Can the Black-Scholes-Merton Model Survive Under Transaction Costs? An Affirmative Answer

By Stylianos Perrakis and Michal Czerwonko[1]

Current draft: June 2010

Keywords: option pricing; option bounds; incomplete markets; stochastic dominance; transaction costs; diffusion processes

---

[1] Both at Concordia University. They wish to thank George Constantinides, Jens Jackwerth, Lars Stentoft, Ioan Oancea and participants at the 2008 Northern Finance Association Conference, the 2008 Quantitative Methods in Finance conference in Sydney, Australia, the 2009 Jerusalem Conference in celebration of Haim Levy's 70th birthday and retirement, and at seminars at McGill University, HEC Montreal, the University of Macedonia at Thessaloniki and the Athens University of Economics and Business for helpful advice and comment. Financial assistance from Canada's Social Sciences and Humanities Research Council and from the Institut de Finance Mathématique (IMF2) is gratefully acknowledged.

Can the Black-Scholes-Merton Model Survive Under Transaction Costs? An Affirmative Answer

By Stylianos Perrakis and Michal Czerwonko

Abstract

We derive a reservation purchase price for a call option under proportional transaction costs. The price is derived in discrete time for any number of periods and for a general distribution of the return of the underlying asset, following the stochastic dominance approach of Constantinides and Perrakis (CP, 2002, 2007). We then consider a lognormal diffusion model of this return, and we formulate a general discrete time trading version of the return that converges to diffusion as the time partition becomes progressively more dense. We show that the CP approach results in a lower bound for European call options that converges to a non-trivial and tight limit that is a function of the transaction cost parameter. This limit defines a reservation purchase price under realistic trading conditions for the call options and becomes equal to the exact Black-Scholes-Merton value if the transaction cost parameter is set equal to zero. We also develop a novel numerical algorithm that computes the CP lower bound for any discrete time partition and converges to the theoretical continuous time limit in a relatively small number of iterations. Last, we extend the lower bound results to American index and American index futures options.

I. Introduction

This paper generalizes the Black-Scholes-Merton (BSM) option pricing model to incorporate proportional transaction costs. It derives a lower bound on the price of a call option in a discrete time setting that, if violated, will create superior returns for investors under realistic trading conditions. It then examines the behavior of this derived bound as the time partition tends to zero, given that the underlying asset's price tends to a lognormal diffusion under such conditions. It is shown that the bound converges to a tight[2] and non-trivial BSM-type expression as the partition of trading time tends to zero, even if the transaction cost parameter stays constant. To our knowledge, this is the only approach to the derivation of the BSM model that can accommodate the introduction of proportional transaction costs *and* produce non-trivial results.

The derived results are part of the stochastic dominance bounds on European and American option prices in the presence of proportional transaction costs of Constantinides and Perrakis (CP, 2002, 2007). These bounds were derived for a general distribution of underlying stock returns in a discrete time context. Hence, their relationship to well-known continuous time models of option pricing is unknown. In this paper these bounds are redefined in a discrete time model of the underlying asset distribution that converges to a lognormal diffusion as the time partition tends to zero. The main result of this paper is that under such conditions the corresponding CP lower bound converges to the BSM model with the price of the underlying asset multiplied by the roundtrip transaction cost. A numerical algorithm is also presented that verifies this convergence and analyzes its properties.

Option pricing models often abstract from both market *incompleteness* and from market *imperfections* such as bid-asked spreads, brokerage fees and execution costs, collectively referred to here as transaction costs. These abstractions are serious, insofar as their relaxation comes at significant theoretical and practical costs. With dynamic market incompleteness the concept of the no-arbitrage option price is undefined, and the

[2] A tight bound is a bound that lies within a distance from the theoretical option value without transaction costs that is comparable to observed bid/ask spreads. As it will be discussed further on, this is the case with the bound derived in this paper.

available option pricing models resort to market equilibrium arguments to derive a solution.[3] It was shown recently that the stochastic dominance bounds generalize these equilibrium models and provide an alternative approach to the problem of market incompleteness.[4]

The problem of market imperfections is more serious insofar as the concept of the no-arbitrage option price is ill-defined, even if the market is dynamically complete. For example, in the Black-Scholes (1973) and Merton (1973) setting, if the market price of an option differs from its theoretical value, the investor buys the underpriced option or writes the overpriced one. The investor perfectly hedges the position by dynamic trading, thereby realizing as arbitrage profit the difference between the market price and the theoretical value. As Merton (1989) first showed, such a dynamic trading policy incurs an infinite volume of trade over the lifetime of an option. Unless transaction costs are assumed away, as they are in the BSM model and in most empirical applications of option pricing, the dynamic trading strategy ends up with trivial prices for the option, equal to the underlying stock price for the long position and to the Merton (1973) lower bound for the short option.[5]

The CP approach derived *equilibrium* (as opposed to no-arbitrage) restrictions on the *range* of the transaction prices of European and American options imposed by a class of traders that were referred to as *utility-maximizing* traders. These traders were assumed to have heterogeneous endowments and be *risk-averse*, with heterogeneous von Neuman-Morgenstern preferences which are otherwise unspecified. Furthermore it was assumed, as in most earlier studies, that trading costs in the underlying security are *proportional* to the value of the underlying security that is being traded. These defining characteristics of utility-maximizing traders apply to a broad spectrum of institutional and individual investors.

[3] See, for instance, Bates (1991), Amin and Ng (1993) and Amin (1993).
[4] See Oancea and Perrakis (2007).
[5] See, for the continuous time, Soner *et al* (1995), and for the binomial model Boyle and Vorst (1992) and Bensaid *et al* (1992).

The CP bounds defined a *range* of prices, such that *any* utility-maximizing trader would be able to exploit a mispricing, net of transaction costs, if the price of the option were to fall outside this range; the frictionless no arbitrage option price lies within the range. The reservation purchase price of an option was defined as the maximum price gross of transaction costs below which a given trader in this class increases her expected utility by purchasing the option. The reservation write price of an option is similarly defined as the minimum price net of transaction costs above which a given trader in this class increases her expected utility by writing the option. For the European call options CP (2002) defined a relatively tight reservation write price that was *independent* of the time partition, and a similarly partition-independent reservation purchase price that was, however, very loose and not particularly useful. For the American call options CP (2007) derived similarly a tight reservation write price and a very loose reservation purchase price.[6] The relationship of these bounds to the Black-Scholes price remained unclear.

CP (2002) derived also several *partition-dependent* call option prices, one reservation write and three reservation purchase ones. Neither the convergence properties of these prices, nor their discrete time values for any given time partition were provided, given the complexity of the resulting expressions. Although the discrete time distribution of the underlying stock price was assumed to have independent and identically distributed (iid) returns, the stochastic process under which the bounds were evaluated as risk neutral expectations was Markovian but with state-dependent returns that were not iid. This presented serious problems in the numerical work for their estimation. This paper presents a novel numerical approach to the estimation of expectations under such state-dependent distributions that may be used in other applications beyond the CP bounds.

In this paper we focus on one of the call option reservation purchase prices, the prices given by Proposition 5 in CP (2002).[7] This price is basically a generalization in a trading model that includes proportional transaction costs of the call option lower bounds derived

[6] The European bounds were tested empirically on S&P 500 index options in Constantinides, Jackwerth and Perrakis (2009). The American bounds were tested on S&P 500 index futures options in Constantinides, Czerwonko, Jackwerth and Perrakis (2009).

[7] It can be shown that the other partition-dependent prices are either inferior to the partition-independent ones, or tend to trivial values as the density of trading increases.

originally by Levy (1985) and Ritchken (1985), and extended to a multiperiod context by Perrakis (1988) and Ritchken and Kuo (1988). We reformulate the CP results, applying them to a case where the iid returns tend to a lognormal distribution as trading becomes progressively denser, as in Oancea-Perrakis (2009). We then show in our main result that the CP bound of Proposition 5 tends at the limit to a Black-Scholes type expression in which the current stock price has been multiplied by the roundtrip transaction costs, and becomes exactly equal to the BSM model when the transaction cost parameter is set to zero. We also show that the numerical algorithm that we develop converges to this limit in a reasonably small number of iterations, thus making the call option lower bound applicable to real life trading under realistic market conditions. Last, we extend the CP (2002) Proposition 5 to American index and index futures options, thus complementing the results of CP (2007). Given the tight and partition-independent upper bound already available from the original CP results, the results of this paper allow the extension of one of the most important models that have ever appeared in financial theory to a universe that recognizes realistic trading conditions and make it suitable for professional applications and empirical work.

In the remainder of this section we complete the literature review of the option pricing models under proportional transaction costs when the underlying asset dynamics follow a diffusion process. Proportional transaction costs were first introduced in the BSM model by Leland (1985), in a continuous time setup. The Leland model was based on imperfect replication of the option in an arbitrarily chosen discretization of the time to option expiration. The accuracy of the approximation of the option payoff and the width of the resulting option bounds were both dependent on the time partition, implying the necessity of a tradeoff between accuracy and costs of replication. Several papers explored this tradeoff, including Grannan and Swindle (1996) and Toft (1996).

The replication approach was also attempted in the binomial model by Merton (1989) and Boyle and Vorst (1992). Bensaid *et al* (1992) introduced the more general notion of super replication in the binomial model and examined the optimality of the exact replication policy, which holds only for options with physical delivery of the underlying asset. Their results were extended by Perrakis and Lefoll (2000, 2004) to American options.

Unfortunately both the continuous time discretization and the binomial approaches ended up with the same dilemma between accuracy and cost and for both option replication and super replication, insofar as the width of the option bounds increased with the time partition defining the size of the binomial tree. As shown in Boyle and Vorst (1992), the option lower bound for a sufficiently fine partition tends to a BSM expression in which the instantaneous variance σ^2 of the underlying asset return is replaced by the expression $\sigma^2\left(1-\frac{2k}{\sigma\sqrt{\Delta t}}\right)$, where k denotes the transaction cost parameter and Δt is the length of the time partition. In Leland (1985) the variance adjustment is smaller by a factor of approximately 0.8 but equally dependent inversely on Δt. It is easy to see that this expression becomes negative with probability 1 as Δt decreases, ending up with an option bound equal to the Merton (1973) arbitrage bound for a call option. A similar trivial result holds also for the option upper bound.[8]

An alternative to replication is the expected utility approach, pioneered by Hodges and Neuberger (1989). In this approach a given investor introduces an option to a portfolio of the riskless bond and the underlying asset and derives a reservation price as the price of the option that makes the investor indifferent between as to including or not the option in her portfolio. This approach was developed rigorously by Davis *et al* (1993), who solved numerically the problem for an investor with an exponential utility and a given risk aversion coefficient. Related contributions to this approach were made by Davis and Panas (1994), Constantinides and Zariphopoulou (1999, 2001), Martellini and Priaulet (2002), and Zakamouline (2006). Most studies assume that the investor's portfolio horizon equals the time to option expiration, an assumption that is both restrictive and unrealistic, given the short maturity of most options. The major drawback of this approach, however, is the dependence of the derived reservation option prices on the

[8] The impossibility of the arbitrage method to produce useful results under proportional transaction costs was shown theoretically by Soner *et al* (1996) in continuous time.

investor risk aversion coefficient. Given the uncertainty prevailing as to the size of that coefficient for the "average" investor,[9] the reservation prices derived by the expected utility approach cannot be generalized to the entire market.

II. The General Model

We adopt the same general setup as in CP (2002, 2007), in which there is a market with several assets with a group of investors who hold portfolios composed of only two of them, a riskless bond and a stock. The stock has the natural interpretation of a stock index.[10] We refer to these investors as *utility-maximizing traders* or simply as "traders". Into this setup we introduce derivative assets in the following sections: a long European call option, a long American call option, and a short European call option.

We assume that each trader makes sequential investment decisions in the primary assets at the discrete trading dates $t = 0,1,...,T'$, where T' is the terminal date and is finite.[11] A trader may hold long or short positions in these assets. A bond with price one at the initial date has price *R, R > 1* at the end of the first trading period, where *R* is a constant. The bond trades do not incur transaction costs.

At date *t*, the *cum dividend* stock price is $\left(1+\gamma_t\right)S_t$, the cash dividend is $\gamma_t S_t$, and the *ex dividend* stock price is S_t, where the *dividend yield* parameters $\left\{\gamma_t\right\}_{t=1,...,T'}$ are assumed to satisfy the condition $0 \le \gamma_t < 1$ and be deterministic and known to the trader at time zero. We assume that $S_0 > 0$ and that the support of the *ex-dividend* rate of return $\frac{S_{t+1}}{S_t} \equiv z$ on the stock is the compact subset $[z_{\min}, z_{\max}]$ of the positive real line.[12] To simplify the

[9] See Kotcherlakota (1996).
[10] There is ample evidence that many US investors follow such an indexing strategy. See Bogle (2005).
[11] The assumption that the time interval Δt between trading dates is one is innocuous: the unit of time is chosen to be such that the time interval between trading dates is one. The continuous time case will be derived as the limit of the discrete time as $\Delta t \to 0$.
[12] In CP (2002, 2007) the support is the entire positive real line. The limits on the support here are necessary because of technical conditions in considering the convergence to continuous time.

notation we also assume that $\gamma_t = \gamma$, constant for all *t*. We also assume that the rates of return are independently distributed with conditional mean return $\bar{z} \equiv E\left[(1+\gamma)z\right]$, known to the trader at time zero. We also assume that

$$\bar{z} > E[z] > R. \tag{2.1}$$

Stock trades incur proportional transaction costs charged to the bond account. At each date t, the trader pays $(1+k_1)S_t$ out of the bond account to purchase one *ex dividend* share of stock and is credited $(1-k_2)S_t$ in the bond account to sell (or, sell short) one share of stock. We assume that $0 \le k_1 < 1$ and $0 \le k_2 < 1$, and we also assume for simplicity that $k_1 = k_2 \equiv k$.

We consider a trader who enters the market at date *t* with dollar holdings v_t in the bond account and w_t / S_t *ex dividend* shares of stock. The endowments are stated net of any dividend payable on the stock at time *t*.[13] The trader increases (or, decreases) the dollar holdings in the stock account from w_t to $w_t' = w_t + \upsilon_t$ by decreasing (or, increasing) the bond account from v_t to $v_t' = v_t - \upsilon_t - k|\upsilon_t|$. The decision variable υ_t is constrained to be measurable with respect to the information up to date *t*. The bond account dynamics is

$$v_{t+1} = \{v_t - \upsilon_t - k|\upsilon_t|\}R + (w_t + \upsilon_t)\gamma z, \quad t \le T'-1 \tag{2.2}$$

and the stock account dynamics is

$$w_{t+1} = (w_t + \upsilon_t)z, \quad t \le T'-1. \tag{2.3}$$

[13] We elaborate on the precise sequence of events. The trader enters the market at date *t* with dollar holdings $v_t - \gamma_t w_t$ in the bond account and w_t / S_t *cum dividend* shares of stock. Then the stock pays cash dividend $\gamma_t w_t$ and the dollar holdings in the bond account become v_t. Thus, the trader has dollar holdings v_t in the bond account and w_t / S_t *ex dividend* shares of stock.

At the terminal date, the stock account is liquidated, $\upsilon_{T'} = -w_{T'}$, and the net worth is $v_{T'} + w_{T'} - k\left|w_{T'}\right|$. At each date *t*, the trader chooses investment υ_t to maximize the expected utility of net worth, $E\left[u\left(v_{T'} + w_{T'} - k\left|w_{T'}\right|\right) | S_t\right]$.[14] We make the plausible assumption that the utility function, $u(.)$, is increasing and concave, and is defined for both positive and negative terminal net worth.[15]

We define the value function recursively as

$$\begin{aligned} &V\left(v_t, w_t, t\right) \\ &= \max_{\upsilon} E\left[V\left(\left\{v_t - \upsilon_t - k\left|\upsilon_t\right|\right\} R + \left(w_t + \upsilon_t\right)\gamma z, \left(w_t + \upsilon_t\right) z, t+1\right)\right] \end{aligned} \tag{2.4}$$

for $t \le T' - 1$ and

$$V\left(v_{T'}, w_{T'}, T'\right) = u\left(v_{T'} + w_{T'} - k\left|w_{T'}\right|\right). \tag{2.5}$$

We assume that the parameters satisfy appropriate technical conditions such that the value function exists and is once differentiable. We denote by υ_t^* the optimal investment decision at date *t* corresponding to the portfolio (v_t, w_t). For future reference, we state that the value function $V(v, w, t)$ is increasing and concave in (v, w), properties inherited from the monotonicity and concavity of the utility function $u(.)$, given that the transaction costs are quasi-linear.[16]

Also for future reference, we define v_t' and w_t' as

---

[14] The results extend routinely to the case that consumption occurs at each trading date and utility is defined over consumption at each of the trading dates and over the net worth at the terminal date. See, Constantinides (1979) for details.

[15] If utility is defined only for non-negative net worth, then the decision variable is constrained to be a member of a convex set that ensures the non-negativity of the net worth. See, Constantinides (1979) for details. This case is studied in Constantinides and Zariphopoulou (1999, 2001). The CP (2002, 2007) bounds apply to this case as well.

[16] See, Constantinides (1979) for details.

$$v_t' = v_t - \upsilon_t^* - k\left|\upsilon_t^*\right| \tag{2.6}$$

and

$$w_t' = w_t + \upsilon_t^* . \tag{2.7}$$

Portfolio (v_t', w_t') represents the new holdings at t following optimal restructuring of the portfolio (v_t, w_t). Equations (2.5), (2.7) and (2.8) and the definition of υ_t^* imply

$$V\left(v_t, w_t, t\right) = V\left(v_t', w_t', t\right) \tag{2.8}$$

Relations (2.1)-(2.9) are sufficient for the CP (2002, 2007) derivations of the bounds. These derivations are done by considering the decision of the investor to open a long (short) option position. The corresponding reservation purchase (write) price of the option is the maximum (minimum) price above (below) which the investor with the open position will have a higher value function than the investor who did not open the option position. Both Proposition 1 of the following section, proven in the appendix, as well as Propositions 3 and 4 for the lower bounds of the American index and index futures call of Section VII, are established by this methodology.

We illustrate the methodology for the derivation of the results of the following section in the case where the transaction cost parameter is set equal to 0 in relations (2.2)-(2.8). In such a case the value function $V\left(v_t, w_t, t\right)$ becomes a function of $v_t + w_t$, say $\Omega(v_t + w_t, t) = E[\Omega(v_t^{'}R + w_t^{'}z, t+1)]$. Let $C\left(S_t z, t+1\right)$ denote the value after one period of a call option expiring at T, $t+1 \le T \le T'$, and assume that the investor may open a long position in the option at a price of C. The investor purchases the option by the zero-net-cost policy of shorting an amount δS_t, $\delta < 1$ of stock and investing the remainder in the riskless asset. The value of C should be sufficiently high, so that the investor would not be able to improve her position for any δ. The value function of the investor with the open position is $\Omega_C(v_t + \delta S_t - C + w_t - \delta S_t, t)$. We have

$$\begin{aligned}&\Omega_C(v_t+\delta S_t-C+w_t-\delta S_t,t)\geq E[\Omega_C(v_t\,'R+(\delta S_t-C)R+(w_t\,'-\delta S_t)z,t+1)]\geq \\ &E[\Omega(v_t\,'R+(\delta S_t-C)R+(w_t\,'-\delta S_t)z+C(S_tz,t+1),t+1)]\end{aligned} \quad . \tag{2.9}$$

In (2.9) the first inequality holds because the optimal restructuring portfolio policy for the investor without the open option position is not necessarily optimal for the option-holding investor and the second inequality because it may not necessarily be optimal for the investor to close her position in the next period. We define $\Delta_t \equiv \Omega_C(v_t+\delta S_t-C+w_t-\delta S_t,t)-\Omega(x_t+y_t,t)$, and we seek a lower bound on C so that $\Delta_t \geq 0$ for any call prices below the bound. Replacing the definitions of Ω and Ω_C and using (2.9) and the concavity property of the value functions we get the relation $\Delta_t \geq E[\Omega_1(z)\underline{H}(\delta,C,z)\big|S_t]$, where Ω_1 is the derivative of Ω with respect to its argument and we have used the concavity of Ω, with

$$\underline{H}(\delta,C,z)=(\delta S_t-C)R-\delta S_tz+C(S_tz,t+1)\,. \tag{2.10}$$

This last function is convex in z, is positive at $z=z_{\min}$ and has at most two zeroes, at $z=x$ and $z=\hat{z}$, while its expectation has a unique maximum in δ for any C. Using the fact that Ω_1 is a decreasing function of z we have

$$\Delta_t \geq \Pr ob(z\leq \hat{z})\Omega_1(x)E[\underline{H}(\delta,C,z)\big|S_t, z\leq \hat{z}]\,. \tag{2.11}$$

This is, however, positive unless the expectation in the right-hand-side is negative. Replacing and maximizing with respect to δ, we get the lower bound

$C\geq \dfrac{1}{R}E[C(S_tz,t+1)\big|z\leq \hat{z}],\;\; E[z\big|z\leq \hat{z}]=R$, which is the lower bound originally derived by Levy (1985) and Ritchken (1985) with different approaches. In the next section and in Appendix A the method presented here is extended to incorporate transaction costs.

III. <u>The Call Option Lower Bound in Discrete Time</u>

Proposition 5 was published without its proof in CP (2002). Given its importance, we restate it in this section, express it in a format suitable for a limiting process in continuous time, and provide its proof in Appendix A. We assume here that the dividend yield $\gamma = 0$. To simplify the notation, we define the following constants:

$$\varphi(k) \equiv (1-k)/(1+k) \text{ and } \beta(k) \equiv 2k/(1+k). \quad (3.1)$$

We also define the following function:

$$I(z) \equiv \begin{cases} 1/(1+k), & z \le 0 \\ 1/(1-k), & z > 0 \end{cases}. \quad (3.2)$$

With these definitions we have the following result, which is a slightly modified version of Proposition 5 in CP (2002).

Proposition 1: *Under the assumptions of the multiperiod economy stated in section 2, the tightest lower bound $\underline{C}(S_t, t)$ on the reservation purchase price of a call option at any time t prior to option expiration is derived recursively from the expressions*[17]

$$\underline{C}(S_{T-1}, T-1) = Max\left(E\left[(S_{T-1}z - K)^+ \mid S_{T-1}, z \le \hat{z}_{T-1} \right] \Big/ R, \left(\varphi(k) S_{T-1} - \frac{K}{R} \right)^+ \right) \quad (3.3)$$

where $\hat{z}_{T-1}$ is implied by the equation

$$E\left[z \mid z \le \hat{z}_{T-1} \right] = \varphi(k) R \quad (3.4)$$

[17] In expression (3.3) the first term in the RHS exceeds the second one by Jensen's inequality. The second term is the Merton (1973) lower bound under transaction costs. In the diffusion case that we examine in the following sections the first term tends to the second as $\Delta t \to 0$. In the numerical work we use only the second term in the algorithm.

If it is the first term within the maximum that is larger in the RHS of (3.4) then the number of shorted shares $g_{T-1}\left(S_{T-1}\right)$ *is equal to*

$$g_{T-1}\left(S_{T-1}\right)=\frac{\left(S_{T-1}\hat{z}_{T-1}-K\right)^{+}-R\underline{C}\left(S_{T-1},T-1\right)}{\left(\hat{z}_{T-1}/\varphi\left(k\right)-R\right)^{+}S_{T-1}} \tag{3.5}$$

Otherwise, if in (3.3) the bound is given by the second term, then we have, depending on whether the term is positive or zero, that $g_{T-1}\left(S_{T-1}\right)=1$ or $g_{T-1}\left(S_{T-1}\right)=0$.

At any time $t<T-1$ *we have*

$$\underline{C}\left(S_{t},t\right)=\frac{E\left[\underline{C}\left(S_{t}z,t+1\right)I\left(z-x_{t}\right)|\,S_{t},z\leq\hat{z}_{t}\right]+\beta\left(k\right)S_{t}E\left[G_{t+1}\left(S_{t}z\right)I\left(z-x_{t}\right)S_{t}z\,|\,S_{t},z\leq\hat{z}_{t}\right]}{RE\left[I\left(z-x_{t}\right)|\,S_{t},z\leq\hat{z}_{t}\right]} \tag{3.6}$$

where $\hat{z}_t$ *is implied by the equation:*

$$\frac{E\left[z\,|\,z\leq\hat{z}_{t}\right]}{\left(1-k\right)E\left[I\left(z-x_{t}\right)|\,z\leq\hat{z}_{t}\right]}=R\,, \tag{3.7}$$

and $g_t\left(S_t\right)$ *is given by*

$$g_{t}\left(S_{t}\right)=\frac{\underline{C}\left(S_{t}\hat{z}_{t},t+1\right)-R\underline{C}\left(S_{t},t\right)}{\varphi\left(k\right)\left(\hat{z}_{t}-R\right)S_{t}}\,, \tag{3.8}$$

and with $G_{t+1}\left(S_{t}z\right)\equiv\left\{g_{t+1}\left(S_{t+1}z\right)\ for\ z\leq x_{t},0\ \ for\ z>x_{t}\right\}$. x_t *is implied by the equation:*

$$R\left[\varphi\left(k\right)g_{t}\left(S_{t}\right)S_{t}-\underline{C}\left(S_{t},t\right)\right]=\varphi\left(k\right)g_{t+1}\left(S_{t}x_{t}\right)S_{t}x_{t}-\underline{C}\left(S_{t}x_{t},t+1\right). \tag{3.9}$$

Proposition 1 is based on the general model of an investor who holds a portfolio of the stock and the riskless bond of the previous section. The investor improves her utility if at any time *t* prior to option expiration she can purchase a call option at a price equal to or lower than $\underline{C}\left(S_t,t\right)$. The purchase is from the riskless bond account, but the proof of the proposition assumes that the investor also shorts an amount equal to $g_t\left(S_t\right)/(1+k)$ shares and invests the proceeds in the bond account, as in relations (2.9)-(2.11). The recursive equation (3.6) that yields the bound requires the simultaneous solution of the system (3.6)-(3.9), that determines the variables $\hat{z}_t$, x_t, $g_t(S_t)$ and $\underline{C}(S_t,t)$ at all times $\tau \in [t,T-1]$ and for all stock prices S_τ. Since all unknown quantities in the system (3.6)-(3.9) are dependent either on x_t or on quantities known at time *t*, this system may be solved by a search over admissible values for x_t. This search is specific to the distribution $f(z)$ of the return. Equations (3.10)-(3.11) below demonstrate the link between x_t and $\hat{z}_t$ under general conditions, which is made specific in Section V for a uniform distribution. The numerical algorithm that solves the recursive equations (3.3)-(3.9) is presented in Section V and Appendix D.

Equations (3.2) and (3.6)-(3.8) may also be formulated in integral form, which facilitates the numerical work. For any type of process we can rewrite equation (3.7) as follows

$$\frac{\left(1+k\right)\int_{z_{\min}}^{\hat{z}_t} zf\left(z\right)dz}{\left(1+k\right)\int_{z_{\min}}^{\hat{z}_t} f\left(z\right)dz-2k\int_{z_{\min}}^{x_t} f\left(z\right)dz}=R \tag{3.10}$$

where $f\left(z\right)$ is the density of the one-period stock return distribution. By differentiating (3.10) with respect to *x*, we have:

$$\hat{z}_t\,'=-\frac{2k}{1+k}\frac{f\left(x_t\right)}{f\left(\hat{z}_t\right)}\frac{R}{\hat{z}_t-R}\,. \tag{3.11}$$

Note that the sign of the derivative in (3.11) is strictly negative, since $\hat{z}_t > R$. We now have the following result, proven in Appendix C.

<u>Lemma 1</u>: Equation (3.9) denotes the first order condition (FOC) for the constrained maximization of (3.6) with respect to x_t, taking into account (3.10).

Next we examine a version of the discrete time returns z that converge to continuous time as trading becomes progressively more dense and $\Delta t \to 0$. We set $T-1 = T - \Delta t$ and we also set everywhere $t+1 = t + \Delta t$. The stock returns become

$$\frac{S_{t+\Delta t}}{S_t} \equiv z = 1 + \mu\Delta t + \sigma\varepsilon\sqrt{\Delta t} \tag{3.12}$$

where $\varepsilon : F(0,1)$ and F is a general distribution with bounded and compact support $\varepsilon \in [\varepsilon_{\min}, \varepsilon_{\max}]$, the counterpart of $[z_{\min}, z_{\max}]$,with density $f(.)$. It can be shown[18] that as $\Delta t \to 0$ (3.12) tends to a lognormal diffusion of the form

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW, \tag{3.13}$$

where dW denotes an elementary Wiener process.

While the returns (3.12) are clearly iid, the stochastic process for the lower bound $\underline{C}(S_t, t)$ described in Proposition 1 is Markovian but non-iid. We seek to show that in spite of this it does converge to a diffusion process whose parameters we shall determine. The weak convergence property that we use stipulates that the limit of the expectation of any bounded continuous function is equal to the expectation of the function with the distribution given by the limiting diffusion process. The criterion for

[18] See Lemma 1 in Oancea and Perrakis (2007). In fact the convergence result is stronger than the version used here, insofar as the process can be multidimensional and the parameters μ and σ can be functions of the state variable S_t.

weak convergence that we use is the *Lindeberg condition*, which was also used by Merton (1982) to develop criteria for the convergence of multinomial processes.

The Lindeberg condition stipulates that, if X_t denotes a discrete time stochastic process then a necessary and sufficient condition that X_t converges weakly to a diffusion, is that for any fixed $\delta > 0$ we must have

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|Y-X\| \geq \delta} Q_{\Delta t}\left(X, dY\right) = 0\,, \tag{3.14}$$

where $Q_{\Delta t}(X, dY)$ is the transition probability from $X_t = X$ to $X_{t+\Delta t} = Y$ during the time interval Δt. Intuitively, it requires that X_t does not change very much when the time interval Δt goes to zero. When the Lindeberg condition is satisfied the following limits define the instantaneous means and covariances of the limiting process

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|Y-X\| < \delta} \left(Y - X\right) Q_{\Delta t}\left(X, dY\right) = \mu\left(X\right) \tag{3.15}$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|Y-X\| < \delta} \left(Y - X\right)^2 Q_{\Delta t}\left(X, dY\right) = \sigma^2\left(X\right) \tag{3.16}$$

In our case we have $X_t = 1, \quad X_{t+\Delta t} = \frac{S_{t+\Delta t}}{S_t} = z$, where z is given by the process (3.12) in the absence of transaction costs. In the next section we apply the Lindeberg condition to the Markovian process described by (3.1)-(3.9) in the presence of transaction costs, given that the stock price evolves according to the process (3.12) that is known to converge to (3.13) when there are no transaction costs.

IV. The Limit of the Lower Bound in Continuous Time

Equation (3.6) shows that the call lower bound $\underline{C}(S_t, t)$ is a discounted recursive expectation of its payoff under a transformed process. Define the density

$$f_x(z) \equiv \frac{I(z-x_t)f(z)}{E[I(z-x_t)]}, \tag{4.1}$$

This is the *transaction cost adjusted* distribution of the return, with z truncated at the value $\hat{z}_t$ in taking expectations in (3.6)-(3.7). In the specialized version of the stock return given by (3.12) $f_x(z)$ becomes $f_x(\varepsilon)$, the distribution of ε is *truncated* at a value $\bar{\varepsilon} \le \varepsilon_{\max}$ with support $[\varepsilon_{\min}, \bar{\varepsilon}]$, and x_t is replaced by $\varepsilon_x \in [\varepsilon_{\min}, \bar{\varepsilon}]$. Hereafter the subscript x in expectations denotes an expectation taken with respect to this transaction cost adjusted and truncated distribution. To simplify the notation define the following expression, which enters in (3-6)-(3.7)

$$E(I) \equiv \left[\frac{1}{1+k}F(\varepsilon_x) + \left(F(\bar{\varepsilon}) - F(\varepsilon_x)\right)\frac{1}{1-k}\right]\frac{1}{F(\bar{\varepsilon})} = E\left[I(z-x_t)\middle|\varepsilon \le \bar{\varepsilon}\right]. \tag{4.2}$$

Setting again $X_t = 1$, $X_{t+\Delta t} = \frac{S_{t+\Delta t}}{S_t} = z$, we observe that in the recursive expressions (3.6) the stock return is replaced by the following process $\overline{X}_{t+\Delta t}$

$$\overline{X}_{t+\Delta t} \equiv \overline{X}^1_{t+\Delta t} + \overline{X}^2_{t+\Delta t} = \frac{z\left|\varepsilon \le \bar{\varepsilon}\right.}{(1-k)E(I)} \tag{4.3a}$$

$$\overline{X}^1_{t+\Delta t} = \frac{z\left|\varepsilon \le \bar{\varepsilon}\right.}{(1+k)E(I)} \text{ for } \varepsilon \le \varepsilon_x, = \frac{z\left|\varepsilon \le \bar{\varepsilon}\right.}{(1-k)E(I)} \text{ for } \varepsilon_x \le \varepsilon \le \bar{\varepsilon}, \tag{4.3b}$$

$$\overline{X}^2_{t+\Delta t} = \beta(k)\frac{z\left|\varepsilon \le \overline{\varepsilon}\right.}{(1-k)E(I)} \quad \text{for } \varepsilon \le \varepsilon_x, \;\; 0 \text{ for } \varepsilon > \varepsilon_x \tag{4.3c}$$

From (3.6) and Lemma A.2 in the appendix it can be easily seen that the call lower bound $\underline{C}(S_t, t)$ given by Proposition 1 is greater than or equal to the following recursive expression for t < T -1

$$\underline{C}(S_t,t) \ge \frac{E_x\left[\underline{C}\left(S_t\overline{X}^1_{t+\Delta t}, t+1\right)\left|\varepsilon \le \overline{\varepsilon}\right.\right]}{R} + \frac{2k}{1-k}\frac{E_x\left[\underline{C}\left(S_t\overline{X}^1_{t+\Delta t}, t+1\right)1_{\varepsilon\le\varepsilon_x}\left|\varepsilon \le \overline{\varepsilon}\right.\right]}{R}, \tag{4.4}$$

where the indicator function $1_{\varepsilon\le\varepsilon_x}$ denotes a quantity that is equal to 0 when $\varepsilon > \varepsilon_x$. Now by applying (4.3abc) to the RHS of (4.4) it may be easily shown that

$$\underline{C}(S_t,t) \ge \frac{E\left[\underline{C}\left(S_t\frac{z}{(1-k)E(I)}, t+1\right)\left|\varepsilon \le \overline{\varepsilon}\right.\right]}{R} = \frac{E\left[\underline{C}\left(S_t\overline{X}_{t+\Delta t}, t+1\right)\left|\varepsilon \le \overline{\varepsilon}\right.\right]}{R} \tag{4.5}$$

for any $t \le T-2$. Furthermore, equation (3.7) now becomes, neglecting the term $o(\Delta t)$

$$\frac{E\left[\varepsilon\left|\varepsilon \le \overline{\varepsilon}\right.\right]}{(1-k)E(I)} = \frac{1+\mu\Delta t + \sigma E\left[\varepsilon\left|\varepsilon \le \overline{\varepsilon}\right.\right]\sqrt{\Delta t}}{(1-k)E(I)} = R = 1 + r\Delta t \tag{4.6}$$

The key issue in applying the Lindeberg condition to evaluate the limiting distribution of the stochastic process $\overline{X}_{t+\Delta t}$ as $\Delta t \to 0$. A major role in this convergence is played by the variable $\varepsilon_x(\Delta t)$, whose limiting behavior determines in turn the limit of the key variable

$\overline{\varepsilon}$. It can be easily seen that (4.6) defines an implicit relation $\overline{\varepsilon}(\varepsilon_x)$ as ε_x varies within its support $\left[\varepsilon_{\min}, \ \overline{\varepsilon}\right]$. For $\varepsilon_x = \varepsilon_{\min}$ in equation (4.6) the value $\overline{\varepsilon}(\varepsilon_x) \equiv \overline{\varepsilon}^*$ is given by

$$E(z \left| \varepsilon \leq \overline{\varepsilon}^* \right.) = 1 + \mu\Delta t + \sigma E(\varepsilon \left| \varepsilon \leq \overline{\varepsilon}^* \right.) \sqrt{\Delta t} = 1 + r\Delta t + o(\Delta t) \tag{4.7}$$

This equation is the martingale probability corresponding to the multiperiod version of the Levy-Ritchken lower bound with stock returns given by (3.12) as the density of trading increases in the absence of transaction costs. In Oancea-Perrakis (2009) it was shown that the recursive discounted expectation of the option payoff under (4.7) tends at the continuous time limit to the Black-Scholes option value. On the other hand, for $\varepsilon_x = \overline{\varepsilon}$ equation (4.6) becomes as in (4.7) but with the RHS multiplied by $\varphi(k)$. An application of the Lindeberg condition (3.16) shows that in such a case the limiting process for $\overline{X}_{t+\Delta t}$ is a diffusion whose volatility tends to zero and defines the trivial Merton (1973) lower bound for the option. Fortunately this does not turn out to be the limiting case, and the convergence of the RHS of (4.5) under the process defined by (4.3) and (4.6) is given by the following proposition, that forms the main result of this paper and is proven in Appendix B.

Proposition 2: *The lower bound of the call option under proportional transaction costs given by Proposition 1 for the discrete stock returns defined by (3.12) tends to the Black-Scholes-Merton option value* $BSM(\varphi(k)S_t)$, *where the stock price has been multiplied by the roundtrip transaction costs and all the other parameters remain unchanged.*

This remarkable result is also the main justification for the title of this paper. The discrete time option bound given by Proposition 2 is a stochastic dominance bound, insofar as any risk averse investor holding a portfolio of the stock and the riskless bond would improve her utility if she can purchase an option at a price at or lower than the Proposition 2 bound. The utility improvement would take place under realistic market conditions that recognize the existence of proportional transaction costs in trading the stock. Unlike similar call option lower bounds derived from arbitrage models that collapse very quickly as the density of trading increases, this bound tends to a relatively tight limit. Last, the bound becomes at the limit equal to the Black-Scholes-Merton expression when the transaction cost parameter is set equal to zero.

The following auxiliary result, applicable to returns that tend to diffusion and given by (3.12), is also of interest and will serve as a verification of the numerical work. It is proven in Appendix C.

Lemma 2: The function $g_t\left(S_t\right)$ tends to $N\left(d_1^*\right)$, the option delta of the Black-Scholes-Merton expression, with the stock price multiplied by the roundtrip transaction cost term $\varphi\left(k\right)$.

Figure 1 display the limiting values of the Proposition 1 lower bound for the following parameters: $K=100$, $\sigma=20\%$, $\mu=8\%$, $r=4\%$, $T=30$, $k=0.5\%$ and 0.2%, stock price range 90-110. As expected, the lower bound is considerably tighter under this reduced transaction cost, while the upper bound is relatively unaffected. Combined with the partition-independent upper bound shown in Proposition 1 of CP (2002),[19] these figure present as tight a spread as it may be feasible to achieve. With our parameter values the two bounds define intervals of [1.95, 2.65] and [2.24, 2.63] for k=0.5% and 0.2% respectively, for an at-the-money BSM value of 2.45, corresponding to spreads of 29% and 16% of the BSM values; for S = 94 the BSM value is 0.44 and the intervals become [0.35, 0.50] and [0.43, 0.5], with spreads of 34% and 15%. For comparison purposes, the observed bid-ask spread in October 2, 2008 around noon on the S&P 500 November options was approximately 10% at the money and 20% at the value S/K = 0.94. In the following section we present the numerical estimation of the discrete time

[19] This upper bound is derived by taking expectations of the terminal payoff under the physical measure, discounting them by the expected return of the stock and dividing the result by the factor $\varphi(k)$.

model of Proposition 1, which will be used to explore the convergence properties of the bound to its continuous-time limit.

[Fig. 1 about here]

V. The Convergence of the Bound to its Continuous Time Limit

To apply Proposition 1 we use as distribution f of the iid random terms ε in (3.12) the uniform distribution with zero mean and unit variance. These last two conditions imply the following error density:

$$f(\varepsilon)=\begin{cases}1/2\sqrt{3}, & \varepsilon\in\left[-\sqrt{3},\sqrt{3}\right]\\ 0 \text{ otherwise}\end{cases}, \tag{5.1}$$

which implies the following density for the one-period return of the underlying:

$$f(z)=\begin{cases}1/2\sqrt{3}\sigma\sqrt{\Delta t}, & z\in\left[z_{\min}, z_{\max}\right]\\ 0 \text{ otherwise}\end{cases}, \tag{5.2}$$

where $z_{\min}$, $\left(z_{\max}\right)=1+\mu\Delta t-(+)\sqrt{3}\sigma\sqrt{\Delta t}$.

For the uniformly distributed disturbances, there exists a closed-form solution for $\hat{z}_t$ in the equation (3.10) for a given *x*. Integrating (3.10) under the uniform density and rearranging yields the following second-order polynomial in $\hat{z}_t$:

$$\hat{z}_t^{\,2}-2R\hat{z}_t+c\left(x_t\right)=0, \tag{5.3}$$

where $c(x_t) = 2R(\varphi(k) z_{\min} + \beta(k) x_t) - z_{\min}^2$. The solution for $\hat{z}_t$ is given by the higher of the two roots of (5.3):

$$\hat{z}_t = R + \sqrt{R^2 - c(x_t)}\,. \tag{5.4}$$

The relation between $\hat{z}_t$ and x_t is shown in Figure 2, which plots relation (5.4) for Δt equal to 1/10 and 1/20 days. To have the two graphs comparable, we scale both independent and independent variable by dividing them by *R*. Note a lower range of both x_t and $\hat{z}_t$ in terms of *R* for the coarser time partition.

[Fig. 2 about here]

In our numerical approach we apply recursive numerical integration, which provides input quantities to the system (3.6)-(3.9) at each time *t*, $t \le T-2$.[20] We solve this system by directly searching for the value of x_t for which (3.6) attains its maximum value, since we know from Lemma 1 that (3.9) is the FOC for the maximization of (3.6). The function $g_t(S_t)$ follows directly from (3.8) for this maximized value of $\underline{C}(S_t, t)$ in (3.6). We detail our numerical approach in Appendix D.

VI. <u>Numerical Convergence Results for Proposition 2</u>

We apply the numerical algorithm described in the previous section to our base case, which uses $k = 0.5\%$ and $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$ and $T = 30$ days. Figure 3 shows the convergence behavior for three different stock prices 98, 100 and 102, with the time partition ranging from 10 to 150. The figure shows clearly that the numerically derived bounds approach the known limit price given by Proposition 2.

[20] At *t* = *T-1* we use the Merton bound in (3.3) with the corresponding value of $g_{T-1}(S_{T-1})$.

[Fig. 3 about here]

Given that continuous trading may in fact be infeasible in practice, it is of interest how close to its limit the lower bound becomes for a 'realistic' time partition. For instance, for daily trading, i.e. 30 subdivisions for the stock prices 98, 100 and 102 the numerical algorithm yields the respective lower bounds of 1.127, 1.909 and 2.967. The corresponding Proposition 2 continuous-time limits are 1.169, 1.954 and 3.011, with differences from the discrete-time values approximately equal to five cents. Note that even for such a coarse subdivision the Proposition 1 discrete time lower bounds are much higher than the corresponding Leland (1985) and Boyle-Vorst (1992) lower bounds. For instance, for a stock price of 100 the call lower bound is 1.28 for Leland and 0.665 for Boyle-Vorst; for half-day trading (60 subdivisions) both bounds collapse to zero.

We also derive relative errors of the convergence to the limit, defined as $1-\underline{C}/BSM\left(\varphi(k)S,.\right)$. In Figure 4, we display these errors for the stock price range from 90 to 110 and for time partitions of 30, 70, 110 and 150. It is clear from Figure 4 that the relative errors tend to zero as the time partition increases, but at a decreasing speed. It is also clear that the convergence speed in terms of relative errors is increasing in the degree of moneyness S/K.

[Fig. 4 about here]

Although systematic results on dollar errors are not shown, we note that the dollar errors decrease, as expected, as the density of time partition increases. These errors peak approximately for at-the-money options. For instance, for the time partition 150 and S = 90, 100 and 110, we find respective errors of 0.002, 0.012 and 0.003, with the respective limiting results for the bound of 0.052, 1.954 and 9.391. The respective Black-Scholes-Merton prices for $k=0$ are: 0.081, 2.451 and 10.433.

We also verify the convergence of our numerical algorithm to the Proposition 2 lower bound as a function of the time to maturity of the option for a *fixed* number time intervals, which implies that the computational time stays approximately constant. It also implies that the size of the time partition Δt increases with maturity. Since this convergence takes place as $\Delta t \to 0$, it is expected that accuracy will decrease as time to maturity increases. Table 1 displays the results for the time to maturity in a range 30-240 days for three different ratios of moneyness, 0.9, 1 and 1.1, and with the number of time intervals used in the computations kept constant at 150 in all cases. The size of the *absolute* errors increases for all options, but the increase is small. In the case of OTM options the absolute increase in errors is dominated by the increase in the value of the option because of the longer maturity, so that the percentage error decreases. For ATM and ITM options the percentage error increases, but the increase is very small. The last column shows the distance of the limit from the BSM value, which as expected decreases as time to maturity increases because of the properties of the BSM function.

[Table 1 about here]

Last, we verify our numerical results by examining the behavior of the *g*-function, which should converge to $N\left(d_1^*\right)$ by Lemma 3, where $d_1^* = d_1\left(\varphi(k)S,.\right)$. Recall that the bound is derived by arguing that whenever the call price is below the lower limit, the investor sells $g_t(S_t) < 1$ shares, purchases the call option and invests the remainder of the proceeds in the riskless asset, which leads to an increase in his expected utility. Figure 5 displays $N\left(d_1^*\right)$ and the *g*-function for the stock price range from 90 to 110 for the time partitions 30 and 150. It is clear that the *g*-function approaches its theoretical limit from above as the time partition increases. To show the convergence of the *g*-function more systematically, we present relative errors from the limit, $1 - g\big/N\left(d_1^*\right)$ for the time partitions of 30, 70, 110 and 150 in Figure 6. These errors clearly decrease at the partition increases, with the convergence speed increasing in the S/K ratio.

[Fig. 5 about here]

[Fig. 6 about here]

VII. Extensions to American Call Index and Index Futures Options

We provide here the extension of the Proposition 1 call lower bound result to American call index and index futures options, along the lines of the CP (2007) results for the call upper bounds for these same options. The proof for Proposition 4 is presented in abridged form in Appendix E, while the proof for Proposition 5 is similar to Propositions 1 and 4 and to the proofs presented in CP (2007), and is omitted.

As in CP (2007), we consider a trader who does not hold the option, with utility function $V\left(v_t, w_t, t\right)$ given in Section 2. By purchasing the American call index option the trader's utility with a long open position in the option becomes[21]

$$
\begin{aligned}
&J\left(v_t, w_t, S_t, t\right) \\
&= \max \begin{bmatrix} V\left(v_t + S_t(1+\gamma) - K, w_t, t\right), \\ \max_j E\left[J\begin{pmatrix} \left\{v_t - j - k\left|j\right| + \left(w_t + j\right)\gamma z\right\} R, \\ \left(w_t + j\right) z, S_t z, t+1 \end{pmatrix} | S_t \right] \end{bmatrix}
\end{aligned}
\tag{7.1}
$$

for $t \le T-1$, and

$$
J\left(v_T, w_T, S_T, T\right) = V\left(v_T + \left[S_T - K\right]^+, w_T, T\right). \tag{7.2}
$$

This formulation recognizes the possibility of early exercise.

As in earlier work, we define the reservation purchase price of the American call as the maximum price below which *any* trader increases his/her expected utility by purchasing

[21] Here we assume $\gamma > 0$, since otherwise there is no early exercise.

the call. For a given trader this reservation purchase price is defined as $Max\left\{C\left|J\left(v_t - C, w_t, S_t, t\right) \geq V\left(v_t, w_t, t\right)\right.\right\}$. It is a price that depends on the utility function of the trader, as well as on her portfolio holdings (v_t, w_t). By definition, a trader who observes a market price lower than her reservation purchase price should establish a long position in the call option. As with Proposition 1, the following results provide a *lower bound*, $\underline{C}\left(S_t, t\right)$, to the reservation purchase prices of *all* traders, which is *independent* of the form of the utility function and the trader portfolio. Consequently, any trader who observes at time t a market price $C \leq \underline{C}(S_t, t)$ should establish a long position in the option.

Proposition 3: *Under the assumptions of the multiperiod economy stated in Section 2, the tightest lower bound* $\underline{C}\left(S_t, t\right)$ *on the reservation purchase price of an American call index option at any time t prior to option expiration is derived recursively from the expressions:*

For any $t \leq T-1$

$$\underline{C}\left(S_t, t\right) = \max\left\{S_t(1+\gamma) - K, N(S_t, t)\right\} \tag{7.3}$$

where the function $N(S_t, t)$ *is defined as follows*

$$N(S_{T-1}, T-1) = \left(\frac{1-k}{1+k} S_{T-1} - \frac{K}{R}\right)^+, \; t = T-1$$

$$N\left(S_t, t\right) = \frac{E\left[\max\left\{\left(1+\gamma\right)\varphi(k) S_t z - K, N\left(S_t z, t+1\right)\right\} I(z - x_t) | S_t, z \leq \hat{z}_t\right]}{RE[I(z - x_t) | S_t, z \leq \hat{z}_t]} +, \tag{7.4}$$

$$\beta(k) \frac{E[S_t z G_{t+1}(S_t z) I(z - x_t) | S_t, z \leq \hat{z}_t]}{RE[I(z - x_t) | S_t, z \leq \hat{z}_t]}, \; t < T-1$$

the function $I(.)$ *is defined in (3.2), the variables* $\hat{z}_t$ *and* x_t *and the function* $g_t(S_t)$ *are defined implicitly by*

$$\frac{E\left[\left(1+\frac{\gamma}{1-k}\right)z \mid S_t, z \le \hat{z}_t\right]}{\left(1-k\right)E\left[I(z-x_t) \mid S_t, z \le \hat{z}_t\right]} = R , \tag{7.5}$$

$$g_t(S_t) = \frac{Max\{\varphi(k)S_t\hat{z}_t(1+\gamma) - K, N(S_t\hat{z}_t, t+1)\} - RN(S_t, t)}{\varphi(k)(\hat{z}_t - R)S_t} \tag{7.6}$$

$$R[\varphi(k)S_t g_t(S_t) - N(S_t, t)] = \varphi(k)S_t x_t g_{t+1}(S_t x_t) - Max\{\varphi(k)S_t x_t(1+\gamma) - K, N(S_t x_t, t+1)\} \tag{7.7}$$

and with $G_{t+1}(S_t x_t) = \left\{g_{t+1}(S_t z),\ \mathrm{z} \le x_t,\ 0\ \ for\ \mathrm{z} > x_t\right\}$.

Note that in the above expressions the function $N(S_t, t)$ has the natural interpretation of the continuation value of the option. The Proposition 3 lower bound is formulated in general terms and converges to the obvious limit of the American option with a transaction cost-adjusted price of the underlying asset as in Proposition 2 if we set the dividend yield per period equal to $\gamma\Delta t$ and replace the instantaneous mean of the ex-dividend stock return z in (3.12) by $(\mu-\gamma)\Delta t$.

Next we develop a lower bound on American call futures options, in which it is assumed for simplicity that the option and the futures contract mature at the same time. As in CP (2007), we assume that the futures prices are linked to the index by the relation

$$F_t = \alpha_t S_t + \eta_t, \quad t \le T , \tag{7.8}$$

where the random variables $\{\eta_t\}$ have zero mean and variance reflecting the basis risk. The call option bound is presented as a function of α_t and the parameter $\underline{\eta}$, defined an as the lower bound to the random variables $\{\eta_t\}$, assumed observable from past data. The value function of the investor who holds the option is similar to (7.1)-(7.2). We may prove the following result.

<u>Proposition 4</u>: *Under the assumptions of the multiperiod economy stated in Section 2, the tightest lower bound $\underline{C}(F_t, S_t, t)$ on the reservation purchase price of an American call index futures option at any time t prior to option expiration is derived recursively from the expressions:*

For any $t \le T-1$

$$\underline{C}(F_t, S_t, t) = Max\{F_t - K, M(S_t, t)\} \tag{7.9}$$

where the function $M(S_t, t)$ *is defined as follows*

$$M(S_{T-1}, T-1) = \left(\frac{1-k}{1+k}\alpha_{T-1}S_{T-1} + \underline{\eta} - \frac{K}{R}\right)^+$$

$$M(S,t) = \frac{E\left[\max\left\{\varphi(k)\alpha_{t+1}Sz_t + \underline{\eta} - K, M(Sz, t+1)\right\} I(z\text{-}x_t)|S_t = S, z \le \hat{z}_t\right]}{RE[I(z\text{-}x_t)|S_t = S, z \le \hat{z}_t]} + \frac{2kE\left[SzG_{t+1}(Sz)I(z\text{-}x_t)|S_t = S, z \le \hat{z}_t\right]}{(1+k)RE[I(z\text{-}x_t)|S_t = S, z \le \hat{z}_t]},\ t < T-1 \tag{7.10}$$

the function $I(.)$ *is defined in (3.2), the variables* $\hat{z}_t$ *and* x_t *and the function* $g_t(S_t)$ *are defined implicitly by*

$$\frac{E[z|S_t,\ z\le\hat{z}_t]}{(1\text{-}k)E[I(z\text{-}x_t)|S_t,\ z\le\hat{z}_t]} \;=\; R \tag{7.11}$$

$$g_t(S_t)=\frac{Max\{\varphi(k)\alpha_{t+1}S_t\hat{z}_t+\underline{\eta}-K,N(S_t\hat{z}_t,t+1)\}-RN(S_t,t)}{\varphi(k)(\hat{z}_t-R)S_t}, \tag{7.12}$$

$$\begin{aligned}&R[\varphi(k)S_tg_t(S_t)-N(S_t,t)]=\\&=\varphi(k)S_tx_tg_{t+1}(S_tx_t)-Max\{\varphi(k)\alpha_{t+1}S_tx_t+\underline{\eta}-K,N(S_tx_t,t+1)\}\end{aligned}, \tag{7.13}$$

and with $G_{t+1}(S_tx_t)=\left\{g_{t+1}(S_tz),\ \mathrm{z}\le x_t,\ 0\ \ for\ \mathrm{z}>x_t\right\}$.

This lower bound may be used in empirical work on American futures options along the lines of Constantinides *et al* (2008).

VIII. <u>The Upper Bound for European Calls</u>

The upper bound on European calls was established in Proposition 1 of CP (2002). It is an expression that is independent of the time partition and for this reason extends without reformulation to continuous time. Nonetheless, it is not an expectation of the option payoff under a martingale probability and for this reason it does not tend to the Black-Scholes price when the transaction cost parameter is set equal to 0. We show in this section that with a suitable redefinition it can also be made equal to the Black-Scholes price as the transaction costs disappear.

The following result from Proposition 1 of CP (2002) establishes the partition-independent upper bound for the call option, assuming again no dividends prior to option expiration

$$\overline{C}_1(S_t,t)=\frac{(1+k)E[(S_T-K)^+|S_t]}{(1-k)\overline{z}},\tag{8.1}$$

where $\overline{z}=E(z)$. In the absence of transaction costs it was shown in Perrakis (1986) and Ritchken and Kuo (1988) that the following recursive result, a payoff expectation under a martingale probability, is the tightest upper bound for a call option

$$\overline{C}(S_T,T)=(S_T-K)^+,\ \overline{C}(S_t,t)=\frac{E^U[\overline{C}(S_tz,t)|S_t]}{R},\tag{8.2}$$

with the superscript denoting an expectation under a martingale probability $U(z)$ defined as follows, with $P(z)$ denoting the physical probability distribution of the return. Clearly, $E^U(z)=R$.

$$U(z)=\begin{cases}P(z) & \text{with probability } \frac{R-z_{\min}}{\overline{z}-z_{\min}}\\ 1_{z_{\min}} & \text{with probability } \frac{\overline{z}-R}{\overline{z}-z_{\min}}\end{cases}\tag{8.3}$$

Under transaction costs this martingale probability can be easily shown to take the following form, *provided* the following inequality holds:[22] $\overline{z}\geq\frac{1+k}{1-k}R$

$$\overline{U}(z)=\begin{cases}P(z) & \text{with probability } \frac{R\frac{1+k}{1-k}-z_{\min}}{\overline{z}-z_{\min}}\\ 1_{z_{\min}} & \text{with probability } \frac{\overline{z}-R\frac{1+k}{1-k}}{\overline{z}-z_{\min}}\end{cases}.\tag{8.4}$$

We then have the following bound

[22] The proof is available from the authors on request. The result can be shown either with an adaptation of the linear programming approach as in Ritchken (1985) or with the CP (2002) approach of a reservation write price.

$$\overline{C}(S_T,T)=(S_T-K)^+,\ \overline{C}_2(S_t,t)=\frac{E^{\overline{U}}[\overline{C}(S_tz,t)|S_t]}{R}. \tag{8.5}$$

Note that $E^{\overline{U}}(z)=R\frac{1+k}{1-k}$.

It can be shown that $\overline{C}_2(S_t,t)\le\overline{C}_1(S_t,t)\Leftrightarrow\bar{z}\ge\frac{1+k}{1-k}R$. This latter inequality is violated with probability 1 as $\Delta t\to 0$ if the return is given by (3.12), tending to diffusion at the limit. On the other hand, for *k=0* the bound $\overline{C}_2(S_t,t)$ becomes equal to the bound (8.2). In Oancea and Perrakis (2009) that bound was shown to tend to the Black-Scholes price for $\Delta t\to 0$. The following obvious result establishes the convergence of the upper bound to the Black-Scholes value as $\Delta t\to 0$ *and* $k\to 0$:

Proposition 5: *Under proportional transaction costs the European call option is bounded from above by the following upper bound*

$$\overline{C}(S_t,t)=\text{Min}\{\overline{C}_1(S_t,t),\overline{C}_2(S_t,t)\}, \tag{8.6}$$

where the quantities within braces are given by (8.1) and (8.5). For returns given by (3.12) and tending to a lognormal diffusion as $\Delta t\to 0$ *and for k>0 at the limit (8.5) tends to a Black-Scholes-Merton expression with the rate of interest replaced by the instantaneous mean* μ *of the stock. For* $k\to 0$ *as well at the limit (8.6) tends to the Black-Scholes option price.*

IX. Conclusions

In this paper we showed that the CP (2002, 2007) stochastic dominance bounds on call option prices are "natural" generalizations of the BSM price in the presence of proportional transaction costs. Although these bounds were derived in discrete time and with an approach that differed from the dominant arbitrage methodology, they converge to the single arbitrage-derived BSM price whenever the discrete time process tends to diffusion and the transaction cost parameter tends to zero. In this paper we focus on the call option lower bound when proportional transaction costs are present. This bound converges to a reasonably tight BSM expression for realistic values of the transaction cost parameter. The convergence was verified empirically through a novel numerical algorithm. This convergence to a useful bound under realistic trading conditions solves a serious problem in one of the most important models in financial theory, a problem originally identified by Merton (1989) that has not had a satisfactory solution till now.

The derived bounds have obvious applications to competitive market-making situations, since they define limits on quoted bid and ask prices under hypothesized underlying asset dynamics. These can be either diffusion as in Proposition 2, or general empirically-derived distributions as in Proposition 1. Other forms of asset dynamics such as jump diffusion or stochastic volatility present important theoretical and computational challenges. It is relatively easy to formulate discrete time versions of the asset dynamics that converge to the desired continuous time distributions, but Proposition 2 does not hold for jump diffusion and Proposition 1 may not hold for stochastic volatility without major modifications. These cases represent major extensions of the results of this paper.

On the empirical side, the bound derived under general conditions in Proposition 1 may be used in non-parametric stochastic dominance tests of option pricing, as in Constantinides, Jackwerth and Perrakis (2008), or Constantinides *et al.* (2009). In these approaches the underlying asset distributions are extracted from past data, and the numerical algorithm described in Section 4 and Appendix D can be easily adapted for the estimation of the Proposition 1 bound. On the other hand, the width both of the theoretically derived bounds and of the observed bid/ask spreads raises serious questions about the widespread empirical practice of estimating the underlying asset's implied risk neutral distribution from a cross section of observed options market prices. Such questions are amplified by the documented mispricing of both the S&P 500 index options and the S&P 500 index futures options in the aforementioned stochastic dominance tests, and warrant a second look at existing econometric methodology.

Appendix

A. Proof of Proposition 1

The proof follows the general approach of CP (2002), that compares the value function $V\left(v_t, w_t, t\right)$ of a trader who does not hold the option with that of an otherwise identical trader with an open long position in a European call option. Let $J\left(v_t, w_t, S_t, t\right)$ denote this latter value function, defined as follows

$$
\begin{aligned}
&J\left(v_t, w_t, S_t, t\right) \\
&= \max_j E\left[J\left(\left\{v_t - j - k\left|j\right|\right\} R + \left(v_t + \upsilon\right)\gamma z, \left(w_t + j\right), S_t z, t+1\right) \middle| S_t \right]
\end{aligned}
\tag{A.1}
$$

for $t \le T-1$ and

$$
J\left(v_T, w_T, S_T, T\right) = V\left(v_T + (S_T - K)^T, w_T, T\right). \tag{A.2}
$$

Note that the optimal investment decision j_t at time t is in general different from the equivalent decision υ_t of the trader who does not hold the option. Since $J\left(v_t, w_t, S_t, t\right)$ is an increasing function in the portfolio holdings, a lower bound on the reservation purchase price for the call option is a lower bound on the call price C such that

$$
\Delta_t \equiv J\left(v_t - C, w_t, S_t, t\right) - V\left(v_t, w_t, t\right) \ge 0 . \tag{A.3}
$$

It is clear that the following relation is a sufficient condition for (A.3) to hold

$$
\Delta_t \ge J\left(v_t + \varphi(k)\delta_t S_t - C, w_t - \frac{\delta_t S_t}{1+k}, S_t, t\right) - V\left(v_t, w_t, t\right) \ge 0 . \tag{A.4}
$$

As noted, we assume that the dividend yield $\gamma = 0$. In (A.4) $\frac{\delta_t}{1+k}$, with $\delta_t \le 1$, denotes a number of shares such that $\varphi(k)\delta_t S_t - C \ge 0$, that were shorted out of the trader's stockholdings and transferred to the bond account. It will be shown that the tightest lower bound on C satisfying (A.3) is found by setting $C = \underline{C}(S_t, t)$ and $\delta_t = g_t(S_t)$. The following auxiliary results are needed for the proof.

<u>Lemma A.1</u>: Let $\Phi(S_t z, t+1)$ denote any monotone increasing function. Then the function $\phi(S_t, x_t, t) \equiv E_x[\Phi(S_t z, t+1) | S_t, z \le \hat{z}_t]$, with the subscript x denoting an expectation over the distribution given by (4.1) and with $\hat{z}_t$ given by (3.7), is maximized in x_t whenever x_t solves the equation $\Phi(S_t x_t, t+1) = E_x[\Phi(S_t z, t+1) | S_t, z \le \hat{z}_t]$.

<u>Proof</u>: Differentiating $\phi(S_t, x_t, t)$ with respect to x_t and taking into account (3.7) we find that the derivative is proportional to the quantity $E_x[\Phi(S_t z, t+1) | S_t, z \le \hat{z}_t] - \Phi(S_t x_t, t+1)$. For $x_t = z_{\min}$ this quantity is obviously positive, while for $x_t = \hat{z}_t$ it becomes negative. Hence, there exists a unique value of x_t solving the equation $\Phi(S_t x_t, t+1) = E_x[\Phi(S_t z, t+1) | S_t, z \le \hat{z}_t]$, and to the left (right) of this value $\phi(S_t, x_t, t)$ is increasing (decreasing), implying that the solution of the equation defines the unique maximum of $\phi(S_t, x_t, t)$, QED.

<u>Lemma A.2</u>: Define the function $H(S_t, t) \equiv \varphi(k) g_t(S_t) S_t - \underline{C}(S_t, t)$. Then we have:

a) $\overline{H}(S_t, t) \equiv Max_{x_t} \{E_x[\varphi(k) g_{t+1}(S_t z) S_t z - \underline{C}(S_t z, t+1) | S_t, z \le \hat{z}_t]\} = H(S_t, t)$.

b) $H(S_t, t)$ is an increasing function of S_t.

Proof: We start from (b), using induction. (b) can be easily seen to hold at $T-1$, since for both values of $\underline{C}(S_{T-1}, T-1)$ in the RHS of (3.3) $H(S_{T-1}, T-1)$ is either equal to 0 or is increasing. Suppose now that (b) holds at $t+1$. Then $H(S_t z, t+1)$ is increasing in $S_t z$ and for any given $(x_t, \hat{z}_t)$ satisfying (3.7) the function $\underline{H}(S_t, x_t, t) \equiv E_x[\varphi(k) g_{t+1}(S_t z) S_t z - \underline{C}(S_t z, t+1) | S_t, z \le \hat{z}_t]$ is increasing in S_t. Similarly, $\overline{H}(S_t, t)$ is also increasing as the maximum of a set of increasing functions $\underline{H}(S_t, x_t, t)$. By Lemma A.1 and equation (3.9), however, both $\overline{H}(S_t, t)$ and $H(S_t, t)$ are equal to $H(S_t x_t, t+1)$ for all S_t, thus proving (a) and completing the proof of (b), QED.

Define now the following function

$$\hat{\underline{C}}(S_{T-1}, T-1) = \underline{C}(S_{T-1}, T-1),$$

$$\hat{\underline{C}}(S_t, x_t, t) = \frac{E\left[\hat{\underline{C}}(S_t \frac{z}{(1-k)E(I)}, t+1) | z \le \hat{z}_t\right]}{(1-k)RE(I)}, \tag{A.5}$$

$$\hat{\underline{C}}(S_t, t) \equiv Max_{x_t} \{\hat{\underline{C}}(S_t, x_t, t)\},$$

$(x_t, \hat{z}_t)$ given by (3.7).

We can now prove the following result on the form of the call lower bound function $\underline{C}(S_t, t)$ given by (3.3)-(3.9).

Lemma A.3: The call option lower bound (3.3)-(3.9) has the following properties:

a) $\underline{C}(S_t, t)$ is an increasing function of S_t.

b) $\underline{C}(S_t, t) \ge \hat{\underline{C}}(S_t, t)$ for all S_t, t.

c) $\hat{\underline{C}}(S_t, t)$ is convex.

d) $\lim_{S_t \to \infty} \hat{\underline{C}}(S_t, t) = \lim_{S_t \to \infty} \underline{C}(S_t, t) = \varphi(k) S_t - \frac{K}{R^{T-t}}$

Proof: (a) can be easily shown to be true by induction. It obviously holds at *T-1*. By Lemma A.2, Lemma 1 and the induction hypothesis it is clear from (3.6) that $\underline{C}(S_t, t)$ is equal to the maximum of the sum of the expectations of two increasing functions, implying that it also holds at t. Similarly, (c) can also be shown easily to be true by induction. It holds at *T-1*, while at *t* $\hat{\underline{C}}(S_t, x_t, t)$ is obviously convex by (A.5) and the induction hypothesis. (c) then holds at *t* since $\hat{\underline{C}}(S_t, t)$ is the maximum of a set of convex functions. To prove (b) we use again induction and we observe from Lemma A.2 that $\varphi(k) g_t(S_t) S_t \geq \underline{C}(S_t, t)$. This, however, implies that $\frac{\beta(k) g_t(S_t) S_t}{1+k} \geq \frac{\beta(k) \underline{C}(S_t, t)}{1-k}$. We now use this last relation and the induction hypothesis to replace in the integrals in the RHS of (3.6) $\underline{C}(S_t z, t+1)$ by $\hat{\underline{C}}(S_t z, t+1)$ and $\frac{g_{t+1}\left(S_{t+1} z\right) S_t z}{1+k}$ by $\frac{\hat{\underline{C}}(S_t z, t+1)}{1-k}$, both smaller quantities by the induction hypothesis and Lemma A.2. We then have $\underline{C}(S_t, t) \geq \hat{\underline{C}}(S_t, x_t, t)$ for all x_t, and by Lemma 1 (b) holds at *t* as well. Last, part (d) follows directly by induction from (3.3)-(3.9) and (A.5), QED.

We may now proceed with the main body of the proof of Proposition 1. We use induction to prove the joint hypothesis that (3.3)-(3.9) define a lower bound on the reservation purchase price C and that (A.4) holds at *t* for $\delta_t = g_t\left(S_t\right)$ and for $C = \underline{C}(S_t, t)$. At *T*-1 it can be easily seen that both parts of the hypothesis hold. Suppose now that they hold at *t+1* and consider (A.3)-(A.4) at *t*. We have, from (A.1)

$$\Delta_t \geq \max_j E\left[J\left(\left\{ v_t - j - k\left|j\right| + \varphi(k)\delta_t S_t - C \right\} R, \left(w_t + j - \frac{\delta_t S_t}{1+k} \right) z, S_t z, t+1 \right) \left| S_t \right. \right] - V\left(v_t, w_t, t\right) \geq$$

$$\geq \max_j E\left[J\left(\left\{ v_t - \upsilon_t - k\left|\upsilon_t\right| + \varphi(k)\delta_t S_t - C \right\} R, \left(w_t + \upsilon_t - \frac{\delta_t S_t}{1+k} \right) z, S_t z, t+1 \right) \left| S_t \right. \right] - V\left(v_t, w_t, t\right) \geq 0$$

(A.6)

In (A.6) we have used the fact that the optimal portfolio revision for the trader who does not hold the option may be suboptimal for the option holder. Since by the induction hypothesis we know that (A.3)-(A.4) hold at *t+1*, we may write

$$\Delta_t \geq E[V\left(\begin{array}{l} v_t' R + (\varphi(k)\delta_t S_t - C)R - \varphi(k) g_{t+1} S_{t+1} z + \underline{C}(S_t z, t+1), \\ \left(w_t' z + \frac{g_{t+1} S_{t+1} z}{1+k} - \frac{\delta_t S_t z}{1+k} \right), t+1 \end{array} \right) \left| S_t \right.] - V\left(v_t, w_t, t\right) \geq 0 .$$

(A.7)

Consider now the term

$$N(\delta_t, C, S_t, z) \equiv (\varphi(k)\delta_t S_t - C)R - \varphi(k) g_{t+1} S_{t+1} z + \underline{C}(S_t z, t+1) \qquad \text{(A.8)}$$

in the RHS of (A.7). By Lemma A.2 this term is decreasing in the return z as it varies within the interval $[z_{\min}, z_{\max}]$. Accordingly, for appropriate choices of the parameters (δ_t, C) there exists a value $z' \in (z_{\min}, z_{\max})$ such that $N(\delta_t, C, S_t, z') = 0$ and $N(\delta_t, C, S_t, z) > (<) 0$ for $z < (>) z'$. We then form the function

$$O(\delta_t, C, S_t, z) = \frac{N(\delta_t, C, S_t, z)}{1+k} + \frac{g_{t+1} S_t z}{1+k} - \frac{\delta_t S_t z}{1+k} =$$

$$= \frac{(\varphi(k)\delta_t S_t - C)R}{1+k} + \frac{\beta(k) g_{t+1} S_t z}{1+k} + \frac{\underline{C}(S_t z, t+1)}{1+k} - \frac{\delta_t S_t z}{1+k}, \; z \leq z' ,$$

(A.9)

$$O(\delta_t, C, S_t, z) = \frac{N(\delta_t, C, S_t, z)}{1-k} + \frac{g_{t+1} S_t z}{1+k} - \frac{\delta_t S_t z}{1+k} =$$

$$= \frac{(\varphi(k)\delta_t S_t - C)R}{1-k} + \frac{\underline{C}(S_t z, t+1)}{1-k} - \frac{\delta_t S_t z}{1+k},\ z > z'$$

The function $O(\delta_t, C, S_t, z)$ represents the efficient transfer of amounts from the bond to the stock account in the middle part of (A.7), which is greater than or equal to the middle part of (A.10) below. It then suffices to show the following result

$$\Delta_t \geq E[V\left(v_t' R, w_t' z + O(\delta_t, C, S_t, z), t+1\right) \middle| S_t] - V\left(v_t, w_t, t\right) \geq 0\,. \tag{A.10}$$

Replacing now $V\left(v_t, w_t, t\right) = E[V(v_t' R, w_t' z, t+1) | S_t]$ into (A.10) we note that by the concavity of $V\left(v_t, w_t, t\right)$ it suffices to show that

$$\Delta_t \geq E[V_w O(\delta_t, C, S_t, z) | S_t] \geq 0\,. \tag{A.11}$$

In (A.11) $V_w \equiv \frac{\partial V}{\partial w}$, evaluated at the points $\left(v_t' R, w_t' z + O(\delta_t, C, S_t, z)\right)$.The function $O(\delta_t, C, S_t, z)$, in addition to the zero that it has at $z = z'$, also has potentially another zero at some value $z = z'' > z'$ for suitable values of the parameters (δ_t, C). To see this note that $O(\delta_t, C, S_t, z)$ is negative in an open neighborhood to the right of $z = z'$ and by Lemma A.3 decreases if $\underline{C}(S_t z, t+1)$ is replaced by $\hat{\underline{C}}(S_t z, t+1)$ in that neighborhood. This latter function is, however, convex in z, implying that it becomes increasing for sufficiently small values of δ_t and $\varphi(k)\delta_t S_t - C$. This value z'', if it exists within the support $z \in (z', z_{\max}]$, solves the equation

$$\frac{(\varphi(k)\delta_t S_t - C)R}{1-k} + \frac{\underline{C}(S_t z, t+1)}{1-k} - \frac{\delta_t S_t z}{1+k} = 0 \, . \tag{A.12}$$

Let $z^* \equiv Min\{z'', z_{\max}\}$ and observe that by concavity V_w is a decreasing function[23] of z. Similarly, $O(\delta_t, C, S_t, z) > 0$ for $z \in [z_{\min}, z']$ and $O(\delta_t, C, S_t, z) < 0$ for $z \in (z', z^*]$. We thus have, for $V_w(z')$ denoting the marginal value function evaluated at z',

$$\Delta_t \geq E[V_w O(\delta_t, C, S_t, z) \big| S_t] \geq V_w(z') E[O(\delta_t, C, S_t, z) \big| S_t, z \leq z^*] \geq 0 \, . \tag{A.13}$$

From (A.8)-(A.9) we see that $E[O(\delta_t, C, S_t, z) \big| S_t, z \leq z^*] \geq 0$ is equivalent to the following lower bound on the option price C

$$\underline{C}(S_t, \delta_t, t) = \frac{E\left[\underline{C}(S_t z, t+1) I(z - z') \mid S_t, z \leq z^*\right]}{RE\left[I(z - z') \mid S_t, z \leq z^*\right]} +$$

$$\frac{\beta(k) S_t E\left[G_{t+1}(S_t z) I(z - z') S_t z \mid S_t, z \leq z^*\right]}{RE\left[I(z - z') \mid S_t, z \leq z^*\right]} + . \tag{A.14}$$

$$\delta_t S_t \left[(1-k) - \frac{E[z \big| z \leq z^*]}{RE\left[I(z - z') \mid S_t, z \leq z^*\right]} \right]$$

In (A.14) the two key values z', z'' are found from the equations (A.12) and $N(\delta_t, C, S_t, z') = 0$ for given (δ_t, C). Maximizing now the RHS of (A.14) with respect to δ_t we find that the maximum occurs when (3.7) is satisfied. The optimal δ_t is equal to $g_t(S_t)$ as given by (3.8) and the resulting maximum lower bound on C is equal to (3.6).

[23] This property was termed the monotonicity condition in CP (2002, 2007). It requires a relatively "small" investment in the option relative to the stockholdings w_t. For a fuller discussion of monotonicity see CP (2007, pp. 80-84).

We have thus shown that (3.6)-(3.9) hold at *t* as well, and that the optimal values satisfy (A.4). This completes the proof, QED.

B. Proof of Proposition 2

We prove Proposition 2 in two steps. First we prove that (3.14) holds for the process $\overline{X}_{t+\Delta t}$, which thus tends to a diffusion. Then we show that in (3.16) the limit is equal to σ^2. Since by (4.4) and (4.5) the option value is the recursive discounted expectation, under a process that is by construction risk neutral, of a terminal payoff given by (3.3) and equal for $\Delta t \to 0$ to $((\varphi(k)S_{T-1} - K)^+$, by the definition of weak convergence the limit is the Black-Scholes value for a stock price multiplied by the roundtrip transaction cost as in Proposition 2.

From (4.3) it is clear that to prove that $\overline{X}_{t+\Delta t}$ satisfies (3.14) it is sufficient to show that $\overline{X}^1_{t+\Delta t}$ satisfies it. We use the approach introduced by Merton (1982) and adapted by Oancea and Perrakis (2009). The transition probability is equal to

$$\frac{f_x(\varepsilon)}{F(\overline{\varepsilon})} = \frac{I(\varepsilon - \varepsilon_x)f(\varepsilon)}{E(I)F(\overline{\varepsilon})} \equiv dF_x(\varepsilon;\overline{\varepsilon}), \tag{B.1}$$

and let $Q_t(\delta)$ the conditional probability that $|\overline{X}_{t+\Delta t} - X_t| > \delta$, given the information available at time t, with $X_t = 1$. Since ε is bounded, define $\breve{\varepsilon} = \max|\varepsilon| = \max(|\varepsilon_{\min}|, |\overline{\varepsilon}|)$. For any $\delta > 0$, define $h(\delta)$ as the solution of the equation

$$\delta = \mu h + \sigma\breve{\varepsilon}\sqrt{h}. \tag{B.2}$$

This equation admits a positive solution

$$\sqrt{h} = \frac{-\sigma\breve{\varepsilon} + \sqrt{\sigma^2\breve{\varepsilon}^2 + 4\mu\delta}}{\mu}. \tag{B.3}$$

For any $\Delta t < h(\delta)$ and for any possible $X_{t+\Delta t}$,

$$| X_{t+\Delta t} - X_t | = | \mu\,\Delta t + \sigma\varepsilon\sqrt{\Delta t}\,| < \mu h + \sigma\breve{\varepsilon}\sqrt{h} = \delta \tag{B.4}$$

so that for any $\varepsilon_x(S_t, t)$ we have

$$Q_t(\delta) = \Pr\left(| X_{t+\Delta t} - X_t |\right) > \delta \equiv 0 \text{ whenever } \Delta t < h \tag{B.5}$$

and hence $\lim_{\Delta t \to 0} \frac{1}{\Delta t} Q_t(\delta) = 0$, implying that (3.14) holds. Hence, the limit of the stock return process for ε distributed according to (B.1) is a diffusion of the form

$$\frac{dS_t}{S_t} = \mu(S_t, k, x)dt + \sigma(S_t, k, x)dW\,. \tag{B.6}$$

Next we seek to find the parameters $\mu(S_t, k, x)$, $\sigma(S_t, k, x)$ of this diffusion by applying (3.14) and (3.16). From (4.6) we get

$$\lim_{\substack{|X_{t+\Delta t} - X_t| < \delta \\ \Delta t \to 0}} \frac{1}{\Delta t}\left[\frac{1 + \mu\Delta t + \sigma E\left[\varepsilon \middle| \varepsilon \le \bar{\varepsilon}\right]\sqrt{\Delta t}}{(1-k)E(I)} - 1\right] = r\,, \tag{B.7}$$

implying that the process $\overline{X}_{t+\Delta t}$ as given by (4.3a) is by construction risk neutral, and $\mu(S_t, k, x) = r$. It remains, therefore, to evaluate $\sigma^2(S_t k, x)$ in (B.6) by applying (3.16).

We first rewrite (4.6) as follows

$$A(\Delta t)+\frac{\sigma E[\varepsilon\,|\,\varepsilon\le\overline{\varepsilon}]\sqrt{\Delta t}}{(1-k)E(I)}=\left[r-\frac{\mu}{(1-k)E(I)}\right]\Delta t,\ \text{where } A(\Delta t)\equiv\frac{1}{(1-k)E(I)}-1 \qquad \text{(B.8)}$$

We also rewrite (3.16), if $E(\overline{X}_{t+\Delta t})$ denotes the expectation given by (4.6) and neglecting the terms $o(\Delta t)$,

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{\|y-x\|<\delta}(Y-E(\overline{X}_{t+\Delta t})+E(\overline{X}_{t+\Delta t})-X)^2 Q_{\Delta t}(x,dy)=\sigma^2\lim_{\substack{\Delta t\to 0\\ \|y-x\|<\delta}}\frac{E[\varepsilon^2\,|\,\varepsilon\le\overline{\varepsilon}]-(E[\varepsilon\,|\,\varepsilon\le\overline{\varepsilon}])^2}{[(1-k)E(I)]^2}$$
(B.9)

To evaluate the limit in the RHS of (B.9) we first prove the following result.

<u>Lemma B.1</u>: We have

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta\\ \Delta t\to 0}}[\frac{\sigma E[\varepsilon\,|\,\varepsilon\le\overline{\varepsilon}]}{(1-k)E(I)}=\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta\\ \Delta t\to 0}}\sigma E[\varepsilon\,|\,\varepsilon\le\overline{\varepsilon}^*(\Delta t)], \qquad \text{(B.10)}$$

where $\overline{\varepsilon}^*(\Delta t)$ was defined in (4.7)

<u>Proof</u>: Since (B.8) shows that $A(\Delta t)$ is at most $O(\sqrt{\Delta t})$, we have, using the definition of $A(\Delta t)$

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta\\ \Delta t\to 0}}A(\Delta t)=\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta\\ \Delta t\to 0}}\frac{1}{(1-k)E(I)}\frac{2k}{1+k}\frac{G(\varepsilon_x)}{G(\overline{\varepsilon})}=0\,. \qquad \text{(B.11)}$$

Since $\bar{\varepsilon}(\varepsilon_x)$ is a decreasing function (B.11) implies that $\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \frac{G(\varepsilon_x)}{G(\bar{\varepsilon}(\varepsilon_x))} = 0$, which is possible only if

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} (\varepsilon_x) = \varepsilon_{\min} \text{ and } \lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \bar{\varepsilon}(\varepsilon_x) = \lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \bar{\varepsilon}^*(\Delta t)\,. \tag{B.12}$$

Dividing now both sides of (B.8) by $\sqrt{\Delta t}$ and passing to the limit, we observe that

$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \left[\frac{A(\Delta t)}{\sqrt{\Delta t}} + \frac{\sigma E[\varepsilon \,|\, \varepsilon \le \bar{\varepsilon}]}{(1-k)E(I)} \right] = 0$. Since the second term within the limit is bounded the first must be bounded as well, and the limit of the sum is equal to the sum of the limits, implying that $\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \left[\frac{A(\Delta t)}{\sqrt{\Delta t}} \right] = \Lambda \ge 0$ and by (4.6) that the limit of the second term satisfies (B.10), QED.

To find the limit in the RHS of (B.9) we consider the definition of $\bar{\varepsilon}^*(\Delta t)$ in (4.7). The following result, whose proof is an alternative to the one of Proposition 2 in Oancea and Perrakis (2009), establishes clearly that the RHS of (B.9) tends to σ^2 and, thus, completes the proof of Proposition 2.

<u>Lemma B.2</u>: We have

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \bar{\varepsilon}^*(\Delta t) = \varepsilon_{\max} \text{ and, thus, } \lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} E[\varepsilon \,|\, \varepsilon \le \bar{\varepsilon}^*(\Delta t)] = E[\varepsilon] = 0\,. \tag{B.13}$$

<u>Proof</u>: From (4.7) we have $E[\varepsilon \,|\, \varepsilon \le \bar{\varepsilon}^*(\Delta t)] = -\frac{(\mu - r)}{\sigma}\sqrt{\Delta t}$, implying that for any $\Delta t > 0$ we have

$$\frac{d(E[\varepsilon \,|\, \varepsilon \le \bar{\varepsilon}^*(\Delta t)])}{d(\Delta t)} = \frac{d(E[\varepsilon \,|\, \varepsilon \le \bar{\varepsilon}^*(\Delta t)])}{d(\bar{\varepsilon}^*(\Delta t))} \frac{d(\bar{\varepsilon}^*(\Delta t))}{d(\Delta t)} < 0\,. \tag{B.14}$$

The first derivative in the RHS of (B.14) is clearly positive, which implies that the second one must be negative, thus proving the first part of (B.13). The first part of Lemma B.2 also implies that

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} E[\varepsilon^2 \,|\, \varepsilon \le \bar{\varepsilon}^* (\Delta t)] = 1 \ , \tag{B.15}$$

The second part of the Lemma follows automatically from the first. Together, they imply that the RHS of (B.9) tends to σ^2 and $\sigma(S_t, k, x) = \sigma$, QED.

C. Proof of Lemmas 1 and 2

By applying the definition of the *I*-function (3.1) to (3.6), taking expectations in the integral form and simplifying we get:

$$\underline{C}(S_t,t) = \frac{(1+k)\int_{z_{\min}}^{\hat{z}_t} \underline{C}(S_t z) f(z)\, dz - 2k\int_{z_{\min}}^{x_t} \underline{C}(S_t z) f(z)\, dz + 2k\varphi(k) S_t \int_{z_{\min}}^{x_t} g_{t+1}(S_t z) z f(z)\, dz}{R\left[(1+k)\int_{z_{\min}}^{\hat{z}_t} f(z)\, dz - 2k\int_{z_{\min}}^{x_t} f(z)\, dz\right]} \tag{C.1}$$

With the use of (3.10), the denominator of (C.1) may be simplified to $(1+k)\int_{z_{\min}}^{\hat{z}_t} z f(z)\, dz$. By denoting by N and D respectively the numerator and denominator of (C.1), it follows:

$$\frac{d\underline{C}(S_t,t)}{dx} = \frac{N'D - ND'}{D^2} = \frac{N'}{D} - \frac{\underline{C}(S_t,t)D'}{D} . \tag{C.2}$$

By equating (C.2) to zero and rearranging, we have the FOC as $\underline{C}(S_t,t) = N'/D'$. From (C.1) we get:

$$N' = (1+k)\hat{z}_t{}'\underline{C}(S_t\hat{z}_t)f(\hat{z}_t) - 2k\underline{C}(S_tx_t)f(x_t) + 2k\varphi(k)S_txg(x_t)f(x_t),$$

(C.3)

and

$$D' = (1+k)\hat{z}_t{}'\hat{z}_tf(\hat{z}_t). \tag{C.4}$$

By substituting for $\hat{z}_t{}'$ from (3.11) and simplifying, we arrive at the following FOC:

$$\underline{C}(S_t,t) = \frac{\underline{C}(S_t\hat{z}_t)}{\hat{z}_t} - \left(\frac{1}{R} - \frac{1}{\hat{z}_t}\right)\left[\varphi(k)g_{t+1}(S_tx_t)S_tx_t - \underline{C}(S_tx_t)\right]. \tag{C.5}$$

The same condition as (B.5) may be derived by substituting for $g_t(S_t)$ from (3.8) into (3.9) and rearranging, which demonstrates that (3.9) is the FOC for maximizing (3.6) or (C.1), QED.

For the proof of Lemma 2, we apply directly the definition (3.8), replacing the returns from (3.12) and the associated expression $\bar{\varepsilon}$ from (4.6).

$$\lim_{\Delta t\to 0} g_t(S_t) = \lim_{\Delta t\to 0}\frac{\underline{C}\left(S_t\left(1+\mu\Delta t+\sigma\bar{\varepsilon}\sqrt{\Delta t}\right),t+\Delta t\right) - \left[1+r\Delta t+o(\Delta t)\right]\underline{C}(S_t,t)}{\varphi(k)\left[(\mu-r)\Delta t+\sigma\bar{\varepsilon}\sqrt{\Delta t}+o(\Delta t)\right]S_t} \tag{C.6}$$

Since both numerator and denominator tend to 0, we take their derivative with respect to $\sqrt{\Delta t}$. By Proposition 2 the first term in the numerator of (C.6) tends to the Black-Scholes-Merton expression with the stock price multiplied by $\varphi(k)$. Hence, the derivative of the numerator of (C.6) becomes equal to the denominator times $N\left(d_1^*\right)$, where $d_1^* = d_1\left(\varphi(k)S_t,.\right)$, QED.

D. Numerical Approach

The numerical procedure described below was designed for a general setup implied by the compact-support process for the underlying return in a discrete-time continuous-state framework. The time variable t denotes the epoch counter, $t = 0,1...T-1,T$, with the corresponding physical time equal to $t\Delta t$ with $\Delta t = T/N$, where N is the time partition and the symbol T is used here to denote the physical time to maturity. In this general setup the underlying return z_t spans $[z^t_{\min}, z^t_{\max}]$ at the epoch t. A natural method to work our computations backward in time is to use recursive numerical integration[24]. To apply this approach we first need to discretize the problem along the state-variable dimension, by equally spacing z_t in each epoch since recursive numerical integration by the Newton-Cotes rules that we use requires equidistant abscissas. A caveat in this step is that the transition to an earlier epoch with the same equidistant spacing as in the present epoch is not easily achieved. We solve this problem by a log-transformation of the state variable, on which we elaborate further on in this section.

The recursive numerical integration is analogous to lattice methods used in the discrete-time discrete-state framework where the expectations are one-step forward realizations of a function of the random variable weighted by the probabilities. In the present setup weights are defined instead as the densities evaluated at equidistant points multiplied by the integration weights, times the integration step. We denote this approach by a 'generalized lattice' or simply 'lattice'. Denote by $I_0\left(h_{t+1}\left(y\right)\right)$ the time-t integral of the $t+1$ function h of a random variable y, given by:

$$I_0\left(h_{t+1}\left(y\right)\right) \cong \sum_{j=0}^{L} \tilde{w}_j \Delta y f\left(y_{\min} + j\Delta y\right) h_{t+1}\left(y_{\min} + j\Delta y\right) \equiv \sum_{j=0}^{L} w_j h_{t+1}\left(y_{\min} + j\Delta y\right), \quad \text{(D.1)}$$

[24] For instance, Andricopoulos at al (2003) used recursive numerical integration to price path-dependent derivatives for the lognormal distribution.

where $\tilde{w}_j$ is the weight for a given integration rule, Δy is the integration step, $f(.)$ is the density function, L is a positive integer satisfying $L\Delta y + y_{\min} \le y_{\max}$, and w_j is the redefined weight. It is clear that functions similar to $I_0\left(h_{t+1}(y)\right)$ will approximate the truncated expectations we need to derive in (3.6) or its equivalent, relation (C.1). Notice that for *L* such that $L\Delta y + y_{\min} = y_{\max}$ in (D.1) we approximate expectations over the full support of *y*, which makes clear the analogy between the recursive numerical integration and discrete-time discrete-state lattice methods with points $y_{\min} + j\Delta y$ replacing the nodes.

The discretized (in log-scale) process of the return of the underlying may be thought of as a recombining lattice method. We space the one-period log-return of the underlying ($\equiv y$) by Δy into *m* increments, where *m* is an odd number[25], with the lowest (highest) increment $y_{\min}$ ($y_{\max}$) satisfying $y_{\min} = \log(z_{\min})$ ($y_{\max} = \log(z_{\max})$). It follows that at the epoch *t* the log-return is spaced by Δy over a segment $[ty_{\min}, ty_{\max}]$ with $t(m-1)+1$ increments; conversely, every state Y_t we consider belongs to the discretized set in this segment. From every Y_t, which, in our notation plays the role of a node[26] in the lattice, *m* states (nodes) spaced by Δy over $[Y_t + y_{\min}, Y_t + y_{\max}]$ may be reached in the subsequent epoch, while going backward in the lattice we may evaluate integrals as in (D.1).

The numerical work for the derivation of the bounds is crucially dependent on the accuracy of the estimation of the key variables $\hat{z}$ and *x*.[27] Even if we limit ourselves to maximizing (B.1) over a set of values of *x* whose logarithms fall exactly on the increments of *y*, the corresponding set of $\hat{z}$'s will in general fall between the nodes.

---

[25] With the exception of the binomial model, an odd number of nodes for the one-period return process of the underlying is necessary for the lattice to recombine.

[26] In principle, we should index Y_t since we use this symbol to also denote the log-transformed (continuous) state variable at time *t*. However, to simplify the notation we skip this indexing while, in what follows, making clear to the reader whenever Y_t is used to denote a typical node for the discretized state variable.

[27] In what follows in this section we suppress the time subscripts on $\hat{z}$ and *x* to simplify the notation.

Approximating $\hat{z}$ by the closest node in the grid to its true value would not yield satisfactory results. To circumvent this problem, we use a non-linear interpolating function, the piece-wise Hermite polynomials.[28] This function, as opposed to the perhaps more widely used splines, has the desirable property of preserving the monotonicity of the data. In our numerical work we use a Matlab function *pchip*, which applies the Fritsch and Carlson (1980) algorithm.

To delineate the search domain for the inversely varying $\hat{z}$ and *x* we let $x_{\max}$ denote the maximum feasible *x*, i.e. the one corresponding to $\hat{z} = R$ in (3.10) and let $\hat{z}_{\max}$ denote the maximum feasible $\hat{z}$, i.e. the one corresponding to $x = z_{\min}$ in (3.10). We then define two numerical integrals for a given node characterized by the log-return Y_t:

$$I_1(L_i) = \sum_{j=0}^{L_i} w_j \underline{C}(y_{\min} + j\Delta y),\ L_i = 1 \ldots L_{\hat{z}}$$

and ,

$$I_2(L_i) = \sum_{j=0}^{L_i} w_j \left[\varphi(k)\exp(Y_t + y_{\min} + j\Delta y)\, g(y_{\min} + j\Delta y) - \underline{C}(y_{\min} + j\Delta y)\right],\ L_i = 1 \ldots L_x \tag{D.2}$$

where the weights w_j are as in (D.1), the time-*t* state argument Y_t were suppressed in $\underline{C}(.)$ and $g(.)$ to simplify the notation, $L_{\bar{\hat{z}}}\Delta y + y_{\min} > \log(\hat{z}_{\max})$ and $L_x \Delta y + y_{\min} > \log(x_{\max})$. These two last conditions ensure that the integrals are computed over sufficiently wide range to interpolate them later on for the variables of interest *x* or $\hat{z}$. It is also apparent that since the minimum value of $\hat{z}_t$ is greater than or equal to the maximum value of *x*, we need to use $L_{\bar{\hat{z}}} + 1$ weights w_j in our numerical work.

The following steps describe our numerical algorithm:

---

[28] See Fritsch and Carlson (1980). Details of the application of their algorithm to our case are available from the authors on request.

1) Fix a set of pairs $(x_s,\ \hat{z}_s)$, $s = 1...n$ linked by (3.10) and spanning the feasible region for *x*, which is $[z_{\min},\ x_{\max}]$. For practical reasons, we consider equal increments for *x* in the log scale in this set ($\equiv \Delta\log(x)$) and to gain on precision we ensure $\Delta\log(x) < \Delta y$. It is apparent that, in general the corresponding points $\hat{z}_s$ will not be equally spaced even in the log-scale.
2) Derive the denominator ($\equiv D_s$) of (C.1) for every pair $(x_s,\ \hat{z}_s)$.
3) For a given node at time *t* characterized by the log-return Y_t derive the $I_1(L_i)$ and $I_2(L_i)$ from (D.2) for every applicable L_i. Use these values as the inputs for the interpolating function. Interpolate for every pair $(x_s,\ \hat{z}_s)$; denote the interpolated results by $I_{1,s}$ and $I_{2,s}$, $s = 1...n$. Now, for a given node we have candidate solutions ($\equiv C_s$) that we write as: $C_s = \left[(1+k) I_{1,s} + 2kI_{2,s}\right] / D_s$, $s = 1...n$.
4) The maximal C_s ($\equiv C_{s*}$) becomes the lower bound for a given node. The estimation of the *g*-function follows from (3.8) with $\underline{C}(S_t\hat{z}, t+1)$ interpolated in log-scale as $\underline{C}(Y_t + \log(\hat{z}_{s*}), t+1)$.
5) Repeat 3 and 4 for every node at time *t*.
6) Proceed to the previous epoch till $t = 0$ is reached.

We use the above algorithm for any epoch $t \le T-2$ while we use the lower bound (3.3) on $\underline{C}(Y_{T-1}, T-1)$ at $t = T-1$ with the corresponding result (3.5) at this time epoch for the *g*-function, i.e. we set $g_{T-1} = 1$ for the nodes with $\underline{C}(Y_{T-1}, T-1) > 0$, 0 otherwise.

As noted earlier, for our generalized lattice we need to perform a log-transformation of the state variable, which necessitates adjusting the form of the density function. For $y \equiv \log(z)$, with *z* distributed uniformly as in (5.1), we have the following density function:

$$f(y) = \begin{cases} \exp(y)/2\sqrt{3}\sigma\sqrt{\Delta t},\ y \in \left[\log\left(z_{\min}\right), \log\left(z_{\max}\right)\right] \\ 0 \text{ otherwise} \end{cases} . \quad \text{(D.3)}$$

Under the log-transformation, we have an additive process for the logarithmic return for which a grid with equal increments suitable for the Newton-Cotes numerical integration can be easily constructed. Now the weights we use to approximate truncated expectations in our lattice become:

$$w_j = \tilde{w}_j \Delta y \exp\left(y_{\min} + j\Delta y\right)/2\sqrt{3}\sigma\sqrt{\Delta t},\ j = 0 \ldots L_{\hat{z}} . \quad \text{(D.4)}$$

In our numerical work we use the tree size $m = 251$ with $n = 250$ candidate values for x located in $\left[z_{\min}, x_{\max}\right]$. Observe, however, that for a time partition N of, say, 100 it will be a formidable task to deal with the resulting number of nodes, given also the fact that for each node we need to compute $2n$ integrals. A numerical technique suitable for the task at hand is the (discrete) Fast Fourier Transform (FFT). We illustrate the technique by presenting a formula which yields integrals of the type I_1 for all nodes at the epoch t for a given L_i:

$$I_1^M\left(L_i\right) = \text{IFFT}\left[\text{FFT}\left(\underline{C}^M\left(t+1\right)\right) \text{e } \text{FFT}\left(MW_{L_i}^M\right)\right], \quad \text{(D.5)}$$

where M denotes the number of nodes at the epoch $t+1$, $M = \left(t+1\right)\left(m-1\right)+1$, FFT denotes a Fast Fourier Transform, IFFT denotes the inverse of FFT, $I_1^M\left(L_i\right)$ and $\underline{C}^M\left(t+1\right)$ are vectors of length M with the latter vector representing the lower bound values at all nodes at the epoch t + 1, e denotes Hadamard product,[29] $W_{L_i}^M$ is a vector $[w_0 \ldots w_{L_i}]'$ padded down with zeros to the length M. The first and last $\left(m-1\right)/2$ entries

[29] For two nx1 vectors a and b, a e b is an nx1 vector containing elementwise products $a_i b_i$.

to the vector $I_1^M\left(L_i\right)$ should be discarded since FFT (IFFT) applied to a vector yields a vector of the same length. $I_2^M\left(L_i\right)$ easily follows by an appropriate substitution in (D.5). To derive the integration weights $\tilde{w}_j$, we apply the Newton-Cotes composite rules,[30] as in equation (D.1). As the base, we use the five-point rule; however, whenever an integer L_i may not fit to this rule, we pad the shortest possible lower-point rule to our base.

E. Proof of Proposition 4

We first show that (7.3) is a lower bound $\underline{C}\left(S_{T-1}, T-1\right)$ on $C\left(S_{T-1}, T-1\right)$. Suppose first that $S_{T-1}(1+\gamma)-K$ is the largest term within braces in the RHS of (7.3). Then replacing in (7.1) we can easily see that

$$
\begin{aligned}
&J\left(x_{T-1}-\underline{C}, y_{T-1}, S_{T-1}, T-1\right) \geq V\left(x_{T-1}+S_{T-1}(1+\gamma)-K-\underline{C}, y_{T-1}, T-1\right) \\
&\geq V\left(x_{T-1}, y_{T-1}, T-1\right)
\end{aligned}.
$$

Suppose now that it is the second term that is the largest in (7.3), with $(\frac{1-k}{1+k}S_{T-1}-\frac{K}{R})^+=\underline{C}$. Let $J^E\left(x_{T-1}, y_{T-1}, S_{T-1}, T-1\right)$ denote the value function of an investor with a long open position in a *European* call option with the same characteristics as the one-period American. Clearly, $J\left(x_{T-1}-\underline{C}, y_{T-1}, S_{T-1}, T-1\right) \geq J^E\left(x_{T-1}-\underline{C}, y_{T-1}, S_{T-1}, T-1\right)$ for any $\underline{C}$. Since $(\frac{1-k}{1+k}S_{T-1}-\frac{K}{R})^+$ is a lower bound on a European option, we have $J\left(x_{T-1}-\underline{C}, y_{T-1}, S_{T-1}, T-1\right) \geq V\left(x_{T-1}, y_{T-1}, T-1\right)$ in this case as well.

We use now induction to prove that $\underline{C}\left(S_t, t\right)$ in (7.3) is a lower bound on the value of the call option for $t \leq T-1$. It was already shown that $\underline{C}\left(S_{T-1}, T-1\right)$ is a lower bound on the

---

[30] For Newton-Cotes integration see, for instance, Davis and Rabinowitz (1966), or any other textbook on numerical integration.

call option. Assume now that $\underline{C}\left(S_{t+1}, t+1\right)$ is a lower bound, and consider the case at *t*. If in (7.3) $\underline{C}\left(S_t, t\right)$ is equal to the first term in braces in the RHS then from (7.1) we have $J\left(x_t - \underline{C}, y_t, S_t, t\right) \geq V\left(x_t - \underline{C} + S_t(1+\gamma) - K, y_t, t\right) \geq V\left(x_t, y_t, t\right)$, and $\underline{C}\left(S_t, t\right)$ is a lower bound on the price of a call option. Assume, therefore, that $\underline{C}\left(S_t, t\right) = N(S_t, t)$. We have, for any $\delta_t \in (0,1]$ such that[31] $\varphi(k)\delta_t S_t > N(S_t, t)$

$$
\begin{aligned}
& J\left(x_t - N(S_t,t), y_t, S_t, t\right) \geq J\left(x_t + \varphi(k)\delta_t S_t - N(S_t,t), y_t - \frac{\delta_t S_t}{1+k}, S_t, t \right) \geq \\
& \max_j E\left[J\begin{pmatrix} \left\{x_t + \varphi(k)\delta_t S_t - N(S_t,t) - j - \max\left[k_1 j, -k_2 j\right]\right\} R \\ +(y_t + j - \dfrac{\delta_t S_t}{(1+k)})\gamma z, \left(y_t - \dfrac{\delta_t S_t}{1+k} + j \right) z, S_{t+1}, t+1 \end{pmatrix} | S_t \right] \geq \\
& \max_j E\left[J\begin{pmatrix} \left\{x_t + \varphi(k)\delta_t S_t - N(S_t,t) - j - \max\left[k_1 j, -k_2 j\right]\right\} R + (y_t + j)\gamma z, \\ (y_t + j) z - \dfrac{\delta_t S_t}{(1+k)}(1 + \dfrac{\gamma}{1-k}) z, S_{t+1}, t+1 \end{pmatrix} | S_t \right] \geq \\
& \max_j E\left[J\begin{pmatrix} x_t' R + \left\{\varphi(k)\delta_t S_t - N(S_t,t)\right\} R, \\ y_t' z - \dfrac{\delta_t S_t}{(1+k)}(1 + \dfrac{\gamma}{1-k}) z, S_{t+1}, t+1 \end{pmatrix} | S_t \right]
\end{aligned}
\qquad (E.1)
$$

We show that the RHS of the above relation is greater than or equal than $V\left(x_t, y_t, t\right)$, in which case we have shown that $\underline{C}\left(S_t, t\right)$ is a lower bound on the call option. We have for the RHS of (E.1), using the definition of the lower bound in (7.3) and the induction hypothesis

[31] It can be shown from the definition of $N(S_t, t)$ that such a value exists.

$$\max_j E\left[J\begin{pmatrix} x_t^{'}R + \left\{\varphi(k)\delta_t S_t - N(S_t,t)\right\}R, \\ y_t^{'}z - \frac{\delta_t S_t}{(1+k)}(1+\frac{\gamma}{1-k})z, S_{t+1}, t+1 \end{pmatrix} | S_t \right] \geq$$

$$E\left[V\begin{pmatrix} x_t^{'}R + \left\{\varphi(k)\delta_t S_t - N(S_t,t)\right\}R + \underline{C}(S_{t+1},t+1), \\ y_t^{'}z - \frac{\delta_t S_t}{(1+k)}(1+\frac{\gamma}{1-k})z, t+1 \end{pmatrix} | S_t \right] \geq$$

$$E\left[V\begin{pmatrix} x_t^{'}R + \left\{\varphi(k)\delta_t S_t - N(S_t,t)\right\}R + Max\{\varphi(k)S_{t+1}(1+\gamma) - K, N(S_{t+1},t+1)\}, \\ y_t^{'}z - \frac{\delta_t S_t}{(1+k)}(1+\frac{\gamma}{1-k})z, t+1 \end{pmatrix} | S_t \right] \geq$$

$$E\left[V\begin{pmatrix} x_t^{'}R + \left\{\varphi(k)\delta_t S_t - N(S_t,t)\right\}R + Max\{\varphi(k)S_{t+1}(1+\gamma) - K, N(S_{t+1},t+1)\} - g_{t+1}(S_{t+1})S_{t+1}\varphi(k), \\ y_t^{'}z - \frac{\delta_t S_t}{(1+k)}(1+\frac{\gamma}{1-k})z + \frac{g_{t+1}(S_{t+1})S_{t+1}}{1+k}, t+1 \end{pmatrix} | S_t \right] \tag{E.2}$$

Set now

$$\begin{aligned} &\left\{\varphi(k)\delta_t S_t - N(S_t,t)\right\}R + Max\{\varphi(k)S_{t+1}(1+\gamma) - K, N(S_{t+1},t+1)\} - g_{t+1}(S_{t+1})S_{t+1}\varphi(k) \\ &\equiv H(S_{t+1}) = H(S_t z) \end{aligned}, \tag{E.3}$$

and then from the RHS of (E.2) we have

$$\begin{aligned} &E\left[V\begin{pmatrix} x_t^{'}R + H(S_t z), \\ y_t^{'}z - \frac{\delta_t S_t}{(1+k)}(1+\frac{\gamma}{1-k})z + \frac{g_{t+1}(S_t z)S_t z}{1+k}, t+1 \end{pmatrix} | S_t \right] \geq \\ &E\left[V\begin{pmatrix} x_t^{'}R, H(S_t z)I(H(S_t z)) + \\ y_t^{'}z - \frac{\delta_t S_t}{(1+k)}(1+\frac{\gamma}{1-k})z + \frac{g_{t+1}(S_t z)S_t z}{1+k}, t+1 \end{pmatrix} | S_t \right] \end{aligned}. \tag{E.4}$$

By the concavity of the indirect utility function we now have that the RHS of (E.4) is greater than or equal to

$$
\frac{E\left[\begin{array}{l} V\left(x_t^{'}R, y_t^{'}z, t+1\right)+V_y\{H(S_tz)I(H(S_tz)) \\ -\frac{\delta_t S_t}{(1+k)}(1+\frac{\gamma}{1-k})z+\frac{g_{t+1}(S_tz)S_tz}{1+k}\}|S_t \end{array}\right] \geq V(x_t, y_t, t)+}{E\left[V_y\{H(S_tz)I(H(S_tz))-\frac{\delta_t S_t}{(1+k)}(1+\frac{\gamma}{1-k})z+\frac{g_{t+1}(S_tz)S_tz}{1+k}|S_t\right]} \quad \text{(E.5)}
$$

By the monotonicity condition the RHS of (E.5) is greater than or equal to[32]

$$
\begin{aligned}
&V(x_t, y_t, t)+ \\
&E[V_y|S_t]E\left[H(S_tz)I(H(S_tz))-\frac{\delta_t S_t}{(1+k)}(1+\frac{\gamma}{1-k})z+\frac{g_{t+1}(S_tz)S_tz}{1+k}\middle|z\leq \hat{z}_t\right]
\end{aligned} \quad \text{(E.6)}
$$

(E.6) is nonnegative unless the expectation in brackets is less than or equal to zero. Re-arranging and defining x_t from the equation $H(S_tx_t)=0$ and $\hat{z}_t$ as in the proof of Proposition 1, we have

$$
\begin{aligned}
N(S_t,t) \geq & \frac{E\left[\frac{g_{t+1}(S_tz)S_tz}{1+k}+Max\left\{\varphi(k)S_tz(1+\gamma)-K, N(S_tz,t+1)\right\}\right.}{RE[I(z-x_t)|z\leq \hat{z}_t]} \\
& -\frac{\left.\varphi(k)g_{t+1}(S_tz)S_tz]I(z-x_t)|z\leq \hat{z}_t\right]}{RE[I(z-x_t)|z\leq \hat{z}_t]} \\
& +\delta_t S_t\left\{\varphi(k)-\frac{1}{(1+k)RE[I(z-x_t)|z\leq \hat{z}_t]}\left(1+\frac{\gamma}{1-k}\right)E[z|z\leq \hat{z}_t]\right\}
\end{aligned} \quad \text{(E.7)}
$$

Maximizing the RHS of (E.7) with respect to δ_t and taking into account the definitions of x_t and $\hat{z}_t$ we get (7.4)-(7.7). From (E.7) the lower bound $\underline{C}(S_t,t)$ as given by (7.3) follows immediately if we add the early exercise condition, QED.

[32] We need here to prove the equivalent of Lemma A.2 and Lemma A.3 in the proof of Proposition 1 to verify the existence of x_t and $\hat{z}_t$. The proofs are identical to those of the lemmas and are omitted.

References

Amin, Kaushik I., 1993, "Jump Diffusion Option Valuation in Discrete Time", *Journal of Finance*, 48, 1833-1863.

Amin, Kaushik I., and Victor K. Ng, 1993, "Option Valuation with Systematic Stochastic Volatility", *The Journal of Finance* 48, 881-910.

Andricopoulos, A., M. Widdicks, P. W. Duck and D. P. Newton, 2003, "Universal Option Valuation Using Quadrature Methods", *Journal of Financial Economics* 67, 447-471.

Bates, D. S., 1991, "The Crash of '87: Was it Expected? The Evidence from Option Markets", *Journal of Finance*, 46, 1009-1044.

Black, F. and M. S. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, 637-654.

Bensaid, B., J.-P. Lesne, H. Pages, and J. Scheinkman, 1992, "Derivative Asset Pricing with Transaction Costs," *Mathematical Finance* 2, 63–86.

Bogle, J. C., 2005, "The Mutual Industry 60 Years Later: For Better or Worse?," *Financial Analyst Journal* 61, 15-24.

Boyle, P. P. and T. Vorst, 1992, "Option Replication in Discrete Time with Transaction Costs," *Journal of Finance* 47, 271-293.

Cerny, A., 2004, "Introduction to Fast Fourier Transform in Finance," *Journal of Derivatives* 12, 73-88.

Constantinides, G. M., 1979, "Multiperiod Consumption and Investment Behavior with Convex Transactions Costs," *Management Science* 25, 1127-37.

Constantinides, G. M., 1986, "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy* 94, 842-862.

Constantinides, G. M., M. Czerwonko, J. Jackwerth and S. Perrakis, 2009, "Are Options on Index Futures profitable for Risk Averse Investors? Empirical Evidence", working paper, University of Chicago, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1106962.

Constantinides, G. M., J. Jackwerth and S. Perrakis, 2009, "Mispricing of S&P 500 Index Options", *Review of Financial Studies*, 22, 1247-1277.

Constantinides, G. M. and S. Perrakis, 2002, "Stochastic Dominance Bounds on Derivatives Prices in a Multiperiod Economy with Proportional Transaction Costs," *Journal of Economic Dynamics and Control* 26, 1323-1352.

Constantinides, G. M. and S. Perrakis, 2007, "Stochastic Dominance Bounds on American Option Prices in Markets with Frictions," *Review of Finance* 11, 71-115.

Constantinides, G. M. and T. Zariphopoulou, 1999, "Bounds on Prices of Contingent Claims in an Intertemporal Economy with Proportional Transaction Costs and General Preferences," *Finance and Stochastics* 3, 345-369.

Constantinides, G. M. and T. Zariphopoulou, 2001, "Bounds on Derivative Prices in an Intertemporal Setting with Proportional Transaction Costs and Multiple Securities," *Mathematical Finance* 11, 331-346.

Davis, M. H. A. and V. G. Panas, 1994, "The Writing Price of a European Contingent Claim under Proportional Transaction Costs," *Computational and Applied Mathematics* 13, 115-157.

Davis, M. H. A., V. G. Panas and T. Zariphopoulou, 1993, "European Option Pricing with Transaction Costs," *SIAM Journal of Control and Optimization* 31, 470-493.

Davis, P. J. and P. Rabinowitz, 1966, *Numerical Integration*, Blaisdell Pub. Co.

Fritsch, F. N. and R. E. Carlson, 1980, "Monotone Piecewise Cubic Interpolation," *SIAM Journal of Numerical Analysis* 17, 238-246.

Grannan, E. R. and G. H. Swindle, 1996, "Minimizing Transaction Costs of Option Hedging Strategies," *Mathematical Finance* 6, 341–364.

Hodges, S. D. and A. Neuberger, 1989, "Optimal Replication of Contingent Claims under Transaction Costs," *Review of Futures Markets*, 8, 222-239.

Kocherlakota, N. R., 1996, "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature* 34, 42-71.

Leland, H. E., 1985, "Option Pricing and Replication with Transactions Costs," *Journal of Finance* 40, 1283-1301.

Levy, H., 1985, "Upper and Lower Bounds of Put and Call Option Value: Stochastic Dominance Approach," *Journal of Finance* 40, 1197-1217.

Martellini, L. and P. Priaulet, 2002, "Competing Methods for Option Hedging in the Presence of Transaction Costs," *Journal of Derivatives* 9, 26-38.

Merton, R. C., 1973, "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science* 4, 141-84.

Merton, R. C., 1989, "On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance," *The Geneva Papers on Risk and Insurance* 14, 225-261.

Oancea, I. M., and S. Perrakis, 2009, "Stochastic Dominance Option Pricing in Discrete and Continuous Time: an Alternative Paradigm", working paper, Concordia University, http://ssrn.com/abstract=1360800.

Perrakis, Stylianos, 1986, "Option Bounds in Discrete Time: Extensions and the Pricing of the American Put", *Journal of Business* 59, 119-141.

Perrakis, S., 1988, "Preference-free Option Prices when the Stock Return Can Go Up, Go Down, or Stay the Same," *Advances in Futures and Options Research* 3, 209-235.

Perrakis, S., and J. Lefoll, 2000, "Option Pricing and Replication with Transaction Costs and Dividends," *Journal of Economic Dynamics and Control* 24, 1527-1561.

Perrakis, S. and J. Lefoll, 2004, "The American Put under Transactions Costs," *Journal of Economic Dynamics and Control* 28, 915-936.

Ritchken, P. H., 1985, "On Option Pricing Bounds," *Journal of Finance* 40, 1219-1233.

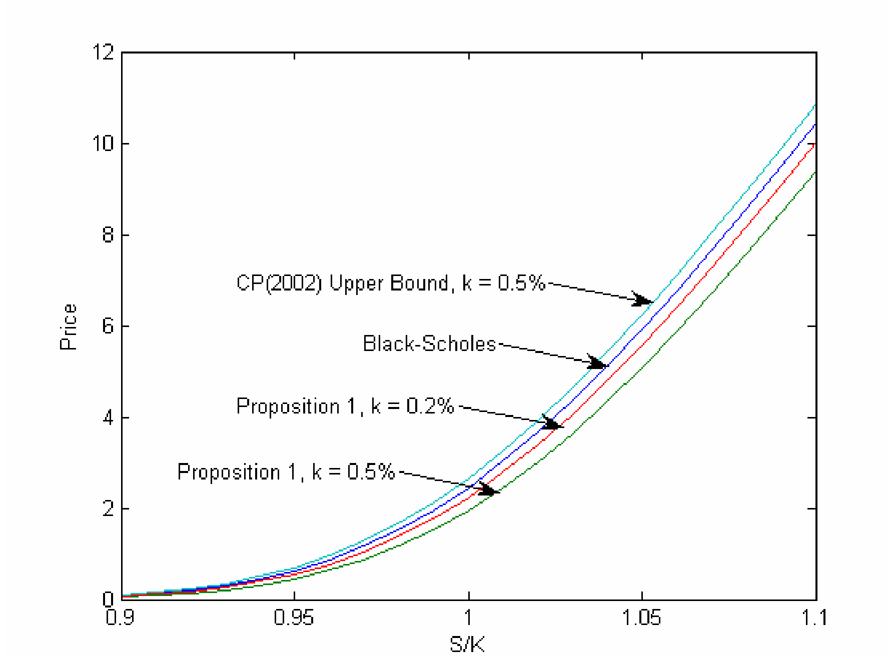
Ritchken, P.H. and S. Kuo, 1988, "Option Bounds with Finite Revision Opportunities," *Journal of Finance* 43, 301-308.

Soner, H. M., S. E. Shreve and J. Cvitanic, 1995, "There is No Nontrivial Hedging Portfolio for Option Pricing with Transaction Costs." *The Annals of Applied Probability* 5, 327-355.

Toft, K. B., 1996, "On the Mean-Variance Tradeoff in Option Replication with Transactions Costs," *Journal of Financial and Quantitative Analysis* 31, 233-263.

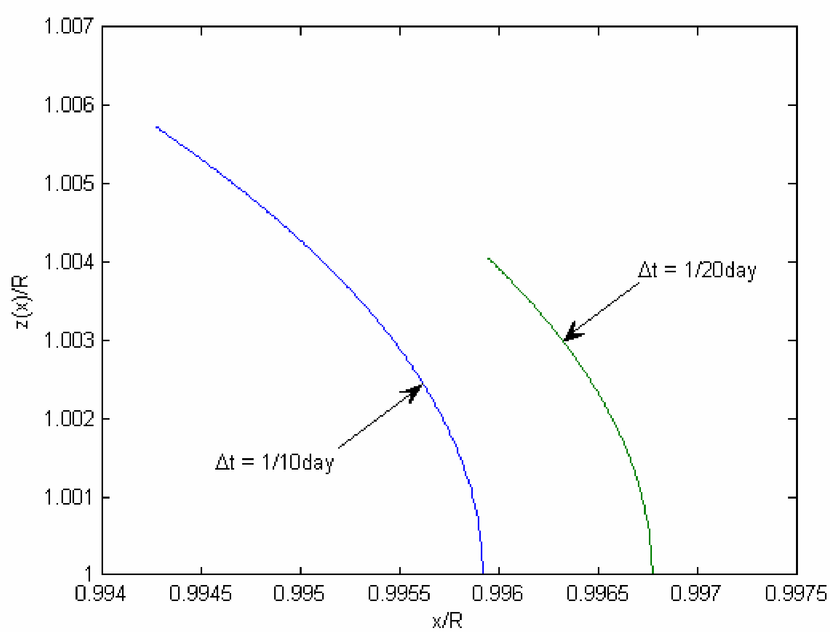
Zakamouline, V., 2006, "European Option Pricing and Hedging with Both Fixed and Proportional Transaction Costs," *Journal of Economic Dynamics and Control* 30, 2006, 1-25.

Figure 1: Continuous-time Limit Results for Proposition 1.



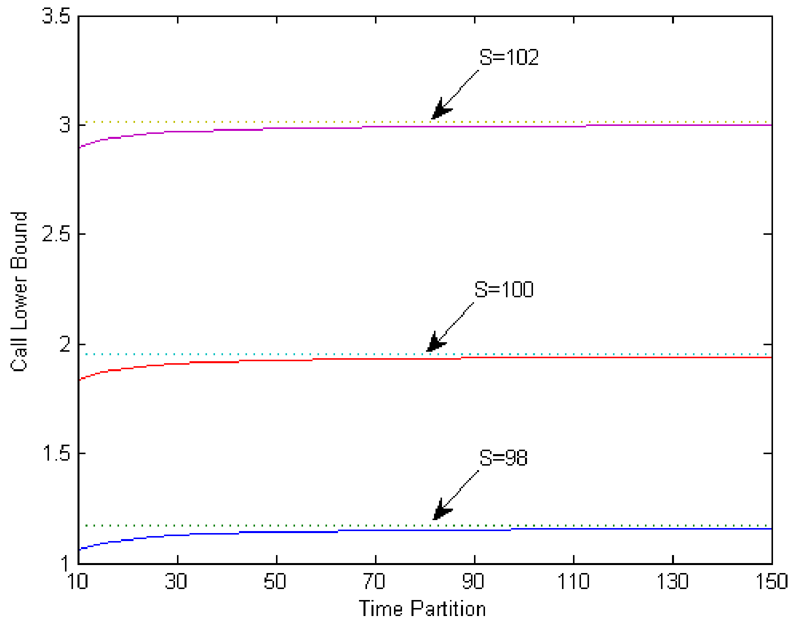
The figure displays the continuous-time limit of the Proposition 1 call lower bound as derived in Proposition 2 for various degrees of moneyness. These results are compared to the CP (2002) call upper bound and to the Black-Scholes-Merton price. The CP (2002) call upper bound for the transaction costs rate $k = 0.2\%$ is not displayed since in the applied scale it would not be distinguishable from the presented bound for $k = 0.5\%$. The parameters are as follows: $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $T = 30$ days.

Figure 2: Behavior of x and $\hat{z}$ in Time Partition



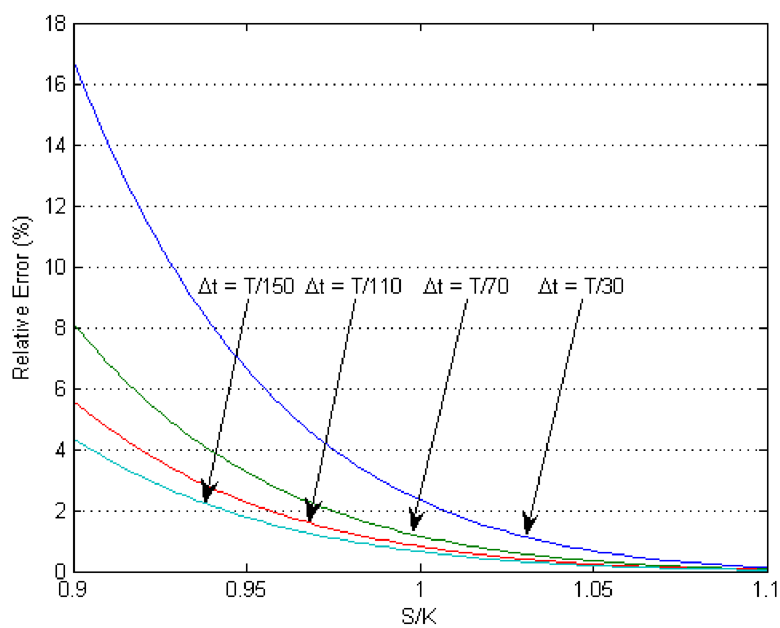
The figure displays the time-partition behavior of the quantities x and $\hat{z}$ as defined by equation (3.10) and derived by equation (5.4) for the uniformly distributed disturbances in (5.1). The displayed quantities were normalized by the riskless return respective to each time partition. The parameters are as follows: $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $k = 0.5\%$.

Figure 3: Convergence of the Proposition 1 Lower Bound to its Continuous-time Limit



The figure displays the convergence behavior of the Proposition 1 lower bound (3.6) to its continuous-time limit given by Proposition 2 and derived for the uniform distribution of the stock returns (5.1). The parameters are as follows: $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $T = 30$, $k = 0.5\%$.

Figure 4: Relative Convergence Errors of the Proposition 1 Lower Bound from its Continuous-time Limit



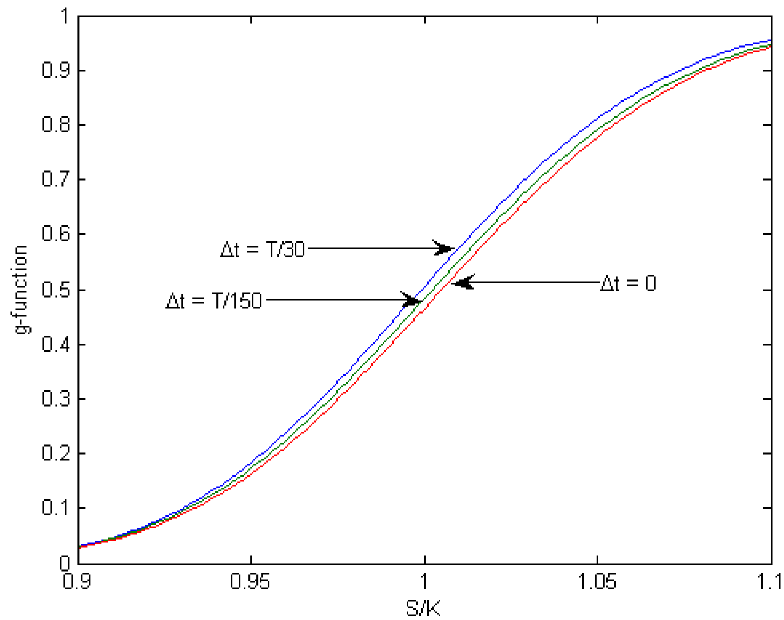
The figure displays the relative convergence errors $1-\underline{C}_5 / BSM\left(\varphi\left(k\right)S,.\right)$ of the Proposition 1 lower bound (3.6) from its continuous-time limit given by Proposition 2 and derived for the uniform distribution of the stock returns (5.1). The parameters are as follows: $K=100$, $\sigma=20\%$, $\mu=8\%$, $r=4\%$, $T=30$, $k=0.5\%$.

Table 1: Convergence of the Proposition 1 Call Lower Bound for a Fixed Number of Time Intervals as a Function of Time to Maturity

S/K	$\underline{C}(S_t)$	$BSM(\varphi(k)S_t)$	$BSM(\varphi(k)S_t)-\underline{C}(S_t)$	$1-\frac{\underline{C}(S_t)}{BSM(\varphi(k)S_t)}$ (%)	$1-\frac{BSM(\varphi(k)S_t)}{BSM(S_t)}$ (%)
			A: T = 30 days		
0.9	0.050	0.052	0.002	4.25	36.46
1	1.942	1.954	0.012	0.62	20.28
1.1	9.388	9.391	0.003	0.03	9.98
			B: T = 60 days		
0.9	0.309	0.318	0.009	2.84	23.31
1	3.020	3.040	0.020	0.67	14.64
1.1	10.093	10.102	0.010	0.10	8.81
			C: T = 120 days		
0.9	1.072	1.096	0.024	2.19	14.92
1	4.643	4.677	0.034	0.73	10.49
1.1	11.476	11.498	0.022	0.19	7.33
			D: T = 240 days		
0.9	2.708	2.761	0.053	1.90	9.70
1	7.119	7.179	0.060	0.83	7.49
1.1	13.886	13.931	0.046	0.33	5.83

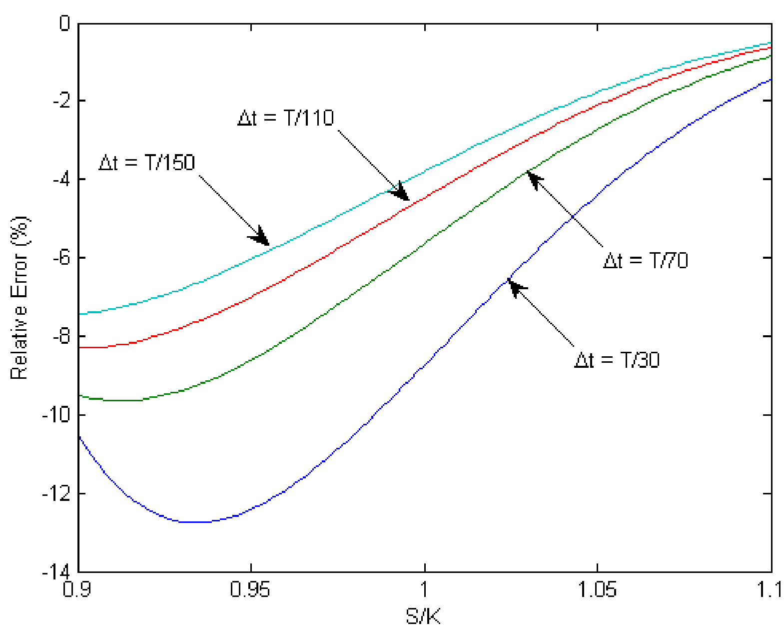
In this table we fixed the number of time intervals at 150 and varied the time to maturity T. Other parameters are as follows: $K=100$, $\sigma=20\%$, $\mu=8\%$, $r=4\%$, $k=0.5\%$.

Figure 5: Convergence of the *g*-function to its Continuous-time Limit



The figure displays the convergence behaviour of the *g*-function (3.8) to its continuous-time limit $N\left(d_1^*\right)$, where $d_1^* = d_1\left(\varphi\left(k\right)S,.\right)$ derived for the uniform distribution of the stock returns (5.1). The parameters are as follows: $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $T = 30$, $k = 0.5\%$.

Figure 6: Relative Convergence Errors of the *g*-function from Continuous-time Limit



The figure displays the relative convergence errors of the *g*-function (3.8) from its continuous-time limit $N\left(d_1^*\right)$, where $d_1^* = d_1\left(\varphi\left(k\right)S,.\right)$ derived for the uniform distribution of the stock returns (5.1). The parameters are as follows: $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $T = 30$, $k = 0.5\%$.