Deterministic regime based Black-Scholes model for short-term option pricing

Ziwen Ye, Thomas Lonon, Ionuţ Florescu

Hanlon Financial Systems Center and Financial Engineering

Stevens Institute of Technology, Hoboken, USA

Abstract

In this paper, we use high frequency equity data to price short-term equity options. To quantify the effects of high frequency trading in the equity markets, we propose a deterministic time regime as well as regime based realized volatility to capture the change. Based on these two estimations, we further propose a deterministic time regime Black-Scholes model. By comparing the result from the previous model structure with the classic Black-Scholes model and observed option bid/ask prices, we conclude that the high frequency data is more informative in short-term option pricing.

Key words: Short-term option pricing, High-frequency data, Black-Scholes model

1 Introduction

The trading volume of short-term options has increased sharply during recent years. The S&P500 weekly options were introduced in 2005 as a way to provide hedging options to investors. In 2010, these short-term S&P 500 options contributed less than 5% (15,133 daily contracts) of the total average daily trading volume of options. This number increased to more than 30% (197,126 contracts) in 2013 [21]. The average daily number of contracts traded in 2017 was 523,170 [19]. The goal of this paper is to provide an effective, accurate methodology to price these options, using high frequency data.

With the increased trading volume in high-frequency trading, market micro-structure is changing [20]. Martens and Zein [18] show that using historical returns calculated from high-frequency data has better predicting performance for financial volatility when compared with the implied volatility. Garcia et al. [10] combines both option prices and high-frequency spot prices when estimating parameters for a stochastic volatility model, which allows them to obtain explicit expansions of the implied volatility. The work by Andersen et al. [1] shows that high-frequency data is robust in pricing tail risk. Inspired by these researchers, we use tick-by-tick equity market data to estimate realized volatility in this work when pricing the short-term options.

Despite its well known shortcomings (constant volatility, independence of increments, lack of leptokurtic behavior), the geometric Brownian motion and the corresponding Black-Scholes model [2] is still used in option pricing. Whether it is using an implied volatility, binomial tree approximation, or an implied value from a different model (the case of SABR), the Black Scholes formula continues to be used in the market today. In this work, we are pricing short maturity options. We need to emphasize that the work is focused on short maturities only as we are going to be using realized variance estimated from price data (i.e., under the objective probability measure). To price options, one needs to use the market price of risk to change between the objective and the equivalent martingale measures. However, in the Black Scholes model, as well as in the regime switching volatility model we are using, the only change is to the drift parameter. By choosing to price short maturity options we are working with drifts that are essentially zero.

A regime switching volatility model is not new. For example, Bollen [3] introduced a two-regime Black-Scholes model. He used Monte-Carlo simulations to show that this model improves pricing accuracy. Ishijima and Kihara [13] used a similar idea but their work can be expanded to multiple regimes. Guo [11] uses a two-state model in option pricing, letting the sentiment analysis determine which state the model was in. Fuh et al. [9] implemented a regime switching model with the regimes determined by a Markov switching model. Then the authors use Monte-Carlo simulations to assess the model performance. Despite the increase in accuracy, they found that their regime-switching model is slower in option pricing compared to the classic Black-Scholes model. Krsteva [14] expands the model to a multi-dimensional regime switching volatility.

The major goal of this study is to introduce a so called "deterministic" regime based model for short-term option pricing. In a traditional stochastic regime switching model, a method of estimating regime and time distribution is devised. Then future projections are random and are based on these estimated distributions. In our case while the regime estimation is similar the future projection is "deterministic". That is once we determine the times and values of the regimes we project an average behavior for the following day. The method to estimate regimes and times is described in Section 2.1, and the corresponding option pricing model is proved in Section 2.2. To check the efficiency of proposed model, we compare the estimated option price with historical data in Section 3.

2 Model description

The main criticism of the geometric Brownian motion and correspondingly the Black and Scholes model is the constancy of the volatility parameter. Numerous expansions have been created all producing better results by considering what essentially are more complex volatility models. We have worked ourselves with stochastic volatility, jump diffusion processes, fractional Brownian motion driven volatility and many other expansions. In this article we are presenting a very noticeable departure from our previous work.

The most important and fundamental difference in this work from other similar work is the assumption we are making on the volatility process. From empirical equity data, we know that the price variability within a short time interval Δt tends to be persistent. Some researchers call this volatility clustering [17, 5]. Further, when we look at the behaviour of intraday stock prices, we notice certain trends that seemed to be fairly consistent. For example, the period shortly after opening almost always corresponds to a higher volatility [4]. We posited for this work that each asset would have these intraday consistencies in the periods of higher and lower volatility.

We use historical prices to determine both the length of the periods, as well as the constant volatility within these periods. These assumptions form the basis of our "deterministic" regime model. As mentioned in the introduction this approach will only produce noticeable results for short-term maturity options.

The time interval Δt is set to 1 minute in this work. We need to use high frequency data and we chose to aggregate the HF data at 1 minute level. This is consistent and inspired by our previous work [22]. The regime periods group multiple such 1 minute intervals, this depends on the historical data. Thus, the choice of the time interval length is not crucial to our work. Nevertheless, in our work we use highly quoted options and highly traded equity. If the procedure is to be applied to equity that isn't highly traded the interval length Δt would probably have to be suitably increased.

2.1 Deterministic time regimes

The model we are using is a simple regime switching model but with a twist - this is why we call it **deterministic** regime switching model. In a classical regime switching model [11, 7, 14], one uses historical data to estimate parameters of the regimes as well as the transition probability matrix between regimes. The probabilities are used when calculating expectations of future values. More precisely, one would use the stationary distribution of regimes in the calculation of option prices (see e.g., [8]). In the current work we are proceeding very differently, we are assuming that during the trading day the regimes duration are fixed and determined from previous days. The regime values are also fixed and determined at the beginning of the trading day from historical estimation. The values and duration of regimes are updated every day.

The assumption we are making is very different from anything existing in literature. Fundamentally, it goes against what we teach in our courses and we wrote in our books. It is simply born from simply observing the equity market. To test its validity we investigate using two types of returns for the period Δt.

First, we use simple return calculated using the close price at every interval. Specifically, we use minute data for the past M days (here $M = 30$). Mathematically, if we denote with $\{P_{t,\mathcal{D}}\}$ the price at close of minute t from day $\mathcal{D}$ the simple return is:

$$SR_{t,\mathcal{D}} = \frac{P_{t+\Delta t,\mathcal{D}} - P_{t,\mathcal{D}}}{P_{t,\mathcal{D}}}, \quad \mathcal{D} \in \{1, \dots, M\}.$$

Second, we are using the activity-weighted return (AWR), introduced in Ye and Florescu [22]. Compared with the simple return, AWR relies on tick-by-tick data:

$$AP_{t,\mathcal{D}} = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} TP_{i,\mathcal{D}} \tag{1}$$

$$AWR_{t,\mathcal{D}} = \frac{AP_{t+\Delta t,\mathcal{D}} - AP_{t,\mathcal{D}}}{AP_{t,\mathcal{D}}} \tag{2}$$

In the formulas above TP_i is the observed price for trade i within interval $[t, t + \Delta t)$. $\mathcal{N}$ is the number of observed trades during the time interval.

Since we use $\Delta t = 1$ minute, we have 390 returns each trading day. For each minute of the day we have exactly M returns from each of the previous days. We use $\{r_{t,\mathcal{D}}\}$ to denote the set of returns at minute t.

Definition 1. A deterministic time regime is a time period during which consecutive returns $\{r_{t,\mathcal{D}}\}$ have the same first two moments.

Based on this definition we determine the regimes using the following algorithm:

1. The first regime of the day is determined on the first minute return: $\{r_{1,\mathcal{D}}\}$

2. For the remaining minutes we use $\{r_{t,\mathcal{D}}\}_{t\in\{2,\ldots,390\}}$:

   (a) We perform a t-test and an F-test on $\{r_{t-1,\mathcal{D}}\}$ and $\{r_{t,\mathcal{D}}\}$ to determine whether they have different mean and standard deviations. The hypotheses for the t-test are:

   H_0 : The expected value of minute t return is the same as the previous regime.

   H_α : The mean values are different.

   The F-test, has the following hypothesis:

   H_0 : The standard deviation of minute t return is the same as the previous regime.

   H_α : The standard deviations are different.

   (b) If the P-values from both tests are not significant, we do not have enough statistical evidence to conclude that t is a different regime. To improve estimation, we merge the t minute data into the previous regime. We then go to the next minute $t = 1$ and repeat from step (a).

   (c) If the P-value from either test is statistically significant, we have enough evidence to conclude that a new regime starts from t. Thus, we set the next regime of the day based on the $\{r_{t,\mathcal{D}}\}$, and we go back to step 1 and set up the new regime.

Remark 1. We note that if the two tests would be independent, the error type 1 probability (concluding the regime changed when in fact it did not) would be around 0.0975. However, it is very clear that the two tests are not independent and clearly the probability of type 1 error should be smaller. Nonetheless, we are aware that the test is detecting new regimes more often than it should. A proper adjustment is complicated and beyond the scope of the paper.

Remark 2. Different types of return may lead to a different number of estimated regimes and regime periods. Tables 1 and 2 show the regimes obtained for SPY on March 6th, 2018, using the activity-weighted return respectively the simple return.

Table 1: The resulting regimes for SPY data using activity-weighted return on March 6th, 2018

Regime	Realized std.	Time intervals	Regime	Realized std.	Time intervals
1	0.00101028	5	12	0.00044217	40
2	0.00069715	24	13	0.00049807	18
3	0.00096870	1	14	0.00057447	2
4	0.00068143	12	15	0.00048320	32
5	0.00059480	56	16	0.00046123	12
6	0.00035984	2	17	0.00053753	21
7	0.00053462	11	18	0.00043654	1
8	0.00045797	11	19	0.00078475	2
9	0.00051905	1	20	0.00050311	31
10	0.00041127	16	21	0.00067752	75
11	0.00043499	16	22	0.00064766	1

Table 2: The resulting regimes for SPY data using simple return on March 6th, 2018

Regime	Realized std.	Time intervals	Regime	Realized std.	Time intervals
1	0.00115839	8	15	0.00057428	6
2	0.00083775	7	16	0.00058914	20
3	0.00077852	8	17	0.00060506	4
4	0.00082103	5	18	0.00056310	12
5	0.00104262	2	19	0.00043842	1
6	0.00079702	12	20	0.00053791	1
7	0.00069331	50	21	0.00069792	17
8	0.00057783	62	22	0.00100232	1
9	0.00050496	9	23	0.00065836	2
10	0.00050110	16	24	0.00060620	30
11	0.00053292	16	25	0.00064020	10
12	0.00061400	18	26	0.00093795	36
13	0.00071667	2	27	0.00088185	29
14	0.00060245	5	28	0.00117858	1

Once we have the deterministic regime R_n we estimate the volatility for that regime using the classical realized variance formula [16, 12, 15]:

$$\sigma_n = \sqrt{\frac{1}{M * k_n} \sum_{\mathcal{D}=1}^{M} \sum_{t=1}^{k_n} r_{t,\mathcal{D}}^2} \tag{3}$$

where k_n denotes the number of returns in regime R_n.

Recall that our goal is to calculate *short maturity* option prices. Due to the short time to maturity it is quite important to have an exact formula to determine the time to expiry at any moment t.

Definition 2. The "remaining life time" τ_n in each of the deterministic time regime R_n can be calculated as:

$$\tau_n = \left(390 * \left\{\frac{\tau}{390}\right\} - \sum_{i=n+1}^{N} k_i)\right)_+ + \left\lfloor \frac{\tau}{390} \right\rfloor * k_n \tag{4}$$

where τ is the exact time to maturity in minutes. We use $\lfloor\cdot\rfloor$ and $\{\cdot\}$ to denote the integer part respectively the fractional part of a real number. A trading day in the equity market contains exactly 390 minutes.

Remark 3. The regime values, as well as time periods, are determined at the end of each trading day, which are fixed values for the following trading days. This is why we call the model "deterministic" time regime model. During the following day,s the option price depends on the current underlying equity price, as well as the time to maturity, which changes every minute.

2.2 Deterministic regime based Black-Scholes model

In this section we are going to calculate the option pricing formula for the deterministic regime switching model. We assume the equity evolves according to a generalized Geometric Brownian Motion:

$$S(T) = S(0) \exp\left\{\int_0^T (r - \frac{\sigma^2(u)}{2}) du + \int_0^T \sigma(u) dW(u)\right\}. \tag{5}$$

In the previous formula $\sigma(u)$ is a piecewise continuous function. We denote:

$$\sigma^2(t) = \sum_{n=1}^{N} \sigma_n^2 \mathbb{1}_{\{t \in R_n\}}$$
$$\sigma(t) = \sum_{n=1}^{N} \sigma_n \mathbb{1}_{\{t \in R_n\}}$$

where $\mathbb{1}$ is the indicator function and n is the index of regimes. The realized volatility in regime n is σ_n. We rewrite Equation 5:

$$\begin{aligned}
S(T) &= S(0) \exp\left\{ \int_0^T (r - \frac{\sigma^2(u)}{2}) du + \int_0^T \sigma(u) dW(u) \right\} \\
&= S(0) \exp\left\{ rT - \sum_{n=1}^{N} \int_0^T \frac{\sigma_n^2}{2} \mathbb{1}_{\{t \in R_n\}} du + \sum_{n=1}^{N} \int_0^T \sigma_n \mathbb{1}_{\{t \in R_n\}} dW(u) \right\} \\
&= S(0) \exp\left\{ rT - \sum_{n=1}^{N} \frac{\sigma_n^2}{2} \tau_n + \sum_{n=1}^{N} \sigma_n W(\tau_n) \right\}
\end{aligned}$$

Recall the notations τ for the time to maturity, and τ_n the remaining life time in regime n.

The call option price is calculated as: $E[e^{-rt}(S(T) - K)_+]$. We simply substitute the stock price and we obtain:

$$\begin{aligned}
E[e^{-rT}(S(T) - K)_+] &= \int_{d^*}^{\infty} e^{-rT} \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{z^2}{2\epsilon^2}} (S(0) e^{rT - \frac{\epsilon^2}{2} + z} - K) dz \\
&= S(0) \int_{d^*}^{\infty} \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{z^2}{2\epsilon^2} - \frac{\epsilon^2}{2} + z} dz - K e^{-rT} \int_{d^*}^{\infty} \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{z^2}{2\epsilon}} dz \\
&= S(0) \int_{d^*}^{\infty} \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{z^2 - 2z\epsilon^2 + \epsilon^4}{2\epsilon^2}} dz - K e^{-rT} \int_{\frac{d^*}{2}}^{\infty} \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{z^2}{2}} dz \\
&= S(0) \int_{d^*}^{\infty} \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{(z-\epsilon^2)^2}{2\epsilon^2}} dz - K e^{-rT} \int_{d^*}^{\infty} \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{z^2}{2\epsilon}} dz \\
&= S(0) \int_{\frac{d* - \epsilon^2}{\epsilon}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - K e^{-rT} \int_{\frac{d^*}{\epsilon}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2\epsilon}} dz \\
&= S(0) N\left(\frac{-d^* + \epsilon^2}{\epsilon} \right) - K e^{-rT} N\left(-\frac{d^*}{\epsilon} \right)
\end{aligned}$$

In the derivation we denoted $d^* = \frac{log \frac{K}{S(0)} - rT + \frac{1}{2} \sum \sigma_n^2 \tau_n}{\sqrt{\sum \sigma_n^2 \tau_n}}$ and $\sum \sigma_n^2 \tau_n = \epsilon^2$.

Finally, if we denote:

$$d_+ = \frac{log\frac{S(0)}{K} + (rT + \frac{\epsilon^2}{2})}{\epsilon}$$
$$d_- = \frac{log\frac{S(0)}{K} + (rT - \frac{\epsilon^2}{2})}{\epsilon}$$

we obtain the call option formula in the deterministic regime switching model:

$$C(S_0, K, \tau_n, \sigma_n, r) = S(0)N(d_+) - Ke^{-rT}N(d_-)$$

Remark 4. The formula above may be viewed as a special case of the option pricing formula in [6, equation (2.23)]. In the citation and in other classic regime switching models [11, 9] based on a hidden Markov process, the option pricing formula use the transition probability matrix or intensities of transition. The present model may be viewed as a particular case. Since the transitions are deterministic, they are only present in the remaining time to maturity formula (4). Furthermore, the quantities $d_\pm$ and ϵ are containing the volatility values and regime times.

Remark 5. We make a final remark about the risk-free rate r. In this section, we derive the option pricing formula in a general form. However, we are only analyzing short-term options, thus the value used in the formula (rT) is generally small. Therefore, the impact coming from the risk-free rate is extremely weak, and this allows us to treat r as 0 throughout this study.

3 Data description and empirical analysis

3.1 Data and model description

Our research focuses on analyzing the options with short maturities. To have a better comparison, we implement our methods on the top 10 traded weekly options at the time of the analysis. The ticker names are recorded in Table 3. SPY has three maturities each week (Monday, Wednesday and Friday), however we only analyze the options expiring on Fridays.

Table 3: 10 equity options we selected in analyzing short-term option pricing results

Ticker	Equity name	Ticker	Equity name
AAPL	Apple Inc.	BAC	Bank of America Corporation
C	Citigroup Inc.	FB	Facebook, Inc.
NVDA	NVIDIA Corporation	QCOM	QUALCOMM Incorporated
TWTR	Twitter, Inc.	SPY	SPDR S&P 500 ETF
XLE	Energy Select Sector SPDR ETF	XLF	Financial Select Sector SPDR ETF

One month length of tick equity data is used when estimating the deterministic time regime. This part of data is obtained through Thomson Reuters DataScope Select. Meanwhile, we download the historical option data from Thomson Reuters Tick History at 1-minute frequency. For the option data, we set up the time window from February 1, 2018 to May 30, 2018 in this research.

Throughout this research, we investigate the short-term option pricing performance using two types of models: the classic Black-Scholes model (BS) and deterministic regime based Black-Scholes model (RBS). Under each model, we are using two types of returns to estimate the realized volatility: the simple return (SR) and the activity-weighted return (AR). Table 4 details the initials we will use for each model in this study.

Table 4: The notation for all 4 types of model and its corresponding explanation

Model name	Model description
BS-SR	The classic Black-Scholes model using simple return to estimate a constant volatility
BS-AR	The classic Black-Scholes model using activity weighted return to estimate a constant volatility
RBS-SR	The deterministic regime based Black-Scholes model. Simple return is used when estimating the time regimes
RBS-AR	The deterministic regime based Black-Scholes model. Activity weighted return is used when estimating the time regimes

3.2 Empirical result and analysis

When we evaluate the model performance, we focus on analyzing the root mean square error (RMSE) between model prediction and the mid price between best bid/ask. The result of two-way analysis of variance (ANOVA) is shown in Table 5. The type of return used (AR or SR) is a significant factor as well as the interaction between the type of return and the type of model used. The model is not significant but the p-value is very close to the significance level 0.05.

Table 5: Two way ANOVA summary table, significant P-values are highlighted with bold letters. We use 87655 observations over 3 months

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Type of return	1	1.36	1.36	8.64	**0.0033**
Model structure	1	0.56	0.56	3.55	0.0594
Interaction	1	1.19	1.19	7.55	**0.0060**
Residuals	87652	13758.33	0.16		

From table above the interaction term is significant. Therefore, we need to analyze each equity separately. For simplification, Table 6 shows a "✓" mark whenever the respective factor is significant at the 0.05 significance level. The interaction term is not significant for any ticker. This considerably simplifies our analysis.

Table 6: Two way ANOVA. Significant factors at 0.05 level for each individual equity options

Ticker	Type of Return	Model Structure	Interaction
AAPL			
BAC	✓	✓	
C	✓		
FB			
NVDA	✓		
QCOM	✓	✓	
TWTR			
SPY	✓		
XLE	✓		
XLF			

Based on Table 6 we separate the 10 equity options into two groups. AAPL, FB, TWTR and XLF show no significant factor at all, therefore there is no preferred model for short-term option pricing. For illustration we present the average RMSE values for each model in Table 7.

Table 7: Mean RMSE values based on each model. The lowest value indicates the corresponding model is prefer for this equity.

Ticker	RBS-AR	RBS-SR	BS-AR	BS-SR	Best Combination
AAPL	0.1983	0.1916	0.1942	0.1992	RBS-SR
FB	0.2944	0.3017	0.2922	0.3236	BS-AR
TWTR	0.1185	0.1099	0.1132	0.1285	RBS-SR
XLF	0.0297	0.0285	0.0294	0.0296	RBS-SR

Looking at the tickers with at least 1 significant factor (Table 8), we perform a multiple pairwise comparison. For BAC and QCOM, both factors are significant and thus the combination of two factors produces better results. For the others: C, NVDA, SPY, and XLE the return type is the significant factor.

Table 8: Mean RMSE values based on each model. The lowest value indicates the corresponding model is prefer for this equity.

Ticker	RBS-AR	RBS-SR	BS-AR	BS-SR	Best Combination
BAC	0.0249	0.0293	0.0257	0.0333	RBS-AR
C	0.0919	0.0971	0.0917	0.1023	AR
NVDA	0.6510	0.6664	0.6425	0.7565	AR
SPY	0.1508	0.1571	0.1517	0.1639	AR
QCOM	0.1801	0.1570	0.1693	0.1519	BS-SR
XLE	0.0789	0.0692	0.0760	0.0693	SR

To further visualize the pricing performance we present in Figure 1 the at the money call and put options for BAC. Specifically, we show the daily evolution of the option price on April 19, 2018. We used the open price for BAC ($29.55) to determine the at-the-money option. We shade the observed bid/ask spread for this option in gray. Clearly, the RBS_AR model has the best performance. For both call and put option, the pricing results from this model are within the gray shaded area. Figure 8 shows similar at-the-money call and put on QCOM. The put option valuation is good for all models but for the call option, the Black

Scholes using simple return (BS_SR) produces numbers within the Bid/Ask spread for most of the day.

Figure 1: Detailed pricing result on BAC option. The gray shaded area is the bid/ask spread. Best combination is the regime switching model using Activity Weighted Return RBS_AR

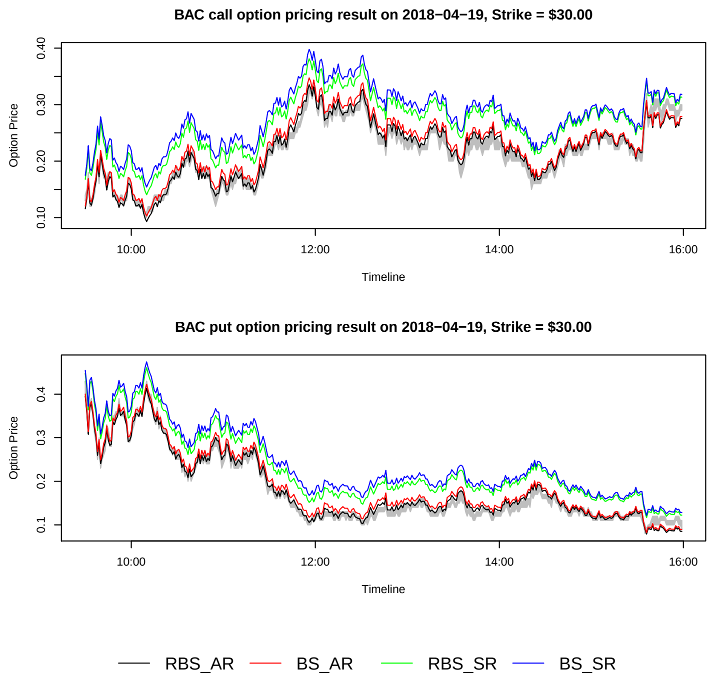


Figure 2: Detailed pricing result on QCOM option. The gray shaded area is the bid/ask spread. Best combination is the Black Scholes model using Simple Return BS_SR

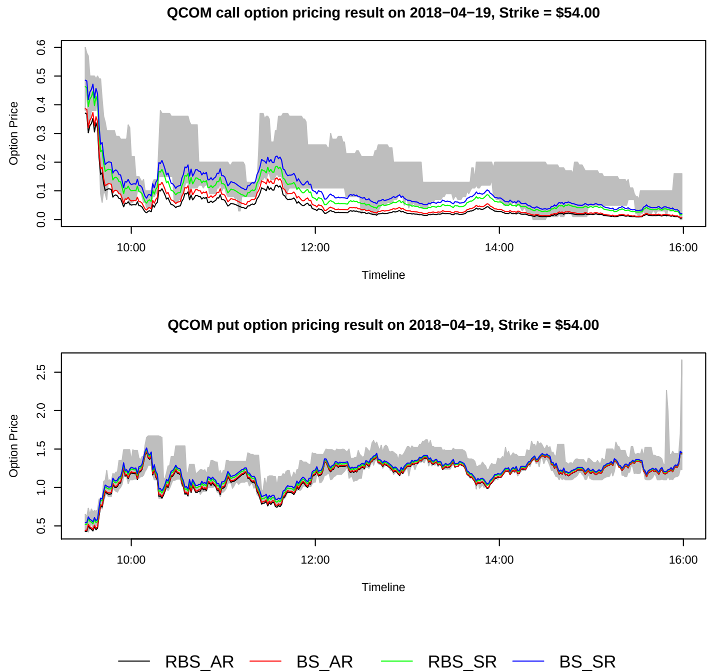
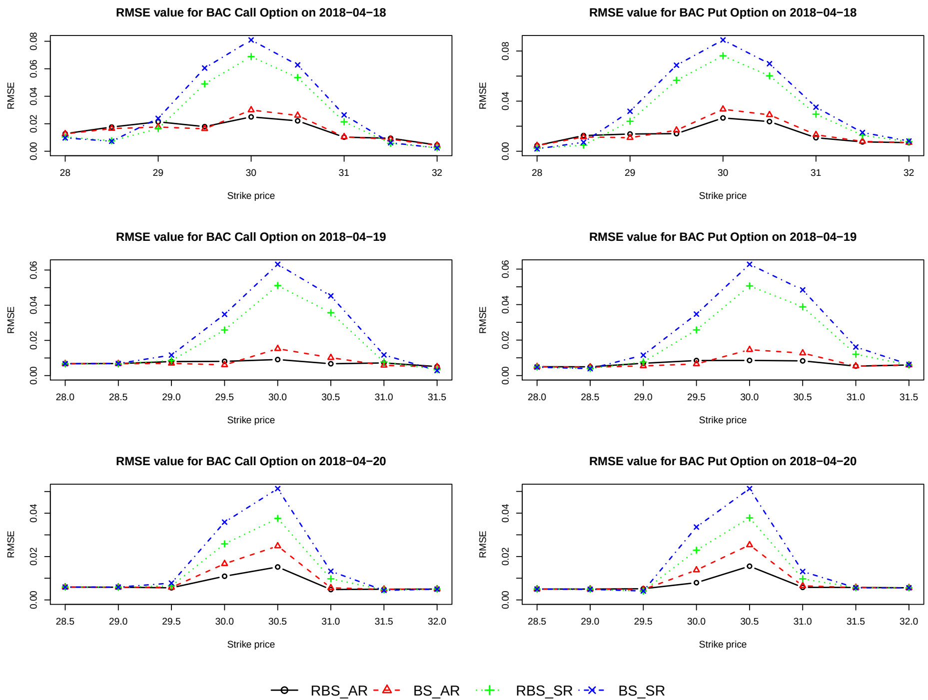


Figure 3 shows the daily RMSE for all strikes on three consecutive days. In this figure, the maturity for all contracts is April 20, 2018. This figure shows that the RBS_AR model has the best performance in both call and put options, for all options. The largest difference is in the at-the-money contracts. This is important because these are the most traded contracts for these short maturity options.

Figure 12 showcases QCOM which if you recall had the combination Black-Scholes/Simple return as best performer. We can see that the difference between the model pricing performance though statistically significant it is small in practice. In fact the RBS_SR and BS_SR both have good performance and when treated separately there is no significant difference between them.

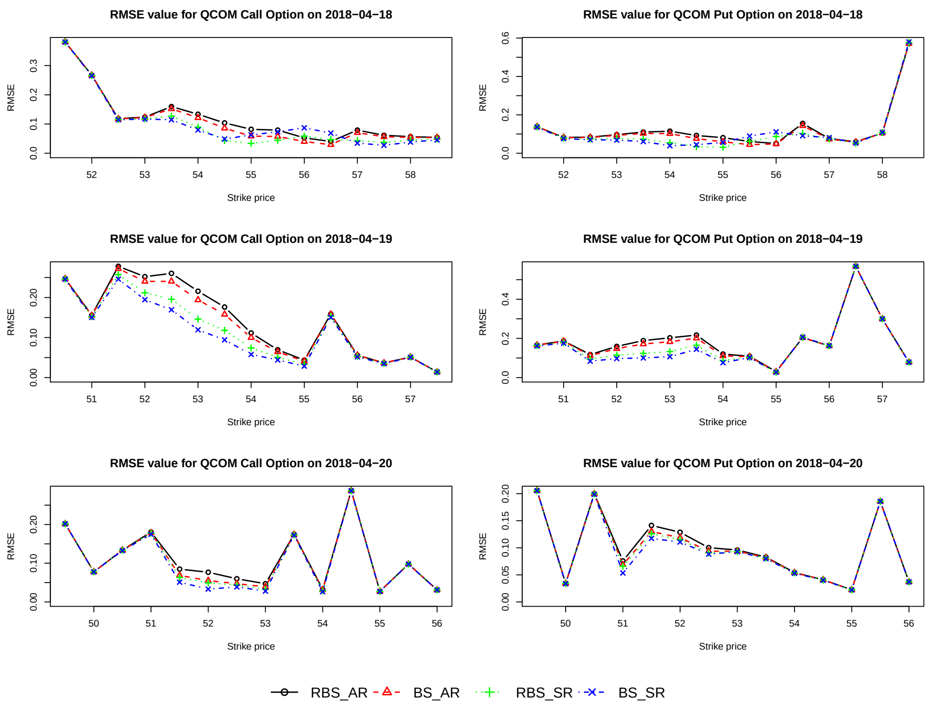
Additionally, we showcases the AAPL and SPY (no significant factors), NVDA and C (1 significant factor) in Appendix B and Appendix C. Although, the statistical analysis told a different story, the plots show the RBS_AR as the best model choice. However, this result is

Figure 3: Short-term pricing result on three consecutive days. The option price for BAC on these three days are \$30.03, \$29.55, and \$30.26. All contracts expire on April 20, 2018



not consistent for consecutive days. In figure C on page 26 we see an explanation why. We can see in that image that the best model changes in later days and that is enough for the ANOVA model to not recognize a best performing model.

Figure 4: Short-term pricing result on three consecutive days. The option price for QCOM on these three days are \$55.14, \$54.00, and \$52.81. All contracts expire on April 20, 2018



4 Conclusion

In this study, we investigate the efficiency of short-term option pricing based on two model structures: the Black-Scholes model (BS) and the deterministic regime based Black-Scholes model (RBS). When estimating the volatility to be used in these models, we use two alternatives: the simple return (SR) and the activity-weighted return (AR). The AR as well as the RBS models are original and introduced by the authors of this work. We apply the 4 combination of these modeling choices to short-term options written on 10 large equity or ETF's. We introduce an ANOVA framework to compare these results - to our surprise this approach is non-existent in finance literature.

As always when working with real data the results we obtain are not definite or conclusive. We generally find that using the activity weighted return (AR) as well as using a deterministic regime model generally improve option pricing accuracy for these short-term options. However, we found that for half of our equities and ETF's we could not draw this conclusion and neither model nor return type made a difference. We believe that this is due to the optimal model choice changing when the option approaches maturity.

Today's markets are fast and every trader anywhere has a Black Scholes based formula which is based on either implied volatility or local volatility or an instantaneous measure of volatility. An option trader typically quotes the price of the option in volatility - that is the Black Scholes implied volatility. When the option price deviates and becomes larger or smaller than the Black Scholes price this will be translated into an increase in volatility which traders often equate with turbulent market. Without a market event that would warrant such increase in volatility, traders are likely to disregard the high vol prices and come back and provide quotes around the Black Scholes price corresponding to historical values of implied volatility.

Based on the empirical result we obtained, we believe that many other factors can influence the choice of the best model, e.g., market events, trading frequency in equity markets, trading frequency in option markets. This is why finding a universal best model for all short term equity option is a hard problem. Nonetheless, the current research message is that looking at the micro cosmos of high frequency data and its dynamics is very relevant to short-term option pricing.

References

[1] Andersen, T. G., N. Fusari, V. Todorov, et al. (2016). The pricing of tail risk and the equity premium: evidence from international option markets. Technical report, Department of Economics and Business Economics, Aarhus University.

[2] Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of political economy 81*(3), 637–654.

[3] Bollen, N. P. (1998). Valuing options in regime-switching models. *Journal of Derivatives 6*, 38–50.

[4] Bozdog, D., I. Florescu, K. Khashanah, and J. Wang (2011). Rare events analysis for high-frequency equity data. *Wilmott 2011*(54), 74–81.

[5] Cont, R. (2007). Volatility clustering in financial markets: empirical facts and agent-based models. In *Long memory in economics*, pp. 289–309. Springer.

[6] Elliott, R. J., L. Chan, and T. K. Siu (2005, Oct). Option pricing and esscher transform under regime switching. *Annals of Finance 1*(4), 423–432.

[7] Florescu, I. and F. Levin (2016). Estimation procedure for regime switching stochastic volatility model and its applications. *Handbook of High-Frequency Trading and Modeling in Finance 9*, 107.

[8] Florescu, I., R. Liu, M. C. Mariani, and G. Sewell (2013). Numerical schemes for option pricing in regime-switching jump diffusion models. *International Journal of Theoretical and Applied Finance 16*(08), 1350046.

[9] Fuh, C.-D., R.-H. Wang, and J.-C. Cheng (2002). Option pricing in a black-scholes model with markov switching. In *Technical Report*. Working Paper, Institute of Statistical Science, Taipei.

[10] Garcia, R., M.-A. Lewis, S. Pastorello, and É. Renault (2011). Estimation of objective and risk-neutral distributions based on moments of integrated volatility. *Journal of Econometrics 160*(1), 22–32.

[11] Guo, X. (2001). Information and option pricings.

[12] Howison, S., A. Rafailidis, and H. Rasmussen (2004). On the pricing and hedging of volatility derivatives. *Applied Mathematical Finance 11*(4), 317–346.

[13] Ishijima, H. and T. Kihara (2005). Option pricing with hidden markov models. *Institute of Finance, Waseda University*.

[14] Krsteva, K. (2014, 12). *Estimation and optimization of linear multi-factor models of stock returns and detection of an underlying regime-switching process.* Ph. D. thesis, Stevens Institute of Technology.

[15] Lian, G., C. Chiarella, and P. S. Kalev (2014). Volatility swaps and volatility options on discretely sampled realized variance. *Journal of Economic Dynamics and Control 47*, 239–262.

[16] Little, T. and V. Pant (2001). A finite difference method for the valuation of variance swaps. In *Quantitative Analysis In Financial Markets: Collected Papers of the New York University Mathematical Finance Seminar (Volume III)*, pp. 275–295. World Scientific.

[17] Lux, T. and M. Marchesi (2000). Volatility clustering in financial markets: a microsimulation of interacting agents. *International journal of theoretical and applied finance 3*(04), 675–702.

[18] Martens, M. and J. Zein (2004). Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets: Futures, Options, and Other Derivative Products 24*(11), 1005–1028.

[19] Moran, M. Record volume for spxw weekly options, products that facilitate targeted strategies.

[20] OHara, M. (2015). High frequency market microstructure. *Journal of Financial Economics 116*(2), 257–270.

[21] Picardo, E. Give yourself more options with weekly and quarterly options.

[22] Ye, Z. and I. Florescu (2019). Extracting information from the limit order book: New measures to evaluate equity data flow. *High Frequency 2*(1), 37–47.

A Risk Netural proof

We define the market price of risk to be:

$$\theta(t) = \frac{\alpha(t) - r}{\sigma(t)} \tag{6}$$

Based on the Girsanov's Theorem, we can define a Brownian motion $\widetilde{W}$ under the risk neutral probability measure $\widetilde{\mathbb{P}}$ as following:

$$\begin{aligned}
\widetilde{W}(t) &= W(t) + \int_0^t \theta(u)du \\
&= W(t) + \int_0^t \frac{\alpha(t) - r}{\sigma(t)}du \\
&= W(t) - r\int_0^t \frac{1}{\sigma(u)}du + \int_0^t \frac{\alpha(u)}{\sigma(u)}du \\
&= W(t) - r\sum_{n=1}^{N}\int_0^t \frac{1}{\sigma_n}\mathbb{1}du + \sum_{n=1}^{N}\int_0^t \frac{\alpha_n}{\sigma_n}\mathbb{1}du
\end{aligned}$$

Which gives us the result:

$$dW(t) = d\widetilde{W}(t) + r\sum_{n=1}^{N}\frac{1}{\sigma_n}\mathbb{1}dt - \sum_{n=1}^{N}\frac{\alpha_n}{\sigma_n}\mathbb{1}dt \tag{7}$$

This allows us to rewrite $S(t)$ as:

$$\begin{aligned}
S(t) &= S(0)exp\left\{\int_0^t (\alpha(u) - \frac{\sigma^2(u)}{2})du + \int_0^t \sigma(u)dW(u)\right\} \\
&= S(0)exp\left\{\int_0^t (\alpha(u) - \frac{\sigma^2(u)}{2})du + \int_0^t \sigma(u)(d\tilde{W}(u) + r\sum_{n=1}^{N}\frac{1}{\sigma_n}\mathbb{1}du - \sum_{n=1}^{N}\frac{\alpha_n}{\sigma_n}\mathbb{1}du)\right\} \\
&= S(0)exp\left\{\int_0^t (\alpha(u) - \frac{\sigma^2(u)}{2})du + \int_0^t \sigma(u)d\tilde{W}(u) + rt - \int_0^t \alpha(u)du\right\} \\
&= S(0)exp\left\{rt - \int_0^t \frac{\sigma^2(u)}{2}du + \int_0^t \sigma(u)d\tilde{W}(u)\right\}
\end{aligned}$$

Where $\alpha(t) = \sum_{n=1}^{N} \alpha_n \mathbb{1}_{\{t \in R_n\}}$

Therefore, we have the discounted stock price:

$$D(t)S(t) = S(0)exp\left\{-\int_0^t \frac{\sigma^2(u)}{2}du + \int_0^t \sigma(u)d\tilde{W}(u)\right\} \tag{8}$$

Now, if we look at the conditional expectation of the discounted stock price under risk neutral measure:

$$\begin{aligned}
\widetilde{E}[D(t)S(t)|\mathscr{F}(s)] &= D(s)T(s)\widetilde{E}\left[\frac{D(t)S(t)}{D(s)S(s)}|\mathscr{F}(s)\right] \\
&= D(s)S(s)\widetilde{E}\left[exp\left\{-\int_s^t \frac{\sigma^2(u)}{2}du + \int_s^t \sigma(u)d\widetilde{W}(u)\right\}\right] \\
&= D(s)S(s)exp\left\{-\int_s^t \frac{\sigma_n^2}{2}\mathbb{1}du\right\} exp\left\{\frac{1}{2}\widetilde{Var}\left[\int_s^t \sigma(u)d\tilde{W}(u)\right]\right\} \\
&= D(s)S(s)
\end{aligned}$$

We can see that the discounted stock price process is a martingale. This confirms the existence of our risk-neutral probability measure and so ensures that there isn't arbitrage in our model. We further claim that since this is a deterministic regime switching model, we also have completeness in our model.

B Daily evolution of the option price and model predictions

Figure 5: Detailed pricing result on AAPL option. The gray shaded area is the bid/ask spread. Best pricing result is from RBS_AR

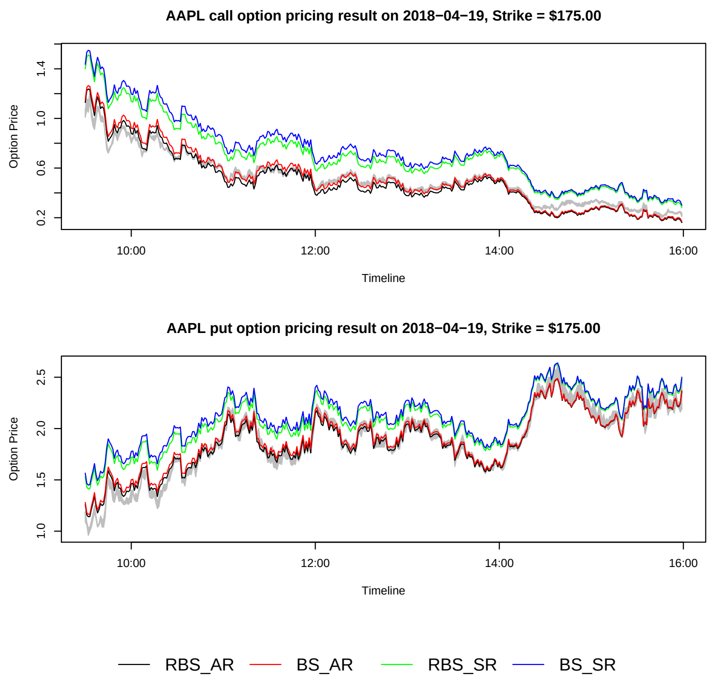


Figure 6: Detailed pricing result on SPY option. The gray shaded area is the bid/ask spread. Best pricing result is from RBS_AR

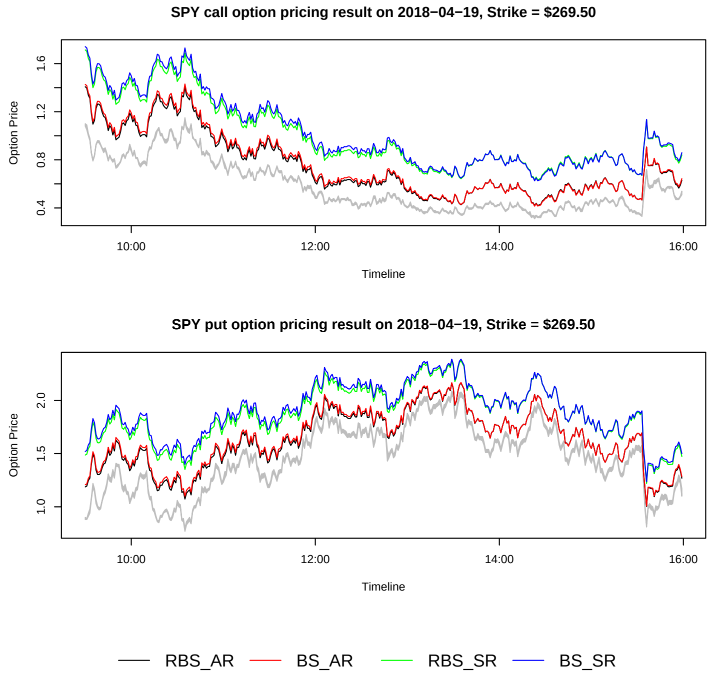


Figure 7: Detailed pricing result on C option. The gray shaded area is the bid/ask spread. Best pricing result is from RBS_AR

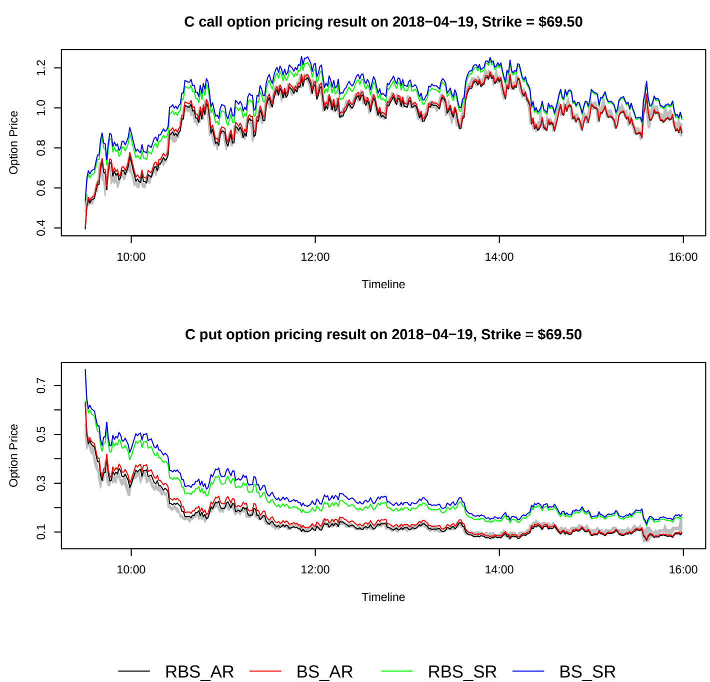
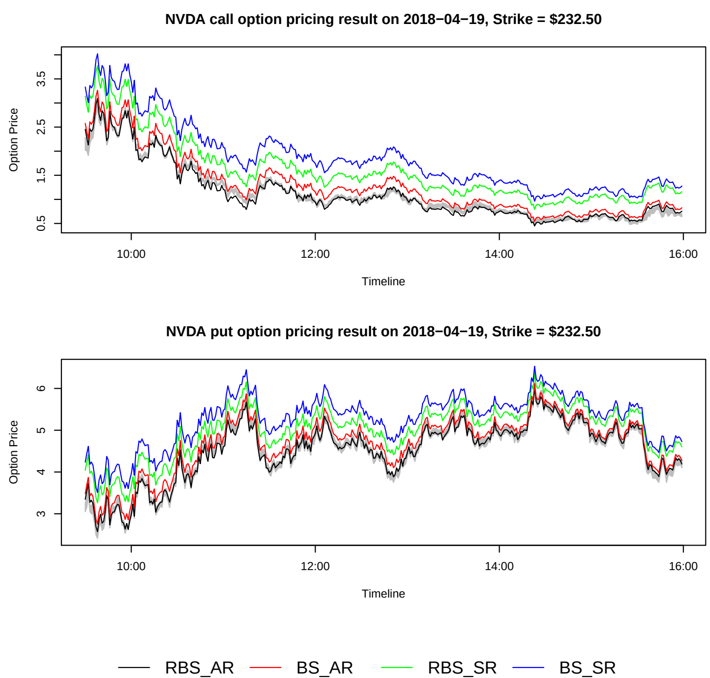


Figure 8: Detailed pricing result on NVDA option. The gray shaded area is the bid/ask spread. Best pricing result is from RBS_AR



C Daily RMSE for all strikes on three consecutive days

Figure 9: Short-term pricing result on three consecutive days. The option price for AAPL on these three days are \$177.81, \$173.76, and \$170.60. All contracts expire on April 20, 2018

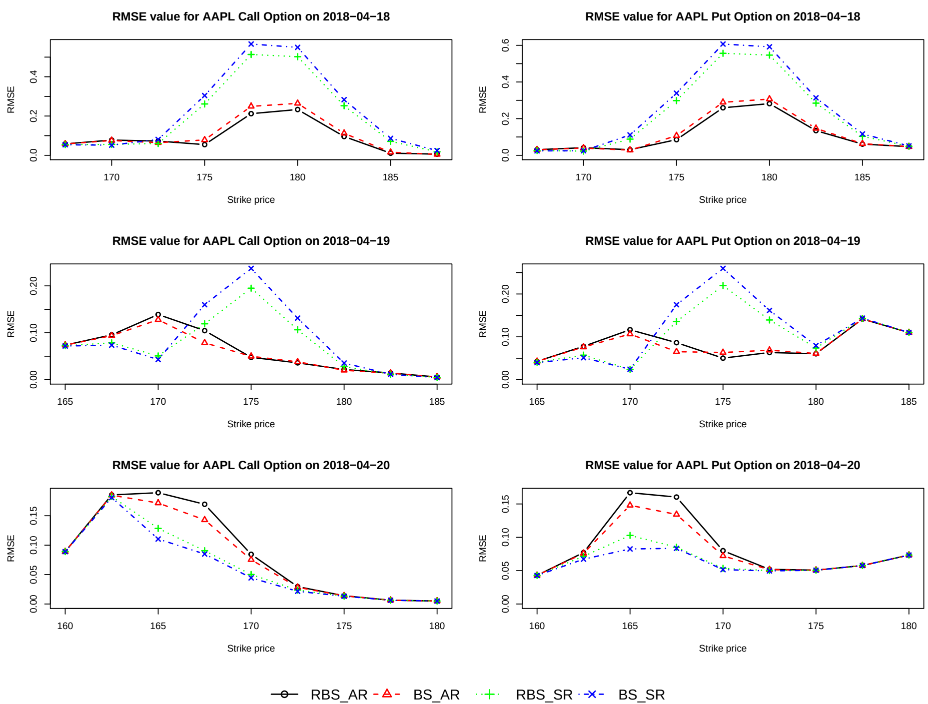


Figure 10: Short-term pricing result on three consecutive days. The option price for SPY on these three days are $270.69, $269.65, and $268.81. All contracts expire on April 20, 2018

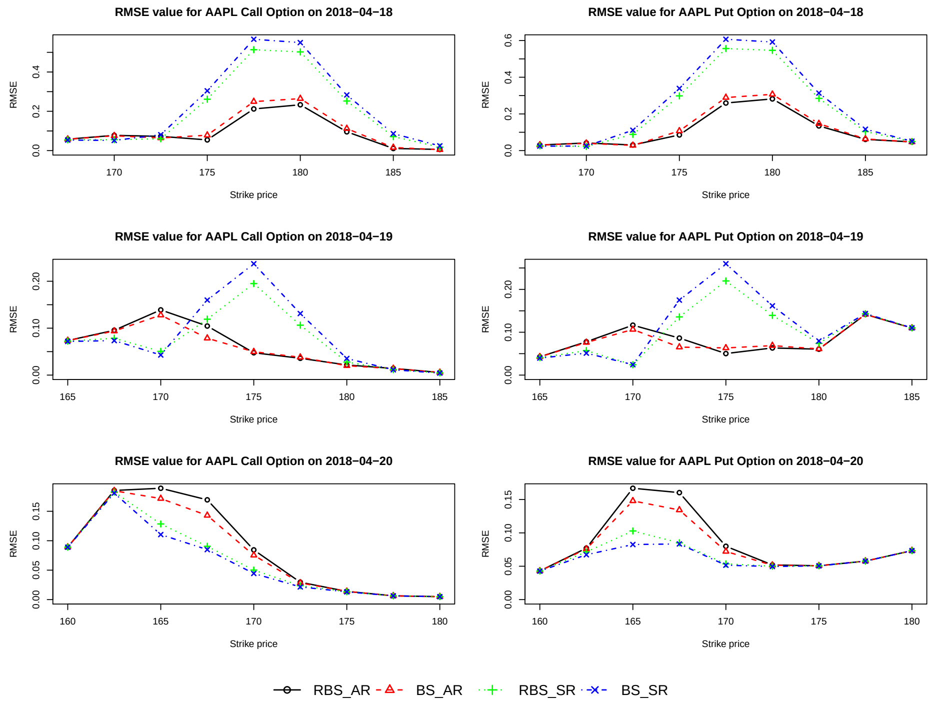


Figure 11: Short-term pricing result on three consecutive days. The option price for C on these three days are $69.81, $69.10, and $70.40. All contracts expire on April 20, 2018

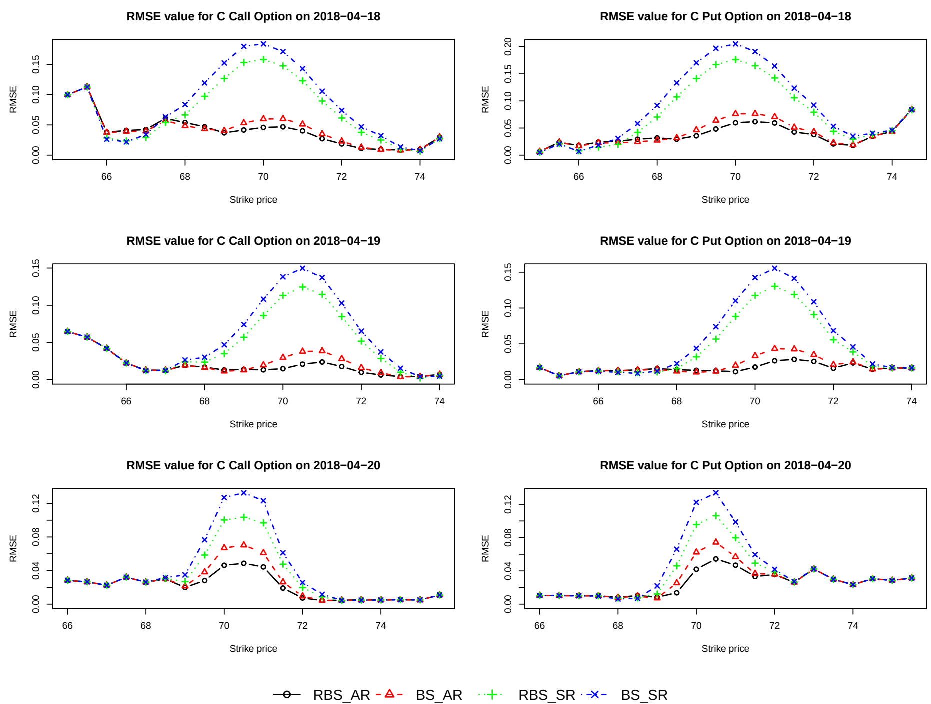


Figure 12: Short-term pricing result on three consecutive days. The option price for NVDA on these three days are \$235.50, \$231.75, and \$228.69. All contracts expire on April 20, 2018