Research Center
Finance & Information Management

Fraunhofer FIT
Project Group
Business & Information
Systems Engineering

Discussion Paper

Scheduling Flexible Demand in Cloud Computing Spot Markets – A Real Options Approach

by

Robert Keller, Lukas Häfner, Thomas Sachs, Gilbert Fridgen

January 2019

appears in: Business & Information Systems Engineering 2019

University of Augsburg, D-86135 Augsburg
Visitors: Universitätsstr. 12, 86159 Augsburg
Phone: +49 821 598-4801 (Fax: -4899)

University of Bayreuth, D-95440 Bayreuth
Visitors: Wittelsbacherring 10, 95444 Bayreuth
Phone: +49 921 55-4711 (Fax: - 844710)

www.fim-rc.de

WI-691

Scheduling Flexible Demand in Cloud Computing Spot Markets – A Real Options Approach

Abstract: The rapid standardization and specialization of cloud computing services have led to the development of cloud spot markets on which cloud service providers and customers can trade in near real-time. Frequent changes in demand and supply give rise to spot prices that vary throughout the day. Cloud customers often have temporal flexibility to execute their jobs before a specific deadline. In this paper, the authors apply real options analysis (ROA), which is an established valuation method designed to capture the flexibility of action under uncertainty. They adapt and compare multiple discrete-time approaches that enable cloud customers to quantify and exploit the monetary value of their short-term temporal flexibility. This paper contributes to the field by guaranteeing cloud job execution of variable-time requests in a single cloud spot market, whereas existing multi-market strategies may not fulfill requests when outbid. In a broad simulation of scenarios for the use of Amazon EC2 spot instances, the developed approaches exploit the existing savings potential up to 40 percent – a considerable extent. Moreover, the results demonstrate that ROA, which explicitly considers time-of-day-specific spot price patterns, outperforms traditional option pricing models and expectation optimization.

Keywords: Cloud Computing; Spot Markets; Temporal Flexibility; Real Options Analysis; Decision Support

1 Introduction

With cloud services' continuously increasing usage and business relevance, their market is becoming increasingly solvent (Keller and König 2014). At the same time, standardization is increasing. This development has allowed users to dynamically adapt their cloud services demand from no to nearly unlimited resources (Mell and Grance 2011). In a rather recent move, Infrastructure-as-a-Service (IaaS) providers, such as *Amazon Web Services* (AWS), reflect the varying demand patterns by offering their services at fluctuating *spot prices* (Karunakaran and Sundarraj 2015), which are volatile throughout the day (Ben-Yehuda et al. 2013). Thereby, such providers seek constant server utilization to avoid idle capacity and large peaks.

In many use cases, customers require the instant delivery of cloud services. Nevertheless, customers may defer jobs, for instance, simulations, rendering jobs, and scientific computations. Whenever customers do not require a cloud service instantly and expect the spot prices to fall, they can defer their demand in order to realize cost savings. The time they are willing to wait for their computing job opens a *window of temporal flexibility*.

Evaluating the cost savings potential of a customer's window of temporal flexibility is a complex task, since cloud spot prices may change frequently, as we will illustrate. Consequently, cloud customers require strategies that take the tradeoff between the costs and the waiting time into consideration (Karunakaran and Sundarraj 2015; Tang et al. 2012). Furthermore, cloud customers may not even be aware of their temporal flexibility. We identify two main obstacles to utilizing temporal flexibility in cloud computing spot markets: First, decision support for customers requires near real-time analytics on when and how long to defer computing jobs given the uncertain price development. Adequate IS or web services are required to help exploit the existing savings potential optimally. Second, deferring jobs requires customers to change their demand behavior, which might inconvenience them. Applying such IS or web services could also incur

costs for process implementation and additional planning, while waiting for jobs could lead to opportunity costs. However, such costs are highly dependent on customers' individual circumstances: the extent of their cloud services dependency, IS infrastructure, employee training, etc. We consequently focus on evaluating objectively measurable savings, because cloud customers need an estimation of their flexibility's current value to weigh it against the incurred expenses.

To address both obstacles, we apply *real options analysis* (ROA), which other IS research domains have established as a valuation method designed to capture the flexibility of action under uncertainty (Amram and Kulatilaka 1999; Benaroch and Kauffman 1999; Trigeorgis 1996). We model a customer's temporal flexibility as a *deferral option*. This real option serves to determine a value for the right to act or to await another opportunity over a period. From this overarching research objective, we derive our research question:

'How can cloud services customers quantify and exploit their demand flexibility's monetary value by using real options analysis and given uncertain short-term price development?'

To address our research question, we adapt and apply multiple option pricing models and process a dataset of Amazon *Elastic Compute Cloud* (EC2) spot prices. Our research objective covers a relevant real-world problem, as cloud customers could profit from decision support on when to purchase cloud services within a temporal flexibility window to optimally exploit their savings potential. Under market principles, such times of day would have lower cloud service demand than the server capacity available. Shifting jobs to these times contributes to balancing the cloud service demand and the supply.

We structure the remainder of this paper as follows: in Section 2, we present related work on cloud computing markets and ROA. In Section 3, we analyze our dataset of EC2 spot prices. In Section 4, we adapt multiple approaches to quantify and exploit the monetary value of short-term temporal flexibility in cloud computing demand. We thereafter evaluate these approaches in a historical simulation and sensitivity analysis in Section 5. Finally, we discuss the results in Section 6 and conclude the paper in Section 7.

2 Cloud Computing Markets and Real Options Analysis

2.1 Current Developments in Cloud Computing Markets

Cloud computing with its pay-as-you-go model and flexible, on-demand resource allocation comprises three major product categories: namely IaaS, Platform as a Service (PaaS), and Software as a Service (SaaS) (Mell and Grance 2011). Keller and König (2014, p. 4) identify three recent trends in cloud computing that "are likely to transform the current cloud landscape":

- increasing standardization, especially viable in IaaS
- increasing SaaS specialization for particular user groups, such as private users or specific industries
- increasing actor dependencies.

These developments specifically occur in emerging cloud marketplaces (Keller and König 2014). Major cloud providers offer standardized products, such as virtual machines with a given operating system, CPU, RAM, and storage. However, especially in the IaaS context, the standardization of cloud computing fosters an oligopolistic market structure, in which the largest two providers (AWS and Microsoft) provide the deployment environment of about 70% of the current applications (Skyhigh Networks 2017). These companies profit from enormous economies of scale, which might, however, stall innovation and progress in the cloud market (Bestavros and Krieger 2014). Nevertheless, recent attempts, such as the Deutsche

Börse Cloud Exchange, the Cloud Commodities Exchange Group, and the Massachusetts Open Cloud Exchange, have opened the IaaS markets to smaller providers, thus increasing the market dynamics. Moreover, standardized *application programming interfaces* (API), which tools like Swagger or CloudStack use, enable the dynamic exchange of commoditized SaaS services, such as weather services (Lewis 2013; Loutas et al. 2011a; Loutas et al. 2011b).

2.2 Cloud Computing Spot Prices

In cloud computing, AWS first introduced spot prices for their computing service Amazon EC2 in 2009. AWS operates EC2 spot instances in 14 locations with about 40 products (Amazon Web Services 2017), which can substitute one another. As AWS' excess capacity, EC2 spot instances are normally cheaper than regular on-demand instances based on a fixed price (Kamiński and Szufel 2015). Similar to spot markets for stocks, electricity, and commodities, a market mechanism brings together demand (bids) and supply (offers) in a Vickrey auction to form EC2 spot prices (Cheng et al. 2016). However, AWS applies a hidden reserve price algorithm to artificially generate a linear dependency between the availability and the spot price that is consistent over multiple instance types and locations (Ben-Yehuda et al. 2013).

Currently, there are different research streams on cloud spot prices. One research stream applies reverse engineering for a better understanding of EC2 spot instances and to deconstruct AWS' spot pricing mechanism (e.g., Ben-Yehuda et al. 2013; Li et al. 2016a). These papers do not provide decision support algorithms. As prices differ between regions, a second research stream analyzes customer strategies to reduce costs by spatially distributing the use of spot instances (e.g., Cheng et al. 2016; Marathe et al. 2014). Since our objective is to study temporal instead of spatial flexibility, we are more closely related to a third research stream focusing on spot price prediction. For example, Baughman et al. (2018) propose a model to predict EC2 spot prices based on long/short-term memory recurrent neural networks. Khandelwal et al. (2017) propose a model based on random forest regression for predicting EC2 spot prices one day and one week ahead. These scholars demonstrate that their non-parametric machine learning approach outperforms previous approaches based on support vector machines (Arevalos et al. 2016) and artificial neural networks (Wallace et al. 2013). Cai et al. (2018) criticize several existing models for being static and neglecting the correlation of sequential cloud spot prices. Instead, these authors propose two Markov regime-switching autoregression models and one autoregressive integrated moving average model that integrate new observable information dynamically to adjust price predictions. These examples are just an excerpt from an extensive research stream, which is, nevertheless, inappropriate for our purposes. Although these studies present sophisticated models for spot price prediction based on (auto)regression and machine learning, their point estimators provide only limited decision support, as they do not consider the type of customer service request and the relevant optimization restrictions.

Vieira et al. (2015, p. 498) distinguish three categories of service requests: "fixed-time requests" without temporal flexibility (e.g., continuous monitoring tasks or websites), "floating-time requests" which can be interrupted and are temporally flexible, and "variable-time requests" which cannot be interrupted, but are temporally flexible. As we aim to quantify and exploit cloud customers' (short-term) temporal flexibility, we will not further consider *fixed-time requests*.

Research not only provides spot price predictions, but also decision support in terms of bidding strategies for floating-time and variable-time requests. *Floating-time requests* require cloud customers to apply complex check-pointing mechanisms and snapshots. Andrzejak et al. (2010) present a probabilistic model that employs temporal flexibility to optimize bidding strategies. By focusing on cost-reliability trade-offs and the selection of instance types, they conclude that cost savings negatively affect execution time (and

vice versa) and that switching from standard or high-memory to high-CPU instance types can save costs. Tang et al. (2012) and Tang et al. (2014) advance this approach by formulating a constrained Markov decision process based on linear programming. These authors improve Andrzejak et al.'s (2010) approach in terms of cost savings and execution time. In these three papers, the researchers set a price threshold and maximize the reliability of long-dated computations (2.6 to 22.6 hours) over a timeframe of several days. Zafer et al. (2012) extend these approaches by proposing a dynamic bidding strategy for floating-time requests with a specific deadline. While their suggested bidding strategy favors the use of EC2 spot instances due to their lower costs, it can only guarantee that jobs will be executed by a fixed deadline if it also uses EC2 on-demand instances.

We aim to contribute to the research of *variable-time requests* that must not be interrupted, such as MapReduce jobs (Dadashov et al. 2014) and other highly parallelized jobs (Kumar et al. 2017). Distributed analytics jobs, for example, those using Hadoop or Spark, are particularly suitable for variable-time requests (Kumar et al. 2017). Zheng et al. (2015) and Tamrakar et al. (2017) analyze the execution of MapReduce jobs, with the former concluding that using spot instances from different markets can reduce costs by 93% compared to regular on-demand cloud instances, but can also increase computation time by 15%. Zheng et al. (2015) and Zafer et al. (2012) model a fixed deadline, but can only guarantee this by using additional EC2 on-demand instances. In terms of the spot markets, they try to balance the trade-off between the costs and the reliability of the job execution.

Extending all previous literature on the topic, we contribute an approach that guarantees to execute variable-time requests in spot markets within a customer's temporal flexibility window. We design the approach to be easier to understand and implement than other approaches, because we reduce the decision complexity to "when to bid" (ignoring "how much to bid") by considering the expected spot price development. We focus on one instance type on one cloud spot market. In contrast to existing literature, we implicitly assume that a customer's bid is high enough for the job execution to be uninterruptible. This assumption is valid for Vickrey auctions, in which a bidder at most pays the common spot price instead of the bid. Our initial motivation also requires our approach to evaluate short-term temporal flexibility while explicitly considering uncertainty. We have therefore chosen to apply ROA, which explicitly suits this requirement (Kleinert and Stich 2010). Undertaking ROA requires the available distribution of possible future spot prices; we therefore need to model spot price development as a stochastic process instead of applying regression models that yield point estimators.

2.3 Real Options Analysis in Information Systems Research

ROA originated from financial option valuation with the aim to evaluate managerial action flexibility that takes uncertainty into consideration. Myers (1977, p. 163) introduced the term *real options* as "opportunities to purchase real assets on possible favorable terms." Real options comprise "discretionary decisions or rights, with no obligation, to acquire or exchange an asset for a specified alternative price" (Trigeorgis 1996, p. xi). IS researchers started applying ROA in the 1990s in order to evaluate managerial flexibility in information technology (IT) investments (Ullrich 2013). Benaroch and Kauffman (1999), for example, study the application of discrete-time and continuous-time option pricing models for evaluating investments in IT infrastructure, emerging technology, application design prototyping, and technology-as-products. These scholars conclude that managers can apply traditional option pricing models to non-traded IT assets without loss of validity. Subsequently, Benaroch and Kauffman (2000) examine a case in order to validate the added value of deferral options for strategic IT investments and elaborate on ROA's advantages instead of traditional IT investment evaluation methods. ROA's application in IS research focusses mainly

on IT investment decisions in general (Chen et al. 2009) or on specific technologies (Lee and Lee 2011; Nwankpa et al. 2016; Wu et al. 2009; Zimmermann et al. 2016).

In our targeted cloud computing research domain, authors apply ROA to migration decisions (Naldi and Mastroeni 2016; Yam et al. 2011), the extension of cloud resources (Alzaghoul and Bahsoon 2013), investment deferral (Alzaghoul and Bahsoon 2014), termination management (Jede and Teuteberg 2016), and risk management regarding cloud services' availability (Allenotor and Thulasiram 2014). Compared to traditional IT investments, infrastructure services in cloud computing are more separable, meeting the ROA requirement of "complete markets" better (Ullrich 2013, p. 335). In line with the development of cloud exchanges, Meinl and Neumann (2009) propose establishing a contract market to enable grid and cloud services' customers and providers to trade real options to reserve resources in advance. Náplava (2016) uses ROA to evaluate external IaaS's additional flexibility compared to that of on-premise solutions. Klaus et al. (2014) develop a model for service providers that evaluates an option to shift excess demand for (e.g., cloud) services to external vendors. This approach determines the business value of shifting flexibility, which decision makers can subsequently use to justify investments in required IS infrastructure.

Our literature review demonstrates ROA applications in IT project and cloud computing business cases. To the best of our knowledge, ROA has not yet been applied to support a cloud service purchase by means of variable-time requests. Kumar et al.'s (2017) research taxonomy of bidding strategy design for cloud spot markets does not list ROA as an already researched method, thus confirming our observation.

Nonetheless, we can build on ROA from other domains. Fridgen et al. (2016) study intraday load-shifting flexibility in the electricity spot market context. These authors propose an ROA-based algorithm to utilize temporal flexibility, adapting and applying the Cox et al. (1979) binomial tree model for discrete-time option valuation. Similar to our approach, they model temporal flexibility as a *deferral option*: Although purchase before a specified deadline is obligatory, this option gives customers the flexibility to decide on their purchase time in order to exploit the cost savings potential of volatile market prices. Although we adapt their model in some respects, we apply, evaluate, and compare multiple discrete-time approaches to ROA in the light of our research question.

3 Cloud Spot Market Data Analysis

We base our study on a time series of Amazon EC2 spot market data, which comprises prices and the associated price changes. Encompassing two years of cloud spot market operation, the data span the period January 1, 2015 to December 30, 2016. We acknowledge Spot Price Archive (Javadi et al. 2011), which downloaded a large dataset ranging from January 2009 to December 2016 via the Amazon EC2 API, as the source of this series of spot prices. More precisely, we analyze historical data from the EC2 spot instance "m1.xlarge" hosted in a North Virginia data center ("us-east-1" region). This type of cloud service encompasses four virtual cores, 15 gigabytes of RAM, 350 gigabytes of hard-disk space, and high network performance (Amazon Web Services 2017).

In Figure 1, we provide an example of the hourly statistics of historical 2016 data.

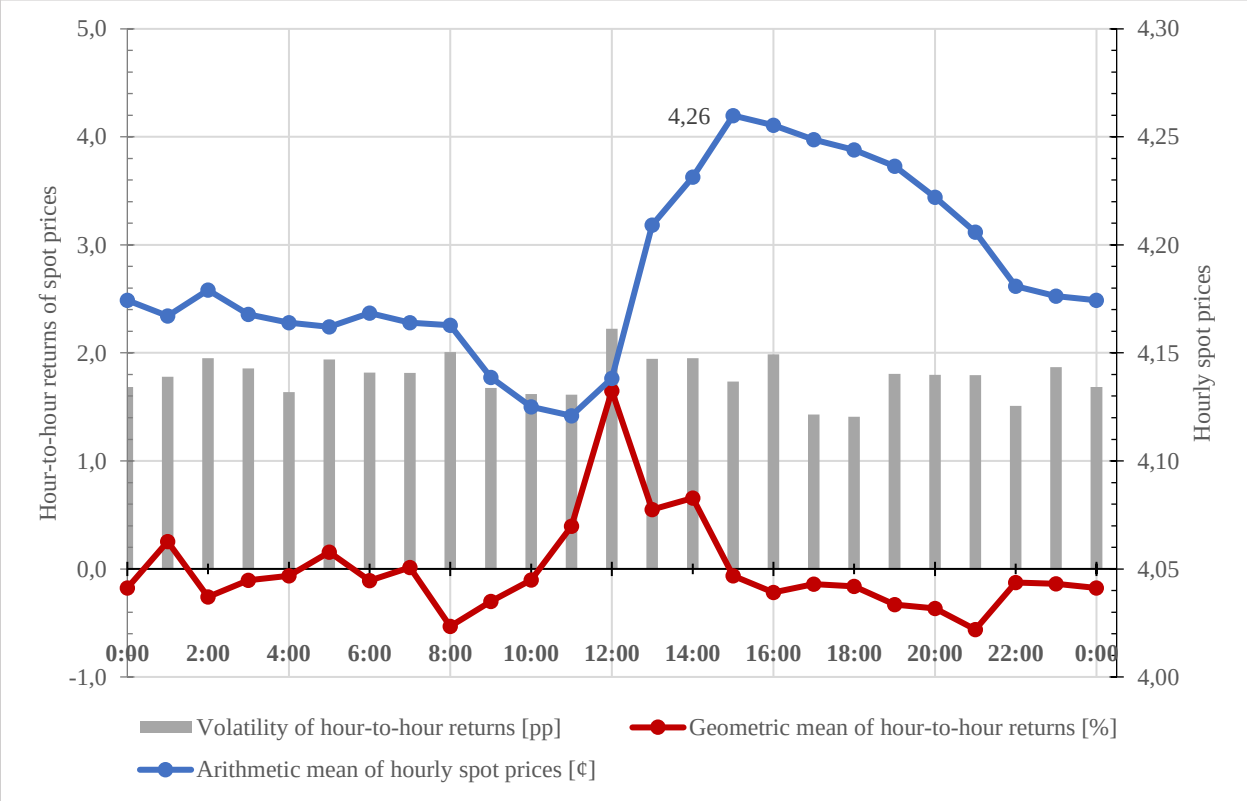


Figure 1. Hourly Statistics of Amazon EC2 Spot Prices

In formulae, we denote references to averaged historical input with a circumflex (ˆ) and the cloud spot price at a given time of day t with S(t). We compute the historical mean cloud spot price $\hat{S}(t)$ at time t:

$$\hat{S}(t) = \frac{\sum_{i=1}^{n} S(t)_i}{n} \qquad (1)$$

More precisely, $\hat{S}(t)$ is the *arithmetic mean* of n historically observed prices at the time of day t. Further, R(t) is the spot price change, or *return*, from t to $t + 1$, which we express relatively:

$$R(t) = \frac{S(t+1)}{S(t)} - 1 \qquad (2)$$

We compute the historical mean return $\widehat{R}(t)$ from n historically observed cloud spot returns:

$$\widehat{R}(t) = \left((1 + R(t)_1) * (1 + R(t)_2) * \dots * (1 + R(t)_n)\right)^{\frac{1}{n}} - 1 \quad (3)$$

Because single returns may be interdependent growth factors, we choose a *geometric mean* over an arithmetic mean, which could yield false results in this case. More precisely, if spot prices at a specific time of day follow a positive or negative growth trend (increase or decrease, on average, over some days, weeks, or months), applying an arithmetic mean of historical returns to forecast spot prices is likely to overestimate the expected developments, especially regarding more than one estimation period (Amenc and Le Sourd 2003).

In continuation, $\widehat{\sigma}(t)$ is the historical standard deviation, or *volatility*, of cloud spot returns. We compute $\widehat{\sigma}(t)$ as the geometric standard deviation:

$$\widehat{\sigma}(t) = e^{\sqrt{\frac{1}{n} * \sum_{i=1}^{n} \left(\ln\left(\frac{1+R(t)_i}{1+\widehat{R}(t)}\right)\right)^2}} \quad (4)$$

Figure 1 indicates that EC2 cloud spot prices for a reference timespan of 24 months are subject to time-of-day-specific patterns of mean prices, mean returns, and return volatilities. We therefore examine the following hypothesis:

Hypothesis 1: One should extend traditional ROA approaches with time-of-day-specific spot price patterns to optimally exploit the monetary value of short-term temporal flexibility in cloud computing demand.

We test Hypothesis 1 by comparing ROA approaches with and without consideration of time-of-day-specific spot price patterns. Moreover, we verify our modeling decision to apply ROA by examing the following hypothesis:

Hypothesis 2: One should not only model the time-of-day-specific mean prices (or returns), but also the return volatilities to optimally exploit the monetary value of short-term temporal flexibility in cloud computing demand.

We test Hypothesis 2 by applying naive expectation optimization as an alternative to ROA. In the following section, we introduce the respective models. Thereafter we evaluate the models on historical EC2 spot market data.

4 Model Development

4.1 Discrete-Time Spot Price Modeling

In this section, we present multiple approaches to support decisions to utilize temporal flexibility in cloud spot markets. We assume a situation in which a customer is temporally flexible (e.g., for some hours) and aims for the lowest possible price in this time window. However, an individual deadline indicating the time at which the customer requires the cloud services at the latest, limits temporal flexibility. Hence, the customer's decision problem is, given the deadline, to defer demand up to the (ex-ante) optimal (cost-minimal) point in time.

Employing ROA, we can model customers' temporal flexibility to defer cloud demand as a deferral option, because they can sell their right to instantly purchase cloud services. This deferral option's value depends specifically on cloud spot prices' (the option's *underlying*) stochastic development and the customer's

deadline (after the deferral option's expiration) at which purchase would be obligatory. The customer may exercise the option (i.e., to purchase cloud services) only once at an arbitrary decision point in time. The deferral option is therefore similar to an American call option in capital markets.

Assumption 1: Until the deferral option expires, a customer can decide in discrete time increments of equal length whether to exercise the option or not.

In Assumption 1 we limit the decision points in time to a finite and equally distributed number for simplicity's sake. Although approaches that allow continuous-time option pricing and decision making (e.g., Black and Scholes 1973) offer more freedom of action, which would make them preferable, they are rather complex. In particular, there are as yet no closed-form solutions for the continuous-time pricing of American call options under consideration of time-of-day-specific mean prices, returns, and return volatilities. Instead, we research discrete-time approaches that are simple, yet accurate enough to considerably exploit a temporally flexible customer's savings potential. To test both hypotheses in consideration of Assumption 1, we have chosen to adapt, apply, and compare the following discrete-time approaches to customer decision support in cloud spot markets:

1. The binomial tree approach of Cox et al. (1979)
2. The binomial tree approach of Tian (1993)
3. Expectation optimization

Cox et al. (1979) were the first authors to develop a discrete-time version of the famous option pricing model by Black and Scholes (1973). They modeled the stochastic movements of an underlying and a matching option as a binomial tree. They prove that this model converges toward the Black-Scholes formula for decreasing-length time increments. Tian (1993) modified Cox et al.'s (1979) binomial tree formulae by matching the discrete-time process's skewness with the continuous-time process. Via numerical simulations on stock prices, Tian demonstrates that this model improves the accuracy of the convergence toward the Black-Scholes model. Although there are other derivatives of Cox et al.'s option pricing model (e.g., Amin 1991; Jarrow and Rudd 1983; Leisen and Reimer 1996), our approaches already provide valuable insights into discrete-time ROA's potential as a tool for decision support in cloud spot markets. Whereas Cox et al. (1979) and Tian (1993) do not model the time-of-day-specific patterns of their underlying, we apply both approaches in their native form and with this model extension (which is required to test Hypothesis 1).

4.2 Binomial Tree Approaches without Time-of-Day Specific Patterns

In the following, we present Cox et al.'s (1979) and Tian's (1993) traditional approaches without consideration of the time-of-day-specific spot price patterns, which we introduce afterward.

Assumption 2: Cloud spot prices are log-normally distributed, while the returns of cloud spot prices are normally distributed.

Following Mazzucco and Dumas (2011), we assume that the returns of cloud spot prices are normally distributed (and that cloud spot prices are therefore log-normally distributed). In respect of EC2 spot prices, this assumption is "adequate but not perfect, as the distribution of the spot prices is more heavily-tailed" (Mazzucco and Dumas 2011, p. 297).

Assumption 3: Cloud customers are risk-neutral in their decisions.

Since both Cox et al. (1979) and Tian (1993) develop their approaches by assuming normally distributed returns and risk-neutral decision makers, we also require these rather technical assumptions. For the sake of our model's simplicity and in the light of our valid results, we consider these limitations adequate.

Cox et al. (1979) and Tian (1993) apply a binomial tree to model their underlying's stochastic process. The tree starts at the current point in time ($t = t_0 = 0$) before forking in discrete time increments into future nodes (i.e., future price levels) up to the option's expiration (denoted $t = T$). Consequently, at each node, with the exception of end nodes, the underlying is expected to move either in an upward or a downward direction. Cox et al. (1979) and Tian (1993) describe the binomial tree by means of the following parameters: $u \leq 1$ and $d \leq 1$ are constant factors for the (expected) extent of the underlying's upward and downward movements within one time increment. Both approaches depend on the historical return volatility $\hat{\sigma}$ and the risk-free interest rate r_f (which are both constant in these traditional models). A condition is that $u * d = 1$ and $u > 1 + r_f > d$. Moreover, $p \leq 1$ is the constant probability of the underlying moving in an upward direction. Conversely, $1 - p$ is the constant probability of a downward movement. The approaches by Cox et al. (1979) and Tian (1993) suggest the following formulae to derive the expected price development in an arbitrary time increment t to $t + 1$:

$$S(t+1)_u = S(t) * u \tag{5}$$

$$S(t+1)_d = S(t) * d \tag{6}$$

In Figure 2, we illustrate an exemplary binomial tree for our underlying (cloud spot prices).

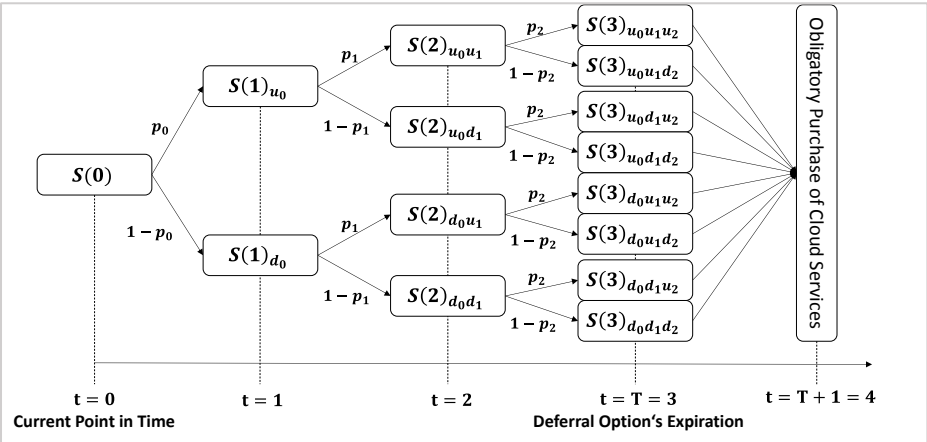


Figure 2. Exemplary Binomial Tree for a Deferral Option with Three Remaining Time Increments

Under consideration of Assumptions 2 and 3, we can apply Cox et al.'s (1979) formulae:

$$u = e^{\hat{\sigma} * \sqrt{\Delta t}} \tag{7}$$

$$d = e^{-\hat{\sigma} * \sqrt{\Delta t}} \tag{8}$$

$$p = \frac{e^{r_f * \Delta t} - d}{u - d} \tag{9}$$

The parameter Δt quantifies the time increments between the decision nodes in the binomial tree, which is $\Delta t = 1$ in our case. Similarly, we can apply Tian's (1993) formulae, which (only) differ in terms of the u and d:

$$u = \frac{V}{2} * e^{r_f * \Delta t} * \left(V + 1 + \sqrt{V^2 + 2V - 3}\right) \quad \text{with } V = e^{\hat{\sigma}^2 * \Delta t} \tag{10}$$

$$d = \frac{V}{2} * e^{r_f * \Delta t} * \left(V + 1 - \sqrt{V^2 + 2V - 3}\right) \quad \text{with } V = e^{\hat{\sigma}^2 * \Delta t} \tag{11}$$

In both approaches, modeling the underlying's binomial tree is the prerequisite for option pricing. In each of the tree's nodes, a cloud customer must decide on whether to exercise the deferral option (i.e., to purchase cloud services) or not (i.e., to wait for another time increment). After exercising the deferral option, the optimization terminates. If the customer does not exercise the deferral option at time $t = T$ at the latest, he/she reaches the individual deadline and must purchase cloud services in the next discrete time step ($t = T + 1$). Technically speaking, modeling a deadline is already an extension of Cox et al.'s (1979) and Tian's (1993) traditional models, which Fridgen et al. (2016) introduced for the former approach. Both approaches start option pricing by analyzing the possible exercise values in the binomial tree's end nodes:

$$C(T) = \max\{X - S(T); 0\} \tag{12}$$

$S(T)$ is the expected cloud spot price at a specific end node in the binomial tree at time T. X is the exercise or strike price of the deferral option, which we explain later. If X is greater than $S(T)$, exercising the option in T is preferable, leaving the deferral option with a value greater than zero; however, if it is not, the customer should wait for one time increment and purchase cloud services at the individual deadline.

For every decision node that is $n \in \{1, \dots, T\}$ periods before T, the customer can compute the deferral option's value by applying the following formula by Cox et al. (1979):

$$C(T - n) = \max\begin{Bmatrix} X - S(T - n); \\ p * C(T - n + 1) + (1 - p) * C(T - n + 1) \end{Bmatrix} \tag{13}$$

Except for the end nodes in T, each decision node receives two values: that of the immediate cloud service purchase (i.e., the deferral option's exertion at that time) and that of deferring the purchasing decision for (at least) one time increment (i.e., the "time value" of exercising it later). The latter requires an algorithm for a probability-weighted valuation, since, from a single decision node's perspective, the tree forks into an upward and downward direction. The maximum of both values constitutes the deferral option's value at the relevant decision node. Note that since both approaches conduct the option pricing from the end nodes in T to root t_0, computing the time values of every decision node for $t = T - n$ can draw on already computed option values in $t = T - n + 1$. The algorithm terminates as soon as it obtains the deferral option's value in t_0 (i.e., the current point in time). Cloud customers can now compare the value of "exercising immediately" and "exercising later," deciding accordingly. If customers decide to wait for the next time increment, they need to update the observable price information and repeat the binomial tree construction and the option evaluation. Note that if customers can only purchase cloud services at certain times (i.e., at certain decision nodes), the deferral option complies with a Bermudan call option (or even with a European call option if they can only decide in $t = T$). Modeling a Bermudan (or European) call option only requires modifying Equation 13 for non-decision nodes by removing the right and value of the immediate exertion.

4.3 Modeling Time-of-Day-Specific Patterns

We follow Fridgen et al. (2016) as follows to model the time-of-day-specific spot price patterns in order to test Hypothesis 1:

- Since we evaluate the monetary value of temporal flexibility in the short term (i.e., a maximum of several hours), the risk-free interest rate is insignificantly low, and we can set $r_f = 0$.

- We consider the time-of-day-specific spot price patterns by assuming *mean reversion*, i.e., for each discrete time step, the spot price is expected to move ("revert") to either the mean price level or according to the mean return, historically observed at the respective time of day. The same applies to volatilities.
- In keeping with both the traditional models created to evaluate options in capital markets, we treat these mean-reverting movements like discrete dividend payments.
- We model binomial parameters time-dependently, i.e., u(t), d(t), and p(t), because of the time-of-day-specific volatility patterns $\widehat{\sigma}(t)$.

While Fridgen et al. (2016) extend the approach by Cox et al. (1979) with mean reversion to the time-of-day-specific mean price and volatility patterns, we also apply Tian's (1993) model and mean reversion to the time-of-day-specific mean return patterns. Financial asset pricing usually exhibits stationary mean returns, but non-stationary mean prices (Rossi and Spazzini 2014), which makes the former preferable for deriving predictions in these markets. Stationarity makes historical data a more appropriate estimator of future movements. As we could not find any related work concerned with stationarity analysis in cloud spot markets, we apply both approaches to model time-of-day-specific patterns and compare them.

In the following, we present relevant extensions of Equations 5 and 6 given the time-of-day-specific mean prices and returns.

Equations 5 and 6 with time-of-day-specific **mean prices** (Fridgen et al. 2016):

$$S(t+1)_{u_t} = S(t) * u(t) + \theta * (\hat{S}(t+1) - S(t)) \quad (14)$$

$$S(t+1)_{d_t} = S(t) * d(t) + \theta * (\hat{S}(t+1) - S(t)) \quad (15)$$

Equations 5 and 6 with time-of-day-specific **mean returns**:

$$S(t+1)_{u_t} = S(t) * u(t) + S(t) * \theta * \widehat{R}(t) \quad (16)$$

$$S(t+1)_{d_t} = S(t) * d(t) + S(t) * \theta * \widehat{R}(t) \quad (17)$$

Parameter $\theta \in [0,1]$ expresses the mean-reversion speed, controlling the speed with which the process reverts to the time-of-day-specific mean price or return patterns. A mean-reversion speed of $\theta = 1$ implies complete mean reversion during one time increment. In contrast, $\theta = 0$ implies no mean reversion.

Additionally, we model the strike price $X(t)$ as the (time-dependent) opportunity costs of exercising the option during the flexibility window before the deadline. Hence, $X(t)$ depicts the expected cloud spot price if the customer were to wait until the obligatory purchase in $T+1$, i.e., $X(t) = S(T+1)$. The deferral option can therefore be interpreted as an option to buy before the individual deadline at relevant opportunity costs $X(t)$. At every decision node in the tree, we compute $X(t)$ as follows (for, respectively, the mean prices and the returns):

$$X(t) = S(t) + \theta * \left(\hat{S}(t+1) - S(t)\right) + \cdots + \theta * \left(\hat{S}(T+1) - S(T)\right) \quad (18)$$

$$X(t) = S(t) + \theta * S(t) * \widehat{R}(t) + \cdots + \theta * S(T) * \widehat{R}(T) \quad (19)$$

Technically, common option pricing approaches assume a constant strike price and ROA literature has been criticized for violating this assumption (Ullrich 2013). Fridgen et al. (2016) therefore keep the strike price constant; however, they sacrifice savings by not allowing an update of the strike price when receiving new

market information. If the strike price can develop stochastically, an option pricing approach must explicitly take the relevant process for deriving the option's value correctly into account. The following reasoning allows us to apply a valid stochastic process for the strike price: As the strike price only depends on one stochastic factor $S(t)$, we obtain exactly one value for $X(t)$ at each decision node in $S(t)$'s binomial tree. Note that our definition of opportunity costs $X(t)$ does not comprise a further inconvenience regarding the customer's willingness to defer the purchase of cloud services, but only takes cost differences into account due to the volatile spot prices and the individual flexibility window.

Table 1 summarizes all the real options approaches that we adapt, apply, and compare.

	Traditional (without time-of-day-specific patterns)	**With time-of-day-specific price patterns**	**With time-of-day-specific return patterns**
Cox et al. (1979)	✓	✓ (Fridgen et al. 2016)	✓
Tian (1993)	✓	✓	✓

Table 1. Real Options Approaches Applied to Schedule Flexible Demand in Cloud Spot Markets

When one applies Cox et al.'s (1979) and Tian 's (1993) traditional approaches, determining the optimal point in time to purchase cloud services is trivial. Following established option pricing theory, by early exercising American call options on underlying assets that pay no dividends (in our case, that do not consider the time-of-day-specific patterns) cannot be optimal (Hull 2014; van Hulle 1988). The same would apply to continuous-time models, such as those of Black and Scholes (1973). Both approaches would therefore not early exercise the option, but instead wait until $t = T$ to decide to either purchase at that time (at a price $S(T)$) or to wait for the deadline at $t = T + 1$ to purchase at a price $S(T + 1)$.

In addition to our real options approaches, we apply naive expectation optimization to test Hypothesis 2. In t_0, naive expectation optimization compares the currently observable price information with the expected prices in each upcoming time step in the flexibility window. The expected prices equal the historically recorded mean prices at the relevant time of day. Expectation optimization suggests that in order to purchase cloud services, customers should choose the time with the lowest expected spot price. Compared to our real options approaches, this naive approach does not consider return volatilities.

5 Evaluation and Sensitivity Analysis

Simulations are a rigorous evaluation technique (Gregor and Hevner 2013). We therefore conducted historical simulations on our EC2 dataset (Section 3) to evaluate our approaches regarding their suitability to quantify and exploit the monetary value of short-term temporal flexibility in cloud computing demand. We implemented our approaches by means of Microsoft Excel with Visual Basic for Application macros and performed statistical tests in R. In randomly assembled scenarios that could have occurred in the past, we analyzed how well our approaches would have realized spot price savings. Our macros followed the following steps in each simulation run:

1. Select an approach (cf. Table 1 or naive expectation optimization).
2. Select a random date and time of day from the historical time series as the starting point (between January 1, 2015 and December 30, 2016).

3. Select a random temporal flexibility window TFW $\in \{1,2,\dots,12\}$ [increments]. Initially, the increment length IL (i.e., the time between two decision nodes) was constant at IL $= 60$ [min].
4. For real options approaches with the time-of-day-specific patterns: Select a random mean-reversion speed $\theta \in \{0, 0.25, 0.5, 0.75, 1\}$ and a reference timespan RTS $\in \{7, 30, 60, 90\}$ [days]. From the chosen starting point in time (2.), look back RTS days in the past to build expectations of the time-of-day-specific price (or return) and the volatility patterns.
5. Run the specific approach's algorithm.
6. After termination (i.e., after the purchase of cloud services), compare the purchase price to the spot price S_0 that was viable at the beginning of the TFW, and which a purchase without temporal flexibility would have realized. Compute the realized absolute and relative savings. With this information, divide the realized absolute savings by the maximum possible absolute savings within the TFW (which the algorithm would have obtained if perfect information were available), in order to compute the exploitation of the existing savings potential.

We distinguish two types of parameters: exogenous (scenario) and endogenous (model) parameters. IL, TFW, and starting time are exogenous parameters drawn to construct a simulation scenario. In contrast, approach selection, RTS, and θ are endogenous parameters. Both parameter types differ in the cloud customers' possibility to freely select endogenous parameters, although they might not be able to influence the exogenous parameters. Hence, in order to maximize their savings, cloud customers try to select endogenous parameters optimally. We conduct and analyze the results of six million simulation scenarios, one million for each approach, which approximates the maximum number of rows in our Microsoft Excel worksheets. Since Cox et al.'s (1979) and Tian's (1993) traditional approaches optimize identically (cf. Section 4.3), we summarize both models in one approach. Table 2 depicts our results.

	Savings with random parameters			**Savings after configuration with optimal θ and RTS**		
	Averaged absolute savings to S_0 [¢]	Averaged relative savings to S_0 [%]	Exploitation of savings potential [%]	Averaged absolute savings to S_0 [¢]	Averaged relative savings to S_0 [%]	Exploitation of savings potential [%]
I. Cox et al. (1979) with price patterns	0.03649	0.80813	21.76075	$\theta = 1$, RTS $= 7$d		
				0.06857	1.51294	40.45341
II. Cox et al. (1979) with return patterns	0.05682	1.25749	33.65950	$\theta = 0.25$, RTS $= 30$d		
				0.06474	1.43051	37.49308
III. Tian (1993) with price patterns	0.03761	0.83261	22.26482	$\theta = 1$, RTS $= 7$d		
				0.07337	1.61352	40.91032
IV. Tian (1993) with return patterns	0.05707	1.26403	33.93849	$\theta = 0$, RTS $= 30$d		
				0.06763	1.49416	38.53289
V. Traditional Cox et al. (1979) and Tian (1993)	0.00929	0.20560	5.51305	Not available		
VI. Expectation Optimization	0.05572	1.23367	33.08806	Not available		

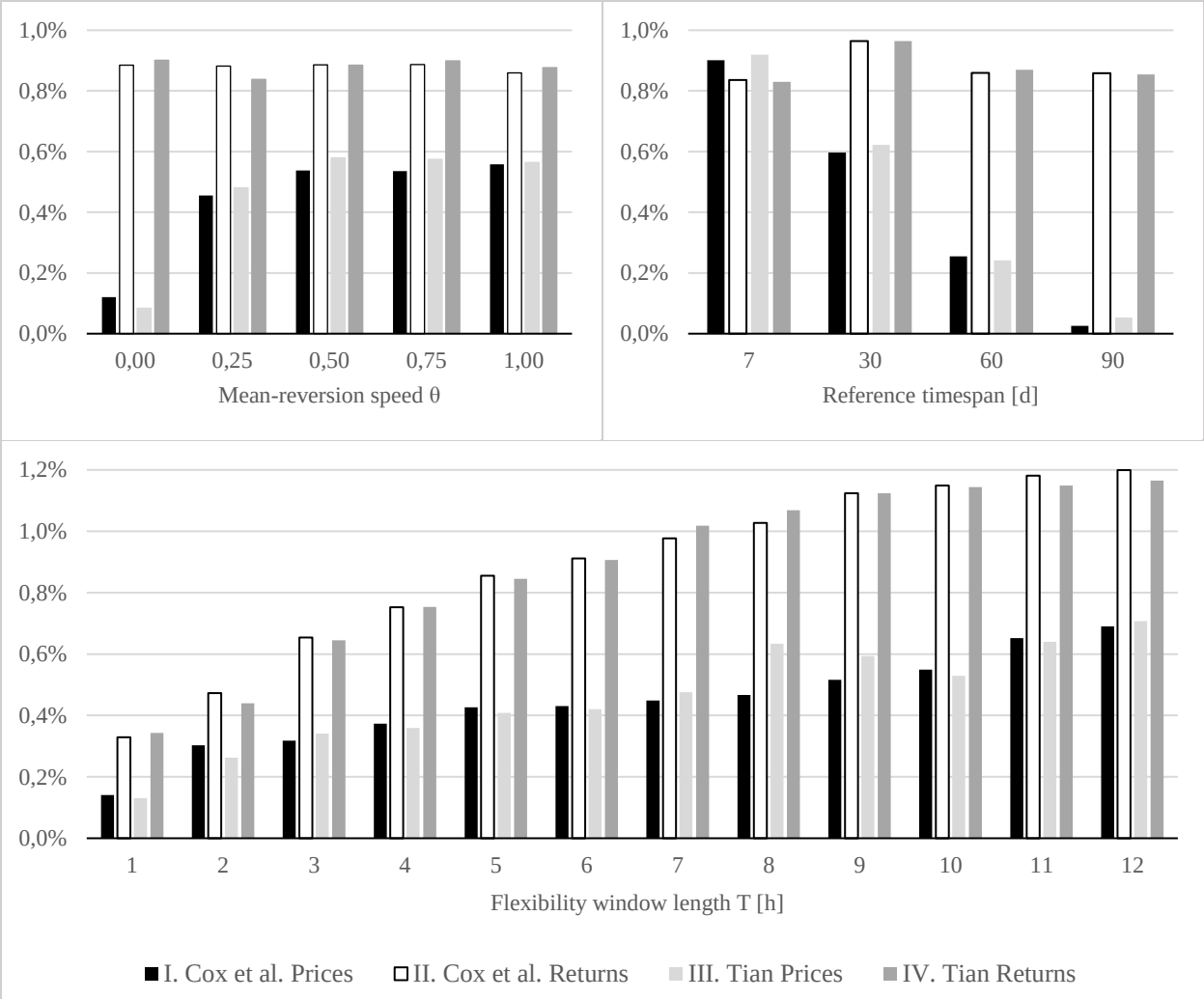
Two-sample t-test: Reject H_0 hypothesis that the mean savings of V ≥ the mean savings of I–IV with optimal θ and RTS → approaches I–IV preferable***
Two-sample t-test: Reject H_0 hypothesis that the mean savings of VI ≥ the mean savings of I–IV with optimal θ and RTS → approaches I–IV preferable***
*** represents a significance level of 0.1%, ** a significance level of 1%, and * a significance level of 5%

Table 2. Evaluation Results of Applied Approaches before and after Configuration of Endogenous Model Parameters

Overall, the results favor Hypotheses 1 and 2. More precisely, statistical two-sample t-tests indicate maintaining the null hypothesis that, after configuration, approaches I–IV yield superior averaged relative savings and exploit more savings potentials than the traditional approaches (V) and the expectation optimization (VI). In contrast to approaches I–IV, V does not model mean reversion, approach VI does not model volatility, and approaches V and VI are impossible to configure without parameters θ and RTS.

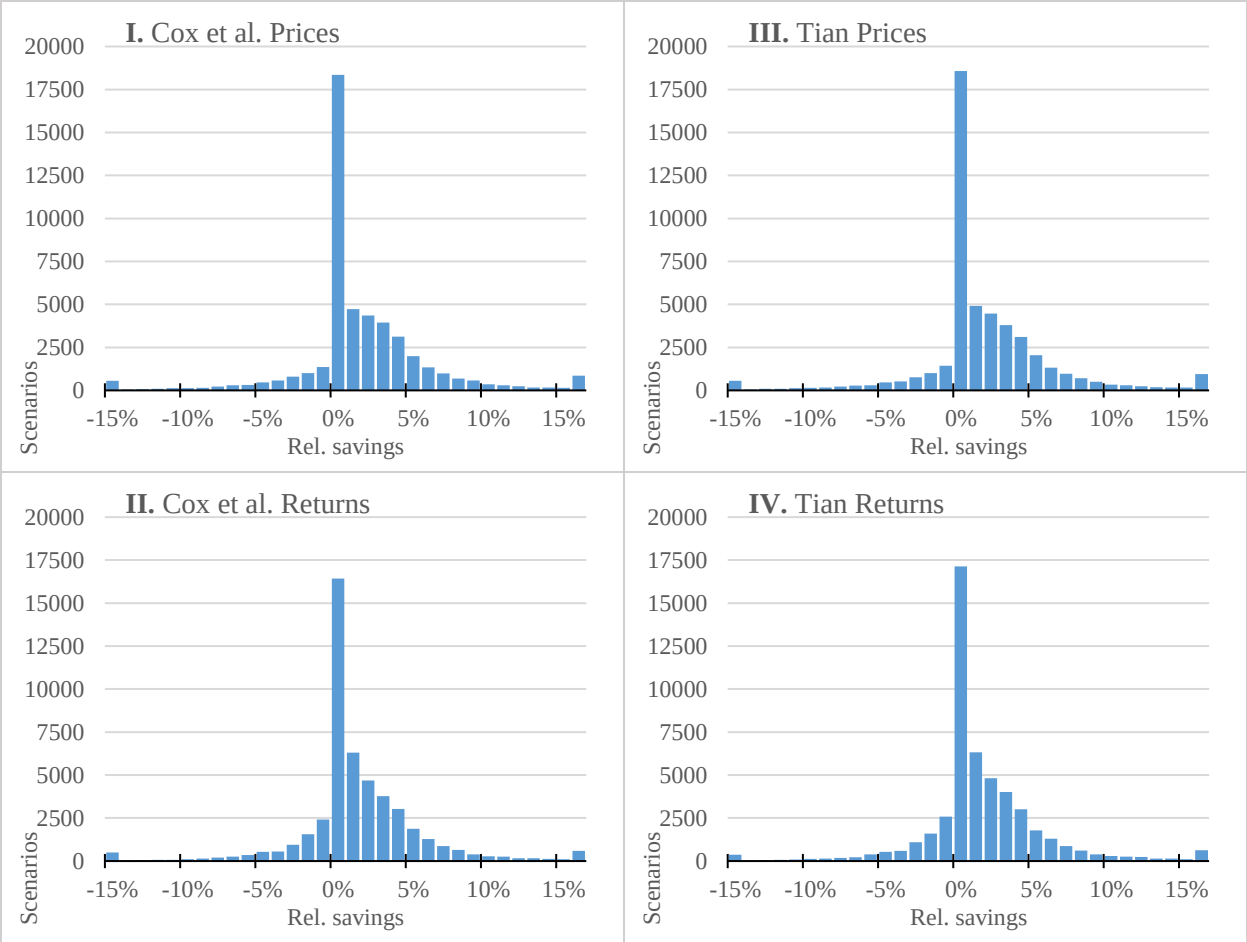
In respect of arbitrary random parameters, Table 2 illustrates that approaches II and IV yield superior averaged savings compared to approaches I and III. However, as this relationship reverses when configuring all four approaches with optimal θ and RTS, the performances of approaches I and III are comparatively more dependent on their parameters. In Figure 3, we show how the averaged relative savings reacted to altering parameters (*univariate sensitivity*).

Figure 3 indicates that the performance of approaches I and III depends significantly on the selection of θ and RTS. More precisely, the performance depends strongly on recent historical price information (shorter RTS), which indicates fast changing price levels in our EC2 dataset. Moreover, since a higher θ improves the results significantly, historical price information seems to be a valuable predictor. The performance of approaches II and IV also depends significantly on the RTS selection. As a longer RTS is optimal in this case, our dataset shows slower changing return levels than price levels. The insignificance of θ indicates that relative savings depend less on the approaches' capability to predict the time-of-day-specific return levels. A longer TFW increases the option values by increasing the action flexibility (Hull 2014), which is in line with common option pricing theory. Figure 4 uses histograms to illustrate these four approaches (after configuration with optimal parameters).



Two-sample t-test: Reject H_0 hypothesis regarding I*** & III*** that the mean savings of ($\theta < 0.5$) ≥ the mean savings of ($\theta \geq 0.5$) → higher θ preferable
Two-sample t-test: Cannot reject H_0 hypothesis regarding II & IV that the mean savings of ($\theta < 0.5$) ≥ the mean savings of ($\theta \geq 0.5$) (no significance)
Two-sample t-test: Reject H_0 hypothesis regarding I*** & III*** that the mean savings of (RTS > 7) ≥ the mean savings of (RTS = 7) → shorter RTS preferable
Two-sample t-test: Cannot reject H_0 hypothesis regarding I & III that the mean savings of (RTS ≠ 30) ≥ the mean savings of (RTS = 30) (no significance)
Two-sample t-test: Cannot reject H_0 hypothesis regarding II & IV that the mean savings of (RTS > 7) ≥ the mean savings of (RTS = 7) (no significance)
Two-sample t-test: Reject H_0 hypothesis regarding II*** & IV*** that the mean savings of (RTS ≠ 30) ≥ the mean savings of (RTS = 30) → RTS = 30 preferable
Two-sample t-test: Reject H_0 hypothesis regarding I–IV*** that the mean savings of (TFW ≤ 6) ≥ the mean savings of (TFW > 6) → longer TFW preferable
*** represents a significance level of 0.1%, ** a significance level of 1%, and * a significance level of 5%

Figure 3. Univariate Parameter Sensitivity of Averaged Relative Savings for Approaches I–IV

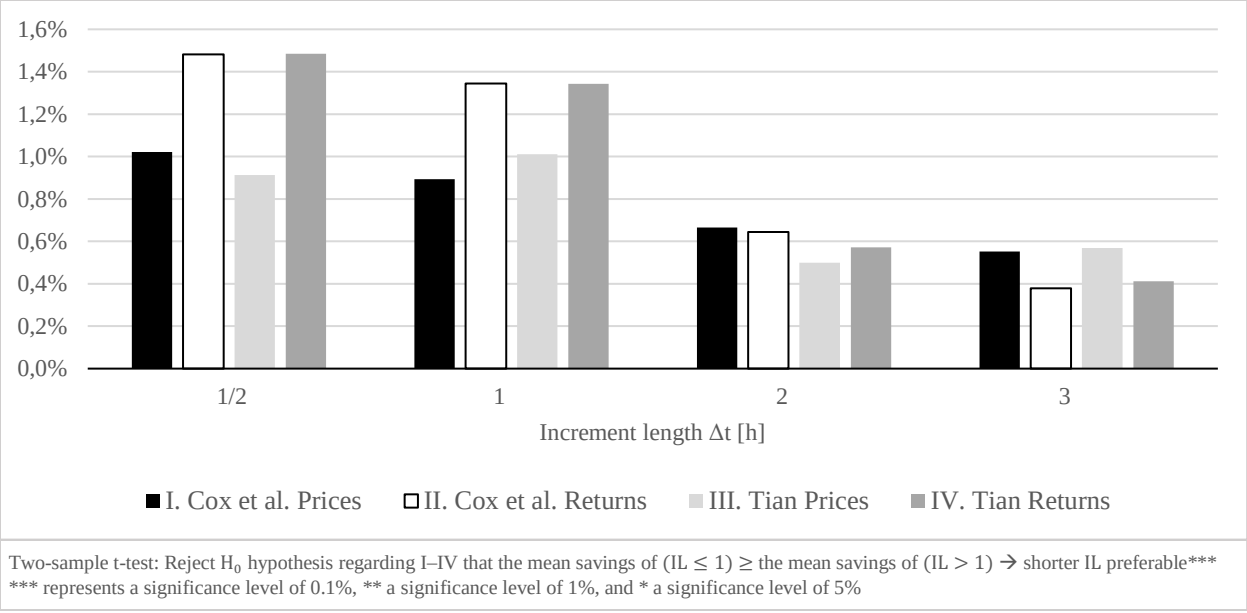


Two-sample t-test: Cannot reject H_0 hypothesis that the mean savings of Cox models ≥ the mean savings of Tian models (no significance)
Two-sample t-test: Reject H_0 hypothesis that the mean savings of return models ≥ the mean savings of price models → price models preferable***
*** represents a significance level of 0.1%, ** a significance level of 1%, and * a significance level of 5%

Figure 4. Histograms of Relative Savings for Approaches I–IV with Optimal θ and RTS

Figure 4 indicates that modeling time-of-day-specific price patterns instead of returns patterns is preferable (but only when configuring these models). According to Table 2, applying approaches following Tian (1993) instead of those following Cox et al. (1979) is preferable, although not statistically significantly. The Tian (1993) approaches may be slightly better performing due to the increasing accuracy of their convergence toward the Black-Scholes model (cf. Section 4.1). The better performance of modeling time-of-day-specific price patterns indicates that historical price information is a better estimator of spot price development over a few hours than return information. However, as approaches I and III's performances decline strongly with longer RTSs, this relation might reverse with longer TFWs (e.g., several weeks). Future research could analyze this hypothesis.

Finally, we run another one million simulated scenarios to test approaches I–IV's sensitivity to IL. We therefore randomize IL $\in \{30,60,120,180\}$ [min], while we keep TFW $= 6$h (a multiple of all IL) and the unconfigured parameters. Figure 5 show that longer ILs tend to yield lower averaged relative savings. This observation is plausible, as a longer IL within a constant TFW reduces the number of decision nodes in the binomial tree and, therefore, the flexibility of action to react to short-term spot price development.



Two-sample t-test: Reject H_0 hypothesis regarding I–IV that the mean savings of (IL ≤ 1) ≥ the mean savings of (IL > 1) → shorter IL preferable***
*** represents a significance level of 0.1%, ** a significance level of 1%, and * a significance level of 5%

Figure 5. Univariate Sensitivity of Averaged Relative Savings to Interval Length for Approaches I–IV

6 Discussion

Our evaluation results could lead to the assumption that an extension of the Tian (1993) model with a mean reversion to time-of-day-specific price patterns is preferable. Such a generalized assumption is not, however, valid, because our results are strongly dependent on our dataset of a single Amazon EC2 spot instance in a specific location, and on our chosen simulation parameters. We actually evaluated representative scenarios and parameter sets to demonstrate that ROA can be a suitable decision support method when customers, given their temporal flexibility and the uncertain spot price development, wish to purchase cloud services at minimal costs.

As a measure of uncertainty, volatility increases a real option's value (Hull 2014). Lower volatility, however, decreases temporal flexibility's value, because it lets one expect fewer savings from spot price movement. When applying ROA to our EC2 dataset, we observed that its return volatilities yielded rather low savings. More precisely, our configured approaches I–IV's relative savings averaged about 1.5 percent. However, this is already equal to exploiting about 40 percent of the existing savings potential (on average, cf. Table 2).

Nonetheless, our results are especially relevant for the following three reasons: First, cloud services are becoming cost-intensive for many companies. For example, if Snap Inc., which recently announced that it would spend $2 billion on Google cloud services over a five-year period (US SEC 2017), achieved realizable savings of 1.5 percent, this would amount to an absolute amount of $6 million per year. Second, other cloud spot instances exhibit higher return volatilities (Ekwe-Ekwe and Barker 2018) and, therefore, higher savings potentials than the one referred to in our dataset. Future research should therefore analyze and compare different cloud spot instances to identify promising application scenarios for our ROA. Third, we expect the return volatilities in multiple cloud spot markets to increase in the future, because the rapid standardization of cloud services should liberalize the market structures further. More cloud providers offering spot prices should also increase the competition and liquidity on the supply side. On the demand

side, recent trends like *cloud bursting*, which prevents peak load in companies' data centers by adding external cloud resources (Lilienthal 2013), will increase demand for cloud services. The latter will lead to trading volumes growing, which will, in turn, increase the return volatility (Wang and Yau 2000).

If cloud customers intend to apply our ROA algorithms within, for instance, their batch job schedulers, they need to identify suitable computation jobs for deferral (e.g., training machine learning models). Moreover, job schedulers must integrate the relevant cloud service provider's API (e.g., Query API for Amazon EC2, or the AWS SDKs) to automatically compare spot prices and the job backlog. This approach takes the boundary conditions of cloud service providers' customers, such as the service level agreements with their own customers, into consideration, which allows them to optimally decide which jobs to outsource to their provider and at what time.

Furthermore, beside to AWS, our ROA is transferable to emerging cloud spot markets: Recently, the Deutsche Börse Cloud Exchange, the Cloud Commodities Exchange Group, and the Massachusetts Open Cloud Exchange have initiated market places that provide spot prices. One could also apply our ROA to other domains, such as electricity and surge pricing, as long as some time-of-day-specific spot price patterns reoccur: Since we build on Fridgen et al.'s 2016 approach, electricity market researchers could inversely utilize our approaches. Surge pricing has also seen the first research on price forecasting (e.g., Laptev et al. 2017).

Cloud providers too can benefit from customers applying our approaches. They could, for instance, categorize spot instance bidders into more and less flexible customers. Flexible customers contribute to an improved server utilization (i.e., less idle resources), as they can "smooth out some of the computation requests with monetary incentives and lead to a more efficient use of Cloud infrastructure" (Li et al. 2016b, p. 7). According to Zhang et al. (2014), this more efficient resource allocation leads to higher provider revenue than fixed-price cloud services, which might be a competitive advantage in the market. To stimulate this benefit, providers could develop business models and provide cloud customers with dedicated decision support tools. However, flexible customers are more likely to avoid providers' price peaks, which may lead to a slight decline in the provider revenue, but could result in higher earnings due to the lower overall costs. Subsequent research could analyze these incentives for cloud providers to support or impede flexible cloud customers.

7 Conclusion, Limitations, and Future Research

The rapid standardization and specialization of cloud computing services have led to the development of cloud spot markets on which cloud service providers and customers can trade in near real-time. The frequent changes in demand and supply give rise to spot prices that vary considerably throughout the day. Depending on the category of a service request, cloud customers often have temporal flexibility to execute their jobs. We apply ROA to the domain of cloud computing spot prices to quantify and exploit the monetary value of short-term temporal flexibility in cloud computing. We adapt different ROA approaches that, at consecutive points in time, decide whether to purchase cloud services immediately or to defer purchase. In our analysis of real-world data from an Amazon EC2 spot instance, we identify time-of-day-specific price patterns. Adapting existing ROA approaches to these patterns, we demonstrate the benefits of such approaches for cloud customers.

Our modeling approaches have technical limitations that subsequent research could address. First, we assume a normal distribution of returns, which does not necessarily hold true for cloud spot prices. Second,

anomalies such as technical issues at the cloud provider might cause immediate and unpredictable price movements (*spikes*) that our stochastic process cannot predict. Third, for reasons of complexity, we limit our research to discrete-time models, although analytical approximations of or numerical solutions for continuous-time models and decision making would offer more action flexibility. Fourth, we limit our discrete-time models to extensions of Cox et al.'s (1979) and Tian's (1993) approaches.

Besides temporal flexibility, cloud customers could also exploit their spatial flexibility, as cloud spot prices still lack liquidity and are not necessarily arbitrage-free given the different providers and locations (Cheng et al. 2016; Fridgen et al. 2017). Further influencing factors, such as the home bias, amplify arbitrage opportunities, which cloud customers could seize by buying and selling cloud capacity. Future research could therefore integrate the optimization of temporal and spatial flexibility.

Cloud customers, service providers, and scholars may embed the proposed ROA in their decision support systems to optimize the execution of variable-time requests in cloud spot markets. This novelty has the potential to not only generate monetary benefits, but to also increase cloud spot markets' adoption.

References

Allenotor D, Thulasiram RK (2014) A Discrete Time Financial Option Pricing Model for Cloud Services. In: IEEE 11th Intl. Conf. on Ubiquitous Intelligence and Computing, Piscataway, NJ, pp 629–636.

Alzaghoul E, Bahsoon R (2013) CloudMTD: Using real options to manage technical debt in cloud-based service selection. In: 4th Intl. Workshop on Managing Technical Debt, pp 55–62.

Alzaghoul E, Bahsoon R (2014) Evaluating Technical Debt in Cloud-Based Architectures Using Real Options. In: 23rd Australian Software Engineering Conference (ASWEC), pp 1–10.

Amazon Web Services (2017) Amazon EC2 Spot Instances Pricing. https://aws.amazon.com/ec2/spot/pricing/. Accessed 2017-05-01.

Amenc N, Le Sourd V (2003) Portfolio theory and performance analysis. Wiley, Chichester, UK.

Amin KI (1991) On the Computation of Continuous Time Option Prices Using Discrete Approximations. The Journal of Financial and Quantitative Analysis 26(4):477. doi:10.2307/2331407.

Amram M, Kulatilaka N (1999) Real options: Managing strategic investment in an uncertain world. Harvard Business School Press, Boston, MA.

Andrzejak A, Kondo D, Yi S (2010) Decision Model for Cloud Computing under SLA Constraints. In: IEEE Intl. Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, pp 257–266.

Arevalos S, Lopez-Pires F, Baran B (2016) A Comparative Evaluation of Algorithms for Auction-Based Cloud Pricing Prediction. In: IEEE Intl. Conference on Cloud Engineering (IC2E), pp 99–108.

Baughman M, Haas C, Wolski R, Foster I, Chard K (2018) Predicting Amazon Spot Prices with LSTM Networks. In: Proceedings of the 9th Workshop on Scientific Cloud Computing (ScienceCloud'18). ACM Press, New York, NY, pp 1–7.

Benaroch M, Kauffman RJ (1999) A Case for Using Real Options Pricing Analysis to Evaluate Information Technology Project Investments. Information Systems Research 10(1):70–86. doi:10.1287/isre.10.1.70.

Benaroch M, Kauffman RJ (2000) Justifying Electronic Banking Network Expansion Using Real Options Analysis. MIS Quarterly 24(2):197. doi:10.2307/3250936.

Ben-Yehuda OA, Ben-Yehuda M, Schuster A, Tsafrir D (2013) Deconstructing Amazon EC2 Spot Instance Pricing. ACM Transactions on Economics and Computation 1(3):1–20. doi:10.1145/2509413.2509416.

Bestavros A, Krieger O (2014) Toward an Open Cloud Marketplace: Vision and First Steps. IEEE Internet Computing 18(1):72–77. doi:10.1109/MIC.2014.17.

Black F, Scholes M (1973) The Pricing of Options and Corporate Liabilities. Journal of Political Economy 81(3):637–654. doi:10.1086/260062.

Cai Z, Li X, Ruiz R, Li Q (2018) Price forecasting for spot instances in Cloud computing. Future Generation Computer Systems 79:38–53. doi:10.1016/j.future.2017.09.038.

Chen T, Zhang J, Lai K-K (2009) An integrated real options evaluating model for information technology projects under multiple risks. International Journal of Project Management 27(8):776–786. doi:10.1016/j.ijproman.2009.01.001.

Cheng HK, Li Z, Naranjo A (2016) Research Note—Cloud Computing Spot Pricing Dynamics: Latency and Limits to Arbitrage. Information Systems Research 27(1):145–165. doi:10.1287/isre.2015.0608.

Cox JC, Ross SA, Rubinstein M (1979) Option pricing: A simplified approach. Journal of Financial Economics 7(3):229–263. doi:10.1016/0304-405X(79)90015-1.

Dadashov E, Cetintemel U, Kraska T (2014) Putting Analytics on the Spot: Or How to Lower the Cost for Analytics. IEEE Internet Computing 18(5):70–73. doi:10.1109/MIC.2014.94.

Ekwe-Ekwe N, Barker A (2018) Location, Location, Location: Exploring Amazon EC2 Spot Instance Pricing Across Geographical Regions. In: 18th IEEE/ACM Intl. Symposium on Cluster, Cloud and Grid Computing (CCGRID), pp 370–373.

Fridgen G, Häfner L, König C, Sachs T (2016) Providing Utility to Utilities: the Value of Information Systems Enabled Flexibility in Electricity Consumption. Journal of the Association for Information Systems 17(8):537 – 563.

Fridgen G, Keller R, Thimmel M, Wederhake L (2017) Shifting load through space–The economics of spatial demand side management using distributed data centers. Energy Policy 109:400–413. doi:10.1016/j.enpol.2017.07.018.

Gregor S, Hevner AR (2013) Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37(2):337–355.

Hull JC (2014) Options, futures, and other derivatives. Pearson, Upper Saddle River, NJ.

Jarrow RA, Rudd A (1983) Option pricing. Irwin, Homewood, IL.

Javadi B, Thulasiramy RK, Buyya R (2011) Statistical Modeling of Spot Instance Prices in Public Cloud Environments. In: IEEE 4th Intl. Conference on Utility and Cloud Computing (UCC), Victoria, NSW, pp 219–228.

Jede A, Teuteberg F (2016) Valuing the Advantage of Early Termination: Adopting Real Options Theory for SaaS. In: 49th Hawaii Intl. Conference on System Sciences, Koloa, HI, pp 4880–4889.

Kamiński B, Szufel P (2015) On optimization of simulation execution on Amazon EC2 spot market. Simulation Modelling Practice and Theory 58:172–187. doi:10.1016/j.simpat.2015.05.008.

Karunakaran S, Sundarraj RP (2015) Bidding Strategies for Spot Instances in Cloud Computing Markets. IEEE Internet Computing 19(3):32–40. doi:10.1109/MIC.2014.87.

Keller R, König C (2014) A Reference Model to Support Risk Identification in Cloud Networks. In: Proceedings of the 35th Intl. Conference on Information Systems (ICIS 2014), Auckland.

Khandelwal V, Chaturvedi A, Gupta CP (2017) Amazon EC2 Spot Price Prediction using Regression Random Forests. IEEE Transactions on Cloud Computing:1. doi:10.1109/TCC.2017.2780159.

Klaus C, Krause F, Ullrich C (2014) Determining the Business Value of Volume Flexibility for Service Providers - A Real Options Approach. In: Proceedings of the 22nd European Conference on Information Systems (ECIS), Tel Aviv.

Kleinert A, Stich V (2010) Valuation of Procurement Flexibility in the Machinery and Equipment Industry Using the Real Option Approach. In: Bernus P, Doumeingts G, Fox M (eds) Enterprise Architecture, Integration and Interoperability. Springer, Berlin, Heidelberg, pp 21–31.

Kumar D, Baranwal G, Raza Z, Vidyarthi DP (2017) A Survey on Spot Pricing in Cloud Computing. Journal of Network and Systems Management 1(10):7. doi:10.1007/s10922-017-9444-x.

Laptev N, Yosinski J, Li LE, Smyl S (2017) Time-series Extreme Event Forecasting with Neural Networks at Uber. In: Time Series Workshop at ICML 2017, Sydney.

Lee Y-C, Lee S-S (2011) The valuation of RFID investment using fuzzy real option. Expert Systems with Applications 38(10):12195–12201. doi:10.1016/j.eswa.2011.03.076.

Leisen DPJ, Reimer M (1996) Binomial models for option valuation - examining and improving convergence. Applied Mathematical Finance 3(4):319–346. doi:10.1080/13504869600000015.

Lewis GA (2013) Role of Standards in Cloud-Computing Interoperability. In: 46th Hawaii Intl. Conference on System Sciences, Wailea, HI, pp 1652–1661.

Li Z, Tärneberg W, Kihl M, Robertsson A (2016a) Using a Predator-Prey Model to Explain Variations of Cloud Spot Price. In: 6th Intl. Conference on Cloud Computing and Services Science, pp 51–58.

Li Z, Zhang H, O'Brien L, Jiang S, Zhou Y, Kihl M, Ranjan R (2016b) Spot Pricing in the Cloud Ecosystem: A comparative investigation. Journal of Systems and Software 114:1–19. doi:10.1016/j.jss.2015.10.042.

Lilienthal M (2013) A Decision Support Model for Cloud Bursting. Business & Information Systems Engineering 5(2):71–81. doi:10.1007/s12599-013-0257-5.

Loutas N, Kamateri E, Bosi F, Tarabanis K (2011a) Cloud Computing Interoperability: The State of Play. In: IEEE 3rd Intl. Conference on Cloud Computing Technology and Science, Athens, GR, pp 752–757.

Loutas N, Kamateri E, Tarabanis K (2011b) A Semantic Interoperability Framework for Cloud Platform as a Service. In: IEEE 3rd Intl. Conference on Cloud Computing Technology and Science, Athens, GR, pp 280–287.

Marathe A, Harris R, Lowenthal D, Supinski BR de, Rountree B, Schulz M (2014) Exploiting redundancy for cost-effective, time-constrained execution of HPC applications on amazon EC2. In: 23rd Intl. Symposium on High-performance Parallel and Distributed Computing (HPDC '14). ACM Press, New York, NY, pp 279–290.

Mazzucco M, Dumas M (2011) Achieving Performance and Availability Guarantees with Spot Instances. In: IEEE 13th Intl. Conference on High Performance Computing and Communications (HPCC), Banff, AB, pp 296–303.

Meinl T, Neumann D (2009) A Real Options Model for Risk Hedging in Grid Computing Scenarios. In: 42nd Hawaii Intl. Conference on System Sciences, Waikoloa, HI, pp 1–10.

Mell PM, Grance T (2011) The NIST definition of cloud computing. National Institute of Standards and Technology, Gaithersburg, MD.

Myers SC (1977) Determinants of corporate borrowing. Journal of Financial Economics 5(2):147–175. doi:10.1016/0304-405X(77)90015-0.

Naldi M, Mastroeni L (2016) Economic decision criteria for the migration to cloud storage. European Journal of Information Systems 25(1):16–28. doi:10.1057/ejis.2014.34.

Náplava P (2016) Evaluation of Cloud Computing Hidden Benefits by Using Real Options Analysis. Acta Informatica Pragensia 5(2):162–179. doi:10.18267/j.aip.92.

Nwankpa J, Roumani Y, Roumani YF (2016) Exploring ERP-enabled Technology Adoption: A Real Options Perspective. Communications of the Association for Information Systems 39:529–555.

Rossi E, Spazzini F (2014) GARCH Models for Commodity Markets. In: Roncoroni A, Fusai G, Cummins M (eds) Handbook of Multi-Commodity Markets and Products. Wiley, Chichester, UK, pp 687–753.

Skyhigh Networks (2017) Custom Applications and IaaS Trends 2017. https://downloads.cloudsecurityalliance.org/assets/survey/custom-applications-and-iaas-trends-2017.pdf. Accessed 2018-08-01.

Tamrakar K, Yazidi A, Haugerud H (2017) Cost Efficient Batch Processing in Amazon Cloud with Deadline Awareness. In: IEEE 31st Intl. Conference on Advanced Information Networking and Applications (AINA), pp 963–971.

Tang S, Yuan J, Li X-Y (2012) Towards Optimal Bidding Strategy for Amazon EC2 Cloud Spot Instance. In: IEEE 5th Intl. Conference on Cloud Computing (CLOUD), Honolulu, HI, pp 91–98.

Tang S, Yuan J, Wang C, Li X-Y (2014) A Framework for Amazon EC2 Bidding Strategy under SLA Constraints. IEEE Transactions on Parallel and Distributed Systems 25(1):2–11. doi:10.1109/TPDS.2013.15.

Tian Y (1993) A modified lattice approach to option pricing. Journal of Futures Markets 13(5):563–577. doi:10.1002/fut.3990130509.

Trigeorgis L (1996) Real Options. The Journal of Finance 51(5):1974–1977.

Ullrich C (2013) Valuation of IT Investments Using Real Options Theory. Business & Information Systems Engineering 5(5):331–341. doi:10.1007/s12599-013-0286-0.

United States Securities and Exchange Commission (US SEC) (2017) Snap Inc.: Form S-1 Registration Statement. https://www.sec.gov/Archives/edgar/data/1564408/000119312517029199/d270216ds1.htm. Accessed 2018-08-01.

van Hulle C (1988) Option pricing methods: an overview. Insurance: Mathematics and Economics 7(3):139–152. doi:10.1016/0167-6687(88)90071-6.

Vieira CCA, Bittencourt LF, Madeira ERM (2015) A Scheduling Strategy Based on Redundancy of Service Requests on IaaS Providers. In: 23rd Euromicro Intl. Conference on Parallel, Distributed, and Network-Based Processing, Turku, pp 497–504.

Wallace RM, Turchenko V, Sheikhalishahi M, Turchenko I, Shults V, Vazquez-Poletti JL, Grandinetti L (2013) Applications of neural-based spot market prediction for cloud computing. In: IEEE 7th Intl. Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), pp 710–716.

Wang GHK, Yau J (2000) Trading volume, bid-ask spread, and price volatility in futures markets. Journal of Futures Markets 20(10):943–970. doi:10.1002/1096-9934(200011)20:10<943::AID-FUT4>3.0.CO;2-8.

Wu F, Li HZ, Chu LK, Sculli D, Gao K (2009) An approach to the valuation and decision of ERP investment projects based on real options. Annals of Operations Research 168(1):181–203. doi:10.1007/s10479-008-0365-7.

Yam C-Y, Baldwin A, Shiu S, Ioannidis C (2011) Migration to Cloud as Real Option: Investment Decision under Uncertainty. In: IEEE 10th Intl. Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Changsha, CN, pp 940–949.

Zafer M, Song Y, Lee K-W (2012) Optimal Bids for Spot VMs in a Cloud for Deadline Constrained Jobs. In: IEEE 5th Intl. Conference on Cloud Computing (CLOUD), Honolulu, HI, pp 75–82.

Zhang Y, Li B, Huang Z, Wang J, Zhu J, Peng H (2014) Strategy-Proof Auction Mechanism with Group Price for Virtual Machine Allocation in Clouds. In: 2nd Intl. Conference on Advanced Cloud and Big Data (CBD), Huangshan, pp 60–68.

Zheng L, Joe-Wong C, Tan CW, Chiang M, Wang X (2015) How to Bid the Cloud. ACM SIGCOMM Computer Communication Review 45(5):71–84. doi:10.1145/2829988.2787473.

Zimmermann S, Müller M, Heinrich B (2016) Exposing and selling the use of web services—an option to be considered in make-or-buy decision-making. Decision Support Systems 89:28–40. doi:10.1016/j.dss.2016.06.006.