Data Mining in Business Analytics Part 13: Forecasting of energy consumption Forecasting of stock option price

Vadim Strijov

University of Siegen,
Faculty III: Business Economics, Business IT and Commercial Law

March 14–18, 2011

Contents

Energy consumption forecasting

Continued, see Part13_EnergyForecasting_Show.pdf.

Problem statement

Let there be given:
$\mathbf{x} = [x_1, \ldots, x_{T-1}]^T$, $x \in \mathbb{R}^1$ — time series,
$t_{\tau+1} - t_\tau = \text{const}$,
k is a period and $T = mk$.
One must:
to forecast the next value x_T.
The reshaped time series is $(m \times k)$**-matrix**

$$X^{\text{combined}} = \begin{pmatrix} x_T & x_{T-1} & \cdots & x_{T-k+1} \\ x_{(m-1)k} & x_{(m-1)k-1} & \cdots & x_{(m-2)k+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{nk} & x_{nk-1} & \cdots & x_{n(k-1)+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_k & x_{k-1} & \cdots & x_1 \end{pmatrix}.$$

The regression problem

$$X^{\text{combined}} = \left(\begin{array}{c|cccc} x_T & x_{T-1} & \dots & x_{T-k+1} \\ \hline \\ x_{(m-1)k} & x_{(m-1)k-1} & \dots & x_{(m-2)k+1} \\ \dots & \dots & \dots & \dots \\ x_{nk} & x_{nk-1} & \dots & x_{n(k-1)+1} \\ \dots & \dots & \dots & \dots \\ x_k & x_{k-1} & \dots & x_1 \end{array}\right).$$

In a nutshell,

$$\left(\begin{array}{c|c} x_T & \mathbf{x}_{\text{test}}^T \\ \hline \\ \mathbf{y} & X \end{array}\right).$$

In terms of linear regression:

$$\mathbf{y} = X\mathbf{w},$$

$$y^* = x_T = \langle \mathbf{x}_{\text{test}}^T, \mathbf{w} \rangle.$$

Model generation

Let there be given:
a set of the functions $G = \{g_1, \dots, g_r\}$, for example
$g_1 = 1$, $g_2 = \sqrt{x}$, $g_3 = x$, $g_4 = x\sqrt{x}$.

The generated regression model $X =$

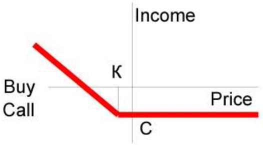
$$\left(\begin{array}{ccc|c|ccc}
g_1 \circ x_{T-1} & \dots & g_r \circ x_{T-1} & \dots & g_1 \circ x_{T-k+1} & \dots & g_r \circ x_{T-k+1} \\
\hline
g_1 \circ x_{(m-1)k-1} & \dots & g_r \circ x_{(m-1)k-1} & \dots & g_1 \circ x_{(m-2)k+1} & \dots & g_r \circ x_{(m-2)k+1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
g_1 \circ x_{nk-1} & \dots & g_r \circ x_{nk-1} & \dots & g_1 \circ x_{n(k-1)+1} & \dots & g_r \circ x_{n(k-1)+1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
g_1 \circ x_{k-1} & \dots & g_r \circ x_{k-1} & \dots & g_1 \circ x_1 & \dots & g_r \circ x_1
\end{array}\right).$$

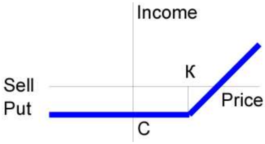
European option

Options are financial instruments that convey the right, but not the obligation, to engage in a future transaction on some underlying security.

$$C_t = F(\sigma, P, B, K, t),$$

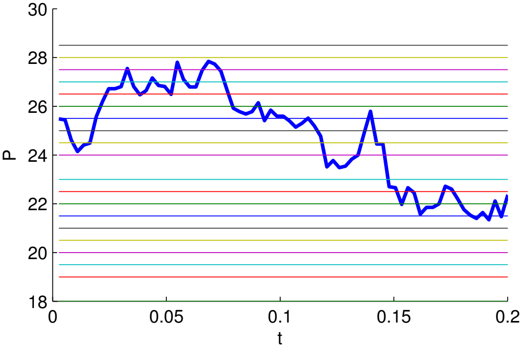
C_t — option price (Call option),
σ — volatility,
P — price of security,
B — risk-free rate,
K — strike price,
t — time to expiration.





$$C_t = \mathcal{N}\left(\frac{ln(\frac{P}{K}) + t(B + \frac{\sigma^2}{2})}{\sigma\sqrt{t}}\right) - Ke^{-Bt}\mathcal{N}\left(\frac{ln(\frac{P}{K}) + t(B - \frac{\sigma^2}{2})}{\sigma\sqrt{t}}\right)$$

Historical price of security

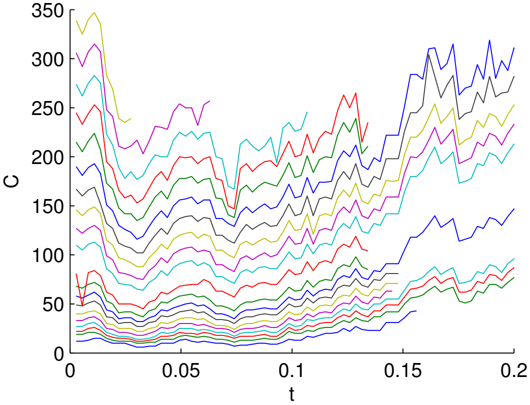


t – time to expiration, years,

P — security price.

Horizontal lines correspond to strike prices K.

Historical prices of options K



t — time to expiration, years,

C — option price.

How to calculate the volatility?

Volatility most frequently refers to the standard deviation of the returns of a financial instrument. It is often used to quantify the risk of the instrument over a time period.

Implied volatility of an option is the volatility implied by the market price of the option based on an option pricing model.

$$\sigma^{\text{imp}} = \arg\min_{\sigma}(C_{\text{hist}} - C(\sigma, P, B, K, t)).$$

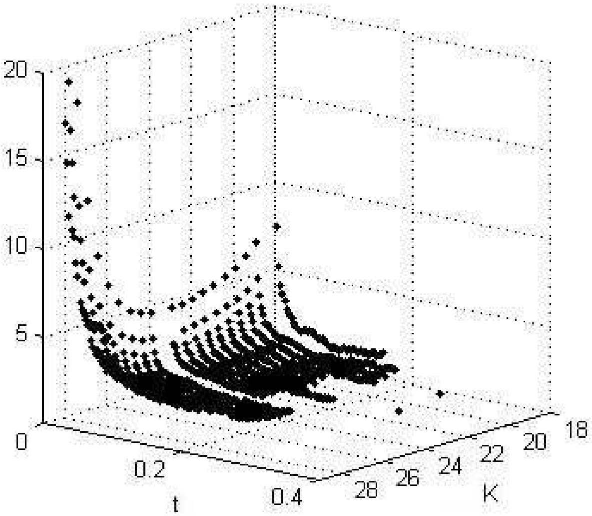
We consider implied volatility as the dependent variable of the regression model.

Our knowledge about volatility helps us to estimate the risk of capital investments.

Implied volatility

The implied volatility depends on the time t and strike price K.

Implied volatility, source data



Volatility model, given by experts

A model for traders at the Russian trade system

$$\sigma = \sigma(\mathbf{w}) = w_1 + w_2(1 - \exp(-w_3 x^2)) + \frac{w_4 \arctan(w_5 x)}{w_5},$$

$$\text{where} \quad x = \frac{\log(K) - \log(C(t))}{\sqrt{t}}.$$

Model assumptions [Daglish, 2006]

- The volatility depends on the option price.
- The volatility proportional to inverse square root of the maturity.

Historical data

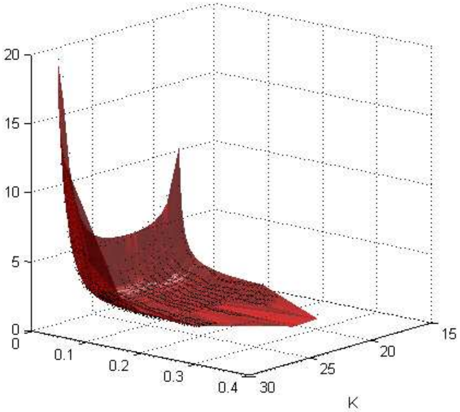
- The following data set was shown:
  - Semi-annual option on Brent Crude Oil, from 02.01.2001 to 26.06.2001. Put option, symbol is CLZ01.
  - quarterly option CNX 100, Deli market index, from 30.09.2007 to 27.12.2007. Call option, symbol CE.
- Basic model by RTS experts:

$$\sigma = \sigma(\mathbf{w}) = w_1 + w_2(1 - \exp(-w_3 x^2)) + \frac{w_4 \arctan(w_5 x)}{w_5},$$

where $x = \frac{\ln K - \ln C(t)}{\sqrt{t}}$.

Non-linear model

$$\sigma = \frac{(w_1 K^2 + w_2 K + w_3)}{\sqrt{t}}$$



Historical data: prices of the options and the security

K= 13.50, 13.00, 12.50, 12.00, 11.50, 11.00

Maturity	K1	K2	K3	K4	K5	K6	Price
-91	0.105	0.16	0.24	0.36	0.56	0.725	11.27
-90	0.105	0.16	0.24	0.35	0.56	0.725	11.29
-87	0.105	0.16	0.21	0.36	0.56	0.725	11.34
-86	0.105	0.16	0.21	0.32	0.56	0.725	11.2
-85	0.105	0.16	0.21	0.32	0.48	0.725	11.18
-84	0.105	0.16	0.215	0.33	0.625	0.725	11.5
-83	0.105	0.16	0.22	0.41	0.625	0.725	11.41
-80	0.105	0.16	0.25	0.42	0.69	0.885	11.48
...	...	...	...	...	...	...	...

Given data

$t \in \{t_1, \ldots, t_\tau, \ldots, t_{64}\} = \mathcal{T}$ is the set of the time ticks,

$K \in \{K_1, \ldots, K_k, \ldots, K_{12}\} = \mathcal{K}$ is the set of the strike prices,

$C = C(t, K)$ are historical option prices

$P = P(t)$ are historical security prices

The desired model is

$$\sigma = f(t, K).$$

Index mapping

Implied volatility

$$\sigma_{t,K} \arg\min_{\sigma}(C_{t,K}^{\text{hist}} - C(\sigma, P_t, B, K, t)).$$

Sample set for regression analysis

$$\sigma_{t,K} \mapsto \sigma^i, i = \tau + k(|\mathcal{T}| - 1),$$

$$(t_i, K_i) \in \mathcal{T} \times \mathcal{K}$$

The regression model

$$\sigma_i' = f(t_i, K_i).$$

Volatility models, toy version

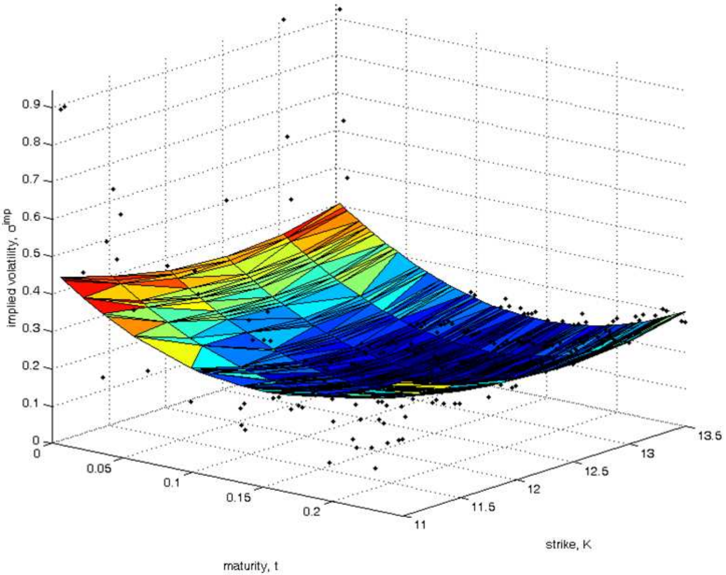
Model-1, polynomial on t, K:

$$f_1(t, K) = w_1 + w_2 t^2 + w_2 tK + w_3 K^2.$$

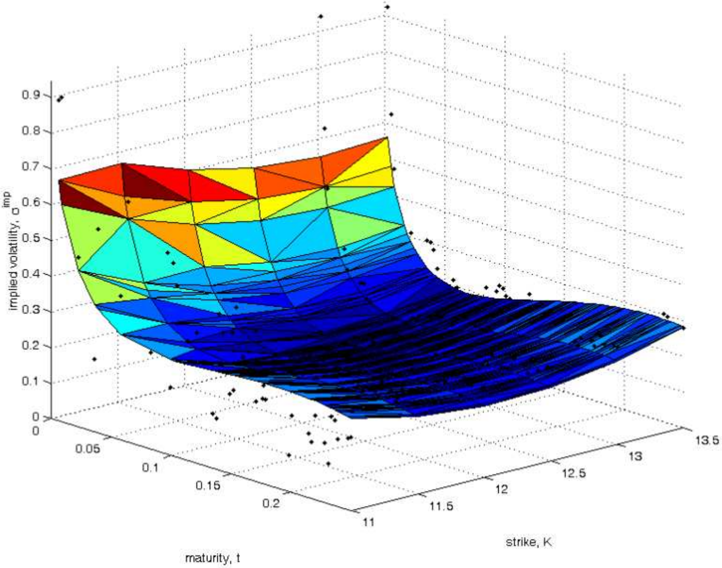
Model-2, fractional power (series) on t, K:

$$f_2(t, K) = w_1 + w_2 t^2 + w_3 K^2 + w_4 \frac{\sqrt{K}}{1 + \exp(t)} + w_5 \frac{\sqrt{tK}\exp(t)}{K}.$$

Volatility model-1 (toy version), polynomial



Volatility model-2 (toy version), fractional power



References

- Hull J. C. Options, Futures and Other Derivatives. USA: Prentice Hall. 2000.
- Also see a small paper on wiki: Option (finance) ↗.