Volatility Uncertainty, Time Decay, and Option Bid-Ask Spreads in an Incomplete Market

PeiLin Hsieh [*], Xiamen University   Robert Jarrow[†], Cornell University

May, 31, 2017

Abstract

This paper documents the fact that in options markets, the (percentage) implied volatility bid-ask spread increases at an increasing rate as the option's maturity date approaches. To explain this stylized fact, this paper provides a market microstructure model for the bid-ask spread in options markets. We first construct a static equilibrium model to illustrate the aforementioned phenomenon where risk averse and competitive option market makers quote bid and ask prices to minimize their inventory risk in an incomplete market with both directional and volatility risk. We extend this model to multi-periods and show that the same phenomenon occurs there as well. Two new implications are generated: a volatility level effect and a volatility variance effect. These implications are empirically tested, and the empirical results confirm the model's validity. Finally, we document the importance of de-trending the maturity effect by showing that the de-trended percentage volatility spread explains future jump intensities better than the original percentage volatility spread.

Key words: option pricing, incomplete market, bid-ask spread, implied volatility, market microstructure, maturity effect

JEL Classification: G10, G13, G14

[*]PeiLin Hsieh is the corresponding author and assistant professor at WISE and SOE, Xiamen University. *Email: ph77@cornell.edu. Tel: 0086-13101438677.* He thanks the grant supported by China National Science Foundation [grant number:71571153], grant supported by Fujian Social Science Foundation [grant number: FJ2015B220], funding by MOE Key Lab of Econometrics and funding by Fujian Key Lab of Statistics.

[†]Robert Jarrow is a Professor at the Department of Economics and the Samuel Curtis Johnson Graduate School of Management, Cornell University. *Email: raj15@cornell.edu. Tel: 607-255-4279. Fax: 607-255-5993*

[‡]The authors would like to thank the editor, associate editor and two referees for their valuable comments which greatly helped to improve the paper.

1 Introduction

This paper documents an intriguing pattern observed in option markets, called the maturity effect. First noticed in the foreign currency (FX) options market, e.g, Chong, Ding and Tan (2003), the maturity effect is when the (percentage) bid-ask implied volatility spreads increase at an increasing rate as an option's maturity approaches. Using daily closing best quotes on S&P Index options over 2001-2010, we confirm the existence of the maturity effect in stock index options.

Our paper makes two contributions to the literature. Although Chong, Ding and Tan (2003) observe the maturity effect for FX options, they compute Black-Scholes-Merton implied volatilities. Since it is well known that the Black-Scholes-Merton model is rejected when pricing FX options, it is an open question whether the observed maturity effect is due to model misspecification. Our first contribution is to show, using model-free implied volatilities, that this is not the case.

Our second contribution is to construct a multi-period equilibrium model to explain the maturity effect. Our model has risk averse and competitive option market makers (hereafter, referred to as MMs) quoting bid and ask prices to minimize their inventory risk in an incomplete market with both directional and volatility risk. It is shown that, although inventory risk decreases as the option's maturity date approaches, an increased risk premium (in term of implied volatility) is needed to compensate the MMs for bearing the unhedgeable risk. This increased risk premium for both bid and ask prices results in the widening of an option's (percentage) bid-ask implied volatility spread as the maturity date approaches.

Two new implications of the model are generated, which are currently untested. The first is that an increase in the level of the underlying's volatility decreases the equilibrium percentage bid-ask implied volatility spread. The second is that, holding the level of the volatility constant, an increase in the volatility's variance increases the equilibrium percentage bid-ask implied volatility spread. Using the same data set employed to document the maturity effect, these additional implications are tested herein. The empirical evidence generally confirms these implications and, thus, the model's validity.

Our paper adds to the market microstructure of option markets literature, which includes among others, Muravyev, Pearson and Broussard (2012), Hsieh, Lee and Yuan (2008), Holowczak, Simaan, and Wu (2007), Chakravarty, Huseying, and Mayhew (2004), Easley, O'Hara and Srinivas (1998), and Pan and Poteshman (2006). In contrast to these papers, which focus on the information content of option trading, our paper focuses on the unhedgeable risks that MMs face in incomplete markets. Additionally, different from Garleanu, Pedersen, and Poteshman (2009) who model the demand side impact on prices, we focus on the supply side of option liquidity providers. Our paper also differs from Chen, Joslin and Ni (2016) who use CRRA utility functions in a complete market to model the optimal consumption choice of investors and dealers. Although they incorporate supply side shocks into the dealers' utility function, risk premium in their model are based on aggregate demands. In contrast, we study an incomplete market where market makers trade-off the expected hedging and inventory holding risks to determine the equilibrium risk premium. Ours is solely a supply side model.

This paper is organized as follows. In section 2 the maturity effect is documented. Section 3 presents the equilibrium model. Next, in section 4, we extend the model to two-periods and then multiple-periods. Section 5 presents the data description and empirical evidence. Finally section 6 concludes.

2 The Maturity Effect

This section presents the *maturity effect*, which is when the (percentage) bid-ask implied volatility spreads increase at an increasing rate as an option's maturity approaches. This pattern was first observed in Chong, Ding and Tan (2003) using Black-Scholes-Merton implied volatilities. Given the Black-Scholes-Merton model is rejected using historical data, it is an open question whether this pattern is a result of model misspecification. Using model free implied volatilities, we show that this is not the case. The existence of the maturity effect motivates the model formulated in the next section to explain the economics underlying this phenomena.

To avoid model misspecification in the computation of implied volatilities, we compute model free implied volatilities (IMV) first proposed by Carr and Madan (1998). Carr and Madan's approach uses only out-of-the-money (OTM) call and put options. A slight modification of their approach is used by the CBOE to compute the VIX index. This paper employs a method similar to the CBOE's, which we call the Carr Madan implied volatility (CM IMV). When constructing the VIX index, the CBOE averages the first and second nearest term contracts.[1] However, here we only investigate the nearest month contract because we are studying the change in volatility spreads over a short time period. We also compute the Black-Scholes-Merton implied volatility (BSM-IMV) for ATM call and put options. We investigate options written on the S&P 500 index from January 2, 2001 to April 18, 2010. Our sample consists of 112 nearest term contracts. For each contract, the trading days range from 12 to 24 days.[2]

In Figure 1, we graph the CM IMV, BSM-IMV, and the daily realized volatility (RV) from 01/02/2005 to 12/31/2009. An IMV is an estimate of the expected volatility over the option's

[1] Because the calculation of the CBOE VIX index is a weight-averaged implied volatility of the two nearest term contracts, the VIX index is an expected volatility for the next 30 days.

[2] In our sample, the contract which expires on January 20, 2001 does not include the records before 2001 so that this contract only has 12 trading days.

remaining life. We use 20-days realized volatility for comparison.[3] We calculate the BSM-IMV for call and put options separately and average the IMV of the four options with strikes closest to the spot price to get the daily BSM-IMV for at-the-money (ATM) calls and BSM-IMV ATM puts. As seen, all the volatility measures are highly correlated reaching a maximum, above 80%, during the 2008 financial crisis. The related descriptive statistics are listed in Table 1 and discussed in a later section.

Figures 2 and 3 document the maturity effect. In each figure, there are 3 charts representing spreads of different volatility measures. The top chart shows the CM IMV spread, the middle and the bottom charts plot the spreads of BSM-IMV of ATM call options and of BSM-IMV ATM put options, respectively. Figure 2 graphs the absolute volatility spreads, while Figure 3 graphs the percentage volatility spreads, which divides the absolute volatility spread by the realized volatility of the corresponding contract period. The vertical dashed lines represent expiration dates, and the dots denote the volatility spreads of 40 consecutive nearest-term contracts from 01/02/2007 to 04/15/2010. Finally, the RV is represented by a bold line for a comparison.

As shown in Figure 2, the implied volatility bid-ask spreads for all IMV measures increase at an increasing rate as the option's maturity approaches. The CM-IMV spread is relatively higher than the BSM-IMV spread, and all the absolute volatility spreads start at approximately $1 \sim 2\%$ and increase to more than 4% at a few days before expiration.[4] As seen, the maturity effect occurs for every contract and for the different volatility measures. The range of the absolute volatility spread is more volatile and increases as the realized volatility increases. Figure 3 contains the percentage implied volatility bid-ask spreads and shows that, in contrast to absolute volatility spreads, the range of the percentage volatility spread is similar over different contracts but larger. The percentage volatility spread starts from about $5 \sim 7\%$ at a month remaining before maturity and increases to over 20%. Similar to the volatility spreads, the percentage spreads increase at an increasing rate. The maturity effect occurs for every contract.

In markets, the dollar spreads paid for a transaction are usually smaller if the transaction prices are lower. Due to discreteness in quoted prices (the tick size), when transaction prices become very small, a minimum discrete dollar bid-ask spread implies that the percentage bid-ask spread increases. Although these effects exist in our option prices, they do not explain the patterns observed in figures 2 and 3 because the volatility measures are independent of the price level. In contrast, however, if the quoted spreads are always at the minimum discrete dollar bid-ask, i.e., 0.05,[5] then the transformed

[3] We use the definition of RV given by Andersen, Bollerslev, Diebold and Ebens (2001).

[4] Chong, Ding and Tan (2003) investigate at-the-money currency options. They show that the volatility spread starts at $2 \sim 6\%$ for options with a 1 year maturity, increases gradually until the last month, increases rapidly in the last month, and finally reaches $8 \sim 16\%$.

[5] S&P options are quoted on a point basis, and the minimum tick is 0.05 for option prices less than 3 and 0.1 for

volatility spreads will increase as a contract approaches its expiration. Figure 4 shows the distribution of the quoted spreads in ticks. There are less than 5 percent of the quotes occurring at the minimum tick. Therefore, the observed pattern can not be attributed to most quotes being at the minimum discrete dollar bid-ask spread.

3 The One-Period Model

This section presents a static model of an options market equilibrium where prices are determined by competitive MMs optimally hedging their inventory risk contingent on the execution of a trade. Markets are assumed to be incomplete because the underlying stock exhibits both directional and volatility risk. MMs, being risk averse, minimize inventory risk by delta hedging, but delta hedging can not eliminate both directional and volatility risks. Equilibrium spreads, therefore, necessarily include compensation (risk premium) for the unhedged risk. We show that this compensation, in conjunction with the time decay embedded in the option's price, generates the maturity effect in the implied volatility bid/ask spreads as an option approaches maturity.

It should be noted that the volatility risk in our incomplete market model implies that the Black-Scholes-Merton (BSM) model is invalid and that vega hedging will not remove an option's volatility risk because the BSM pricing formula does not reflect the sensitivity of the option's value to changes in volatility (see Chatterjea and Jarrow [27] for an elaboration of this comment).

3.1 The Economy

This section describes the economy. Trading in the economy are a riskless asset, a stock, and a call option on the stock with strike price K and maturity time T. Let the stock price at time 0 be denoted P_0, the stock's time 0 volatility σ_0, and the annualized time to maturity τ, measuring time length from 0 to T. We let $P_0 > 0$ and $\sigma_0 = \sigma > 0$. We assume two volatility states, so the stock's volatility realization at time T is denoted $\sigma_\tau \in \{\sigma_L, \sigma_H\}$,[6] and the stock price can increase or decrease at time T for each volatility level, i.e.

$$P_T = P_0 e^{\{\mu\tau + \sigma_\tau W_\tau\}},$$

where μ is the expected return per unit time, and W_τ is Wiener process. This assumption states that the MMs believe $\ln[E(\frac{P_T}{P_0})]$ is distributed $N(\mu_L, \sigma_L)$ with probability $1 - \phi$ and $N(\mu_H, \sigma_H)$

prices larger than 10.

[6] For simplicity of the derivation, we assume $\sigma_H = \sigma \cdot H$ and $\sigma_L = \sigma \cdot L$ for a fixed $\sigma > 0$ where $L < 1 < H$, see sections 3.6 and 3.7.

with the probability ϕ. These are the MM's subjective probabilities.

The collection of competitive and homogeneous MMs quote bid and ask prices for the traded option. Let the option's equilibrium ask price at time 0 be denoted $C_0(a)$ and the bid price be denoted $C_0(b)$. The call option's payoff at time T is

$$C_T(\sigma_\tau) = \max[P_T(\sigma_\tau) - K, 0]. \tag{1}$$

3.2 Arbitrage Pricing

We assume that the stock and option market are arbitrage-free. The first fundamental theorem of asset pricing implies that there exist risk-neutral probabilities such that

1. the stock price equals its expected (discounted) value under the risk-neutral probabilities, i.e.

$$\begin{aligned} P_0 &= e^{-r\tau}\left\{(1-\Phi)\int_{-\infty}^{\infty} P_t \cdot f(P_t,\sigma_L,\tau)dP_t + \Phi\int_{-\infty}^{\infty} P_t \cdot f(P_t,\sigma_H,\tau)dP_t\right\} \\ &= e^{-r\tau}\left\{(1-\Phi)\,E_L(P_t) + \Phi E_H(P_t)\right\}, \end{aligned} \tag{2}$$

   where Φ is the risk-neutral probability that the volatility is (σ_H), $f(P_t, \sigma_L, \tau)$ is the conditional risk-neutral probability density function of the stock price given the volatility is (σ_L), and $f(P_t, \sigma_H, \tau)$ is the conditional risk-neutral probability density function given the volatility is (σ_H); and

2. the ask and bid call option prices equal their expected (discounted) values under the risk-neutral probabilities, i.e.

$$\begin{aligned} C_0(s) &= e^{-r\tau}\left\{(1-\Phi_s)\,E_L\left[\max\left(P_T-K,0\right)\right] + \Phi_s E_H\left[\max\left(P_T-K,0\right)\right]\right\} \\ &= e^{-r\tau}\left\{(1-\Phi_s)\,E_L\left[C_T\right] + \Phi_s E_H\left[C_T\right]\right\}, \end{aligned} \tag{3}$$

   where Φ_s is the ask or the bid price ($s \in \{a, b\}$) risk-neutral probability that the volatility is (σ_H).

We note that the conditional risk-neutral probabilities $\{f(\sigma_L, \tau), f(\sigma_H, \tau)\}$ are unique because, given the volatility, the market is complete. Since the unconditional market is incomplete, however, the risk-neutral probabilities $\{\Phi, \Phi_a, \Phi_b\}$ are non-unique and not determined by arbitrage arguments alone.[7] Indeed, any call price between the quotes satisfying the no-arbitrage conditions corresponds

[7] This is due to the second fundamental theorem of asset pricing.

to an acceptable Φ. To determine the unique arbitrage-free ask/bid call option prices, an equilibrium model is needed. Such a model is given in the next section.

To facilitate a solution, we assume that conditional upon the volatility's realization $\{\sigma_L, \sigma_H\}$, the MMs' subjective probabilities satisfy the following condition, called the *independence assumption*:

$$E_L(\frac{P_T}{P_0})e^{-r\tau} = 1 \quad \text{and} \quad E_H(\frac{P_T}{P_0})e^{-r\tau} = 1. \tag{4}$$

This condition essentially states that the expected stock price doesn't depend on the volatility, and that the expected return from holding the underlying asset equals the risk free rate.[8] This assumption best describes an economy where traders' risk premium are independent of the level of an asset's volatility. In the Appendix, we show that after we impose this assumption (4), both $E_L[\max(P_T - K, 0)]$ and $E_H[\max(P_T - K, 0)]$ satisfy the Black-Scholes-Merton formula.

3.3 The Market Maker's Problem

Because a mean-variance objective provides tractable results, many papers employ it to study asset allocation and hedging in an incomplete market. For example, in futures and forward markets, Lioui and Poncet (2000) and In and Kim (2006) use this objective to study optimal hedging policy, while Brooks, Henry, and Persand (2002) use it to investigate an interest rate market with stochastic volatility. In option markets, it has been used by Bakshi, Cao, and Chen (1997). Viswanathan and Wang (2002) use it to model the MMs' quoting behavior for limit order books. Last, Basak and Chabakauri (2010, 2012) study both dynamic portfolio choice and dynamic hedging using a minimum-variance objective function. Following this literature, we use a mean-variance objective function to study a MM's optimal hedging strategy. We first do this for a one-period model to illustrate the results in the simplest setting, and then we generalize to a two-period and a multi-period model.

As such, we assume that the options market is populated by a collection of homogeneous MMs with mean-variance utility functions, who delta hedge their option positions using the underlying stock to minimize both directional and volatility risk. They behave in this manner because they realize they are uninformed traders. Given markets are incomplete, the delta hedge cannot remove all of the position's risk.

MMs quote bid and ask option prices to compensate themselves for the inventory risk they bear contingent on the transaction. To simplify the analysis, since the MMs can hedge the option's

[8] The justification for this statement is the following: $E_L(\frac{P_T}{P_0})e^{-r\tau} = e^{\mu_L\tau + \frac{\sigma_L^2\tau}{2} - r\tau} = 1 \Rightarrow \mu_L\tau + \frac{\sigma_L^2\tau}{2} = r\tau$ and $E_H(\frac{P_T}{P_0})e^{-r\tau} = e^{\mu_H\tau + \frac{\sigma_H^2\tau}{2} - r\tau} = 1 \Rightarrow \mu_H\tau + \frac{\sigma_H^2\tau}{2} = r\tau$.

directional risk given the volatility, we assume that their conditional beliefs equal the risk-neutral beliefs, $\{f(\sigma_L,\tau), f(\sigma_H,\tau)\}$. Combined with the independence assumption expression (4), this implies that the MMs use these risk-neutral probabilities to price "directional risk." We let the MM's unconditional probability beliefs over the volatility's realizations (σ_L, σ_H) be denoted by $(1-\phi, \phi)$. These beliefs differ from the corresponding risk-neutral probabilities Φ given earlier.

The MM's provide liquidity to investors. Given a price, they quote option quantities to maximize their expected utility, i.e. they choose the trade quantity to solve the following optimization problem:

$$\max_{q,\ \Delta_a}\left(qE(W_a)-\frac{1}{\gamma}Var(qW_a)-cq\right) \quad \text{where} \tag{5}$$

$$W_a(\sigma_T)=C_0(a)e^{r\tau}-C_T(\sigma_T)+\Delta_a\left[P_T\left(\sigma_T\right)-P_0e^{r\tau}\right]$$

is the profit from the delta hedged option portfolio at time T for $\sigma_T\in\{\sigma_L,\sigma_H\}$, c is a transaction or opportunity cost for a trade, $\gamma>0$ is a risk aversion coefficient, q is the quantity quoted, and Δ_a is the number of shares held in the underlying asset (the delta).[9] Note that $C_0(a)-\Delta_a P_0$ is the amount invested in the portfolio at time 0.

For simplicity, our discussion focuses on the maximization problem of writing a call option and choosing the ask price. The problem of buying a call and determining the bid price is analogous, and the results are reported below. All derivations are given in Appendix B.

Because of expression (4) and the fact that the MM's conditional beliefs equal the risk-neutral conditional probabilities, the stock's expected payoff is zero. This implies that the MM focuses on minimizing volatility risk. In this regard, one can show that the MM's expected payoff $E(W_a)$ is the option's selling price minus the expected fair value of the option conditional upon knowledge of the volatility, i.e.

$$E(W_a)=C_0(a)e^{r\tau}-\phi E_H\left[C_T\right]-(1-\phi)E_L\left[C_T\right] \quad \text{where} \tag{6}$$

$$E_H\left[C_T\right]\equiv\int_{-\infty}^{\infty}\max\left(P_t-K,0\right)\cdot f(P_t,\sigma_H,\tau)dP_t, \text{ and}$$
$$E_L\left[C_T\right]\equiv\int_{-\infty}^{\infty}\max\left(P_t-K,0\right)\cdot f(P_t,\sigma_L,\tau)dP_t.$$

The key insight here is that the MM's expected profit does not depend on the shares held in the underlying stock.

[9] Constructing the optimization problem for the bid price is analogous to the ask price, and the portfolio's profit at the maturity date is

$$W_b(\sigma_T)=-C_0(b)e^{r\tau}+C_T(\sigma_T)+\Delta_b\left[P_T\left(\sigma_T\right)-P_0e^{r\tau}\right].$$

The optimization gives the sign for Δ_b. The expected profit from buying options is not the same as writing options. The expected profit for writing is $C_0(a)e^{r\tau}-\phi E_H\left[C_T\right]-(1-\phi)E_L\left[C_T\right]$ and for buying it is $\phi E_H\left[C_T\right]+(1-\phi)E_L\left[C_T\right]-C_0(b)e^{r\tau}$.

The optimal solution for writing call options satisfies the following first order conditions.

$$\begin{aligned} &E(W_a) - \tfrac{2q}{\gamma} Var(W_a) - c = 0 \\ &\tfrac{\partial Var(W_a)}{\partial \Delta_a} = 0 \end{aligned} \tag{7}$$

where

$$\begin{aligned} Var(W_a) = {} & \phi Var_H [W_a] + (1-\phi) Var_L [W_a] \\ & + \phi(1-\phi)(E_L [C_T] - E_H [C_T])^2 \end{aligned} \tag{8}$$

The solution is

$$\begin{aligned} &q^* = \tfrac{\gamma\{C_0(a) - \phi E_H[C_T] - (1-\phi)E_L[C_T] - c\}}{2Var(W_a)} \\ &\text{if } C_0(a) > \phi E_H [C_T] + (1-\phi) E_L [C_T] + c \\ &q^* = 0 \\ &\text{if } C_0(a) \le \phi E_H [C_T] + (1-\phi) E_L [C_T] + c \end{aligned} \tag{9}$$

Moreover, the optimal delta Δ_a^* is determined independently of q^* because the expected return from holding the stock is zero. Therefore, the stock holdings only affect the portfolio's variance. The optimal delta positions are therefore

$$\Delta_a^* = \tfrac{(1-\phi)Cov_L[P_T, C_T] + \phi Cov_H[P_T, C_T]}{(1-\phi)Var_L(P_T) + \phi Var_H(P_T)} \quad \text{and} \tag{10}$$

$$\Delta_b^* = -\tfrac{(1-\phi)Cov_L[P_T, C_T] + \phi Cov_H[P_T, C_T]}{(1-\phi)Var_L(P_T) + \phi Var_H(P_T)} \tag{11}$$

where Δ_a^* and Δ_b^* are the optimal delta hedging positions for writing and buying a call option, respectively. The derivation is shown in Appendix B-2. Intuitively, and as shown in the appendix, Δ_a^* (Δ_b^*) is approximately 0.5 (-0.5) for at-the-money call options, and that Δ_a^* (Δ_b^*) approaches 1 (-1) when the strike price approaches 0.

3.4 Equilibrium under Perfect Competition

We assume that there are N MMs quoting option prices on a competitive exchange, subject to exchange rules. We assume that the exchange rules require that a MM's quotes must be for Q shares. This quoting quantity is set for investor convenience. This fixed quoting quantity is not the same as the actual market demands nor the equilibrium trading volume. Market demand and trading volume are determined by the interaction among the MMs and the investors, which is formally outside the structure of the model.

The competitive market assumption is captured by the condition that, in equilibrium, the market marker's optimal utility equals her reservation utility level, which is assumed to be zero. Combined, these two conditions imply that the equilibrium call option's ask price $C_0(a)$ satisfies

$$E(W_a) - \frac{Q}{\gamma}Var(W_a) - c = 0, \ i.e. \tag{12}$$

which is

$$\begin{aligned} C_0(a)e^{r\tau} &= \phi E_H\left[C_T\right] + (1-\phi)E_L\left[C_T\right] + \frac{Q}{\gamma}Var(W_a) + c \\ &= E\left[C_T\right] + \frac{Q}{\gamma}Var(W_a) + c \end{aligned} \tag{13}$$

Similarly, the equilibrium bid price $C_0(b)$ is

$$E(W_b) - \frac{Q}{\gamma}Var(W_b) - c = 0, \ i.e. \tag{14}$$

$$C_0(b)e^{r\tau} = E\left[C_T\right] - \frac{Q}{\gamma}Var(W_b) - c \tag{15}$$

Expression (13) shows that the equilibrium ask price is composed of three parts: (1) the option's expected price, $E\left[C_T\right]$, (2) the required risk premium for inventory risk, $\frac{Q}{\gamma}Var(W_a)$, and (3) the compensation for transaction costs, c. A similar interpretation holds for the option's equilibrium bid price.[10] The dollar spread of the quoted prices is $\frac{Q}{\gamma}Var(W_a)+\frac{Q}{\gamma}Var(W_b) + 2c$.

3.5 Risk-Neutral Probabilities

This section characterizes the unique risk-neutral ask and bid probabilities determined by the options market equilibrium. In this regard, the ask price risk-neutral probability, Φ_a, is determined as the solution to

$$C_0(a)e^{r\tau} = \phi E_H\left[C_T\right] + (1-\phi)E_L\left[C_T\right] + \frac{Q}{\gamma}Var(W_a) + c \tag{16}$$

$$= \mathbb{E}^*\left[\mathrm{C_T}\right] = \Phi_a E_H\left[C_T\right] + (1-\Phi_a)E_L\left[C_T\right] \tag{17}$$

[10] This equilibrium solution is common for agents with mean-variance utility functions. In Viswanathan and Wang (2006), their strategic bid equilibrium for the limit order book in a stock market is $p_i(x) = \overline{v} - \frac{\sigma_v^2}{\gamma}\left[\frac{\theta+(N-1)x}{N(1+\theta)-1}\right]$ where p is the price for stock i, $\overline{v}$ is the mean asset value, σ_v^2 is the variance of the asset value, γ is the risk attitude, N is the number of market makers, x is the submitted order, and θ is a parameter indicating the subjective belief for the order arrival intensity.

which is

$$\begin{aligned}\Phi_a &= \frac{\left(\phi E_H\left[C_T\right]+(1-\phi)E_L\left[C_T\right]+\frac{Q}{\gamma}Var(W_a)+c\right)-E_L\left[C_T\right]}{E_H\left[C_T\right]-E_L\left[C_T\right]}\\ &= \phi+\frac{1}{E_H\left[C_T\right]-E_L\left[C_T\right]}\left[\frac{Q}{\gamma}Var(W_a)+c\right]\end{aligned} \tag{18}$$

where $\mathbb{E}^*[\cdot]$ corresponds to expectation under the risk-neutral probability measure.

A similar expression holds for the bid price risk-neutral probability as well. Obviously, the risk neutral probability Φ_a is higher than the subjective probability because the option seller requires a risk premium for being unhedged. The more volatile the incomplete hedging outcomes, the larger Φ_a is relative to ϕ. This result is consistent with the variance risk premium in Carr and Wu (2008). It is also consistent with Coval and Shumway (2001), Bakshi and Kapadia (2003), and Broadie, Chernov and Johannes (2009) who show that an option's purchase price includes a risk premium and that a long option position has negative profits (even after delta hedging).

3.6 Maturity Dependence

The purpose of this section is to obtain comparative statics relating to equilibrium ask/bid option prices' *implied volatilities* and the option's maturity τ. As noted before, given the independence assumption 4, it is shown in Appendix B-1 that the the formula of $E_H\left[C_T\right]$ and $E_L\left[C_T\right]$ is the Black-Scholes-Merton model. Next, define the *ask and bid implied volatilities*, denoted $A\cdot\sigma$ and $B\cdot\sigma$ for constants $A, B>0$, respectively as the solutions to the following equations:

$$\begin{aligned}C_0(a) &= \mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+r\tau+(A\sigma)^2\tau/2}{A\sigma\sqrt{\tau}}\right)-K\cdot e^{-r\tau}\cdot\mathbf{N}(\frac{\ln(\frac{P_0}{K})+r\tau-(A\sigma)^2\tau/2}{A\sigma\sqrt{\tau}})\\ C_0(b) &= \mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+r\tau+(B\sigma)^2\tau/2}{B\sigma\sqrt{\tau}}\right)-K\cdot e^{-r\tau}\cdot\mathbf{N}(\frac{\ln(\frac{P_0}{K})+r\tau-(B\sigma)^2\tau/2}{B\sigma\sqrt{\tau}})\end{aligned} \tag{19}$$

where $C_0(a)$ and $C_0(b)$ are the equilibrium prices. It can also be shown that the unconditional variance of the position is[11]

$$\begin{aligned}Var(W_a) = \quad &\phi Var_H\left[W_a\right]+(1-\phi)Var_L\left[W_a\right]\\ &+\phi(1-\phi)(E_L\left[C_T\right]-E_H\left[C_T\right])^2\end{aligned} \tag{20}$$

where Var_H and Var_L denote the variance of the portfolio given the volatility. Substitution yields[12]

$$\begin{aligned}&Var_H\left[W_a\right]=Var_H\left\{C_0(a)e^{r\tau}-C_T+\Delta_a\left[P_T-P_0e^{r\tau}\right]\right\}\\ &=C+A_1\cdot\mathbf{N}(2\sigma H\sqrt{\tau}+z_H)+A_2\cdot\mathbf{N}(\sigma H\sqrt{\tau}+z_H)+A_3\cdot\mathbf{N}(z_H)\end{aligned} \tag{21}$$

[11]The derivation is provided in Appendix B-1.

[12]In the proposition, we assume that $\sigma_H=\sigma\cdot H$ and $\sigma_L=\sigma\cdot L$ to simplify the derivation.

where

$$\begin{cases} C = P_0^2 \Delta_a^2 e^{2r\tau} \left[e^{\sigma^2 H^2 \tau} - 1 \right] \\ A_1 = P_0^2 e^{2r\tau + \sigma^2 H^2 \tau} (1 - 2\Delta_a) \\ A_2 = -P_0^2 e^{2r\tau} \mathbf{N}(\sigma H \sqrt{\tau} + z_H) - 2P_0 K e^{r\tau} \mathbf{N}(-z_H) + 2\Delta_a P_0 e^{r\tau}(K + P_0 e^{r\tau}) \\ A_3 = K^2 \mathbf{N}(-z_H) - 2\Delta_a P_0 K e^{r\tau} \\ z_H = \frac{\ln(K) - E_H(\frac{P_T}{P_0})}{\sigma H \sqrt{\tau}} = \frac{\ln(K) - e^{r\tau}}{\sigma H \sqrt{\tau}} . \end{cases}$$

3.7 Equilibrium Implications

This section shows that the previous model is consistent with the observed call option bid/ask implied volatility spread increasing at an increasing rate as the time to maturity approaches. In addition, new testable implications of our model are generated. These implications follow from various partial derivatives of the equilibrium bid and ask implied volatilities using expressions (6), (12) and (21).

Proposition 1 (The Maturity Effect)

For small τ, the bid-ask implied volatility spread of ATM options increases as maturity decreases ($\frac{\partial A}{\partial \tau} < 0, \frac{\partial B}{\partial \tau} > 0$ and $\frac{\partial (A-B)}{\partial \tau} < 0$).

The proposition confirms the maturity effect documented in section 2. Figure 5 illustrates the intuition behind this result. Point E_1 gives the first equilibrium where the MMs' expected profit, $E(W_a(E_1))$ valued at $A_1\sigma$, equals the required risk premium. As the maturity date approaches, expected profits $E(W_a)$ decrease to a level where the modified risk premium increases. To continue quoting, MMs need to increase the selling volatility to $A_2\sigma$ so that $E(W_a(E_2))$ increases to reflect the increased risk premium at the second equilibrium point E_2. As a result, as the time to maturity approaches, implied volatilities increase.

Two considerations explain our result. First, the volatility decays proportionately to the square root of the time remaining to maturity, which implies a non-linear rate of decay in the option's price. Second, MMs have mean-variance utility functions, implying a linear risk premium. Combining these two effects, expression (22) shows the relation between the time decay of the expected profits and the time decay of the risk premium charged by the market markers.

$$\frac{\partial E(W_a)}{\partial \tau} - \frac{Q}{\gamma} \frac{\partial Var(W_a)}{\partial \tau} = \text{Decay of Extra Profits} + \text{Decay of Risk Premium} \tag{22}$$

First, we compute the time decay of the extra profits:

$$\begin{aligned} \frac{\partial E(W_a)}{\partial \tau} &= \frac{K\sigma}{2\sqrt{\tau}} \left[\mathbf{n}(z_A)A - \phi \cdot \mathbf{n}(z_H)H - (1-\phi) \cdot \mathbf{n}(z_L)L \right] \\ &+ rP_0 e^{r\tau} \left[\mathbf{N}(z_A + \sigma A\sqrt{\tau}) - \phi \cdot \mathbf{N}(z_H + \sigma H\sqrt{\tau}) - (1-\phi) \cdot \mathbf{N}(z_L + \sigma L\sqrt{\tau}) \right]. \end{aligned}$$

To compute $\frac{\partial Var(W_a)}{\partial \tau}$, an intermediate calculation is

$$\begin{aligned}\frac{\partial Var_H}{\partial \tau} &= D + B_1\mathbf{N}(2\sigma H\sqrt{\tau} + z_H) + B_2\mathbf{N}(\sigma H\sqrt{\tau} + z_H) + B_3\mathbf{N}(z_H) \\ &+ \left[K^2\mathbf{N}(z_H) - KP_0e^{r\tau}\mathbf{N}(\sigma H\sqrt{\tau} + z_H) - \Delta_a K^2 + \Delta_a KP_0e^{r\tau}\right]\frac{\sigma H}{\sqrt{\tau}}\mathbf{n}(z_H)\end{aligned}$$

where

$$\begin{cases} D = P_0^2\Delta_a^2\left[(2r + \sigma^2H^2)e^{2r\tau+\sigma^2H^2\tau} - 2re^{2r\tau}\right] \\ B_1 = P_0^2(2r + \sigma^2H^2)(1 - 2\Delta_a)e^{2r\tau+\sigma^2H^2\tau} \\ B_2 = -2rP_0^2e^{2r\tau}\mathbf{N}(\sigma H\sqrt{\tau} + z_H) - 2rKP_0e^{r\tau}\mathbf{N}(-z_H) + 2r\Delta_a KP_0e^{r\tau} + 4r\Delta_a P_0^2e^{2r\tau} \\ B_3 = -2r\Delta_a KP_0e^{r\tau}\ . \end{cases}$$

Finally, the last term in the unconditional variance (20) is

$$\begin{aligned}&\frac{\partial(E_H - E_L)^2}{\partial \tau} \\ &= [E_H\{C_T\} - E_L\{C_T\}] \times \frac{\partial(E_H\{C_T\} - E_L\{C_T\})}{\partial \tau} \\ &= [E_H\{C_T\} - E_L\{C_T\}] \times \left[\frac{K\sigma}{2\sqrt{\tau}}(\mathbf{n}(Z_H)H - \mathbf{n}(Z_L)L) + rP_0e^{r\tau}(\mathbf{N}(z_H + \sigma H\sqrt{\tau}) - \mathbf{N}(z_L + \sigma L\sqrt{\tau}))\right].\end{aligned}$$

For ATM call options, i.e. $K \doteqdot P_0e^{r\tau}$, we have

$$\mathbf{n}(z_H) = \mathbf{n}(z_L) = \mathbf{n}(z_A) \to \frac{1}{\sqrt{2\pi}} \quad and \quad \mathbf{N}(z_H) = \mathbf{N}(z_L) = \mathbf{N}(z_A) \to \frac{1}{2}, \quad when \quad \tau \to 0$$

This generates the limiting behavior of $\frac{\partial E(W_a)}{\partial \tau}$ and $\frac{\partial Var(W_a)}{\partial \tau}$ as follows.

$$\frac{\partial E(W_a)}{\partial \tau} = \frac{K\sigma}{\sqrt{2\pi\tau}}(A - \phi H - (1 - \phi)L) + o(\frac{1}{\sqrt{\tau}}) \tag{23}$$

$$\begin{aligned}&\frac{\partial Var(W_a)}{\partial \tau} \\ &= \phi P_0^2\sigma^2H^2(\Delta_a^2 - \Delta_a + \frac{1}{2} - \frac{1}{2\pi}) + (1 - \phi)P_0^2\sigma^2L^2(\Delta_a^2 - \Delta_a + \frac{1}{2} - \frac{1}{2\pi}) \\ &\quad +\phi(1 - \phi)\frac{K^2\sigma^2(H - L)^2}{4\pi} + o(1)\end{aligned} \tag{24}$$

As $\tau \to 0$, expression (23) approaches positive infinity, while expression (24) approaches a constant, which implies that $\frac{\partial A}{\partial \tau} < 0$.[13] Our next result proves that the maturity effect increases and at an increasing rate as the maturity date approaches.

Proposition 2 (The Increasing Rate of the Maturity Effect)

For small τ, the bid-ask implied volatility spread of ATM options increases at an increasing rate as the maturity date approaches ($\frac{\partial^2 A}{\partial \tau^2} > 0$, $\frac{\partial^2 B}{\partial \tau^2} < 0$ and $\frac{\partial^2 (A-B)}{\partial \tau^2} > 0$).

[13] The derivation is in Appendix B-3.

We first show the following inequality.

$$\frac{\partial^2 A}{\partial \tau^2} = -\frac{\partial \left[\frac{\partial f}{\partial \tau} / \frac{\partial f}{\partial A}\right]}{\partial \tau} = \frac{\frac{\partial f}{\partial \tau}\frac{\partial^2 f}{\partial A \partial \tau} - \frac{\partial^2 f}{\partial \tau^2}\frac{\partial f}{\partial A}}{\left(\frac{\partial f}{\partial A}\right)^2} > 0$$

where $f(A, \sigma, L, H, \tau; \phi) = E(W_a) - \frac{Q}{\gamma}Var(W_a) - c = 0$. Consider the numerator. Using the options pricing formula at (19), we have

$$\frac{\partial f}{\partial A} = K\sigma\sqrt{\tau} \cdot \mathbf{n}(z_A) > 0$$
$$\frac{\partial^2 f}{\partial A \partial \tau} = K\sigma\sqrt{\tau} \cdot \mathbf{n}(z_A)\left[\frac{1}{2\tau} - z_A \times \frac{\partial z_A}{\partial \tau}\right]$$

For ATM call options, we have

$$\frac{\partial^2 E(W_a)}{\partial \tau^2} = \frac{-K\sigma}{2\sqrt{2\pi\tau^3}}\left[A - \phi H - (1-\phi)L\right] + o(\frac{1}{\sqrt{\tau^3}}) \tag{25}$$

$$\frac{\partial Var(W_a)}{\partial \tau} = O(1) \tag{26}$$

$$\frac{\partial^2 f}{\partial A \partial \tau} = \frac{K\sigma \cdot \mathbf{n}(z_A)}{2\sqrt{\tau}} + o(\frac{1}{\sqrt{\tau}}). \tag{27}$$

As $\tau \to 0$, (25) becomes negative, expression (26) approaches a positive constant, and so $\frac{\partial^2 f}{\partial \tau^2}$ is negative. Also, as $\tau \to 0$, expression (27) becomes positive. Therefore, as time approaches the maturity date, the ask volatility increases at an increasing rate.

The previous two results show that our model is able to generate the patterns observed in section 2. We now provide some new testable implications of our model.

Proposition 3 (The Volatility Level Effect)

For small τ, the bid-ask implied volatility spread of ATM options decreases as the underlying's volatility level σ increases ($\frac{\partial A}{\partial \sigma} < 0$, $\frac{\partial B}{\partial \sigma} > 0$ and $\frac{\partial (A-B)}{\partial \sigma} < 0$).

To prove this result, we need to show the following inequality.

$$\frac{\partial A}{\partial \sigma} = -\frac{\frac{\partial f}{\partial \sigma}}{\frac{\partial f}{\partial A}} < 0$$

Again, we have shown that $\frac{\partial f}{\partial A} = K\sigma\sqrt{\tau} \cdot \mathbf{n}(z_A) > 0$; therefore, we only need to determine the sign of $\frac{\partial f}{\partial \sigma} = \frac{\partial E(W_a)}{\partial \sigma} - \frac{Q}{\gamma}\frac{\partial Var(W_a)}{\partial \sigma}$.

For $\frac{\partial E(W_a)}{\partial \sigma}$, we have

$$\frac{\partial E(W_a)}{\partial \sigma} = K\left[\mathbf{n}(z_A)A\sqrt{\tau} - \phi \cdot \mathbf{n}(z_H)H\sqrt{\tau} - (1-\phi)\cdot \mathbf{n}(z_L)L\sqrt{\tau}\right]. \tag{28}$$

For $\frac{\partial Var(W_a)}{\partial \sigma}$, we first calculate $\frac{\partial Var_H}{\partial \sigma}$.

$$\begin{aligned}\tfrac{\partial Var_H}{\partial \sigma} &= 2P_0^2\Delta_a^2 H^2\tau\sigma e^{2r\tau+\sigma^2H^2\tau} + 2(1-2\Delta_a)P_0^2H^2\tau\sigma e^{2r\tau+\sigma^2H^2\tau}\mathbf{N}(2\sigma H\sqrt{\tau}+z_H) \\ &+2\left[K^2\mathbf{N}(z_H) - KP_0e^{r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H) - \Delta_a K^2 + \Delta_a KP_0e^{r\tau}\right]H\sqrt{\tau}\cdot \mathbf{n}(z_H)\end{aligned}$$

We also need to calculate $\frac{\partial (E_H - E_L)^2}{\partial \sigma}$.

$$\begin{aligned}\tfrac{\partial (E_H-E_L)^2}{\partial \sigma} &= [E_H\{C_T\} - E_L\{C_T\}] \times \tfrac{\partial (E_H\{C_T\} - E_L\{C_T\})}{\partial \sigma} \\ &= [E_H\{C_T\} - E_L\{C_T\}] \times K\left[\mathbf{n}(z_H)H\sqrt{\tau} - \mathbf{n}(z_L)L\sqrt{\tau}\right]\end{aligned}$$

For ATM call options, i.e. $K \doteqdot P_0e^{r\tau}$, we have

$$\mathbf{n}(z_H) = \mathbf{n}(z_L) = \mathbf{n}(z_A) \to \frac{1}{\sqrt{2\pi}} \quad and \quad \mathbf{N}(z_H) = \mathbf{N}(z_L) = \mathbf{N}(z_A) \to \frac{1}{2}, \quad when \quad \tau \to 0$$

The limiting behavior (28) and $\frac{\partial Var(W_a)}{\partial \sigma}$ is as follows:

$$\tfrac{\partial E(W_a)}{\partial \sigma} = \tfrac{K\sqrt{\tau}}{\sqrt{2\pi}}(A - \phi H - (1-\phi)L) + o(\sqrt{\tau})\tilde{\ }O(\sqrt{\tau}) \quad \text{and} \quad \tfrac{\partial Var(W_a)}{\partial \sigma} = O(\tau)$$

As $\tau \to 0$, (28) dominates $\frac{\partial Var(W_a)}{\partial \sigma}$. Therefore, $\frac{\partial f}{\partial \sigma} > 0$ as the maturity date approaches, which implies that $\frac{\partial A}{\partial \sigma} < 0$.

To confirm this result in market prices, we average the percentage implied volatility bid-ask spread using the different measures for each option contract and plot the average percentage spreads in Figure 6**.** The bold line denotes the realized volatility during the contract period, while the dashed line represents the averaged percentage volatility of CM-IMV, BSM-IMV ATM call and put options, respectively in the top, middle and bottom charts. As seen in the sample period from 01/02/2007 to 04/15/2010, when RV is near its maximum, the BSM-IMV spreads are at their lowest levels. The realized volatility are lower at the beginning and the end of this sample period, where it can be observed that the locally lowest percentage spreads are relatively higher. This documents a negative relationship between the percentage implied volatility bid-ask spread and the level of the realized volatility. This confirms the model's implication that MMs do not increase the implied volatility bid-ask spread proportional to an increase in the level of the volatility.

Next, we study the relation between the percentage implied volatility bid-ask spread and the stochastic volatility's variance. We define the variance of the volatility as

$$Var(v) = \phi(\sigma H)^2\tau + (1-\phi)(\sigma L)^2\tau - E(v)^2 \quad \text{where} \tag{29}$$

$E(v) = \phi\sigma H\sqrt{\tau} + (1-\phi)\sigma L\sqrt{\tau}$ and $v \in \{\sigma L, \sigma H\}$.

It is easy to show that changing the parameters $\{\sigma, L, H\}$, changes the variance of the volatility:

$$\begin{aligned}
\tfrac{\partial Var(v)}{\partial H} &= 2\phi(1-\phi)\sigma^2\tau(H-L) > 0, \\
\tfrac{\partial Var(v)}{\partial L} &= 2\phi(1-\phi)\sigma^2\tau(L-H) < 0, \\
\tfrac{\partial Var(v)}{\partial \sigma} &= 2\sigma\tau\phi(1-\phi)(L-H)^2 > 0.
\end{aligned}$$

As seen, the variance of the volatility increases as the largest volatility level increases or the lowest volatility level declines. Additionally, an increase in the volatility level also increases the variance of the volatility.[14] Given these insights, we can prove the following proposition.

Proposition 4 (The Variance of the Volatility Effect)

For small τ, an increase in the variance of the volatility due to increasing H and/or decreasing L results in a larger implied bid-ask volatility spread ($\frac{\partial(A-B)}{\partial H} > 0, \frac{\partial(A-B)}{\partial L} < 0$). If (H, L) move in the same direction and with the same magnitude, the volatility spread and volatility uncertainty remain unchanged ($\frac{\partial(A-B)}{\partial H} + \frac{\partial(A-B)}{\partial L} = 0$). An increase in the variance of the volatility due to an increase in the volatility level σ results in a lower percentage implied volatility bid-ask spread ($\frac{\partial(A-B)}{\partial \sigma} < 0$).

Proposition 4 states that a change in the implied volatility bid-ask spread is positively correlated with the variance of the stochastic volatility, after controlling for the volatility's level (due to ΔH and ΔL). However, it also implies that an increase (decrease) in the volatility's variance does not always imply an increase (decrease) in the implied volatility bid-ask spread due to a changing volatility level $\Delta\sigma$, which is the reason to control for the volatility's level in any test of proposition 4.

4 The Multiple-Period Model

In this section we extend the one-period model to a multiple-period model. We show the multiple-period models generate similar patterns in option ask/bid implied volatility spreads as obtained in the single period model.

4.1 Two-Periods

Here we let X_1 and X_2 denote the stock's log price return at time 1 and time 2, respectively where X_t follows a Bernoulli-type random normal distribution. That is, for each period, $X_t \tilde{} N(\mu_t, V_t)$ and with probability ϕ, $(\mu_t, V_t) = (\mu_H\Delta\tau, \sigma_H^2\Delta\tau)$ and with probability $(1-\phi)$, $(\mu_t, V_t) = (\mu_L\Delta\tau, \sigma_L^2\Delta\tau)$. $\Delta\tau$ is

[14] Here we do not discuss volatility uncertainty over ϕ, because it is similar to the case where H and L shift in opposite directions.

the length of each time period. Because our independence assumption implies $\mu_H \Delta\tau + \frac{\sigma_H^2 \Delta\tau}{2} - r\Delta\tau = 0$ and $\mu_L \Delta\tau + \frac{\sigma_L^2 \Delta\tau}{2} - r\Delta\tau = 0$, the amount of stock held in the hedged portfolio does not affect the portfolio's return but only its variance. Moreover, hedging doesn't affect the expected value of the option. Consequently, the option's equilibrium price at each time t can be calculated directly from the conditional risk neutral probabilities, allowing us to focus on volatility risk.

Given the independence assumption (4), mean-variance optimization reduces to the variance minimization problem

$$\underset{\Delta_0,\Delta_1}{Min}\ Var_0[W_{a,2}]$$

where $W_{a,2}$ denotes the value of portfolio containing a written call option which expires at period 2. The total variance of a dynamic hedge given Hamilton-Jacobi-Bellman (HJB) equation for our discrete-time problem is

$$Var_0[W_{a,2}] = E_0\left[Var_0\left(C_{1,2} - \Delta_0 P_1\right) + Var_1\left(C_{2,2} - \Delta_1 P_2\right)\right], \tag{30}$$

where $C_{1,2}$ is the "fair value" at time 1 of a option expiring at time 2 and $C_{2,2} = (P_2 - K)^+$. This equation shows that under minimum-variance criteria the total variance of hedging error is the sum of expected hedging error variance of each period. In general, solving a variance minimization problem involves the issue of time-inconsistency of dynamic hedges; i.e., the $t+1$ hedging strategy will change the evaluation measure at t. In simpler terms, the future hedging strategy changes the equilibrium option prices today. Basak and Chabakarui (2012) gives a sufficient condition, which is the same as our assumption (4), which eliminates this time-inconsistency problem. They also provide the general solutions for the hedge ratios Δ_t at each time t in a dynamic hedging setting,[15] which are

$$\Delta_t^* = \frac{Cov_t(C_{t+1}, P_{t+1})}{Var_t(P_{t+1})}.$$

As given $E_H(\frac{P_T}{P_0}) = E_L(\frac{P_T}{P_0}) = e^{r\Delta\tau}$, algebra yields[16]

$$\Delta_1^* = \frac{(1-\phi)Cov_{1,L}(C_{2,2}, P_2) + \phi Cov_{1,H}(C_{2,2}, P_2)}{P_1^2 e^{2r\Delta\tau}\left[\phi e^{\sigma_H^2} + (1-\phi)e^{\sigma_L^2} - 1\right]}$$

$$\Delta_0^* = \frac{(1-\phi)Cov_{0,L}(C_{1,2}, P_1) + \phi Cov_{0,H}(C_{1,2}, P_1)}{P_0^2 e^{2r\Delta\tau}\left[\phi e^{\sigma_H^2} + (1-\phi)e^{\sigma_L^2} - 1\right]}$$

[15] The general solution is derived by Basak and Chabakarui (2012). We provide the same derivation in Appendix C-1.

[16] Given the covariance decomposition,
$\Delta_1^* = \frac{(1-\phi)Cov_{1,L}(C_{2,2}, P_2) + \phi Cov_{1,H}\left(C_{2,2}, P_2\right) + Cov_1(E_1(C_{2,2}|V), E_1(P_2|V))}{\phi Var_{1,H}(P_2) + (1-\phi)Var_{1,L}(P_2)}$.
Since $E(P_2|V)$ is constant, the covariance can be expressed as $\frac{(1-\phi)Cov_{1,L}(C_{2,2}, P_2) + \phi Cov_{1,H}\left(C_{2,2}, P_2\right)}{\phi Var_{1,H}(P_2) + (1-\phi)Var_{1,L}(P_2)}$.

Here, at every t, the "equilibrium call" price at $t+1$ is needed to derive the optimal Δ_t^*. Given the independence assumption, the equilibrium options' prices are the expected fair value for $C_{1,2}$ and $C_{0,2}$.

$$\begin{aligned} &C_{1,2} = e^{-r\Delta\tau} E_1[(P_2-K)^+] \\ &= e^{-r\Delta\tau}\left\{\phi E_1\left[(P_2-K)^+ \,|X_1=x_1, V_2=\sigma_H\right] + (1-\phi)E_1\left[(P_2-K)^+ \,|X_1=x_1, V_2=\sigma_L\right]\right\} \\ &= \phi BSM_{x_1,H} + (1-\phi)BSM_{x_1,L}, \end{aligned}$$

where x_1,σ_L and σ_H are the realization of X_1 and V_2. $BSM_{x_1,H}$ and BSM_{x_1L} are

$$\begin{aligned} BSM_{x_1,H} &= P_1 \cdot \mathbf{N}\left(\frac{\ln(\frac{P_1}{K})+r_H^+}{\sigma_H\sqrt{\Delta\tau}}\right) - e^{-r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_1}{K})+r_H^-}{\sigma_H\sqrt{\Delta\tau}}\right) \\ BSM_{x_1,L} &= P_1 \cdot \mathbf{N}\left(\frac{\ln(\frac{P_1}{K})+r_L^+}{\sigma_L\sqrt{\Delta\tau}}\right) - e^{-r}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_1}{K})+r_L^-}{\sigma_L\sqrt{\Delta\tau}}\right). \end{aligned}$$

where $r_H^+ = r\Delta\tau + \frac{\sigma_H^2\Delta\tau}{2}$, $r_L^+ = r\Delta\tau + \frac{\sigma_L^2\Delta\tau}{2}$, $r_H^- = r\Delta\tau - \frac{\sigma_H^2\Delta\tau}{2}$and $r_L^- = r\Delta\tau - \frac{\sigma_L^2\Delta\tau}{2}$.

$C_{1,2}$ is similar formula to that obtained in the 1 period model. The fair value of $C_{0,2}$ is

$$C_{0,2} = e^{-r\Delta\tau}E_0\left[C_{1,2}\right] = e^{-2r\Delta\tau}E_0\left[(P_2-K)^+\right] = \phi^2 BSM_{H,H} + 2\phi(1-\phi)BSM_{H,L} + (1-\phi)^2 BSM_{H,L}$$

where

$$\begin{aligned} BSM_{H,H} &= P_0\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+r_H^+*2}{\sigma_H\sqrt{2\Delta\tau}}\right) - e^{-2r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+r_H^-*2}{\sigma_H\sqrt{2\Delta\tau}}\right) \\ BSM_{H,L} &= P_0\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+r_H^++r_L^+}{\sqrt{\sigma_H^2\Delta\tau+\sigma_L^2\Delta\tau}}\right) - e^{-2r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+r_H^-+r_L^-}{\sqrt{\sigma_H^2\Delta\tau+\sigma_L^2\Delta\tau}}\right) \\ BSM_{L,L} &= P_0\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+r_L^+*2}{\sigma_L\sqrt{2\Delta\tau}}\right) - e^{-2r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+r_L^-*2}{\sigma_L\sqrt{2\Delta\tau}}\right). \end{aligned}$$

The derivation is in Appendix C-2.

4.2 N-Periods

This subsection extends the 2-period model to N-periods. We first derive the general equalities for the fair call values and the conditional covariances at time t, allowing us to calculate the optimal hedging ratios and get numerical solutions for the total variance of the hedging error. The derivation is given in Appendix C-3.

Proposition 4 (The fair call options price for bernoulli type volatility)

$$\begin{aligned} &C_{t,T} = \sum_{i=0}^{n}\binom{n}{i}\phi^i(1-\phi)^{n-i}\left\{P_t\cdot\mathbf{N}(d_1) - e^{-nr\Delta\tau}K\cdot\mathbf{N}(d_2)\right\} \\ &d_1 = \frac{\ln(\frac{P_t}{K})+i\cdot r_H^+ + (n-i)\cdot r_L^+}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n-i)\sigma_L^2\Delta\tau}} \qquad , \\ &d_2 = \frac{\ln(\frac{P_t}{K})+i\cdot r_H^- + (n-i)\cdot r_L^-}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n-i)\sigma_L^2\Delta\tau}},\ n=\frac{T}{\Delta\tau} \end{aligned}$$

The covariance equation is[17]

$$
\begin{aligned}
&Cov_t(C_{t+1,T}\ ,P_{t+1}) \\
&= \phi \sum \binom{n^*}{i} \phi^i (1-\phi)^{n^*-i} \left\{ P_t^2 e^{2r_H^+} \mathbf{N}\left(d_{1,H}\right) - P_t e^{-(n^*-1)r\Delta\tau} K \cdot \mathbf{N}\left(d_{2,H}\right) \right\} \\
&\quad +(1-\phi) \sum \binom{n^*}{i} \phi^i (1-\phi)^{n^*-i} \left\{ P_t^2 e^{2r_L^+} \mathbf{N}\left(d_{1,L}\right) - P_t e^{-(n^*-1)r\Delta\tau} K \cdot \mathbf{N}\left(d_{2,L}\right) \right\} \\
&\quad - \textstyle\sum_{i=0}^{n^*+1} \binom{n^*+1}{i} \phi^i (1-\phi)^{n^*+1-i} \cdot \left\{ P_t^2 e^{2r\Delta\tau} \mathbf{N}\left(d_1^*\right) - P_t e^{-(n^*-1)r\Delta\tau} K \cdot \mathbf{N}\left(d_2^*\right) \right\}.
\end{aligned}
$$

and

$$
\begin{aligned}
d_{1,H}^* &= \frac{\ln(\frac{P_t}{K}) + i \cdot r_H^+ + (n^*-i) \cdot r_L^+ + \left[r_H^+ + \sigma_H^2 \Delta\tau \right]}{\sqrt{(i+1) \cdot \sigma_H^2 \Delta\tau + (n^*-i) \sigma_L^2 \Delta\tau}} \\
d_{2,H}^* &= \frac{\ln(\frac{P_t}{K}) + i \cdot r_H^- + (n^*-i) \cdot r_L^- + r_H^+}{\sqrt{(i+1) \cdot \sigma_H^2 \Delta\tau + (n^*-i) \sigma_L^2 \Delta\tau}} \\
d_{1,L}^* &= \frac{\ln(\frac{P_t}{K}) + i \cdot r_H^+ + (n^*-i) \cdot r_L^+ + \left[r_L^+ + \sigma_L^2 \Delta\tau \right]}{\sqrt{i \cdot \sigma_H^2 \Delta\tau + (n^*-i+1) \sigma_L^2 \Delta\tau}} \\
d_{2,L}^* &= \frac{\ln(\frac{P_t}{K}) + i \cdot r_H^- + (n^*-i) \cdot r_L^- + r_L^+}{\sqrt{i \cdot \sigma_H^2 \Delta\tau + (n^*-i+1) \sigma_L^2 \Delta\tau}} \\
d_1 &= \frac{\ln(\frac{P_t}{K}) + (i+1) \cdot r_H^+ + (n^*-i+1) \cdot r_L^+}{\sqrt{i \cdot \sigma_H^2 \Delta\tau + (n^*-i+1) \sigma_L^2 \Delta\tau}} \\
d_2 &= \frac{\ln(\frac{P_t}{K}) + (i+1) \cdot\cdot r_H^- + (n^*-i+1) \cdot r_L^-}{\sqrt{i \cdot \sigma_H^2 \Delta\tau + (n^*-i+1) \sigma_L^2 \Delta\tau}} \\
n^* &= \frac{T-(t+1)}{\Delta\tau} > 0,\ T > t+1.
\end{aligned}
$$

Next, we apply our propositions to derive the dynamic hedging plan, and we provide numerical results for the total variance $Var_t[W_{a,T}]$ at each time t. For the numerical results, we set $\sigma = 0.2$, $\sigma \cdot H = 0.22$, $\sigma \cdot L = 0.18$, $r = 0.04$, $\phi = 0.5$, $P_t = 1500$, $c = 0.4$ and fix 10 days to maturity.[18]

Figure 7 shows the variance of the hedging error for different number of options hedged $n = 1, 2, 3, \ldots, 10$. Our result at Chart (a) indicates that for 10-day tenor contracts $Var_0[W_{a,10}]$ drops from about 233.57 to 55.49 when the number of hedged options increases from 1 to 5, but the variance only reduces by about 25 when the number of hedged options increases from 5 to 10. Apparently the marginal benefit of hedging is decreasing when hedging frequency increases. Second, fixing the number of hedged options to 10, for different maturities, Chart (b) shows $Var_t[W_{a,10}]$ for different tenors. We see that as days to maturity declines, the variance steadily decreases. The variance is about 30 for 10-days tenor and drops by approximately 3 when DTM (days to maturity) decreases by 1.

To consider the case where traders hedge at a daily frequency, we let the ratio of DTM to the total number of hedged options equal 1 and we study the total variance of N-day tenors, hedged daily. As shown in Chart (c) of Figure 7, the hedging error variance increases as the days to maturity

[17]

In the one-period model, $n^* = 0$. Therefore, the covariance for the one-period model uses equation (9) with $Cov\left[Y, (Y-k)^+\right]$ as calculated in appendix B.

[18]Again, we use $\sigma_H = \sigma \cdot H$ and $\sigma_L = \sigma \cdot L$ in the simulations and the calibration.

increase. Finally, Chart (d) shows the call prices for different days to maturity. Compared to a flat pattern at Chart (c), the ATM call option price (denoted by solid dark line) decreases faster than the variance of the hedging error. The dashed line shows the ratio of the standard deviation of the hedging error to the option price. This ratio increases from 0.257 at 10 DTM to 0.751 at 1 DTM. In short, though both the hedging variance and the option's price decrease when time approaches the expiration date, the rates of their decrease are very different. And, it is this disproportional decay rate that leads to the previously discussed patterns of option bid-ask spreads in implied volatility measures.

4.3 The Implied Volatility Spread

The equilibrium condition remains the same as in the single period model. Let $C_{t,T}(a)$ denote the fair value of the ask call price and $Var_t^*(W_{a,T})$ the variance of hedging error delivered by the optimal daily hedging strategy. Without loss of generality, we let $Q = 1$ and $c = 0.4$ for equilibrium analyses (13). Hence, for each time t, the MMs quote at a price which satisfies the equations

$$C_{t,T}(a) - C_{t,T} - \frac{1}{\gamma}Var_t(W_{a,T}) - c = 0 \quad (31)$$

$$C_{t,T} - C_{t,T}(b) - \frac{1}{\gamma}Var_t(W_{b,T}) - c = 0 \text{ and } t = 1, 2, 3...T, \quad (32)$$

where the equilibrium ask price $C_{t,T} + \frac{1}{\gamma}Var_t(W_{a,T})$ and bid price $C_{t,T} - \frac{1}{\gamma}Var_t(W_{b,T})$ are inverted to obtain the implied volatilities $A\sigma$ and $B\sigma$, respectively.

In our numerical analyses, we use the same parameter values listed in the previous section and we assume that $\frac{1}{\gamma} = 0.01$ and 0.05. Figure 8 summarizes the results. Chart (e) shows the daily equilibrium IMV $A\sigma$ and $B\sigma$ for tenors of 10 days to 1 days. The dashed lines represent the IMV of bid and ask prices, respectively, when $\frac{1}{\gamma}$ is 0.01, and the solid lines are the IMV of ask and bid prices at $\frac{1}{\gamma} = 0.05$. Given $\sigma = 0.2$, the ask IMV is increasing, and the bid IMV is decreasing. Chart (f) shows the difference in the IMV, and we observe an increasing IMV spread with an increasing magnitude. Charts (g) and (h) use the multipliers A and B to present the same information. These results support propositions 1 and 2 in the one period model.

We next study the volatility level effect and variance of volatility effect. Setting $\frac{1}{\gamma} = 0.01$, we observe that the level effect happens a couple few days before expiration. In Figure 9, we show the price and variance of the final 10 hours in the option's life. Chart (i) shows the ATM call prices for volatility 0.2 and 0.3. The dashed line is the call price valued at $\sigma = 0.3$, and it is always higher than the solid line which denotes the call price at $\sigma = 0.2$. Chart (j) shows the variance of the hedging error at two different volatility levels. A shown, the higher volatility level the higher the variance.

However, in Chart (k), the solid lines, representing the bid and ask IMV multipliers at $\sigma = 0.2$, lie outside dashed lines which denote the multipliers at $\sigma = 0.3$. In other words, even though the hedging variance is higher at a 0.3 volatility, the multiplier of the ask (bid) IMV at a 0.2 volatility is larger (lower) than that at a 0.3 volatility. Shown in Chart (l), when DTM is very small the percentage spread is larger as the volatility level decreases. These results confirm the implications of proposition 3.

5 Data and the Empirical Evidence

We use the Options Metric's database focusing on SP index options. Our sample contains daily best closing quotes, and the time period is from January 2, 2001 to April 17, 2010. The contracts are European style options, and the nearest maturities expire every third Saturday with the settlement value determined by the opening selling price on the third Friday. The nearest maturity contracts have a strike interval of 5 points, and the minimum tick size for quotes is \$0.05 if the traded price is below \$3 dollars and \$0.1 if the traded price is above \$3.

Two implied volatility measures, the Black-Scholes-Merton implied volatility (BSM-IMV) and the Carr and Madan model-free implied volatility (CM-IMV), are used to calculate the volatility spreads. Table 1 provides descriptive statistics of BSM-IMV, CM-IMV, and bid-ask spreads in term of those two measures. As shown in Panel (1) of Table 1, CM-IMV has larger means than at-the-money BSM-IMV. The average CM-IMV for bid and ask prices are 21.29% and 23.81%, respectively, while the average BSM-IMV is 17.48% (20.81%) and 19.58% (23.27%) for ATM Call (Put) options. Since out-of-the money (OTM) options commonly trade at relatively larger implied volatilities, the higher CM-IMV mean may be the result of the CM-IMV calculation containing deep OTM options. We also find that the ATM put options have larger BSM IMV means than do call options. Moreover, the standard deviation, skewness and kurtosis are also larger for put options.

Panel (B) presents the statistics for the "spread" calculated using the 3 volatility measures. The implied volatility spreads are 1.99%, 2.25% and 2.44% for BSM-IMV Call, BSM-IMV Put, and CM-IMV. For the BSM-IMVs, the spreads for the put options do not have larger skewness and kurtosis than the spreads for call options, although the mean and standard deviation of put spreads are still larger. Because the CM-IMV calculation includes OTM options, the mean of the CM-IMV spread is relatively larger than the BSM-IMV spreads of ATM options, but the skewness and kurtosis are lower. Interestingly, if we compare the percentage spreads of BSM-IMV and CM-IMV, it is surprising to find the kurtosis of percentage volatility spreads is much lower than that of the volatility spreads. In the bottom three rows of Panel (B), the kurtosis for the percentage spread of implied volatility for BSM-IMV call and put options are 8.46 and 9.98, and 11.99 for the kurtosis of CM-IMV spread, while they are 65.5, 64.33 and 51.47 for volatility spreads. Because the kurtosis of the BSM-IMV spread and CM-IMV spread are much larger, ordinary econometric estimation will work better for percentage implied volatility spreads.

5.1 Calibration

In this subsection, we confirm that our theoretical model fits well. We use the call contracts expiring on November, 22, 2014 to calibrate the model.[19] For each of the final 25 trading days, dollar spreads of call options having a delta between 0.4 to 0.6 are collected. To calibrate the 5 parameters $(r, A\sigma, B\sigma, \phi, \gamma)$ we compute

$$Min \sum_k \left[\frac{2}{\gamma} Var_t(W_{k,T}) + 0.05 - \ Spread_{t,k} \right]^2 ,$$

where k is the strike and 0.05 is the transaction cost (also minimum tick).[20]

Assuming that the MMs delta hedge once per day, the calibration results are contained in Figure 10. In Panel (m), (o) and (p), we provide the real spreads and the estimated theoretical spreads for options which have strikes nearest to the spot price. In Panel (m), the theoretical spread fits the market dollar spread well and both lines moves together. As the index keeps increasing during the final 25 trading days, Panel (n) shows that the traded volatility is decreasing and remains at a relatively low level (about 15%). Panel (o) transforms the dollar spread into a percentage volatility spread. We see the market and theoretical multiplier spreads are increasing as the expiration date approaches. Finally, we compare the theoretical and market ask and bid prices in Panel (p). This panel shows that our theoretical (bid and ask) prices are relatively larger than market prices. Again, this may be the result of using OTM call options in our calibration.

5.2 The Econometric Model

To test Propositions 1 - 4, we apply panel regression to capture the covariance relationship between consecutive contracts. Our first estimation model is

$$SpPct_{c,t} = \alpha + \theta \cdot \frac{1}{TM_{t,c}} + \beta_1 \cdot RV_c + \beta_2 \cdot VRP_{c,t} + \beta_3 \cdot Jump\_intensity_{c,t} + v_c + \varepsilon_{c,t}, \quad t = 1, 2, 3, 4.. \quad (33)$$

where the dependent variable, $SpPct_{t,c}$, is the percentage volatility bid-ask spread multiplied

[19]Our empirical work covers the time period January 2, 2001 to April 17, 2010, where we used OTM calls and puts for the CM IMV calculations and 4 ATM strikes to derive the BSM IMV.

[20]For each day, we use the initial values $(r, \sigma \cdot L, \sigma \cdot H, \phi, \gamma) = (0.04, VIX_{t-1} + 0.1, VIX_{t-1} - 0.1, 0.05, 0.02)$ in the calibration.

by 100. The subscript c indicates the contract of a given expiration date, and t denotes the date. We use the realized volatility RV during the contract period as the denominator when calculating $SpPct$. By doing so, the realized volatility RV reflects the volatility level effect, and any change of $SpPct$'s risk premium is due to a change in the numerator, the volatility spread. TM refers to the annualized time to maturity, and the variance risk premium VRP is used as a proxy for the volatility's variance. VRP is defined as

$$VRP_{c,t} = IMV_{c,t} - RV_{c,t}, \tag{34}$$

see Carr and Wu (2009). The $RV_{c,t}$ in equation (34) is different from the RV_c in equation (33). RV_c is per contract over the sample period, while $RV_{c,t}$ is the past annualized 30-day realized volatility, computed every day. $IMV_{c,t}$ is implied volatility.

In analogy to a stock's bid-ask spread which reflects the stock's risk to a MM, the volatility spread measures the options' risk to a MM. It is well understood that jumps in the underlying stock hinders a trader's ability to hedge and results in an incomplete market. Jarrow, Lando and Yu (2005) show that if a stock's jump risk is diversifiable, then a unique option price still results. Therefore, it is of interest to study the relationship between the percentage spread and the index's jump risk. In this regard, we add the independent variable

$$Jump\_intensity_t = \sum_{i=1}^{M} I(r_{t,i} > k),$$

where I is a indication function, which equals 1 if the 5-minute return r_i is greater than a threshold value k and 0 otherwise. M is the number of 5-minute intervals within a day. We let k equal the smallest tenths value which lies above the largest 4% of the total 5-minute returns.

Propositions 1 and 2 correspond to the maturity effect, i.e. the bid-ask percentage volatility spread should increase at an increasing rate (convex shape) as the option's maturity approaches. To confirm the time to maturity effect, the coefficient θ should be positive in equation (33). Proposition 3 concerns the volatility level effect, i.e. the percentage volatility spread should decrease as the volatility level increases. Since the realized volatility RV_c is included in equation (33) to capture this effect, we expect the coefficient β_1 to be negative. Finally, Proposition 4 states that after controlling for the volatility level, a change in the volatility's variance (uncertainty) should be positively correlated with changes in the percentage volatility spread; hence in equation (33), β_2 or β_3 should be positive.

In addition, because traders in the options market often know particular news arrival times, e.g. the date of a earnings announcement, trading information may reflect market information. To investigate this possibility, we study whether the percentage spread can predict the future jump intensity by estimating the following regression

$$\begin{aligned} Jump\_intensity_{c,t+1} = \alpha + \beta_1 \cdot DeTrend\_SpPct_{c,t} + \beta_2 \cdot SpPct_{c,t} + \beta_3 \cdot VRP_{c,t} \\ +\beta_4 \cdot RV_{c,t} + \beta_5 \cdot Jump\_intensity_{c,t} + \varepsilon_{c,t} \ , \end{aligned} \tag{35}$$

where $DeTrend\_SpPct_{c,t}$ is de-trended percentage volatility spread.[21] We also consider the original percentage spread $SpPct_{c,t}$ to see if the de-trended volatility spread performs better than original spread in explaining future jump intensities.

5.3 Empirical Results

Table 2 presents the estimation results of expression (33). Panels (A), (B) and (C) are the results for the volatility measures CM-IMV, BSM-IMV (ATM) call and BSM-IMV (ATM) put options, respectively. Because VRP and $Jump\_intensity$ are both significant, we find that RV becomes significant when the $Jump\_intensity$ is added as a control variable. Because the realized volatility and the jump intensity are highly correlated but have opposite effects, we focus on the regression models which include the jump intensity.

As seen, the coefficient θ is significantly positive; hence the maturity effect exists. And, if we only control for the jump intensity, a 1% increase in the level of the realized volatility decreases $SpPct$ by 0.139%, 0.178% and 0.247% for CM-IMV, BS-IMV call and BS-IMV-put; and by 0.068%, 0.121%, and 0.194% if both the volatility risk premium and the jump intensity are included as regressors. An increase in the jump intensity enlarges the spread by 0.282%, 0.244% and 0.247% for 3 different implied volatility measures before including the volatility risk premium. They enlarge the spread by 0.178%, 0.161% and 0.17% after VRP is included in model. Finally a 1% increase in VRP results in an increase in the percentage volatility spread by 0.244%, 0.197% and 0.184% for the different volatility measures. The R^2 in Panel (A) reaches 0.43, while the largest R^2 are 0.54 and 0.67 in Panels (B) and (C).

Having confirmed our theoretical propositions, we now show the importance of de-trending the maturity effect. First, we examine the jump's contemporaneous effect on the percentage spread and include as regressors the daily realized volatility $RV_{c,t}$, the realized volatility level RV_c, the volatility risk premium $VRP_{c,t}$, and the jump intensity $Jump\_intensity_{c,t}$. Then, we run the panel regression separately on $DeTrend\_SpPct_{c,t}$ and $SpPct_{c,t}$. As shown in Table 3, all independent

[21] We first run the regression $SpPct_{c,t} = \alpha + \theta \cdot \frac{1}{TM_{t,c}}$. After deriving the estimated $\widehat{\theta}$, we subtract $\widehat{\theta} \cdot \frac{1}{TM_{t,c}}$ from $SpPct_{c,t}$ to derive $Detrend\_SpPct_{c,t}$.

variables have daily contemporaneous effect on the spreads, and, if the spread is de-trended, then the jump intensity becomes more significant in explaining the variation of $DeTrend\_SpPct_{c,t}$.

Finally, Table 4 shows that both $DeTrend\_SpPct_{c,t}$ and $SpPct_{c,t}$ predict the future jump intensity in accordance with expression (35). For all the volatility measures, the de-trended spreads always perform better. If the regressor contains only either $DeTrend\_SpPct_{c,t}$ or $SpPct_{c,t}$, both explain the next day's jump intensity. However, if we use the detrended and non-detrended spreads together, only $DeTrend\_SpPct_{c,t}$ is significant in explaining the future jump intensity. Finally, we add $VRP_{c,t}$, $RV_{c,t}$, and $VRP_{c,t}$ as control variables, and the results show that $DeTrend\_SpPct_{c,t}$ is significant, but $SpPct_{c,t}$ is not.

6 Conclusion

This paper empirically documents the fact that the percentage implied volatility bid-ask spread increases at an increasing rate as an option's maturity date approaches. This maturity effect is confirmed using quotes from options on the CBOE S&P index over the time period 2001/01/02 - 2010/04/17. This effect is validated using model-free implied volatilities as well as Black-Schole-Merton implied volatilities.

We construct an equilibrium model in an incomplete market with volatility risk to explain this phenomena. The equilibrium model has risk averse and competitive option MMs quoting bid and ask prices to minimize their inventory risk. Two additional testable implications of the model are generated. The first is that an increase in the level of the underlying asset's volatility decreases the percentage bid-ask implied volatility spread. The second is that, holding the level of the volatility constant, an increase in the volatility's variance increases the percentage bid-ask implied volatility spread. These implications are also empirically tested herein. The empirical evidence generally confirms these implications and the model's validity.

References

[1] Andersen, T., T. Bollerslev, F. X. Diebold and H. Ebens. (2001). *"The Distribution of Realized Stock Return Volatility"*, Journal of Financial Economics. 61(1): 43-76

[2] Bakshi, G., C. Cao and Z. Chen. (1997). *"Empirical Performance of Alternative Option Pricing Models"*, Journal of Finance, 52, 2003--2049.

[3] Bakshi, G., and N. Kapadia. (2003). *"Delta-hedged Gains and the Negative Volatility Market Risk Premium"*, Review of Financial Studies, 16, 527--66.

[4] Basak, S., and G. Chabakauri. (2010). *"Dynamic Mean-variance Asset Allocation"*, Review of Financial Studies, 23(8), 2970--3016.

[5] Basak, S. and G. Chabakarui. (2012). *"Dynamic Hedging in Incomplete Markets: A Simple Solution"*, Review of Financial Studies, 25(6), 1845-1896.

[6] Black, F. and M. Scholes. (1972). *"The pricing of options and corporate liabilities"*, Journal of Political Economy, 81, 637-659.

[7] Bollerslev, Tim. (1986). *"Generalized Autoregressive Conditional Heteroskedasticity"*, Journal of Econometrics, 31:307-327

[8] Broadie, M., M. Chernov and M. Johannes. (2009). *"Understanding Index Option Returns"*, Review of Financial Studies, 22(11), 4493-4529.

[9] Brooks, C., O. T. Henry and G. Persand. (2002). *"The Effect of Asymmetries on Optimal Hedge Ratios"*, Journal of Business 75, 333--352.

[10] Cao, M. and J. Wei. (2010). *"Option market liquidity: Commonality and other characteristics"*, Journal of Financial Market, 13, 20-48

[11] Carr, P. and D. Madan. (2001). *"Towards a Theory of Volatility Trading"*, Handbooks in Mathematical Finance: Option Pricing, Interest Rates and Risk Management, eds. J. Cvitanic, E. Jouini and M. Musiela, Cambridge University Press, 458-476

[12] Carr, P. and L. Wu. (2009). *"Variance Risk Premiums"*, Review of Financial Studies, 22(3), 1311-1341.

[13] CBOE Volatility Index White Paper.

[14] Chakravarty, S., H Gulen and S Mayhew. (2004). *"Informed trading in stock and optioton markets"*, Journal of Finance, 1235-1257

[15] Chen, H., S. Joslin, and S. Ni. (2016). *"Demand for Crash Insurance, Intermediary Constraints, and Risk Premia in Financial Markets"*, working paper.

[16] Chong, B.S., D.K. Ding and K.H. Tan. (2003). *"Maturity Effect on bid-ask Spreads of OTC Currency Options"*, Review of Quantitative Finance and Accounting, 21, 5-15.

[17] Coval, J., and T. Shumway. (2001). *"Expected Option Returns"*, Journal of Finance 56, 983--1009.

[18] Dufour, JM, R. Garcia and A. Taamouti. (2010). *"Measuring High-Frequency Causality between Returns, Realized Volatility and Implied Volatility"*, Discussion Paper, McGill University (Department of Economics), CIREQ and CIRANO.

[19] Easley, D., M. O'Hara and P. Srinivas. (1998). *"Option volume and stock prices: Evidence on where informed traders trade"*, Journal of Finance, vol. 53, no. 2, pp. 431-465.

[20] Engle, R.F. (1982). *"Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation"*, Econometrica, vol. 50, No. 4, pp. 987-1007.

[21] Engle, R.F. (1995). ARCH Models. Oxford University Press, Oxford.

[22] Garleanu, N., L. H. Pedersen and A. M. Poteshman. (2009). *"Demand-Based Option Pricing"*, Review of Financial Studies, 22(10), 4259-4299.

[23] Hasbrouck, J. (1995). *"One security, many markets: Determining the contributions price discovery"*, Journal of Finance, vol. 50, no. 4, pp. 1175-1199

[24] Holowczak, R., S. Yusif and L. Wu. (2007). *"Price Discover in the U.S. Stock and Stock Options Markets: A Portfolio Approach"*, Review of Derivatives Research 9, Issue 1, pp. 37-65

[25] Hsieh, W.L.G., C.S. Lee and S.F. Yuan. (2008). *"Price discovery in the options markets: An application of put-call parity"*, Journal of Futures Markets, vol. 28, no. 4, pp. 354-375.

[26] In, F., and S. Kim. (2006). *"The Hedge Ratio and the Empirical Relationship Between the Stock and Futures Markets"*, Journal of Business 79, 799-820.

[27] Jarrow, R. and A. Chatterjea (2013*). An Introduction to Derivative Securities, Financial Markets, and Risk Management*, W.W. Norton and Co., N.Y.

[28] Jarrow, R., D. Lando and F. Yu (2005). Default Risk and Diversification: Theory and Empirical Implications. Mathematical Finance, Vol. 15, No. 1, 1-26.

[29] Jiang, G.J. and Y. S. Tian. (2005). *"The Model-Free Implied Volatility and Its Information Content"*, Review of Financial Studies 18(4), 1305-1342.

[30] Lioui, A. and P. Poncet. (2000). *"The Minimum Variance Hedge Ratio Under Stochastic Interest Rates"*, Management Science 46, 658--668.

[31] Muravyev, D, N.D. Pearson and J.P. Broussard. (2012). *"Is There Price Discovery in Equity Options?"*, Journal of Financial Economics.

[32] Newey, Whitney K and Kenneth D West. (1987). *"A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix"*, Econometrica 55 (3): 703--708

[33] Pan, J. and A.M. Poteshman. (2006). *"The information in option volume for future stock prices"*, Review of Financial Studies, 19(3), 871-911.

[34] Robert C. Merton. (1973). *"Theory of rational option pricing"*, Bell Journal of Economics and Management Science, 4, 141-183.

[35] Viswanathan, S. and J.D. Wang. (2002). *"Market Architecture: Limit-Order Books Versus Dealership Markets"*, Journal of Financial Markets, vol. 5, 126-167.

[36] Wei, J. and J.G. Zheng. (2010). *"Trading activity and bid--ask spreads of individual equity options"*, Journal of Banking & Finance, 34, 2897-2916

Appendix A

The Carr Madan Implied Volatility (CM IMV)

The following is the CBOE formula for the VIX index, denoted as CM IMV.

$\sigma_{CM}^2 = \frac{2}{\tau}\sum_i \frac{\Delta K_i}{K_i^2} e^{r\tau} Q(\tau, K_i) - \frac{1}{\tau}[\frac{F}{K_0} - 1]^2$

$CM\ IMV = \sigma_{CM} * 100$

τ	: Time to expiration
F	: Forward index level derived from put-call parity
K_0	: First strike price below the forward index level, F
K_i	: Strike price of the i^{th} out-of-the-money option; a call if $K_i > K_0$, and a put if $K_i < K_0$; both put and call if $K_i = K_0$
ΔK_i	$= \frac{K_{i+1} - K_{i-1}}{2}$
r	: Risk-free spot rate of interest
$Q(\tau, K_i)$	: The ask price/bid price for the option with strike K_i

In the VIX index calculation, $Q(K_i)$ is the midpoint of the bid-ask spread for each option with strike price K_i. The forward index level is:

$$F = strike\ price + e^{r\tau}(\ Call\ price - Put\ price\ )$$

where the strike price selected is that for which the absolute difference between the call and put prices is the smallest.

In our paper, we use the strike price that is closest to the spot price to calculate the forward index level, sometimes called the effective forward price. The original formula proposed by Carr and Madan doesn't include the term involving $[\frac{F}{K_0} - 1]$.

Appendix B

Appendix B-1 (Expected Profit and Variance)

Portfolio Expected Profit Given σ_Y:

Assume $y = \ln Y \sim N(\mu_Y, \sigma_Y^2)$ and $Z \sim N(0,1)$, then

$$X = \sigma_Y Z + \mu_Y, Y = exp\,(\sigma_Y Z + \mu_Y)$$

The calculation of $E\,[Y \cdot I(Y \geq y)]$ and $Var\,[Y \cdot I(Y \geq y)]$ is detailed as follows:

$$\begin{aligned}
E\,[Y \cdot I(Y \geq y)] =& E\,[\exp(\sigma_Y Z + \mu_Y) \cdot I(\exp(\sigma_Y Z + \mu_Y) \geq y)] \\
=& E\left[\exp(\sigma_Y Z + \mu_Y) \cdot I(Z \geq \frac{\ln(y) - \mu_Y}{\sigma_Y})\right] \\
=& \int_z^\infty \exp(\sigma_Y u + \mu_Y) \cdot \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}) du; Let\ z = \frac{\ln(y) - \mu_Y}{\sigma_Y} \\
=& \exp(\frac{\sigma_Y^2}{2} + \mu_Y) \int_z^\infty \frac{1}{\sqrt{2\pi}} \exp(-\frac{(u - \sigma_Y)^2}{2}) du \\
=& \exp(\frac{\sigma_Y^2}{2} + \mu_Y) \cdot \mathbf{N}(\sigma_Y - z)
\end{aligned}$$

$$
\begin{aligned}
E\left[Y^2 \cdot I(Y \geq y)\right] =& E\left[\exp(2\sigma_Y Z + 2\mu_Y) \cdot I(\exp(\sigma_Y Z + \mu_Y) \geq y)\right] \\
=& \int_z^{\infty} \exp(2\sigma_Y u + 2\mu_Y) \cdot \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}) du; Let\ z = \frac{\ln(y) - \mu_Y}{\sigma_Y} \\
=& \exp(2\sigma_Y^2 + 2\mu_Y) \int_z^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{(u-2\sigma_Y)^2}{2}) du \\
=& \exp(2\sigma_Y^2 + 2\mu_Y) \cdot \mathbf{N}(2\sigma_Y - z)
\end{aligned}
$$

$$
\begin{aligned}
Var\left[Y \cdot I(Y \geq y)\right] =& E\left[Y^2 \cdot I(Y \geq y)\right] - \{E\left[Y \cdot I(Y \geq y)\right]\}^2 \\
=& \exp(\sigma_Y^2 + 2\mu_Y) \left\{\exp(\sigma_Y^2) \cdot \mathbf{N}(2\sigma_Y - z) - [\mathbf{N}(\sigma_Y - z)]^2\right\}
\end{aligned}
$$

Let $y \to 0$, that is $z \to -\infty$, we have

$$
\begin{aligned}
E\left[Y\right] &= \exp(\frac{\sigma_Y^2}{2} + \mu_Y) \\
Var\left[Y\right] &= \exp(\sigma_Y^2 + 2\mu_Y) \left[\exp(\sigma_Y^2) - 1\right]
\end{aligned}
$$

Let $C_\tau = (P_0 Y - P_0 k)^+$. The expectation and variance of the portfolio value can be explicitly calculated below. let $z = \frac{\ln(k) - \mu_Y}{\sigma_Y}$, Δ is the holded stocks for hedging, $P_0 k$ is the strike price, and $B = p_c - \Delta_a P_0$ is the amount of money market account at very beginning.

$$
\begin{aligned}
E\left[W_a\right] =& E[\Delta_a \cdot P_\tau - C_\tau + Be^{r\tau}] \\
=& E[\Delta_a \cdot P_0 Y - (P_0 Y - P_0 k)^+ + Be^{r\tau}] \\
=& \Delta_a \cdot P_0 \cdot \exp(\frac{\sigma_Y^2}{2} + \mu_Y) - P_0 \left[\exp(\frac{\sigma_Y^2}{2} + \mu_Y) \cdot \mathbf{N}(\sigma_Y - z) - k \cdot \mathbf{N}(-z)\right] + Be^{r\tau}.
\end{aligned}
$$

If we express the mean and volatility in annualized term, we can show the formula for C_τ is the same as Black-Scholes-Merton pricing model. Let $\mu_Y = \mu\tau$, $\sigma_Y^2 = \sigma^2\tau$ and $K = P_0 k$. Given the expectation operator is the risk neutral probability, we have $e^{(\frac{\sigma^2\tau}{2} + \mu\tau)} = e^{r\tau}$. Then the following equation for C_τ is exactly the same as Black-Scholes-Merton Formula after discounted by risk-free rate.

$$
\begin{aligned}
& P_0 \left[\exp(\frac{\sigma_Y^2}{2} + \mu_Y) \cdot \mathbf{N}(\sigma_Y - z) - k \cdot \mathbf{N}(-z)\right] \\
&= P_0 e^{r\tau} \cdot \mathbf{N}\left(\frac{\sigma_Y^2 - \ln(k) + \mu_Y}{\sigma_Y}\right) - P_0 k \cdot \mathbf{N}(\frac{-\ln(k) + \mu_Y}{\sigma_Y}) \\
&= P_0 e^{r\tau} \cdot \mathbf{N}\left(\frac{\sigma_Y^2/2 - \ln(k) + \ln(e^{\mu_Y + \sigma_Y^2/2})}{\sigma_Y}\right) - P_0 k \cdot \mathbf{N}(\frac{-\sigma_Y^2/2 - \ln(k) + \ln(e^{\mu_Y + \sigma_Y^2/2})}{\sigma_Y}) \\
&= P_0 e^{r\tau} \cdot \mathbf{N}\left(\frac{\ln(\frac{P_0 e^{\mu_Y + \sigma_Y^2/2}}{P_0 k}) + \sigma_Y^2/2}{\sigma_Y}\right) - K \cdot \mathbf{N}\left(\frac{\ln(\frac{P_0 e^{\mu_Y + \sigma_Y^2/2}}{P_0 k}) - \sigma_Y^2/2}{\sigma_Y}\right) \\
&= P_0 e^{r\tau} \cdot \mathbf{N}\left(\frac{\ln(\frac{P_0}{K}) + r\tau + \sigma_Y^2\tau/2}{\sigma_Y\sqrt{\tau}}\right) - K \cdot \mathbf{N}(\frac{\ln(\frac{P_0}{K}) + r\tau - \sigma_Y^2\tau/2}{\sigma_Y\sqrt{\tau}})
\end{aligned}
$$

The variance of the portfolio is

$$
\begin{aligned}
Var\left[W_a\right] =& Var[\Delta_a \cdot P_\tau - C_\tau + B_a \cdot e^{r\tau}] \text{ and } B_a = C_0^a - \Delta_a P_0 \\
=& Var[\Delta_a \cdot P_0 Y] + Var\left[(P_0Y - P_0k)^+\right] - 2\Delta_a \times cov(P_0Y, (P_0Y - P_0k)^+) \\
=& \left[\Delta_a \cdot P_0\right]^2 Var[Y] + P_0^2 Var\left[(Y-k)^+\right] - 2\Delta_a P_0^2 \times cov(Y, (Y-k)^+)
\end{aligned}
$$

$$
\begin{aligned}
Var\left[W_b\right] =& Var[\Delta_b \cdot P_\tau + C_\tau + B_b \cdot e^{r\tau}] \text{ and } B_b = -C_0^b - \Delta_b P_0 \\
=& Var[\Delta_b \cdot P_0 Y] + Var\left[(P_0Y - P_0k)^+\right] + 2\Delta_b \times cov(P_0Y, (P_0Y - P_0k)^+) \\
=& \left[\Delta_b \cdot P_0\right]^2 Var[Y] + P_0^2 Var\left[(Y-k)^+\right] + 2\Delta_b P_0^2 \times cov(Y, (Y-k)^+)
\end{aligned}
$$

We need to calculate $Var\left[(Y-k)^+\right]$ and $cov(Y, (Y-k)^+)$.

$$
\begin{aligned}
Var\left[(Y-K)^+\right] =& Var\left[(Y-k) \cdot I(Y \geqslant k)\right] \\
=& Var\left[Y \cdot I(Y \geqslant k) - k \cdot I(Y \geqslant k)\right] \\
=& Var\left[Y \cdot I(Y \geqslant k)\right] + k^2 Var\left[I(Y \geqslant k)\right] - 2k \cdot cov\left[Y \cdot I(Y \geqslant k), I(Y \geqslant k)\right] \\
=& \exp(\sigma_Y^2 + 2\mu_Y)\left[\exp(\sigma_Y^2)\mathbf{N}(2\sigma_Y - z) - \left[\mathbf{N}(\sigma_Y - z)\right]^2\right] \\
&+ k^2 \mathbf{N}(z)\mathbf{N}(-z) - 2k \cdot \exp(\frac{\sigma_Y^2}{2} + \mu_Y)\mathbf{N}(\sigma_Y - z)\mathbf{N}(z)
\end{aligned}
$$

$$
\begin{aligned}
& Cov\left[Y, (Y-k)^+\right] \\
&= Cov\left[Y, (Y-k) \cdot I(Y \geqslant k)\right] \\
&= E\left[Y(Y-k) \cdot I(Y \geqslant k)\right] - \exp\left(\tfrac{\sigma_Y^2}{2} + \mu_Y\right)\left[\exp\left(\tfrac{\sigma_Y^2}{2} + \mu_Y\right)\mathbf{N}(\sigma_Y - z) - k \cdot \mathbf{N}(-z)\right] \\
&= E\left[Y^2 \cdot I(Y \geqslant k)\right] - k \cdot E[Y \cdot I(Y \geqslant k)] - \exp\left(\tfrac{\sigma_Y^2}{2} + \mu_Y\right) \cdot \left[\exp\left(\tfrac{\sigma_Y^2}{2} + \mu_Y\right)\mathbf{N}(\sigma_Y - z) - k \cdot \mathbf{N}(-z)\right] \\
&= \exp\left(2\sigma_Y^2 + 2\mu_Y\right)\mathbf{N}(2\sigma_Y - z) - \left(k + \exp\left(\tfrac{\sigma_Y^2}{2} + \mu_Y\right)\right) \cdot \exp\left(\tfrac{\sigma_Y^2}{2} + \mu_Y\right)\mathbf{N}(\sigma_Y - z) \\
&+ k \cdot \exp\left(\tfrac{\sigma_Y^2}{2} + \mu_Y\right)\mathbf{N}(-z)
\end{aligned}
$$

Plugging $Var\left[(Y-k)^+\right]$ and $cov(Y, (Y-k)^+)$ into $Var\left[W_a\right]$, and rearranging the terms, we obtain

$$
Var\left[W_a\right] = Var\left[W_b\right] = P_0^2\left[C + A_1 \cdot \mathbf{N}(2\sigma_Y - z) + A_2 \cdot \mathbf{N}(\sigma_Y - z) + A_3 \cdot \mathbf{N}(-z)\right]
$$

where

$$
\begin{aligned}
C &= \Delta_a^2 \exp\left(\sigma_Y^2 + 2\mu_Y\right)\left[\exp\left(\sigma_Y^2\right) - 1\right] \\
A_1 &= \exp\left(2\sigma_Y^2 + 2\mu_Y\right)(1 - 2\Delta_a) \\
A_2 &= -\exp(\tfrac{\sigma_Y^2}{2} + \mu_Y)\left[\exp(\tfrac{\sigma_Y^2}{2} + \mu_Y)\mathbf{N}(\sigma_Y - z) + 2k \cdot \mathbf{N}(z) - 2\Delta_a\left(k + \exp(\tfrac{\sigma_Y^2}{2} + \mu_Y)\right)\right] \\
A_3 &= k\left[k \cdot \mathbf{N}(z) - 2\Delta_a \exp(\tfrac{\sigma_Y^2}{2} + \mu_Y)\right]
\end{aligned}
$$

Portfolio Mean and Variance given a Bernoulli type Volatility.

- We assume a Bernoulli type random volatility, which is a probability with independence assumption (4) for different annual volatility levels, σ_L and σ_H. In our math derivation, we let $\sigma_L=\sigma \cdot L$ and $\sigma_H=\sigma \cdot H$. We also use (L, H) in subscripts to denote the volatilities (σ_L, σ_H).

- Traders assign the subjective probability $(1-\phi)$ and ϕ to volatility σL and σH, respectively.

Expected Profit given a Bernoulli type Volatility

For a "written" and a "long" call option position, after traders apply delta hedging, the final wealth of W_a and W_b are

$$W_a = \Delta_a \cdot P_T - C_T + (C_a^0 - \Delta_a \cdot P_0) \cdot e^{r\tau}$$

$$W_b = \Delta_b \cdot P_T + C_T + (-C_b^0 - \Delta_b \cdot P_0) \cdot e^{r\tau}$$

where C_a^0 and C_b^0 are selling price and buying price respectively, C_T is final option payoff ($C_T = Max[P_T - K, 0]$), and Δ are the hedging positions. $C_T \geqslant 0$ and is cash-outflow for a written call, while $C_a^0 \geqslant 0$ and is cash inflow for a written call. Conversely, C_T and C_b^0 are cash-inflow and cash-outflow respectively for a long position. A positive (negative) Δ means a long (short) position of the stock; therefore, if Δ_a is positive , $\Delta_a \cdot P_T$ is the amount of cash-inflow from selling stocks at P_T and$-\Delta_a \cdot P_0$ is the cash-outflow to buy stocks at P_0. Given assumption (4),

$$E(P_T) = P_0 E(\frac{P_T}{P_0}) = P_0 \left\{(1-\phi)E_L(\frac{P_T}{P_0}) + \phi E_H(\frac{P_T}{P_0})\right\} = P_0 \left\{(1-\phi)e^{r\tau} + \phi e^{r\tau}\right\} = P_0 e^{r\tau}.$$

And the expected final wealth for selling a call option and purchasing a call options are,

$$\begin{aligned} E(W_a) =& \phi E_H \left\{C_a^0 e^{r\tau} - C_T + \Delta_a[P_T - P_0 e^{r\tau}]\right\} + (1-\phi)E_L \left\{C_a^0 e^{r\tau} - C_T + \Delta_a[P_T - P_0 e^{r\tau}]\right\} \\ =& C_a^0 e^{r\tau} - \phi E_H \left\{C_T\right\} - (1-\phi)E_L \left\{C_T\right\} = C_a^0 e^{r\tau} - E\left\{C_T\right\} \end{aligned}$$

$$\begin{aligned} E(W_b) =& \phi E_H \left\{-C_b^0 e^{r\tau} + C_T + \Delta_b[P_T - P_0 e^{r\tau}]\right\} + (1-\phi)E_L \left\{-C_b^0 e^{r\tau} + C_T + \Delta_b[P_T - P_0 e^{r\tau}]\right\} \\ =& E\left\{C_T\right\} - C_b^0 e^{r\tau} \end{aligned}$$

Using Appendix B-1 formula, we can derive the option price formula which is the same as Black-Schole-Merton Model. The expected prices under different volatility levels are,

$$\begin{aligned} E_H\left[C_\tau\right] =& e^{r\tau}\left[P_0 \cdot \mathbf{N}(z_H + \sigma H\sqrt{\tau}) - Ke^{-r\tau} \cdot \mathbf{N}(z_H)\right] \quad and \quad z_H = \frac{\ln \frac{P_0}{K} + (r - \frac{1}{2}\sigma^2 H^2)\tau}{\sigma H\sqrt{\tau}} \\ E_L\left[C_\tau\right] =& e^{r\tau}\left[P_0 \cdot \mathbf{N}(z_L + \sigma L\sqrt{\tau}) - Ke^{-r\tau} \cdot \mathbf{N}(z_L)\right] \quad and \quad z_L = \frac{\ln \frac{P_0}{K} + (r - \frac{1}{2}\sigma^2 L^2)\tau}{\sigma L\sqrt{\tau}} \end{aligned}$$

where $z_H = \frac{\ln(K) - E_H(\frac{P_T}{P_0})}{\sigma_J}$.Therefore, the expected final wealth is

$$E(W_a) = C_a^0 e^{r\tau} - e^{r\tau}\left\{P_0\left[\phi \cdot \mathbf{N}(z_H + \sigma H\sqrt{\tau}) + (1-\phi) \cdot \mathbf{N}(z_L + \sigma L\sqrt{\tau})\right] - Ke^{-r\tau}\left[\phi \cdot \mathbf{N}(z_H) + (1-\phi) \cdot \mathbf{N}(z_L)\right]\right\}$$

Variance given a Bernoulli type Volatility

$$\begin{aligned} Var(W_a) = & \phi Var_H \left\{C_a^0 e^{r\tau} - C_T + \Delta_a[P_T - P_0 e^{r\tau}]\right\} + (1-\phi)Var_L \left\{C_a^0 e^{r\tau} - C_T + \Delta_a[P_T - P_0 e^{r\tau}]\right\} \\ & + \phi(1-\phi)(E_L - E_H)^2 \end{aligned}$$

where $E_L = E_L \left\{C_a^0 e^{r\tau} - C_T + \Delta_a[P_T - P_0 e^{r\tau}]\right\}$ and $E_H = \left\{C_a^0 e^{r\tau} - C_T + \Delta_a[P_T - P_0 e^{r\tau}]\right\}$.

To calculate $Var_H \left\{C_a^0 e^{r\tau} - C_T + \Delta_a[P_T - P_0 e^{r\tau}]\right\}$, we can rewrite $Var_H \left\{C_a^0 e^{r\tau} - C_T + \Delta_a[P_T - P_0 e^{r\tau}]\right\}$ as

$$Var_H \left\{C_a^0 e^{r\tau} - C_T + \Delta_a[P_T - P_0 e^{r\tau}]\right\} = \Delta_a^2 Var_H \{P_T\} + Var_H \{C_T\} - 2 \cdot \Delta_a \cdot Cov_H (P_T, C_T)$$

$\ln(\frac{P_\tau}{P_0})$ follows $N\left((r - \frac{1}{2}\sigma^2 H^2)\tau, \sigma^2 H^2 \tau\right)$. Then, the variance of P_τ is [22]

$$Var_H \{P_T\} = E_H \left\{P_T^2\right\} - [E_H(P_T)]^2 = P_0^2 e^{2r\tau} \left[e^{\sigma^2 H^2 \tau} - 1\right].$$

The variance of C_τ is

$$\begin{aligned}
Var_H(C_\tau) =& Var_H \left[(P_\tau - K) \cdot I(P_\tau \geq K)\right] \\
=& Var_H \left[P_\tau \cdot I(P_\tau \geq K) - K \cdot I(P_\tau \geq K)\right] \\
=& Var_H[P_\tau \cdot I(P_\tau \geq K)] + K^2 \cdot Var_H \left[I(P_\tau \geq K)\right] - 2K \cdot Cov_H[P_\tau \cdot I(P_\tau \geq K), I(P_\tau \geq K)] \\
=& P_0^2 e^{2r\tau} \left\{e^{\sigma^2 H^2 \tau} \mathbf{N}(2\sigma \cdot H\sqrt{\tau} + z_H) - \left[\mathbf{N}(\sigma \cdot H\sqrt{\tau} + z_H)\right]^2\right\} + K^2 \mathbf{N}(z_H)\mathbf{N}(-z_H) \\
& - 2P_0 K e^{r\tau} \mathbf{N}(\sigma H\sqrt{\tau} + z_H)\mathbf{N}(-z_H)
\end{aligned}$$

The co-variance of P_τ and C_τ is

$$\begin{aligned}
Cov_H (P_\tau, C_\tau) =& Cov_H \left(P_T, [P_T - K] \cdot I(P_T \geq K)\right) \\
=& E_H \left\{{P_T}^2 \cdot I(P_T \geq K)\right\} - K \cdot E_H \left\{P_T \cdot I(P_T \geq K)\right\} - E_H(P_T) \cdot E_H \left[(P_T - K)^+\right] \\
=& P_0^2 e^{2r\tau + \sigma^2 H^2 \tau} \mathbf{N}\left(2\sigma H\sqrt{\tau} + z_H\right) - \left(P_0 K e^{r\tau} + P_0^2 e^{2r\tau}\right) \mathbf{N}\left(\sigma H\sqrt{\tau} + z_H\right) + P_0 K e^{r\tau} \mathbf{N}(z_H)
\end{aligned}$$

Therefore,

$$Var_H \left\{C_a^0 e^{r\tau} - C_T + \Delta_a \left[P_T - P_0 e^{r\tau}\right]\right\} = C + A_1 \mathbf{N}(2\sigma H\sqrt{\tau} + z_H) + A_2 \mathbf{N}(\sigma H\sqrt{\tau} + z_H) + A_3 \mathbf{N}(z_H)$$

$$\begin{cases}
C = P_0^2 \Delta_a^2 e^{2r\tau} \left[e^{\sigma^2 H^2 \tau} - 1\right] \\
A_1 = P_0^2 e^{2r\tau + \sigma^2 H^2 \tau}(1 - 2\Delta_a) \\
A_2 = -P_0^2 e^{2r\tau} \mathbf{N}(\sigma H\sqrt{\tau} + z_H) - 2P_0 K e^{r\tau} \mathbf{N}(-z_H) + 2\Delta_a P_0 e^{r\tau}(K + P_0 e^{r\tau}) \\
A_3 = K^2 \mathbf{N}(-z_H) - 2\Delta_a P_0 K e^{r\tau}
\end{cases}$$

The derivation for Var_L is the same. And the unconditional Variance for a bernoulli type random volatility is:

$$\begin{aligned}
Var(W_a) =& \phi Var_H \{-C_T + \Delta_a \cdot P_T\} + (1 - \phi) Var_L \{-C_T + \Delta_a \cdot P_T\} \\
& + \phi(1 - \phi)(E_L [C_T] - E_H [C_T])^2
\end{aligned}$$

[22] For the general log-normal random variable $\ln(X) \sim N(\mu, \sigma^2)$, we have the following general results:
(1)The First and Second Moment of X.

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2} \quad and \quad E(X^2) = e^{2\mu + 2\sigma^2}.$$

(2)The Restricted First and Second Moment of X.

$$E\left[X \cdot I(X \geq x)\right] = E(X) \cdot \Phi\left(\sigma + \frac{-\ln(x) + \mu}{\sigma}\right) \quad and \quad E\left[X^2 \cdot I(X \geq x)\right] = E(X^2) \cdot \Phi\left(2\sigma + \frac{-\ln(x) + \mu}{\sigma}\right).$$

$$
\begin{aligned}
Var(W_b) = {} & \phi Var_H \left\{ C_T + \Delta_b \cdot P_T \right\} + (1-\phi) Var_L \left\{ C_T + \Delta_b \cdot P_T \right\} \\
& + \phi(1-\phi)(E_L \left[C_T \right] - E_H \left[C_T \right])^2
\end{aligned}
$$

The derivation of $Var\,(W_b)$ is the same as we did for $Var\,(W_a)$. Variance are the same, because $\Delta_a = -\Delta_b$, the proof of which is shown next.

Appendix B-2 (The Optimal Delta)

The portfolio value at the expiration date W_a and W_b can have two volatility realizations $(\sigma L, \sigma H)$ with subjective probability $1-\phi$ and ϕ respectively. The variance of W_a and W_b can be written as

$$
\begin{aligned}
Var(W_a) = & (1-\phi) Var_L[W_a] + \phi Var_H[W_a] + (1-\phi)\phi \left[E_L(W_a) - E_H(W_a) \right]^2 \\
= & (1-\phi) \left\{ \Delta_a^2 Var_L(P_T) + Var_L \left[C_T \right] - 2\Delta_a Cov_L(P_T, C_T) \right\} \\
& + \phi \left\{ \Delta_a^2 Var_H(P_T) + Var_H \left[C_T \right] - 2\Delta_a Cov_H(P_T, C_T) \right\} + (1-\phi)\phi \left[E_L(W_a) - E_H(W_a) \right]^2
\end{aligned}
$$

$$
\begin{aligned}
Var(W_b) = & (1-\phi) Var_L[W_b] + \phi Var_H[W_b] + (1-\phi)\phi \left[E_L(W_b) - E_H(W_b) \right]^2 \\
= & (1-\phi) \left\{ \Delta_b^2 Var_L(P_T) + Var_L \left[C_T \right] + 2\Delta_b Cov_L(P_T, C_T) \right\} \\
& + \phi \left\{ \Delta_b^2 Var_H(P_T) + Var_H \left[C_T \right] + 2\Delta_b Cov_H(P_T, C_T) \right\} + (1-\phi)\phi \left[E_L(W_b) - E_H(W_b) \right]^2
\end{aligned}
$$

The necessary conditions to minimize $Var\,[W_a]$ and $Var\,[W_b]$ are

$$
\begin{aligned}
& 0 = \tfrac{\partial Var(W_a)}{\partial \Delta} \\
& = (1-\phi) \left\{ \Delta_a Var_L(P_T) - Cov_L(P_T, (P_T - K)^+) \right\} + \phi \left\{ \Delta_a Var_H(P_T) - Cov_H(P_T, (P_T - K)^+) \right\} \\
& + (1-\phi)\phi \left\{ \Delta_a E_L(P_T) - E_L \left[(P_T - K)^+ \right] - \Delta_a E_H(P_T) + E_H \left[(P_T - K)^+ \right] \right\} \left[E_L(P_T) - E_H(P_T) \right]
\end{aligned}
$$

$$
\begin{aligned}
& 0 = \tfrac{\partial Var(W_b)}{\partial \Delta} \\
& = (1-\phi) \left\{ \Delta_b Var_L(P_T) + Cov_L(P_T, (P_T - K)^+) \right\} + \phi \left\{ \Delta_b Var_H(P_T) + Cov_H(P_T, (P_T - K)^+) \right\} \\
& + (1-\phi)\phi \left\{ \Delta_b E_L(P_T) + E_L \left[(P_T - K)^+ \right] - \Delta_b E_H(P_T) - E_H \left[(P_T - K)^+ \right] \right\} \left[E_L(P_T) - E_H(P_T) \right]
\end{aligned}
$$

Therefore, the optimal Δ_a^* and Δ_b^* are

$$
\begin{aligned}
& \Delta_a^* = \tfrac{E(Cov([P_T - K]^+, P_T | V))}{E[Var(P_T)|V]} \\
& = \tfrac{(1-\phi)Cov_L(P_T, (P_T - K)^+) + \phi Cov_H\left(P_T, (P_T - K)^+\right) + (1-\phi)\phi[E_L(P_T) - E_H(P_T)]\left\{E_L\left[(P_T - K)^+\right] - E_H\left[(P_T - K)^+\right]\right\}}{(1-\phi)Var_L(P_T) + \phi Var_H(P_T) + (1-\phi)\phi[E_L(P_T) - E_H(P_T)]^2}
\end{aligned}
$$

$$
\begin{aligned}
& \Delta_b^* = -\tfrac{E(Cov([P_T - K]^+, P_T | V))}{E[Var(P_T)|V]} \\
& = -\tfrac{(1-\phi)Cov_L(P_T, (P_T - K)^+) + \phi Cov_H\left(P_T, (P_T - K)^+\right) + (1-\phi)\phi[E_L(P_T) - E_H(P_T)]\left\{E_L\left[(P_T - K)^+\right] - E_H\left[(P_T - K)^+\right]\right\}}{(1-\phi)Var_L(P_T) + \phi Var_H(P_T) + (1-\phi)\phi[E_L(P_T) - E_H(P_T)]^2}
\end{aligned}
$$

Appendix B-3 (Comparative Statics for τ)

Let $f(A, \sigma, L, H, \tau; \phi) = E(W_a) - \frac{Q}{\gamma} Var(W_a) - c = 0$. We show the following propositions.

When time to maturity decreases, the ask volatility increases for ATM options, i.e.

$$\frac{\partial A}{\partial \tau} = -\frac{\frac{\partial f}{\partial \tau}}{\frac{\partial f}{\partial A}} < 0$$

Assume the ask volatility corresponds to the ask price C_a^0 in the risk neutral probability measure, i.e.

$$C_a^0 = P_0 \cdot \mathbf{N}(\sigma A\sqrt{\tau} + z_A) - Ke^{-r\tau} \cdot \mathbf{N}(z_A) \quad and \quad z_A = \frac{\ln \frac{P_0}{K} + (r - \frac{1}{2}A^2\sigma^2)\tau}{\sigma A\sqrt{\tau}}$$

We have

$$\frac{\partial f}{\partial A} = e^{r\tau} \times \frac{\partial C_a^0}{\partial A} > 0.$$

Now, we derive the equations for $\frac{\partial f}{\partial \tau}$ and $\frac{\partial A}{\partial \tau}$.

$$\frac{\partial f}{\partial \tau} = \frac{\partial E(W_a)}{\partial \tau} - \frac{Q}{\gamma}\frac{\partial Var(W_a)}{\partial \tau}$$

To calculate $\frac{\partial E(W_a)}{\partial \tau}$, we first calculate $\frac{\partial\left(e^{r\tau}C_0^a\right)}{\partial \tau}$.[23]

$$\frac{\partial\left(e^{r\tau}C_0^a\right)}{\partial \tau} = \frac{K \cdot \mathbf{n}(z_A) \cdot \sigma A}{2\sqrt{\tau}} + rP_0e^{r\tau} \cdot \mathbf{N}(z_A + \sigma \cdot A\sqrt{\tau})$$

Since $\frac{\partial C_H}{\partial \tau}$ and $\frac{\partial C_L}{\partial \tau}$ are similar to $\frac{\partial\left(e^{r\tau}C_0^a\right)}{\partial \tau}$ with H, L replacing a respectively, we can finally write $\frac{\partial E(Wa)}{\partial \tau}$ as

$$\begin{aligned}\frac{\partial E(W_a)}{\partial \tau} = \quad & \frac{K\sigma}{2\sqrt{\tau}}\left[\mathbf{n}(z_A)A - \phi \cdot \mathbf{n}(z_H)H - (1-\phi)\cdot \mathbf{n}(z_L)L\right] \\ & + rP_0e^{r\tau}\left[\mathbf{N}(z_A + \sigma A\sqrt{\tau}) - \phi \cdot \mathbf{N}(z_H + \sigma H\sqrt{\tau}) - (1-\phi)\cdot \mathbf{N}(z_L + \sigma L\sqrt{\tau})\right]\end{aligned}$$

To calculate $\frac{\partial Var(W_a)}{\partial \tau}$, we first calculate $\frac{\partial Var_H}{\partial \tau}$ and the details follow.

$$\begin{aligned}\frac{\partial Var_H}{\partial \tau} = \quad & \frac{\partial C}{\partial \tau} + \frac{\partial A_1}{\partial \tau}\mathbf{N}(2\sigma H\sqrt{\tau} + z_H) + \frac{\partial A_2}{\partial \tau}\mathbf{N}(\sigma H\sqrt{\tau} + z_H) + \frac{\partial A_3}{\partial \tau}\mathbf{N}(z_H) \\ & + A_1\frac{\partial \mathbf{N}(2\sigma H\sqrt{\tau}+z_H)}{\partial \tau} + A_2\frac{\partial \mathbf{N}(\sigma H\sqrt{\tau}+z_H)}{\partial \tau} + A_3\frac{\partial \mathbf{N}(z_H)}{\partial \tau}\end{aligned}$$

We have

[23] We have the basic results regarding the normal density function.

$$\mathbf{n}(z_A + \sigma A\sqrt{\tau}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\sigma A\sqrt{\tau}+z_A)^2} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\sigma^2A^2\tau - \frac{1}{2}z_A^2 - \sigma A\sqrt{\tau}z_A} = \frac{K}{P_0}e^{-r\tau}\mathbf{n}(z_A)$$
$$\mathbf{n}(z_A + 2\sigma A\sqrt{\tau}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(2\sigma A\sqrt{\tau}+z_A)^2} = \frac{1}{\sqrt{2\pi}}e^{-2\sigma^2A^2\tau - \frac{1}{2}z_A^2 - 2\sigma A\sqrt{\tau}z_A} = \frac{K^2}{P_0^2}e^{-2r\tau-\sigma^2A^2\tau}\mathbf{n}(z_A)$$

$\frac{\partial C}{\partial \tau} = P_0^2\Delta^2 \left[(2r+\sigma^2H^2)e^{2r\tau+\sigma^2H^2\tau} - 2re^{2r\tau}\right]$

$\frac{\partial A_1}{\partial \tau}\mathbf{N}(2\sigma H\sqrt{\tau}+z_H) = P_0^2(2r+\sigma^2H^2)(1-2\Delta)e^{2r\tau+\sigma^2H^2\tau}\mathbf{N}(2\sigma H\sqrt{\tau}+z_H)$

$A_1\frac{\partial \mathbf{N}(2\sigma H\sqrt{\tau}+z_H)}{\partial \tau} = K^2(1-2\Delta)\left(\frac{\sigma H}{\sqrt{\tau}}+\frac{\partial z_H}{\partial \tau}\right)\mathbf{n}(z_H)$

$$
\begin{aligned}
&\frac{\partial A_2}{\partial \tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H)\\
&= -KP_0e^{r\tau}\left(\frac{\sigma H}{2\sqrt{\tau}}+\frac{\partial z_H}{\partial \tau}\right)\mathbf{n}(z_H)\mathbf{N}(\sigma H\sqrt{\tau}+z_H) - 2rP_0^2e^{2r\tau}\left[\mathbf{N}(\sigma H\sqrt{\tau}+z_H)\right]^2\\
&+2KP_0e^{r\tau}\frac{\partial z_H}{\partial \tau}\mathbf{n}(z_H)\mathbf{N}(\sigma H\sqrt{\tau}+z_H) - 2rKP_0e^{r\tau}\mathbf{N}(-z_H)\mathbf{N}(\sigma H\sqrt{\tau}+z_H)\\
&+2r\Delta KP_0e^{r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H) + 4r\Delta P_0^2e^{2r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H)
\end{aligned}
$$

$$
\begin{aligned}
&A_2\frac{\partial \mathbf{N}(\sigma H\sqrt{\tau}+z_H)}{\partial \tau}\\
&= \left\{-KP_0e^{r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H) - 2K^2\mathbf{N}(-z_H) + 2\Delta K^2 + 2\Delta KP_0e^{r\tau}\right\}\left(\frac{\sigma H}{2\sqrt{\tau}}+\frac{\partial z_H}{\partial \tau}\right)\mathbf{n}(z_H)
\end{aligned}
$$

$\frac{\partial A_3}{\partial \tau}\mathbf{N}(z_H) = \left(-K^2\frac{\partial z_H}{\partial \tau}\mathbf{n}(z_H) - 2r\Delta KP_0e^{r\tau}\right)\mathbf{N}(z_H)$

$A_3\frac{\partial \mathbf{N}(z_H)}{\partial \tau} = \left(K^2\mathbf{N}(-z_H) - 2\Delta P_0Ke^{r\tau}\right)\frac{\partial z_H}{\partial \tau}\mathbf{n}(z_H)$

Adding up all the parts and rearranging, we finally have[24]

$$
\begin{aligned}
\frac{\partial Var_H}{\partial \tau} = &\quad D + B_1\mathbf{N}(2\sigma H\sqrt{\tau}+z_H) + B_2\mathbf{N}(\sigma H\sqrt{\tau}+z_H) + B_3\mathbf{N}(z_H)\\
&+\left[K^2\mathbf{N}(z_H) - KP_0e^{r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H) - \Delta K^2 + \Delta KP_0e^{r\tau}\right]\frac{\sigma H}{\sqrt{\tau}}\mathbf{n}(z_H)
\end{aligned}
$$

$$
\begin{cases}
D = P_0^2\Delta^2\left[(2r+\sigma^2H^2)e^{2r\tau+\sigma^2H^2\tau} - 2re^{2r\tau}\right]\\
B_1 = P_0^2(2r+\sigma^2H^2)(1-2\Delta)e^{2r\tau+\sigma^2H^2\tau}\\
B_2 = -2rP_0^2e^{2r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H) - 2rKP_0e^{r\tau}\mathbf{N}(-z_H) + 2r\Delta KP_0e^{r\tau} + 4r\Delta P_0^2e^{2r\tau}\\
B_3 = -2r\Delta KP_0e^{r\tau}
\end{cases}
$$

To calculate $\frac{\partial Var(W_a)}{\partial \tau}$, we also need to calculate $\frac{\partial (E_H-E_L)^2}{\partial \tau}$.

$$
\begin{aligned}
\frac{\partial (E_H-E_L)^2}{\partial \tau} &= [E_H\{C_T\} - E_L\{C_T\}] \times \frac{\partial (E_H\{C_T\}-E_L\{C_T\})}{\partial \tau}\\
&= [E_H\{C_T\} - E_L\{C_T\}] \times \left[\frac{K\sigma}{2\sqrt{\tau}}(\mathbf{n}(Z_H)H - \mathbf{n}(Z_L)L) + rP_0e^{r\tau}(\mathbf{N}(z_H+\sigma\cdot H\sqrt{\tau}) - \mathbf{N}(z_L+\sigma\cdot L\sqrt{\tau}))\right]
\end{aligned}
$$

We discuss the limiting behavior of $\frac{\partial f}{\partial \tau}$ for ATM, OTM and ITM call options in turn.

(1) For ATM call option, i.e. $K = P_0e^{r\tau}$, we have

$$\mathbf{n}(z_H) = \mathbf{n}(z_L) = \mathbf{n}(z_A) \to \frac{1}{\sqrt{2\pi}} \quad and \quad \mathbf{N}(z_H) = \mathbf{N}(z_L) = \mathbf{N}(z_A) \to \frac{1}{2}, \quad when \quad \tau \to 0$$

24 $\frac{\partial Var_L}{\partial \tau}$ can be written in the similar format with L replacing H .

Then we have the limiting behavior $\frac{\partial E(W_a)}{\partial\tau}$ and $\frac{\partial Var(W_a)}{\partial\tau}$ [25]as follows,

$$\frac{\partial E(W_a)}{\partial\tau} = \frac{K\sigma}{2\sqrt{2\pi\tau}}(A-\phi H-(1-\phi)L)+o(\frac{1}{\sqrt{\tau}})$$
$$\frac{\partial Var(W_a)}{\partial\tau} = \phi P_0^2\sigma^2H^2(\Delta_a^2-\Delta_a+\frac{1}{2}-\frac{1}{2\pi})+(1-\phi)P_0^2\sigma^2L^2(\Delta_a^2-\Delta_a+\frac{1}{2}-\frac{1}{2\pi})+\phi(1-\phi)\frac{K^2\sigma^2(H-L)^2}{4\pi}+o(1)$$

As $\tau\to 0$, $\frac{\partial E(W_a)}{\partial\tau}$approaches positive infinity [26] , while $\frac{\partial Var(W_a)}{\partial\tau}$ is a constant. Therefore, $\frac{\partial f}{\partial\tau}>0$ as time approaches maturity date, which implies that $\frac{\partial A}{\partial\tau}<0$.

(2) For OTM call option, i.e. $K>P_0e^{r\tau}$, we have

$$\frac{\mathbf{n}(z_H)}{\sqrt{\tau}}=\frac{\mathbf{n}(z_L)}{\sqrt{\tau}}=\frac{\mathbf{n}(z_A)}{\sqrt{\tau}}\to 0\quad and\quad \mathbf{N}(z_H)=\mathbf{N}(z_L)=\mathbf{N}(z_A)\to 1,\quad when\quad \tau\to 0$$

Then, the limiting behavior of $\frac{\partial E(W_a)}{\partial\tau}$ and $\frac{\partial Var(W_a)}{\partial\tau}$ is as follows,

$$\begin{aligned}\frac{\partial E(W_a)}{\partial\tau} &= o(1)\\ \frac{\partial Var(W_a)}{\partial\tau} &= \phi P_0^2\Delta_a^2\sigma^2H^2+(1-\phi)P_0^2\Delta_a^2\sigma^2L^2+o(1)\end{aligned}$$

---

[25]The last term of $\frac{\vartheta Var_H}{\vartheta\tau}$ is messy, and needs special attention. The detailed derivation, assuming $\tau\to 0$, follows.

$$\left[K^2\mathbf{N}(z_H)-KP_0e^{r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H)-\Delta K^2+\Delta KP_0e^{r\tau}\right]\frac{\sigma H}{\sqrt{\tau}}\mathbf{n}(z_H)=\left[K^2\mathbf{N}(-\frac{1}{2}\sigma H\sqrt{\tau})-K^2\mathbf{N}(\frac{1}{2}\sigma H\sqrt{\tau})\right]\frac{\sigma H}{\sqrt{\tau}}\mathbf{n}(z_H)$$
$$=-\sigma^2H^2K^2\times\frac{\mathbf{N}(\frac{1}{2}\sigma H\sqrt{\tau})-\mathbf{N}(-\frac{1}{2}\sigma H\sqrt{\tau})}{\sigma H\sqrt{\tau}}\times\mathbf{n}(-\frac{1}{2}\sigma H\sqrt{\tau})$$

As time τ goes to 0, we have

$$\lim_{\tau\to 0}\frac{\mathbf{N}(\frac{1}{2}\sigma H\sqrt{\tau})-\mathbf{N}(-\frac{1}{2}\sigma H\sqrt{\tau})}{\sigma H\sqrt{\tau}}=\mathbf{n}(0)=\frac{1}{\sqrt{2\pi}}\quad and\quad \lim_{\tau\to 0}\mathbf{n}(-\frac{1}{2}\sigma H\sqrt{\tau})=\mathbf{n}(0)=\frac{1}{\sqrt{2\pi}}$$

Therefore,

$$\lim_{\tau\to 0}\left[K^2\mathbf{N}(z_H)-KP_0e^{r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H)-\Delta K^2+\Delta KP_0e^{r\tau}\right]\frac{\sigma H}{\sqrt{\tau}}\mathbf{n}(z_H)=-\frac{1}{2\pi}\sigma^2H^2K^2$$

The limiting behavior of $\frac{\vartheta(E_H-E_L)^2}{\vartheta\tau}$ is also tricky, and the derivation is

$$\begin{aligned}\frac{\vartheta(E_H-E_L)^2}{\vartheta\tau} &= [E_H\{C_\tau\}-E_L\{C_\tau\}]\times\frac{K\sigma}{2\sqrt{\tau}}(\mathbf{n}(Z_H)H-\mathbf{n}(Z_L)L)+o(1)\\ &= 2P_0e^{r\tau}\left[\mathbf{N}(\frac{1}{2}\sigma H\sqrt{\tau})-\mathbf{N}(\frac{1}{2}\sigma L\sqrt{\tau})\right]\times\frac{K\sigma}{2\sqrt{\tau}}(\mathbf{n}(Z_H)H-\mathbf{n}(Z_L)L)+o(1)\\ &= \frac{1}{2}KP_0e^{r\tau}\sigma^2(H-L)^2\times\frac{\mathbf{N}(\frac{1}{2}\sigma H\sqrt{\tau})-\mathbf{N}(\frac{1}{2}\sigma L\sqrt{\tau})}{\frac{1}{2}\sigma(H-L)\sqrt{\tau}}\times\mathbf{n}(0)+o(1)\end{aligned}$$

As time τ goes to 0, we have

$$\lim_{\tau\to 0}\frac{\vartheta(E_H-E_L)^2}{\vartheta\tau}=\frac{K^2\sigma^2(H-L)^2}{4\pi}+o(1)$$

[26]To compensate for the hedging uncertainty and transaction cost, we have $E(W_a)>0$. For ATM call option,

$$E(W_a)=K[2\cdot\mathbf{N}(\frac{1}{2}\sigma A\sqrt{\tau})-1]-\phi\cdot K[2\cdot\mathbf{N}(\frac{1}{2}\sigma H\sqrt{\tau})-1]-(1-\phi)K[2\cdot\mathbf{N}(\frac{1}{2}\sigma L\sqrt{\tau})-1]>0$$
$$\Rightarrow\mathbf{N}(\frac{1}{2}\sigma A\sqrt{\tau})-\phi\cdot\mathbf{N}(\frac{1}{2}\sigma H\sqrt{\tau})-(1-\phi)\cdot\mathbf{N}(\frac{1}{2}\sigma L\sqrt{\tau})>0$$
$$\Rightarrow\left(\mathbf{N}(0)+\frac{1}{2}\sigma A\sqrt{\tau}\mathbf{n}(0)\right)-\phi\cdot\left(\mathbf{N}(0)+\frac{1}{2}\sigma H\sqrt{\tau}\mathbf{n}(0)\right)-(1-\phi)\cdot\left(\mathbf{N}(0)+\frac{1}{2}\sigma L\sqrt{\tau}\mathbf{n}(0)\right)+o(\sqrt{\tau})>0$$
$$\Rightarrow A-\phi H-(1-\phi)L+o(1)>0\Rightarrow A-(1-\lambda)H-\lambda L>0\quad as\quad \tau\to 0$$

As $\tau \to 0$, $\frac{\partial E(W_a)}{\partial \tau}$ becomes 0, while $\frac{\partial Var(W_a)}{\partial \tau}$ becomes a constant positive number. Therefore, $\frac{\partial f}{\partial \tau} < 0$ as time approaches maturity date, which implies that $\frac{\partial A}{\partial \tau} > 0$.

(3) For ITM call option, i.e. $K < P_0 e^{r\tau}$, we have

$$\frac{\mathbf{n}(z_H)}{\sqrt{\tau}} = \frac{\mathbf{n}(z_L)}{\sqrt{\tau}} = \frac{\mathbf{n}(z_A)}{\sqrt{\tau}} \to 0 \quad and \quad \mathbf{N}(z_H) = \mathbf{N}(z_L) = \mathbf{N}(z_A) \to 1, \quad when \quad \tau \to 0$$

Then, the limiting behavior $\frac{\partial E(W_a)}{\partial \tau}$ and $\frac{\partial Var(W_a)}{\partial \tau}$ is

$$\begin{aligned} \frac{\partial E(W_a)}{\partial \tau} &= o(1) \\ \frac{\partial Var(W_a)}{\partial \tau} &= \phi P_0^2 (1-\Delta_a)^2 \sigma^2 H^2 + (1-\phi) P_0^2 (1-\Delta_a)^2 \sigma^2 L^2 + o(1) \end{aligned}$$

As $\tau \to 0$, $\frac{\partial E(W_a)}{\partial \tau}$ becomes 0, while $\frac{\partial Var(W_a)}{\partial \tau}$ becomes a positive constant. Therefore, $\frac{\partial f}{\partial \tau} < 0$ as time approaches maturity date, which implies that $\frac{\partial A}{\partial \tau} > 0$.

Appendix B-4 (Second Derivative of τ)

$$\frac{\partial^2 A}{\partial \tau^2} = -\frac{\partial \left[\frac{\partial f}{\partial \tau} / \frac{\partial f}{\partial A}\right]}{\partial \tau} = \frac{\frac{\partial f}{\partial \tau}\frac{\partial^2 f}{\partial A \partial \tau} - \frac{\partial^2 f}{\partial \tau^2}\frac{\partial f}{\partial A}}{\left(\frac{\partial f}{\partial A}\right)^2}$$

Based on B-S model, we have

$$\begin{aligned} \frac{\partial f}{\partial A} &= K\sigma\sqrt{\tau}\mathbf{n}(z_A) > 0 \\ \frac{\partial^2 f}{\partial A \partial \tau} &= K\sigma\sqrt{\tau}\mathbf{n}(z_A)\left[\frac{1}{2\tau} - z_A \times \frac{\partial z_A}{\partial \tau}\right] \end{aligned}$$

In Proposition 1, we have already discussed the limiting behavior of $\frac{\partial f}{\partial \tau}$ for ATM ,OTM and ITM call options, and now we continue on discussing the limiting behavior of $\frac{\partial^2 f}{\partial \tau^2}$ for the three different cases.

(1) For ATM call option, we have

$$\begin{aligned} \frac{\partial^2 E(W_a)}{\partial \tau^2} &= -\frac{K\sigma}{4 \cdot \sqrt{2\pi\tau^3}}(A - \phi H - (1-\phi)L) + o(\frac{1}{\sqrt{\tau^3}}) \\ \frac{\partial Var(W_a)}{\partial \tau} &= O(1) \\ \frac{\partial^2 f}{\partial A \partial \tau} &= \frac{K\sigma \mathbf{n}(z_A)}{2\sqrt{\tau}} + o(\frac{1}{\sqrt{\tau}}) \end{aligned}$$

As $\tau \to 0$, $\frac{\partial^2 E(W_a)}{\partial \tau^2}$ becomes negative, $\frac{\partial Var(W_a)}{\partial \tau}$ tends to a positive constant, so $\frac{\partial^2 f}{\partial \tau^2}$ is negative. Also, as $\tau \to 0$, $\frac{\partial^2 f}{\partial A \partial \tau}$ becomes positive. Therefore, we can conclude that as time approaches maturity date, the ask volatility will increase at an increasing rate.

(2) For OTM call option, we have

$$\frac{\partial^2 E(W_a)}{\partial \tau^2} = O(\frac{\mathbf{n}(z_A)}{\sqrt{\tau^5}}) \quad \frac{\partial^2 Var(W_a)}{\partial \tau^2} = O(1) \quad and \quad \frac{\partial^2 f}{\partial A \partial \tau} = O(\frac{\mathbf{n}(z_A)}{\sqrt{\tau^3}})$$

As $\tau \to 0$, $\frac{\partial^2 E(W_a)}{\partial \tau^2}$ becomes 0, $\frac{\partial^2 Var(W_a)}{\partial \tau^2}$ tends to a positive constant, so $\frac{\partial^2 f}{\partial \tau^2}$ is a negative constant . The order of $\frac{\partial f}{\partial A}$ is $O(\mathbf{n}(z_A)\sqrt{\tau})$, while the order of $\frac{\partial^2 f}{\partial A \partial \tau}$ is $O(\frac{\mathbf{n}(z_A)}{\sqrt{\tau^3}})$, which implies that $\frac{\partial^2 f}{\partial A \partial \tau}$ dominates $\frac{\partial f}{\partial A}$. Therefore, as time approaches the maturity date, the ask volatility will decrease at a decreasing rate.

(3) For ITM call option, we have the same conclusion as for the OTM call option. The insight lies in noting that for OTM and ITM call option, the normal density function of z is an infinitesimal in any order of τ, i.e. $\frac{\mathbf{n}(z)}{\tau^m} \to 0$ as $\tau \to 0$, for all $m \in R$.

Appendix B-5 (The Volatility Level Effect)

Consider

$$\frac{\partial A}{\partial \sigma} = -\frac{\frac{\partial f}{\partial \sigma}}{\frac{\partial f}{\partial A}}$$

In Proposition 2, we proved that $\frac{\partial f}{\partial A} = K\sigma\sqrt{\tau}\mathbf{n}(z_A) > 0$. Hence, we only need to determine the sign of $\frac{\partial f}{\partial \sigma} = \frac{\partial E(W_a)}{\partial \sigma} - \frac{Q}{\gamma}\frac{\partial Var(W_a)}{\partial \sigma}$.

For $\frac{\partial E(W_a)}{\partial \sigma}$, we have

$$\frac{\partial E(W_a)}{\partial \sigma} = K\left[\mathbf{n}(z_A)A\sqrt{\tau} - \phi\cdot\mathbf{n}(z_H)H\sqrt{\tau} - (1-\phi)\cdot\mathbf{n}(z_L)L\sqrt{\tau}\right]$$

For $\frac{\partial Var(W_a)}{\partial \sigma}$, we first calculate $\frac{\partial Var_H}{\partial \sigma}$ using the Chain Rule.

$$\begin{aligned}\frac{\partial Var_H}{\partial \sigma} =& \frac{\partial C}{\partial \sigma} + \frac{\partial A_1}{\partial \sigma}\mathbf{N}(2\sigma H\sqrt{\tau}+z_H) + \frac{\partial A_2}{\partial \sigma}\mathbf{N}(\sigma H\sqrt{\tau}+z_H) + \frac{\partial A_3}{\partial \sigma}\mathbf{N}(z_H)\\ &+ A_1\frac{\partial \mathbf{N}(2\sigma H\sqrt{\tau}+z_H)}{\partial \sigma} + A_2\frac{\partial \mathbf{N}(\sigma H\sqrt{\tau}+z_H)}{\partial \sigma} + A_3\frac{\partial \mathbf{N}(z_H)}{\partial \sigma}\end{aligned}$$

For every part of $\frac{\partial Var_H}{\partial \sigma}$, we have

$\frac{\partial C}{\partial \sigma} = 2P_0^2\Delta_a^2H^2\sigma\tau e^{2r\tau+\sigma^2H^2\tau}$

$\frac{\partial A_1}{\partial \sigma}\mathbf{N}(2\sigma\cdot H\sqrt{\tau}+z_H) = 2(1-2\Delta_a)P_0^2H^2\tau\sigma e^{2r\tau+\sigma^2H^2\tau}\mathbf{N}(2\sigma H\sqrt{\tau}+z_H)$

$A_1\frac{\partial \mathbf{N}(2\sigma H\sqrt{\tau}+z_H)}{\partial \sigma} = K^2(1-2\Delta_a)\left(2H\sqrt{\tau}+\frac{\partial z_H}{\partial \sigma}\right)\mathbf{n}(z_H)$

$\frac{\partial A_2}{\partial \sigma}\mathbf{N}(\sigma H\sqrt{\tau}+z_H) = P_0Ke^{r\tau}\left(\frac{\partial z_H}{\partial \sigma} - H\sqrt{\tau}\right)\mathbf{n}(z_H)\mathbf{N}(\sigma H\sqrt{\tau}+z_H)$

$A_2\frac{\partial \mathbf{N}(\sigma H\sqrt{\tau}+z_H)}{\partial \sigma} = \left\{-KP_0e^{r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H) - 2K^2\mathbf{N}(-z_H) + 2\Delta_aK^2 + 2\Delta_aKP_0e^{r\tau}\right\}\left(H\sqrt{\tau}+\frac{\partial z_H}{\partial \sigma}\right)\mathbf{n}(z_H)$

$\frac{\partial A_3}{\partial \sigma}\mathbf{N}(z_H) = -K^2\frac{\partial z_H}{\partial \sigma}\mathbf{n}(z_H)\mathbf{N}(z_H)$

$A_3\frac{\partial \mathbf{N}(z_H)}{\partial \sigma} = \left(K^2\mathbf{N}(-z_H) - 2\Delta_aP_0Ke^{r\tau}\right)\frac{\partial z_H}{\partial \sigma}\mathbf{n}(z_H)$

Adding up all the parts and rearranging, we finally have[27]

$$\begin{aligned}\frac{\partial Var_H}{\partial \sigma} &= 2P_0^2\Delta^2H^2\tau\sigma e^{2r\tau+\sigma^2H^2\tau} + 2(1-2\Delta_a)P_0^2H^2\tau\sigma e^{2r\tau+\sigma^2H^2\tau}\mathbf{N}(2\sigma H\sqrt{\tau}+z_H)\\ &+2\left[K^2\mathbf{N}(z_H) - KP_0e^{r\tau}\mathbf{N}(\sigma H\sqrt{\tau}+z_H) - \Delta_aK^2 + \Delta_aKP_0e^{r\tau}\right]H\sqrt{\tau}\mathbf{n}(z_H)\end{aligned}$$

To calculate $\frac{\partial Var(W_a)}{\partial \sigma}$, we also need to calculate $\frac{\partial (E_H-E_L)^2}{\partial \sigma}$ and the details follow.

$$\begin{aligned}\frac{\partial (E_H-E_L)^2}{\partial \sigma} &= [E_H\{C_\tau\} - E_L\{C_\tau\}]\times\frac{\partial (E_H\{C_\sigma\}-E_L\{C_\tau\})}{\partial \sigma}\\ &= [E_H\{C_\tau\} - E_L\{C_\tau\}]\times K\left[\mathbf{n}(z_H)H\sqrt{\tau} - \mathbf{n}(z_L)L\sqrt{\tau}\right]\end{aligned}$$

[27] $\frac{\vartheta Var_L}{\vartheta\sigma}$ can be written in the similar format with L replacing H .

We will discuss the limiting behavior of $\frac{\partial f}{\partial \sigma}$ for ATM, OTM and ITM call options in turn.

(1) For ATM call option, i.e. $K = P_0 e^{r\tau}$, we have

$$\mathbf{n}(z_H) = \mathbf{n}(z_L) = \mathbf{n}(z_A) \to \frac{1}{\sqrt{2\pi}} \quad and \quad \mathbf{N}(z_H) = \mathbf{N}(z_L) = \mathbf{N}(z_A) \to \frac{1}{2}, \quad when \quad \tau \to 0$$

Then we have the limiting behavior $\frac{\partial E(W_a)}{\partial \tau}$ and $\frac{\partial Var(W_a)}{\partial \tau}$ as follows,

$$\frac{\partial E(W_a)}{\partial \sigma} = \frac{K\sqrt{\tau}}{\sqrt{2\pi}}(A - \phi H - (1-\phi)L) + o(\sqrt{\tau})\tilde{\ }O(\sqrt{\tau}) \quad and \quad \frac{\partial Var(W_a)}{\partial \sigma} = O(\tau)$$

As $\tau \to 0$, $\frac{\partial E(W_a)}{\partial \sigma}$ dominates $\frac{\partial Var(W_a)}{\partial \sigma}$. Therefore, $\frac{\partial f}{\partial \sigma} > 0$ as time approaches maturity date, which implies that $\frac{\partial A}{\partial \sigma} < 0$.

(2) For OTM call option, i.e. $K > P_0 e^{r\tau}$, we have

$$\frac{\partial E(W_a)}{\partial \sigma} = O(\mathbf{n}(z)\sqrt{\tau}) \quad and \quad \frac{\partial Var(W_a)}{\partial \sigma} = O(\tau)$$

Since $\mathbf{n}(z)$ is an infinitesimal of $o(\tau^n)$ for any n when $z \to \propto$, $\frac{\partial Var(W_a)}{\partial \sigma}$ dominates $\frac{\partial E(W_a)}{\partial \sigma}$, which implies that $\frac{\partial f}{\partial \sigma} < 0$ as $\tau \to 0$. Therefore, $\frac{\partial A}{\partial \sigma} > 0$ as time approaches maturity date.

(3) For ITM call option, i.e. $K > P_0 e^{r\tau}$, we use the similar reasoning as in OTM, and we reach the same conclusion.

Appendix C

Appendix C-1 (The optimal dynamic hedging strategy)

We first compute the variance over two periods. According to the law of total variance, the total variance for a two-periods model is

$$\begin{aligned}
& Var_0\left(W_{a,2}\right) \\
& = E_0[Var_1(W_{a,2})] + Var_0\left[E_1(W_{a,2})\right] \\
& \quad \because Var_1(W_{a,2}) = E_1[Var_2(W_{a,2})] + Var_1\left[E_2(W_{a,2})\right] \ \ and\ Var_2(W_{a,2}) = 0 \\
& = E_0[Var_0[E_1(W_{a,2})] + Var_1\left[E_2(W_{a,2})\right]] \\
& = E_0\left[Var_0\left(-C_{1,2} + \Delta_0 P_1\right) + Var_1\left(-C_{2,2} + \Delta_1 P_2\right)\right]
\end{aligned}$$

where

$$\begin{aligned}
& Var_0\left\{E_1(W_{a,2})\right\} \\
& = Var_0\left\{\left[-e^{-r\Delta\tau} \cdot E_1(P_2 - K)^+ + \Delta_0 e^{-r\Delta\tau} \cdot E_1(P_2)\right]\right\} \\
& = Var_0\left\{\left[-C_{1,2} + \Delta_0 P_1\right]\right\}
\end{aligned}$$

$$Var_1\left\{E_2(W_{a,2})\right\} = Var_1[-E_2(P_2 - K)^+ + \Delta_1 P_2] = Var_1\left[-C_{2,2} + \Delta_1 P_2\right]$$

Therefore,

$$\begin{aligned}
& Var_0\left(W_{a,2}\right) \\
& = E_0\left\{Var_0[-C_{1,2} + \Delta_0 P_1] + Var_1\left[-C_{2,2} + \Delta_1 P_2\right]\right\}
\end{aligned}$$

We let $C_{1,2}$ denote the equilibrium call price at time 1 where $C_{1,2}e^{r\Delta\tau} = E_1(P_2 - K)^+$. In addition, $C_{2,2} = (P_2 - K)^+$.

Now we derive the optimal hedging strategy. We follow the methodology in Basak and Chabakauri (2010) and apply dynamic programming to the value function J_t , which is defined as

$$J = Var_t(W_{a,2}).$$

The law of total variance yields a recursive representation for the value function.

$$J_t = \min_{\Delta_t} \Big\{ E_t(J_{t+\Delta\tau}) + Var_t[-E_{t+\Delta\tau}([P_T - K]^+ + \Delta_t P_{t+\Delta\tau})] \Big\}.$$

where Δ_t is the stock holding and $\Delta_\tau$is time interval. We first check optimization for period 1.

$$J_0 = \min_{\Delta_0} \Big\{ E_0(J_1) + Var_0[-E_1([P_2 - K]^+ + \Delta_0 P_1)] \Big\} = \min_{\Delta_0} \{E_0(J_1) + Var_0[-C_{1,2} + \Delta_0 P_1]\}.$$

By F.O.C, we get optimal $\Delta_0$as,

$$\Delta_0^* = \tfrac{Cov_0(C_{1,2}, P_1)}{Var_0[P_1]}.$$

We continue to get optimal Δ_1 for period 2.

$$J_1 = \min_{\Delta_1} E_1(J_2) + Var_1[-E_2([P_2 - K]^+ + \Delta_1 P_2)] = \min_{\Delta_0} E_0(J_2) + Var_1[-C_{2,2} + \Delta_1 P_2].$$

The solution is

$$\Delta_1^* = \tfrac{Cov_1(C_{2,2}, P_2)}{Var_1[P_2]}.$$

The general solution for multiple-periods model is also provided by Basak and Chabakauri (2012). To get analytical solution for Δ^*, we advance to compute covariance. Given in each period we have two possible realizations $(\mu_H \Delta\tau, \sigma_H^2 \Delta\tau)$ and $(\mu_L \Delta\tau, \sigma_L^2 \Delta\tau)$ with a bernoulli random arrival rate, law of total covariance yields

$$\begin{aligned}
&Cov_1(C_{2,2}, P_1) \\
&= (1-\phi) Cov_{1,L}\left[C_{2,2}, P_2\right] + \phi Cov_{1,H}\left[C_{2,2}, P_2\right] \\
&+ Cov_1 \Big[(1-\phi) E_{1,L}\left(C_{2,2}\right) + \phi E_{1,H}\left(C_{2,2}\right), (1-\phi) E_{1,L}\left(P_2\right) + \phi E_{1,H}\left(P_2\right) \Big].
\end{aligned}$$

Given $E_1(P_2|V) = P_1 e^{r\Delta\tau}$ is constant by assumption (4), we simplify the covariance as

$$\begin{aligned}
Cov_1(C_{2,2}, P_2) &= (1-\phi) Cov_{1,L}\left[C_{2,2}, P_2\right] + \phi Cov_{1,H}\left[C_{2,2}, P_2\right], \\
Cov_0(C_{1,2}, P_1) &= (1-\phi) Cov_{0,L}\left[C_{1,2}, P_1\right] + \phi Cov_{0,H}\left[C_{1,2}, P_1\right].
\end{aligned}$$

Appendix C-2 (The optimal dynamic hedging strategy)

Here we compute $C_{1,2}$ and $C_{0,2}$.

$$\begin{aligned}
&C_{1,2} = E_1[(P_2 - K)^+] \\
&= e^{-r\Delta\tau} \Big\{ \phi E_1 \Big[(P_2 - K)^+ | X_1 = x_1, V_2 = \sigma_H \Big] + (1-\phi) E_1 \Big[(P_2 - K)^+ | X_1 = x_1, V_2 = \sigma_L \Big] \Big\} \\
&= \phi BSM_{x_1,H} + (1-\phi) BSM_{x_1,L},
\end{aligned}$$

where x_1, σ_H, σ_L are the realizations of X_1 and V_2.

$$BSM_{x_1,H} = P_1 \cdot \mathbf{N}\left(\frac{\ln(\frac{P_1}{K}) + [r\Delta\tau + \frac{\sigma_H^2 \Delta\tau}{2}]}{\sigma_H \sqrt{\Delta\tau}} \right) - e^{-r\triangle\tau} K \cdot \mathbf{N}\left(\frac{\ln(\frac{P_1}{K}) + [r\Delta\tau - \frac{\sigma_H^2 \Delta\tau}{2}]}{\sigma_H \sqrt{\Delta\tau}} \right)$$

$$BSM_{x_1,L} = P_1 \cdot \mathbf{N}\left(\frac{\ln(\frac{P_1}{K}) + [r\Delta\tau + \frac{\sigma_L^2 \Delta\tau}{2}]}{\sigma_L \sqrt{\Delta\tau}} \right) - e^{-r\triangle\tau} K \cdot \mathbf{N}\left(\frac{\ln(\frac{P_1}{K}) + [r\Delta\tau - \frac{\sigma_L^2 \Delta\tau}{2}]}{\sigma_L \sqrt{\Delta\tau}} \right).$$

$$
\begin{aligned}
&C_{0,2} = e^{-2r\Delta\tau}E_0[(P_2-K)^+] \\
&= e^{-2r\Delta\tau}\left\{\phi^2 E_0\left(E_1\left[(P_2-K)^+ |X_1, V_2=\sigma_H\right]|V_1=\sigma_H\right) + (1-\phi)\phi E_0\left(E_1\left[(P_2-K)^+|X_1,V_2=\sigma_H\right]|V_1=\sigma_L\right)\right\} \\
&+e^{-2r\triangle\tau}\left\{(\phi(1-\phi)E_0\left(E_1\left[(P_2-K)^+|X_1,V_2=\sigma_L\right]|V_1=\sigma_H\right) + (1-\phi)^2E_0\left(E_1\left[(P_2-K)^+|X_1,V_2=\sigma_L\right]|V_1=\sigma_L\right)\right\} \\
&= \phi^2 BSM_{H,H} + 2\phi(1-\phi)BSM_{H,L} + (1-\phi)^2 BSM_{L,L}
\end{aligned}
$$

The following is the derivation for $BSME_{H,H}$, $BSME_{H,L}$and $BSM_{L,L}$. Here in Appendix C2, for the notation convenience, we let $\Delta\tau = 1$.

$$
\begin{aligned}
&E_0\left(E_1\left[(P_2-K)^+|X_1,V_2=\sigma_H\right]|V_1=\sigma_H\right) \\
&= E_0(E_1[P_2\cdot I(P_2>K)|X_1,V_2=\sigma_H]|V_1=\sigma_H) - E_0(E_1[K\cdot I(P_2>K)|X_1,V_2=\sigma_H]|V_1=\sigma_H)
\end{aligned}
$$

Let $f_{X_1,H}$ the normal density function. The first term can be derived as

$$
\begin{aligned}
&E_0(E_1[P_2\cdot I(P_2>K)|X_1,V_2=\sigma_H]|V_1=\sigma_H) \\
&= \int_{X_{1,H}} P_0 e^{X_{1,H}+X_{2,H}}\Pr(X_{2,H}>\ln(\tfrac{K}{P_0})-X_{1,H}|X_{1,H})f_{X_1,H}dX_{1,H} \\
&\quad \because X_{2,H} = u_H + \sigma\cdot H\varepsilon_2 \\
&= \int_{X_{1,H}} P_0 e^{X_{1,H}+u_H+\sigma_H\varepsilon_2}\Pr(\varepsilon_2 > \frac{\ln(\frac{K}{P_0})-X_{1,H}-u_H}{\sigma_H}|X_{1,H})f_{X_1,H}dX_{1,H}. \\
&\quad let\ Z = \frac{\ln(\frac{K}{P_0})-X_{1,H}-u_H}{\sigma_H} \\
&= \int_{X_{1,H}} P_0 e^{X_{1,H}+u_H}\left[\int_{\varepsilon_2>Z} e^{\sigma_H\varepsilon_2}\frac{1}{\sqrt{2\pi}}e^{-\frac{\varepsilon_2^2}{2}}d\varepsilon_2\right]f_{X_1,H}dX_{1,H} \\
&= \int_{X_{1,H}} P_0 e^{X_{1,H}+u_H+\frac{\sigma^2H^2}{2}}\left[\int_Z \frac{1}{\sqrt{2\pi}}e^{-\frac{(\varepsilon_2-\sigma_H)^2}{2}}d\varepsilon_2\right]f_{X_1,H}dX_{1,H} \\
&= \int_{X_{1,H}} P_0 e^{X_{1,H}+u_H+\frac{\sigma_H^2}{2}}\left[1-\mathbf{N}(\frac{\ln(\frac{K}{P_0})-X_{1,H}-u_H}{\sigma_H}-\sigma_H)\right]f_{X_1,H}dX_{1,H} \\
&= \int_{X_{1,H}} P_0 e^{u_H+\sigma\varepsilon_1+u_H+\frac{\sigma_H^2}{2}}\left[\mathbf{N}(\frac{X_{1,H}-\left[\ln(\frac{K}{P_0})-u_H-\sigma_H^2\right]}{\sigma_H})\right]\frac{1}{\sqrt{2\pi}\sigma_H}e^{-\frac{(X_{1,H}-u_H)^2}{2\sigma_H^2}}d[\sigma_H\varepsilon_1] \\
&\quad Note\,\mathbf{N}(\frac{X_{1,H}-[\ln(\frac{K}{P_0})-u_H-\sigma_H^2}{\sigma_H}) = \Pr(\varepsilon_2 > \frac{\ln(\frac{K}{P_0})-X_{1,H}-u_H-\sigma_H^2}{\sigma_H}). \\
&= P_0 e^{2[u+\frac{\sigma_H^2}{2}]}\int_{\varepsilon_1}\left[\mathbf{N}(\frac{\ln(\frac{P_0}{K})+2u_H+\sigma_H^2}{\sigma_H}+\varepsilon_1)\right]\frac{1}{\sqrt{2\pi}}e^{-\frac{(\varepsilon_1-\sigma_H)^2}{2}}d\varepsilon_1 \\
&\quad apply\,theorem: \int \mathbf{N}(m+s\varepsilon_1)\frac{1}{\sqrt{2\pi}}e^{-\frac{(\varepsilon_1-g)^2}{2}}d\varepsilon_1 = \mathbf{N}\left(\frac{\frac{m}{s}+g}{\sqrt{(\frac{1}{s})^2+1}}\right). \\
&= P_0 e^{2[u_H+\frac{\sigma_H^2}{2}]}\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+2u_H+2\sigma_H^2}{\sqrt{2}\sigma_H}\right).
\end{aligned}
$$

Then we derive the second term.

$$
\begin{aligned}
&E_0(E_1[K\cdot I(P_2>K|X_1,V_2=\sigma_H)|V_1=\sigma_H]) \\
&= K\int_{X_{1,H}}\Pr(X_{2,H}>\ln(\tfrac{K}{P_0})-X_{1,H}|X_{1,H})f_{X_1,H}dX_{1,H} \\
&= K\int_{X_{1,H}}\Pr(\varepsilon_2 > \frac{\ln(\frac{K}{P_0})-X_{1,H}-u_H}{\sigma_H}|X_{1,H})f_{X_1,H}dX_{1,H} \\
&= \int_{X_{1,H}} K\left[\mathbf{N}(\frac{-\ln(\frac{K}{P_0})+X_{1,H}+u_H}{\sigma_H})\right]\frac{1}{\sqrt{2\pi}\sigma_H}e^{-\frac{(X_1-u_H)^2}{2\sigma_H^2}}dX_{1,H} \\
&= \int_{\varepsilon_1} K\left[\mathbf{N}(\frac{\ln(\frac{P_0}{K})+2u_H}{\sigma_H}+\varepsilon_1)\right]\frac{1}{\sqrt{2\pi}}e^{-\frac{(\varepsilon_1)^2}{2}}d\varepsilon_1 \\
&= K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+2u_H}{\sqrt{2}\sigma_H}\right)
\end{aligned}
$$

Given $u_H + \frac{(\sigma_H)^2}{2} = r$,

$$\begin{aligned}
&e^{-2r}\left\{E_0\left(E_1\left[(P_2-K)^+|X_1, V_2=\sigma_H\right]|V_1=\sigma_H\right)\right\}\\
&= P_0\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+[r+\frac{\sigma_H^2}{2}]*2}{\sigma_H\sqrt{2}}\right) - e^{-2r}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+[r-\frac{\sigma_H^2}{2}]*2}{\sigma_H\sqrt{2}}\right)\\
&= BSM_{H,H}
\end{aligned}$$

Using the same procedure, we can derive

$$\begin{aligned}
&e^{-2r}\left\{E_0\left(E_1\left[(P_2-K)^+|X_1, V_2=\sigma_L\right]|V_1=\sigma_L\right)\right\}\\
&= P_0\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+[r+\frac{\sigma_L^2}{2}]*2}{\sigma_L\sqrt{2}}\right) - e^{-2r}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+[r-\frac{\sigma_L^2}{2}]*2}{\sigma_L\sqrt{2}}\right)\\
&= BSM_{L,L}
\end{aligned}$$

and

$$\begin{aligned}
&e^{-2r}\left\{E_0\left(E_1\left[(P_2-K)^+|X_1, V_2=\sigma_H\right]|V_1=\sigma_L\right)\right\}\\
&= P_0\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+[r+\frac{\sigma_H^2}{2}+r+\frac{\sigma_L^2}{2}]}{\sqrt{\sigma_H^2+\sigma_L^2}}\right) - e^{-2r}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+[r-\frac{\sigma_H^2}{2}+r-\frac{\sigma_L^2}{2}]}{\sqrt{\sigma_H^2+\sigma_L^2}}\right)\\
&= BSM_{L,H}
\end{aligned}$$

and

$$\begin{aligned}
&e^{-2r}\left\{E_0\left(E_1\left[(P_2-K)^+|X_1, V_2=\sigma L\right]|V_1=\sigma H\right)\right\}\\
&= P_0\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+[r+\frac{\sigma_H^2}{2}+r+\frac{\sigma_L^2}{2}]}{\sqrt{\sigma_H^2+\sigma_L^2}}\right) - e^{-2r}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_0}{K})+[r-\frac{\sigma_H^2}{2}+r-\frac{\sigma_L^2}{2}]}{\sqrt{\sigma_H^2+\sigma_L^2}}\right)\\
&= BSM_{H,L}
\end{aligned}$$

Appendix C-3 (Proof of Proposition 5)

Let $Y = \sum_{j=t+\Delta\tau}^{T} X_j$, $P_T = P_0 e^Y$ with each X_j a Bernoulli-type random normal distribution.
Therefore, we know $C_{t,T} = \sum_{i=0}^{n}\binom{n}{i}\phi^i(1-\phi)^{n-i}E_t\left[(P_T-K)^+|Var(Y) = i\cdot(\sigma_H)^2\Delta\tau + (n-i)\cdot(\sigma_L)^2\Delta\tau\right]$

$$\begin{aligned}
&C_{t,T} = \textstyle\sum_{i=0}^{n}\binom{n}{i}\phi^i(1-\phi)^{n-i}\left\{P_t\cdot\mathbf{N}(d_1) - e^{-nr\Delta\tau}K\cdot\mathbf{N}(d_2)\right\}\\
&d_1 = \frac{\ln(\frac{P_t}{K})+i\cdot[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}]+(n-i)\cdot[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}]}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n-i)\sigma_L^2\Delta\tau}} \qquad ,\\
&d_2 = \frac{\ln(\frac{P_t}{K})+i\cdot[r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}]+(n-i)\cdot[r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}]}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n-i)\sigma_L^2\Delta\tau}}, n = \frac{T}{\Delta\tau}
\end{aligned}$$

Next,

$$\begin{aligned}
&Cov_t(C_{t+1,T}\ , P_{t+1})\\
&= E_t\left[Cov_t(C_{t+1,T}, P_{t+1}|V_{t+1})\right] + Cov_t\left[E_t(C_{t+1,T}|V_{t+1}), E_t(P_{t+1}|V_{t+1})\right]\\
&\quad \because E_t(P_{t+1}|V_{t+1}) \text{ is constant by our assumption.}\\
&= E_t\left[Cov_t(C_{t+1,T}, P_{t+1}|V_{t+1})\right]\\
&= \phi\left\{E_t(C_{t+1,T}P_{t+1}|V_{t+1}=\sigma_H) - E_t(C_{t+1,T}|V_{t+1}=\sigma_H)E(P_{t+1}|V_{t+1}=\sigma_H)\right\}\\
&\quad +(1-\phi)\left\{E_t(C_{t+1,T}P_{t+1}|V_{t+1}=\sigma_L) - E_t(C_{t+1,T}|V_{t+1}=\sigma_L)E(P_{t+1}|V_{t+1}=\sigma_L)\right\}
\end{aligned}$$

The detail of first term of covariance is

$$E_t(C_{t+1,T}P_{t+1}|V_{t+1}=\sigma_H)$$
$$=E_t\left\{\sum_{i=0}^{n^*}\binom{n^*}{i}\phi^i(1-\phi)^{n^*-i}\left\{P_{t+1}^2\mathbf{N}\left(d_1\right)-P_{t+1}e^{-n^*r\Delta\tau}K\mathbf{N}\left(d_2\right)\right\}|V_{t+1}=\sigma_H\right\},$$

where $n^*=\frac{T-(t+\Delta\tau)}{\Delta\tau}$. And we advance to express the terms as

$$E_t\left\{P_{t+1}^2\mathbf{N}\left(d_1\right)-P_{t+1}e^{-n^*r\Delta\tau}K\mathbf{N}\left(d_2\right)|V_{t+1}=\sigma_H\right\}$$
$$=E_t\left[\begin{array}{c} P_t^2e^{2u_H\Delta\tau+2\sigma_H\sqrt{\Delta\tau}\varepsilon_{t+1}}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}\right]+u_H\Delta\tau}{\sqrt{i\cdot(\sigma_H)^2\Delta\tau+(n^*-i)(\sigma_L)^2\Delta\tau}}+\varepsilon_{t+1}\right) \\ -P_te^{-n^*r\Delta\tau}e^{u_H\Delta\tau+\sigma_H\sqrt{\Delta\tau}\varepsilon_{t+1}}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}\right]+u_H\Delta\tau}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}+\varepsilon_{t+1}\right)\end{array}|V_{t+1}=\sigma_H\right]$$

$$\because\int\mathbf{N}(m+s\varepsilon)\frac{1}{\sqrt{2\pi}}e^{-\frac{(\varepsilon_1-g)^2}{2}}d\varepsilon_1=\mathbf{N}\left(\frac{\frac{m}{s}+g}{\sqrt{(\frac{1}{s})^2+1}}\right)$$
$$=P_t^2e^{2r\Delta\tau+\sigma_H^2\Delta\tau}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}\right]+[r\Delta\tau+\frac{3\sigma_H^2\Delta\tau}{2}]}{\sqrt{(i+1)\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}\right)$$
$$-P_te^{-(n^*-1)r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}\right]+[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}]}{\sqrt{(i+1)\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}\right).$$

Let $r_H^+=r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}$, $r_L^+=r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}$, $r_H^-=r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}$ $and\, r_L^-=r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}$.

$$E_t\left\{P_{t+1}^2\mathbf{N}\left(d_1\right)-P_{t+1}e^{-n^*r\Delta\tau}K\mathbf{N}\left(d_2\right)|V_{t+1}=\sigma_H\right\}$$
$$=P_t^2e^{2\cdot r_H^+}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot r_H^++(n^*-i)\cdot r_L^++\left[r_H^++\sigma_H^2\Delta\tau\right]}{\sqrt{(i+1)\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}\right)$$
$$-P_te^{-(n^*-1)r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot r_H^-+(n^*-i)\cdot r_L^-+\left[r_H^+\right]}{\sqrt{(i+1)\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}\right).$$

Similarly, details inside $E_t(C_{t+1,T}P_{t+1}|V_{t+1}=\sigma_L)$ are

$$E_t\left\{P_{t+1}^2\cdot\mathbf{N}(d_1)-P_{t+1}e^{-n^*r\Delta\tau}K\cdot\mathbf{N}(d_2)|V_{t+1}=\sigma_L\right\}$$

$$=E_t\left[\begin{array}{c} P_t^2e^{2u_L\Delta\tau+2\sigma_L\sqrt{\Delta\tau}\varepsilon_{t+1}}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}\right]+u_L\Delta\tau}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}+\varepsilon_{t+1}\right) \\ -P_te^{-n^*r\Delta\tau}e^{u_L\Delta\tau+\sigma_L\sqrt{\Delta\tau}\varepsilon_{t+1}}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}]\right]+u_L\Delta\tau}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n-i)\sigma_L^2\Delta\tau}}+\varepsilon_{t+1}\right)\end{array}|V_{t+1}=\sigma_L\right]$$

$$=P_t^2e^{2r\Delta\tau+\sigma_L^2\Delta\tau}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}\right]+[r\Delta\tau+\frac{3\sigma_L^2\Delta\tau}{2}]}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i+1)\sigma_L^2\Delta\tau}}\right)$$
$$-P_te^{-(n^*-1)r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}\right]+[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}]}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i+1)\sigma_L^2\Delta\tau}}\right)$$

$$=P_t^2e^{2\cdot r_L^+}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot r_H^++(n^*-i)\cdot r_L^++\left[r_L^++\sigma_L^2\Delta\tau\right]}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i+1)\sigma_L^2\Delta\tau}}\right)$$
$$-P_te^{-(n^*-1)r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot r_H^-+(n^*-i)\cdot r_L^-+\left[r_L^+\right]}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i+1)\sigma_L^2\Delta\tau}}\right).$$

The second term $E_t(C_{t+1,T}|V_{t+1}=\sigma_H)E(P_{t+1}|V_{t+1}=\sigma_H)$ and $E_t(C_{t+1,T}|V_1=\sigma_H)E(P_{t+1}|V_{t+1}=\sigma_L)$ can be expressed separately as

$$E_t(C_{t+1,T}|V_{t+1}=\sigma_H)E(P_{t+1}|V_{t+1}=\sigma_H)$$
$$=\sum_{i=0}^{n^*}\binom{n^*}{i}\phi^i(1-\phi)^{n^*-i}E_t\left\{\left[P_{t+1}\cdot\mathbf{N}(d_1)-e^{-n^*r\Delta\tau}K\cdot\mathbf{N}(d_2)\right]|V_{t+1}=\sigma_H\right\}P_te^{u_H\Delta\tau+\frac{(\sigma_H)^2\Delta\tau}{2}}$$

and

$$E_t\left\{\left\{P_{t+1}\cdot\mathbf{N}(d_1)-e^{-n^*r\Delta\tau}K\cdot\mathbf{N}(d_2)\right\}|V_{t+1}=\sigma_H\right\}\cdot P_te^{u_H\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}}$$

$$=P_te^{r\Delta\tau}E_t\left[\begin{array}{c} P_te^{u_H\Delta\tau+\sigma_H\sqrt{\Delta\tau}\varepsilon_{t+1}}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}\right]]+(n^*-i)\cdot\left[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}\right]+u_H\Delta\tau}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}+\varepsilon_{t+1}\right) \\ -e^{-n^*r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}]+u_H\Delta\tau\right]}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}+\varepsilon_{t+1}\right)\end{array}|V_{t+1}=\sigma_H\right]$$

$$=P_t^2e^{2r\Delta\tau}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+(i+1)\cdot\left[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}\right]}{\sqrt{(i+1)\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}\right)$$
$$-P_te^{-(n^*-1)r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+(i+1)\cdot\left[r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}\right]}{\sqrt{(i+1)\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}\right)$$

$$=P_t^2e^{2r\Delta\tau}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+(i+1)\cdot r_H^++(n^*-i)\cdot r_L^+}{\sqrt{(i+1)\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}\right)$$
$$-P_te^{-(n^*-1)r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+(i+1)\cdot r_H^-+(n^*-i)\cdot r_L^-}{\sqrt{(i+1)\cdot\sigma_H^2\Delta\tau+(n^*-i)\sigma_L^2\Delta\tau}}\right)$$

Similarly

$$E_t(C_{t+1,T}|V_1=\sigma_L)E(P_{t+1}|V_1=\sigma_L)$$
$$=\sum_{i=0}^{n^*}\binom{n^*}{i}\phi^i(1-\phi)^{n^*-i}E_t\left\{\left[P_{t+1}\cdot\mathbf{N}(d_1)-e^{-n^*r\Delta\tau}K\cdot\mathbf{N}(d_2)\right]|V_{t+1}=\sigma_L\right\}\cdot P_te^{u_L\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}}$$

and

$$E_t\left\{\left[P_{t+1}\cdot\mathbf{N}(d_1)-e^{-n^*r\Delta\tau}K\cdot\mathbf{N}(d_2)\right]|V_{t+1}=\sigma_L\right\}\cdot P_te^{u_L\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}}$$
$$=P_te^{r\Delta\tau}E_t\left[\begin{array}{c}P_te^{u_L\Delta\tau+\sigma_L\sqrt{\Delta\tau}\varepsilon_{t+1}}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}\right]+u_L\Delta\tau}{\sqrt{i\cdot\sigma H^2\Delta\tau+(n^*-i)\sigma L^2\Delta\tau}}+\varepsilon_{t+1}\right)\\-e^{-n^*r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i)\cdot\left[r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}\right]+u_L\Delta\tau}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n-i)\sigma_L^2\Delta\tau}}+\varepsilon_{t+1}\right)\end{array}|V_{t+1}=\sigma_L\right]$$

$$=P_t^2e^{2r\Delta\tau}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau+\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i+1)\cdot\left[r\Delta\tau+\frac{\sigma_L^2\Delta\tau}{2}]\right]}{\sqrt{i\cdot(\sigma_H)^2\Delta\tau+(n^*-i+1)(\sigma_L)^2\Delta\tau}}\right)$$
$$-P_te^{-(n^*-1)r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot\left[r\Delta\tau-\frac{\sigma_H^2\Delta\tau}{2}\right]+(n^*-i+1)\cdot\left[r\Delta\tau-\frac{\sigma_L^2\Delta\tau}{2}\right]}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i+1)\sigma_L^2\Delta\tau}}\right)$$

$$=P_t^2e^{2r\Delta\tau}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot r_H^++(n^*-i+1)\cdot r_L^+}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i+1)\sigma_L^2\Delta\tau}}\right)$$
$$-P_te^{-(n^*-1)r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+i\cdot r_H^-+(n^*-i+1)\cdot r_L^-}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i+1)\sigma_L^2\Delta\tau}}\right).$$

Combining,

$$\phi E_t(C_{t+1,T}|V_1=\sigma_H)E(P_{t+1}|V_1=\sigma_H)+(1-\phi)E_t(C_{t+1,T}|V_1=\sigma_L)E(P_{t+1}|V_1=\sigma_L)$$
$$=\sum_{i=0}^{n^*+1}\binom{n^*+1}{i}\phi^i(1-\phi)^{n^*+1-i}\cdot P_t\left[\begin{array}{c}P_te^{2r\Delta\tau}\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+(i+1)\cdot r_H^++(n^*-i+1)\cdot r_L^+}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i+1)\sigma_L^2\Delta\tau}}\right)\\-e^{-(n^*-1)r\Delta\tau}K\cdot\mathbf{N}\left(\frac{\ln(\frac{P_t}{K})+(i+1)\cdot\cdot r_H^-+(n^*-i+1)\cdot r_L^-}{\sqrt{i\cdot\sigma_H^2\Delta\tau+(n^*-i+1)\sigma_L^2\Delta\tau}}\right)\end{array}\right]$$

The analytical solution for covariance is

$$
\begin{aligned}
&Cov_t(C_{t+1,T}\ , P_{t+1}) \\
&= \phi \sum \binom{n^*}{i} \phi^i (1-\phi)^{n^*-i} \left\{ P_t^2 e^{2r_H^+} \mathbf{N}\left(d_{1,H}\right) - P_t e^{-(n^*-1)r\Delta\tau} K \cdot \mathbf{N}\left(d_{2,H}\right) \right\} \\
&\quad + (1-\phi) \sum \binom{n^*}{i} \phi^i (1-\phi)^{n^*-i} \left\{ P_t^2 e^{2r_L^+} \mathbf{N}\left(d_{1,L}\right) - P_t e^{-(n^*-1)r\Delta\tau} K \cdot \mathbf{N}\left(d_{2,L}\right) \right\} \\
&\quad - \sum_{i=0}^{n^*+1} \binom{n^*+1}{i} \phi^i (1-\phi)^{n^*+1-i} \cdot \left\{ P_t^2 e^{2r\Delta\tau} \mathbf{N}\left(d_1^*\right) - P_t e^{-(n^*-1)r\Delta\tau} K \cdot \mathbf{N}\left(d_2^*\right) \right\}.
\end{aligned}
$$

and

$$
d_{1,H} = \frac{\ln(\frac{P_t}{K}) + i \cdot r_H^+ + (n^*-i) \cdot r_L^+ + \left[r_H^+ + \sigma_H^2 \Delta\tau \right]}{\sqrt{(i+1) \cdot \sigma_H^2 \Delta\tau + (n^*-i) \sigma_L^2 \Delta\tau}}
$$

$$
d_{2,H} = \frac{\ln(\frac{P_t}{K}) + i \cdot r_H^- + (n^*-i) \cdot r_L^- + \left[r_H^+ \right]}{\sqrt{(i+1) \cdot \sigma_H^2 \Delta\tau + (n^*-i) \sigma_L^2 \Delta\tau}}
$$

$$
d_{1,L} = \frac{\ln(\frac{P_t}{K}) + i \cdot r_H^+ + (n^*-i) \cdot r_L^+ + \left[r_L^+ + \sigma_L^2 \Delta\tau \right]}{\sqrt{i \cdot \sigma_H^2 \Delta\tau + (n^*-i+1) \sigma_L^2 \Delta\tau}}
$$

$$
d_{2,L} = \frac{\ln(\frac{P_t}{K}) + i \cdot r_H^- + (n^*-i) \cdot r_{\sigma_L}^- + \left[r_L^+ \right]}{\sqrt{i \cdot \sigma_H^2 \Delta\tau + (n^*-i+1) \sigma_L^2 \Delta\tau}}
$$

$$
d_1^* = \frac{\ln(\frac{P_t}{K}) + (i+1) \cdot r_H^+ + (n^*-i+1) \cdot r_L^+}{\sqrt{i \cdot \sigma_H^2 \Delta\tau + (n^*-i+1) \sigma_L^2 \Delta\tau}}
$$

$$
d_2^* = \frac{\ln(\frac{P_t}{K}) + (i+1) \cdot r_H^- + (n^*-i+1) \cdot r_L^-}{\sqrt{i \cdot \sigma_H^2 \Delta\tau + (n^*-i+1) \sigma_L^2 \Delta\tau}}
$$

Table 1: Descriptive Statistics

Panel (A): Descriptive Statistics for Volatility							
		Mean	Min	Max	Std	Skewness	Kurtosis
BSM-IMV (ATM)	Call (bid)	17.48%	4.59%	85.21%	10.19%	2.14	7.34
	Call (ask)	19.58%	4.68%	117.50%	13.96%	2.56	11.57
	Put (bid)	20.81%	7.75%	91.76%	10.56%	2.09	6.63
	Put (ask)	23.27%	9.48%	132.54%	12.34%	2.54	10.90
CM-IMV (OTM+ATM)	Call+Put (bid)	21.29%	8.54%	97.14%	10.84%	2.14	7.12
	Call+Put (ask)	23.81%	9.68%	129.78%	12.56%	2.56	10.85
Realized Volatility	30 Days	18.57%	5.72%	88.50%	12.35%	2.59	8.75

Panel (B): Descriptive Statistics for Volatility Spread							
		Mean	Min	Max	Std	Skewness	Kurtosis
BSM-IMV Spread (ATM)	Call (ask-bid)	1.99%	0.36%	28.42%	1.89%	6.30	65.5
	Put (ask-bid)	2.25%	0.40%	32.12%	2.06%	6.22	64.33
CM-IMV Spread	ask-bid	2.44%	0.83%	29.64%	1.94%	5.56	51.47
P. BSM-IMV Spread	Call	11.17%	1.60%	52.95%	5.59%	2.33	8.46
	Put	13.14%	1.59%	75.20%	7.46%	2.52	9.98
P. CM-IMV Spread		14.43%	3.24%	88.90%	6.94%	2.34	11.99

Table 2: Estimation Results (1)

Model : $SpPct_{c,t} = \alpha + \theta \cdot \frac{1}{TM_{t,c}} + \beta_1 \cdot RV_c + \beta_2 \cdot VRP_{c,t} + \beta_3 \cdot Jump\_intensity_{c,t} + v_c + \varepsilon_{c,t}$

α	θ	β_1	β_2	β_3	$adj-R^2$
Panel (A) : Dependent Variable: Percentage CM-IMV spread					
6.2	0.037	0.024	0.314		0.40
(0.625)***	(0.002)***	(0.028)	(0.018)***		
8,90	0.046	-0.139		0.282	0.38
(0.625)***	(0.002)***	(0.033)***		(0.019)***	
7.377	0.038	-0.068	0.244	0.178	0.43
(0.626)***	(0.002)***	(0.029)**	(0.019)***	(0.020)***	
Panel (B): Dependent Variable: Percentage BSM-IMV (ATM Call) spread					
α	θ	β_1	β_2	β_3	$adj-R^2$
7.631	0.072	-0.038	0.261		0.53
(0.717)***	(0.002)***	(0.032)	(0.022)***		
9.916	0.08	-0.178		0.244	0.53
(0.799)***	(0.002)***	(0.037)***		(0.023)***	
8.692	0.074	-0.121	0.197	0.161	0.54
(0.707)***	(0.002)***	(0.033)***	0.023***	(0.024)***	
Panel (C): Dependent Variable: Percentage BSM-IMV (ATM Put) spread					
α	θ	β_1	β_2	β_3	$adj-R^2$
9.108	0.113	-0.106	0.251		0.66
(0.858)***	(0.002)***	(0.038)**	(0.024)***		
11.37	0.121	-0.247		0.247	0.66
(0.938)***	(0.002)***	(0.043)***		(0.025)***	
10.23	0.115	-0.194	0.184	0.170	0.67
(0.847)***	(0.002)***	(0.039)***	(0.026)***	(0.027)***	

The t values in tables are already adjusted by Newy-West variance and covariance estimator.

Table 3: Estimation Results (2)

Model : $\begin{cases} SpPct_{c,t} = \alpha + \beta_1 \cdot RV_{c,t} + \beta_2 \cdot RV_c + \beta_3 \cdot VRP_{c,t} + \beta_4 \cdot Jump\_intensity_{c,t} + \varepsilon_{c,t} \ , \\ DeTrend\_SpPct_{c,t} = \alpha + \beta_1 \cdot RV_{c,t} + \beta_2 \cdot RV_c + \beta_3 \cdot VRP_{c,t} + \beta_4 \cdot Jump\_intensity_{c,t} + \varepsilon_{c,t} \ , \end{cases}$

Panel (A) : Percentage CM-IMV spread						
Dependent Variable	α	β_1	β_2	β_3	β_4	R^2
$DeTrend\_SpPct_{c,t}$	5.48 (0.78)***	0.35 (0.03)***	-0.35 (0.04)***	0.35 (0.02)***	0.1 (0.02)***	0.24
$SpPct_{c,t}$	6.9 (0.88)***	0.39 (0.03)***	-0.36 (0.04)***	0.4 (0.02)***	0.08 (0.02)***	0.26

Panel (B): Percentage BSM-IMV (ATM Call) spread						
Dependent Variable	α	β_1	β_2	β_3	β_4	R^2
$DeTrend\_SpPct_{c,t}$	6.92 (0.87)	0.33 (0.03)***	-0.38 (0.05)***	0.3 (0.03)***	0.08 (0.03)***	0.14
$SpPct_{c,t}$	9.68 (0.99)***	0.36 (0.03)***	-0.39 (0.05)***	0.34 (0.03)***	0.07 (0.03)***	0.14

Panel (C): Percentage BSM-IMV (ATM Put) spread						
Dependent Variable	α	β_2	β_3	β_4	β_5	R^2
$DeTrend\_SpPct_{c,t}$	9.17 (0.92)***	0.22 (0.04)***	-0.37 (0.05)***	0.23 (0.03)***	0.13 (0.03)***	0.09
$SpPct_{c,t}$	13.3 (1.07)***	0.26 (0.04)***	-0.39 (0.05)***	0.29 (0.03)***	0.1 (0.03)***	0.10

Table 4: Estimation Results (3)

Model : $\left\{ \begin{array}{ll} Jump\_intensity_{c,t+1} = & \alpha + \beta_1 \cdot DeTrend\_SpPct_{c,t} + \beta_2 \cdot SpPct_{c,t} \\ & +\beta_3 \cdot VRP_{c,t} + \beta_4 \cdot RV_{c,t} + \beta_5 \cdot Jump\_intensity_{c,t} + \varepsilon_{c,t} \end{array} \right.$

Panel (A) : Dependent Variable: Jump Intensity at t+1, Spread Measurement: Percentage CM-IMV spread

α	β_1	β_2	β_3	β_4	β_5	$adj-R^2$
2.352 (0.628)***	0.229 (0.027)***					0.04
2.378 (0.633)***		0.176 (0.023)***				0.03
2.23 (0.634)***	0.17 (0.049)***	0.059 (0.041)				0.04
-3.828 (0.441)***	0.102 (0.051)**	-0.188 (0.047)***	0.345 (0.025)***	0.338 (0.02)***	0.316 (0.024)***	0.42

Panel (B): Dependent Variable: Jump Intensity at t+1, Spread Measurement: Percentage BSM-IMV (ATM Call) spread

α	β_1	β_2	β_3	β_4	β_5	$adj-R^2$
3.015 (0.629)***	0.138 (0.024)***					0.02
3.1 (0.629)***		0.086 (0.017)***				0.01
2.895 (0.635)***	0.105 (0.033)***	0.033 (0.024)				0.02
-3.957 (0.447)***	0.053 (0.033)*	-0.104 (0.027)***	0.332 (0.024)***	0.335 (0.02)***	0.309 (0.024)***	0.42

Panel (C): Dependent Variable: Jump Intensity at t+1, Spread Measurement: Percentage BSM-IMV (ATM Put) spread

α	β_1	β_2	β_3	β_4	β_5	$adj-R^2$
3.194 (0.631)***	0.112 (0.021)***					0.01
3.347 (0.625)***		0.054 (0.013)***				0.01
3.072 (0.637)***	0.089 (0.026)***	0.022 (0.016)				0.02
-4.067 (0.461)***	0.046 (0.025)*	-0.067 (0.018)***	0.325 (0.024)***	0.331 (0.021)***	0.305 (0.024)***	0.41

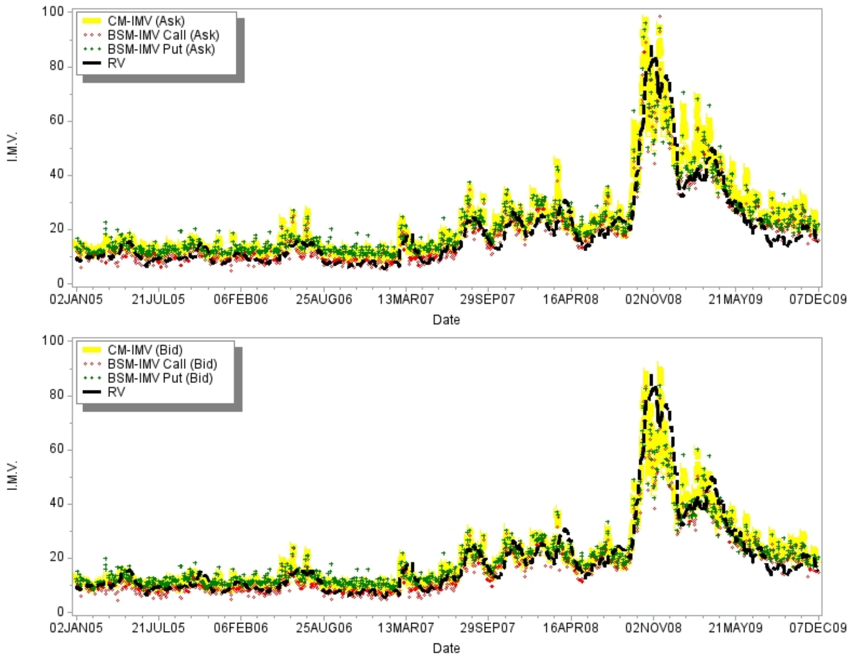

Figure 1: The Implied Volatility

Figure 1 graphs the 3 measures of implied volatilities, including CM IMV, BSM-IMV ATM Call and BSM-IMV ATM Put. The light bold line represents CM IMV, the circle signs denote BSM-IMV ATM Call, and the plus signs are BSM-IMV ATM Put. The top chart draws the IMV for bid prices of S&P index options, while the bottom chart graphs the IMV for ask prices. The 20-days realized volatility (RV) is given by the dashed line in both charts. To compute the RV, we use 5-minute high frequency returns for a historical time window of 20 days. As shown in this figure, the volatility measures are highly correlated, reaching their largest values, approximately 80%, during the 2008 financial crisis and declining to 20% in the late 2009.

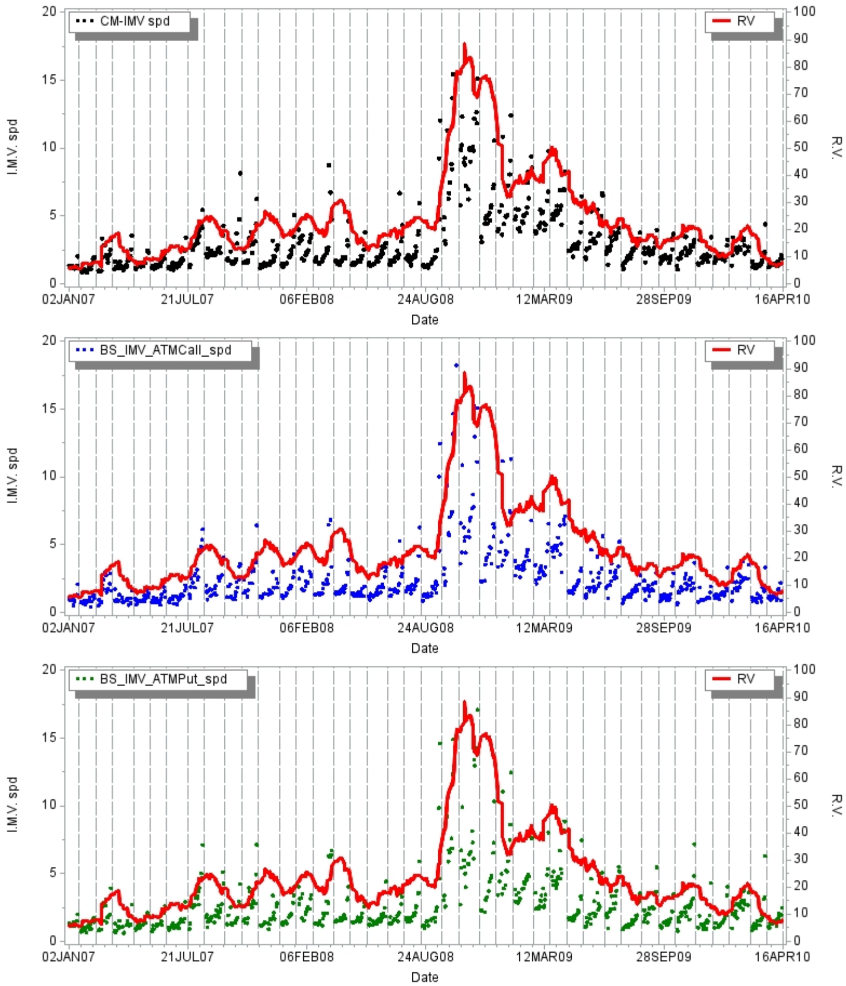

Figure 2: The Implied Volatility Spread

Figure 2 graphs the implied volatility spreads from 01/02/2007 to 04/15/2010. The first chart contains the CM-IMV spreads, the second and third charts are the IMV spreads of ATM call and ATM put, respectively. The vertical dashed lines are the expiration date, while the solid bold line is the realized volatility. As each dot represents the implied volatility spread, it can be observed that the volatility spread increases at an increasing rate as the expiration date approaches.

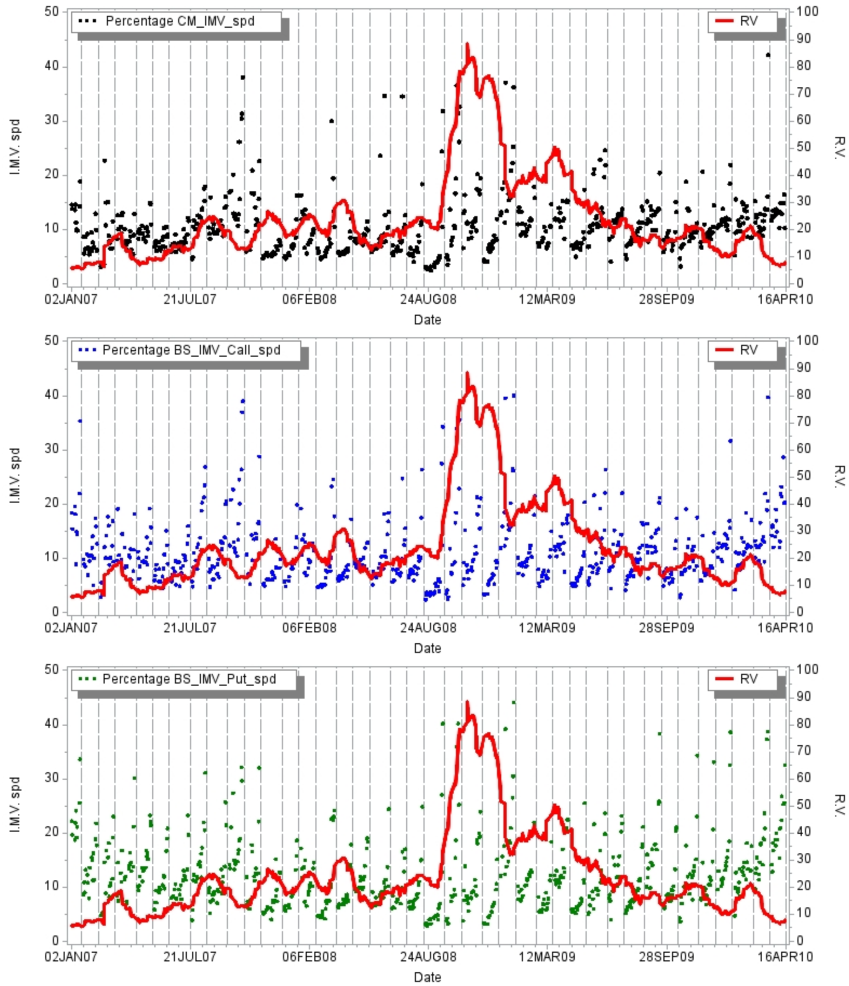

Figure 3: The Percentage Implied Volatility Spread

Figure 3 graphs the percentage implied volatility spread. The dots denote the percentage volatility spreads, while the other signs represent the same variables in figure 2. The patterns are similar to those in Figure 2, although the width of the spreads differs.

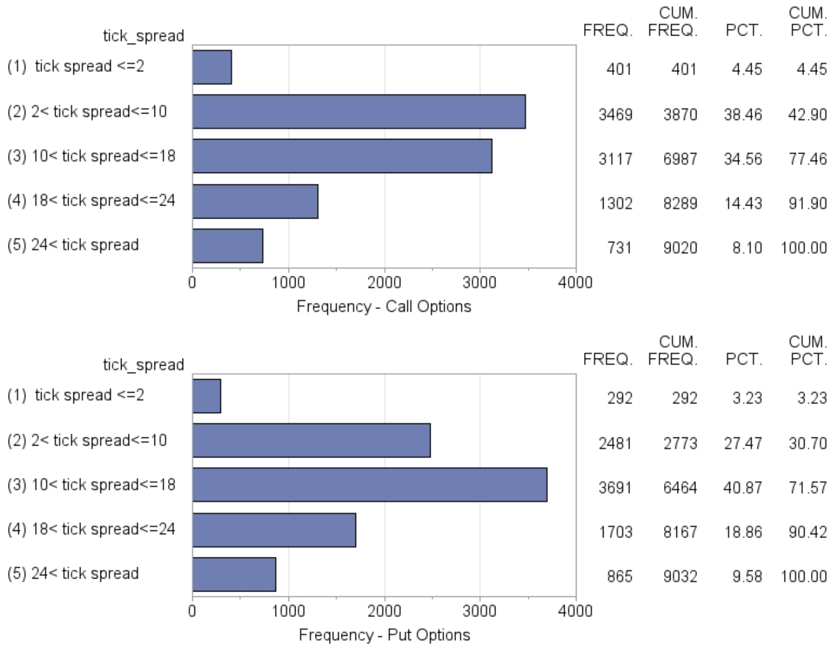

Figure 4: Minimum Spread in Ticks

Figure 4 show the spreads in terms of ticks for ATM options. The top chart display the frequency of tick spreads for call options, and the bottom one demonstrates the same information of put options. The first column divides the whole sample into by the number of tick spreads of closing prices. Those groups include: (1) tick spread ≤ 2, (2) $2 <$ tick spread ≤ 10, (3) $10 <$ tick spread ≤ 18 (3) $18 <$ tick spread ≤ 24, and (5) $24 <$ tick spread. The second column is the histogram graphically representing the number of the observations of each group, and the numerical number of frequency is shown in column 4. The percentage and the accumulated percentage of total observations number are shown in column 4 and 5.

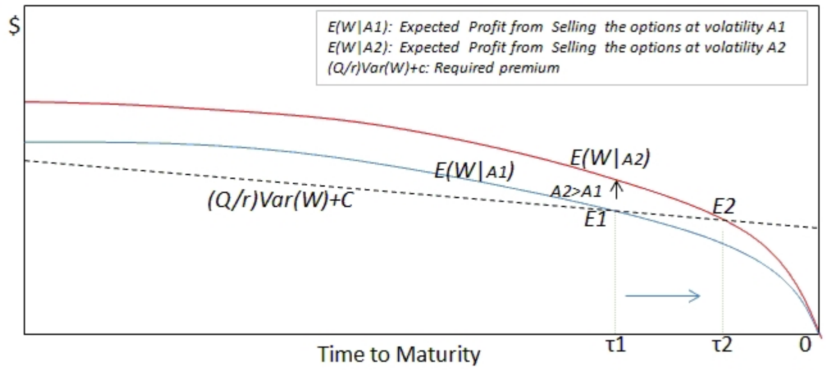

Figure 5: Quotation Dynamic

Figure 5 illustrates the maturity effect. E_1 is the first equilibrium selling the options at volatility A1. The expected profit decays faster than the required risk premium so that expected profits fall below the risk premium as the time to maturity decreases. To compensate, MMs increase the selling volatility to A2, shifting the expected extra profit curve upward, resulting in the equilibrium E_2. The result is that the implied volatility increases as the time to maturity approaches.

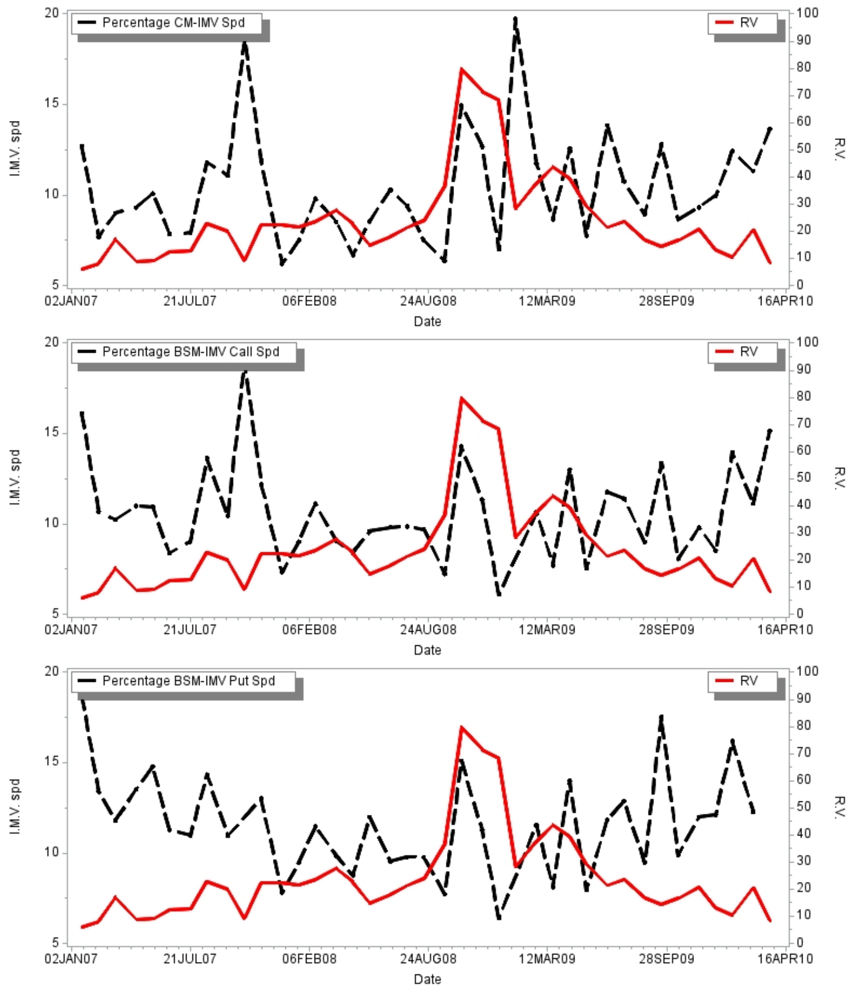

Figure 6: Averaged Volatility Percentage Spread and Volatility Level

In Figure 6, a negative relationship between percentage volatility spread and volatility level is exhibited. The dashed line is the volatility spread, and the bold line is the realized volatility. Because the increasing pattern obscures the level effect, the negative relationship is not easily observed if every data point is displayed. Here we averaged the daily data and we plot the averaged daily volatility percentage spread for each contract at the expiration date. As seen, especially in the bottom chart, when the realized volatility level is high, the spread in the percentage of implied volatilities tends to be low.

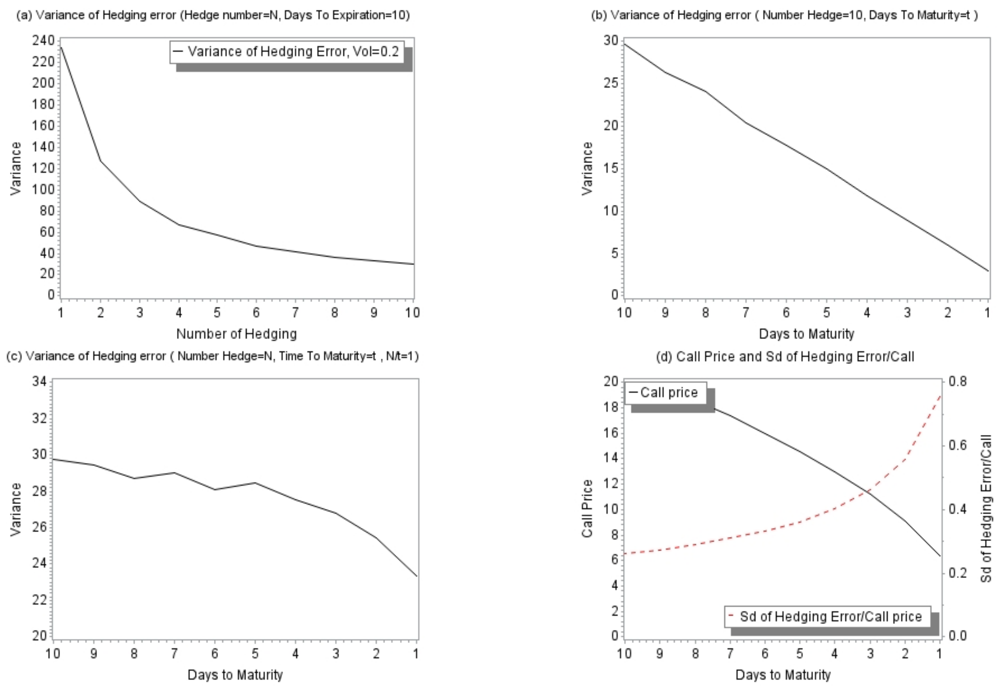

Figure 7: Total Variance and Call Prices for Different Dynamic Hedging Schemes
Figure 7 shows the numerical results for the total variance and the call price under different dynamic hedging plans. The model parameters are σ=0.2, σ_H =0.22, σ_L =0.18, r =0.04 and Po =K =1500.

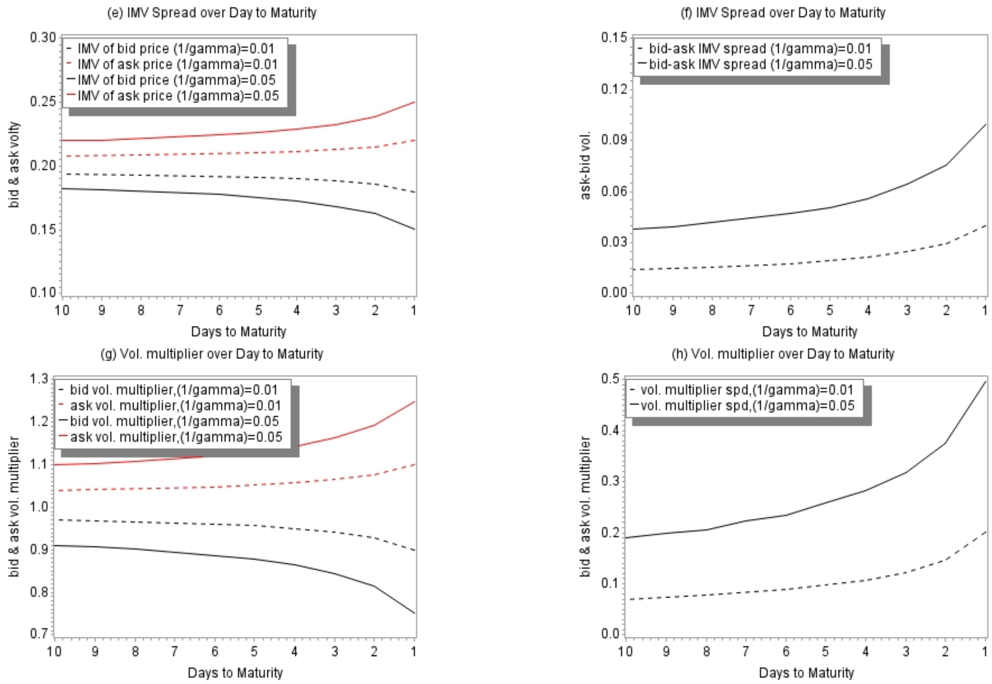

Figure 8: The IMV spread and Percentage IMV spread

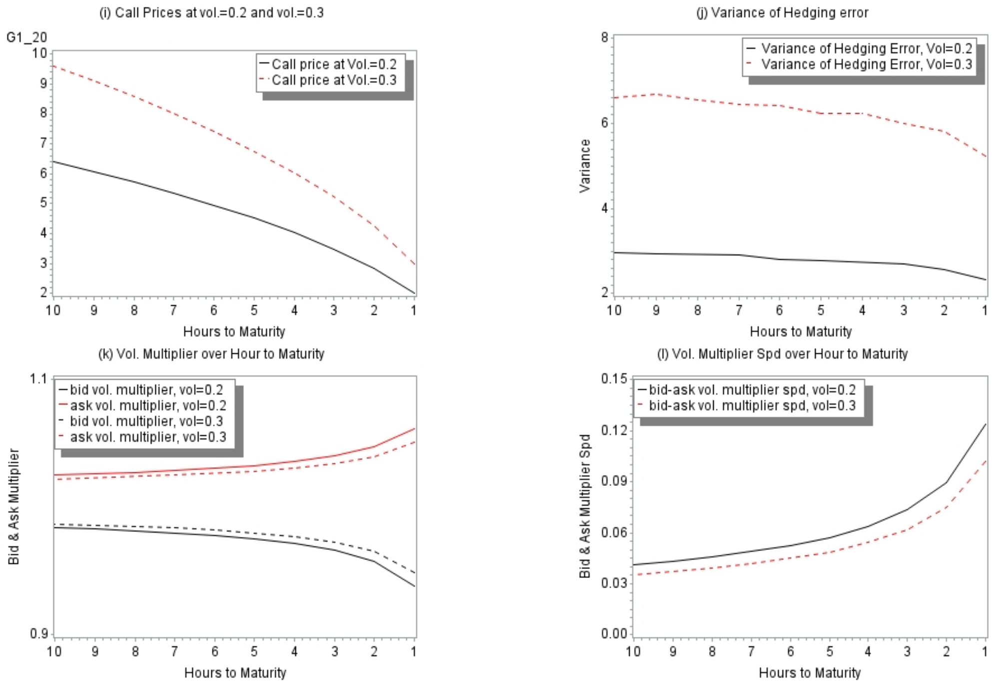

Figure 9: The Total Variance of the Hedging Error at Different Volatility Levels

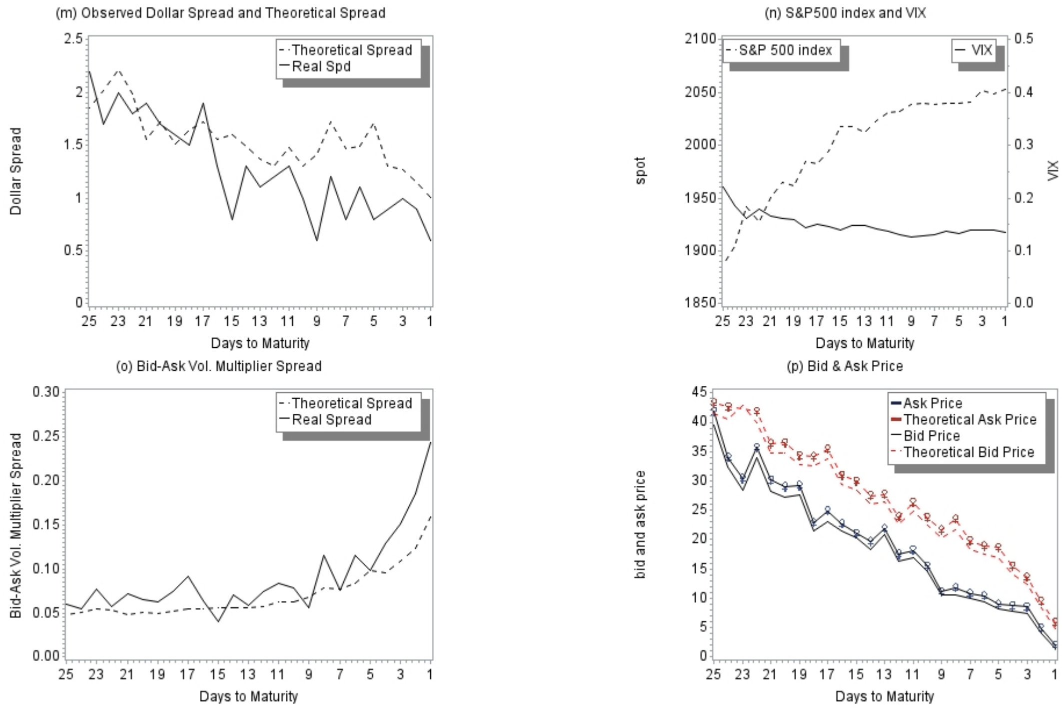

Figure 10: Calibrated Results for ATM call Option Spreads