Estimating Changing Volatility in Cash Flow Simulation Based Real Option Valuation with Regression Sum of Squares Error Method

Tero Haahtela
Aalto University, School of Science and Technology
P.O. Box 15500, FI-00076 Aalto, Finland
+358 50 577 1690 tero.haahtela@aalto.fi

Abstract

This paper presents a practical volatility estimation method for cash flow simulation based real option valuation with changing volatility. During cash flow simulation, present value of the future cash flows and their corresponding cash flow state variable values are recorded for all time periods. Then, for each time period, regression analysis is used for relating the present value to the cash flow state variables of the same time period. Each regression equation provides an estimate of the expected present value as a function conditioned on the resolution of all uncertainties up to that time. Then, basic regression statistics of Pearson’s correlation $R^2$ and sum of squares error (or standard error) for each equation provide all the information required for estimating how the standard deviation and volatility of the stochastic process change over time. The method is computationally efficient because it requires only one pass of simulation runs regardless of the number of time periods. It can also handle negative underlying asset values. Straightforward calculation of required regression statistics and their availability in any statistical software and even in spreadsheet programs make this approach very easy to apply for a practitioner.

JEL Classification: G31, G13, D81

Keywords: Real options, cash flow simulation, volatility estimation
1 Introduction

Real options analysis is a framework for valuing managerial flexibility under uncertainty. It has adapted advanced methods from financial derivatives valuation and made valuation of projects with several sequential and parallel decision alternatives more accessible. Difficulty in volatility estimation, however, has been one of the reasons for the slow acceptance of the new valuation framework. Volatility is probably the most difficult input parameter to estimate in real options analysis (Mun 2003), which is also the case with financial options. However, volatility estimation in case of financial options is easier because of the observable historical data and future price information. With real options, especially if related to R&D, there is not necessarily such information available (Lint & Pennings 1999, Newton & Pearson 1994). Therefore, volatility estimation has to be based on some other method.

One alternative is to use Monte Carlo simulation for the gross present value and volatility estimation. According to Trigeorgis (1996), present value calculations may help in finding the correct volatility estimation for the project. In this approach, forecast data for future cash values with probabilities is converted into an estimated underlying asset value and volatility (Newton & Pearson, 1994). The cash flow model is usually a gross present value (i.e. a conventional net present value less investments) where uncertainties related to parameters are presented as objective or subjective distributions with different correlations, time series and other constraints. After the simulation, the mean and the standard deviation of the rate of return, i.e. volatility, are calculated.

Monte Carlo simulation on cash flow calculation consolidates a high-dimensional stochastic process of several correlated variables into a low-dimension process. Most common assumption is the univariate geometric Brownian motion process. Firstly, Monte Carlo simulation technique is applied to develop a probability distribution for the rate of return. Then, volatility parameter \( \sigma \) (or other parameters for other stochastic processes than gBm) of the underlying asset can be estimated with alternative approaches, of which the most common is presumably calculating the standard deviation of the simulated probability distribution for the rate of return.

Several authors have suggested different variations of applying Monte Carlo simulation on cash flow calculation to estimate the volatility. The existing cash flow simulation based volatility estimation methods are the logarithmic present value approach of Copeland & Antikarov (2001) and Herath & Park (2002), conditional logarithmic present value approach of Brandão, Dyer & Hahn (2005), two-level simulation and least-squares regression methods of Godinho (2006), and generalized risk-neutral volatility estimation over different time periods (Hull 2006). All these methods are based on the same basic idea. Monte Carlo simulation technique is applied to develop a probability distribution for the rate of return. Then, the volatility parameter \( \sigma \) of the underlying asset is estimated by calculating the standard deviation of the rate of return.
The previously mentioned methods have different strengths and qualities. Different aspects related to them are theoretical correctness, computational efficiency, capability to handle negative underlying asset values, capability to estimate changing volatilities, ability to separate ambiguity from volatility, and ease of use from the intuitive aspects of understanding the logic behind the volatility estimation as well as ease of use in terms of available software to do the estimation. Luckily, not all of these aspects are relevant in all practical valuation cases. Therefore, usability of the methods depends heavily on the case they are applied.

The approach presented in this paper can be considered as a mixture of the previously mentioned methods. The purpose is to sustain and combine many good qualities of them while keeping the procedure as straightforward as possible for a practitioner. In the method presented in this paper, present value of the cash flows and the cash flow state variable values are recorded for each time period during cash flow simulation runs. Then a regression analysis is run for relating the PV for each year to the corresponding cash flow state variables. Each regression equation provides an estimate of the expected present value as a function conditioned on the resolution of all uncertainties up to that time. Then, basic regression statistics of Pearson’s correlation $R^2$ and sum of squares error (or standard error) for each equation provide all the information required for estimating how the standard deviation and the volatility change over time. The method is computationally efficient as it requires only a single pass of simulation runs, it is easy to understand and use, and it can handle changing volatility and negative underlying asset values.

After introductory section 1, the next section 2 discusses in detail earlier cash flow simulation based methods, because the procedure presented in this paper is heavily based on their ideas and results. Section 3 presents the intuition behind the suggested method, the idea of proportion of unsolved and solved uncertainty during the project’s timespan. The main contribution of this paper is in section 4 that describes the regression sum of squares error and $R^2$ based method for estimating changing volatility in cash flow simulation based real option valuation. Section 5 illustrates the use of this method in a case example. Finally section 6 concludes the paper.

2 Cash flow simulation based volatility estimation methods

Logarithmic present value approach of Copeland & Antikarov (2001)

The first detailed explanation of the use of Monte Carlo simulation for volatility estimation based on cash flows was presented by Copeland & Antikarov (2001). This logarithmic present value approach in terms of Mun (2003, 2006) is based on several assumptions. Firstly, it relies on marketed asset disclaimer and Samuelson’s proof (1965) that correctly estimated rate of return of any asset follows random walk regardless of the pattern of the cash flows. Secondly, approach is based on the idea that an investment with real options should be valued as if it was a traded asset in markets even though it would not be publicly listed. Thirdly, the present value of the cash flows of the project without flexibility is the best unbiased estimate of the market value of the
project were it a traded asset. This is called the *marketed asset disclaimer* assumption. Therefore, simulation of cash flows should provide a reliable estimate of the investment’s volatility.

The idea of Copeland & Antikarov (2001) logarithmic present value approach is – similarly to most other consolidated volatility approaches - analogous to stock price simulations where the theoretical stock price is the sum of all future dividend cash payments, and with real options, these cash payments are the free cash flows. The sum of free cash flows’ present value \( PV_0 \) at time zero is the current stock price (asset value), and at time one, the stock price \( PV_1 \) in the future. As stock price at time zero is known while the future stock price is uncertain, only the uncertain future stock price is simulated. As a result, Copeland & Antikarov (2001) approach uses Monte Carlo simulation on project’s present value to develop a hypothetical distribution of one period returns. On each simulation trial run, the value of the future cash flows is estimated at two time periods, one for the first time period and another for the present time. The cash flows are discounted and summed to the time 0 and 1, and the following logarithmic ratio is calculated according to Equation (1):

\[
z = \ln \left( \frac{PV_1 + FCF_1}{PV_0} \right)
\]

where \( PV_1 \) means present value at time \( t = 1 \), \( FCF_1 \) means free cash flow at time 1, and \( PV_0 \) project’s present value at the beginning of the project at time \( t=0 \). Then, volatility \( \sigma \) is defined as the standard deviation of \( z \).

Modifications to this method include duplicating the cash flows and simulating only the numerator cash flows while keeping the denominator value constant. This reduces measurement risks of auto-correlated cash flows and negative cash flows (Mun 2002). Whereas simulating logarithmic cash flows gives a distribution of volatilities and therefore also a distribution of different real options values, this alternative gives a single-point estimate.

Although the fundamental idea in Copeland & Antikarov (2001) approach is correct, it has one technical deficiency. The method would be appropriate volatility estimate if the \( PV_1 \) were period 1’s expected NPV of subsequent cash flows and this volatility would reflect the resolution of a single year’s uncertainty and its impact on expectations for future cash flows. However, in Copeland & Antikarov’s solution this \( PV_1 \) is the NPV of a particular realization of future cash flows that is generated in the simulation, and therefore the calculated standard deviation is the outcome of all future uncertainties (Smith 2005). Therefore the approach overestimates the volatility.

**Logarithmic present value approach of Herath & Park (2002)**

Herath & Park (2002) volatility estimation is very similar to Copeland & Antikarov (2001) and is based on the same Equation (1). However, whereas in Copeland & Antikarov (2001) only the
numerator is simulated and the denominator is kept constant, Herath & Park (2002) applies simulation to both the numerator and denominator with different independent random variables: “...both \( PV_0 \) and \( PV_1 \) are independent random variables. Therefore, a different set of random number sequences has to be generated when calculating \( PV_0 \) and \( PV_1 \)”. However, this alternative has the same over-estimation problem as the original Copeland & Antikarov (2001), and by having a non-constant denominator, the approach actually worsens the situation and over estimation of the volatility.

**Conditional volatility estimation of Brandão, Dyer & Hahn (2005)**

Other authors have resolved the original problem of Copeland & Antikarov’s approach. Conditional volatility estimation of Brandão, Dyer & Hahn (2005b), based on comments of Smith (2005), and similarly to Godinho (2006), suggest an alternative where the Copeland & Antikarov (2001) simulation model is changed so that only the first year’s cash flow \( FCF_1 \) is stochastic, and the following cash flows \( FCF_2, \ldots, FCF_n \) are specified as expected values conditional on the outcomes of \( FCF_1 \). Thus, the only variability captured in \( PV_1 \) is due to the uncertainty resolved up to that point. The method works well, if the conditional future values are straightforward to calculate or estimate. Then, the standard deviation of the following Equation (2) is used to estimate the volatility \( \sigma \) of the rate of return:

\[
z = \ln\left( \frac{FCF_1 + PV_1(E_1(FCF_2), ..., E_1(FCF_n)|FCF_1)}{PV_0} \right)
\]

(2)

The deficiency with the method is that it may be hard to compute the expected future values given the values simulated in earlier periods. This is true especially for both auto- and cross-correlated input variables in cash flow simulations.

**Two-level simulation of Godinho (2006)**

Two-level simulation of Godinho (2006) is also based on the idea of conditionality in expected cash flows given stochastic \( C_1 \). In comparison with CVE, it works also in situations where conditional outcomes given \( C_1 \) cannot be calculated analytically. Firstly, the simulation is done for the project behavior in the first year. Secondly, project behavior given the first year information is simulated for the rest of the project life cycle. Thirdly, average cash flows after the first year are used to calculate \( PV_1 \), which is then used to calculate a sample of \( z \). Finally, volatility of \( z \) (standard deviation) is calculated. The method mostly suffers from required computation time. This is because the calculation is iterative, meaning that after each first year simulation, a large number of second stage simulation is required. Therefore, the total number of simulations is the product of first and second stage simulation runs. In practice, whereas other methods compute the volatility within a few seconds even with larger models, this procedure requires at least several minutes of computation time with present computers and algorithms. Secondly, the method requires somewhat programming skills.

Inspired by Longstaff & Schwartz (2001), Godinho (2006) presents least squares regression method for volatility estimation. This procedure consists of two simulations. In the first simulation, the behavior of the project value and the first year information is simulated. Then, $PV_1$ is explained with linear regression with first year information as follows according to Equation (3):

$$ PV_1 = a_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n $$

Then, in the second simulation round, volatility is calculated as the standard deviation of $z$

$$ z = \ln \left( \frac{PV_1}{PV_0} \right) $$

Often a good and straightforward approximation is to use first year’s free cash flow $FCF_1$ as the explaining variable with intercept term. Then, in the second simulation round, only first year cash flow is simulated, and the estimation model is used to calculate the expected value of $PV_1$ for calculating the sample $z$.

Generalized risk neutral valuation approach

There is another very effortless method for finding the volatility. It is based on the assumptions and qualities of the gBm and its lognormal underlying asset value distribution. Very similar thoughts are presented by Smith (2005) suggesting that correct parameterization for the mean value and volatility could be found by changing the volatility until the underlying asset mean and standard deviation match the simulated cash flow and its standard deviation. However, if common gBm assumptions hold, this can actually be solved analytically. Given that $PV_0$ is known, and it is possible to simulate future cash flows, true or risk-neutral distribution of the cash flows in future can be simulated. As well as discounting all the cash flows to the present value, they can also be undiscounted to their future value. Because the present value of cash flows is known ($PV_0$) as well as its simulated undiscounted future value with standard deviation, it is possible to find the volatility parameter analytically without any unnecessary additional computations and simulations. It is known that in a risk neutral world, the asset value increases with time according to the Equation (5), and that the standard deviation of the process increases according to the Equation (6).

$$ S_t = S_0 e^{rt} \quad (5) $$

$$ std(S) = S_0 e^{rt} \sqrt{e^{\sigma^2 t} - 1} = S_0 e^{rt} \sqrt{e^{2\sum_{i=1}^{t} \sigma_i^2} - 1} \quad (6) $$
Therefore, if the standard deviations of the underlying asset process at certain time points are known, it is possible to compute the average volatility for each time period. Starting from the beginning of the process, each $\sigma_i$ can be calculated according to Equation (7) as

$$\sigma_i = \sqrt{\frac{\ln \left( \frac{\text{std}(S)^2_i}{S_0 e^{\mu T}} \right) + 1}{t_i} - \sum_{i=0}^{l-1} \sigma_i^2 t_i}$$

(7)

The information which is required is the length of time period $T$ for volatility estimation, value of the asset $S_0$ in the beginning, interest rate $\mu$, and $\text{St.Dev}_e$ standard deviation of the asset value at the end of volatility estimation period. Interest rate $\mu$ does not change the volatility estimation as long as the same interest rate is used both in cash flow simulation and when computing the volatility. Therefore, even if the risky expected return is used in volatility estimation, the option valuation still follows risk-neutral pricing with risk-free interest rate used.

**Least squares regression method with risk-neutral approach of Haahtela (2010)**

Haahtela (2010) further extends the idea of Godinho (2006) and properties of generalized risk-neutral valuation for estimating changing volatilities for different time periods\(^1\). The approach uses ordinary least squares regression for estimating $PV_t$ with cash flow simulation state variables as explaining input parameters. However, instead of the common approach of directly estimating volatility as the standard deviation of the rate of return as $z = \ln(S_{t+1}/S_t)$, $PV$ regression or response surface estimators are used directly for estimating both the underlying asset value and its standard deviations at different time points. Figure 1 illustrates this. Based on this information, volatilities are estimated for different time periods. After the cash flow simulation and constructing the regression equations, the actual procedure for the standard deviation calculation is presented in Haahtela (2010) as follows:

1) Calculate the estimated expected value for $\hat{PV}_t$ from the cash flow parameters $X_{i,k}$ generated in the simulation trial run using the regression equation.
2) Calculate the differences between values predicted by the regression estimator and the values actually observed from the realized simulations.
3) a) Square the differences,
   b) add them together,
   c) divide them by the number of observation, and
   d) take the square root.

---

\(^1\) Volatility estimation was not the main topic of the paper, it was only presented as an example of that it is at least somehow possible to estimate changing volatilities based on simulated cash flows, the idea that was presented to support the main topic of the paper.
After that, volatility for different time periods is calculated period by period according to Equation (7). Yet the method provides sound volatility estimation, it is not very usable or comfortable in practice. If there are changes in any cash flow simulation parameters, the whole volatility estimation procedure has to be repeated again for all the time periods.

3 Resolving of the uncertainty during investment project

Before going to the actual regression solution procedure, it is good to illustrate the idea behind the calculation. Usually, uncertainty reveals over time and we have better understanding and knowledge about the situation and the investment viability during later stages of the project. Also, the proportional uncertainty is likely to decline over time, especially in case of R&D and learning. Therefore, volatility is often time changing and declining. Finally, all the uncertainty has revealed and there is no uncertainty and volatility left anymore.
Figure 2: Unsolved uncertainty and solved uncertainty during investment. Usually uncertainty resolves faster in the early stages of the project.

Figure 2 illustrates this situation, and the same is demonstrated in the forthcoming example. In the beginning, we estimate the uncertainty of the investment project (in terms of variance) to be 12,000. During each time period of the process, more and more of the uncertainty is revealed. At the end, all the uncertainty has been revealed. Now, if we were able to construct good forward-looking value estimators $PV_1, ..., PV_n$ for each time period, we could estimate the overall uncertainty solved (and unsolved) for each time point, and this information can then be used to calculate the volatility for each time period according to Equation (7).

For simplicity, we assume in this example use of arithmetic Brownian motion and the interest rate is assumed to be zero. Arithmetic Brownian motion is normally distributed with mean $\mu$ and variance $\sigma^2$. Sum of normally distributed variables is

$$
\sum N(m_i, \sigma_i^2) = N(\mu_1, \sigma_1^2) + \ldots + N(\mu_n, \sigma_n^2) = N(\mu_1 + \ldots + \mu_n, \sigma_1^2 + \ldots + \sigma_n^2)
$$

(8)

The standard deviation of the arithmetic Brownian motion is:

$$
Std(S_t) = S_0 e^{rt} \sqrt{\sum \sigma_i^2 t_i}
$$

(9)

The example has only four annual cash flows, each with expected value of 100. These cash flows are all normally distributed with expected value of 100 and the standard deviation for time periods 1…4 are 80, 60, 40, and 20. Therefore, based on the properties of normal distribution, the sum of these cash flows is 400 and the variance is 12,000, which is the same as our present value of cash flows because the interest rate is assumed to be zero.
Table 1: Example of how uncertainty reveals over time

<table>
<thead>
<tr>
<th>Time period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash flow</td>
<td>N(100; 80²)</td>
<td>N(100; 60²)</td>
<td>N(100; 40²)</td>
<td>N(100; 20²)</td>
</tr>
<tr>
<td>Variance (variation, uncertainty) of the time period</td>
<td>6 400</td>
<td>3 600</td>
<td>1 600</td>
<td>400</td>
</tr>
<tr>
<td>Solved (explained) uncertainty after time period</td>
<td>6 400</td>
<td>10 000</td>
<td>11 600</td>
<td>12 000</td>
</tr>
<tr>
<td>Unsolved uncertainty after time period</td>
<td>5 600</td>
<td>2 000</td>
<td>400</td>
<td>0</td>
</tr>
<tr>
<td>Proportion of total variance of 12 000</td>
<td>0,533</td>
<td>0,300</td>
<td>0,133</td>
<td>0,033</td>
</tr>
<tr>
<td>Cumulative proportion of solved uncertainty</td>
<td>0,533</td>
<td>0,833</td>
<td>0,967</td>
<td>1,000</td>
</tr>
<tr>
<td>Proportion of unsolved uncertainty</td>
<td>0,467</td>
<td>0,167</td>
<td>0,033</td>
<td>0,000</td>
</tr>
<tr>
<td>Volatility of the time period</td>
<td>20% (80/400)</td>
<td>15% (60/400)</td>
<td>10% (40/400)</td>
<td>5% (20/400)</td>
</tr>
</tbody>
</table>

If we consider this cash flow calculation period by period, we can recognize how the variance uncertainty reveals over time. Table (1) shows this. After we know the realization of the first cash flow, N(100; 80²), actually 6 400 (or 53,3 %) of the overall variance and uncertainty of 12 000 has resolved, leaving 5 600 of the uncertainty unsolved. During the second time period, 3 600 of the total variance is solved, and thus the cumulative portion of solved uncertainty is 10 000, leaving 2 000 left as unsolved uncertainty. After the third time period, 1 600 more is solved, and in the final stage, the remaining 400 is resolved. As a result, amount of solved uncertainty increases over time until all the uncertainty is solved, while the amount of unsolved uncertainty diminishes. This is precisely the same as demonstrated in Figure 2.

When the variance and standard deviation change between time periods are known, this information can be used for the volatility estimation. Starting from the last time period 4, and using Equation (9) with the given parameters Std(S₄) = 20, S₀ = 400, r = 0 %, and t₄ = 1, we get σ₄ = 0.05. This means that standard deviation is 5 % of the expected value. This is also quite self-evident, given we already knew that standard deviation (20) is 5% of the expected value (400). Then, having this information available, we can step backwards into time period 3 and knowing also Std(S₃) calculate the volatility σ₃. This is 0.10 which is also consistent with standard deviation of 40 being 10 % of mean value 400. With the same backward logic, respectively, we can calculate σ₂ = 0.15 and σ₁ = 0.20.

4 Regression based volatility estimation

As mentioned in the example of previous section, if we are able to construct good forward-looking value estimators PV₁…PVₙ with the information of their expected values and standard deviations for each time period, we can estimate the overall uncertainty solved (and unsolved)
for each time point, and this information could then be used to volatility estimation. One alternative to do this is to use ordinary least squares regression equations.

In statistics and econometrics, ordinary least squares (OLS) or linear least squares is a method for estimating the unknown parameters in a linear regression model. The method is a standard approach to the approximate solution of overdetermined systems, i.e. sets of equations in which there are more equations than unknowns. The most important application is in data fitting. OLS method minimizes the sum of squared distances, also called residuals or squares of errors, between the observed responses in the dataset, and the responses predicted by the linear approximation model. The resulting estimator can be expressed by a simple formula, especially in the case of a single regressor on the right-hand side.

Regression analysis includes any techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. Regression analysis helps us understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables — that is, the average value of the dependent variable when the independent variables are held fixed. In all cases, the estimation target is a function of the independent variables called the regression function. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function, which can be described by a probability distribution.

Regression divides the sum of squares (SS) in Y, total variation, into two parts: the sum of squares predicted due regression (SSDR) and the sum of squares error (SSE), so that $SS = SSDR + SSE$. SSDR is the sum of the squared deviations of the predicted scores from the mean predicted, and the SSE is the sum of the squared errors of prediction. SSDR therefore describes the variation that regression model can explain and SSE is the unexplained variation. Square root of the SSE is standard error.

A common measure of the regression model is a coefficient of determination, $R^2$. It is the proportion (percentage) of the sum of squares explained due regression divided by the total sum of squares, i.e. $R^2 = SSDR/SS$. $R^2$ is the same as Pearson’s correlation measure between the predicted and observed values. As a result, if we know $R^2$ and standard error, we can also calculate the explained variation, the square root of SSDR, according to:

$$Excluded variation = \sqrt{\frac{Std. Error^2}{1 - R^2}} - Std. Error^2$$  \hspace{1cm} (10)
As a result, we can use regression equations to estimate the expected value and the standard deviation of PV, time period by time period, by linking the explaining cash flow variables to the explanatory estimators of PV. These explaining variables can be any parameters in the cash flow calculation. Often sufficient explanatory variables for estimating PV are free cash flows of the given and the previous years\(^2\), i.e. FCF\(_0\),…,FCF\(_t\). As a result, we can calculate PV\(_0\) according to Equation (11). This increases for the following time periods with the risk-free interest rate according to Equation (12).

\[
\hat{PV}_0 = \sum_{t=1}^{T} \frac{CF_t}{(1 + r)^t}
\]

\[
PV_T = PV_0e^{rT}
\]

Then, we can construct the regression estimators as suggested for each time period according to the following equations:

\[
\hat{PV}_1 = \alpha_1 + \beta_{1,1}FCF_1
\]

\[
\hat{PV}_2 = \alpha_2 + \beta_{1,2}FCF_1 + \beta_{2,2}FCF_2
\]

\[
\vdots
\]

\[
\hat{PV}_t = \alpha_t + \beta_{1,t}FCF_1 + \beta_{2,t}FCF_2 + \cdots + \beta_{t,t}FCF_t
\]

Now we can use Monte Carlo simulation on our example cash flow model. Before the cash flow simulation run, we set the free cash flows of the time periods (FCF\(_t\)) and the present values of the project (PV\(_t\)) as output parameters. These parameter values are then saved during the simulation run so that they can be used in regression calculations after the simulation. Then we use regression Equations (13) for forecasting the expected values and the standard errors of PVs’.

There are many different alternatives to conduct the simple OLS regression analysis. Even the basic tools provided with the common spreadsheet programs (e.g. MS Excel or OpenOffice Calc) are sufficient for this purpose\(^3\). As an example of this, the following Table 2 is taken from the Excel summary output regression statistics report for the PV’s of the previous example. As we can see from the results in Table 2, numbers of R Square are very close to those of Cumulative proportion of solved uncertainty in Table 1.

---

\(^2\) Haahrtaela (2010) and Haahrtaela (2011) have more details about constructing good regression estimators.

\(^3\) In practice, majority of the potential users would rather use other software (e.g. SPSS, SAS, R, Statistix, Matlab) for the data analysis, because more advanced software is required anyway for many other common tasks related to data analysis and other financial analysis.
Table 2: Regression statistics of the example.

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th>year 1</th>
<th>year 2</th>
<th>year 3</th>
<th>year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.730</td>
<td>0.912</td>
<td>0.983</td>
<td>1.000</td>
</tr>
<tr>
<td>R Square</td>
<td>0.532</td>
<td>0.831</td>
<td>0.966</td>
<td>1.000</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.532</td>
<td>0.831</td>
<td>0.966</td>
<td>1.000</td>
</tr>
<tr>
<td>Standard Error</td>
<td>74.48</td>
<td>44.78</td>
<td>20.00</td>
<td>0.000</td>
</tr>
<tr>
<td>Observations</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
</tr>
</tbody>
</table>

The only information needed of the summary output regression statistics are the correlation coefficient R Square and Standard Error that are used in the calculation of the explained variation according to Equation (10). These results are presented in Table 3. The explained variation is indeed the standard deviation of the underlying asset process at the particular time point, i.e. how much of the uncertainty is revealed until that time point from the earlier time points. With this information, we can calculate the volatilities for the different time periods according to the Equation (7). Also SSDR, square of explained variation, and SSE, square of standard error, are presented in the table for comparing the simulated results with the calculated results presented earlier in Table 1.

Table 3: Explained variation and the volatilities for different time periods.

<table>
<thead>
<tr>
<th>Explained variation</th>
<th>79.47</th>
<th>99.30</th>
<th>107.09</th>
<th>(108,9)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSDR</td>
<td>6315</td>
<td>9860</td>
<td>11467</td>
<td>(11860)*</td>
</tr>
<tr>
<td>SSE</td>
<td>5547</td>
<td>2005</td>
<td>400</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Volatility</th>
<th>19.87 %</th>
<th>14.89 %</th>
<th>10.02 %</th>
<th>4.96 %</th>
</tr>
</thead>
</table>

* observed directly from the simulation results

As we can see from the numbers in Table 3, results of the SSDR are very similar to the results of Solved (explained) uncertainty, and the results of SSE are close to the numbers of Unsolved uncertainty in Table 1. If more simulation runs are used, the regression based results converge gradually to the accurate results of Table 1. The results thus also show that the logic and intuition behind the procedure is in line with the idea of resolving uncertainty during the project timespan presented in Section 3.

In the following section the use of the regression based volatility method is illustrated with actual case example. We can’t calculate directly comparative results such as presented in Table 1 because of the serial- and cross-correlations among the cash flow input parameters, and therefore modeling the forthcoming cash flow parameters conditional on earlier realizations of several
input parameters is in practice impossible. Suggested regression based method is not sensitive to that. It can also be adjusted for stochastic interest rate. Another advantage of the approach is that it can be used with different stochastic processes (e.g. with arithmetic, geometric, or displaced diffusion process).

5 Case example of method use

The following example illustrates the use of the approach. The input parameters (unit sales price, sales quantity, and variable unit costs) are presented as normal distributions, and they have also correlations with each other. For example, if the sales price is low in 2011, it is also likely lower also in the forthcoming years. Also fixed costs have some normal variation. These uncertain input parameters of the cash flow calculation are presented with gray background color. As such, this is a typical simplified cash flow calculation that has several partly correlated parameters and some uncertainty.

Table 4: Cash flow calculation of the case example.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit sales price</td>
<td>2,00</td>
<td>1,80</td>
<td>1,62</td>
<td>1,46</td>
<td>1,31</td>
<td>1,18</td>
<td></td>
</tr>
<tr>
<td>Sales quantity</td>
<td>25 000</td>
<td>37 500</td>
<td>37 500</td>
<td>28 125</td>
<td>21 094</td>
<td>15 820</td>
<td></td>
</tr>
<tr>
<td>Variable unit costs</td>
<td>1,70</td>
<td>1,36</td>
<td>1,16</td>
<td>0,98</td>
<td>0,88</td>
<td>0,80</td>
<td></td>
</tr>
</tbody>
</table>

**Cash flow calculation**

| Revenue | 50 000 | 67 500 | 60 750 | 41 006 | 27 679 | 18 683 |
| Variable costs | -42 500 | -51 000 | -43 350 | -27 636 | -18 654 | -12 591 |
| Fixed costs | 5 000 | 5 000 | 5 000 | 5 000 | 5 000 | 5 000 |
| Depreciation | -833 | -833 | -833 | -833 | -833 | -833 |
| EBIT | 11 667 | 20 667 | 21 567 | 17 538 | 13 192 | 10 259 |
| Taxes | -4 667 | -8 267 | -8 627 | -7 015 | -5 277 | -4 104 |
| Depreciation | 833 | 833 | 833 | 833 | 833 | 833 |
| Free cash flows (FCFt) | 7 833 | 13 233 | 13 773 | 11 356 | 8 748 | 6 988 |
| Discounted cash flows (11%) | 7 057 | 10 740 | 10 071 | 7 480 | 5 192 | 3 736 |
| Present value | 44 276 |

The ordinary cash flow calculation model requires one additional row that describes how the expected present value of the discounted cash flows PV0 increases over time according to the risk-free interest rate. Before the cash flow simulation, the free cash flows of the time periods (FCFt) and the present values of the project (PVt) are set as output parameters. These parameter
values are saved during the simulation run so that they can be later used in regression calculations.

After the simulation, we can investigate the overall uncertainty of the project. The following Figure 3 shows the simulated underlying asset terminal value distribution discounted to the present value. We know that the underlying asset expected value in a risk-neutral world increases according to the risk-free interest rate, but we do not know how the uncertainty changes over the time periods. Based on the assumption of generalized risk-neutral valuation and assuming the stochastic process to follow geometric Brownian motion⁴, we can calculate according to the Equation (7) the annualized volatility to be 18.4 %⁵. However, we want to know how the uncertainty and thus volatility evolves over time, and therefore we use the regression based volatility estimation presented earlier on.

![Distribution for Present Value](image)

Mean  44 321  
St.Dev  20 981  

Shifted lognormal distribution fitting parameters  
Mean  97 877  
Shift  -53 556  
St.Dev.  20 981  

**Figure 3:** Probability distribution of the case example present value PVₜ. Also basic statistics of mean and standard deviation are presented as well as the parameterization for the shifted lognormal distribution.

Then we use the regression approach similarly to the approach presented in section 4. The regression estimator Equations (13) were used with each PVₜ as explanatory variable and the free cash flows of the corresponding and the previous years as the explaining variables for each case.

After that, the regression analysis is done, and we get the following regression analysis summary output results presented in Table 5. Based on the R Square (proportion of uncertainty solved) and Standard Error (uncertainty left), we calculate the explained variation for each time period according to Equation (10). After that, volatility for each time period is calculated according to the Equation (7). The method and the calculation of the results were confirmed by using the approach of Haahrta (2010) for the volatility estimation. The results given by both methods are

⁴ Later, this assumption of gBm is also omitted and displaced diffusion process is assumed instead  
⁵ Given S₀ = 44 321, t = 6, Std(S₀) = 20 981, and rᵣ = 5%
exactly the same when calculated using the same simulated data set. However, use of the approach presented in this paper is much easier and faster to apply for a practitioner.

However, as the distribution of the present value (Figure 3) shows, the shape of it is between a normal and a lognormal distribution, and there are also negative underlying asset values. Therefore, common assumption of the geometric Brownian motion does not hold. One viable alternative is use of displaced diffusion process suggested by Rubinstein (1983) that has an underlying asset terminal value distribution described by a three parameter (mean, deviation, locus) shifted lognormal distribution. Displaced diffusion process with different values of \( S_\theta \) (mean) and \( \theta \) (shift) is capable to model stochastic processes that are between multiplicative (lognormal) and additive (arithmetic) processes, and it also allows negative underlying asset values. The relationship between volatility \( \sigma \) and shifted volatility \( \sigma_{dd} \) is presented in equation 14. The results of shifted volatility are also presented in Table 5.

\[
S\sigma = (S_\theta + \theta)\sigma_{dd}
\]  

(14)

Table 5: Summary output of the regression analysis and the calculation of explained variation, volatility and shifted volatility.

<table>
<thead>
<tr>
<th>SUMMARY OUTPUT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regression Statistics</strong></td>
</tr>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>Explained variation</td>
</tr>
<tr>
<td>Volatility ( \sigma )</td>
</tr>
<tr>
<td>Shifted volatility ( \sigma_{dd} )</td>
</tr>
</tbody>
</table>

6 Discussion and conclusions

This paper presented a practical volatility estimation method for cash flow simulation based real option valuation cases with changing volatility. During cash flow simulation runs, present value of the cash flows and the corresponding cash flow state variable values are recorded for each time period. Then a regression analysis is run for relating the PV for each year to the cash flow
state variables. Each regression equation provides an estimate of the expected present value as a function conditioned on the resolution of all uncertainties up to that time. Then, basic regression statistics of Pearson’s correlation $R^2$ and sum of squares error (or standard error) for each equation provide all the information required for estimating how the standard deviation of the stochastic process and the volatility change over time.

The method is computationally very efficient as it requires only one pass of simulation runs regardless of the number of time periods. Straightforward calculation of required regression statistics and their availability in any statistical software and even in spreadsheet programs make this approach very easy to apply for a practitioner. Also the intuitive logic behind the procedure and the capability to handle negative underlying asset values can be regarded strengths of the method. As such the approach is sufficiently robust for a cash flow simulation based volatility estimation.

The method presented in this paper is already quite good in practical terms. It takes into account the stochastic time-dependence of volatility and the level of underlying asset price level. The latter of these means that the model is capable of taking into account the common characteristic of the underlying asset to have lower volatility with higher underlying asset values. The next logical step for improving the accuracy of volatility estimation is to use a complete state-space dependent volatility structure. This is very close to the logic of applying implied volatility term structure method were simulated cash flows determine the underlying asset price state-space. However, this approach requires more computation and forward-looking estimation and is not that easy to implement in practice in comparison with the method suggested in this paper.

One alternative between the two previously mentioned alternatives (time-varying volatility for displaced diffusion process and implied volatility term structure approach) is to model the change in volatility as a function of several underlying parameters. One alternative to determining these functions would be to use statistical software on simulated data to investigate what kind of response surface functions could be feasible. However, finding a good function (or a set of functions for different time periods) for describing state-dependent volatility may actually be somewhat difficult. On the other hand, this approach allows direct linking of the cash flow calculation parameters to the parameters describing underlying asset value process. This may be a useful feature for sensitivity analysis purposes if someone wants to investigate how changes in cash flow calculation parameters affect the project value with real options.

However, similarly to the complete state-space implied volatility approach, this approach is not either a significant improvement over the approach suggested in this paper. This is mostly because the cash flow simulation follows law of large numbers. According to central limit theorem, the sum of a sufficiently large number of independent random variables of fairly finite mean and variance, is approximately normally distributed. If the process is multiplicative, i.e. the
random variables are multiplied, the result is lognormally distributed. Typical cash flow
calculation has both of these properties i.e. they are sums of several cash flow variables, with
some of these variables having multiplicative correlated properties over the time periods.
Therefore, when we sum up a large number of cash flows over multiple simulation runs, the
terminal value distribution is often very close to the shape of the displaced lognormal
distribution. Therefore, the method presented in this paper, especially when applied to displaced
diffusion process, is sufficiently accurate volatility estimation method, given the realities of

Completely another question is how much we are actually interested in to know the volatility
changes overtime. The consolidated cash flow simulation based volatility measure is not an
observable market variable that could be used for hedging purposes. As such, volatility
parameter may not be even needed in valuation. If the approach based on simulated implied
value distribution approach is used with an implied binomial lattice, calculating the volatility
parameter for each node does not provide significant advantages over the approach of directly
estimating the risk-neutral implied probabilities for the transitions7. For example Wang and Dyer
(2010) use this approach without calculating volatility parameter for each node. Another
advantage of this approach is that it offers very flexible distribution assumption for project
values. The shortcoming of this approach is that it is more complex to apply in comparison with
the method of this paper. On the other hand, such approach is better for such cases were the
underlying asset is clearly tracktable, which is the case for many projects related to natural
resource investments. If this assumption does not hold, the approach presented here gives similar
results with less effort.

However, even if the procedure presented here may technically be feasible, it still has many non-
computational drawbacks common to the similar cash flow simulation based methods. Firstly,
we are consolidating a high-dimensional simulated process into a very low-dimensional process.
As also discussed in Brandão et al. (2005) and Smith (2005), the applicability of such approaches
is usually case dependent. Secondly, the high-dimensional process is based on a simulated cash
flow, not on a real observable process. As such, it may have highly subjective estimates,
especially when estimating the correlations between different cash flow variables. Thirdly, the
whole valuation may appear as a black box for the decision-maker after the consolidation,
because it is harder to monitor and understand how different parameters and their changes affect
the consolidated process. This may significantly reduce the capability to exercise the real options

6 This does not mean that the reality would follow this kind of pattern. It is just a consequence of how sampling
based methods work.
7 If we know the risk-neutral transition probabilities between the nodes and assume some stochastic process (most
often geometric Brownian motion), we can easily calculate the implied volatilities for each node.
optimally. Haahtela (2011) mitigates this problem by linking the primary cash flow calculation parameters to the option valuation parameters with regression equations.

There are several computational issues that could be done to improve technically the approach presented here. However, both from the academic and practitioners’ viewpoint it would be far more interesting and relevant to investigate, based on several real life cases form different industries, how and when the consolidated approach based methods are of practical use in comparison with some other valuation methods.

7 References


