Simulation-based estimation of state-dependent project volatility

Pedro Godinho[a]

[a]CeBER and Faculty of Economics of the University of Coimbra, Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal. Phone: +351+239790571. Email: pgodinho@fe.uc.pt

Abstract

Project volatility is an essential parameter for real options analysis, and it may also be useful for risk analysis. Many volatility estimation procedures only consider the volatility in the first year of the project. Others consider that different years may have different values of the project volatility. This paper takes into account that volatility may change not only with time but also with the state of the project. Two possible definitions for the project volatility are considered, the log-variance and the variance of the project value, and two simulation-based procedures are proposed for estimating state-dependent volatility: two-level simulation and one and a half level simulation. Computational experiments show that both the procedures perform better than the method proposed by Copeland and Antikarov, and that the one and a half level simulation procedure leads to the most accurate estimations of project volatility.

Keywords: Simulation; Project volatility; Real options

This is an Accepted Manuscript of an article published by Taylor & Francis in The Engineering Economist on 25 October 2017 (online publication), available online: http://www.tandfonline.com/10.1080/0013791X.2017.1384523 .

1. Introduction

Despite having enjoyed some significant success in academe, the adoption of real options analysis by firms has been quite slow (e.g., Triantis, 2005, Block, 2007, Baker *et al.*, 2011). Since real options models have been around for some decades, and they allow practitioners to avoid some severe drawbacks of the classical discounted cash flows methodologies (see, e.g., Trigeorgis, 1993), such a slow rate of adoption may seem surprising. However, it may be explained by the mathematical sophistication that is required to understand and use most real options models (Baker *et al.*, 2011).

Mun (2002, p. 142) points out that it is advantageous to use lattice approaches in management discussions, since they are intuitive and easy to understand. Brandão *et al.* (2005, p. 85) point out that binomial decision trees may have even more intuitive appeal. The quality of results obtained with lattice or decision tree approaches depends, among other things, on an accurate modeling of underlying asset volatility. If there is only one significant source of project risk and that source of risk is a traded asset, then such an asset may be used as the single underlying asset of the project, and market data may be used for estimating its volatility (e.g., Smit, 1997). However, most real life projects have several sources of risk, and some of them are not traded. If many risk sources are simultaneously modeled, the problem becomes multi-dimensional, and the use of lattices or decision trees may become impractical. Copeland and Antikarov (2001) propose using the project without options as the underlying asset of the analysis in a lattice-based approach. Such an approach may lead to some sub-optimal decisions (if the optimal decisions would depend on the values of all state variables), but it allows the problem to remain single-dimensional. This is, therefore, a useful approach, since it provides a simple way of incorporating risk and managerial flexibility in the analysis, allowing a more accurate valuation than traditional discounted cash flows methodologies.

Copeland and Antikarov's (2001) approach requires an estimate of the project volatility, as is also the case with other real options models. The authors propose a procedure for estimating project volatility based on Monte Carlo simulation. In fact, Monte Carlo simulation models have been widely used in project valuation, since they provide an accessible way to incorporate the impacts of random events and random variables in project value, as well as complex relations between variables that often occur in investment projects. Such issues, which might become intractable by other methods, are easily handled by Monte Carlo simulation, and transparently translated into statistics concerning the project outcomes, like the project value or the volatility.

The volatility estimation procedure proposed by Copeland and Antikarov (2001) has since been shown to grossly over-estimate the true project volatility (Brandão *et al.*, 2012, Godinho, 2006, Smith, 2005). Other authors have also addressed the estimation of project volatility, either using Monte Carlo-based procedures (e.g., Brandão *et al.*, 2012, Fleten *et al.*, 2010, Godinho, 2006, Haahtela, 2011, Herath and Park, 2002, Ruhrmann *et al.*, 2014) or performing analytical calculations based on a pre-defined model of the project (e.g., Costa Lima and Suslick, 2006a & 2006b, Davis, 1998). Nicholls *et al.* (2014) compare several methods for estimating volatility, emphasizing the difficulty of finding the best estimates. Given these difficulties, the authors suggest using the breakeven volatility, the minimum volatility that is necessary in order to justify the project, as a guidance for the decisions concerning the project.

Analytical calculation of project volatility provides accurate results, but it is only applicable when the project fits the pre-defined underlying model (Godinho, 2006). Many real life projects have complex cash flow structures, which are not amenable to the analytical calculation of volatility. The analytical estimation of volatility may also be impractical for the development of software tools aimed at supporting capital investment decisions. Monte-Carlo based procedures, on the other hand, are more flexible, providing ways of estimating volatility whenever it is possible to build a Monte Carlo model of the project.

Most procedures for volatility estimation only consider the first year volatility. However, the results obtained by other authors (e.g., Brandão *et al.*, 2012) empirically show that volatility changes over time in many cases. So, a complete real options analysis must take into account the operational flexibility in the remaining years, and the way it affects the first year decisions concerning the project. Davis (1998) estimates the volatility in different years. Since the author uses an analytical procedure, it can only be applied to projects fitting his pre-defined model. Brandão *et al.* (2012) propose procedures for estimating volatility in any year t. The main procedure proposed by the authors assumes that the analyst is able to write analytical expressions for the expected value of the cash flows occurring after t, conditioned on information available at time t. A second procedure proposed by the authors (Brandão *et al.*, 2012, p. 647) relies instead on the ability to identify a set of variables that completely determine the project value such that the expected value of each variable at each time point can be reasonably estimated using just the previous value of that same variable. Haahtela (2011) proposes a procedure for estimating the volatility in different years, based on the residual sum of squares of a linear regression. The same author (Haahtela, 2007, 2012) considers the separation between volatility and ambiguity (uncertainty for which no precise data is available) in the context or simulation-based estimation methods.

Volatility may change not only with time but also with the state of the project. In the context of option valuation, several authors have considered different models that assume that volatility may be state-dependent. Examples of such models are the constant elasticity of variance model (see, e.g., Cox and Ross, 1976, and Beckers, 1980; Sbuelz, 2005, considers such a model in the context of real options) and models of the AutoRegressive Conditional Heteroskedasticity (ARCH) type (see, e.g., Kallsen and Taqqu, 1998).

This paper recognizes that volatility may change with the state of the project, the state of the project being defined as the vector composed by the random variables on which the cash flows depend, like demand, prices of market-traded inputs and outputs, stochastic fixed costs, etc. This means that, in each year after starting the project, the best estimate of the future project volatility depends on time, and also on the remaining underlying stochastic variables. Failing to recognize the state dependence of volatility in real options analysis will lead to sub-optimal decisions, as well as an inaccurate valuation of the project.

This paper discusses the estimation of volatility for any given project state, based on Monte Carlo models of the project. As far as the author is aware, there are no previous works on the estimation of state-dependent project volatility based on a simulation model of the project – the others works mentioned before, which consider state dependent volatility, assume a specific form for such volatility (constant elasticity of variance, ARCH type, etc.) This work analyses how volatility may be estimated by Monte Carlo simulation when no such specific form is assumed.

The main contribution of this paper is to propose two procedures for estimating project volatility in a given state. These procedures are based on other methods proposed in the literature, sometimes in different contexts. The paper also presents computational tests to compare the performance of the proposed procedures in a small number of projects.

Real options analysis is used as the main motivation for volatility estimation, but it must also be stressed that accurate estimation of state-dependent volatility may also be useful for several other purposes. For example, accurate volatility estimation may be useful for some kinds of risk analysis (e.g., calculation of value-at-risk or conditional value-at-risk), or even for defining an appropriate discount rate for projects undertaken by some private companies (particularly companies with few owners with undiversified portfolios and restrictions in the access to credit markets).

The paper is structured as follows. After this introduction, Section 2 defines the concepts of volatility that are used in the paper. Section 3 defines some simple projects for which it is possible to analytically estimate the volatility. Section 4 presents the procedures that can be used to estimate volatility in given project states, as well as some computational tests for comparing their performance. Finally, the conclusions are presented in Section 5. The Appendixes contains a mathematical result concerning one of the volatility estimation procedures and examples of the implementation of the procedures in Microsoft Excel, using the @Risk add-in.

2. Concepts of project volatility

Project volatility is a measure of the uncertainty over expected project returns. Volatility is usually defined as the variance, or standard deviation, of the project value (or of the logarithm of the project value). In this paper, variance is used as the measure for volatility (of course, if the standard deviation is necessary, it is only necessary to calculate the square root of the variance). This variance corresponds to changes in the expected project value that take place in a given period (usually in a given year). These changes may be due to the impact of the events occurring in that period both on that period's cash flows and on the expected value of subsequent cash flows. The problem with some simulation-based volatility estimation procedures (e.g., the ones proposed by Copeland and Antikarov, 2001 and Herath and Park, 2002) is that instead of capturing the impact of the events occurring in that period on the expected value of subsequent cash flows, they also incorporate changes in subsequent cash flows due to events occurring in posterior periods (see, e.g., Godinho, 2006). For the purpose of accurate volatility estimation these latter changes amount to random variability, and should not be considered as part of the variance of project value in the considered period.

Several real options models assume that project value follows a geometric Brownian motion, and in this case the variance of the logarithm of project value is a convenient measure of volatility. This measure is hereafter termed the log-variance, and it is presented in Subsection 2.1.

In spite of being widely used, the log-variance has a very important drawback: it can only be used if the project value does not become negative. So, another measure is considered, which can be applied to all projects: the simple variance of project value. This measure has not been as widely used as the log-variance, but it may be very convenient, for example to approximate the local project behavior by an Arithmetic Brownian Motion (ABM). This measure is simply termed variance, and it is discussed in Subsection 2.2.

2.1. The log-variance

Let us consider an investment project with a known initial investment outlay F_0 and future uncertain cash flows F_t, $t = 1,\ldots,N$, and a continuously compounded discount rate r. The expected value of the project at time t, PV_t, is defined as the expected value of the cash flows that will occur after time t, discounted to time t. The value of the project is conditioned by the values, at time t, of the random variables on which the cash flows depend, like demand, prices of market-traded inputs and outputs, stochastic fixed costs, etc. The vector with these random variables will be termed the state of the project, and denoted by $\boldsymbol{\psi}_t$. The project value may be written as:[1]

$$PV_t = \sum_{s=t+1}^{N} E\left(F_s \mid \boldsymbol{\psi}_t\right) \cdot e^{-r(s-t)}, \tag{1}$$

where $E(\cdot)$ denotes the expected value. The net value of the project at time t, NPV_t, is defined as the sum of the project value at time t with the time-t cash flow. By noticing that $F_t = E\left(F_t \mid \boldsymbol{\psi}_t\right)$, we may write:

$$NPV_t = PV_t + F_t = \sum_{s=t}^{N} E\left(F_s \mid \boldsymbol{\psi}_t\right) \cdot e^{-r(s-t)} \tag{2}$$

Let k_t be a random variable that represents the continuously compounded rate of return on the project between time t-1 and time t. Then:

$$NPV_t = PV_{t-1} \cdot e^{k_t} \tag{3}$$

From expression (3) it follows that:

$$k_t = \ln\left(\frac{NPV_t}{PV_{t-1}}\right) = \ln\left(NPV_t\right) - \ln\left(PV_{t-1}\right) \tag{4}$$

In order to measure the uncertainty from time t-1 to time t, it must be acknowledged that $\boldsymbol{\psi}_{t-1}$ will be known at time t-1. Therefore, the relevant variance, which is denoted as $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$, must be conditioned on $\boldsymbol{\psi}_{t-1}$:

[1] In fact, PV_t is a function of $\boldsymbol{\psi}_t$. In order to avoid notational clutter, this dependence is not represented explicitly. The same will be done for NPV_t.

$$LVar_t\left(\boldsymbol{\psi}_{t-1}\right) = \operatorname{var}\left(k_t \mid \boldsymbol{\psi}_{t-1}\right) = \operatorname{var}\left(\ln\left(NPV_t\right) \mid \boldsymbol{\psi}_{t-1}\right), \tag{5}$$

where var(·) is the variance and the second equality takes into account that PV_{t-1} is completely determined by $\boldsymbol{\psi}_{t-1}$. If a single value for the variance of the project value from *t*-1 to *t* was to be estimated, like it is done by Brandão *et al.* (2012) or Haahtela (2011), it would just be necessary to calculate the expected value of $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$. This expected volatility is denoted as $LVar_t^e$:

$$LVar_t^e = E\left(LVar_t\left(\boldsymbol{\psi}_{t-1}\right)\right) = E\left(\operatorname{var}\left(\ln\left(NPV_t\right) \mid \boldsymbol{\psi}_{t-1}\right)\right). \tag{6}$$

By applying the law of total variance, it may be seen that if the variance of k_t is not conditioned on $\boldsymbol{\psi}_{t-1}$, an upward bound on $LVar_t^e$ will be obtained.

It may be questioned whether $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$ or $LVar_t^e$ should be used when building a project value model. Let us consider real options models. A real options model usually intends to determine the optimal strategy (or at least a nearly optimal strategy) for managing the project, and to calculate the project value in case that strategy is followed. If the state variables can be observed, or if there is some observable information that allows conditioning the distribution of the state variables, then the use of $LVar_t^e$ will result in a distortion of project behavior (since, when time *t*-1 is reached and state $\boldsymbol{\psi}_{t-1}$ is observed, the best estimate of its variance is $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$). To illustrate this, consider the decisions made at time *t*-1 in a project with an abandonment option that allows selling the project for a value V_{Ab}. Additionally assume that in the states $\boldsymbol{\psi}_{t-1}$ for which $PV_{t-1} < V_{Ab}$, we have $LVar_t\left(\boldsymbol{\psi}_{t-1}\right) < LVar_t^e$. If a project model uses $LVar_t^e$ instead of $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$, the variance of project value considered at time *t*-1 will be larger than the true variance whenever $PV_{t-1} < V_{Ab}$. So, if the project reaches time *t*-1 with $PV_{t-1} < V_{Ab}$, the probability of the project value rising above V_{Ab} again is over-estimated, therefore leading to an abandonment threshold that will be smaller than the true optimal threshold.

Therefore, if the variance $LVar_t^e$ is used on a real options model, the maximization of the project value will not be achieved. This means that the conditional variance should be used in real options models. A similar reasoning will apply to other uses of stochastic project value models, and $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$ will provide a more accurate depiction of project behavior.

2.2. The variance of project value

The use of log-variance assumes that the project dynamics follows (3), at least approximately. This kind of dynamics may be appropriate for assets that are traded in capital markets, and whose values will never become negative. However, it will usually be less applicable to investment projects, whose value may become negative and may not have a lognormal distribution, and therefore it has been criticized by several authors (e.g., Wang and Dyer, 2010). When the project value may become negative, the expression (4) for calculating k_t will not be correctly defined. Other measures of the project rate of return, based on ratios, might be used instead (see, e.g, Costa Lima and Suslick, 2006b). However, such measures will also become incorrectly defined when there is a strictly positive probability of the project value becoming negative.

In order to avoid such drawbacks associated with measures of the relative change in project value, a measure of the absolute change in project value may be used: the unanticipated change in project value between time *t*-1 and time *t*, that is, the difference between the net value at *t* and the present value at *t*-1, compounded to *t*. Such a measure may be particularly useful for some approximations of the dynamics of project value, e.g., when the local dynamics of project value is approximated by an Arithmetic Brownian Motion (ABM). Arithmetic Brownian motions have, in fact, been used by several authors to model the dynamics of cash flows or project values (e.g., Alexander *et al.*, 2012, Bar-Ilan, 2000, Lahmann, 2013, Wolbert-Haverkamp and Musshoff, 2014).

The unanticipated change in project value between time *t*-1 and time *t* is denoted by Δ_t. Assuming that the continuously compounded discount rate is r:

$$\Delta_t = NPV_t - PV_{t-1} \cdot e^r \tag{7}$$

For the same reason that was presented in the previous subsection, the variance of Δ_t should be conditioned to $\boldsymbol{\psi}_{t-1}$. Since PV_{t-1} is determined by $\boldsymbol{\psi}_{t-1}$ and r_t is known, this variance (denoted by $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$)[2] may be defined as:

$$Var_t\left(\boldsymbol{\psi}_{t-1}\right) = var\left(\Delta_t \mid \boldsymbol{\psi}_{t-1}\right) = var\left(NPV_t \mid \boldsymbol{\psi}_{t-1}\right) \tag{8}$$

[2] In order to clarify the context in which the variance is being calculated, var(·) is used for the variance of a random variable and *Var*(·) is used specifically for the variance of the value of a project.

The expected variance between time t-1 and time t is denoted by $LVar_t^e$ and defined as:

$$Var_t^e = E\left(Var_t\left(\boldsymbol{\psi}_{t-1}\right)\right) = E\left(var\left(NPV_t \mid \boldsymbol{\psi}_{t-1}\right)\right) \tag{9}$$

Following the argument presented in the previous subsection, $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$ should be preferred to Var_t^e when building project value models. Therefore, this paper will focus on estimating $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$. Notice that the proposed procedures may also be easily used to estimate Var_t^e.

As explained before, the use of the log-variance assumes that the project dynamics follows (3), at least approximately. Such an assumption is more valid for assets traded in capital markets than for investment projects, but several authors use it in this latter context. When that assumption is valid and the rate of return on the project follows a normal distribution, it is possible to use the well-known relationships between the first two moments of a normal and a lognormal distribution to establish a relation between $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$ and $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$:

$$LVar_t\left(\boldsymbol{\psi}_{t-1}\right) = \ln\left(1 + \frac{Var_t\left(\boldsymbol{\psi}_{t-1}\right)}{E\left(NPV_t \mid \boldsymbol{\psi}_{t-1}\right)}\right) \tag{10}$$

Expression (10) will be useful for estimation procedures that can handle the variance but not the log-variance, as is the case with the one and a half level simulation procedure proposed in this paper.

3. Examples

This section defines two example projects that have been originally considered in other papers, to which the volatility estimation procedures will later be applied. The considered projects are simple projects, for which it is possible to analytically calculate the volatility. Notice that the proposed procedures are particularly useful for complex projects, for which it is impracticable to analytically calculate the volatility. However, in order to test the procedures, it is convenient to use them in projects for which the volatility can be analytically calculated, in order to be able to assess the estimation error.

For each of the projects, a given year will be used for the purpose of assessing the accuracy of the estimation procedures, and the analytical expression for the volatility in that year will be determined. Time 0 corresponds to the beginning of the project, so there is no information about the way the project is evolving. Volatility in year 1 is based on the state of the project in time 0. Since there is no information about how the project is evolving in time 0, year 1 will be of little interest for assessing procedures aiming at estimating state-dependent volatility. Volatility in year 2 is conditioned by the state of the project at the end of year 1, so there is already information about how the project is evolving, and there are also uncertain cash flows remaining in the project. Year 2 will, therefore, provide an interesting assessment of the estimation procedures, and it will be used for Project 1. In the case of Project 2, the analytical calculation of state-dependent variance in years 2 and 3 is quite cumbersome, so year 4 will be used instead.

3.1. Project 1

Project 1 is very simple, based on a project originally considered by Godinho (2006), and designed in such a way that its log-variance is constant (that is, independent of time and of the project state). The project consists of producing 100 units of a market-traded commodity that has a current price of \$1/unit. The continuously compounded rate of return on the commodity is normally distributed with a mean $\mu = 10\%$/year and a standard deviation $\sigma = 15\%$/year. The only cost of the project is the initial investment and, since we are only interested in estimating volatility, the value of that cost is irrelevant. The 100 units of the commodity will only be available three years after starting the project, and the rate of return shortfall is null for the commodity (meaning that there are no benefits or costs from physically holding the commodity before year 3). The continuously compounded risk-adjusted discount rate for the project is the average annual commodity price increase, $r = 11.125\%$.

The commodity price at year t is denoted by P_t. The only state variable of this project is the commodity price, so $\boldsymbol{\psi}_{t-1} \equiv \left[P_{t-1}\right]$. In the assessment of the volatility estimation procedures, the volatility of this project for year 2 is considered, so it is only necessary to analyze the volatility of NPV_2 given the state of the project at the end of year 1.

For the log-variance, we have:

$$\begin{aligned} LVar_2\left(P_1\right) &= \operatorname{var}\left(\ln\left(NPV_2\right) \mid P_1\right) \\ &= \operatorname{var}\left(\ln\left(100e^{-0.11125}E\left(P_3 \mid P_2\right)\right) \mid P_1\right) \\ &= \operatorname{var}\left(\ln\left(100e^{-0.11125}\left(P_2 e^{0.11125}\right)\right) \mid P_1\right) \\ &= \operatorname{var}\left(\ln\left(P_2\right) \mid P_1\right) \\ &= 0.15^2 = 0.0225 \end{aligned} \tag{11}$$

In (11), the first equality comes from (5). The second one comes from the application of (2) to calculate NPV_2. To get the expression in the third line, we replace the conditional expectation of P_3 by its analytical expression. The fourth equality comes from changing the logarithm of the product into a sum of logarithms and removing the constant terms thus obtained (since they have null variance). Finally, given the geometric Brownian motion defined for the price, we know that the conditional distribution of $\ln(P_2)$ is normal, with standard deviation 0.15.

So, in this project, the log-variance will be constant for all years and for all values of the state variable: it is always $LVar_t = 0.0225$. As for the variance of the project value, it changes from year to year, and it is a function of the state variable P_{t-1}. For $t = 2$:

$$\begin{aligned} Var_2(P_1) &= \text{var}(NPV_2 \mid P_1) \\ &= \text{var}\left(100e^{-0.11125}E(P_3 \mid P_2) \mid P_1\right) \\ &= \text{var}\left(100e^{-0.11125}\left(P_2 e^{0.11125}\right) \mid P_1\right) \\ &= 10^4 \,\text{var}(P_2 \mid P_1) \\ &= 10^4 P_1^2 \left(e^{0.0225}-1\right)e^{0.2225} \end{aligned} \tag{12}$$

The first equality comes from (8). The second one comes from the application of (2) to calculate NPV_2. In the third one, we replace the expectation of P_3 by its analytical expression. The fourth equality comes from the moving the constant to outside the variance expression. Finally, given the geometric Brownian motion defined for the price, we know that the conditional distribution of P_2 is lognormal with parameters $0.1+\ln(P_1)$ (geometric mean) and 0.0225 (geometric variance). So, the variance is $P_1^2\left(e^{0.0225}-1\right)e^{0.2225}$ (see, e.g, Montgomery and Runger, 2011, p. 144).

3.2. Project 2

Project 2 was originally analyzed by Cobb and Charnes (2004), and also considered by Godinho (2006). It is an investment project that produces cash flows for five years. In each year t (t = 1,..,5), the relevant sources of uncertainty are the unit contribution margin X_t and the annual demand D_t. Data used to define the values of the before-tax cash flows is presented in Table 1. In Table 1, $N(\mu, var)$ represents the normal distribution with mean μ and variance *var*, and $T(a,b,c)$ represents the triangular distribution with minimum *a*, mode *b*, and maximum *c*.

The discount rate of the project is 12% and the tax rate is 40% (these values are non-stochastic). The required initial investment is irrelevant for the estimation of project volatility.

Table 1 – Data used for defining the before-tax cash flows of Project 2.

Year	Unit Contribution Margin (X_t)	Annual Demand (D_t)	Fixed Expenses
1	$X_1 \sim N(50, 10)$	$D_1 \sim T(95, 100, 105)$	\$4,250
2	$X_2 \sim N(60, 15)$	$D_2 \sim T(82.5, 100, 117.5)$	\$4,500
3	$X_3 \sim N(70, 21)$	$D_3 \sim T(70, 100, 130)$	\$4,750
4	$X_4 \sim N(80, 28)$	$D_4 \sim T(57.5, 100, 142.5)$	\$5,000
5	$X_5 \sim N(90, 36)$	$D_5 \sim T(45, 100, 155)$	\$5,250

Cobb and Charnes (2004) examine several different scenarios for the correlations between the random variables X_t and D_t. This paper considers that there is serial correlation in the unit price, and consequently in the unit contribution margin, with a correlation coefficient of 0.6 between X_t and X_{t+1} ($t = 1,...,4$). In this paper, we consider the volatility of this project in year 4, for the purpose of assessing the volatility estimation procedures.

For this project, there is a strictly positive (although small) probability of the project value becoming negative. In such a case, the log-variance will be improperly defined. Therefore, only the variance is considered for the effect of assessing the volatility estimation procedures.

The state variables are denoted by X_t and D_t, so $\boldsymbol{\psi}_{t-1} \equiv \begin{bmatrix} X_{t-1} & D_{t-1} \end{bmatrix}$. Since the unit contribution margin is auto-correlated, the volatility may be different for different states. In fact, we get:

$$
\begin{aligned}
Var_4\left(X_3,D_3\right) &= \operatorname{var}\left(NPV_4 \mid X_3,D_3\right) \\
&= \operatorname{var}\left(F_4+E\left(F_5 \mid X_4,D_4\right)e^{-0.12} \mid X_3,D_3\right) \\
&= \operatorname{var}\left(0.6\left(X_4D_4-5,000\right)+0.6\left(E\left(X_5 \mid X_4\right)E\left(D_5\right)-5,250\right)e^{-0.12} \mid X_3,D_3\right) \\
&= 0.36\operatorname{var}\left(X_4D_4+100e^{-0.12}\left(90+0.6\frac{\sqrt{36}}{\sqrt{28}}\left(X_4-80\right)\right) \mid X_3,D_3\right) \\
&= 0.36\operatorname{var}\left(X_4\left(D_4+60\sqrt{\frac{9}{7}}e^{-0.12}\right) \mid X_3,D_3\right) \qquad (13) \\
&= 0.36\left(\begin{array}{l}\operatorname{var}\left(X_4 \mid X_3\right)E\left(D_4+60\sqrt{\frac{9}{7}}e^{-0.12}\right)^2+ \\ +\left(E\left(X_4 \mid X_3\right)^2+\operatorname{var}\left(X_4 \mid X_3\right)\right)\operatorname{var}\left(D_4\right)\end{array}\right) \\
&= 167,796+52.020\left(X_3+45.470\right)^2
\end{aligned}
$$

The first equality comes from (8). The second one comes from the first equality of (2). The third equality is obtained by replacing the F_4 and F_5 by the expressions that define them, using basic properties of the expected value and taking into account that are X_5 and D_5 are independent, X_5 and D_4 are independent and D_5 is independent of both X_4 and D_4. The fourth equality comes from some basic properties of the variance (scaling and invariance to constant location parameters), replacing $E\left(D_5\right)$ by its value (100) and using the properties of the bivariate normal distribution to calculate $E\left(X_5 \mid X_4\right)$. In the fifth line, the invariance to a constant location parameter is used once again, and a rearrangement of the resulting expression is performed. Next, we take into account that X_4 and D_4 are independent, use the expression for the variance of the product of independent random variables, and use the independence of D_4 from both X_3 and D_3, and the independence of X_4 from D_3 to simplify the resulting expression. In the last line, the expected values and variances are replaced by their values, the conditional expected values and conditional variances are calculated by resorting to the properties of the bivariate normal distribution, and the expression thus obtained is simplified.

3.3. Project volatility as a function of the state of the project

As shown in (11), the second-year variance of the value of Project 1 is a function of the state variable, that is, the commodity price. Figure 1 depicts the relation between the commodity price in the beginning of the second year and the second-year variance of the project value, calculated using the analytical expression (12).

Figure 1 shows that the variance of project value may change significantly with the state variable. For large values of the state variable, the variance may be more about four times as high as for small values of this variable.

Following the discussion in Subsection 2.1, if a single value is estimated for the second-year volatility, this value will be an average of the values plotted in Figure 1, previously denoted as Var_2^e . Since abandonment options are usually valuable for low project values, if the average second-year volatility was used to value an abandonment option, the volatility value used in the model would be larger than the real volatility in the range of state variable values for which this option is valuable. Therefore, the value of the option would be over-estimated. On the other hand, since expansion options are usually valuable for high project values, if the average second-year volatility was used to value an expansion option, the volatility value used in the model would be smaller than the real volatility in the range of state variable values for which this option is valuable, therefore under-estimating the value of the option. In both cases, the options would be incorrectly valued.

Figure 1: Second-year variance of the value of Project 1, as a function of the state variable (the commodity price), for values of the state variable between the 2nd and the 99th percentiles of its probability distribution.

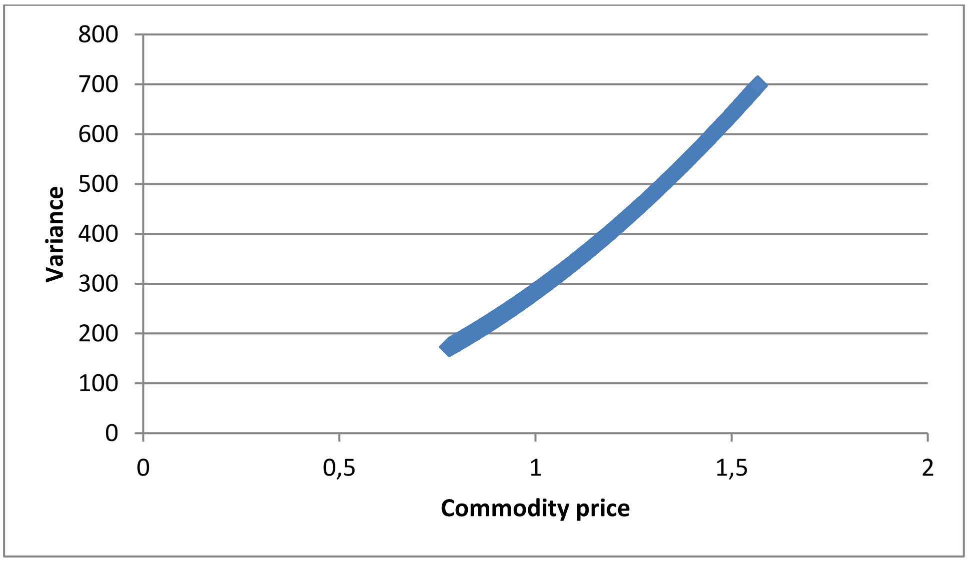
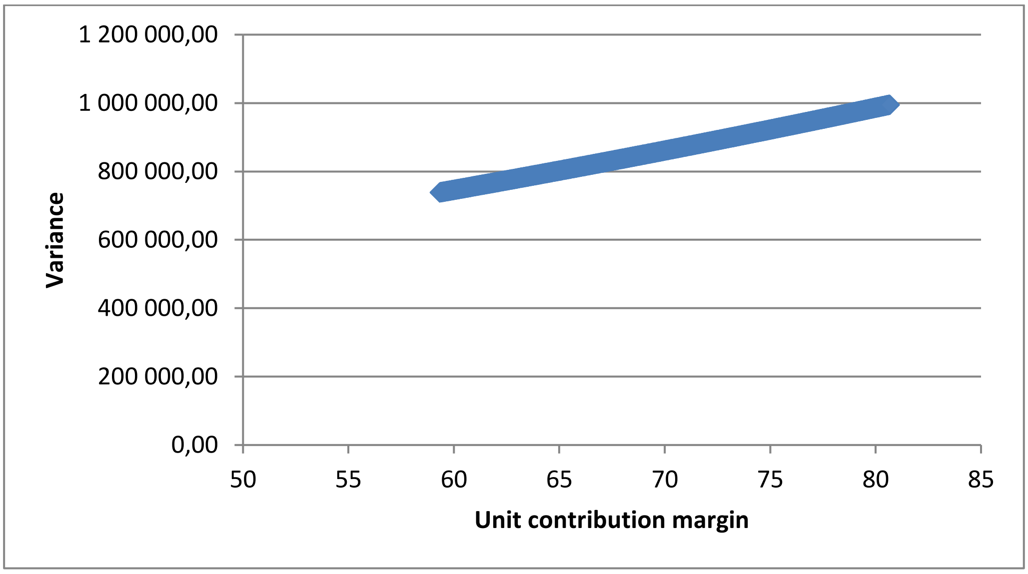


Figure 2 shows a similar issue with the fourth-year variance of the value of Project 2. The figure depicts the variance as a function of a state variable (the unit contribution margin). In this case, the variance also changes significantly with the value of the considered state variable – for large values of the unit contribution margin, the variance may be about 35% higher than for small values of this variable. This would lead to the problems discussed before in the valuation of real options.

Many real options models use the project value as the single state variable, instead of considering the underlying factors that drive project value (the "true" state variables), usually assuming that the volatility is constant. Brandão *et al.* (2012) show that volatility often changes over time and the analysis presented in this subsection shows that volatility also changes with the state of the project, so using a constant volatility may lead to important errors in real option valuation. The immediate conclusion is that it is preferable to define models that use the "true" state variables of the project, instead of using the project value as the underlying variable. However, this will often lead to very complex models, either difficult or impossible to apply.

Figure 2: Fourth-year variance of the value of Project 2, as a function the unit contribution margin, for values of this state variable between the 2nd and the 99th percentiles of its probability distribution.



An alternative may be to take into account that the "true" state variables that define the project volatility also define project value, and this will usually lead to a strong relation between project value and project volatility. In both the examples considered in this paper, this relation will be direct, in the sense that project value uniquely defines the project volatility. This means that it is possible to define charts that are similar to the ones in Figure 1 and 2, with the project value on the X-axis, instead of the state variable. In most real examples, such a direct relationship will not exist: the same project value may lead to different values of the volatility. However, in such cases, there will usually be a significant correlation between project value and volatility: a high project value will usually lead to a higher or lower volatility, according to the characteristics of the project. The availability of volatility estimation procedures will allow the calculation of an expected volatility as a function of project value, which, although being an approximation, should lead to much better results than the use of a constant volatility per period[3]. Therefore, if it is necessary or convenient to use project value instead of the "true" state variables in real options models, volatility (or expected volatility) should be modelled as a function of the project value, instead of using a constant volatility.

4. Estimating project volatility for a given state

Some procedures for calculating the project volatility in a given period (*t*) and for a given project state ($\boldsymbol{\psi}_{t-1}$) will now be presented. If $\boldsymbol{\psi}_{t-1}$ is the project state at time *t*-1, it may seem that we must "simply" simulate the behavior of the project from *t*-1 to *t*, calculate NPV_t and then define the volatility as the variance of the simulated NPV_t's (or the variance of the logarithm of NPV_t, if we are interested in the log-variance). The difficulty with this simple procedure lies in calculating NPV_t. By simulating the project behavior from *t*-1 to *t*, we will have the state of the project at time *t*, $\boldsymbol{\psi}_t$. Although NPV_t is uniquely determined by $\boldsymbol{\psi}_t$, there is usually no simple way of calculating NPV_t given $\boldsymbol{\psi}_t$. As can be seen in (2), the calculation of NPV_t requires defining the expected values of all future cash flows as a function of the state $\boldsymbol{\psi}_t$. In some very simple cases, it may be possible to determine an analytical expression for NPV_t as a function of $\boldsymbol{\psi}_t$ (the main procedure used in Brandão *et al.*, 2012, relies on the ability of defining such an analytical expression), but in many real projects that will be impractical. Another possible simplifying assumption for estimating NPV_t as a function of $\boldsymbol{\psi}_t$ consists of assuming that $\boldsymbol{\psi}_t$ can be defined as a set of variables such that the expected value of each variable at each time point can be reasonably estimated using just the previous value of that same variable (as implicit in the second procedure of Brandão *et al.*, 2012). This may also raise significant difficulties, since often costs and revenues are subject both to common influencing factors and to specific sources of uncertainty, and in such cases reasonable estimations of the expected value should incorporate previous values of multiple variables. Therefore, this paper

[3] This average volatility per year was denoted before by Var_t^e or $LVar_t^e$.

proposes procedures that do not rely on the ability of being able to define an analytical expression for NPV_t as a function of $\boldsymbol{\psi}_t$ or assuming that the uncertainty in NPV_t can be expressed by independent dynamics of influencing variables.

The first procedure, termed "two-level simulation", is based on a procedure proposed by Godinho (2006) for estimating the first-year volatility, and estimates NPV_t by using a new simulation. The second procedure is based on results presented by Sun *et al.* (2011), and it is termed "one and a half level simulation". This procedure is also based on a two level simulation, but it uses a direct estimator of the volatility, without resorting to the intermediate estimation of NPV_t. This estimator is able to achieve accurate results when the number of iterations of the second level simulation is small (thus the term "one and a half level simulation"). One drawback of such a procedure is that it is not able to directly estimate the log-variance – just the variance. However, as explained in Section 2, if it is assumed that the project value follows a lognormal distribution (as it usually is whenever log-variance is used), log-variance can be estimated easily by using (10).

Let us now introduce some notation to be used in the procedures. In both procedures, samples of the project cash flows $F_t, F_{t+1}, \ldots,$ are generated by simulation, and there are two simulations, one of them (the second-level simulation) nested into the other (the first-level simulation). Each sample of F_t is generated by the first-level simulation and then the samples of $F_{t+1}, F_{t+2}, \ldots$ are generated by the second-level simulation that takes place within each iteration of the first-level simulation. n_1 and n_2 are the numbers of iterations in the first- and second-level simulations, respectively. Sometimes it is convenient to clarify in which iteration of the first- or second-level simulations a given sample of a cash flow is generated. If a sample of F_t is generated in the *i*-th iteration of the first-level simulation, the notation $F_{t,i}$ is used to identify the value of that particular sample. For the samples of cash flows generated by the second-level simulation, it may be necessary to identify both the first- and second-level iterations in which they are generated. If a sample of F_{t+s} $(s = 1, \ldots, N - t)$ is generated in the *j*-th iteration of the second level simulation that takes place within the *i*-th iteration of the first-level simulation, the notation $F_{t+s,i,j}$ is used to identify its particular value. In order to simplify the description of the procedures, the values of the samples of the cash flows will hereafter be simply referred to as cash flows generated by the simulations.

The expected value of the time-$t+s$ cash flow, for the i-th iteration of the first-level simulation, is:

$$\overline{F_{t+s,i}} = \frac{\sum_{j=1}^{n_2} F_{t+s,i,j}}{n_2} \tag{14}$$

The cash flow $F_{t,i}$, along with the cash flows generated in an iteration of the second level simulation, lead to a time-t net present value defined by that specific second level simulation. It is denoted by $NPV_{t,i,j}$, and it can be calculated as:

$$NPV_{t,i,j} = F_{t,i} + \sum_{s=1}^{N-t} F_{t+s,i,j} \cdot e^{-r(s-t)} \tag{15}$$

In order to calculate an estimate of the net present value defined by an iteration of the first-level simulation, the expected values of the second-level cash flows must be used. Such an estimate is denoted by $\overline{NPV_{t,i}}$, and it can be calculated as:

$$\overline{NPV_{t,i}} = F_{t,i} + \sum_{s=1}^{N-t} \overline{F_{t+s,i}} \cdot e^{-r(s-t)} = \frac{\sum_{j=1}^{n_2} NPV_{t,i,j}}{n_2} \tag{16}$$

The expected time-t net present value, based on the both levels of the simulation, is denoted by $\overline{\overline{NPV_t}}$ and calculated as:

$$\overline{\overline{NPV_t}} = \frac{\sum_{i=1}^{n_1} \overline{NPV_{t,i}}}{n_1} \tag{17}$$

In the case of the one and a half level simulation procedure, some additional notation follows Sun *et al.* (2011), on whose results this procedure is based.

4.1. Two-level simulation

The underlying idea of the two-level simulation procedure is estimating NPV_t by performing a new simulation. This means that there is a first-level simulation, in which the project behavior is simulated from t-1 to t, thus generating the project state at year t, $\boldsymbol{\psi}_t$. In each iteration i of this first-level simulation, a complete second-level simulation is also performed. This second-level simulation allows the estimation of the expected value of the cash flows $F_{t+1}, F_{t+2}, \ldots$, thus providing an estimate of NPV_t for the state $\boldsymbol{\psi}_t$ defined by the first-level iteration.

This procedure will now be illustrated with a very small example that refers to its application to estimate the second-year volatility of Project 1. In order to keep the example small, the first-level simulation will only have five iterations, and the second-level simulation will have four. The values used in this example are shown in Table 2.

Table 2 – Values used in the example that illustrates two-level simulation.

First-level iteration (i)	$P_{2,i}$ (simulated)	Second-level iteration (j)	$P_{3,i,j}$ (simulated)	$F_{3,i,j}$ ($100 \cdot P_{3,i,j}$)	$\overline{F_{3,i}}$ (average of $F_{3,i,j}$)	$\overline{NPV_{2,i}}$ ($F_{2,i} + \overline{F_{3,i}} \cdot e^{-0.11125}$)
1	\$1.05	1	\$1.07	\$107.00	\$110.00	\$98.42
		2	\$1.13	\$113.00		
		3	\$1.15	\$115.00		
		4	\$1.05	\$105.00		
2	\$1.20	1	\$1.30	\$130.00	\$126.25	\$112.96
		2	\$1.25	\$125.00		
		3	\$1.23	\$123.00		
		4	\$1.27	\$127.00		
3	\$1.14	1	\$1.13	\$113.00	\$117.75	\$105.35
		2	\$1.04	\$104.00		
		3	\$1.16	\$116.00		
		4	\$1.38	\$138.00		
4	\$1.29	1	\$1.31	\$131.00	\$145.75	\$130.40
		2	\$1.55	\$155.00		
		3	\$1.27	\$127.00		
		4	\$1.70	\$170.00		
5	\$1.08	1	\$1.19	\$119.00	\$117.50	\$105.13
		2	\$1.19	\$119.00		
		3	\$1.35	\$135.00		
		4	\$0.97	\$97.00		

Simulated values ($P_{2,i}$ and $P_{3,i,j}$) are assumed to be generated randomly.
In this project, the second-year cash flow ($F_{2,i}$) is always \$0, so it is not shown in the table.

As explained, the variance will depend on the values of the state variables at the beginning of the period considered for estimating volatility (in this case, at the beginning of the second year of the project, which is time point t = 1). The only state variable of Project 1 is the price of the commodity, which will be assumed to be \$1.10/unit at t = 1 for the purpose of this example.

In each first-level iteration, i, a commodity price is simulated for year 2 (denoted by $P_{2,i}$). We will assume that the year-2 price simulated in the first iteration is \$1.05/unit. The second-year cash flow should also be calculated at this stage. This cash flow is always $0 for this project (therefore, it is not presented in Table 2).

Then, in each iteration of the first-level simulation, a complete second-level simulation is performed. This second-level simulation generates values for the state variable until the end of the project (in this example, just for year 3), and calculates the corresponding cash flows. For this example, we will assume that the simulated year-3 prices are $1.07, $1.13, $1.15 and $1.05, in the four iterations of the second-level simulation, with corresponding cash flows of $107, $113, $115 and $105, respectively. These cash flows are used to estimate the year-2 net present value of the project for the first-level iteration (in this case $\overline{NPV_{2,1}}$, since we are in iteration i = 1). To accomplish this, the average of the year-3 cash flows is calculated, and this average (an estimate of the year-3 expected cash flow) is discounted to year 2 and added to the year-2 cash flow (which is always $0 in this example), in order to calculate $\overline{NPV_{2,1}}$.

A new iteration is now started, and $\overline{NPV_{2,2}}$ is calculated. The process is repeated for all iterations of the first-level simulation, as shown in Table 2. For this example, we will assume that the values presented in Table 2 would be obtained, the most relevant outputs of the simulation being the values of $\overline{NPV_{2,i}}$, in the last column of the table. The project volatility, measured as the variance of project value in the second year of the project, is just the variance that can be estimated from the values in that column. In this case, we obtain a variance of 150.88. For the log-variance, it suffices to calculate the logarithms of the values in the last column of Table 2, and calculate their variance. In this example, we would obtain an estimate of 0.01149.

The complete procedure is described in Algorithm 1. The main lines of an implementation of this procedure using Microsoft Excel and the @Risk add-in can be found in Appendix 1.

Algorithm 1 – Two-level simulation procedure

$n_1 \leftarrow$ number of iterations of the first-level simulation

$n_2 \leftarrow$ number of iterations of the second-level simulation

For i = 1 to n_1

  Simulate the project behavior from t-1 to t and the year-t cash flow, $F_{t,i}$

  For j = 1 to n_2

Simulate the project behavior from *t* until the last period

Store the generated cash flows , which will be denoted as $F_{t+1,i,j}, F_{t+2,i,j},...$

Next *j*

Calculate the expected value of the cash flows simulated in the second-level simulation: $\overline{F_{t+1,i}} = \sum_{j=1}^{n_2} F_{t+1,i,j} \Big/ n_2$, $\overline{F_{t+2,i}} = \sum_{j=1}^{n_2} F_{t+2,i,j} \Big/ n_2$,...

Calculate $\overline{NPV_{t,i}} = F_{t,i} + \sum_{s=1}^{N-t} \overline{F_{t+s,i}} \cdot e^{-r(s-t)}$

Next *i*

The variance of $\left\{\overline{NPV_{t,i}}\right\}, i = 1,...,n_1,$ is an estimator for $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$

(the variance of $\left\{\ln\left(\overline{NPV_{t,i}}\right)\right\}, i = 1,...,n_1,$ is an estimator for $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$)

Such two-level simulations are used in project risk management models (see, e.g., Lan *et al.*, 2007). Godinho (2006) proposes such a procedure for estimating the unconditional first-year log-variance.

One important problem with this procedure is its computational burden. Since each iteration of the first-level simulation is followed by a complete simulation of the second level, a very large number of iterations of the second level must be performed. Moreover, reducing the number of iterations of the second-level simulation may lead to a bias in the results. This can be easily seen if the variance measure $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$ is taken into account. In Algorithm 1, $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$ is estimated as the variance of $\left\{\overline{NPV_{t,i}}\right\}$, for values of $\overline{NPV_{t,i}}$ generated assuming that the state of the project at time *t*-1 is $\boldsymbol{\psi}_{t-1}$. If the law of total variance is applied to the variance of $\left\{\overline{NPV_{t,i}}\right\}$:

$$\begin{aligned} \operatorname{var}\left(\overline{NPV_{t,i}} \mid \boldsymbol{\psi}_{t-1}\right) &= \operatorname{var}\left(E\left(\overline{NPV_{t,i}} \mid \boldsymbol{\psi}_t\right)\right) + E\left(\operatorname{var}\left(\overline{NPV_{t,i}} \mid \boldsymbol{\psi}_t\right)\right) \\ &= Var_t\left(\boldsymbol{\psi}_{t-1}\right) + E\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t\right) \Big/ n_2 \end{aligned} \tag{18}$$

Expression (18) shows that this procedure will provide a biased estimator for $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$ when n_2 is finite and some of the cash flows occurring after time *t* are stochastic. The value $E\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t\right)$ corresponds to random variability that does not translate into volatility, since it occurs outside the time window for which volatility is considered. In two-level simulation, such variability is mitigated, since it is divided by the number of iterations of the second level, n_2. If n_2 is large, the bias will become negligible; however, if n_2 is small, the bias may become significant. This means that the second-level simulation should utilize a large number of iterations.

4.2. One and a half level simulation

One and a half level simulation is a procedure based on the results presented by Sun *et al.* (2011), which uses a two-level simulation with a small number of iterations in the second level in order to estimate a variance that is, in fact, equivalent to $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$. There are three main differences between this procedure and the two-level simulation procedure presented before:

- Instead of estimating $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$ as the variance of $\left\{\overline{NPV_{t,i}}\right\}$, which would lead to a biased estimation (as shown before, based on (18)), an ANOVA estimator, demonstrated to be unbiased under mild conditions, is used;
- The number of second-level iterations is not defined as a parameter of the procedure, but instead it is calculated in a way that minimizes the variance of the estimator;
- The results in which this procedure is based do not extend to the estimation of log-variance; however, if it is assumed that that the project dynamics approximately follows (3) and the continuously compounded rate of return on the project follows a normal distribution (as usually happens when the log-variance is of interest), it is possible to use (10) in order to estimate $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$ by resorting to estimates of $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$ and $E\left(NPV_t \mid \boldsymbol{\psi}_{t-1}\right)$ (for the latter estimate, the value $\overline{\overline{NPV_t}}$, defined in (17) and easily calculated within this procedure, can be used).

The estimator used for $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$, denoted by $V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)$, is calculated as (see Sun *et al.*, 2011, eq. 9):

$$V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)=\frac{1}{n_1-1}\sum_{i=1}^{n_1}\left(\overline{NPV_{t,i}}-\overline{\overline{NPV_t}}\right)^2-\frac{1}{n_1 n_2\left(n_2-1\right)}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left(NPV_{t,i,j}-\overline{NPV_{t,i}}\right)^2 \qquad (19)$$

In order to determine the optimal number of first- and second-level iterations in one and a half level simulation, Sun *et al.* (2011) assume the existence of a given computational budget *C*, and divide this budget by the two simulation levels in order to minimize the variance of the estimator (in the case considered in this paper, the variance of $V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)$). The authors show that there is a finite, constant, asymptotically optimal level of second-level iterations, $n_2{**}$, independent of the budget *C*. So, as this computational budget grows to infinity, the policy of setting the number of second-level iterations to $n_2{**}$ is at least as good as any policy of setting the second-level size as a function of *C*. In light of some computational tests, Sun *et al.* recommend the usage of $n_2{**}$ second-level iterations regardless of the total computational budget. This way, the number of first-level iterations is defined by the computational budget and by $n_2{**}$.

In this paper, the recommendation of Sun *et al.* (2011) is followed, and an estimator of the asymptotically optimal level of second-level iterations, $n_2{**}$, is used. However, it should be noticed that the expression used by Sun *et al.* to calculate $n_2{**}$ is derived assuming that the computational cost of generating a first-level scenario (that is, the cost of generating the project state $\boldsymbol{\psi}_t$) is null. This implies that all the computational cost is due to the second-level iterations, allowing the authors to simplify the expressions they use. In the application of one and a half level simulation to the estimation of project volatility, the generation of the project state $\boldsymbol{\psi}_t$ corresponds to the calculation of the state variables' values and of the time-*t* cash flow, and each second-level iteration will correspond to the generation of the state variables' values and of the cash flows from year *t*+1 until the end of the project. If the project volatility is being estimated for a period near the end of the project life, the computational cost of the generating $\boldsymbol{\psi}_t$ will be relevant, in relative terms, so it will not be reasonable to follow the assumption of Sun *et al.*. Therefore, a more general expression for calculating $n_2{**}$ will now be obtained.

Assume that the cost of performing a second-level iteration is used as the unit of measurement for the computational cost, and that the cost of generating the project state $\boldsymbol{\psi}_t$ is γ. To illustrate the meaning of γ, assume that the first-level generation of the project state consists on generating a random value and a cash flow, and that in each second-level iteration it is necessary to generate three random values and three cash flows. Assuming additionally that each cash flow and each random value requires the same computational effort, it can be seen that a second-level iteration requires three times the computational effort of generating $\boldsymbol{\psi}_t$. Since the unit of measure is a second-level iteration, $\gamma = 1/3$.

So, if there are n_1 iterations of the first level and n_2 iterations of the second level, the computational cost is given by:

$$C = n_1 n_2 + n_1 \gamma \tag{20}$$

For a given time-*t* project state, $\boldsymbol{\psi}_t$, define τ as the difference between the expected value of NPV_t given then specific state $\boldsymbol{\psi}_t$ and the expectation of NPV_t calculated over all possible states $\boldsymbol{\psi}_t$:

$$\tau = E\left(NPV_t \mid \boldsymbol{\psi}_t\right) - E\left(NPV_t \mid \boldsymbol{\psi}_{t-1}\right) \tag{21}$$

For a second-level iteration of a first-level iteration that defined the state $\boldsymbol{\psi}_t$, define ε as the difference between the time-*t* sum of discounted cash flows and the expected value of NPV_t given state $\boldsymbol{\psi}_t$:

$$\varepsilon = \sum_{s=t}^{T} F_s \cdot e^{-r(s-t)} - E\left(NPV_t \mid \boldsymbol{\psi}_t\right) \tag{22}$$

According to Sun *et al.* (2011), eq. 10, for given values of n_1 and n_2, the variance of the estimator $V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)$ can be defined as:

$$\begin{aligned}
\operatorname{var}\left(V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)\right) = {} & \frac{1}{n_1} E\left(\tau^4\right) - \frac{\left(n_1 - 3\right)}{n_1\left(n_1 - 1\right)} Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2 + \frac{2}{n_1^2 n_2^2\left(n_1 - 1\right)} \sigma_\varepsilon^4 \\
& + \frac{2\left(n_1 + 1\right)}{n_1^2 n_2\left(n_1 - 1\right)} \sigma_\varepsilon^2 Var_t\left(\boldsymbol{\psi}_{t-1}\right) + \frac{2}{n_1^2 n_2^3} E\left(\varepsilon^4\right) \\
& + \frac{2\left(n_2^2 + n_1 n_2^2 - 4n_2 + 3\right)}{n_1^2 n_2^3\left(n_2 - 1\right)} E\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t\right)^2\right) \\
& + \frac{4n_1 + 2}{n_1^2 n_2} E\left(\tau^2 \varepsilon^2\right) + \frac{4}{n_1^2 n_2^2} E\left(\tau \varepsilon^3\right)
\end{aligned} \tag{23}$$

Similarly to Sun *et al.* (2011), we aim at finding an asymptotical approximation to the minimization of $\operatorname{var}\left(V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)\right)$. Assuming a fixed computational budget *C* whose relation with n_1 and n_2 is defined by (20), the following expression may be equivalently minimized:

$$h\left(n_1, n_2\right) = C \cdot \operatorname{var}\left(V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)\right) = \left(n_1 n_2 + \gamma n_1\right) \operatorname{var}\left(V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)\right) \tag{24}$$

In order to obtain an asymptotical approximation, we make *C* tend to infinity. For a fixed n_2 (whose optimal value we want to find out), this is the same as making n_1 tend to infinity. So we define:

$$
\begin{aligned}
h(n_2) &= \lim_{n_1 \to \infty} h(n_1, n_2) \\
&= \frac{n_2 + \gamma}{n_2} \left\{ n_2 \left[E\left(\tau^4\right) - Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2 \right] + \frac{2}{n_2 - 1} E\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t \right)^2 \right) + 4E\left(\tau^2 \varepsilon^2\right) \right\}
\end{aligned}
\tag{25}
$$

For the relevant values of n_2, that is, for $n_2 > 1$, the minimum of $h(n_2)$ will be the root of the following equation (see Appendix 2 for the derivation of this result):

$$
\begin{aligned}
&\left[E\left(\tau^4\right) - Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2 \right] (n_2 - 1)^4 + 2\left[E\left(\tau^4\right) - Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2 \right] (n_2 - 1)^3 + \\
&+ \left[E\left(\tau^4\right) - Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2 - 4\gamma E\left(\tau^2 \varepsilon^2\right) - 2E\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t \right)^2 \right) \right] (n_2 - 1)^2 - \\
&- 4(1+\gamma) E\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t \right)^2 \right) (n_2 - 1) - 2(1+\gamma) E\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t \right)^2 \right) = 0
\end{aligned}
\tag{26}
$$

The left side of (26) is a quartic in $(n_2 - 1)$ and the equation can be solved by several different methods (for an example see Borwein and Erdelyi, 1995, p. 4). Notice that, since it was not assumed that the computational cost of generating a first-level scenario is negligible, we arrived at a more complex expression than the one derived by Sun *et al.* (2011).

The number of second-level iterations must be integer, and the root of (26) will usually be non-integer. However, defining $n_2{**}$ as the nearest integer that is greater than one will provide a good approximation to the minimization of $h(n_2)$.

An important issue concerns four quantities that are necessary for defining (26) and are not known at the outset: $E\left(\tau^4\right)$, $Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2$, $E\left(\tau^2\varepsilon^2\right)$ and $E\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t\right)^2\right)$. In fact, one of the quantities we need is the squared value of the quantity that this procedure aims to estimate. This paper follows the approach of Sun *et al.* (2011), and uses a pilot simulation to obtain rough estimates of these quantities, which are then entered into (26) in order to define the number of iterations that allow an accurate estimate of $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$. This pilot simulation uses a small portion of the computational budget (e.g., 10% of this budget), and it will be a two-level simulation with an arbitrarily chosen small second-level size n_2'. The first-level size, n_1', will be defined by the portion of the computational budget devoted to this pilot simulation and by the second-level size n_2'. Sun *et al.* propose estimators for the above mentioned quantities, based on this pilot simulation:

$$
\hat{E}\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t \right)^2 \right) = \frac{1}{n_1'} \sum_{i=1}^{n_1'} \left(\sum_{j=1}^{n_2'} \frac{{NPV_{t,i,j}}^2}{n_2' - 1} - \frac{n_2' \, \overline{NPV_{t,i}}^2}{n_2' - 1} \right)^2
\tag{27}
$$

$$V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)=\frac{\sum_{i=1}^{n_1'}\left(\overline{NPV_{t,i}}-\overline{NPV_t}\right)^2}{n_1'-1}-\frac{\sum_{i=1}^{n_1'}\sum_{j=1}^{n_2'}\left(NPV_{t,i,j}-\overline{NPV_{t,i}}\right)^2}{n_1'n_2'\left(n_2'-1\right)} \tag{28}$$

$$\hat{E}\left(\tau^2\varepsilon^2\right)=V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)\cdot\frac{\sum_{i=1}^{n_1'}\sum_{j=1}^{n_2'}\left(NPV_{t,i,j}-\overline{NPV_{t,i}}\right)^2}{n_1'\left(n_2'-1\right)} \tag{29}$$

$$\begin{aligned}\hat{E}\left(\tau^4\right)=&\frac{n_1'^4}{\left(n_1'-1\right)^4+\left(n_1'-1\right)}\times\\&\times\left[\frac{1}{n_1'}\sum_{i=1}^{n_1'}\left(\overline{NPV_{t,i}}-\overline{\overline{NPV}}_t\right)^4-\frac{3\left(n_1'-1\right)\left(2n_1'-3\right)}{n_1'^3}V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)^2-\right.\\&\left.-6\frac{\left(n_1'-1\right)^4+\left(n_1'-1\right)}{n_1'^4n_2'}\hat{E}\left(\tau^2\varepsilon^2\right)\right]\end{aligned} \tag{30}$$

So, after the pilot simulation, (27)-(30) are used to estimate the required quantities, and (26) is used to calculate the number of second-level iterations. A two-level simulation is performed and (19) is used to estimate $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$.

This procedure, particularly the pilot simulation component of the procedure, will now be illustrated with a very small example, which concerns to the estimation of the second-year volatility of Project 1. For the pilot simulation, we assume that both the first-level and the second-level simulations have four iterations. The values used for the pilot simulation are shown in Table 3.

Table 3 – Values used for the pilot simulation of the example that illustrates one and a half level simulation.

First-level iteration (i)	$P_{2,i}$ (simulated)	Second-level iteration (j)	$P_{3,i,j}$ (simulated)	$F_{3,i,j}$ ($100 \cdot P_{3,i,j}$)	$NPV_{2,i,j}$ ($F_{2,i}+F_{3,i,j}\cdot e^{-0.11125}$)	$\overline{NPV_{2,i}}$ (average of $NPV_{2,i,j}$)
1	$1.54	1	$1.50	$150.00	$134.21	$152.10
		2	$1.79	$179.00	$160.15	
		3	$1.76	$176.00	$157.47	
		4	$1.75	$175.00	$156.58	
2	$0.93	1	$0.92	$92.00	$82.31	$97.52
		2	$1.13	$113.00	$101.10	
		3	$1.02	$102.00	$91.26	
		4	$1.29	$129.00	$115.42	
3	$1.22	1	$1.43	$143.00	$127.94	$119.67
		2	$1.28	$128.00	$114.52	

		3	$1.52	$152.00	$136.00	
		4	$1.12	$112.00	$100.21	
4	$1.29	1	$1.39	$139.00	$124.37	$119.44
		2	$1.06	$106.00	$94.84	
		3	$1.52	$152.00	$136.00	
		4	$1.37	$137.00	$122.58	

Simulated values ($P_{2,i}$ and $P_{3,i,j}$) are assumed to be generated randomly.
In this project, the second-year cash flow ($F_{2,i}$) is always $0, so it is not shown in the table.

The pilot simulation starts by generating a value for the state variables at the beginning of the second year of the project (t =1). The only state variable of Project 1 is the price of the commodity, which will be assumed to be $1.10/unit at t = 1. In each first-level iteration, i, a commodity price is simulated for year 2 (denoted by $P_{2,i}$), and the corresponding second-year cash flow is calculated. We will assume that the year-2 price simulated in the first iteration is $1.54/unit, and the second-year cash flow is always $0 for this project.

In each iteration of the first-level simulation, a complete second-level simulation is performed. This second-level simulation generates values for the state variable in year 3 and calculates the corresponding cash flows. For this example, we will assume that the simulated year-3 prices are $1.50, $1.79, $1.76 and $1.75, in the four iterations of the second-level simulation, with corresponding cash flows of $150, $179, $176 and $175, respectively. In this pilot simulation, it is also necessary to calculate $NPV_{2,1,j}$ in all iterations of the second level simulation. This corresponds to discounting the year-3 cash flow to the second year, leading to the values of $134.21, $160.15, $157.47 and $156.58. $\overline{NPV_{2,1}}$ can now be calculated as the average of these for values, leading to $\overline{NPV_{2,1}} = \$152.10$.

A new iteration is now started, and the values of $NPV_{2,2,j}$ and $\overline{NPV_{2,2}}$ are calculated. The process is repeated for all iterations of the first-level simulation, as shown in Table 3. After the simulation is concluded, $\overline{\overline{NPV_2}}$ can be calculated as the average of the values $\overline{NPV_{2,i}}$. For the values in Table 3, we obtain $\overline{\overline{NPV_2}} = \$122.18$.

The values of $NPV_{2,i,j}$, $\overline{NPV_{2,i}}$ and $\overline{\overline{NPV_2}}$ allow the calculation of $\hat{E}\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t\right)^2\right)$, $V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)$, $\hat{E}\left(\tau^2\varepsilon^2\right)$ and $\hat{E}\left(\tau^4\right)$, by using (27)-(30). For the values in Table 3, the following values are obtained: $\hat{E}\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t\right)^2\right) = 53,746.00$,

$\hat{Var}_t(\boldsymbol{\psi}_{t-1}) = 449.61$, $\hat{E}(\tau^2\varepsilon^2) = 100{,}829.05$ and $\hat{E}(\tau^4) = 307{,}787.80$. These values are substituted into (26) to reach the following equation:

$$105{,}642 \cdot (n_2 - 1)^4 + 211{,}285 \cdot (n_2 - 1)^3 - 405{,}166(n_2 - 1)^2 - 429{,}968(n_2 - 1) - 214{,}984 = 0 \quad (31)$$

It is necessary to find the only root of (31) that is larger than 1. For that we can resort to the Microsoft Excel Solver, for example. We reach a root $n_2{*} = 2.78$. We will use the closest integer to $n_2{*}$, which is $n_2{**} = 3$, as the number of second-level iterations for the main simulation. The number of first-level iterations can be calculated from the budget defined for the simulation, using (20).

The simulation process in the main simulation is similar to the one in the pilot simulation, and it will not be presented in detail. In fact, $NPV_{2,i,j}$, $\overline{NPV_{2,i}}$ and $\overline{\overline{NPV_2}}$ are calculated for all first- and second-level iterations of the main simulation, and they are used in (19) to estimate the variance of the project value. If the log-variance is necessary, it can be estimated by using (10), with $\overline{\overline{NPV_2}}$ as an estimate for $E(NPV_t \mid \boldsymbol{\psi}_{t-1})$, as long as the project dynamics follows (3) and the continuously compounded rate of return on the project follows a normal distribution.

The complete procedure is described in Algorithm 2. The main lines of an implementation of this procedure using Microsoft Excel and the @Risk add-in can be found in Appendix 3.

Algorithm 2 – One and a half level simulation procedure

$C \leftarrow$ computational budget for the simulations, in terms of second-level iterations
$\gamma \leftarrow$ computational cost of generating a project state $\boldsymbol{\psi}_t$, in terms of second-level iterations
$\alpha \leftarrow$ percentage of the computational budget to be used in the pilot simulation
$n_2' \leftarrow$ number of second-level iterations in the pilot simulation
Calculate $n_1' = \lfloor \alpha C / (n_2' + \gamma) \rfloor$
For *i* = 1 to n_1'
Simulate the project behavior from *t*-1 to *t* and the year-*t* cash flow, $F_{t,i}$

For j = 1 to n_2'

Simulate the project behavior from t until the last period, and calculate the respective cash flows, which will be denoted as $F_{t+1,i,j}, F_{t+2,i,j}, \ldots$

Calculate $NPV_{t,i,j}$, according to (15)

Next j

Use (16) to calculate $\overline{NPV_{t,i}}$

Next i

Use (17) to calculate $\overline{\overline{NPV_t}}$

Use (27)-(30) to estimate $E\left(\tau^4\right)$, $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$, $E\left(\tau^2\varepsilon^2\right)$ and $E\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t\right)^2\right)$

Substitute the estimates of $E\left(\tau^4\right)$, $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$, $E\left(\tau^2\varepsilon^2\right)$ and $E\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t\right)^2\right)$ into (26), and calculate the only root of that equation that is larger than 1 (one), $n_2{}^*$

Let $n_2{}^{**}$ be the closest integer to $n_2{}^*$ and calculate $n_1{}^{**} = \left\lfloor (1-\alpha)C/\left(n_2{}^{**}+\gamma\right) \right\rfloor$

For i = 1 to $n_1{}^{**}$

Simulate the project behavior from t-1 to t and the year-t cash flow, $F_{t,i}$

For j = 1 to $n_2{}^{**}$

Simulate the project behavior from t until the last period, and calculate the respective cash flows, which will be denoted as $F_{t+1,i,j}, F_{t+2,i,j}, \ldots$

Use (15) to calculate $NPV_{t,i,j}$

Next j

Use (16) to calculate $\overline{NPV_{t,i}}$

Next i

Use (17) to calculate $\overline{\overline{NPV_t}}$

Use (19) to estimate $Var_t\left(\boldsymbol{\psi}_{t-1}\right)$

Expression (10) can be used to estimate $LVar_t\left(\boldsymbol{\psi}_{t-1}\right)$, using $\overline{\overline{NPV_t}}$ as an estimate for $E\left(NPV_t \mid \boldsymbol{\psi}_{t-1}\right)$

4.3. Computational tests

4.3.1. Definition of the tests

Computational tests were defined for comparing the performance of the procedures described in the previous subsections. The procedures were implemented in C computer language, using Microsoft Visual Studio. The projects described in Section 3 were used in order to compare the procedures for volatility estimation. Since the Copeland and Antikarov (2001) procedure has been widely used by different authors, it was also included in the analysis. Notice that, as shown before, this procedure is expected to over-estimate volatility.

The same computational budget for the simulations was allocated to each procedure and one thousand (1,000) initial project states were simulated for the moment before the beginning of the volatility estimation year ($\boldsymbol{\psi}_{t-1}$, with $t=2$ in the case of Project 1 and $t=4$ in the case of Project 2). The volatility was then estimated for each of these states, using the defined computational budget and each of the three procedures. The projects were defined in a way that it is possible to calculate the theoretically correct value of the volatility (this was done in Subsections 3.1 and 3.2), so we are able to compare the estimated volatility with a value that we know that is theoretically correct.

For defining and assigning computational budgets, it was assumed that in each project the simulation effort was proportional to the number of cash flows it is necessary to simulate. For each procedure, three different computational budgets were considered: 10,000, 100,000 and 1,000,000 simulated cash flows. In the cases of the two-level and the one and a half level simulations, the computational budget is shared by first- and second-level iterations, and in the case of the Copeland and Antikarov procedure, it is used by the single level of simulation. For the two-level and the one and a half level simulations, a first-level iteration will simulate one cash flow (the second-year cash flow, in the case of Project 1, and the fourth-year cash flow, for Project 2) and a second-level iteration will also simulate one cash flow (the third and fifth year cash flows for projects 1 and 2, respectively). This means that the cost of a first-level iteration is equal to the cost of a second-level iteration, consequently $\gamma=1$ for both projects.

In the case of the one and a half level simulation procedures, the way the computational budget is split by the two simulation levels is defined by the method: the number of second-level iterations is defined by (26), and (20) can then be used to calculate the number of first-level iterations. However, for the pure two-level simulation procedure, there is no a priori rule for the number of iterations in each level. In each application of this procedure, a ratio between the number of first- and second-level iterations is defined, $\alpha = n_1/n_2$. Two values for this ratio are considered: $\alpha = 1$ and $\alpha = 10$. The former value represents an equal number of first- and second-level iterations, while the latter represents an increase in the number of first-level iterations at the expense of the number of second-level ones. This latter value, $\alpha = 10$ was chosen because in preliminary tests it behaved well in the estimation of volatility near the end of the project.

In the one and a half level simulation procedure, 10% of the computational budget is allocated to the pilot estimation, similarly to Sun *et al.* (2011). The number of second-level iterations used in this pilot estimation is 15.

In the case of Project 1, both the log-variance and the variance were estimated. In the case of Project 2, only the variance was estimated, since the project value may be negative.

4.3.2. Results and analysis

Table 4 presents the results of the tests. The accuracy of the estimations is assessed by the Mean Absolute Error (MAE) and by the Mean Absolute Percentage Error (MAPE). Notice that, in these examples, the magnitude of the volatility is very different for the two concepts of volatility (log-variance and variance) and for the two projects. As explained before, the log-variance is 0.0225 for Project 1 and Figure 1 shows the variance of Project 1 ranging from around 200 to around 700, while Figure 2 shows the variance of Project 2 ranging from around 750,000 to around 1,000,000. So, a 10% error in volatility estimation may correspond to absolute errors of different magnitudes for the different projects and different concepts of volatility. In order to show values that are easier to read, in Table 4 the mean absolute error was multiplied by 1000 in the case of the log-variance of Project 1 and by 0.001 in the case of the variance of Project 2.

Table 4 – Estimation of the project volatility using the proposed procedures, for the three example projects.

	Project 1		Project 2
	Log-variance[4]	Variance	Variance

[4] As explained in the text, the log-variance cannot be directly estimated by 1½-level simulation. To estimate it with this procedure, it is assumed that the dynamics of the project value follows (3), with the rate of return on the project following a normal distribution (this assumption is valid for Project 1). Under this assumption, the estimates of the mean and variance of project value are substituted into expression (10) to estimate the log-variance.

Budget	Method	MAE ($x10^3$)	MAPE	MAE	MAPE	MAE ($x10^{-3}$)	MAPE
10^4	Copeland & Antikarov	22.50	99.98%	370.61	102.32%	1225.78	142.49%
	2-level (α=1)	2.48	11.02%	44.99	12.37%	90.95	10.61%
	2-level (α=10)	1.44	6.38%	27.32	7.56%	62.23	7.19%
	1½-level	0.73	3.26%	12.45	3.42%	30.76	3.57%
10^5	Copeland & Antikarov	22.50	100.02%	367.53	102.32%	1230.88	142.26%
	2-level (α=1)	1.44	6.38%	25.67	7.15%	49.06	5.70%
	2-level (α=10)	0.77	3.44%	14.84	4.11%	29.68	3.44%
	1½-level	0.24	1.06%	3.91	1.10%	8.53	0.99%
10^6	Copeland & Antikarov	22.50	100.01%	371.22	102.28%	1228.38	142.36%
	2-level (α=1)	0.78	3.49%	13.91	3.83%	27.45	3.18%
	2-level (α=10)	0.42	2.01%	7.73	2.15%	16.59	1.93%
	1½-level	0.08	0.34%	1.29	0.36%	2.58	0.30%

Budget: computational budget for the simulations; MAE: Mean Absolute Error; MAPE: Mean Absolute Percentage Error; 2-level: two-level simulation procedure; 1½ level: one and a half level simulation procedure; ; Copeland & Antikarov: procedure proposed by Copeland and Antikarov (2001) ; α: ratio between the number of first- and second-level iterations.

Both the two-level and the one and a half level simulation procedures produce different estimation errors that depend on the computational budget. As expected, the mean estimation error for these methods always decreases when the computational budget increases. The results also show clearly that the one and a half level simulation procedure performs much better than the two-level simulation. As for the Copeland and Antikarov procedure, it produces much higher errors that do not show a clear decrease when the computational budget increases. This is due to the fact that the most important component of the estimation errors is the bias induced by this procedure, which does not decrease with the computational budget. In fact, the results of Godinho (2006) show that this procedure is expected to have an upward bias of 100% when estimating the first-year log-variance of a 2-year project with characteristics similar to Project 1. Since in Project 1 we are also estimating the log-variance of the project in the year before the last year of the project, and since the log-variance is constant in this project, the bias will be similar in this application of the procedure. In fact, a percentage error of around 100% is always found when this procedure is applied to estimate the second-year log-variance of Project 1.

Another interesting result concerns the comparison between the results of the two-level simulation procedure for different ratios between the numbers of first- and second-level iterations (α). For these projects, a larger number of first-level iterations (α=10) always performs better. Remember that in the projects being analysed, the project volatility is estimated for the year immediately before the last year of the project. In these cases, the volatility of the values obtained in the second-level iterations will be limited, so the bias due to this volatility (the bias due to the last term of (18)) will be small and it is better to use a larger number of first-level iterations. If volatility was being calculated for previous years, the results would likely be quite different, in the sense of a larger α leading to worse results, since the bias in volatility estimation would be more significant.

5. Conclusions and future research

This article addressed the estimation of state-dependent project volatility by employing Monte Carlo simulation. Two possible definitions of the concept of project volatility were considered and it was shown that volatility may change significantly with the project state, which means that expected volatility may also change significantly with project value. It was argued that using a single average volatility per period may introduce important biases in the valuation of real options.

Two procedures were proposed for estimating state-dependent volatility: two-level simulation and one and a half level simulation. Both procedures produce volatilities that depend on the project state, showing their ability to capture the change of volatility with the state of the project. Considering two simple projects and defining identical computational budgets for the procedures (based on the number of cash flows it is necessary to simulate), the accuracy of the estimates was compared for both procedures, and with the Copeland and Antikarov (2001) procedure. Both procedures proposed in this paper perform much better than the Copeland and Antikarov procedure, and the one and a half level simulation procedure produces better results than the two-level simulation procedure.

The use of real options models has been an important motivation for developing procedures for the estimation of project volatility, but this research also has important consequences in other contexts related to project risk analysis. If volatility increases when project value decreases, an initial decrease in value will tend to produce higher fluctuations in project value, fattening the left tail of the distribution of project value. As a consequence, risk measures based on this left tail, like value-at-risk or conditional value-at-risk, will lead to larger values. So, in this case, using the average project volatility will lead to the under-estimation of such measures. On the contrary, if volatility tends to decrease when project value decreases, using an average project volatility instead of state-dependent volatility will lead to over-estimating such measures.

Having accurate estimates of project volatility is particularly important for applying the lattice-based approach proposed by Copeland and Antikarov (2001). In order to use this approach, project volatility must be defined as a function of the project value, and not as a function of other state variables. As argued at the end of Subsection 3.3, there will usually be a strong relation between project value and project volatility: in the example projects considered in this paper, volatility is uniquely defined by the project value and, in both the projects, variance increases monotonically with the project value. Even if no such direct relationship exists between project value and volatility, defining expected volatility as a function of project value, while still being an approximation, is expected to lead to much better results than the use of a constant volatility per period.

A possible approach to define such a relationship between project value and expected volatility is to start by considering several samples of the state variables and estimate both the project values and the volatilities associated with those samples. The project values and volatilities thus estimated may then be used to build a function that maps a project value to an expected volatility. Defining the best ways to sample the values of state variables and the best ways to build the volatility function are left for future research.

The Copeland and Antikarov approach assumes that the project value follows a geometric Brownian motion with constant log-variance. This assumption is usually too strong, and inapplicable to projects whose value may become negative. However, it allows the use of recombining lattices. When volatility changes with time and/or project state, lattices no longer recombine, so the approach must be modified. Some authors have already proposed some enhancements to the Copeland and Antikarov approach that do not rely on recombining lattices (e.g., the approaches based on binomial trees proposed by Brandão and Dyer, 2005 and Brandão *et al.*, 2005). The availability of methods for estimating time and state-dependent project volatility opens the way to the development of new approaches that do not require the geometric Brownian motion assumption: for example, approaches that locally approximate the behavior of project value by an arithmetic Brownian motion, and then resort to nonrecombining lattices or binomial trees. The development of such approaches is another promising way of future research.

Acknowledgments

The author gratefully acknowledges the useful comments and suggestions received from two anonymous referees and the area editor.

References

- Alexander, D. R., Mo, M. and Stent, A.F. (2012). Arithmetic Brownian motion and real options. *European Journal of Operational Research*, *219(1)*, 114-122.
- Bar-Ilan, A. (2000). Investment with an arithmetic process and lags. *Managerial and Decision Economics, 21(5)*, 203-206.
- Baker, H.K., Dutta, S. and Saadi, S. (2011). Management Views on Real Options in Capital Budgeting. *Journal of Applied Finance, 21(1)*, 18-29.
- Beckers, S. (1980). The constant elasticity of variance model and its implications for option pricing. *The Journal of Finance, 35(3)*, 661-673.
- Block, S. (2007). Are 'Real Options' Actually Used in the Real World? *The Engineering Economist*, *52(3)*, 255-267.
- Borwein, P. and Erdelyi, T. (1995). *Polynomials and Polynomial Inequalities*. New York: Springer-Verlag.
- Brandão, L.E. and Dyer, J.S. (2005). Decision Analysis and Real Options: A Discrete Time Approach to Real Option Valuation. *Annals of Operations Research, 135*, 21-39.
- Brandão, L.E., Dyer, J.S. and Hahn, W.J. (2005). Using binomial decision trees to solve real-option valuation problems. *Decision Analysis*, *2(2)*, 69-88.
- Brandão, L.E., Dyer, J.S. and Hahn, W.J. (2012). Volatility Estimation for Stochastic Project Value Models. *European Journal of Operational Research, 220(3)*, 642-648.
- Cobb, B.R. and Charnes, J.M. (2004). Real Options Volatility Estimation with Correlated Inputs. *The Engineering Economist*, *49(2)*, 119-137.
- Copeland, T. and Antikarov, V. (2001). *Real Options: A Practitioners Guide*. New York: Texere.

- Costa Lima, G.A. and Suslick, S.B. (2006a). Estimating the Volatility of Mining Projects Considering Price and Operating Cost Uncertainties. *Resources Policy 31*, 86-94.
- Costa Lima, G.A. and Suslick, S.B. (2006b). Estimation of Volatility of Selected Oil Production Projects. *Journal of Petroleum Science and Engineering, 54(3-4)*, 129-139.
- Cox, J.C. and Ross, S.A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics, 3(1)*, 145-166.
- Davis, G.A. (1998). Estimating Volatility and Dividend Yield when Valuing Real Options to Invest or Abandon. *The Quarterly Review of Economics and Finance, 38(3)*, 725-754.
- Fleten, S.-E., Heggedal, A.M. and Siddiqui, A. (2011). Transmission capacity between Norway and Germany: a real options analysis. *Journal of Energy Markets, 4(1)*, 121-147.
- Godinho, P. (2006). "Monte Carlo Estimation of Project Volatility for Real Options Analysis. *Journal of Applied Finance, 16(1)*, 15-30.
- Haahtela, T. (2007). Separating Ambiguity and Volatility in Cash Flow Simulation Based Volatility Estimation. Available at SSRN: http://ssrn.com/abstract=968226 [Accessed 30 January 2017].
- Haahtela, T. (2011). Estimating Changing Volatility in Cash Flow Simulation Based Real Option Valuation with Regression Sum of Squares Error Method. Available at SSRN: http://ssrn.com/abstract=1864905 [Accessed 30 January 2017].
- Haahtela, T. (2012). Increasing uncertainty in cash flow simulation-based volatility estimation for real options: Actual increase in volatility or symptom of excess unresolved ambiguity uncertainty. Lecture Notes in Management Science, 4, 145-153.
- Herath, H.S. and Park, C.S. (2002). Multi-Stage Capital Investment Opportunities as Compound Real Options. *The Engineering Economist, 47(1)*, 1-27.
- Kallsen, J. and Taqqu, M.S. (1998). Option Pricing in ARCH-type Models. *Mathematical Finance, 8(1)*, 13-26.
- Lahmann, A.D. (2013). The Arithmetic Brownian Motion in Corporate Valuation. Available at SSRN: http://ssrn.com/abstract=2150667 [Accessed 30 January 2017].
- Lan, H., Nelson, B.L. and Staum, J. (2007). Two-Level Simulations for Risk Management. In S.G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J.D. Tew and R.R.

Barton (Eds.), *Proceedings of the 2007 INFORMS Simulation Society Research Workshop* (pp. 102-107). Hanover: INFORMS.

- Montgomery, D.C. and Runger, G.C. (2011). *Applied Statistics and Probability for Engineers*, 5th edition, Hoboken: John Wiley and Sons.
- Mun, J. (2002). *Real Options Analysis: Tools and Techniques for Valuing Strategic Investments and Decisions*. Hoboken: John Wiley and Sons.
- Nicholls, G.M., Lewis, N.A., Zhang, L. and Jiang Z. (2014). Breakeven volatility for real option valuation. *Engineering Management Journal*, *26(2)*, 49-61.
- Ruhrmann, S., Hochdörffer, J. and Lanza, G. (2014). A methodological approach to evaluate supplier development based on real options. *Production Engineering, 8(3),* 373-382.
- Sbuelz, A. (2005). Investment under Higher Uncertainty when Business Conditions Worsen. *Finance Letters, 3(3)*.
- Smit, H.T.J. (1997). Investment Analysis of Offshore Concessions in the Netherlands. *Financial Management*, *26(2),* 5-17.
- Smith, J. (2005). Alternative Approaches for Solving Real Options Problems (Comment on Brandão *et al.* 2005). *Decision Analysis*, *2(2)*, 89-102.
- Sun, Y., Apley, D.W. and Staum, J. (2011). Efficient Nested Simulation for Estimating the Variance of a Conditional Expectation. *Operations Research, 59(4),* 998-1007.
- Triantis, A. (2005). Realizing the Potential of Real Options: Does Theory Meet Practice? *Journal of Applied Corporate Finance, 17(2),* 8-16.
- Trigeorgis, L. (1993). Real Options and Interactions with Financial Flexibility. *Financial Management*, *22(3),* 202-224.
- Wang, T. and Dyer, J. S. (2010). Valuing multifactor real options using an implied binomial tree". *Decision Analysis, 7(2),* 185-195.
- Wolbert-Haverkamp, M. and Musshoff, O. (2014). Are short rotation coppices an economically interesting form of land use? A real options analysis. *Land Use Policy, 38*, 163-174.

Appendix 1

Main lines of the implementation of two-level simulation in Microsoft Excel with the @Risk add-in

Two-level simulation can be easily implemented in Microsoft Excel, using a simulation add-in like @Risk. Figure 3 exemplifies such an implementation, for estimating the variance and the log-variance of Project 1. A number of 1,000 first-level iterations and 1,000 second-level iterations is assumed in this example.

Figure 3: Spreadsheet for estimating the volatility of Project 1 in the second year using two-level simulation.

	A	B	C	D	E
1	**Spreadsheet for estimating the volatility of project 1 in the second year using two-level simulation**				
3	**Input data (value of the state variable in the end of the first year)**				
4					
5	Year-1 commodity price		$ 1.10		
6	Discount rate		11.125%		
7	Number of second level iterations		1000		
9		**Calculations**		**Results**	
10		**1st level**			
11	Year-2 commodity price	$ 1.00		Variance	345.14
12	Mean year-3 cash flow	$ 111.03		Log-variance	2.258%
13	NPV_2	$ 99.34			
14	LN(NPV_2)	4.60			
16		**2nd level**			
17	Iteration	Year-3 commodity price	Year-3 cash flow		
18	1	$ 1.02	$ 102.36		
19	2	$ 1.00	$ 100.19		
20	3	$ 1.28	$ 128.05		
21	4	$ 1.22	$ 122.19		
22	5	$ 1.11	$ 111.02		
23	6	$ 1.26	$ 125.65		

Cells C5 to C7 are used to insert the basic input data. At any time, the spreadsheet will have a complete iteration of the first-level simulation. In order to achieve this, the second-year commodity price (the state variable) is generated in B11 (the formula used is C5*EXP(RiskNormal(10%;15%)). Then, the second level simulation is performed in the lines 18 and below. The number of lines for which the third year commodity price and its corresponding cash flow are calculated should match the number defined in C7 (meaning that these values should be calculated for lines 18 to 1017). The third year commodity price can be simulated by using the formula B11*EXP(RiskNormal(10%;15%), and the corresponding cash flow is determined by multiplying this value by 100.

This second level simulation allows the calculation of the expected year-3 cash flow (the average of the values from C18 to C1017, which is calculated in B12) and of the year-2 net present value for this iteration (the expected year-3 cash flow, discounted to year 2 by using the rate inserted in C6). To calculate the log-variance, the logarithm of this net present value must also be calculated (it is shown in B14).

The @Risk function RiskVariance can be then used: the formula RiskVariance(B13) is used in E11 to estimate the variance, and RiskVariance(B14) is used in E12 to estimate the log-variance. Then, the number of first-level iterations (assumed to be 1,000 in this example) must be inserted in the "Iterations" field of the @Risk ribbon, and the simulation may be started.

This spreadsheet is available in an online supplement.

Appendix 2

Mathematical result for 1 ½ level simulation (derivation of (26))

Following (25), $h(n_2)$ is defined as

$$h(n_2)=\frac{n_2+\gamma}{n_2}\left\{n_2\left[E\left(\tau^4\right)-Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2\right]+\frac{2}{n_2-1}E\left(\mathrm{var}\left(NPV_{t,i,j}\mid\boldsymbol{\psi}_t\right)^2\right)+4E\left(\tau^2\varepsilon^2\right)\right\}$$

Calculating the first derivative of $h(n_2)$ and rearranging the terms, we obtain

$$\begin{aligned}h'(n_2)=&\left\{\left[E\left(\tau^4\right)-Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2\right](n_2-1)^4+2\left[E\left(\tau^4\right)-Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2\right](n_2-1)^3+\right.\\&+\left[E\left(\tau^4\right)-Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2-4\gamma E\left(\tau^2\varepsilon^2\right)-2E\left(\mathrm{var}\left(NPV_{t,i,j}\mid\boldsymbol{\psi}_t\right)^2\right)\right](n_2-1)^2-\\&\left.-4(1+\gamma)E\left(\mathrm{var}\left(NPV_{t,i,j}\mid\boldsymbol{\psi}_t\right)^2\right)(n_2-1)-2(1+\gamma)E\left(\mathrm{var}\left(NPV_{t,i,j}\mid\boldsymbol{\psi}_t\right)^2\right)\right\}\Big/\left[n_2^{\,2}(n_2-1)^2\right]\end{aligned}$$

We are interested in the situation $n_2>1$, so the first order condition for an optimum can be written as (26). $E\left(\tau^4\right)\big/Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2$ is the kurtosis of the distribution of NPV_t, so it must be at least 1 (one). This means that $E\left(\tau^4\right)\geq Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2$, with the equality occurring only if NPV_t takes just two possible values, with equal probability. Assuming that $E\left(\tau^4\right)>Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2$ it can be seen that the first two coefficients of the polynomial in the numerator of $h'(n_2)$ are positive, and the last two are negative (it is not possible to say whether the coefficient of $(n_2-1)^2$ is positive or negative). This means that the coefficients of this polynomial will only have one sign change, and the Laplace rule implies that the polynomial has at most a root for $n_2>1$. $E\left(\tau^4\right)>Var_t\left(\boldsymbol{\psi}_{t-1}\right)^2$ also implies that $\lim_{n_2\to+\infty}h'(n_2)>0$, and we can easily see that $\lim_{n_2\to 1}h'(n_2)=-\infty$. Since $h'(n_2)$ is continuous for $n_2>1$, $h'(n_2)=0$ has exactly one root, n_2*, for $n_2>1$. $h'(n_2)$ is increasing at n_2*, since it crosses the abscissa axis from below. Therefore n_2* is a minimum for $h(n_2)$.

Appendix 3

Main lines of the implementation of one and a half level simulation in Microsoft Excel with the @Risk add-in

One and a half level simulation can also implemented in Microsoft Excel, using a simulation add-in like @Risk. For this method, it is preferable to use two spreadsheets: one for the pilot simulation and another one for the main simulation. Figure 4 shows a spreadsheet used for the pilot simulation and Figure 5 shows a spreadsheet for the main simulation, both of them considering the estimation of the second-year variance and log-variance of Project 1. A computational budget of 100,000 is assumed in this example, with 10% of the budget to be used in the pilot simulation. A number of 15 second-level iterations is assumed for the pilot simulation, meaning that there will be 625 first-level iterations in the pilot simulation.

Figure 4: Spreadsheet for the pilot simulation used for estimating the second-year volatility of Project 1 with one and a half level simulation.

	A	B	C	D	E	F	G
1	**Spreadsheet for performing pilot simulation in project 1 (for one and a half level simulation)**						
2							
3		**Input data**					
4	Year-1 commodity price		$ 1.10				
5	Discount rate		11.125%				
6	First level iterations in the pilot (n1)		625				
7	Second level iterations in the pilot (n2)		15				
8	Gamma		1.0				
9	Total budget (pilot+main simulation)		100000				
10							
11			**Calculations**				
12			**1st level (to be used in second level)**				
13	Year-2 commodity price		$ 1.11				
14			**1st level (to be used in parameter estimation)**				
15	Mean(NPV_2,i)		$ 105.48		E(Var(NPV_2,i,j|psi_t)^2)		161131.92
16	Mean(NPV_2,i)^2		11126.50		Var_t(psi_t-1)		353.19
17	Mean(NPV_2,i)^3		1173648.66		E(tau^2*epsilon^2)		124540.18
18	Mean(NPV_2,i)^4		123799096.45		E(tau^4)		419318.959
19	Mean(NPV_2,i^2)		11310.66				
20	(Sum((NPV_2,i^2)/(n2-1))-n2/(n2-1)*Mean(NPV_2,i)^2)^2		38930.16		Coefficients of the quartic equation		
21					Coef. of (n_2-1)^4		294575.384
22	Mean(Mean(NPV_2))		$ 122.99		Coef. of (n_2-1)^3		589150.768
23	Mean(Mean(NPV_2)^2)		15502.22		Coef. of (n_2-1)^2		-525849.168
24	Mean(Mean(NPV_2)^3)		2002451.84		Coef. of (n_2-1)^1		-1289055.35
25	Mean(Mean(NPV_2)^4)		265048008.34		Coef. of (n_2-1)^0		-644527.676
26	Mean(Mean(NPV_2^2))		15831.32				
27			**Parameter estimation**				
28	Initial n_2		2.552042768		n_2 (for main simulation)		3
29	Value of the quartic		0.00		n_1 (for main simulation)		22500
30							
31		**2nd level simulation for the pilot**					
32	Iteration	Year-3 commodity price	Year-3 cash flow	NPV_2,i,j	NPV_2,i,j^2		
33	1	$ 1.11	$ 111.35	$ 99.62	9924.89		
34	2	$ 0.95	$ 95.44	$ 85.39	7291.76		
35	3	$ 1.28	$ 127.66	$ 114.21	13045.04		
36	4	$ 1.27	$ 127.08	$ 113.70	12928.39		

Cells C4 to C9 are used to insert the basic input data. The second-year commodity price (the state variable) for the first-level simulation is generated in C13, and the second-level simulation is performed in the lines 33 and below. Since the second-level simulation has 15 iterations, lines 33 to 47 must be used for this simulation. Apart from generating the year-3 commodity price and cash flow, the second-level simulation must also calculate $NPV_{2,i,j}$ and $NPV_{2,i,j}^{\;2}$, which will be necessary for calculating the coefficients of the quartic present in the left side of (26).

This second-level simulation allows the calculation of $\overline{NPV_{2,i}}$ (the average of the values from D33 to D47, which is calculated in C15) and of the average of $NPV_{2,i,j}^{\;2}$ (the average of the values from E33 to E47, which is calculated in C19). The 2nd, 3rd and 4th powers of $\overline{NPV_{2,i}}$ are calculated in cells C16 to C18.

Cell C20 contains the squared value within parenthesis on the right side of (27). Averaging this summation over the first level iterations (by using the @Risk function RiskMean) allows the calculation of $\hat{E}\left(\operatorname{var}\left(NPV_{t,i,j} \mid \boldsymbol{\psi}_t\right)^2\right)$, in cell G15.

Cells C22 to C26 use the RiskMean function calculated over the values of cells C15 to C19. This way, C22 contains $\overline{\overline{NPV_2}}$, cells C23 to C25 contain the averages of the 2nd, 3rd and 4th powers of $\overline{NPV_{2,i}}$ and C26 contains the average, over the first level iterations, of the average of $NPV_{2,i,j}^{\;2}$. The values on these cells can be used to calculate $V\hat{a}r_t\left(\boldsymbol{\psi}_{t-1}\right)$, $\hat{E}\left(\tau^2\varepsilon^2\right)$ and $\hat{E}\left(\tau^4\right)$, in cells G16, G17 and G18, respectively.

By using the values in cells G15 to G18, the coefficients of the quartic of the left side of (26) can be easily calculated, in cells G21 to G25.

It is now necessary to perform the pilot simulation. This can be done by inserting the number of first-level iterations (assumed to be 625 in this example) in the "Iterations" field of the @Risk ribbon, and clicking the "Start Simulation" icon.

The pilot simulation is used to calculate the number of second-level iterations in the main simulation, n_2, but to accomplish that another step is still necessary, after the pilot simulation. An arbitrary initial value is inserted in C28 and the quartic is inserted in C29, written as a function of the value in C28 and using the coefficients calculated in cells G21 to

G25. Then, the Excel solver is used to find the value of $n_2 > 1$ (in cell C28) that makes the quartic take the value zero. The rounded value of C28 is then presented in G28 as the number of second-level iterations in the main simulation, and (20) is used to calculate the number of first-level iterations in G29.

These numbers of first- and second-level iterations are used in the main simulation. The spreadsheet for this main simulation is shown in Figure 5.

Figure 5: Spreadsheet for the main simulation used for estimating the second-year volatility of Project 1 with one and a half level simulation.

	A	B	C	D	E	F	G
1	**Spreadsheet for performing main simulation in project 1 (for one and a half level simulation)**						
2							
3			**Input data**				
4	Year-1 commodity price		$ 1.10				
5	Discount rate		11.125%				
6	Number of first level iterations		22500				
7	Number of second level iterations		3				
8							
9			**Calculations**			**Results**	
10		**1st level (to be used in second level)**					
11	Year-2 commodity price		$ 1.46			Variance	344.07
12		**1st level (to be used in parameter estimation)**				Log-Variance	0.0225
13	Mean(NPV_2,i)		$ 133.28				
14	Mean(NPV_2,i)^2		17764.20				
15	Mean(NPV_2,i^2)		17939.76				
16							
17	Mean(Mean(NPV_2))		122.944525				
18	Mean(Mean(NPV_2)^2)		15576.74024				
19	Mean(Mean(NPV_2^2))		15811.40907				
20							
21			**2nd level**				
22	Iteration	Year-3 commodity price	Year-3 cash flow	NPV_2,i,j	NPV_2,i,j^2		
23	1	$ 1.45	$ 145.50	$ 130.18	16946.88		
24	2	$ 1.33	$ 132.81	$ 118.83	14120.52		
25	3	$ 1.69	$ 168.59	$ 150.84	22751.89		

Cells C4 to C7 are used to insert the basic input data (including the numbers of iterations coming from the pilot simulation). The second-year commodity price (the state variable) for the first-level simulation is generated in C1, and the second-level simulation is performed in the lines 23 and below (lines 23 to 25, since we reached a number of 3 iterations for the second level simulation in the pilot simulation). The structure of the second level simulation is identical to the one used in the pilot simulation.

Cells C13 to C15 calculate $\overline{NPV_{2,i}}$, its squared value, and the average of $NPV_{2,i,j}{}^2$. Cells C17 to C19 average these values over the iterations of the first level, using the RiskMean function. G11 estimates the variance by using (19), resorting to the values in C17 to C19, and G12 estimates the log-variance by using (10) and using the value in C17 as an estimate of $E\left(NPV_t \mid \boldsymbol{\psi}_{t-1}\right)$.

The number of first-level iterations (assumed to be 22,500 in this example) must be inserted in the "Iterations" field of the @Risk ribbon, and the main simulation may be started.

These spreadsheets are available in an online supplement.

Author information

Pedro Manuel Cortesão Godinho is an Assistant Professor at the Faculty of Economics of the University of Coimbra. He is currently coordinator of the Master's Program in Management of that University and Sub-Director of the Center for Business and Economics Research (CeBER). He holds a PhD in Management/Systems Sciences in Organisations, a Master in Financial Economics and a degree in Informatics Engineering from the University of Coimbra. His research interests include project appraisal, real options, game theory, project management and portfolio theory. He is the author or co-author of about sixty refereed publications and two books.