Predicting CAPM residuals volatility with aggregate public sentiment

Summary

- Part 1 : Predicting macro-economic statistics with public sentiment analysis
- Part 2 : CAPM model : critics and refinement
- Part 3 : Results

Part 1 : Predicting macro-economic statistics with public sentiment analysis

Twitter may predicts the stock market

Claims

- Research paper from Johan Bollen et al. (2010)
- Predicts with a 3 day lag the Dow Jones Industrial Average movement with a precision of 86.7%
- 10 millions tweets selected among 500 millions tweets from March to December 2008

Twitter may predicts the stock market

Methodology

- Each tweet is analyzed with two mood assessment tools
- Opinion Finder measures positive versus negative mood from text content
- GPOMS measures 6 different mood dimensions from text content : *Calm, Alert, Sure, Vital, Kind and Happy*
- Strong correlation between Calm scores and DJIA

Twitter may predicts the stock market

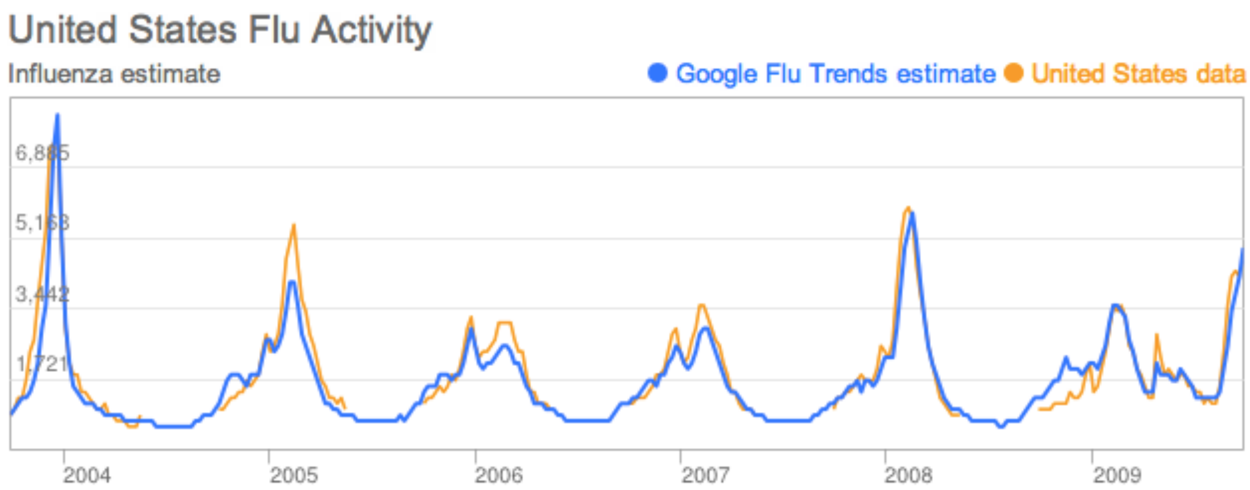
Observations and critics

- Why Calm score is relevant for the stock markets ?
- No all tweets from the collection were made in the United States
- Raw approach : no use of common accepted model in the financial industry
- Strong increase of volume of tweets (350 millions per month against 50 millions per month when the analysis was made)

Predicting macro-events with Google

Google Flu Trends

- Strong correlation was found between an aggregation of search queries on Google and the flu activity (reported by US centers for disease control)



Market anxiousness and volatility

Definitions

- Volatility is defined as the standard deviation of the stock returns
- Implied Volatility is derived from Black-Scholes options pricing formula for example

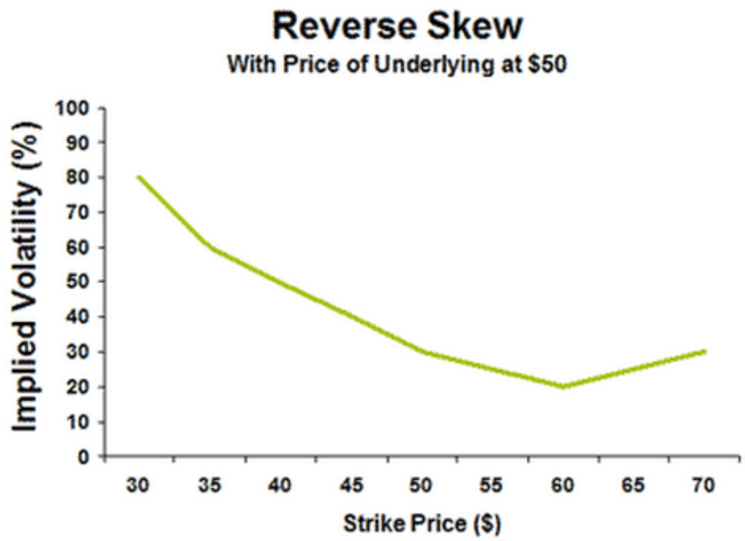
$$C(S,t) = N(d_1)S - N(d_2)Ke^{-r(T-t)}$$

$$d_{1,2} = \frac{\ln\left(\frac{S}{K}\right) + \left(r \pm \frac{\sigma_{implied}^2}{2}\right)(T-t)}{\sigma_{implied}\sqrt{T-t}}$$

Market anxiousness and volatility

Signature

- Volatility skewness (equity options market) is the signature of the relationship between market anxiousness and volatility



Part 2 : CAPM, critics and refinement

Capital Asset Pricing Model

The equation

- Introduced by Sharpe (1964) and Linter (1965)
- Predicts the relationship between between risk and the return of a portfolio

$$R_i - R_F = \alpha_i + \beta_i \cdot (R_M - R_F) + \varepsilon$$

- Ri return of an asset or a specific portfolio
- Rf risk-free rate
- Rm return of the market
- ε error

Capital Asset Pricing Model

Interpretation and resolution

- The return of an asset or a portfolio is proportional to the covariance between its return and the market returns

$$\beta_i = \frac{\text{cov}(R_i, R_M)}{\text{var}(R_M)}$$

- Resolution with the classical OLS estimator

$$Y = X\beta + \varepsilon$$

$$\hat{\beta} = (X'X)^{-1} X'Y$$

Capital Asset Pricing Model

Critics

- Widely used for simplicity but CAPM fails in general to predict returns with accuracy
- βi only reflects the market variation and not the nature of the stocks returns
- Fama and French observed that some stocks performed better than the market : small caps and stocks with a high-book-to-market ratio

Capital Asset Pricing Model

Fama French three factors model

- Introduced in 1993

$$R_i - R_F = \alpha_i + \beta_i \cdot (R_M - R_F) + b_s SMB + b_v HML + \varepsilon$$

  - SMB or Small Minus Big : measure of the historic excess returns of small caps against big caps
  - HML or High Minus Low : measure of the historic excess returns of value portfolio against growth portfolio
  - ε abnormal Fama French return

Capital Asset Pricing Model

Critics

- The classic OLS estimator supposes that the residual ε has a constant variance (spherical variance assumption)

$$E(\varepsilon'\varepsilon \mid X) = \sigma^2 Id$$

- In case of heteroskedasticity, this assumptions will not hold and we will have to apply a generalized least square estimator

Estimation of the error

ARCH/GARCH effect in abnormal returns

- It is possible to use ARCH/GARCH model to predict the variance of the residual from the CAPM regression

$$R_i - R_F = \alpha_i + \beta_i \cdot (R_M - R_F) + \sigma_t z_t$$

  - σt is variance of the residual
  - zt is a strong white noise (random gaussian independent with mean zero and variance one)

- In that case, abnormal returns are independent between each others but their variance follows a specific process

Estimation of the error

Different estimation of beta with GLS

- The Generalized Least Squares estimator is :

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

with $\Omega = E(\varepsilon'\varepsilon \mid X)$

- In case of ARCH/GARCH effects, Ω is diagonal and easy to inverse

$$\Omega = \begin{pmatrix} \sigma_1 & & & & \\ & .. & & 0 & \\ & & .. & & \\ & 0 & & .. & \\ & & & & \sigma_T \end{pmatrix}$$

Estimation of the error

ARCH (q) process

- An ARCH (Auto Regressive Conditional Heteroskedasticity) process models the effect of volatility clustering : resilience of volatility level over time
- The variance of the innovation is a linear function of the size of the squared previous innovations

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$$

Estimation of the error

GARCH (p,q) process

- A GARCH process is an ARCH process combined with an ARMA process (Auto Regressive Moving Average)
- More complex but allows more flexibility

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

- Introduced because ARCH models required in general long lags in the conditional variance equation

Estimation of the error

EGARCH (p,q) process

- Aims to capture the asymmetric effect : a positive returns shock generate less volatility than a negative return

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left(\frac{\varepsilon_t}{\sqrt{\sigma_t}} \right) + \sum_{i=1}^{q} \alpha_i^* \left[\left| \frac{\varepsilon_t}{\sigma_t} \right| - \mu \right] + \sum_{j=1}^{p} \beta_j \ln \sigma_{t-j}^2$$

Excess volatility

CAPM equation for the excess volatility

- The equation for CAPM is the following :

$$R_i - R_F = \alpha_i + \beta_i \cdot (R_M - R_F) + \sigma_t z_t$$

- We can take the variance of the following equation :

$$\mathrm{var}(R_i) = \mathrm{var}(R_M) + \sigma_t^2$$

- It is possible to predict the short-term evolution of σt because it follows a specific process

Refinement of the estimation of σ_t

Observations

- It is possible to track down the excess volatility of an asset based on volatility clustering
- The innovations must be explained by other hidden variables
- We want to see if additional information from aggregation of Google queries can provide a better estimate of this excess volatility

Refinement of the estimation of σ_t

Refinement proposed

- We add a Google Trends component to the regression of the ARCH/GARCH estimation.
- The objective is to try to explain more explicitly the variance of the innovation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{k=1}^{r} \gamma_k G_{t-k}$$

- Gt is the component relative to Google Trends or micro blogging activity such as Twitter

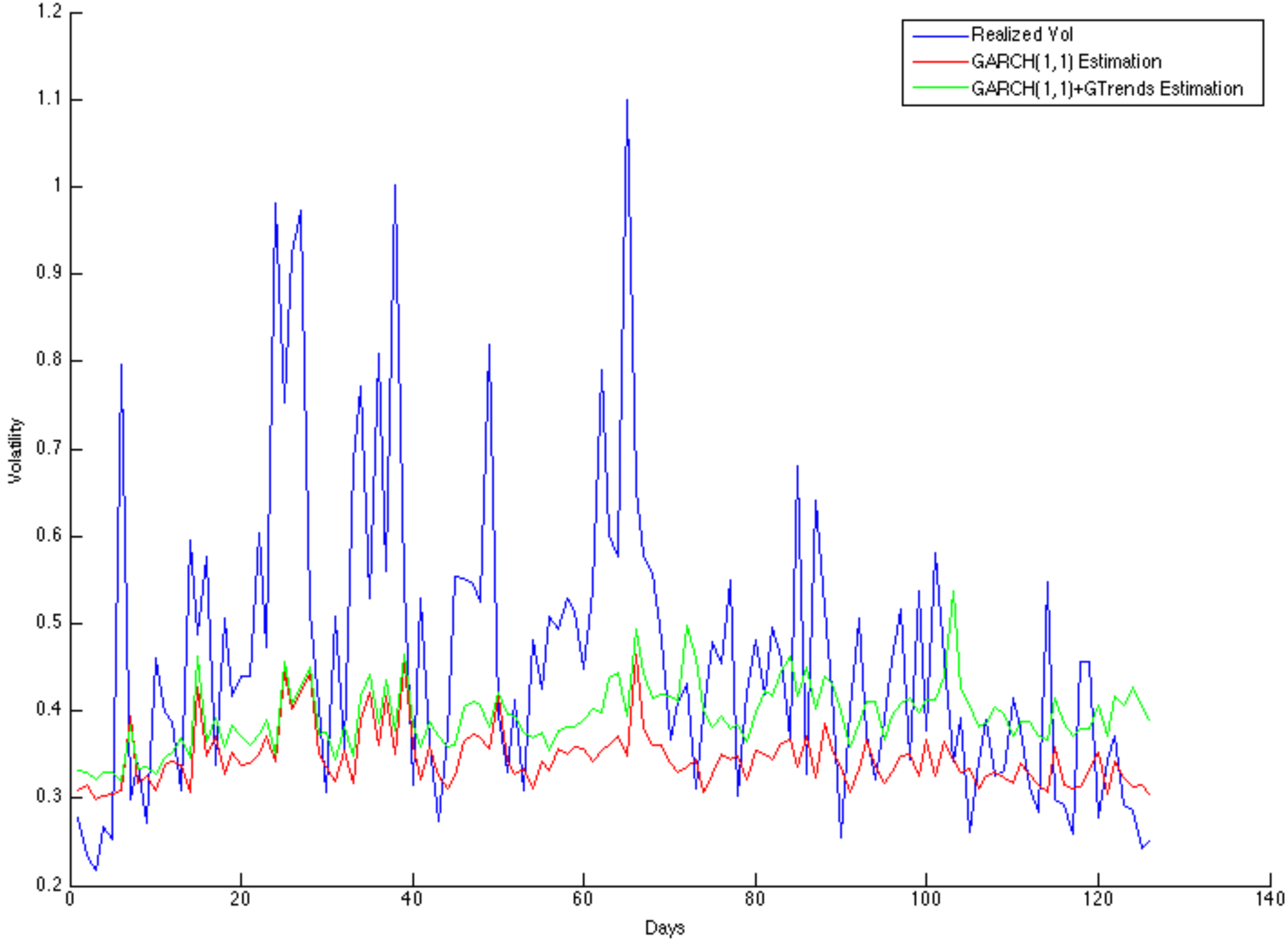
Part 3 : (Very) Preliminary results

Data

Trades and Quotes data

- AMD stock: TAQ data from 01/01/2011 to 12/31/2011
- Compute a daily volatility for price
- Split the year into an estimation period and a simulation period of equal length

Data


Realized Vol
GARCH(1,1) Estimation
GARCH(1,1)+GTrends Estimation
Volatility
Days
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
20
40
60
80
100
120
140

Conclusion