% interval confidence, the null hypothesis that both models are equivalent is strongly rejected. The positivity of this statistics indicates that model 1 is favored over model 2.
Using actual data

<table>
<thead>
<tr>
<th></th>
<th>QML</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EGARCH GH</td>
<td>APARCH GH</td>
<td>EGARCH MN</td>
</tr>
<tr>
<td>EGARCH GH</td>
<td>–</td>
<td>0.57</td>
<td>-1.07</td>
</tr>
<tr>
<td>APARCH GH</td>
<td>–</td>
<td>–</td>
<td>-1.10</td>
</tr>
<tr>
<td>EGARCH MN</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>APARCH MN</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ML</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EGARCH GH</td>
<td>APARCH GH</td>
<td>EGARCH MN</td>
</tr>
<tr>
<td>EGARCH GH</td>
<td>–</td>
<td>21.20</td>
<td>2.42</td>
</tr>
<tr>
<td>APARCH GH</td>
<td>–</td>
<td>–</td>
<td>-15.86</td>
</tr>
<tr>
<td>EGARCH MN</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>APARCH MN</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>REC</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EGARCH GH</td>
<td>APARCH GH</td>
<td>EGARCH MN</td>
</tr>
<tr>
<td>EGARCH GH</td>
<td>–</td>
<td>-0.09</td>
<td>-3.88</td>
</tr>
<tr>
<td>APARCH GH</td>
<td>–</td>
<td>–</td>
<td>-0.81</td>
</tr>
<tr>
<td>EGARCH MN</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>APARCH MN</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Results

What we learned from the estimation and the horse race:

1. The recursive estimation method outperforms the remaining estimation strategies

2. The estimation obtained with the recursive approach appears to be all the more correct as they are stable across the 4 pairs of models

3. When using such estimates, most of the models are equivalent, for an equivalent number of parameters
I. Modeling historical volatility

I-3. Intra-day based volatility measures and the distribution of assets’ returns
Basic representation for a stochastic process

Any process $S_t$ can be represented the following way:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t + J_t dN_t \quad (53)$$

Deterministic part \hspace{3cm} Diffusice part \hspace{3cm} Jump part

Most of the time series models’ approach are actually mixing the diffusive and the jump part of the processes:

$$r_t = \mu + \sigma t \epsilon_t, \quad (54)$$
$$\epsilon_t \sim IID \text{(Possibly skewed and leptokurtic)} \quad (55)$$

How helpful is it to disentangle jumps from pure volatility process?
Realized variance


Quadratic variation

\[
[p]_{t}^{[2]} = p \lim_{\Delta \to 0} \frac{\text{Int}(t/\Delta)}{\Delta} \sum_{j=1}^{\text{Int}(t/\Delta)} \left( p_{j\Delta} - p_{(j-1)\Delta} \right)^{2}
\]  

(56)

Captures the total quadratic variation = QV of the continuous part of the process + the squared jumps (of finite size).

Estimates:

Cut the trading day into \( M \) parts of size \( \delta \), then.

\[
[\hat{p}]_{t}^{[2]} = \sum_{j=2}^{M} \left( p_{j\Delta} - p_{(j-1)\Delta} \right)^{2}
\]

(57)

yields the day \( t \) realized variance. Yet incorporate the squared jumps.
Bipower variation

Joint probability of a jump over two consecutive trading intervals is smaller. A bipower variation estimate is

$$[\hat{p}_b]_t^{[2]} = \sum_{j=2}^{M-1} \left| p_j \Delta - p_{(j-1)} \Delta \right| \times \left| p_{(j+1)} \Delta - p_{(j)} \Delta \right|$$  (58)

The difference between both measures often used as a proxy for the jump activity in a trading day. On bipower variation, main reference is:


- For an extensive use of it for empirical finance, see Jumps, CoJumps and Macro Announcements (with Jérome Lahaye and Chris Neely).
Data snapshot: high frequency data from 05.08.2008 to 08.08.2008.
Realized estimates from 05.08.2008 to 08.08.2008.
Bipower estimates from 05.08.2008 to 08.08.2008.
Jump activity from 05.08.2008 to 08.08.2008.
The intuition would say to use data at the lowest frequency possible, so as to be as close as possible to the limit of the estimator.

However, market quotes are affected by specific phenomenon labelled as ”microstructure noise”, related to the organization of the market, such as:

- infrequent trading
- seasonality
- price manipulation
- ...

The higher the frequency used to compute both formula and the higher the likelihood that the estimator is affected by a microstructure noise.
Testing for the optimal sampling frequency

The first attempt is the **signature plot**: plot the average volatility w.r.t. the sampling frequency. See Hansen and Lund (2004).

The estimator is said to be free from microstructure noise when volatility converges to a certain level.

Various tests designed to compute the optimal frequency to compute realized and bipower variations measure of volatility, as Awartani et al. (2009)’s test

\[
ZT_{BPV} = \frac{BPV(M) - BPV(N)}{\sqrt{QV(N)}} \times \sqrt{N \times T}
\]  

(59)

with \(QV\) being the realized quarticity. The literature gets quite technical in this respect.
Figure 1: Volatility signature plots for RF, based on ask quotes (circles); bid quotes (crosses); mid quotes (triangles), and transaction prices (dots). The left column is for AA and the right column is for MSFT. The two top rows are based on calendar time sampling, in contrast to the bottom rows that are based on tick time sampling. The results for 2000 are the panels in rows 1 and 3 and those for 2004 are in rows 2 and 4. The horizontal line represents an estimate of the average IV, $\hat{\sigma}^2 = \frac{RF_{\text{Naive}}}{\text{Daily}}$, that is defined in Section 4.2. The shaded area about $\hat{\sigma}^2$ represents an approximate 95% confidence interval for the average volatility.
Figure 1: Volatility signature plots for RF, based on ask quotes (circles); bid quotes (crosses); mid quotes (triangles), and transaction prices (dots). The left column is for AA and the right column is for MSFT. The two top rows are based on calendar time sampling, in contrast to the bottom rows that are based on tick time sampling. The results for 2000 are the panels in rows 1 and 3 and those for 2004 are in rows 2 and 4. The horizontal line represents an estimate of the average IV, $\sigma^2 \equiv \frac{RF_{\text{ACNW}10}}{\text{NWB}}$, that is defined in Section 4.2. The shaded area about $\sigma^2$ represents an approximate 95% confidence interval for the average volatility.
What can we do with that?

1. Increasing our understanding about the jump activity in financial returns. Is it weak? Is it strong? Does prices move only by jumps?
3. Risk management: forecasting the density of returns with or without this jump component? See Maheu and McCurdy (2009).
5. ...
Estimation of the Heston volatility dynamics

In the Heston framework: Cox-Ingersoll-Ross process for the volatility:

\[ dV_t = \kappa(\theta - V_t)dt + \sigma \sqrt{V_t}dW_t, \]  \hspace{1cm} (60)

Conditional distribution of the volatility is a non-centered gamma distribution, up to a change of scale.

\[ \rho = \exp(-\kappa), \quad \delta = \frac{2\kappa\theta}{\sigma^2}, \quad c_{\Delta t} = \frac{\sigma^2}{2\theta} (1 - \exp(-\theta \Delta t)) \]  \hspace{1cm} (61)

Then:

\[ \mathbb{E}[V_{t+\Delta t} | V_t] = c_{\Delta t} \delta + \rho^{\Delta t} V_t \]  \hspace{1cm} (62)
\[ \mathbb{V}[V_{t+\Delta t} | V_t] = c_{\Delta t}^2 \delta + 2\rho^{\Delta t} c_{\Delta t} V_t \]  \hspace{1cm} (63)

Thus, it is possible to estimate the CIR dynamic for the volatility from the following discrete time model:

\[ V_{t+\Delta t} = c_{\Delta t} \delta + \rho^{\Delta t} V_t + u_{t,\Delta t} \]  \hspace{1cm} (64)
\[ \mathbb{E}[u_{t,\Delta t} | V_t] = 0, \quad \mathbb{E}[u_{t,\Delta t}^2 | V_t] = c_{\Delta t}^2 \delta + 2\rho^{\Delta t} c_{\Delta t} V_t \]  \hspace{1cm} (65)

Can be estimated in a consistent way by Generalized Least Squares (GLS). Use jointly the method of moments and GLS to estimate the whole parameters.
Estimates

Using the method of moment estimates, we obtain:

<table>
<thead>
<tr>
<th></th>
<th>$\kappa$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realized</td>
<td>121</td>
<td>16%</td>
<td>0.1%</td>
<td>-48%</td>
</tr>
<tr>
<td>Bipower</td>
<td>117</td>
<td>10%</td>
<td>0.07%</td>
<td>-24%</td>
</tr>
</tbody>
</table>

Clearly:

- depending on the vol measure you are using results are different
- the elimination of jumps reduces the vol of vol parameter
- also an impact on the persistence parameter
<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma$</th>
<th>$\rho$</th>
<th>$\lambda$</th>
<th>$\mu_j$</th>
<th>$\sigma_j$</th>
<th>$\mu_v$</th>
<th>$\rho_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.056</td>
<td>0.783</td>
<td>0.270</td>
<td>0.920</td>
<td>-0.273</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.017)</td>
<td>(0.023)</td>
<td>(0.101)</td>
<td>(0.201)</td>
<td>(0.114)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Jiang and Knight (sample: 1990-1999)

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma$</th>
<th>$\rho$</th>
<th>$\lambda$</th>
<th>$\mu_j$</th>
<th>$\sigma_j$</th>
<th>$\mu_v$</th>
<th>$\rho_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1406</td>
<td>0.0255</td>
<td>10.6997</td>
<td>4.7140</td>
<td>-0.4105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0573)</td>
<td>(0.0039)</td>
<td>(4.5364)</td>
<td>(0.5660)</td>
<td>(0.229)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chacko and Viceira (sample: 1990-1999)

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma$</th>
<th>$\rho$</th>
<th>$\lambda$</th>
<th>$\mu_j$</th>
<th>$\sigma_j$</th>
<th>$\mu_v$</th>
<th>$\rho_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1121</td>
<td>0.0230</td>
<td>14.2819</td>
<td>5.1505</td>
<td>-0.6299</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0279)</td>
<td>(0.0021)</td>
<td>(2.2702)</td>
<td>(0.4722)</td>
<td>(0.2997)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chacko and Viceira (sample: 1980-2000)

II. Modelling risk neutral volatility

II-1. Implied volatility’s stylized facts
Measuring implied volatility

Implied volatility is an monotonously increasing transform of option prices.

In the Black Scholes model (1973), the call option prices are functions of a few parameters, among of which volatility

\[
C(K, \sigma, \tau, S_t, r_f) = SN(d_1) - Ke^{-r\tau}N(d_2)
\]  

(66)

Given option market prices, it is possible to invert numerically the previous formula to obtain the volatility implied by both the market price of the option and of the other elements involved into the formula:

\[
\sigma_{BS} = C(K, \sigma, \tau, S_t, r_f)^{-1}
\]  

(67)

Obtained usually by minimizing a quadratic criterion such that

\[
\min_{\sigma}(C(K, \sigma, \tau, S_t, r_f) - C_M)^2
\]  

(68)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Last</th>
<th>Change</th>
<th>Volume</th>
<th>Strike</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAX1</td>
<td>4861.38</td>
<td>-29.34</td>
<td>156000</td>
<td>4000</td>
</tr>
<tr>
<td>DAX2</td>
<td>1064.22</td>
<td>1.40</td>
<td>12350</td>
<td>4000</td>
</tr>
<tr>
<td>DAX3</td>
<td>123.65</td>
<td>0.20</td>
<td>2145</td>
<td>4000</td>
</tr>
<tr>
<td>DAX4</td>
<td>234.56</td>
<td>0.30</td>
<td>3456</td>
<td>4000</td>
</tr>
<tr>
<td>DAX5</td>
<td>567.89</td>
<td>1.10</td>
<td>7890</td>
<td>4000</td>
</tr>
<tr>
<td>DAX6</td>
<td>890.12</td>
<td>2.20</td>
<td>1234</td>
<td>4000</td>
</tr>
<tr>
<td>DAX7</td>
<td>1234.56</td>
<td>3.30</td>
<td>2345</td>
<td>4000</td>
</tr>
<tr>
<td>DAX8</td>
<td>567.89</td>
<td>4.40</td>
<td>3456</td>
<td>4000</td>
</tr>
<tr>
<td>DAX9</td>
<td>890.12</td>
<td>5.50</td>
<td>7890</td>
<td>4000</td>
</tr>
</tbody>
</table>

**Strike Prices:**
- 34.40
- 52.50
- 50.00
- 55.50
- 42.30
- 37.40
- 32.20
- 41.00
- 25.00
Market Smiles on 08-28-08

Forw ard Money ness

Annualized vol.

0.08 0.33 0.83 1.33 1.82
Dividends in options

When considering dividends, the put-call parity formula is then given by:

\[ C^M(t, T, K) - P^M(t, T, K) = S_t e^{-d_t(T-t)} - K e^{-r_t(T-t)} \]  \hspace{1cm} (69)

where \( r_t \) is the risk free rate on time \( t \) corresponding to the maturity of the option price, and \( d_t \) is the expected dividend on date \( t \) for a maturity \( T \). The expression of the dividends, as a function of the market prices of calls and puts is then:

\[ d_t = \frac{1}{T-t} \log \left( \frac{S_t}{C^M(t, T, K) - P^M(t, T, K) + K e^{-r_t(T-t)}} \right) \]

The adapted Black-Scholes formula becomes:

\[ C^{BS}(t, T, K) = S e^{-d_t(T-t)} N(d_1) - K e^{-r_t(T-t)} N(d_2) \]

\[
\begin{align*}
  d_1 & = \frac{\log \left( \frac{S_t e^{-d_t(T-t)}}{K e^{-r_t(T-t)}} \right) + \frac{\sigma}{2} \sqrt{T-t}}{\sigma \sqrt{T-t}} \\
  d_2 & = d_1 - \sigma \sqrt{T-t}.
\end{align*}
\]
2006-2007 SPX Dividends recovered from the put-call parity relation

-1%
0%
1%
2%
3%
4%
5%
6%
7%
8%
Jan-06 
Feb-06 
Mar-06 
Apr-06 
May-06 
Jun-06 
Jul-06 
Aug-06 
Sep-06 
Oct-06 
Nov-06 
Dec-06 
Jan-07 
Feb-07 
Mar-07 
Apr-07 
May-07 
Jun-07 
Jul-07 
Aug-07 
Sep-07 
Oct-07
Listing implied volatility’s stylized facts

1. For a given time to maturity, implied volatilities are a convex function of the moneyness

2. This convex shape is sharper for short term maturities (<1M)

3. The term structure of at-the-money implied volatility is upward slopping most of the time

4. There are three statistical factors that explain the volatility surface moves.

5. The persistence of volatility depends on the option’s moneyness
The three factors of the volatility surface

Cont and Da Fonseca (2002) propose a decomposition of the variations of the volatility surface using a Karhunen-Loève decomposition. This decomposition resembles the Principal Component Analysis usually performed over bond yields.

Two main results:

1. The surface is driven by 3 factors: level, asymmetry and convexity.

2. All of these factors are persistent factors, close to an AR(1) process.
pages 45-60.

The persistence and the moneyness

AR(1) representation for the implied volatility for a given strike:

\[ \sigma_t(k) = \phi_0 + \phi_1 \sigma_{t-1}(k) + \epsilon_t(k) \]  

(71)

\( \phi_1 \) represent the persistence of the process. It is related to the "half-life" the following way:

\[ HL = \frac{\log \frac{1}{2}}{\log \phi_1} \]  

(72)

Provide of measure of the time needed for volatility shocks to mean-revert. See e.g. the survey in Michael Rockinger & Maria Semenova, 2005.

Estimated from DAX and FTSE options, \( \phi_1 \) is found to vary, depending on the moneyness.
Interpolation methodologies

The implied volatility surface is the function of the moneyness and the time to maturity. For calibration and time series analysis purposes, having the term structure of volatility each day is essential. Surfaces of this kind are obtained through interpolation and smoothing of the existing volatility values:

1. Classic interpolation between two points of the surface

2. Cubic polynomials

\[ \sigma_{imp}(m) = \alpha_0 + \alpha_1 m + \alpha_2 m^2 + \alpha_3 m^3 + \epsilon \]  

Estimated by OLS. Consistent with smile and skew.
3. Kernel smoothing:

\[
\sigma(m, \tau) = \frac{\sum_i \sum_j K(m - m_i) K(\tau - \tau_i) \sigma(m_i, \tau_j)}{\sum_i \sum_j K(m - m_i) K(\tau - \tau_i)} \tag{74}
\]

\(K(.)\) is a smoothing function names "kernel".

Usually the "Gaussian kernel" is chosen

\[
K(x) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{1}{2}\frac{x^2}{h^2}\right) \tag{75}
\]

\(h\) is a smoothing parameter. When \(h\) is too big, the smoothing is too strong and when it is too low, the estimation function is bumpy.
II. Modelling risk neutral volatility

II-2. The model free measure to risk neutral volatility
A model free measure of volatility

A new measure of the implied volatility in option prices:


- Jiang and Tian, 2005, The model free implied volatility and its information content. RFS.

**Key idea:** when dealing with implied volatility, which moneyness to be considered?

\[
\mathbb{E} \left[ \int_0^T \left( \frac{dF_t}{F_t} \right)^2 \right] = 2 \int_0^\infty \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2} dK
\]

with \( F \) denoting the computation of expectations with respect to the forward probability measure.

**Remarks:**

- Truncation error:

\[
2 \int_0^\infty \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2} dK \approx 2 \int_{K_{\min}}^{K_{\max}} \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2} dK
\]
Rule of the thumb: 2 to 3 SD around the moneyness. Very sensitive to the informational content of option prices.

• Trapezoidal integration:

\[
2 \int_{K_{\text{min}}}^{K_{\text{max}}} \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2} \, dK \approx \sum_{i=2}^{n} \left[ g(T, K_i) + g(T, K_{i-1}) \right] \Delta K
\]

\[
g(T, K) = \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2}
\]

• Using calls on the spot asset:

\[
\mathbb{E} \left[ \int_0^T \left( \frac{dF_t}{F_t} \right)^2 \right] = 2 \int_0^\infty \frac{C(T, K) \exp(r_f \times T) - \max(0, S_0 \exp(r_f \times T) - K)}{K^2} \, dK
\]

• Robust to jumps.

Illustration using a dataset of options on the carbon market from Jan. 2nd 2008 to Dec. 31st 2008. Truncation for moneyness 1.4 to 2.6. This calculation is used for the VIX.
Panel A: The B–S implied volatility (parameter set I)
Panel C: The B–S Implied volatility (parameter set II)
Volatility trading

As pointed out in Leippold, Egloff and Wu (2008), the risk neutral volatility is usually higher than the historical volatility: the risk premium on volatility is negative. Hence, the trading of variance swaps

\[
\text{Marked to market} = \text{Notional amount} \times (\sigma_{\text{realized}}^2 - \sigma_{\text{strike}}^2)
\]

Squared volatility because as variance is a linear function of state variable in most asset pricing models.

Again, this variance swaps can be used for hedging purposes or for speculation purposes.
II. Modelling risk neutral volatility

II-3. Modelling risk neutral volatility in continuous time
What continuous time can do for you

A large part of modern finance is based on the continuous time framework. Does it mean that discrete time can’t do anything for asset pricing purposes?

- Continuous time is able to accommodate rather complex type of distribution, thanks to the introduction of stochastic volatility and jumps

- It makes it possible to test different shapes for the risk premium, as they can be written as function of the state variables

- Continuous time models usually provide (almost) closed form expression for asset prices
Continuous time’s pitfalls

However, by solving certain issues in finance, continuous time models actually raise new issues.

- Most of the time, all the state variables are stochastic, making the derivation of a conditional likelihood impossible.
- For the same reason, filtrating the unobservable processes get really tricky, requiring the use of complicated filters.
- Conditional volatility can hardly be estimated, making the use of such models almost impossible when it comes to risk management.
The Heston model

**Heston (1993) process**: continuous time diffusion defined as:

\[
\begin{align*}
\frac{dS_t}{S_t} &= \mu dt + \sqrt{V_t} dW^S_t \\
\frac{dV_t}{V_t} &= \kappa (\theta - V_t) dt + \sigma \sqrt{V_t} dW^V_t \\
\langle dW^S_t, dW^V_t \rangle &= \rho dt
\end{align*}
\] (77-79)

Why such a model?

- Volatility is stochastic
- Displays fat tails
- \(\rho\) gives control over leverage effects/skewness of returns
- Mean reverting process like for GARCH
- Variance never reaches 0 whenever \(2\frac{\theta \kappa}{\sigma^2} > 1\) (Feller condition)
Reasons for its success

• Consistent with most of the stylized facts in option prices;
  – Smile
  – Skew
  – Smile term structure
• Easy to compute option prices because the characteristic function is exponential affine.
  \( \text{Heston} \in \text{affine process class of Duffie and Kan (1996)}: \)
  – The drift is an affine function of state variables
  – The variance is also an affine function of state variable
  Then \( \mathbb{E}[e^{i\omega \log S_T}|V_t, S_t] = e^{A(t,T)+B(t,T) \log S_t+C(t,T)V_t} \), with \( A, B \) and \( C \) deterministic functions of time and \( \omega \).
• Two possibilities to use this characteristic function to compute option prices in way that is faster than MC methods. See Carr and Madan (1999).
Estimation challenges

**Main problem:** In the Heston model, the volatility is driven by its own stochastic disturbance
Thus: vol is not observable and we cannot use the GARCH trick to filtrate it out of the returns’ dynamics.

**Possible solutions:**

- Estimation based on a high frequency approach: use the ultra high frequency data to make volatility observable.
Option pricing with affine processes

This process belongs to the class of the exponential affine models of Duffie and Kan (1996). With this class of models, computing option prices can be very accurately done using the fast Fourier transform (FFT) of Carr and Madan (1998). These methods use the closed form expression of the characteristic function of the Heston’s process. The European call option price is given by:

\[ C_T(k) = \int_k^\infty e^{-rT} (e^s - e^k) q(s) ds \]  

where \( K \) is the strike price and then \( k \) the moneyness, and \( q(.) \) the pricing probability function. Carr and Madan (1998) modify this price in the following way:

\[ c_T(k) = \exp\left(\alpha k\right) C_T(k), \quad \alpha \in \mathbb{R}^+_+. \]  

\( \alpha \) is the "dampening factor", a positive constant which ensures that the following sum exists. They show that the price could then be expressed as:

\[ C_T(k) = \frac{\exp\left(-\alpha k\right)}{\pi} \int_0^\infty e^{-iwk} \psi_T(w) dw, \]
with $\psi_T(\omega) = \int_{-\infty}^{\infty} e^{iwk} c_T(k) dk$. The latter expression turns out to be pretty simple, when expressed using the characteristic function of the log asset price:

$$\psi_T(\omega) = \frac{\phi(i\omega + \alpha + 1)e^{-rT}}{\alpha^2 + \alpha + i(\alpha + 1)w - w^2},$$

(83)

with $\phi(\omega) = \mathbb{E}[e^{\omega s_T}|s_t, V_t]$, the moment generating function associated to the log asset price. The inverse Fourier transform can be accurately computed using a Fast Fourier Transform.
Adding jumps to the volatility process

The price process is the following:

\[
ds_t = (\mu - \frac{1}{2} V_t)dt + \sqrt{V_t}dW_t
\]

\[
dV_t = \kappa(\theta - V_t)dt + \sigma \sqrt{V_t}dZ_t + J_t dN_t,
\]

with \( J_t \) exponentially distributed, with a parameter \( \nu \).

Again, this is an affine process and an affine characteristic function can be derived.

Adding jumps to volatility makes it possible to have thicker tails and to fit the short term smile in a better way.

It also makes it possible to replicate the dependence of the persistence of implied volatility over moneyness.
<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>SV+JD</th>
<th>SV+PJ</th>
<th>SV+PJ+JD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V(0)$</td>
<td>0.5265</td>
<td>0.3647</td>
<td>0.4982</td>
<td>0.3568</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.4854</td>
<td>0.1563</td>
<td>0.4871</td>
<td>0.1579</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>4.0183</td>
<td>2.8018</td>
<td>4.3739</td>
<td>2.6306</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>1.9751</td>
<td>0.9359</td>
<td>2.0642</td>
<td>0.9115</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.9453</td>
<td>-0.7372</td>
<td>-0.9900</td>
<td>-0.9629</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.1024</td>
<td></td>
<td></td>
<td>0.0977</td>
</tr>
<tr>
<td>$\beta$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td></td>
<td>0.0159</td>
<td>0.0039</td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td></td>
<td>0.6902</td>
<td>1.6523</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td></td>
<td>0.0069</td>
<td>0.0071</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4.1: Parameter estimates for 4 prototype dynamics.
The double Heston process

Short term and long term properties of the smile are different, both in terms of level and persistence.
Creating a two-factor model for volatility makes it possible to replicate such features. Hence the Christoffersen, Heston and Jacobs (2004)'s model:

\[
\begin{align*}
    dS_t &= \sqrt{v_t^{(1)}} dW_t^{(1)} + \sqrt{v_t^{(2)}} dW_t^{(2)} \\
    dv_t^{(i)} &= \kappa_i (\theta_i - v_t^{(i)}) dt + \sigma^{(i)} dZ_t^{(i)} \\
    d\langle Z^{(i)}, W^{(i)} \rangle_t &= \rho^{(i)} dt \\
    d\langle Z^{(i)}, z^{(j)} \rangle_t &= 0 dt
\end{align*}
\]

The 0 correlation between vol risk factors is mandatory to preserve the affine properties of the process.
Bridging the gap between discrete and continuous time models

Biggest problems the GARCH literature is facing:

- Handling changes in data frequencies is not obvious
- Pricing options without Monte Carlo methods is nearly impossible
- Parameters specification is difficult to analyse

Over recent years, a next gen of DT models able to cope with some of these difficulties:

Option Pricing with Exponential Affine SDF

Assumption on pricing kernel:

\[ M_{t,T} = \beta \left( \frac{S_T}{S_t} \right)^\gamma \tag{90} \]

Like in Black Scholes case. Constraints to rule out arbitrage opportunities:

\[
\begin{align*}
\mathbb{E}[M_{t,T}] &= e^{-r(T-t)} \\
\mathbb{E}[M_{t,T}S_T|S_t] &= S_t \\
\phi(\gamma)\beta &= e^{-r(T-t)} \\
\phi(\gamma + 1)\beta &= 1 \\
\beta &= \phi(\gamma + 1)^{-1} \\
\gamma &\text{ solves } \log \phi(\gamma) - \log \phi(\gamma + 1) + r(T-t) = 0
\end{align*} \tag{91, 92, 93} \]

The Heston and Nandi process

The process

\[
\log S_t = \log S_{t-\Delta} + r + \lambda h_t + \sqrt{h_t} \epsilon_t \\

h_t = \omega + \beta h_{t-\Delta} + \alpha \left( \epsilon_{t-\Delta} - \gamma \sqrt{h_{t-\Delta}} \right)^2
\]

\( \epsilon_t \sim N(0, 1). \)

Remarks:

- \( r \) is the risk free rate
- \( \lambda h_t \) is the risk premium:

\[
\mu - r = \lambda h_t
\]

Market price of risk is proportional to \( \sqrt{h_t} \) (CIR case)

\[
\frac{\mu - r}{\sqrt{h_t}} = \lambda \sqrt{h_t}
\]

- Leverage effect \( Cov_{t-\Delta}(\log S_t, h_{t+\Delta}) = -2\alpha \gamma h_t \neq 0. \)
Pricing options in the Heston Nandi framework

Change in probability measure

Find $\lambda$ such that

\[ E^Q[S_{t+\Delta}] = S_te^r \]  \hspace{1cm} (98)

\[ \Leftrightarrow \lambda^* = -\frac{1}{2} \]  \hspace{1cm} (99)

Conclusions

- Both RN and Hist. distributions of next period returns considering today are gaussian
- Conditional pricing kernel for $t + \Delta | t$ is exponential affine
- Over one time bucket $\Rightarrow$ obeys the Black Scholes world.
- For the remaining dates: we don’t know the distribution / pricing kernel but we know the conditional Laplace transform.
Convergence of the process to a continuous time process

This process weakly converges to a Heston process for specific parameters settings, once the time bucket impact on parameters is taken into account:

\[ \gamma = \frac{\gamma^*}{\sqrt{\Delta}}, \beta = \beta^* \Delta, \alpha = \alpha^* \Delta, \omega = \omega^* \Delta \]  

(100)

Introducing these parameters:

\[ h_{t+\Delta} = \omega^* \Delta + \beta^* h_t \Delta + \alpha^* \Delta \left( \epsilon_t - \frac{\gamma^*}{\sqrt{\Delta}} \sqrt{h_t} \right)^2 \]  

(101)

\[ = (\omega^* + (\beta^* + \alpha \gamma^2) h_t) \Delta + \alpha(\sqrt{\Delta \epsilon_t})^2 - \alpha \gamma \sqrt{h_t \epsilon_t} \]  

(102)

\[ h_{t+\Delta} - h_t = \kappa \Delta \left( \theta - h_t \right) \Delta + \sigma \Delta \sqrt{h_t \epsilon_t} + \alpha(\sqrt{\Delta \epsilon_t})^2 \]  

(103)

When \( \Delta \) approaches 0:

\[ dh_t = \kappa (\theta - h_t) dt + \sigma \sqrt{h_t} dW_t \]  

(104)

\( \Rightarrow \) Cox Ingersoll Ross process.
Characteristic function of the Heston and Nandi process

Exponential Affine Characteristic function:

$$\mathbb{E}[e^{\phi \log S_T} | S_t] = e^{A_t + \phi \log S_t + B_t \sigma_t^2} \tag{105}$$

with $A$ and $B$ recursively defined through Riccati equations:

$$A_t = A_{t+\Delta} + \phi r + B_{t+\Delta} \omega - \frac{1}{2} \log(1 - 2\alpha B_{t+\Delta}) \tag{106}$$

$$B_t = \phi(\gamma - \frac{1}{2}) - \frac{\gamma^2}{2} + \beta B_{t+\Delta} + \frac{1}{2}(\phi - \gamma)^2 \frac{1}{1 - 2\alpha B_{t+\Delta}} \tag{107}$$

with terminal conditions:

$$A_T = 0 \tag{108}$$

$$B_T = 0 \tag{109}$$

Computations possible because $\mathbb{E}_t[e^{a(z-b)^2}] = e^{-\frac{1}{2} \log(1-2a)} + \frac{ab^2}{1-2a}$, with $z \in N(0, 1)$. 
Quadratic Heston Nandi model

- 0.3 years
- 0.5 years
- 1 years

% of pricing error vs. Moneyness
Heston Nandi model (opt.)

- 0.3 years
- 0.5 years
- 1 years
### Parameter estimates

#### Time Series Estimates

<table>
<thead>
<tr>
<th></th>
<th>$\omega$</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>$\gamma$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>5.04E-07</td>
<td>0.664</td>
<td>1.496E-06</td>
<td>462.32</td>
<td>0.91</td>
</tr>
<tr>
<td>5% quantile</td>
<td>3.60E-07</td>
<td>0.657</td>
<td>1.476E-06</td>
<td>461.54</td>
<td>0.20</td>
</tr>
<tr>
<td>95% quantile</td>
<td>6.32E-07</td>
<td>0.669</td>
<td>1.518E-06</td>
<td>463.03</td>
<td>1.69</td>
</tr>
</tbody>
</table>

#### Option Implied Estimates

<table>
<thead>
<tr>
<th></th>
<th>$\omega$</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>$\gamma$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>4.33E-05</td>
<td>0.589</td>
<td>7.97E-07</td>
<td>458.20</td>
<td>5.10</td>
</tr>
<tr>
<td>5% quantile</td>
<td>5.67E-06</td>
<td>0.589</td>
<td>-8.59E-07</td>
<td>452.57</td>
<td>0.20</td>
</tr>
<tr>
<td>95% quantile</td>
<td>1.95E-04</td>
<td>0.589</td>
<td>1.62E-06</td>
<td>463.10</td>
<td>10.73</td>
</tr>
</tbody>
</table>
Christoffersen and alii’s model

Problem with HN model: No enough time invariant skewness

⇒ Present a new TS model based on inverse Gaussian distribution

\[
\begin{align*}
\log S_{t+\Delta} &= \log S_t + r\Delta + v h_{t+\Delta} + \eta y_{t+\Delta} \\
h_{t+\Delta} &= w + bh_t + cy_t + \frac{ah_t^2}{y_t} \\
y_{t+\Delta} &\sim IG\left(\frac{h_{t+\Delta}}{\eta^2}\right)
\end{align*}
\]

Features?

- \( y \)’s support is \( \mathbb{R}^+ \Rightarrow \eta < 0 \Rightarrow \) Negative time invariant skewness

- \( \mathbb{E}_t[y_{t+\Delta}] = \mathbb{V}_t[y_{t+\Delta}] = \frac{h_{t+\Delta}}{\eta^2} \Rightarrow \) variance and expectation of log asset price are affine functions of \( h_{t+\Delta} \)

- Nests the HN model for particular choices for parameters
Specificity of the Inverse Gaussian distribution

Density $\mu, \lambda > 0$

$$f(x; \mu \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp \left[-\frac{\lambda(x - \mu)^2}{2\mu^2 x}\right]$$  \hspace{1cm} (113)

**Inverse Gaussian:** first time such that a BM with drift $v$ and vol $\sigma$ reaches $\alpha$ is distributed as $IG\left(\frac{\alpha}{v}, \frac{\alpha^2}{\sigma^2}\right)$

**Distribution of a sum of IG:**

$$X_i \sim IG(\mu_0 w_i, \lambda_0 w_i^2)$$ \hspace{1cm} (114)

$$S = \sum_{i=1}^{n} X_i \sim IG(\mu_0 \bar{w}, \lambda_0 \bar{w}^2), \bar{w} = \sum_{i=1}^{n} w_i$$ \hspace{1cm} (115)

**Scaling**

$$X \sim IG(\mu, \lambda) \Rightarrow tX \sim IG(t\mu, t\lambda)$$ \hspace{1cm} (116)

**In CHJ:** $\lambda = \mu^2$ and $\mu = \delta$
Characteristic Function of the CHJ process

**Characteristic function** is exponential affine function of state variables:

\[
\mathbb{E}_t[S_t^\phi] = e^{\phi \log S_t + A(t,T) + B(t,T)h(t+\Delta)}
\]

With deterministic coefficients:

\[
A(t,T) = A(t + \Delta, T) + \phi r \Delta + wB(t + \Delta) - \frac{1}{2} \log (1 - 2a\eta^4B(t + \Delta)) \tag{118}
\]

\[
B(t,T) = bB(t + \Delta) + \phi v + \eta^{-2} - \eta^{-2} \sqrt{(1 - 2a\eta^4B(t + \Delta))(1 - 2cB(t + \Delta, T) - 2\eta\phi)} \tag{119}
\]

\[
A(T,T) = B(T,T) = 0 \tag{120}
\]

**Nice result obtained because**

\[
\mathbb{E}[e^{\phi y + y}] = \frac{\delta}{\sqrt{\delta^2 - 2\theta}} \exp \left( \delta - \sqrt{(\delta^2 - 2\theta)(1 - 2\phi)} \right) \tag{121}
\]

⇒ Key ingredient to full discrete time option pricing models.
III. Bridging the gap between historical and risk neutral distributions

III-1. Risk aversion and asset pricing
Introduction to State Price Density

Let $S_t$ be the price of a spot asset.

When $t > \text{today}$ ⇒ the price of this asset is uncertain.

This uncertainty is quantified though probability distribution functions.

**The subjective distribution**: expectations of the representative agent for the future price of this asset.

**The historical distribution**: historical dynamics of asset prices. Physical or objective distribution.

**The risk neutral distribution**: usually, distribution that makes agents neutral toward risk and allows used to give a fair price to assets.

The truth: RND is just:

- A pricing measure: PD consistent with current derivative prices
- Equivalent to Subj/historical distribution
- That rules out arbitrage opportunities
- No consensus around uniqueness!
Risk premium = 0

Risk Neutral distribution: used for asset pricing using no arbitrage arguments

Historical distribution: financial asset time series distribution

Information processing?

Subjective distribution: reflects the expectations of the representative agent

Risk aversion correction
Changes in the risk aversion lead to...

Changes in hedging strategies on the ECX market that impact...

Option prices ↔ Implied volatilities

Depend on strikes

Change in the risk neutral distribution

Changes in the risk aversion!!

Change in the historical distribution

Term structure of futures value

- Expectation
- Volatility
- Skewness
- Kurtosis
State Price Density

SPD has several names:

- State Price Density
- Risk Neutral Density
- Pricing Density
- Equivalent Martingale Measure (Harrison and Kreps (1979))

Here, we denote it $q_{t,T}(S_T)$ this probability distribution density.

Formally, we have:

$$
\mathbb{E}^Q[g(S_t)|\mathcal{F}_t] = B(\tau) \int_0^\infty g(S_T)q_{t,T}(S_T)dS_T
$$

(122)
Can be written as:

$$
\mathbb{E}^Q[g(S_t)\mid \mathcal{F}_t] = \int_0^\infty g(S_T)p_{t,T}(S_T) \times B(\tau)\frac{q_{t,T}(S_T)}{p_{t,T}(S_T)} dS_T
$$

(123)

$$
= \mathbb{E}^P [S_T M_{t,T}(S_T)]
$$

(124)

where $M_{t,T}(S_T)$ has different names:

- SPD per unit prob.
- Pricing kernel
- Stochastic discount factor

Remark:

- $q_{t,T}(S_T)$ is the forward price of an Arrow Debreu security that pays 1 is the state at time $T$ is $S_T$ (so to speak)
- $B(\tau) = B(T - t) = e^{-r(T-t)}$ is the price of zero coupon bond paying 1 at time $T$ with certainty (no credit risk)
Empirical Pricing Kernel

EPK is linked to the difference between:

- the historical distribution, incorporating risk in its dynamics
- the risk neutral distribution, dumping risk in its dynamics

⇒ from one to another, need to handle risk correction.

Remarks:

1. For the Gaussian framework, you know it is just done through a change in the drift.

2. For more general cases... we don’t really know. Most of the time: change in the drift, in an awkwardly defined way...

The problem: nobody knows what is really at work in the market.
Most likely, most of the first three moments actually changes:

- change in the **drift**, as documented in the literature related to the BS model

- change in the **volatility**: volatility trading exists, based on the spread between historical and implied vol:
  - Implied vol = huge problem to handle through models
  - But IV is easy to measure through inversing the BS formula
  - Problem: which strike to chose to symbolize the RN vol?
  - Historical vol: hard to measure: rolling? GARCH based? Non parametric estimates?

- change in the **skewness**: growing literature on the change in the skewness.
  Not enough skewness in the time series when compared to option prices ⇒ **biggest mispricing errors usually in the left tail**.
  Hedges against drop in the market for long positions in the underlying. Crash-o-phobia.
  Like for volatility: how to measure it? Model dependency?

- Maybe **higher order moments**?
Risk Aversion and Option Prices

Link between RN and Subj: risk aversion in the market.

From wiki: Risk aversion is the reluctance of a person to accept a bargain with an uncertain payoff rather than another bargain with a more certain, but possibly lower, expected payoff.

In Jackwerth (2001):

\[
\text{Risk Neutral Distrib.} = \text{Subjective Distrib.} \times \text{Risk Correction}
\]  \hfill (125)

Recall that Arrow Pratt risk aversion measure for the future state \( S_T, T > t \) is given by:

\[
RA(S_T) = -\frac{U''(S_T)}{U'(S_T)}
\]  \hfill (126)

It is linked to the pricing kernel:

\[
RA(S_T) = -d \log M_{t,T}(S_T) = \frac{p'(S_T)}{p(S_T)} - \frac{q'(S_T)}{q(S_T)}
\]  \hfill (127)
Remarks:

- In a Gaussian framework: risk correction (setting market price of risk \( \sigma^{-1}(\mu - r_f) \) to 0) is the same thing as risk aversion correction.

- Wrong in a more general framework!! The pricing kernel does more than changing the drift!

In a one period framework

\( \Rightarrow \) Empirical option pricing is a difficult art.
Proofs of the previous relations

Assume a one period general framework.

Proof based on the assumed existence of holly market spirit known as representative agent
⇒ More like an "average trader" in the market, with a utility function to represent it.

Settings:
Let $U(.)$ be the utility function of the representative agent of the economy that is considered.
Program : Max on total discounted utility by parameterizing the saving $X$:

$$F(X) = U(C_0) + \beta \mathbb{E}^H [U(C_1(X))].$$

(128)

Initial endowment $W = C_0 + X$

At time 0, saving $X$ invested in a financial asset whose certain price $S_0$.

Delivers $S_1$ at time 1, yielding an risky return $R = \frac{S_1}{S_0} - 1$. 
At time 1, the uncertain consumption of the agent is:

\[ C_1 = X (1 + R). \]  (129)

Introducing budget constraint in the program:

\[ \max_X F(X) = U(W_0 - X) + \beta \mathbb{E}^H [U(X (1 + R))]. \]  (130)

The first order conditions of the optimization problem yield:

\[ \beta \mathbb{E}^H \left[ S_1 \frac{U'(C_1(S_1))}{U'(C_0)} \right] = S_0. \]  (131)

Relation true for any asset ⇒ for a pure discount bond:

\[ \beta \mathbb{E}^H \left[ \frac{U'(C_1(S_1))}{U'(C_0)} \right] = e^{-r}. \]  (132)
Combining equations (131) and (132):

\[ S_0 = e^{-r} \int S_1 \frac{U'(C_1(S_1))}{U'(C_0)} p(S_1) dS_1 = e^{-r} \int S_1 \xi(S_1) p(S_1) dX = e^{-r} \mathbb{E}^Q[S_1], \] (133)

Remarks:

- Relation is model-free and distribution-free.
- \( \beta \) represents a psychologic discount factor.
- \( \mathbb{E}^H[.] \) underlines the fact that the utility expectation is computed in respect with the subjective distribution of \( R \).
- We can note that:
  - \( p(S_1) \xi(S_1) \geq 0 \) a.s..
  - \( \int h(S_1) \xi(S_1) dS_1 = 1 \),
- \( e^{-r} \xi(S_1) \) is the SDF.
The final link:

Changing the notations:

$$\xi(S_1) = \lambda U'(C_1(S_1)),$$

with:

$$\lambda = \frac{1}{U'(C_0)\mathbb{E}H\left[\frac{U'(C_1(S_1))}{U(S_0)}\right]}$$

Taking log on both sides:

$$\log q(S_1) = \log \lambda + \log U'(S_1) + \log p(S_1)$$

By differentiating we get:

$$RA(S_1) = -\frac{U''(S_1)}{U'(S_1)} = \frac{p'(S_1)}{p(S_1)} - \frac{q'(S_1)}{q(S_1)}$$
Projected pricing kernels

**Problem:** the pricing kernel is only an indirect function of $S_1$, through the final consumption of the representative agent, since:

$$M_{0,1}(C_1(S_1)) = e^{-r} \frac{U'(C_1(S_1))}{U'(C_0)} \mathbb{E}^H \left[ \frac{U'(C_1(S_1))}{U'(C_0)} \right]$$

(138)

Worst: no consensus on the data to include in the pricing kernel (see Rosenberg and Engle on this point).

**Solution:** integrate out the remaining risk factors in the economy / projection of the pricing kernel on the space solely spanned by the asset.

$Z_1$ contains the remaining risk factors.
\[ S_0 = \mathbb{E}[\xi(S_1, Z_1) S_1] \quad (139) \]
\[ = \int_{S_1} \int_{Z_1} \xi(Z_1, S_1) S_1 p(Z_1, S_1) dS_T dZ_1 \quad (140) \]
\[ = \int_{S_1} \int_{Z_1} \xi(Z_1, S_1) p(Z_1, S_1 | S_1) S_1 p(S_1) dZ_1 dS_1 \quad (141) \]
\[ = \int_{S_1} \mathbb{E}_{Z_1}[\xi(Z_1, S_1)] S_1 p(S_1) dS_1 \quad (142) \]
\[ = \int_{S_1} \tilde{\xi}(S_1) S_1 p(S_1) dS_1 \quad (143) \]
\[ = \mathbb{E}[\tilde{\xi}(S_1) S_1] \quad (144) \]

with \( \tilde{\xi}(S_1) = \mathbb{E}_{Z_1}[\xi(Z_1, S_1)] \) the projected pricing kernel.

The stochastic discount factor only depends on the asset price, and not on the representative agent consumption or any other state variable \( \Rightarrow \) more parsimonious modeling.
A special popular case: the power utility function

Power utility:

\[ U(x) = \lambda x^{1-\gamma} \]  \hspace{1cm} (145)

Risk aversion:

\[ RA(x) = -\frac{-(1 - \gamma)\lambda\gamma x^{-(\gamma+1)}}{(1 - \gamma)\lambda x^{-\gamma}} = \frac{\gamma}{x} \]  \hspace{1cm} (146)

Constant relative risk aversion:

\[ RRA(x) = x \times RA(x) = \gamma \]  \hspace{1cm} (147)

Remarks:

- With this utility function: risk perception gets down when wealth gets higher.
- Quite realistic
- Very popular in finance (actually nests BS, Markowitz, CAPM, ICAPM...
Implied pricing kernel?

\[
\frac{U'(S_T)}{U'(S_t)} = \left(\frac{S_T}{S_t}\right)^{-\gamma}
\]  

(148)

Now, need for assumptions: \( \log \frac{S_T}{S_t} \sim N(\mu, \sigma) \)

\[
\mathbb{E} \left[ \frac{U'(S_T)}{U'(S_t)} \right] = \mathbb{E} \left[ e^{-\gamma \log \frac{U'(S_T)}{U'(S_t)}} \right] \\
= e^{-\mu \gamma + \frac{\gamma^2 \sigma^2}{2}}
\]

(149)

(150)

The pricing kernel:

\[
M_{t,T}(S_T) = e^{-r(T-t)-\mu \gamma + \frac{\gamma^2 \sigma^2}{2} \left(\frac{S_T}{S_t}\right)^{-\gamma}}
\]

(151)

\[
= \exp(A(t, T) - \gamma \log \frac{S_T}{S_t})
\]

(152)

Exponential affine function of the returns!
What about the Black Scholes world?

In the BS world:

- Under $P$, $\log \frac{S_T}{S_t} \sim N(\frac{\mu - \sigma^2}{2}(T - t), \sigma \sqrt{T - t})$

- Under $Q$, $\log \frac{S_T}{S_t} \sim N((r - \frac{\sigma^2}{2})(T - t), \sigma \sqrt{T - t})$

**Pricing kernel:**

$$M_{t,T}(S_T) = e^{\frac{(\mu-r)(\mu+r-\sigma^2)}{2\sigma^2} (T-t) - r(T-t)} \left( \frac{S_T}{S_t} \right)^{\frac{r-\mu}{\sigma^2}}$$ (153)

$\Rightarrow$ Just power utility with relative risk aversion equal to $\frac{r-\mu}{\sigma^2}$!

**Remark:**

- Aversion not proportional to market price of risk but to $\frac{M_{PR}}{\sigma}$: risk aversion/correction not equal!

- If $r < \mu$ downward sloping
Historical and RN distributions

Moneyness

Densities

Historical Distribution

Risk Neutral Distribution

Pricing Kernel

Moneyness

Absolute risk aversion

ARA
III. Bridging the gap between historical and risk neutral distributions

III-2. Estimation strategies
Three estimation strategies

After Bertholon, Monfort and Pegoraro (2008), three possible estimation strategies

1. The direct approach: the Black Scholes approach. Start from the historical distribution, assume a shape for the risk premium and the risk neutral distribution is a by-product of the approach.

2. The risk neutral constrained direct modelling: choose a specification for the risk neutral and the historical distribution and obtain the risk premium as a by-product.

3. The back modelling approach: choose a specification for the risk neutral distribution and for the risk aversion and obtain the historical distribution as a by-product.
How good is the direct approach?

To quants, the direct approach clearly makes no sense:
- Vanilla option prices already are in the market
- All they need is a stable model with parameters to calibrate...
- ... so that they can compute the fair price of more elaborated contingent claims.

To statisticians, it is clearly different:
- They like to prove that quants are wrong about their models
- They are interested in the empirical historical behavior of asset prices

But the specification of the model under the historical distribution is clearly not enough!
- Consistency with time varying moments is not enough
- Consistency with the unconditional moments is not enough
- Need to handle too:
  - change in these moments from $\mathbb{P}$ to $\mathbb{Q}$
  - the correct specification of the risk premium

More: how crucial is that to understanding the rationale behind?
How to recover EPK?

**Global idea**: use the relation between RN, Hist. distributions and PK to recover one of them.

2. Engle and Rosenberg 2002: GARCH GJR for Hist and orthogonal polynomial for the EPK
3. Aït Sahalia and Lo 2000: Nadaraya Watson Kernel estimator for RN distribution and historical distributions

**Main conclusion**: EPK is not a monotonous function of the future wealth (expect in the BAEM case!)
A first glance at EPK

A benchmark model: the non parametric approach.
Build on the link between implied vol. and RND.
Breeden & Litzenberger relation:

\[
\frac{\partial^2 C(\tau, K)}{\partial K^2} = e^{r\tau} q(S_\tau | S_\tau = K)
\] (154)

⇒ strike by strike state price density!

Another way to put it: Ait Sahalia and Lo 2000
- Butterfly strategy: sell two calls at strike \( K + \epsilon \) and buy two call at strike \( K - \epsilon \).
- By \( \frac{1}{\epsilon} \) shares of this portfolio.
- When \( \epsilon \to 0 \), final payoff degenerates to a Dirac mass at \( S_T = K \) (Arrow Debreu asset)
- Using no arbitrage arguments, its price is \( e^{-r\tau} q(S_T | S_T = K) \epsilon \).
- Actual price of \( \frac{1}{\epsilon} \) of this portfolio?

\[
P = \lim_{\epsilon \to 0} \frac{1}{\epsilon^2} \left(-2C(\tau, K) + C(\tau, K + \epsilon) + C(\tau, K - \epsilon)\right)
\] (155)

\[
= \frac{\partial^2 C(\tau, K)}{\partial K^2}
\] (156)
Very useful result! Since implied volatility is also related to the strikes, by using the right formula to express its dependency ⇒ Recover SPD easily through the "BS filter".

![Graphs showing elementary payoffs and final payoffs for different epsilon values.](image-url)
A Common approach:

\( \sigma_{BS}(K) \) is a function of the strike price \( K \) and we use the BS formula at each strike point

\[
e^{r\tau} \frac{\partial^2 C(\tau, K, \sigma_{BS}(K))}{\partial \sigma^2} = q(S_T|S_T = K)
\]  

(157)

Different applications:

- Ait Sahalia and Lo 2000: Nadaraya Watson Kernel Estimate

\[
\hat{\sigma}(\tau, K) = \frac{\sum_{i=1}^{n_K} \sum_{j=1}^{n_\tau} \sigma_{BS}(K_i, \tau_j) Ker \left( \frac{K-K_i}{h_K} \right) Ker \left( \frac{\tau-\tau_j}{h_\tau} \right)}{\sum_{i=1}^{n_K} \sum_{j=1}^{n_\tau} Ker \left( \frac{K-K_i}{h_K} \right) Ker \left( \frac{\tau-\tau_j}{h_\tau} \right)}
\]  

(158)

- Shimko 1993, Briere 2006: cubic function estimated by OLS

With the right model for historical distribution ⇒ makes it possible to compute the EPK and make comments on it.
Jackwerth's results:

Figure 3
Risk aversion functions across wealth.
Rosenberg and Engle’s results:

Fig. 5. Empirical pricing kernel using orthogonal polynomial specification in mid-June 1991–1995. We use a four-term expansion: \( M_t(r_{t+1}) = \theta_0, T_0(r_{t+1})\exp[\theta_1, T_1(r_{t+1}) + \theta_2, T_2(r_{t+1}) + \theta_3, T_3(r_{t+1})] \). The pricing kernel estimates using this specification assign greater weight to large negative S&P 500 return states and lesser weight to large positive S&P 500 return states than do the estimates using the power specification.
Ait Sahalia and Lo’s results:
BAEM’s results:

**Illustration:** Chevallier Ielpo 2008 on Carbon Market.
• New market but derivatives
• Spot cannot be used ⇒ use future instead
• Average smile for mat = 1.3 years and average EPK
• Impact of allowances on market risk aversion.
Dec. 2008 ECX Future
Barone-Adesi et alii approach in steps

1. Step 1: under the Historical distribution GARCH GJR model: stochastic vol and leverage effect.
2. Step 2: Estimation by QML with a large sample.
3. Step 3: Recover the residuals: not gaussian, asymmetric distribution, fat tails and neg. skewness.
4. Step 4: Under the RN distribution, another GARCH GJR model, with different parameter values. Conditional distribution under $\mathbb{Q}$? Unknown: directly sample this distribution from the residuals. Aka FHS (Filtering Historical Simulations). Residuals = filtered (not T.V. component).
5. Step 5: Compute option prices using MC method as an average of simulated future payoff.

Finally, we have at hand:

(a) Historical distribution for any point of time
(b) RN dynamics for any horizon
⇒ deduce a semi parametric approach to pricing kernel estimation.
<table>
<thead>
<tr>
<th>Year</th>
<th>$\omega \times 10^6$</th>
<th>$\beta$</th>
<th>$\alpha \times 10^3$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
<td>Mean</td>
<td>Std.</td>
</tr>
<tr>
<td>2002</td>
<td>1.39</td>
<td>0.20</td>
<td>0.93</td>
<td>0.01</td>
</tr>
<tr>
<td>2003</td>
<td>1.28</td>
<td>0.30</td>
<td>0.93</td>
<td>0.01</td>
</tr>
<tr>
<td>2004</td>
<td>1.05</td>
<td>0.06</td>
<td>0.93</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>$\omega^* \times 10^6$</th>
<th>$\beta^*$</th>
<th>$\alpha^* \times 10^3$</th>
<th>$\gamma^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
<td>Mean</td>
<td>Std.</td>
</tr>
<tr>
<td>2002</td>
<td>3.56</td>
<td>2.98</td>
<td>0.86</td>
<td>0.09</td>
</tr>
<tr>
<td>2003</td>
<td>3.42</td>
<td>4.91</td>
<td>0.85</td>
<td>0.20</td>
</tr>
<tr>
<td>2004</td>
<td>1.80</td>
<td>2.03</td>
<td>0.85</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Lessons from BAEM paper

Measure of pricing error: Absolute Mispricing Error

\[ AME = \left| \frac{P_{\text{model}} - P_{\text{market}}}{P_{\text{market}}} \right| \]  \hspace{1cm} (159)

With FHS:

1. Biggest errors are for the left tail: no enough skewness in historical data

2. Since not enough skewness, slope of PK not steep enough $\Rightarrow$ flatter risk aversion than what is really observed in markets

3. Best model ever for right tail

4. No detail for market momentum

5. Computer intensive

6. Empirical results: more time varying skewness in option price and more time invariant skewness
Does the conditional distribution matters more than the change in parameters?

The lesson from BAEM's paper is that the change the dynamics from $\mathbb{P}$ to $\mathbb{Q}$ is empirically changing.

**Key questions:**

- How much does this change matters?
- Is it more important than the conditional distribution of the returns?

A sketch of answers: Chorro, Guégan and Ielpo, 2008, Option pricing under GARCH models with Generalized Hyperbolic innovations.

1. a new option pricing model based on the Generalized Hyperbolic distribution
2. an empirical testing strategy for the assumption of the exponential affine pricing kernel
The option pricing model

Under the historical probability $P$

- the bond price process $(B_t)_{t \in \{0, 1, \ldots, T\}}$
  \[ B_t = B_{t-1} e^r, \quad B_0 = 1, \]  \hspace{1cm} (160)

- the stock price process $(S_t)_{t \in \{0, 1, \ldots, T\}}$
  \[ Y_t = \log \left( \frac{S_t}{S_{t-1}} \right) = r + m_t + \sqrt{h_t} z_t, \quad S_0 = s, \]  \hspace{1cm} (161)

where $z_t \sim GH(\lambda, \alpha, \beta, \delta, \mu)$. 


The option pricing model

Option pricing using the pricing kernel
Basic asset pricing results states that

\[ C_t(T, K) = \mathbb{E}^Q \left[ e^{-\int_t^T r_s ds} (S_T - K)^+ \right] \]

Which is equivalent to

\[ C_t(T, K) = \mathbb{E}^P \left[ e^{-\int_t^T r_s ds} M_{t,T}(S_T) (S_T - K)^+ \right] , \]

with \( M_{t,T}(S_T) \) the pricing kernel over the period \([t, T]\). That is

\[ M_{t,T}(S_T) = \frac{q_{t,T}(S_T)}{p_{t,T}(S_T)} \]
The option pricing model

The Pricing Kernel

\[ M_{t,t+1} = e^{\theta_{t+1}Y_{t+1} + \beta_{t+1}} \] (162)

where \( \theta_{t+1} \) and \( \beta_{t+1} \) are \( \mathcal{F}_t \) measurable random variables

The resulting Risk Neutral distribution

1. The RN distribution is unique
2. The RN distribution is again a conditional GH distribution

the stock price process \((S_t)_{t \in \{0,1,\ldots,T\}}\)

\[ Y_t = \log \left( \frac{S_t}{S_{t-1}} \right) = r + \sqrt{h_t} z_t, \quad S_0 = s, \] (163)

where \( z_t \rightleftharpoons GH(\lambda, \alpha, \beta + \theta_t \sqrt{h_t}, \delta, \mu) \).
Econometric considerations

1. The choice for the variance dynamics is led by the quest for leverage effects
   - Asymmetric GARCH (GJR) process:
     \[ h_t = \omega_0 + \alpha \epsilon_{t-1}^2 + \beta h_{t-1} + \gamma \max(0, -\epsilon_{t-1})^2 \]
   - Nelson (1991) exponential GARCH process:
     \[ \log h_t = \omega_0 + \alpha (|\epsilon_{t-1}| + \gamma \epsilon_{t-1}) + \beta \log h_{t-1} \]

2. Estimation strategy: two steps estimation procedure
   - QML estimation of the dynamic variance process
   - ML estimation of the GH parameters using the residuals estimated on the previous step
Competing models

We compare the performance for option pricing purposes of our model with respect to several other models:

1. The Black Scholes model (benchmark)
2. The Heston and Nandi (2000) model
3. The Duan (1995) model
4. Models conditionally mixture of Gaussian distributed
   (a) GJR GARCH model
   (b) Nelson GARCH model
5. Models conditionally Normal Inverse Gaussian distributed
   (a) GJR GARCH model
   (b) Nelson GARCH model
6. Models conditionally Hyperbolic distributed
   (a) GJR GARCH model
   (b) Nelson GARCH model
7. Models conditionally GH distributed
   (a) GJR GARCH model
   (b) Nelson GARCH model
The dataset

We use data coming from different indexes:

- CAC, DAX, FTSE and SP500 stock indexes
- To estimate the historical distribution, index spot closing price from 16/07/1987 to 03/01/2008
- The option prices used for option errors measurement: datasets including each available contracts from 01/02/2006 to 10/26/2007.

Once estimated, option prices computation through:

- Monte Carlo method (10000 runs)
- Duan and Simonato (1998) Martingale restriction

Various sample sizes: 2500, 3500, 4000...
The dataset

Various strategies:
Compare the average absolute pricing errors across models:

$$\sum_t \sum_K \left| \frac{P_t(K) - P_t^{theo}(K)}{P_t(K)} \right|$$

Either using:
- the exponential affine pricing kernel + exact martingalisation (risk neutral approach)
- only exact martingalisation (historical approach)
Comparison between RN and Historical option pricing

<table>
<thead>
<tr>
<th></th>
<th>.8-.9</th>
<th>.9-1</th>
<th>1-1.1</th>
<th>1.1-1.2</th>
<th>&gt;1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson MN Historical</td>
<td>0.073</td>
<td>0.035</td>
<td>0.090</td>
<td>0.205</td>
<td>0.421</td>
</tr>
<tr>
<td>Nelson NIG Historical</td>
<td>0.074</td>
<td>0.037</td>
<td>0.096</td>
<td>0.216</td>
<td>0.490</td>
</tr>
<tr>
<td>Nelson GH Historical</td>
<td><strong>0.070</strong></td>
<td>0.032</td>
<td>0.112</td>
<td>0.225</td>
<td>0.401</td>
</tr>
<tr>
<td>Nelson MN RN</td>
<td>0.074</td>
<td>0.038</td>
<td>0.092</td>
<td>0.223</td>
<td><strong>0.359</strong></td>
</tr>
<tr>
<td>Nelson NIG RN</td>
<td>0.071</td>
<td><strong>0.031</strong></td>
<td>0.108</td>
<td>0.257</td>
<td>0.497</td>
</tr>
<tr>
<td>Nelson GH RN</td>
<td>0.073</td>
<td>0.034</td>
<td><strong>0.090</strong></td>
<td>0.212</td>
<td>0.474</td>
</tr>
</tbody>
</table>
Lessons from the equity story

- Moving beyond Black Scholes is an obvious necessity
- Black Scholes however remains a very useful tool to understand how the option market moves.
- Less parameter specification than the uncond. disturbance.
- Volatility structure might not be what you should focus on.