The Stochastic Dominance Valuation of Options under Transaction Costs. Extensions, Implementation and Empirical Tests.

Michal Czerwonko

A Thesis

in

The Department of Finance

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy at

Concordia University

Montréal, Québec, Canada

April 2008

Your file Votre référence
ISBN: 978-0-494-37780-2
Our file Notre référence
ISBN: 978-0-494-37780-2

ABSTRACT

The Stochastic Dominance Valuation of Options under Transaction Costs. Extensions, Implementation and Empirical Tests.

Michal Czerwonko, Ph. D.

Concordia University, 2008

In the first essay American call and put options on the S&P 500 index futures that violate the stochastic dominance bounds of Constantinides and Perrakis (CP, 2007) over 1983-2006 are identified as potentially profitable investment opportunities. Call bid prices more frequently violate their upper bound than put bid prices do, while evidence of underpriced calls and puts over this period is scant. In out-of-sample tests, the inclusion of short positions in such overpriced calls, puts, and, particularly, straddles in the market portfolio is shown to increase the expected utility of any risk averse investor and also increase the Sharpe ratio, net of transaction costs and bid-ask spreads. The results are strongly supportive of mispricing and also strongly supportive of the CP bounds as screening mechanisms for mispriced options.

The second essay introduces a result for call lower bound more powerful that the one applied in the first part of this thesis. The Proposition 5 call lower bound in Constantinides and Perrakis (2002) is shown to have a non-trivial limit as the time interval tends to zero. This establishes the bound as the first call lower bound known in the literature on derivative pricing in the presence of transaction costs with a non-trivial limit. The bound is shown to be tight even for a low number of time subdivisions. Novel numerical methods to derive recursive expectations under a Markovian but non-identically distributed stochastic process are presented.

The third essay relaxes an assumption in the first part of this thesis on the optimal trading policy in the presence of transaction costs. We derive the boundaries of the region of no transaction when the risky asset follows a mixed jump-diffusion instead of a simple diffusion process. These boundaries are shown to differ from their diffusion counterparts in relation to the jump intensity for lognormally distributed jump size. A general numerical approach is presented for iid risky asset returns in discrete time. An error in an earlier published work on the region of no transaction for discretized diffusions is demonstrated and corrected results are presented. Comparative results with a recent study on the same topic are presented and it is shown that the numerical algorithm has equally attractive approximation properties to the unknown continuous time limit.

TABLE OF CONTENTS

LIST OF FIGURES ... vi
LIST OF TABLES ... vii
Introduction ... 1
  References ... 7
Essay I: The Stochastic Dominance Efficiency of the S&P 500 Index Futures Options Market ... 10
  1 Restrictions on Futures Option Prices Imposed by Stochastic Dominance ... 14
  2 Empirical Design ... 19
  3 Empirical Results ... 23
  4 Concluding Remarks ... 35
  Appendix A: Data ... 36
  Appendix B: Calibration of the index return tree and numerical approach ... 37
  Appendix C: Trading policy ... 45
  Appendix D: The Sharpe ratio and the Davidson-Duclos (2000, 2006) tests ... 46
  Appendix E: Conditional versus unconditional tests ... 48
  References ... 51
Essay II: The Black-Scholes-Merton Model under Proportional Transaction Costs ... 53
  1 Introduction ... 54
  2 Stochastic Dominance Restrictions on European Call Price ... 57
  3 Proposition 5 Call Lower Bound ... 62
  4 Limiting Results ... 65
  5 Numerical Approach ... 75
  6 Results ... 86
  7 Concluding Remarks ... 92
  References ... 93
Essay III: Portfolio Selection with Transaction Costs and Jump-Diffusion Asset Dynamics ... 96
  1 Introduction ... 97
  2 Optimal Portfolio Policy under Proportional Transaction Costs ... 100
  3 Numerical Analysis ... 110
  4 Results ... 118
  5 Concluding Remarks ... 124
  References ... 125

LIST OF FIGURES

Figure 1: Illustration of Upper and Lower Bounds on Call and Put Options 24
Figure 2: Time Distribution of Observed Violations 25
Figure 3: Continuous-time Limit Results for Proposition 5 74
Figure 4: Behaviour of x and $\hat{z}$ in Time Partition 84
Figure 5: Convergence of the Proposition 5 Lower Bound to the Black-Scholes Price for $k = 0$ 86
Figure 6: Convergence of the Proposition 5 Lower Bound to Continuous-time Limit......... 87
Figure 7: Relative Convergence Errors of the Proposition 5 Lower Bound from Continuous-time Limit 88
Figure 8: Convergence of the g-function to Continuous-time Limit 90
Figure 9: Relative Convergence Errors of the g-function from Continuous-time Limit....... 91
Figure 10: No Transaction Region for Diffusion and Mixed Process 119
Figure 11: No Transaction Region for Liu-Loewenstein (2008) Parameters 122

LIST OF TABLES

Table 1 24
Table 2 26
Table 3 27
Table 4 29
Table 5 30
Table 6 31
Table 7 32
Table 8 33
Table 9 34
Table 10 42
Table 11 42
Table 12 109
Table 13 120
Table 14 121
Table 15 123

Introduction

The three essays of this dissertation cover three distinct aspects of the design and implementation of stochastic dominance (SD)[1] option pricing under realistic trading conditions. All three fill gaps into existing theory and practice as it has evolved slowly over the almost quarter of a century since it first appeared in the literature. The order of their presentation has more to do with the timing of their writing than with the organization of their material, and all three stand on their own as independent pieces. The purpose of this short introduction is to relate stochastic dominance option pricing to the arbitrage-based approach to contingent claims valuation that has dominated both theory and practice since the seminal Black-Scholes-Merton (BSM, 1973) studies, and to identify the gaps in existing knowledge that the three essays fill within the stochastic dominance literature.

Stochastic dominance was originally developed in economics as a generalization of expected utility decision-making under risk.[2] In its original form it developed rules for pairwise rankings of two risky prospects. These rules were applicable to all risk averse investors and to all discrete or continuous distributions. Its early applications in finance were in identifying efficient (undominated) portfolios and mutual funds out of a finite set of alternatives,[3] which were identified by testing all possible pairs of alternatives. Needless to say, this constitutes a severe limit to its applicability, since it has not been extended to the identification of the efficient set of feasible portfolios out of a given set of assets.[4]

Stochastic dominance was also applied to option pricing, although it was not initially identified as such. The two prospects whose distributions were compared consisted of a portfolio containing the underlying asset and the riskless asset,[5] and another portfolio that had a long or short position in an option. This allowed the derivation of upper and lower bounds on the prices of the options that depended on the entire distribution of the underlying asset's returns. The bounds were applicable to any distribution and were by construction independent of investor preferences. They initially assumed no intermediate trading till

[1] Hereafter whenever we refer to stochastic dominance we imply second degree dominance.
[2] See Hadar and Russell (1969), Hanoch and Levy (1969) and Rothschild and Stiglitz (1970).
[3] See the survey article by Levy (1992).
[4] See Post (2003) for an interesting attempt to bypass this difficulty.
[5] In fact the requirement that the option holder's portfolio contain no other risky asset save the underlying and the option is only a sufficient condition. The bounds are valid as long as the pricing kernel is monotone with respect to the underlying asset's return, a property that holds in most equilibrium models.

option expiration but were eventually extended to multiperiod trading under independent and identically distributed (iid) returns.[6]

The relationship between the option bounds and the arbitrage-based option prices, especially the BSM price, was not clear at the outset. The BSM model was derived under complete markets assumptions, under lognormal diffusion asset dynamics in continuous time and as the limit of a two-state discrete time model. Extensions that combine the arbitrage approach with market equilibrium considerations allowed the derivation of option prices under more complex assumptions about asset dynamics that incorporate jumps and/or stochastic volatility. Such extensions were based on rather restrictive assumptions, which were necessary in order to derive a unique option price. On the other hand the extension of the stochastic dominance option bounds to continuous time under suitable limit conditions did not take place till recently, in spite of some earlier partial results.[7] These extensions showed that the stochastic dominance bounds did tend both at the limit to the same unique option price that was derived under arbitrage in the complete markets case; for lognormal diffusion this was the BSM model. In the more interesting cases of jumps and stochastic volatility the limit forms of the bounds were expressions that generalized the corresponding arbitrage-market equilibrium option models that derived a unique option price.

An entirely different extension of the stochastic dominance approach was the incorporation of proportional transaction costs in the market equilibrium conditions under which the option prices were derived. Such costs cannot be included in the arbitrage-based models, since these models are based on continuous trading to eliminate possible arbitrage opportunities. As Merton (1989) first pointed out, even a small trading cost parameter would result in infinite transactions costs over the life of the option. The theoretical failure of arbitrage to accommodate this realistic feature of financial markets is common to both continuous time and the binomial-based discrete time models. Attempts to bypass this difficulty have resulted at best in approximations with unknown accuracy and in expressions that converge to the trivial arbitrage bounds originally derived by Merton (1973).[8]

---

[6] See Perrakis and Ryan (1984), Rithcken (1985), Levy (1985), Perrakis (1986) and Ritchken and Kuo (1988).

[7] See Oancea and Perrakis (2007), as well as Perrakis (1988, 1993).

[8] See Leland (1985), Boyle and Vorst (1992), Bensaid *et al* (1992) and Perrakis and Lefoll (1997, 2000).

Under stochastic dominance the starting point for option pricing is the investor holding the underlying and the riskless asset. This investor maximizes an unspecified risk averse expected utility of terminal wealth over a finite horizon, or a similarly unspecified time-additive expected utility of consumption in an infinite horizon. In the presence of proportional transaction costs it was shown by Constantinides (1979) that for any asset dynamics the optimal portfolio policy of the investor consists of a no trading zone in which she does not rebalance her portfolio, while she restructures to the nearest limit of that zone whenever the portfolio composition goes outside the limits. In the specific case of an investor with a constant proportional risk aversion (CPRA) utility Constantinides (1979) also showed that the no trading zone was a cone with possibly time-varying edges. This model was applied to diffusion asset dynamics for the risky asset in an infinite time by Constantinides (1986) and Davis and Norman (1990). In finite time it was estimated numerically in a discrete time model tending to diffusion by Genotte and Jung (1994), and extended theoretically to continuous time diffusion by an approximate method by Liu and Lowenstein (2002). More recently Liu and Lowenstein (2008) extended their earlier model to mixed jump-diffusion asset dynamics.[9]

In such models it was shown by Constantinides and Perrakis (CP, 2002) that an investor who introduces a long or short European option into her portfolio may increase her utility if the option price satisfies a given bound that is *independent* of the investor utility. This bound constitutes a reservation write or a reservation purchase price for the option, allowing any risk averse holder or writer to increase her expected utility by taking the corresponding zero net cost position in the option. These bounds are functions of the hypothesized probability distribution of the underlying asset returns and are valid for any type of asset dynamics that results in iid returns per period, even though they may be extended to non-iid returns in several important cases at some computational cost. Constantinides and Perrakis (2007) extended some of the European bounds to American options and American futures options.

The CP (2002, 2007) bounds, unlike the earlier arbitrage ones, are reasonably tight and exact and based on relatively innocuous assumptions. They are linear descendants of the

[9] There have been attempts to derive exact option prices from such models, by introducing an option in an investor's portfolio that is liquidated at option expiration. Unfortunately the resulting option reservation prices are functions of investor characteristics such as wealth and attitude towards risk. See Davis *et al* (1993).

earlier multiperiod stochastic dominance bounds, in which the transaction cost parameter has been introduced in every instance of intermediate trading of the underlying asset. Their main limitation stems from the fact that the bounds require the existence of a class of investors holding only the underlying asset and the riskless asset, as well as the option. While this raises questions about the validity of the bounds for stock options, their applicability to index options and index futures options cannot be denied, given widespread evidence that indexing is a popular policy followed by many US investors.[10] The bounds also result in closed form expressions for the reservation purchase and write prices of the options only for cases where the payoff is convex, which makes them suited primarily for plain vanilla call and put options.

Thus, the stochastic dominance approach to option pricing emerges as an alternative paradigm to arbitrage, able to fill important gaps in both theory and practice. There are, nonetheless, several questions that are raised with respect to their usefulness. The first one is the paucity of the available bounds results for the transaction cost case. CP (2002) derived only two bounds, a reservation write (upper bound) for the call option and a reservation purchase (lower bound) for the put option, that were independent of the partition of the time to option expiration into trading intervals. These were also the results that were extended to American options in CP (2007). Several other results were also derived in CP (2002), but they were complex time-recursive expressions that had never been evaluated, let alone applied to real data. There are, therefore, legitimate questions about their usefulness, since a distinct possibility exists that they may suffer the fate of the arbitrage-based results and collapse to trivial values as the time partition becomes progressively more dense.

It is in this context, of the stochastic dominance option bounds under proportional transaction costs, that the three essays of this dissertation must be viewed. Essay 2 extends the available results for European options in the all-important case of the lognormal diffusion process for the dynamics of the risky asset, by examining the limiting process of one of the time partition-dependent results of CP (2002) that develops a lower bound for a call option. The convergence of this limiting process is confirmed by a novel numerical algorithm that can also be used for other cases. It thus provides, in combination with another one of the results of CP (2002) a reasonably tight interval of admissible values for call

[10] See Bogle (2005).

options under conditions identical to BSM but with realistic trading assumptions. This, to our knowledge, is the only generalization of the BSM model under transaction costs able to provide empirically useful results.

While the SD bounds do provide theoretical benchmarks for option values in important cases where the arbitrage methods fail, there is an almost total lack of empirical results that may validate their use. In Constantinides, Jackwerth and Perrakis (2008) the European version of the CP bounds was examined in the context of S&P 500 index options and several violations were noted. Nonetheless, these bounds were estimated with a variety of in-sample estimation techniques, and there was no indication on whether these violations gave indeed rise to superior trading opportunities or were simply the result of improper estimation techniques.

Essay 1 of this dissertation addresses these problems, by applying the CP (2007) bounds to S&P 500 index futures options data from the Chicago Mercantile Exchange. A set of option prices with approximately equal times to expiration over the 1983-2006 period are selected, and for each one of them the corresponding bound is evaluated using a distribution drawn from observable returns of the S&P 500 index and its futures options till that time. Market bid or ask prices violating these bounds are identified. These violating options are then introduced into the portfolio of an investor who holds only the index and the riskless asset. At option expiration the returns of the liquidated investor portfolio with and without the option are noted. The time series of these pairs of returns are then compared in econometric tests of second degree SD.

The econometric techniques are applications of the SD tests developed by Davidson and Duclos (2000, 2007). They test for dominance of the populations from which the samples of the two time series originate. Of particular interest is the Davidson-Duclos (2007) test, which tests the null of absence of dominance; rejection of the null implies dominance of one series over another with a very high probability. The results of these out-of-sample tests are strongly supportive of the hypothesis that the returns series with the option stochastically dominates the series without the option. Dominance is confirmed with two alternative SD tests, with several alternative methods of estimating the bounds, and with several risk aversion coefficients of the investor.

Although the CP bounds are independent of investor preferences and of asset dynamics, their implementation can only be investor- and distribution-specific. In the tests carried out in Essay 1 the investor was assumed to have CPRA time-additive utility of consumption and the asset dynamics were diffusion, resulting in an adaptation of the Constantinides (1986) model. This is a limit on the generality of the SD tests, which cannot be extended easily, given the lack of available models of investor portfolio selection in the presence of transaction costs. As noted above, these models are limited to diffusion asset dynamics, with a very recent extension to jump-diffusion. Accordingly, Essay 3 of this thesis is a contribution to this topic, by developing a numerical method for the computation of the optimal portfolio for an investor holding a portfolio of a risky and a riskless asset, who maximizes the expected CPRA utility of terminal wealth. The model is formulated in discrete time, but the asset dynamics tend to a mixed jump-diffusion process at the continuous time limit. The results show that the jump component has a relatively small effect on the no transaction region, provided the total volatility of the mixed process is kept equal to that of the diffusion. The essay also examines the Genotte-Jung (1994) numerical results and identifies an error in that paper's derivations that can seriously bias the results. Last, the numerical algorithm presented in this third essay can be easily adapted to any empirical distribution with iid returns.

In summary, the three essays of this thesis extend the stochastic dominance approach to option pricing under transaction costs in three important ways. They provide an additional useful result in the form of a lower bound for a call option in the fundamental lognormal diffusion case. They provide an empirical verification of the bounds as screening devices in order to identify mispriced options and show that trading on these options generates superior results. Last, they extend the portfolio selection literature for investors in the presence of transaction costs, which can then be used to verify the stochastic dominance of trading strategies involving mispriced options under alternative asset dynamics specifications.

References

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

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

Bogle, J. C., 2005, "The Mutual Fund 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.

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., J. Jackwerth and S. Perrakis, 2008, "Mispricing of S&P 500 Index Options," forthcoming in the *Review of Financial Studies*.

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.

Davis, M. H. A. and A. R. Norman, 1990, "Portfolio Selection with Transaction Costs," *Mathematics of Operations Research* 15, 676-713.

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.

Davidson, R. and J.-Y. Duclos, 2000, "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica* 68, 1435-1464.

Davidson, R. and J.-Y. Duclos, 2006, "Testing for Restricted Stochastic Dominance," Working Paper, McGill University.

Genotte, G. and A. Jung, 1994, "Investment Strategies under Transaction Costs: The Finite Horizon Case," *Management Science* 40, 385-404.

Hadar, J. and W. R. Russell, 1969, "Rule For Ordering Uncertain Prospects," *American Economic Review* 59, 25-34.

Hanoch, G. and H. Levy, 1969. "The Efficiency Analysis Of Choices Involving Risk," Review of Economic Studies 36, 335-346.

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.

Levy, H., 1992, "Stochastic Dominance end Expected Utility: Survey and Analysis," Management Science 38, 555-593.

Liu, H. and M. Loewenstein, 2002, "Optimal Portfolio Selection with Transaction Costs and Finite Horizons," *Review of Financial Studies* 15, 805-835.

Liu, H. and M. Loewenstein, 2007, "Optimal Portfolio Selection with Transaction Costs and 'Event Risk'," working paper.

Merton, Robert C., 1973, "Theory of Rational Option Pricing," *Bell Journal of* Economics and Management Science 4, 141-183.

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, 2007, "Stochastic Dominance and Option Pricing in Discrete and Continuous Time: an Alternative Paradigm," Concordia University working paper.

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., 1993, "Options for Multinomial Stock Returns for Diffusion and Jump Processes," *Canadian Journal of Administrative Sciences* 10, 68-82.

Perrakis, S., and J. Lefoll, 1997, "Derivative Asset Pricing with Transaction Costs: an Extension," *Computational Economics* 10, 359-376.

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 Peter P. J. Ryan, 1984, "Option Bounds in Discrete Time," *The Journal of Finance 39*, 519-525.

Post T., 2003, "Empirical Tests for Stochastic Dominance Efficiency," *Journal of Finance* 58, 1905-1932.

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.

Rothschild, M., and J. E. Stiglitz, 1970, "Increasing Risk: I, a Definition," *Journal of Economic Theory*, 225-243.

Essay I: The Stochastic Dominance Efficiency of the S&P 500 Index Futures Options Market

Abstract

American call and put options on the S&P 500 index futures that violate the stochastic dominance bounds of Constantinides and Perrakis (2007) over 1983-2006 are identified as potentially profitable investment opportunities. Call bid prices more frequently violate their upper bound than put bid prices do, while evidence of underpriced calls and puts over this period is scant. In out-of-sample tests, the inclusion of short positions in such overpriced calls, puts, and, particularly, straddles in the market portfolio is shown to increase the expected utility of any risk averse investor and also increase the Sharpe ratio, net of transaction costs and bid-ask spreads. The results are strongly supportive of mispricing.

A large body of finance literature addresses the mispricing of options. Rubinstein (1994), Jackwerth and Rubinstein (1996), and Jackwerth (2000), among others, observed a steep index smile in the implied volatility of S&P 500 index options that suggests that out-of-the-money (OTM) puts are too expensive. Indeed, a common hedge-fund policy is to sell OTM puts. Coval and Shumway (2001) found that buying zero-beta at-the-money (ATM) straddles loses money. Constantinides, Jackwerth, and Perrakis (2007) provided empirical evidence that both OTM puts and calls on the S&P 500 index are mispriced by showing that they violate stochastic dominance bounds put forth by Constantinides and Perrakis (2002).

In this essay, we provide out-of-sample tests of option mispricing, net of transaction costs and bid-ask spreads. Specifically, we identify American call and put options on the S&P 500 index futures that violate the stochastic dominance bounds of Constantinides and Perrakis (2007) as potentially profitable investment opportunities. In out-of-sample tests over 1983-2006, we show that trading policies that exploit these violations provide higher Sharpe ratios than policies without option trading. We also show that the expected utility of any risk averse investor, net of transaction costs and bid-ask spreads, increases when exploiting such option trading. Below we highlight novel features of our approach.

First, we use the Chicago Mercantile Exchange (CME) data base on S&P 500 futures options, 1983-2006, which is clean and spans a long period. Much of the earlier empirical work on the mispricing of index options is based on data on the S&P 500 index options that comes from two principal sources: the Berkeley Options Database (1986-1995) that provides relatively clean transaction prices, but misses important events over the past 12 years, such as the 1998 liquidity crisis, the dot-com bubble, and its 2001 burst; and the OptionMetrics (1996-2006) data base which, however, is of uneven quality and only contains end-of-day quotes.

Second, we identify mispriced options with a screening mechanism that uses *minimal* assumptions about market equilibrium. This mechanism is based on the stochastic dominance bounds of Constantinides and Perrakis (2007). These bounds identify reservation purchase and reservation write prices such that *any* risk averse investor may increase her expected utility by including the option that violates these bounds in her portfolio. The bounds are valid for any distribution of the underlying asset and accommodate jumps. They also recognize the possibility of early exercise of American options.

The only necessary assumption about the market for the validity of these bounds is that there exists a class of traders holding portfolios containing only the S&P 500 index and the riskless asset.[11] Ample evidence exists that this assumption holds for US markets. Numerous surveys have shown that a large number of US investors follow indexing policies in their investments. Bogle (2005) reports that in 2004 index funds accounted for about one third of equity fund cash inflows since 2000 and represented about one seventh of equity fund assets. The S&P 500 index is not only the most widely quoted market index, but has also been available to investors through exchange traded funds for several years. We find that any such investor would improve her utility by including in her portfolio an option identified as mispriced by the stochastic dominance bounds.

Third, we assess the profitability of our trading policy by employing the powerful statistical tests of stochastic dominance by Davidson and Duclos (2000 and 2006) which can deal with option returns even in a setting where we do not make assumptions about the preferences of the investors. These tests compare the profitability of the optimal trading policies of a generic S&P 500 index investor with and without the option in a setting that recognizes the possibility of early exercise of the futures option. These profitability comparisons are valid from the perspective of any risk averse investor. By contrast, the ubiquitous Sharpe ratio measure of portfolio performance is valid only from the perspective of a mean-variance investor and suffers from well known problems when used to assess non-normal returns such as those encountered in portfolios that include options.

Finally, both the bounds employed in detecting mispriced options and the portfolio returns explicitly take into consideration bid-ask spreads and trading costs. Once a trading opportunity is detected, we execute the trade by buying at the next ask price or selling at the next bid price.

We use historical data on the underlying S&P 500 index returns in order to estimate the bounds. We use several empirical estimates of the underlying return distribution, all of them observable at the time the trading policy is implemented. For each one of these

[11] The mean-variance portfolio theory that gives rise to the Sharpe ratio measure of portfolio performance is based on the stronger assumption that every investor holds the market portfolio (and the risk free asset).

estimates we evaluate the corresponding bounds over the period 1983-2006, and then identify the observed S&P 500 futures options prices that violate them. For each violation, we identify the optimal trading policy of a generic investor with and without the mispriced option, using the observed path of the underlying asset till option expiration and recognizing realistic trading conditions such as possible early exercise and transaction costs. We identify the profitability of the pair of policies for each observed violation, and then conduct stochastic dominance comparison tests over the entire sample of violations. We find a substantial number of violations of the upper bounds, but relatively few violations of the lower bounds. Since the frequency of violations of the lower bounds is too low for statistical inference, we focus on violations of the upper bounds. The results are strongly supportive of mispricing.

The essay is organized as follows. In Section 1, we present the restrictions on futures option prices imposed by stochastic dominance and discuss the underlying assumptions. In Section 2, we present the empirical design and, in Section 3, the present the empirical results. We conclude in Section 4.

1 Restrictions on Futures Option Prices Imposed by Stochastic Dominance

We assume that market agents are heterogeneous and investigate the restrictions on option prices imposed by one particular class of agents that we simply refer to as "traders". We allow for other agents to participate in the market but this allowance does not invalidate the restrictions on option prices imposed on traders.

We consider a market with several types of financial assets. First, we assume that traders invest only in two of them, a bond and a stock with natural interpretation as a market index.[12] Subsequently, we assume that traders can invest in a third asset as well, an American call or put option on the index futures. The bond is risk free and has total return R. The stock has *ex dividend* stock price S_t at time t and pays cash dividend γS_t, where the dividend yield γ is deterministic. The total return on the stock, $(1+\gamma)(S_{t+1}/S_t)$, is assumed *i.i.d.* with mean R_S. The call or put option on the index futures has strike K and expiration date T. The underlying futures contract is cash-settled and has maturity T^F, $T^F \geq T$. We assume that the futures price F_t is linked to the stock price by the approximate cost-of-carry relation $F_t = (1+\gamma)^{-(T^F-t)} R^{T^F-t} S_t + \varepsilon_t, \quad t \leq T^F, \quad |\varepsilon_t| \leq \bar{\varepsilon}$, where the basis risk variables $\{\varepsilon_t\}$ are distributed independently of each other and of the stock price series $\{S_t\}$.

Transfers to and from the cash account (bond trades) do not incur transaction costs. Stock trades decrease the bond account by transaction costs equal to the absolute value of the dollar transaction, times the proportional transaction costs rate, $k, 0 \leq k < 1$. Option trades incur transaction costs, exchange fees, and price impact which are incorporated in what we refer to as their bid and ask prices.

We assume that traders maximize generally heterogeneous, state-independent, increasing, and concave utility functions. We further assume that each trader's wealth at the

[12] Essentially, we model buy-and-hold investors who trade infrequently and incur low transaction costs. At least for large investors who earn a fair return on their margin, transaction costs are even lower in the index futures market than the stock market. In practice, however, buy-and-hold investors invest in the stock and bond markets because of the inconvenience and cost of the frequent rolling over of short-term futures contracts and the illiquidity of long-term futures and forward contracts. Note that we do not restrict other market agents from trading futures contracts and other financial assets.

end of each period is monotone increasing in the stock return over the period. For example, a trader who holds 100 shares of stock and a net short position in 200 call options violates the monotonicity condition, while a trader who holds 200 shares of stock and a net short position in 200 call options satisfies the condition. Essentially, we assume that the traders have a sufficiently large investment in the stock, relative to their net short position in call options, such that the monotonicity condition is satisfied.

We do not make the restrictive assumption that all market agents belong to the class of utility-maximizing traders. Thus, our results are robust and unaffected by the presence in the market of agents with beliefs, endowments, preferences, trading restrictions, and transaction costs schedules that differ from those of the utility-maximizing traders modeled in this essay.

A trader enters the market at time zero with x_0 dollars in bonds and y_0 dollars in *ex dividend* shares of stock. We consider two scenarios. In the first scenario, the trader may trade the bond and stock but not the options. The trader makes sequential investment decisions at discrete trading dates t ($t = 0, 1 \ldots, T'$), where $T', T' \geq T^F \geq T$, is the finite terminal date. The trader's objective is to maximize expected utility, $E[u_{T'}(W_{T'})]$, where $W_{T'}$ is the trader's net worth at date T'.[13] Utility is assumed to be concave and increasing and defined for both positive and negative terminal worth, but is otherwise left unspecified. We refer to this trader as the *index* (and bond) trader, IT, and denote her maximized expected utility by $V_0^{IT}(x_0, y_0)$.[14]

In the second scenario, the trader enters the market at time zero with x_0 dollars in bonds and y_0 dollars in *ex dividend* shares of stock, but immediately writes an American futures call option with maturity $T, T \leq T^F$, where C are the net cash proceeds from

[13] Alternatively, the objective may be the maximization of the discounted sum of the utility of consumption $u_t(c_t)$ at each trading date, including the terminal date. In this case, the terminal date may be finite or infinite. Although the Constantinides and Perrakis (2007) bounds are derived under the terminal wealth objective, they remain valid without any reformulation under the alternative objective.

[14] In Essay 2, we provide more details on the investor's investment problem and also define the value function.

writing the call.[15] We assume that the trader may not trade the call option thereafter.[16] At each trading date t ($t = 0, 1 \dots, T$) the trader is informed whether or not she has been assigned (that is, assigned to act as the counterparty of the holder of a call who exercises the call at that time). If the trader has been assigned, the call position is closed out, the trader pays $F_t - K$ in cash, and the value of the cash account decreases from x_t to $x_t - (F_t - K)$. The trader makes sequential investment decisions with the objective to maximize expected utility, $E[u_{T'}(W_{T'})]$. We refer to this trader as the *option* (plus index and bond) trader, OT, and denote her maximized expected utility by $V_0^{OT}(x_0 + C, y_0)$.

For a given pair (x_0, y_0), we define the reservation write price of a call as the value of C such that $V_0^{OT}(x_0 + C, y_0) = V_0^{IT}(x_0, y_0)$. The interpretation of C is the write price of the call at which the trader with initial endowment (x_0, y_0) is indifferent between writing the call or not. Constantinides and Perrakis (2007) stated a tight upper bound on the reservation write price of an American futures call option that is independent of the trader's utility function and initial endowment and independent of the early exercise policy on the calls:

$$\overline{C}(F_t, S_t, t) = \frac{1+k_1}{1-k_2} \max\left[N(S_t, t),\ F_t - K\right], \qquad t \leq T. \qquad (1)$$

where k_1 and k_2 are respectively proportional transaction costs on stock purchase and sale. In what follows, we set $k_1 = k_2 = k$. The continuation value, the function $N(S,t)$ is defined as follows:

---

[15] The reservation write price of a call is derived from the perspective of a trader who is marginal in the index, the bond, and only one type of call or put option at a time. Therefore, these bounds allow for the possibility that the options market is segmented.

[16] The reservation write price of a call is derived under this constrained policy. Under this policy, the investor increases her expected utility by writing a call at price $\overline{C}$ and refraining from trading the call thereafter. If the constraint on trading the call is relaxed, the policy which the investor follows under the constraint policy remains feasible and increases her expected utility by writing a call at price $\overline{C}$. Therefore, $\overline{C}$ remains an upper bound on the reservation write price of a call. Whereas the upper bound may be tightened when the constraint on trading the call is relaxed, there is no known tighter bound that is preference free. For further discussion on this point, see Constantinides and Perrakis (2007).

$$N(S,t) = (R_S)^{-1} E[\max\{(1+\gamma)^{-(T^F-t)} R^{T^F-t} S_t + \overline{\varepsilon} - K, N(S_{t+1}, t+1)\} | S_t = S], \quad t \le T-1$$
$$= 0, \quad t = T. \tag{2}$$

The interpretation of the call upper bound is as follows. If we observe a call bid price above the reservation write price $\overline{C}$, then any trader (as defined in this essay) can increase her expected utility by writing the call.

If we further assume that the trader can buy a call at a price $\overline{C}(F_t, S_t, t)$ or less and trade the futures and do so costlessly, we obtain the following put upper bound:[17]

$$\overline{P}(F_t, S_t, t) = \overline{C}(F_t, S_t, t) - R^{-(T-t)} F_t + K, \quad t \le T. \tag{3}$$

The interpretation of the put upper bound is as follows. If we observe a put bid price above the reservation write price $\overline{P}$, then any trader can increase her expected utility by writing the put.

Constantinides and Perrakis (2007) also stated a tight lower bound on the reservation purchase price of an American futures put option. The cash payoff of the put exercised at time t is $K - F_t$, $t \le T$. As in the case of a call option, we define the reservation purchase price of a put as the value of P such that the trader with initial endowment (x_0, y_0) is indifferent between purchasing the put or not. The following is a tight lower bound on the reservation purchase price of an American futures put option that is independent of the trader's utility function and initial endowment:

$$\underline{P}(F_t, S_t, t) \equiv \max\left[K - F_t, \frac{1-k}{1+k} M(S_t, t) \right], \quad t \le T. \tag{4}$$

The function $M(S, t)$ is defined as follows:

[17] We prove equation (3) by noting that an investor achieves an arbitrage profit by buying a call at $\overline{C}(F_t, S_t, t)$, writing a put at $P, P > \overline{P}(F_t, S_t, t)$, selling one futures, and lending $K - R^{-(T-t)} F_t$. In the proof, we ignore the daily marking-to-market on the futures until the exercise of the put or the options' maturity, whichever comes first.

$$M(S,t)=(R_S)^{-1}E\left[\max\left[K-(1+\gamma)^{-(T^F-t)}R^{T^F-t}S_t-\bar{\varepsilon},\ M(S_{t+1},t+1)\right]\mid S_t=S\right],\quad t\le T-1$$
$$=0,\quad t=T. \tag{5}$$

If we observe a put ask price below the reservation purchase price $\underline{P}$, then any trader can increase her expected utility by buying the put.

If we further assume that the trader can write a put at price $\underline{P}(F_t,S_t,t)$ or more, and trade the futures and do so costlessly, then we obtain the following call lower bound, with corresponding interpretation:[18]

$$\underline{C}(F_t,S_t,t)=\underline{P}(F_t,S_t,t)+R^{-(T-t)}F_t-K,\quad t\le T. \tag{6}$$

If we observe a call ask price below the reservation purchase price $\underline{C}$, then any trader can increase her expected utility by buying the call.

[18] We prove equation (6) by noting that an investor achieves an arbitrage profit by writing a put at $\underline{P}(F_t,S_t,t)$, buying a call at $C, C<\underline{C}(F_t,S_t,t)$, selling one futures, and lending $K-R^{-(T-t)}F_t$.

2 Empirical Design

We describe our empirical design, starting with a description of the data, the calibration of a tree of the daily index return, and the construction of the portfolio of the index trader (who does not trade in the option) and of the option trader. This allows us to introduce the well-known Sharpe ratio test and we discuss the problems associated with using this test. To address problems with the Sharpe ratio test, we introduce tests based on second order stochastic dominance.

2.1 Data and estimation

We obtain the time-stamped quotes of the 30-calendar-day S&P 500 futures options and the underlying 1-month futures for the period February 1983-July 2006 from the Chicago Mercantile Exchange (CME) tapes. This results in 247 sampling dates. We obtain the interest rate as the three-month T-bill rate from the Federal Reserve Statistical Release. The data sources are described in further detail in Appendix A.

We set the mean index return at 4% plus the observed 3-month T-bill rate instead of estimating the mean index return from the data in order to mitigate statistical problems in estimating the mean.[19] We implement this by adding a constant to the observed logarithmic index returns so that their sample mean equals the above target. We estimate the 3rd and 4th moments of the index return as their sample counterparts over the preceding 90 days.

Finally, we estimate both the unconditional and conditional volatility of the index return as follows. We estimate the unconditional volatility as the sample standard deviation over the period January 1928 to January 1983. We estimate the conditional volatility in three different ways: (1) the sample standard deviation over the preceding 90 trading days; (2) the at-the money (ATM) implied volatility (IV) on the preceding day, adjusted by the mean prediction error for all dates preceding the given date (typically some 3%), where we drop from the preceding days all 21 pre-crash observations; and (3) the EGARCH of Nelson (1991) volatility using EGARCH coefficients estimated for S&P 500 daily returns over

[19] Short-horizon forecasts of the *conditional* mean equity premium are notoriously unreliable. Fama and French (2002), Constantinides (2002), and Dimson, Marsh and Staunton (2006) estimated the adjusted *unconditional* mean equity premium to be 4-6% per year. For our main results, we set the mean return at 4% plus the observed 3-month T-bill rate. We also report results when we set the mean return at 6% plus the observed 3-month T-bill rate.

January 1928 to January 1983 applied to residuals observed over the 90 days preceding each sample date to form projections of the volatility realized till the option expiry date.[20] In Table 1, we report statistics of the prediction error of the above volatility estimates. The best overall predictors are the adjusted ATM IV and the 90-day historical volatility.

2.2 Calibration of the index return tree and calculation of the option bounds

We model the path of the daily index return till the option expiration on a T-step tree, where T is the number of trading days in that particular month.[21] The tree is recombining with m branches emanating from each node. Every month we calibrate the tree by choosing the number of branches and the return at each node to match the first four moments of the daily index return distribution, as described in Appendix B, which also details numerical techniques applied to deal with multibranching.

The upper and lower bounds on the call and put prices are given in equations (1)-(6). We numerically calculate the bounds by iterating backwards on the calibrated tree.

2.3 Portfolio construction and trading

For each path drawn from the estimated return distribution, we employ the following trading policies. For the index trader (who manages a portfolio of the index and the risk free asset in the presence of transaction costs), we employ the optimal trading policy, as derived in Constantinides (1986) and extended in Perrakis and Czerwonko (2007) to allow for dividend yield on the stock. Essentially, this policy consists of trading only to confine the ratio of the index value to the bond value, y_t / x_t, within a no-transactions region, defined by lower and upper boundaries. We derive these boundaries for the following parameter values: one-way transaction cost rate on the index of 0.5%; annual return volatility of the index of 0.1856, the sample volatility over 1928-1983; interest rate equal to the observed 3-month T-bill date;

[20] To derive the GARCH coefficients we apply the approach of Nelson (1991) with the (1,1) structure. Applying GARCH (1,1) model did not resulted in significant changes in our results.

[21] For example, if the 3rd Friday of July is on July 27, we record the price of the July option on June 27, which is 30 calendar days earlier. (If June 27 is a holiday, we record the price on June 26.) If there are 21 trading days between June 27 and July 27, we model the path of the daily index return till the option expiration on a 21-step tree.

risk premium 4%; and constant relative risk aversion coefficient of 2.[22] For this set of parameters, the lower and upper boundaries are $y_0 / x_0 = 1.2026$ and 1.5259, respectively. At the beginning of each month and before the trader trades in options, we set $x_0 = 73{,}300$ and $y_0 = 100{,}000$, which corresponds to the midpoint of the no-transactions region, $y_0 / x_0 = 1.3642$.

For the option trader (who manages a portfolio of the option, index, and the risk free asset in the presence of transaction costs), we employ the trading policy which is optimal for the index trader but is generally suboptimal for the option trader. Recall that the goal is to demonstrate that there exist profitable investment opportunities for the option trader. Given this goal, it suffices to show that there exist profitable investment opportunities for the option trader even if the option trader follows a generally suboptimal policy. We set x_0 and y_0 to the same values as for the index trader. However, this portfolio composition changes, depending on the assumed position in futures options, as explained in Appendix C.

We focus on the cases where the basis risk bound, $\bar{\varepsilon}$, is 0.5% of the index price. Over the years 1990-2002, 95% of all observations have basis risk less than 0.5% of the index price. For reference purposes, we also consider the case $\bar{\varepsilon} = 0$. As to be expected, when we suppress the basis risk, the bounds are tighter and there appear to be more violations.

2.4 Description of the empirical tests

For each one of our methods of estimating the bounds, we obtain 247 monthly portfolio returns for the index trader and the option trader, respectively. Our goal is to test whether the portfolio profitability of the index and option traders are statistically different in the months in which we observe violations of the bounds.

In our first set of tests, we compare the Sharpe ratios of the two portfolios. Despite the well-known limitations of the Sharpe ratio, we report these results because the Sharpe

[22] We clarify that the upper and lower stochastic dominance bounds on option prices apply to any risk averse trader, independent of her particular degree of risk aversion. In our empirical work, we make an assumption about the relative risk aversion coefficient in order to calculate the boundaries of the no-transactions region for a specific trader. We present results for the value of 2 and 10 for the relative risk aversion coefficient.

ratio is one of the most popular measures of portfolio performance.[23] We use the approach of Jobson and Korkie (1981) with the Memmel (2003) correction that accounts for different variances of the two portfolios. Details of the test are described in Appendix D.

In our second set of tests, we compare the returns of the two portfolios in terms of the criterion of stochastic dominance, which states that the dominating portfolio is preferred by any risk-averse trader, independent of distributional assumptions such as normality and preference assumptions such as quadratic utility. Specifically, we test the null hypothesis $H_0 : OT \not\succ_2 IT$, which states the option trader's portfolio return does not stochastically dominate the index trader's portfolio return, against the alternative hypothesis $H_A : OT \succ_2 IT$, which states the option trader's portfolio return stochastically dominates the index trader's portfolio return. We report the results of tests proposed by Davidson and Duclos (2006), using the algorithm developed by Davidson (2007).

An earlier test, proposed by Davidson and Duclos (2000), tests the null hypothesis $H_0 : OT \succ_2 IT$ against the alternative, which is that either $IT \succ_2 OT$ or that neither one of the two distributions dominates the other. Hence, rejection of the null hypothesis fails to rank the two distributions in the absence of information on the power of the test, which is generally not available. We report results of this test as well because it has certain statistical advantages over the Davidson and Duclos (2006) test. Appendix D provides details on both tests; here we provide only a short discussion of the difference between the population and sample stochastic dominance.

Second order stochastic dominance in a population is defined by a classic integral condition on two distribution functions, i.e. in our notation $\int_{\underline{z}}^{z} [F_{IT}(x) - F_{OT}(x)]dx \geq 0$, >0 for some z, with $\underline{z}$ denoting the lower not necessarily finite limit of the joint support of the two distributions.[24] For a sample of observations, this integral criterion is not a sufficient statistic to determine stochastic dominance; however, its discrete sample equivalent forms the numerator of T-statistics for both Davidson and Duclos (2000 and 2006) tests.

---

[23] The Sharpe ratio ignores moments of the return distribution beyond the mean and variance and this is theoretically justified only the special cases where either investors have quadratic utility or the portfolio returns are normally distributed. The latter assumption is obviously violated in portfolios that include options.
[24] See, for instance, Hanoch and Levy (1969).

3 Empirical Results

In Section 3.1, we describe the empirical results. We compare the portfolio return of an option trader who writes overpriced calls, or puts, or straddles at their bid price with the portfolio return of an index trader who does not trade in the options over the period 1983-2006. In out-of-sample tests, we find that the return of an option writer stochastically dominates the index trader's return, net of transaction costs and the bid-ask spread. We also find that the Sharpe ratio of the index trader's return is higher than the Sharpe ratio of the index trader's return and is often statistically significant. In Section 3.2, we establish that the empirical results are robust.

In what follows, for trading in options we consider the quote directly following the one violating a given bound (market order).[25]

3.1 Results

In Figure 1, we plot the four bounds for one-month options, expressed in terms of the implied volatility[26], as a function of the moneyness, K / F_0. We set the underlying volatility $\sigma = 20\%$ and the error from the cost of carry $\bar{\varepsilon} = 0$. The figure also displays the 95% and 5% confidence intervals, derived by bootstrapping the 90-day distribution. Regarding the upper bounds, we observe that the call upper bound is tighter than the put upper bound. Also, the call and put upper bounds are tighter when the (*K/F*) ratio is high, that is when the calls are OTM or the puts are ITM. Regarding the lower bounds, we observe that the put lower bound is tighter than the call lower bound. Also, the call and put lower bounds are tighter when the (*K/F*) ratio is low, that is when the calls are ITM or the puts are OTM.

[25] We also considered trading at the following quote given it violates a given bound (limit order); however, we don't report the results for this approach since they were not significantly different.

[26] To derive the implied volatility, we used the Kamrad and Ritchken (1991) trinomial model.

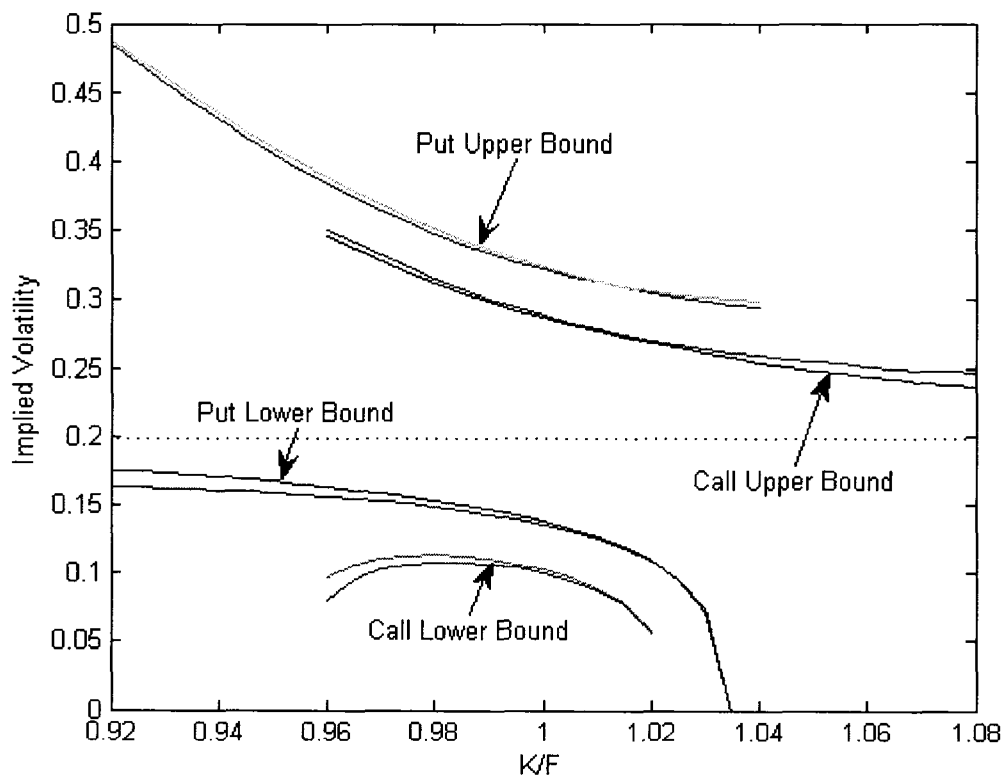
Table 1

Prediction Error of Monthly Volatility, 1983-2006

Prediction mode	Mean	Median	St. dev.	Skew.	Ex. Kurt.
Unconditional	0.0429	0.0649	0.0680	-1.7300	3.8296
90-day	0.0095	0.0076	0.0595	0.2687	5.2490
Adjusted IV	-0.0005	0.0002	0.0496	-0.2625	3.4680
GARCH	0.0177	0.0185	0.0531	0.0936	7.8302

The unconditional volatility is the sample standard deviation over the period January 1928 to January 1983. The 90-day volatility is the sample standard deviation over the preceding 90 trading days. The adjusted IV is the ATM IV on the preceding day, adjusted by the mean prediction error for all dates preceding the given date, where we drop from the preceding days all 21 pre-crash observations. The GARCH volatility is the volatility using GARCH coefficients estimated for S&P 500 daily returns over January 1928 to January 1983 and applied to residuals observed over the 90 days preceding each sample date to form projections of the volatility realized till the option expiry date.

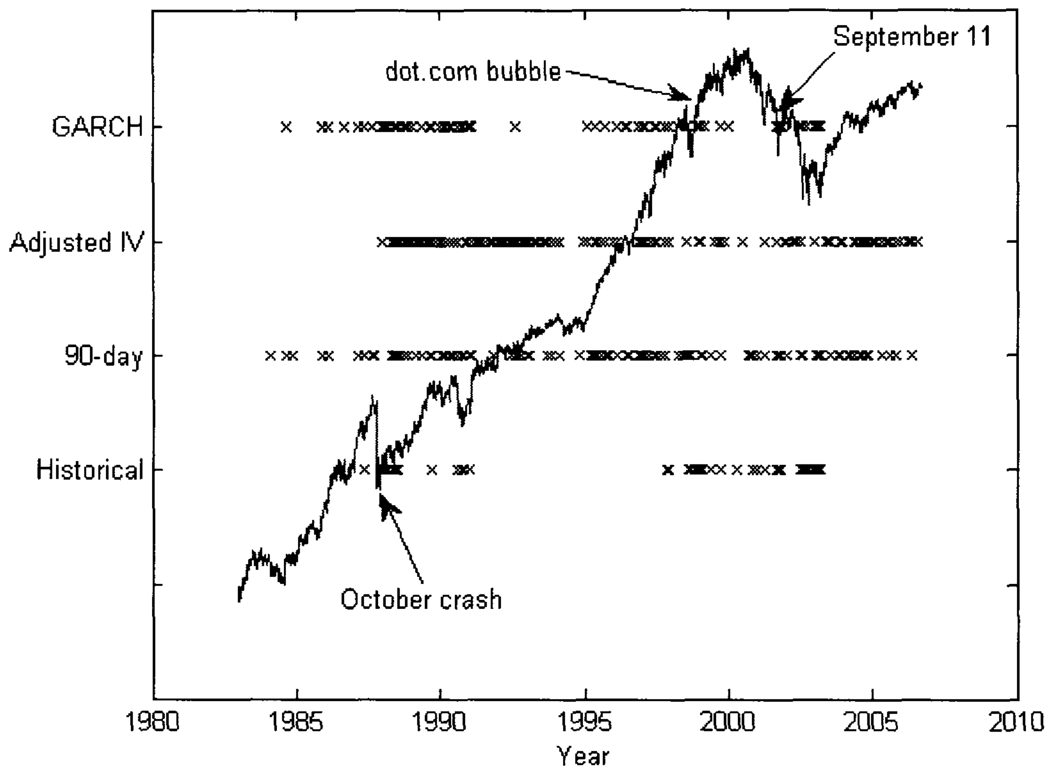
Figure 1: Illustration of Upper and Lower Bounds on Call and Put Options



Bound were derived for $\sigma = 0.20$ imposed on a 90-day distribution for a date in our sample. 95% and 5% CI were derived by bootstrapping the 90-day distribution and exemplify the bounds dependence on the third and fourth distribution moments.

In Figure 2, for every sample month, we plot the frequency of actual violations of the upper call bound.[27] We may observe in Figure 2 that the violations persist in time relatively consistently for all volatility estimation modes and that all these modes coincide at indicating violations after significant decreases in the index, i.e. when we may expect the implied volatility of the market option prices to be high.

Figure 2: Time Distribution of Observed Violations



The figure displays the violations of the call upper bound against 247 dates with app. monthly periodicity for the period of February 1983-July 2006. For the adjusted IV volatility estimation mode, the first 21 dates are not in the sample. The line across the plot is the natural logarithm of the S&P 500 index.

In Table 2, we present the cases of call and put bid prices violating their upper bound, when we set the basis risk bound at 0.5% of the index price. We do not present the cases of call and put ask prices violating their lower bound because we do not have a sufficient number of such violations to be able to draw statistical inference, as we observed in Figure 2. We observe that we have a higher frequency of violations of the upper call

[27] We don't display similar frequencies for the three remaining bounds since for these bounds we observed few violations.

bound than of the upper put bound. This may be partly explained by the fact that the upper call bound is tighter than the upper put bound, as we observed in Figure 1.

Table 2

Returns of Options Trader and Index Trader

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) p-value		DD (2006) p-value
			$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \nsucc_2 IT$
		A: Call Upper Bound			
Unconditional	44 (247)	0.074	>0.1	<0.01	0
90-day	101 (247)	0.060	>0.1	<0.01	0
Adjusted IV	120 (226)	0.090*	>0.1	<0.01	0
GARCH	65 (247)	0.096*	>0.1	<0.01	0
		B: Put Upper Bound			
Unconditional	23 (247)	0.122*	>0.1	>0.1	0
90-day	15 (247)	0.011	>0.1	>0.1	0.239
Adjusted IV	4 (226)	n/a	n/a	n/a	n/a
GARCH	8 (247)	n/a	n/a	n/a	n/a

Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbol * denotes a difference in the Sharpe ratios significant at the 10% level in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. Maximal t-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and ν set to the number of dates for which violations are observed.

The Sharpe ratio of the call trader's return is uniformly higher than the Sharpe ratio of the index trader's return, irrespective of the mode of predicting the volatility as an input to the call upper bound. When the call upper bound is calculated using the adjusted IV or the GARCH volatility, the difference in Sharpe ratios exceeds 9% annually and is statistically significant at the 10% level. There are far fewer violations of the put upper bound and, therefore, the results are statistically weaker. Nevertheless, when using the unconditional prediction of volatility as an input to the put upper bound, we find 23 violations of the put upper bound and the put trader's portfolio has a Sharpe ratio that exceeds the index trader's portfolio by 12.2%, statistically significant at the 10% level.

These Sharpe ratio preliminary results motivate and reinforce our main results on stochastic dominance which are discussed next.

The DD (2000) test does not reject the hypothesis $H_0 : OT \succ_2 IT$, which states that the option trader's return dominates the index trader's return; and rejects the hypothesis $H_0 : IT \succ_2 OT$, which states that the index trader's return dominates the option trader's return. The DD (2006) test strongly rejects the null hypothesis $H_0 : OT \not\succ_2 IT$, which states that either the index trader's return dominates the option trader's return or that neither distribution dominates the other. The p-values of the hypothesis $H_0 : IT \not\succ_2 OT$ are equal to one and are not reported here.

Table 3

Returns of Straddles Trader and Index Trader

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$			DD (2000) *p*-value		DD (2006) *p*-value
		Straddle	Call	Put	$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \not\succ_2 IT$
Unconditional	34 (247)	0.264^{***}	0.180^{**}	0.203^{**}	>0.1	>0.1	0.033
90-day	67 (247)	0.160^{**}	0.075	0.102^{*}	>0.1	>0.1	0.003
Adjusted IV	71 (226)	0.361^{***}	0.206^{**}	0.224^{***}	>0.1	<0.05	0.002
GARCH	40 (247)	0.401^{***}	0.188^{**}	0.230^{***}	>0.1	>0.1	0.010

Equally weighted average of all violating options equivalent to one call and one put per share was traded at each date. Trades were executed whenever there was a call violating the upper bound and a put traded at the same strike for the same date. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% level, respectively. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \not\succ_2 OT$ are equal to one and are not reported here. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and ν set to the number of dates for which violations are observed.

Next, we explore the performance of the policy of writing overpriced calls through the policy of writing straddles. Straddles are popular trading policies and have been investigated in the literature. For example, Coval and Shumway (2001) show that a long ATM straddle on the S&P 500 index or the S&P 100 index produces substantial negative returns.[28] Each month, we look for call bid prices that lie above the upper call bound. If we find at least one call bid prices that lie above the upper call bound *and* if we find at least one put bid price (irrespective of whether the put bid price violates the put upper bound or not) we proceed as follows. We short equal fractions of the calls that violate the call upper bound, such that the fractions add up to one; we short equal fractions of the puts for which we have bid prices, such that the fractions add up to one; and we sell one futures on the index. The results are reported in Table 3. The annualized Sharpe ratio differentials are large and significant at the 5% or 1% level. These results are consistent with the results of Coval and Shumway (2001). The DD (2000) test does not reject the hypothesis $H_0 : OT \succ_2 IT$. It often rejects the hypothesis $H_0 : IT \succ_2 OT$, but not consistently so. Finally, the DD (2006) test strongly rejects the hypothesis $H_0 : OT \nsucc_2 IT$. We conclude that the results in Table 3 are consistent with those in Table 2.

3.2 Robustness tests

In Tables 4-9, we demonstrate that the results of Tables 2 and 3 are robust. Table 4 differs from Table 2 only in that the basis risk is set at zero, $\bar{\varepsilon} = 0$, instead of bounding the basis risk by $\bar{\varepsilon} = 0.5\%$. There are now more options across the board violating the bounds because all the bounds become tighter: the upper bounds are lowered and the lower bounds are raised. We present the cases of call and put bid prices violating their upper bound. We do not present results for the cases when the call and put ask prices violate their lower bound because we still do not have a sufficient number of such violations to be able to make statistical inference.

Since the upper call and put bounds are lower, the options trader is less selective than before in writing options that violate their upper bounds and we find that the differences of the Sharpe ratios are smaller in Table 4 than in Table 2. However, since there are more observations in Table 4, the differences of the Sharpe ratios are statistically more significant than in Table 2. The DD (2000) test does not reject the hypothesis $H_0 : OT \succ_2 IT$ and rejects the hypothesis $H_0 : IT \succ_2 OT$. Finally, the DD (2006) test

[28] The paper of Coval and Shumway (2001)) focuses on the relation between the CAPM beta and the return of straddles and, as such, differs from our goal of measuring the performance of straddles, net of bid-ask spreads and through the broader criterion of stochastic dominance.

strongly rejects the hypothesis $H_0 : OT \not\succ_2 IT$. We conclude that the results in Table 4 are consistent with those in Table 2.

Table 5 differs from Table 3 on straddles only in that the basis risk is set at zero, $\bar{\varepsilon} = 0$, instead of bounding the basis risk by $\bar{\varepsilon} = 0.5\%$. Again, we conclude that the results in Table 5 are consistent with those in Table 3.

Table 4

Returns of Options Trader and Index Trader—without Futures Basis Risk

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) p-value $H_0 : OT \succ_2 IT$	DD (2000) p-value $H_0 : IT \succ_2 OT$	DD (2006) p-value $H_0 : OT \not\succ_2 IT$
		A: Call Upper Bound			
Unconditional	71 (247)	0.061	>0.1	<0.01	0
90-day	159 (247)	0.079*	>0.1	<0.01	0
Adjusted IV	197 (226)	0.086**	>0.1	<0.01	0
GARCH	114 (247)	0.100*	>0.1	<0.01	0
		B: Put Upper Bound			
Unconditional	37 (247)	0.049	>0.1	<0.05	0
90-day	50 (247)	0.057	>0.1	<0.01	0
Adjusted IV	65 (226)	0.092	>0.1	<0.01	0
GARCH	39 (247)	0.135*	>0.1	<0.01	0

The table differs from Table 2 only in that the basis risk is set at zero, $\bar{\varepsilon} = 0$, instead of bounding the risk by $\bar{\varepsilon} = 0.5$. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols * and ** denote a difference in the Sharpe ratios significant at the 10% and 5% level, respectively, in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \not\succ_2 OT$ are equal to one and are not reported here. Maximal t-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and ν set to the number of dates for which violations are observed.

Table 6 differs from Table 2 only in that the relative risk aversion coefficient is set at 10 instead of 2. Since the upper and lower stochastic dominance bounds on option prices are independent of the trader's utility, we observe the same number of violations in Table 6 as we do in Table 2. The change in the risk aversion coefficient does change the boundaries of the no-transactions region and, therefore, the trading policy of the index trader and the option trader. The Sharpe ratio differences are substantially higher in Table 6 but these differences are not statistically significant. The stochastic dominance results in writing calls are as strong in writing calls and stronger in writing puts.

Table 7 differs from Table 2 only in that the expected premium on the index is set at 6% instead of 4%. For an increase in the risk premium, the bounds become looser, i.e. the upper bounds increase and the lower bounds decrease; therefore, we observe fewer violations in Table 7 than in Table 2. The Sharpe ratio differences are comparable to those in Table 2 but these differences are not statistically significant. The stochastic dominance results in writing calls are as strong in writing calls and stronger in writing puts. We conclude that the results in the main Table 2 are robust to the assumption that the expected premium on the index is 4%.

Table 5

Returns of Straddles Trader and Index Trader—without Futures Basis Risk

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$			DD (2000) p-value		DD (2006) p-value
		Straddle	Call	Put	$H_0: OT \succ_2 IT$	$H_0: IT \succ_2 OT$	$H_0: OT \nsucc_2 IT$
Unconditional	52 (247)	0.174^{**}	0.099	0.118^{*}	>0.1	>0.1	0.019
90-day	132 (247)	0.208^{***}	0.104^{*}	0.128^{**}	>0.1	<0.01	0
Adjusted IV	167 (226)	0.240^{***}	0.181^{***}	0.203^{***}	>0.1	<0.01	0
GARCH	101 (247)	0.315^{***}	0.194^{***}	0.237^{***}	>0.1	<0.05	0

The table differs from Table 3 only in that the basis risk is set at zero, $\bar{\varepsilon} = 0$, instead of bounding the risk by $\bar{\varepsilon} = 0.5$. Equally weighted average of all violating options equivalent to one call and one put per share was traded at each date. Trades were executed whenever there was a call violating the upper bound and a put traded at the same strike for the same date. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% level, respectively. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0: IT \nsucc_2 OT$ are equal to one and are not reported here. Maximal t-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and ν set to the number of dates for which violations are observed.

Table 8 differs from Table 2 only in that the seven observations which include the date of the October crash and the following six months are excluded. The Sharpe ratio differences are comparable to those in Table 2 but these differences are not statistically significant. The stochastic dominance results in writing calls and puts are the same as in

Table 2. These results are quite natural. Unless one has exactly the day of the crash (-20% and the day thereafter with a lot of recovery), one does see much over the month of October (October as a whole was quite flat). The same seems to hold for the subsequent month. While prices were rather high (high IV for entering positions), the market was already somewhat calmer. This would mean a slight tendency options which could be sold at high prices. But then again, all this is on 7 out of 247 dates. Thus any effect is really small. All this is borne out in the table.

Table 6

Returns of Options Trader and Index Trader—with Risk Aversion Coefficient 10

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) *p*-value		DD (2006) *p*-value
			$H_0: OT \succ_2 IT$	$H_0: IT \succ_2 OT$	$H_0: OT \nsucc_2 IT$
A: Call Upper Bound					
Unconditional	44 (247)	0.155	>0.1	<0.01	0
90-day	101 (247)	0.161	>0.1	<0.01	0
Adjusted IV	120 (226)	0.227	>0.1	<0.01	0
GARCH	65 (247)	0.183	>0.1	<0.01	0
B: Put Upper Bound					
Unconditional	23 (247)	0.226	>0.1	>0.1	0.005
90-day	15 (247)	0.066	>0.1	>0.1	0.069
Adjusted IV	4 (226)	n/a	n/a	n/a	n/a
GARCH	8 (247)	n/a	n/a	n/a	n/a

The table differs from Table 2 only in that the risk aversion coefficient is set to 10, instead of 2. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbol * denotes a difference in the Sharpe ratios significant at the 10% level in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0: IT \nsucc_2 OT$ are equal to one and are not reported here. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with $k = 20$ and ν set to the number of dates for which violations are observed.

The bounds that are used in identifying mispriced options in our empirical work are calculated with parameter inputs which are point estimates and vary for each time point of our sample for all but the historical method of estimating the bounds. These varying parameters imply that the screening rules for mispriced options become conditional on the time point of our sample. Since the earlier tests do not recognize this conditionality, we develop in Appendix E an alternative set of tests that explicitly take into account the time varying nature of our sample and conclude that conditional and unconditional tests lead to same conclusions. The results are reported in Table 9 and discussed in Appendix E. They are also consistent with the main results of Table 2 and supportive of the mispricing hypothesis, even though they are derived with a different method.

Table 7

Returns of Options Trader and Index Trader—with Equity Risk Premium 6%

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) *p*-value		DD (2006) *p*-value
			$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \not\succ_2 IT$
		A: Call Upper Bound			
Unconditional	37 (247)	0.094	>0.1	<0.01	0
90-day	85 (247)	0.039	>0.1	<0.01	0
Adjusted IV	97 (226)	0.081	>0.1	<0.01	0
GARCH	58 (247)	0.084	>0.1	<0.01	0
		B: Put Upper Bound			
Unconditional	23 (247)	0.118	>0.1	>0.1	0
90-day	10 (247)	0.040	>0.1	>0.1	0.094
Adjusted IV	3 (226)	n/a	n/a	n/a	n/a
GARCH	5 (247)	n/a	n/a	n/a	n/a

The table differs from Table 2 only in that the risk premium is set to 6%, instead of 4%. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbol * denotes a difference in the Sharpe ratios significant at the 10% level in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \not\succ_2 OT$ are equal to one and are not reported here. Maximal t-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and ν set to the number of dates for which violations are observed.

Table 8

Returns of Options Trader and Index Trader—without the Crash Period

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) *p*-value		DD (2006) *p*-value
			$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \nsucc_2 IT$
		A: Call Upper Bound			
Unconditional	39 (241)	0.051	>0.1	<0.01	0
90-day	100 (241)	0.072	>0.1	<0.01	0
Adjusted IV	118 (220)	0.081	>0.1	<0.01	0
GARCH	60 (241)	0.081	>0.1	<0.01	0
		B: Put Upper Bound			
Unconditional	19 (241)	0.080	>0.1	>0.1	0.100
90-day	15 (241)	0.011	>0.1	>0.1	0.239
Adjusted IV	4 (220)	n/a	n/a	n/a	n/a
GARCH	8 (241)	n/a	n/a	n/a	n/a

This table differs from Table 2 only in that the seven observations which include the date of the October crash and the following six months were excluded. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbol * denotes a difference in the Sharpe ratios significant at the 10% level in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. Maximal t-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with $k = 20$ and ν set to the number of dates for which violations are observed.

Table 9

Returns of Options Trader and Index Trader—Non-Stationary Distribution

Panel A: Observed Option Prices

Volatility prediction mode	Call Upper Bound		Put Upper Bound	
	# months with viol. (# months)	Proportion $OT \succ_2 IT$	# months with viol. (# months)	Proportion $OT \succ_2 IT$
Unconditional	44 (247)	0.614**	23 (247)	0.870***
90-day	101 (247)	0.614***	15 (247)	0.800***
Adjusted IV	120 (226)	0.633***	4 (226)	n/a
GARCH	65 (247)	0.600**	8 (247)	n/a

Panel B: Option Prices on the Bounds

Volatility prediction mode	Call Upper Bound Proportion $OT \succ_2 IT$	Put Upper Bound Proportion $OT \succ_2 IT$
Unconditional	0.859***	0.952***
90-day	0.819***	0.964***
Adjusted IV	0.789***	0.960***
GARCH	0.855***	0.976***

This table shows the proportion of stochastic dominance tests in which the conditional bootstrapped distribution of the option trader's wealth dominated that of the index trader, as described in Appendix E. In panel A, the set of mispriced options is the same as in Table 2. In Panel B, the options are written with a price equal to the corresponding bound. The significance levels are for a binomial sign test that the indicated proportion exceeds 50%. The symbols ** and *** indicate significance respectively at 5% or better and 1% or better.

4 Concluding Remarks

We search for mispriced American call and put options on the S&P 500 index futures by employing stochastic dominance upper and lower bounds on the prices of options. We identify call and put bid prices on index futures that violate the upper bounds and call and put ask prices that violate the lower bounds. We find a substantial number of violations of the upper bounds, but relatively few violations of the lower bounds. Since the frequency of violations of the lower bounds is too low for statistical inference, we focus on violations of the upper bounds.

We compare the portfolio return of an option trader who writes overpriced calls or puts at their bid price with the portfolio return of an index trader who does not trade in the options over the period 1983-2006. In out-of-sample tests, our main result is that the return of a call or put writer stochastically dominates (in second order) the index trader's return, net of transaction costs and the bid-ask spread. The dominance holds under a variety of methods in estimating the underlying return distribution. It also holds with or without the assumption that the portfolio returns are drawn from the same distribution each period.

We also find that the Sharpe ratio of the call trader's return is uniformly higher than the Sharpe ratio of the index trader's return and is often statistically significant. The Sharpe ratio of the put trader's return is uniformly higher than the Sharpe ratio of the index trader's return but the results are less statistically significant. Finally, the policy of writing straddles produces returns that strongly stochastically dominate the index trader's return and have substantially higher Sharpe ratios. The results are supportive of the hypothesis that the options identified by violations of the CP (2007) bounds are mispriced.

Appendix A: Data

We obtain the time-stamped quotes of the one-month S&P 500 futures options and the underlying one-month futures for the period February 1983-July 2006 from the CME tapes. From the futures prices, we calculate the implied S&P 500 index prices by applying the cost-of-carry relation $F_t = (1+\gamma)^{-(T^F-t)} R^{T^F-t} S_t + \varepsilon_t$, assuming away basis risk, $\varepsilon_t \equiv 0$.[29] We obtain the daily dividend record of the S&P 500 index over the period 1928-2006 from the S&P 500 Information Bulletin and convert it to a constant dividend yield for each 30-day period. Before April 1982, dividends are estimated from monthly dividend yields. We obtain the interest rate as the three-month T-bill rate from the Federal Reserve Statistical Release. We estimate the variance of the basis risk, $\mathrm{var}(\varepsilon_t)$, from the observed futures prices and the intraday time-stamped S&P 500 record obtained from the CME.

We rescale the index price S_t by the multiplicative factor $100{,}000/S_0$ so that the index price at the beginning of each 30-day period is 100,000. Accordingly, we rescale the futures price, index futures option price, and strike by the same multiplicative factor.

We consider options maturing in 30 calendar days, which results in 247 sampling dates.[30] Since the first maturity of serial options was in August 1987, the first 19 periods occur with quarterly periodicity. Overall, we record 36,921 raw call quotes and 42,881 raw put quotes. After eliminating obvious data errors, we apply the following filters: minimum 15 cents for a bid quote and 25 cents for an ask quote; *K/F* ratio within 0.96-1.08 for calls and within 0.92-1.04 for puts; and matching the underlying futures quote within 15 seconds. Part of the data is lost due to the CME rule of flagging quotes, i.e. bids (asks) are flagged only if a bid (ask) is higher (lower) than the preceding bid (ask); in addition, no transaction data is flagged. We recover a large part of the data by analyzing the sequence between consecutive bid-ask flags; however, this recovery is not possible in all cases. As a result of the applied filters, we obtain 29,822 quotes for calls and 30,281 quotes for puts in our final

[29] Recall that our goal is to compare the investment policies, of the index trader and the option trader. Since both policies stipulate approximately the same stock component, the effect of this component cancel each other out. Also, it is a common empirical approach to derive the index value from the index futures; see, for example, Jackwerth and Rubinstein (1996).

[30] The 30-day rule eliminates the occurrence of the October crash from our sample. Therefore, we use one 40-day period to have the crash (the 248th observation) and verified that the inclusion of the crash does not alter our results.

sample. These quantities translate into roughly 60 data points for all strikes for either bid or ask prices for an average day.

Appendix B: Calibration of the index return tree and numerical approach

Estimation of the Constantinides-Perrakis (2007) bounds requires a numerical method that is flexible enough to mirror several data characteristics. In this section we present the derivation of a recombining multinomial lattice method that is flexible enough to meet the above requirement, which we detail in the following paragraphs.

The data whose distribution is to be mirrored in the estimation method consists of S&P 500 daily returns. Since it is known that the estimation of the expected return from the stock market entails well known difficulties since the work of Merton (1980), the first requirement for our method is that a lattice should assume a given mean return. The second requirement is that a lattice should assume a given second moment, which may or may not to be estimated from a given sample of the S&P daily returns. The last two requirements are that a lattice should assume the third and fourth moment of a given sample of the S&P daily returns. Here we propose a solution that results in the exact match of the first three moments, with only small errors in kurtosis. We demonstrate the size of those errors by calibrating the lattice derived by our method to a large set of the S&P daily returns.

Besides the use of physical distributions, the task at hand differs from the lattice methods commonly used in option pricing in other important aspects. In risk-neutral pricing the limiting distribution is known, which implies that any errors on the moments of the one-period distribution are of little importance as long as the time partition is sufficiently fine so as to ensure a weak convergence to the risk-neutral lognormal distribution. In our approach we do not adopt any specific limiting distribution. We also wish to keep a sufficiently coarse time partition in order to avoid market microstructure issues. We choose a partition of one (trading) day and we focus on matching the distribution moments for this specific partition.

The scope of our method is potentially wider than the estimation of the Constantinides-Perrakis (2007) bounds. It may be applied whenever recursive discrete-time expectations under an empirical distribution need to be taken, provided that the first four moments are sufficient in order to characterize a given distribution. An alternative to the presented method would be to use Edgeworth binomial tree (Rubinstein, 1998). However, this last approach may result in negative probabilities outside relatively narrow ranges for the third and fourth moment.

The number of branches in a recombining lattice M must satisfy $M = 2m+1$, where m is a positive integer. Thanks to the recombination property, at each time t there is $2mt+1$ nodes in a given lattice. The odd number of branches is not sufficient for the lattice to recombine, which occurs when the nodes are equally spaced in a log scale; or, equivalently, the ratio of any two adjacent returns is constant.

In the first step of our algorithm we pick a value for the number of branches M and group the observed returns in a histogram with M bins of equal length (on the log scale) such that the extreme bins are centered on the extreme observed returns. The center of each bin then becomes a state in the equally spaced tree, with the ordered states and the corresponding probabilities denoted respectively as x_i and p_i, $i = 1...M$. Note that, given our data set we end up with states in our lattice that have zero probabilities since S&P returns clearly have outliers.[31] We don't investigate here the precise recombination pattern of a lattice with zero-probability states; we observe, however, that the number of zero-probability states remains relatively constant as the number of convolutions of the lattice with itself increases, resulting in a decreasing proportion of such states as the time period increases.

Instead of a histogram, we could build our lattice by discretizing a kernel-smoothed distribution. However, since the kernel smoothing would involve more parameter choices,[32] we retain a preference for the histogram approximation.

We denote the target logarithmic return by $\hat{\mu}$ with the corresponding target return $\exp(\hat{\mu})$, the target return variance by $\hat{\sigma}^2$, and the sample skewness and kurtosis respectively by $\hat{\mu}_3$ and $\hat{\mu}_4$. We wish to adjust the lattice derived by the histogram so that the exponent of it will match the desired four moments, $\exp(\hat{\mu})$, $\hat{\sigma}^2$, $\hat{\mu}_3$, and $\hat{\mu}_4$. In the following paragraphs, we present an approach that will match the first three moments exactly by analytical means, with the fourth moment matched approximately by varying M,

[31] In order to have positive probabilities for all states in the lattice the lattice should have at times distances between states that preclude the goal of precise matching of the four moments.

[32] A critical parameter in the kernel density estimation is the kernel bandwidth. In addition, since the density estimate of the log-returns covers the real line, the scope of the discretized distribution would need to be chosen.

the number of branches in the lattice. First, consider an affine transformation of the log-returns x_i into $ax_i + b$. It can be easily shown that the transformed log-returns remain equidistant. We may use this transformation to impose any desired first two moments $\exp(\hat{\mu})$ and $\hat{\sigma}^2$ by simply solving two moment conditions:

$$\begin{aligned} &\sum_{i=1}^{M} p_i \exp(ax_i + b) - \exp(\hat{\mu}) = 0 \\ &\sum_{i=1}^{M} p_i \left[\exp(ax_i + b)\right]^2 - \exp(\hat{\mu})^2 - \hat{\sigma}^2 = 0 \end{aligned} \quad . \qquad \text{(B.1)}$$

These two non-linear equations in (B.1) may be easily solved numerically for the 'stretch' and 'shift' parameters a and b. For instance, if we wish to keep the sample variance in our lattice, the parameter a will simply correct for a rounding error on the variance from the histogram while b will set the mean at its target.[33] Note that in the second condition we don't need to write the expectations in the second term since the first condition ensures the target value for these expectations.

The third moment condition is more difficult to achieve since we have exhausted the means for transforming the x_i's without discarding the equidistant property. We reach the equality of the third moment by suitably varying the probabilities p_i. If, for instance, the third moment of the x_i's is too low relative to the target $\hat{\mu}_3$ we may try to solve our problem by adding a positive constant c to the right-tail probabilities and then normalizing all the probabilities to 1, while deriving this constant c jointly with a and b from a system of three equations. However, since at the same time we also need to solve for the parameters a and b,[34] it will not be necessarily the case that a positive c will provide a feasible solution to the system. Conversely, we may end up with an acceptable solution for c by subtracting a positive constant from the left-tail probabilities, provided we perform this operation only on

[33] It is clear that changing the mean has no effect on the variance.
[34] It is apparent that adjusting to the desired mean will change the third moment.

positive probabilities.[35] To cater for all contingencies, we search for a solution both in the right and left tails of the x_i's and we verify which solution results in positive probabilities. If both solutions conform to this condition, we take the one which results in a lower error on the fourth moment.

We define two transformed distributions p^* and p_* as follows:

$$p_i^* = \frac{p_i + c\underline{1}_{(i \geq n^*)}\underline{1}_{(p_i \neq 0)}}{\sum_{i=1}^{M}\left(p_i + c\underline{1}_{(i \geq n^*)}\underline{1}_{(p_i \neq 0)}\right)}$$

and , (B.2)

$$p_{*i} = \frac{p_i + c\underline{1}_{(i \leq n_*)}\underline{1}_{(p_i \neq 0)}}{\sum_{i=1}^{M}\left(p_i + c\underline{1}_{(i \leq n_*)}\underline{1}_{(p_i \neq 0)}\right)}$$

where $\underline{1}_{(.)}$ is the indicator function, n^* (n_*) is the index to this x_i which brackets from above (below) the target expected log-return $\hat{\mu}$. The first indicator function ensures that the constant c is added only to the probabilities in the chosen tail of the distribution; the second one ensures that the constant c is added only to the positive probabilities. The denominator in either line in (B.2) standardizes a given transformed distribution to 1.

We defined in (B.2) the distributions under which we may set the third moment equal to the sample moment $\hat{\mu}_3$. This operation will simply correct for a rounding error arising from the use of the histogram. For an adjustment in the right tail of the distribution we solve the following three moment conditions:

[35] Our numerical work indicated that adjusting only positive probabilities yields superior solutions in term of errors on the fourth moment.

$$\sum_{i=1}^{M} p_i^* \exp(ax_i + b) - \exp(\hat{\mu}) = 0$$

$$\sum_{i=1}^{M} p_i^* \left[\exp(ax_i + b)\right]^2 - \exp(\hat{\mu})^2 - \hat{\sigma}^2 = 0 \quad , \tag{B.3}$$

$$\sum_{i=1}^{M} p_i^* \left[\exp(ax_i + b) - \exp(\hat{\mu})\right]^3 - \hat{\mu}_3 \hat{\sigma}^3 = 0$$

where the p_i^*'s are as defined in (B.2), which definition will explicitly introduce the third variable c into the set of three non-linear equations above. These three equations may be solved numerically. For an adjustment in the left tail, substitute p_{*i} for p_i^*, as defined in (B.2).

An exact match to the fourth moment by adding a fourth equation to the system (B.3) would be a difficult task. For instance, a four-parameter solution might pose numerical difficulties, or finding all non-negative probabilities might be not feasible in some cases. To avoid these problems, we resort to varying *M*, the number of nodes in the lattice. With each new *M* the initial distribution derived from a histogram changes providing some variability in the fourth moment after the adjustments resulting from solving (B.3). After a search over a range of *M*'s we pick this distribution which has the lowest absolute difference between its kurtosis and the sample kurtosis $\hat{\mu}_4$. It turns out that this search procedure ends up with acceptably small errors in matching $\hat{\mu}_4$ for the data that we use.

In the following paragraphs we present the performance of our lattice construction method. At each of 248 dates we sample past returns over 30, 90 and 365 calendar days and measure relative errors on the target moments of the lattice. In our results, at each date we impose an annualized expected return of 1.08.[36] We present the results separately for second moments equal to 0.15^2 and 0.25^2. The third and fourth moments we attempt to match are the sample moments. The last parameter we vary is the maximum number of branches, which bounds the search for the lattice which provides the best match for the sample fourth moment.

[36] Varying the first moment has no distinguishable effect on the performance.

Table 10

Proportional Errors on the Fourth Moment ($\sigma = 0.15$)

# of Calendar Days in Sample	Mean # of Branches	$100\left\lvert\frac{\mu_4 - \hat{\mu}_4}{\hat{\mu}_4}\right\rvert$			
		Mean	Median	90th Percent.	Max
Panel A: $M \leq 101$					
30	68.1	0.083	0.010	0.035	13.077
90	66.4	0.035	0.012	0.048	1.043
365	62.6	0.019	0.009	0.038	0.264
Panel B: $M \leq 201$					
30	124.6	0.069	0.002	0.014	13.077
90	122.4	0.016	0.003	0.021	0.611
365	100.0	0.013	0.004	0.020	0.264
Panel C: $M \leq 1001$					
30	415.7	0.067	0.000	0.007	13.077
90	420.3	0.010	0.001	0.011	0.611
365	379.9	0.009	0.001	0.013	0.264

Table 11

Proportional Errors on the Fourth Moment ($\sigma = 0.25$)

# of Calendar Days in Sample	Mean # of Branches	$100\left\lvert\frac{\mu_4 - \hat{\mu}_4}{\hat{\mu}_4}\right\rvert$			
		Mean	Median	90th Percent.	Max
Panel A: $M \leq 101$					
30	63.5	0.048	0.012	0.067	2.577
90	59.8	0.055	0.018	0.146	0.721
365	60.6	0.103	0.018	0.202	1.651
Panel B: $M \leq 201$					
30	105.3	0.036	0.005	0.052	2.577
90	108.3	0.029	0.007	0.078	0.615
365	91.9	0.073	0.009	0.112	1.224
Panel C: $M \leq 1001$					
30	275.7	0.033	0.002	0.048	2.577
90	330.1	0.016	0.002	0.045	0.315
365	340.3	0.018	0.003	0.037	0.586

Tables 10 and 11 display the absolute proportional errors on the fourth moment for the second moments 0.15^2 and 0.25^2, respectively. We don't present similar errors on the first three moments since we could always solve the system of equations (3) within a numerical accuracy, which translates to the highest recorded absolute proportional error of an order of 1e-8, with 90% of these errors within of 1e-9.

The results in Tables 10 and 11 demonstrate the excellent performance of our method. In Table 10 we don't find any 90$^{\text{th}}$ percentile error above 0.05%. Except for the 30-day sampling interval, the maximum error in Table 10 is around 1%. In Table 11 we can observe that imposing a higher second moment causes a decrease in the performance; however, the errors on the fourth moment remain low. Last, note that gains from increasing the lattice size are limited since the performance changes only marginally for a five-fold increase in the maximum number of branches in panels B and C of Tables 10 and 11.

The method proposed in this essay may result in a large number of branches, which is not attractive from the point of view of recursive algorithms applied in option pricing, especially if the exercise style is American. In the following paragraphs we introduce the (discrete) Fast Fourier Transform (FFT), a computational tool suitable to deal with the problem of the lattice size. First, we define two distinct terms: a *discrete Fourier transform* is an operation which results, for instance, in the probability characteristic function if applied to a discrete density; a *discrete fast Fourier transform* is an efficient numerical algorithm used to derive discrete Fourier transforms. We will not concern ourselves with the latter while applying it in numerical work since it is a standard tool in a number of matrix algebra packages; we present applications of the former to taking recursive conditional expectations. In what follows, we denote the (fast) discrete Fourier transform as FFT.

An application of FFT to option pricing may be found in Cerny (2004). This paper, however, is entirely focused on European options, which doesn't require taking recursive conditional expectations. We introduce below a simple generalization of this application.

For any two N-dimensional vectors $x=[x_1,\dots,x_N]'$ and $y=[y_1,\dots,y_N]'$ we define a circular convolution, an operation that results in a new N-dimensional vector z:

$$z \equiv x \circ y \Leftrightarrow z_i = \sum_{k=1}^{N} x_{i-k+1_{(i-k\le 0)}N}\, y_k \quad . \tag{B.4}$$

The indexing of the elements of x ensures that N is added whenever $i-k$ is non-positive, which results in proper indices to the elements of x. As we demonstrate below, the number of computations necessary to derive the vector z may be greatly reduced with the use of FFT. First, however, we present an example demonstrating that (B.4) may be applied to derive recursive expectations. Consider, for instance, a trinomial model with two time partitions with the probabilities of one-period increasingly ordered states denoted as p_1, p_2 and p_3. After two periods, we have five terminal payoffs $y=[y_1,...,y_5]'$, which can be any function of the underlying, where the states of the underlying were set in the increasing order. Consider the following vector $x=[0,0,p_3,p_2,p_1]'$. It is easy to show that applying (B.4) to x and y will result in the vector z whose top three entries are one-period conditional expectations of the terminal payoffs. If we now apply (B4) to x and z, we obtain a five-element vector whose top entry is the expectations of the terminal payoffs as of time zero. Note that at the middle date nothing prevents us from discounting the entries to z or comparing them with any function of the states of the underlying at this date. This simple example demonstrates two important aspects of the use of the circular convolution to recursively derive conditional expectations at the initial date: first, we may use constant size vectors; second, our final result is the top entry to a vector that results from multiple circular convolutions of the constant vector containing properly ordered probabilities with time-varying one-step ahead vector function of the underlying.

The above example easily generalizes to an algorithm for any number of periods and any lattice size. All we need to do is to reverse the order of the one-period probabilities and to pad them from above with zeros to the size of the vector of terminal payoffs. Then we may proceed blindly with backward recursion, knowing that the top entry in the vector z is the desired quantity after T circular convolutions, with T denoting the number of periods to maturity. Discounting or comparing one-period conditional expectations with the present-time payoffs may be easily handled.

The circular convolution would not be of much use in numerical work without FFT since it needs even more computations that solving a multi-period lattice by 'standard' vector operations. Let F (F^{-1}) denote FFT (inverse FFT). We have:

$$x \circ y = F^{-1}\left(\sqrt{N}F^{-1}(x) \times F(y)\right), \tag{B.5}$$

where $\times$ denotes a vector element-by-element multiplication.[37] FFT can be executed very fast by a number of matrix algebra packages. Note that since the vector x will remain constant, $F^{-1}(x)$ needs to be derived only once. Therefore, at any time we need to perform just three algebraic operations: one vector element-by-element multiplication, one FFT and one inverse FFT.

For completeness sake, we present below two important properties of FFT:

i. Inversion: $$F^{-1}F(x) = FF^{-1}(x) = x, \tag{B.6}$$

ii. Linearity: $$F(\alpha x + \beta) = \alpha F(x) + \beta. \tag{B.7}$$

Appendix C: Trading policy

We consider calls with moneyness (*K/S*) within the range 0.96-1.08 and puts within the range 0.92-1.04. If we observe n call bid prices violating the call upper bound, each with different strike price, the option trader writes $1/n$ calls of each type with the underlying futures corresponding to the index value of y_0. The trader transfers the proceeds to the bond account: $x = x_0 + \sum_{i=1}^{n} C_i / n$ and $y = y_0$.

If we observe n put ask prices violating the put lower bound, each with different strike price, the option trader buys $1/n$ puts of each type and finances the purchase out of the bond account: $x = x_0 - \sum_{i=1}^{n} P_i / n$ and $y = y_0$.

However, when there is a violation of the upper put bound and the option trader writes puts, the trader also sells one futures contract for each written put. The intuition for this policy may be gleaned from the observation that the combination of a written put and a short futures amounts to a synthetic short call. In fact, the upper put bound in equation (3) is derived from the upper call bound in equation (2) through the observation that if we can

[37] The proof of this result may be found in Cerny (2004).

write a put at a sufficiently high price we violate the upper call bound by writing a synthetic call.[38]

Finally, when there is a violation of the lower call bound and the option trader buys calls, the trader also sells one futures contract for each purchased call. The intuition is the same as above.

The early exercise policy of a call is based on the function N in equation (2). The early exercise policy of a put is based on the function M in equation (5). However, whenever the option trader is short an option, each period we derive the functions N and M based on the *forward-looking* distribution of daily returns, i.e. these functions are derived under the empirical distribution of the daily index returns between the option trade and the option maturity. Effectively, we endow the counterparty of the option trader with information on the 2nd, 3rd, and 4th moments of the forward distribution, while imposing the first moment. The early exercise policy of a call or put is simplified by the observation that the decision is a function only of time and the ratio of the strike price to the index level.

Appendix D: The Sharpe ratio and the Davidson-Duclos (2000, 2006) tests

For the Sharpe ratio tests, we use the approach of Jobson and Korkie (1981) with the Memmel (2003) correction that accounts for different variances of the two portfolios. Specifically, given the sample of N realizations of the index trader's (IT) and option trader's (OT) portfolio outcomes with $\hat{\mu}_{OT}$, $\hat{\mu}_{IT}$, $\hat{\sigma}^2_{OT}$, $\hat{\sigma}^2_{IT}$, $\hat{\sigma}_{IT,OT}$ as their estimated excess means, variances, and covariances, we test the hypothesis $H_0 : \hat{\mu}_{OT}\hat{\sigma}_{IT} - \hat{\mu}_{IT}\hat{\sigma}_{OT} \leq 0$ with the test statistic $\hat{z}$, which is asymptotically standard normal:

$$\hat{z} = \frac{\hat{\mu}_{OT}\hat{\sigma}_{IT} - \hat{\mu}_{IT}\hat{\sigma}_{OT}}{\sqrt{\hat{\theta}}} \tag{D.1}$$

where

[38] In implementing the trading policy of either writing puts or buying calls, the option trader buys or sells a futures contract as well and this violates the assumption made in Section 1 that the option trader does not trade in futures. Even when we relax the assumption on trading in futures, in practice, traders manage their portfolio by trading in the index because of the inconvenience and cost of the frequent rolling over of short-term futures contracts and the illiquidity of long-term futures and forward contracts.

$$\hat{\theta}=\frac{1}{N}\left(2\hat{\sigma}_{IT}^{2}\hat{\sigma}_{OT}^{2}-2\hat{\sigma}_{IT}\hat{\sigma}_{OT}\hat{\sigma}_{IT,OT}+\frac{1}{2}\hat{\mu}_{OT}^{2}\hat{\sigma}_{IT}^{2}+\frac{1}{2}\hat{\mu}_{IT}^{2}\hat{\sigma}_{OT}^{2}-\frac{\hat{\mu}_{IT}\hat{\mu}_{OT}}{\hat{\sigma}_{IT}\hat{\sigma}_{OT}}\hat{\sigma}_{IT,OT}^{2}\right). \quad \text{(D.2)}$$

DD (2000) provide a test of the null hypothesis $H_0 : OT \succ_2 IT$ in terms of the maximal and minimal values of the extremal test statistic, $T(z)$. The null is not rejected, if the maximal value of the statistic is positive and statistically significant *and* the minimal value of the statistic is either positive or negative and statistically not significant.

The variable z denotes the logarithm of end-of-the-month wealth of a trader, where the subscripts *IT* and *OT* distinguish between the index trader and the option trader. The statistic $T(z)$ is defined as follows:

$$T(z)=\frac{D_{IT}^{2}(z)-D_{OT}^{2}(z)}{\sqrt{\hat{V}^{2}(z)}} \quad \text{(D.3)}$$

where

$$D_{I}^{2}(z)=\frac{1}{N}\sum_{i=1}^{N}(z-W_{Ii})^{+} \quad \text{(D.4)}$$

$$\hat{V}^{2}(z)=\hat{V}_{IT}^{2}(z)+\hat{V}_{OT}^{2}(z)-2\hat{V}_{IT,OT}^{2}(z) \quad \text{(D.5)}$$

$$\hat{V}_{I}^{2}(z)=\frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}(z-W_{Ii})_{+}^{2}-D_{I}^{2}(z)^{2}\right],\ I=IT,OT \quad \text{(D.6)}$$

and

$$\hat{V}_{OT,IT}^{2}(z)=\frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}(z-W_{ITi})^{+}]^{2}[(z-W_{OTi})^{+}-D_{IT}^{2}(z)D_{OT}^{2}(z)\right]. \quad \text{(D.7)}$$

The maximal and minimal values of the statistic are calculated as a maximum and minimum of (D.3) over a set of points of z as explained below. Stoline and Ury (1979) provide tables for the distribution of the maximal and minimal value of $T(z)$, which is not standard at the levels 1, 5 and 10%. In principle, the number of points in this joint support over which the test may be performed needs to be restricted since a 'large' number of these points violates the independence assumption between the $T(z)$s. Therefore, we compute these statistics for 20 points equally spaced in the log-transformed joint support of W_{IT} and W_{OT}, which corresponds to $k = 20$ in the Stoline and Ury (1979) tables.

DD (2006) provide a test of the null hypothesis $H_0 : OT \not\succ_2 IT$. The test statistic is the same as in DD (2000), except that instead of the extremal T-statistic we are now interested in the *minimal* T-statistic. This statistic is computed for the values of z that are sample points within the restricted interval, i.e. in this interval we have coupled log-transformed observations of W_{IT} and W_{OT}. As opposed to the DD (2000) test, there is no restriction on the number of these points and we compute the minimal value of $T(z)$ in the restricted interval.[39] If the minimal value is negative, the null of non-dominance is accepted. Otherwise, there exists a bootstrap approach for the derivation of the p-values for the null hypothesis, which is described in detail in DD (2006) and Davidson (2007). In our tests, we use 999 bootstrap replications in order to derive the p-values in the tables.

There is a cost in adopting the DD (2006) null, because, as it can be analytically shown, this null cannot be rejected over the entire support of the sample distribution. DD (2006) overcame this problem by restricting the interval over which the null may be rejected to the interior of the support, excluding points at the edges. They then showed by simulation that inferences on the basis of this restricted interval constitute the most powerful available inference on the existence of stochastic dominance. We follow their suggestion on the method for restricting the interval, which we also test on simulated data.[40]

Appendix E: Conditional versus unconditional tests

For each time point of our sample we generate artificially samples of stock return paths drawn from a bootstrapped distribution constructed from the (approximately 22) observed daily stock returns till option expiration *for each one* of the 247 dates t=1,..., 247 in our data period. Such a distribution represents the information that the trader would have used to estimate the bounds had she been able to observe it. For each stock return path of the

[39] It may be shown that $T(z)$ is monotonic between the sample points; therefore the minimal value of $T(z)$ may be found only at a sample point.

[40] Details on the restrictions of the test interval are available from the authors on request.

bootstrap we then compute the wealth indices for OT and IT, W_{OT} and W_{IT}, and generate two distributions of these quantities at each date in our period and for each method of estimating the bounds; recall that all these estimation methods use only quantities that can be observed by the trader before adopting her option position. Hence, evidence that the returns of the option trader dominate the returns of the index trader for "most" of the time points of our sample validates the use of the observable distribution for estimating the bounds in lieu of the unobservable distribution of the actual stock return paths.

Since these bootstrapped samples are large, we can treat the samples as the entire populations, applying a direct stochastic dominance test based on the integral condition that defines stochastic dominance.[41] This integral condition takes the following form, for $J = OT$, IT and for $\underline{z}$ denoting the lower limit of the joint support of the two distributions F_{OT}, F_{IT}:

$$D_{IT}^2(z) - D_{OT}^2(z) \geq 0, \;\; \text{where } D_J^2(z) = \int_{\underline{z}}^{z} (z - x)\, dF_J(x) \tag{E.1}$$

In the particular case in which we observe the paired wealth levels W_{OT} and W_{IT} from a sample of size N with values W_{Ji}, $i = 1,\ldots,N$, $J = OT, IT$ the test statistic $D_J^2(z)$ becomes

$$D_J^2(z) = \frac{1}{N}\sum_{i=1}^{N}\left(z - W_{Ji}\right)^{+} \tag{E.2}$$

For the bootstrapped distribution, we calculate the SD2 test statistic from (E.1)-(E.2) for the hypothesis $H_0 : OT_t \succ IT_t \;\; t = 1,\ldots, 247$ as above and decide on acceptance/rejection at a chosen significance level α (say 5%). Next, we set the variable Z_t to equal one if $\text{Prob}\{H_0 \text{ false}\} \leq \alpha$, and zero otherwise. The hypothesis $\Pr\text{ob}\{OT_t \succ IT_t\} > 0.5$ for any t, against the alternative $\Pr\text{ob}\{OT_t \succ_2 IT_t\} \leq 0.5$ for any t, is accepted if $\sum_1^{247} Z_t \geq \beta$,

[41] See, for instance, Hanoch and Levy (1969). Condition (9) can be easily shown, through integration by parts, to be equivalent to the better-known form of the integral condition used in most SD studies.

where β is chosen according to the desired significance level from the binomial distribution with probability $p = ½$.

In Table 9, panel A, we present the results of these conditional tests. The upper panel tests the hypothesis $\Pr ob\{OT_t \succ IT_t\} > 0.5$ for the observed option bid prices that violate the call and put upper bounds in equations (1)-(4) under the same conditions as Table 2. The results are strongly supportive of the null hypothesis in all but one case for which there are too few observations, and are in full agreement with the results of the unconditional test of Table 2. Similar results also hold for the options that violate the option upper bounds under the conditions of Table 3, with the basis risk set equal to 0.[42]

In Table 9, panel B, we present the results of tests of the hypothesis $\Pr ob\{OT_t \succ IT_t\} > 0.5$ for the artificial set of options written at the upper bounds of the call and put options, as in the upper panel of Table 6. Again, the results are strongly supportive of the null hypothesis in all cases, with the observed probabilities $\Pr ob\{OT_t \succ IT_t\}$ greater than 65% in all but one case and always significantly greater than 50%.[43] Hence, conditional and unconditional tests agree here as well. Similar results (available upon request) establish the validity of the hypothesis $\Pr ob\{OT_t \succ IT_t\} > 0.5$ for call options purchased at the lower bound of equation (6), while for put options purchased at the lower bound of equations (4) and (5) the hypothesis is verified in all cases except for $K / F < 0.98$, again as in the unconditional tests.

[42] The results are available from the authors on request.
[43] Similar, although slightly weaker, results also hold for the option upper bounds for the case where there is no basis risk in computing the bounds.

References

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

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

Bollerslev, T., 1986, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

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

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

Constantinides, G. M., 2002, "Rational Asset Prices," *Journal of Finance* 57, 1567-1591.

Constantinides, G. M., J. C. Jackwerth, and S. Perrakis, 2007, "Mispricing of S&P 500 Index Options," *Review of Financial Studies*, forthcoming.

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.

Coval J. D. and T. Shumway, 2001, "Expected Option Returns," *Journal of Finance* 56, 983-1009.

Davidson, R., 2007, "Testing for Restricted Stochastic Dominance: Some Further Results," Working Paper, McGill University.

Davidson, R. and J.-Y. Duclos, 2000, "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica* 68, 1435-1464.

Davidson, R. and J.-Y. Duclos, 2006, "Testing for Restricted Stochastic Dominance," Working Paper, McGill University.

Dimson, E., P. Marsh, and M. Staunton, 2006, "The Worldwide Equity Premium: A Smaller Puzzle," Working paper, London Business School.

Fama, E. F., and K. R. French, 2002, "The Equity Premium," *Journal of Finance* 57, 637-660.

Hull, J. C., 2006, *Options, Futures, and Other Derivatives*, Prentice Hall.

Jackwerth, J. C., 2000, "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies* 13, 433-451.

Jackwerth, J. C., 2004, "Option-Implied Risk-Neutral Distributions and Risk Aversion," ISBN 0-943205-66-2, Research Foundation of AIMR, Charlottesville, USA.

Jackwerth, J. C. and M. Rubinstein, 1996, "Recovering Probability Distributions from Option Prices," *Journal of Finance* 51, 1611-1631.

Jobson J. D. and B. M. Korkie, 1981, "Performance Hypothesis Testing with the Sharpe and Treynor Measures," *Journal of Finance* 4, 889-908.

Kamrad, B. and P. Ritchken, 1991, "Multinomial Approximating Models for Options with k State Variables," *Management Science* 37, 1640-1652.

McDonald, R. L., 2006, *Derivatives Markets*, Addison-Wesley.

Memmel, C., 2003, "Performance Hypothesis Testing with the Sharpe Ratio," *Finance Letters* 1, 21-23.

Merton, R. C., 1973, "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science* 4, 141-183.

Merton, R. C., 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics* 8, 323-361.

Nelson, D. B., 1991, "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica* 59, 347-370.

Perrakis, S. and M. Czerwonko, 2006, "Transaction Costs and Stochastic Dominance Efficiency in the Index Futures Options Market," Working paper, Concordia University.

Rubinstein, M., 1994, "Implied Binomial Trees," *Journal of Finance* 49, 771-818.

Rubinstein, M., 1998, "Edgeworth Binomial Trees," *Journal of Derivatives* 5, 20-27.

Stoline, M. R. and H. K. Ury, 1979, "Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons among Means," *Technometrics* 21, 87-93.

Essay II: The Black-Scholes-Merton Model under Proportional Transaction Costs

Abstract

Proposition 5 call lower bound in Constantinides and Perrakis (2002) is shown to have a non-trivial limit as the time interval tends to zero. This establishes the bound as the first call lower bound known in the literature on derivative pricing in the presence of transaction costs with a non-trivial limit. The bound is shown to be tight even for a low number of time subdivisions. Novel numerical methods to derive recursive expectations under non-identically distributed stochastic process are presented.

1 Introduction

Stochastic dominance bounds on European and American option prices in the presence of proportional transaction costs were derived by 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 essay we redefine the CP 2002 European call option lower bound in a discrete time model of the underlying asset distribution that converges to a lognormal diffusion as the time partition tends to zero. We then show that the corresponding lower bound converges to tight and non-trivial Black-Scholes type expression as the partition of trading time tends to zero, even if the transaction cost parameter remains constant. Further, this bound converges to the conventional Black-Scholes value when that parameter is set equal to zero.

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.[44] For the American call options CP (2007) derived similarly a tight reservation write price and a very loose reservation purchase price. In all cases the derived CP reservation prices *did not* converge to the prices that would prevail in a complete and frictionless market if the transaction cost parameters were set equal to zero. Their relationship, therefore, to the Black-Scholes price remained unclear.

[44] These bounds were empirically applied in Constantinides, Jackwerth and Perrakis (2008) to the market of S&P 500 index options.

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 essay 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 essay we focus on one of the call option reservation purchase prices, the prices given by Proposition 5 in CP (2002).[45] This price is basically a generalization of the call option lower bounds derived originally by Levy (1985) and Ritchken (1985), and extended to a multiperiod context by Perrakis (1988) and Ritchken and Kuo (1988), in a trading model that includes proportional transaction costs. We reformulate the CP results, applying them to a case where the iid returns tend to a lognormal distribution as trading becomes progressively more dense, as in Oancea-Perrakis (2007). 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 developed 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.

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

[45] 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.

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 the binomial approach ended up with the same dilemma as the continuous time discretization, insofar as the width of the option bounds increased with the time partition defining the size of the binomial tree.

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). The major drawback of this approach is the dependence of the derived reservation option prices on the investor risk aversion coefficient. Given the uncertainty prevailing as to the size of that coefficient for the "average" investor,[46] the reservation prices derived by the expected utility approach cannot be generalized to the entire market.

[46] See Kocherlakota (1996).

2 Stochastic Dominance Restrictions on European Call Price

We adopt the same general setup as in CP (2007), with a few important differences. Although we consider a market with several assets, we focus on 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.[47] We refer to these investors as *utility-maximizing traders* or simply as "traders". We do not make the restrictive assumption that *all* investors participating in the market belong to the class of utility-maximizing traders. Thus our results are unaffected by the presence of traders with different objectives and preferences and facing a different transaction costs schedule than that of the utility-maximizing traders. Into this setup we introduce a long 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.[48] 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 $(1+\gamma_t)S_t$, the cash dividend is $\gamma_t S_t$, and the *ex dividend* stock price is S_t, where the *dividend yield* parameters $\{\gamma_t\}_{t=1,...,T'}$ 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 rate of return on the stock, $(1+\gamma_{t+1})\frac{S_{t+1}}{S_t}$ is the compact subset $[z_{\min}, z_{\max}]$ of the positive real line.[49] To simplify the 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

[47] There is ample evidence that many US investors follow such an indexing strategy. See Bogle (2005).

[48] 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 $\Delta t \to 0$.

[49] In CP (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.

$$z_1 \equiv E\left[(1+\gamma)\frac{S_{t+1}}{S_t}\right]. \tag{2.1}$$

known to the trader at time zero. We also assume that $z_1 > E\left[\frac{S_{t+1}}{S_t}\right] > R$.

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 $0 \le k_2 < 1$. We also define the mean return with the dividend reinvested in the stock, net of transaction costs, long and short as

$$z_2 \equiv E\left[\left(1+\frac{\gamma}{1+k_1}\right)\frac{S_{t+1}}{S_t}\right],\ z_3 \equiv E\left[\left(1+\frac{\gamma}{1-k_2}\right)\frac{S_{t+1}}{S_t}\right]. \tag{2.2}$$

In practice, the distinction between z_1 and z_2 or z_3 is negligible, given that both the dividend yield (γ) and the transaction costs rates (k_1, k_2) are small.

We consider a trader who enters the market at date *t* with dollar holdings x_t in the bond account and y_t / S_t *ex dividend* shares of stock. The endowments are stated net of any dividend payable on the stock at time *t*.[50] The trader increases (or, decreases) the dollar holdings in the stock account from y_t to $y_t' = y_t + \upsilon_t$ by decreasing (or, increasing) the bond account from x_t to $x_t' = x_t - \upsilon_t - \max\left[k_1\upsilon_t, -k_2\upsilon_t\right]$. The decision variable υ_t is constrained to be measurable with respect to the information up to date *t*. The bond account dynamics is

$$x_{t+1} = \left\{x_t - \upsilon_t - \max\left[k_1\upsilon_t, -k_2\upsilon_t\right]\right\}R + (y_t + \upsilon_t)\frac{\gamma S_{t+1}}{S_t}, \quad t \le T'-1 \tag{2.3}$$

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

and the stock account dynamics is

$$y_{t+1} = (y_t + \upsilon_t)\frac{S_{t+1}}{S_t}, \quad t \le T'-1. \tag{2.4}$$

At the terminal date, the stock account is liquidated, $\upsilon_{T'} = -y_{T'}$, and the net worth is $x_{T'} + y_{T'} - \max[-k_1 y_{T'}, k_2 y_{T'}]$. At each date t, the trader chooses investment υ_t to maximize the expected utility of net worth, $E\left[u\left(x_{T'} + y_{T'} - \max[-k_1 y_{T'}, k_2 y_{T'}]\right) | S_t\right]$.[51] 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.[52]

We define the value function recursively as

$$\begin{aligned} &V(x_t, y_t, t) \\ &= \max_{\upsilon} E\left[V\left(\{x_t - \upsilon - \max[k_1\upsilon, -k_2\upsilon]\}R + (y_t + \upsilon)\frac{\gamma S_{t+1}}{S_t}, (y_t + \upsilon)\frac{S_{t+1}}{S_t}, t+1\right)\right] \end{aligned} \tag{2.5}$$

for $t \le T'-1$ and

$$V(x_{T'}, y_{T'}, T') = u\left(x_{T'} + y_{T'} - \max[-k_1 y_{T'}, k_2 y_{T'}]\right). \tag{2.6}$$

We assume that the parameters satisfy appropriate technical conditions such that the value function exists and is once differentiable with respect to x_t and y_t. We denote by υ_t^* the optimal investment decision at date t corresponding to the portfolio (x_t, y_t). The value

[51] 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.

[52] If utility is defined only for non-negative net worth, then the decision variable is constrained to be a member of a convex set, A, 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.

function $V(x,y,t)$ is increasing and concave in (x,y), properties inherited from the monotonicity and concavity of the utility function $u(.)$, given that the transaction costs are quasi-linear.[53]

We define x_t' and y_t' as

$$x_t' = x_t - \upsilon_t^* - \max\left[k_1\upsilon_t^*, -k_2\upsilon_t^*\right] \tag{2.7}$$

and

$$y_t' = y_t + \upsilon_t^*. \tag{2.8}$$

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

$$V(x_t, y_t, t) = V(x_t', y_t', t) \tag{2.9}$$

Relations (2.1)-(2.9) are sufficient for the CP (2002, 2007) derivations of the bounds. For a given pair (x_t, y_t), we define the reservation purchase price of a call as the value of C such that

$$V(x_t + (1-k_2)g_t S_t - C, y_t - g_t S_t, t) = V(x_t, y_t, t), \tag{2.10}$$

where $g_t < 1$ is the number of shares of the stock sold. The interpretation of C is the purchase price of the call at which the trader with initial endowment (x_t, y_t) is indifferent between purchasing the call or not. CP (2002) in Proposition 5 stated a partition-dependent bound[54] on the reservation purchase price of a European call option that is independent of

[53] See Constantinides (1979) for details.

[54] The interpretation of this bound is that when an investor observes a call price below the bound, she increases her expected utility by purchasing the call financed by shorting g_t shares, with the remainder of proceeds invested in the riskless bond.

the trader's utility function and initial endowment. A key ingredient to this result is optimizing the number of sold shares g_t which we elaborate in later sections.

Many of the results for the option bounds in incomplete markets without transaction costs, however, have been derived in a framework first developed by Ritchken (1985), which relies on the pricing kernel and the first order utility maximization conditions of the trader. The two approaches are equivalent. This approach relies on the partial derivatives of the value function $V(x,y,t)$, which by marginal analysis satisfy the following conditions, if we set $V(t) \equiv V(x_t, y_t, t)$ to simplify the notation:

$$V_x(t) > 0, \quad V_y(t) > 0, \qquad t = 0,...,T,...T' \tag{2.11}$$

$$(1-k_1)V_x(t) \le V_y(t) \le (1+k_2)V_x(t), \qquad t = 0,...,T,...T' \tag{2.12}$$

$$V_x(t) = RE\left[V_x(t+1)\right], \qquad t = 0,...,T,...T'-1 \tag{2.13}$$

$$V_y(t) = E\left[\frac{S_{t+1}}{S_t}V_y(t+1) + \gamma_t \frac{S_{t+1}}{S_t}V_x(t+1)\right], \qquad t = 0,...,T,...T'-1 \tag{2.14}$$

$$V_x(x,y,t),\ V_y(x,y,t) \text{ are nonincreasing in } x \text{ and } y \text{ respectively.} \tag{2.15}$$

(2.11) and (2.15) stem from the monotonicity and concavity of the value function. (2.12) reflects the ability of the trader to transfer funds between the bond and stock accounts by incurring transaction costs. (2.13) and (2.14) are conditions on the marginal rates of substitution, respectively between the bond accounts and between the stock and the bond and stock accounts, at dates t and $t+1$.

3 Proposition 5 Call Lower Bound

In this section, we introduce the CP Proposition 5 call lower bound and present a general idea for the numerical solution. We set $k = k_1 = k_2$, a one-way transaction costs rate, $k \in [0,1)$. To simplify the notation, we define the following constants: $\varphi(k) \equiv (1-k)/(1+k)$ and $\alpha(k) \equiv 2k/(1+k)$. We also define the following function:

$$I(z) \equiv \begin{cases} 1/(1+k), & z \leq 0 \\ 1/(1-k), & z > 0 \end{cases}. \tag{3.1}$$

We define the Proposition 5 lower bound for the following discrete-time stock process:

$$\frac{S_{t+\Delta t}}{S_t} \equiv z_{t+\Delta t} = 1 + \mu\Delta t + \sigma\varepsilon\sqrt{\Delta t}\ , \tag{3.2}$$

where $\varepsilon \sim F(0,1)$ and F is a general distribution with zero mean and unit variance with density $f(.)$. We further assume that ε is bounded with compact support, i.e. $\varepsilon \in [\varepsilon_{\min}, \varepsilon_{\max}]$. The returns are iid, which implies that the time subscript can be omitted and the limit of the distribution is a diffusion, which we demonstrate in the following section. We also define Z_t as the time-t state variable, be it the stock price or, if we define the bound for the stock price $S_0 = 1$ and $K' = K / S_0$, be it the time-t stock return.

One period prior to option expiration the lower bound is defined as:

$$\underline{C}_5(Z_{T-1}, T-1) = E\left[(Z_{T-1}z - K)^+ \mid Z_{T-1}, z \leq \hat{z}_{T-1}\right] \Big/ R, \tag{3.3}$$

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

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

and the number of shorted shares $g_{T-1}(Z_{T-1})$ is equal to

$$g_{T-1}(Z_{T-1}) = \frac{(Z_{T-1}\hat{z}_{T-1} - K)^{+} - R\underline{C}_5(Z_{T-1}, T-1)}{(\hat{z}_{T-1}/\varphi(k) - R)^{+} Z_{T-1}} \tag{3.5}$$

Note that as Δt gets small, (3.4) may not have a solution since R will approach 1 while $\varphi(k)$ remains constant. We address this problem in the section below.

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

$$\underline{C}_5(Z_t, t) = \frac{E\left[\underline{C}_5(Z_t z, t+1) I(z-x) \mid Z_t, z \le \hat{z}\right] + \alpha(k) Z_t E\left[G_{t+1}(x) I(z-x) z \mid Z_t, z \le \hat{z}\right]}{RE\left[I(z-x) \mid Z_t, z \le \hat{z}\right]}, \tag{3.6}$$

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

$$\frac{E\left[z \mid z \le \hat{z}\right]}{(1-k)E\left[I(z-x) \mid z \le \hat{z}\right]} = R, \tag{3.7}$$

where we suppressed the dependence of $\hat{z}$ on x to simplify the notation. $g_t(Z_t)$ is given by:

$$g_t(Z_t) = \frac{\underline{C}_5(Z_t\hat{z}, t+1) - R\underline{C}_5(Z_t, t)}{\varphi(k)(\hat{z} - R)Z_t}, \tag{3.8}$$

and with $G_{t+1}(x) \equiv \begin{cases} g_{t+1}(z), & z \le x \\ 0, & z > x \end{cases}$. The value of x is implied by the equation:

$$R\left[\varphi(k) g_t Z_t - \underline{C}_5(Z_t, t)\right] = \varphi(k) g_{t+1}(x) Z_t x - \underline{C}_5(Z_t x, t+1), \tag{3.9}$$

where from the proof of Proposition 5 in CP 2002 we know that $x \leq R$.

We state now a generic program solving for $\underline{C}_5(Z_t, t)$ in numerical work by noting that once we have derived $\underline{C}_5(Z_{t+1}, t+1)$, the only quantity of interest is *x*. Select a candidate value for *x*, which implies a value for $\hat{z}$ by (3.7) and therefore generates a candidate solution for $\underline{C}_5(Z_t, t)$ by (3.6), which in turn generates a candidate solution for g_t. Verify whether the condition (3.9) holds for the quadruple of candidate values for *x*, $\hat{z}$, $\underline{C}_5(Z_t, t)$ and g_t since all other quantities in (3.9) are known at time *t*. Search for *x* till the condition (3.9) is satisfied.

4 Limiting Results

For the stock return process defined in (3.2), we examine the limiting form of the lower bounds derived in Propositions 5 of CP (2002) as $\Delta t \to 0$.

In the presence of proportional transaction costs Proposition 5 of CP (2002) shows that the call lower bound $\underline{C}_5(S_t, t)$ is a discounted recursive expectation of its payoff under a transformed process. Define the density

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

where $I(z-x)$ is defined in (3.1), and x belongs to the support of the distribution of z. In Proposition 5 the stock returns are still given by (3.2) and equation (3.1) still defines the function $I(z)$ from the associated value of x, but now the distribution of ε is *truncated* at a value $\bar{\varepsilon} \le \varepsilon_{\max}$. The truncated support for ε ($\equiv \varepsilon_x$), therefore, has $\varepsilon_x \in [\varepsilon_{\min}, \bar{\varepsilon}]$. It is known that $x \le R = 1 + r\Delta t$, from which we get

$$x = 1 + \mu\Delta t + \sigma\varepsilon_x\sqrt{\Delta t} \le 1 + r\Delta t \Rightarrow \varepsilon_x \le -\frac{\mu - r}{\sigma}\sqrt{\Delta t}.$$

(4.2)

We also define here the following expression

$$E(I) \equiv \left[\frac{1}{1+k}F(\varepsilon_x) + \left(F(\bar{\varepsilon}) - F(\varepsilon_x)\right)\frac{1}{1-k}\right]\frac{1}{F(\bar{\varepsilon})} = E\left[I(z)\middle|\varepsilon \le \bar{\varepsilon}\right] \tag{4.3}$$

Equation (4.1) still defines a transformation of the original density $f(.)$, with the difference that the denominator is now replaced by (4.3) and the support of the distribution is now $[\varepsilon_{\min}, \bar{\varepsilon}]$. Let

$$\frac{f(\varepsilon)}{F(\bar{\varepsilon})} \equiv dF(\varepsilon;\bar{\varepsilon}), \tag{4.4a}$$

and

$$\frac{I(z_{t+\Delta t})\frac{f(\varepsilon)}{F(\bar{\varepsilon})}}{E(I)} \equiv dF_x(\varepsilon;\bar{\varepsilon}). \tag{4.4b}$$

We shall examine the weak convergence of the stock return process given by (3.1) when the error term is distributed according to the distribution described in (4.1) with the distribution of ε truncated at a value $\bar{\varepsilon} \le \varepsilon_{\max}$; this is the return process described by the term $X_{t+\Delta t}$ below. Note that here the returns are Markovian but not iid because of the dependence of the distribution on x, implying that the time subscript cannot be omitted. The convergence criterion is the Lindeberg condition, which is a necessary and sufficient condition for the convergence to a diffusion. According to this condition, the limit of the expectation of any bounded continuous function is equal to the expectation of the function with the limiting distribution.

The Lindeberg condition stipulates that, if X_t denotes a discrete multidimensional 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

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|Y-X\| \ge \delta} Q_{\Delta t}(X, dY) = 0, \tag{4.5}$$

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

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|Y-X\|<\delta} (Y_i - X_i) Q_{\Delta t}(X, dY) = \mu_i(X) \tag{4.6}$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|Y-X\|<\delta} (Y_i - X_i)(Y_j - X_j) Q_{\Delta t}(X, dY) = \sigma_{ij}(X), \tag{4.7}$$

where the symbol $\|.\|$ denotes the Euclidean norm of a vector. In our case $X_t = 1$, $X_{t+\Delta t} = \dfrac{S_{t+\Delta t}}{S_t} = z_{t+\Delta t}$, and (4.5) can be demonstrated very simply, by the same proof as in Lemma 1 of Oancea-Perrakis (2007), itself an adaptation of Merton (1982). Denote $Q_t(\delta)$ the conditional probability that $|X_{t+\Delta t} - X_t| > \delta$, given the information available at time t, namely given $X(S_t, t)$. Since ε is bounded, define $\hat{\varepsilon} = \max|\varepsilon| = \max(|\varepsilon_{\min}|, |\overline{\varepsilon}|)$. For any $\delta > 0$, define $h(\delta)$ as the solution of the equation

$$\delta = \mu h + \sigma\hat{\varepsilon}\sqrt{h}. \tag{4.8}$$

This equation admits a positive solution

$$\sqrt{h} = \frac{-\sigma\hat{\varepsilon} + \sqrt{\sigma^2\hat{\varepsilon}^2 + 4\mu\delta}}{\mu}. \tag{4.9}$$

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

$$|X_{t+\Delta t} - X_t| = |\mu\Delta t + \sigma\varepsilon\sqrt{\Delta t}| < \mu h + \sigma\hat{\varepsilon}\sqrt{h} = \delta \tag{4.10}$$

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

$$Q_t(\delta) = \Pr(|X_{t+\Delta t} - X_t|) > \delta \equiv 0 \text{ whenever } \Delta t < h \tag{4.11}$$

and hence $\lim_{\Delta t \to 0} \frac{1}{\Delta t} Q_t(\delta) = 0$. Hence, the limit of the stock return process for ε distributed according to (4.1), with ε truncated at $\bar{\varepsilon}$ is a diffusion of the form

$$\frac{dS_t}{S_t} = \mu(S_t)dt + \sigma(S_t)dW. \tag{4.12}$$

We use the subscript x to denote expectations under (4.4b), with the absence of a subscript denoting expectation under (4.4a). The value of x is given by equation (5.24) of CP (2002). The successive returns under which expectations are taken and which are defined in (4.13abc) below are Markovian but not iid.

We set $X_t = 1$, $X_{t+\Delta t} = \frac{S_{t+\Delta t}}{S_t}$ and we use the following definition of the return $X_{t+\Delta t}$:

$$X_{t+\Delta t} \equiv X^1_{t+\Delta t} + X^2_{t+\Delta t} = \frac{z \,|\, \varepsilon \le \bar{\varepsilon}}{(1-k)E(I)}, \tag{4.13a}$$

$$X^1_{t+\Delta t} = \frac{z \,|\, \varepsilon \le \bar{\varepsilon}}{(1+k)E(I)} \text{ for } \varepsilon \le \varepsilon_x, = \frac{z \,|\, \varepsilon \le \bar{\varepsilon}}{(1-k)E(I)} \text{ for } \varepsilon_x \le \varepsilon \le \bar{\varepsilon}, \tag{4.13b}$$

$$X^2_{t+\Delta t} = \varphi(k)\frac{z \,|\, \varepsilon \le \bar{\varepsilon}}{(1+k)E(I)} \text{ for } \varepsilon \le \varepsilon_x, \ 0 \text{ for } \varepsilon > \varepsilon_x. \tag{4.13c}$$

The call lower bound $\underline{C}_5(S_t, t)$ in the presence of transaction costs is greater than or equal to[55] the following recursive expression for $t < T-1$

$$\underline{C}_5(S_t,t) \ge \frac{E_x\left[\underline{C}_5\left(S_t X^1_{t+\Delta t}, t+1\right) \,|\, \varepsilon \le \bar{\varepsilon}\right]}{R} + \frac{2k}{1-k}\frac{E_x\left[\underline{C}_5\left(S_t X^1_{t+\Delta t}, t+1\right)1_{\varepsilon \le \varepsilon_x} \,|\, \varepsilon \le \bar{\varepsilon}\right]}{R}, \tag{4.14}$$

[55] The extra term in (5.19) of CP (2002) over and above the first term in the RHS of (4.14) can be easily shown to be smaller than or equal to the second term in the RHS of (4.14).

where the indicator function $1_{\varepsilon\le\varepsilon_x}$ denotes a quantity that is equal to 0 when $\varepsilon > \varepsilon_x$. The RHS of (4.14) can be easily shown to be equal to

$$\underline{C}_5(S_t,t) \ge \frac{E\left[\underline{C}_5\left(S_t\frac{z}{(1-k)E(I)},t+1\right)\middle|\varepsilon\le\bar{\varepsilon}\right]}{R} = \frac{E\left[\underline{C}_5(S_tX_{t+\Delta t},t+1)\middle|\varepsilon\le\bar{\varepsilon}\right]}{R} \tag{4.15}$$

for any $t \le T-2$. Further, the truncation of the return distribution is defined by the equation

$$\frac{E\left[\varepsilon\middle|\varepsilon\le\bar{\varepsilon}\right]}{(1-k)E(I)} = \frac{1+\mu\Delta t+\sigma E\left[\varepsilon\middle|\varepsilon\le\bar{\varepsilon}\right]\sqrt{\Delta t}}{(1-k)E(I)} = R = 1+r\Delta t \tag{4.16}$$

We rewrite (4.16) as follows

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

(4.16a)

It is easy to see that $A(\Delta t)$ is at most $O\left(\sqrt{\Delta t}\right)$. Further, we note from (4.3) and (4.16) that *for any given* Δt the relation (4.16) defines a range of values for $\bar{\varepsilon}$ as a *decreasing* function $\bar{\varepsilon}(\varepsilon_x)$ of the values that ε_x takes within its own range,[56] $\varepsilon_x \in [\varepsilon_{\min}, \bar{\varepsilon})$. For $\varepsilon_x = \varepsilon_{\min}$ we have $\bar{\varepsilon}(\varepsilon_{\min}) = \bar{\varepsilon}^*(\Delta t)$ and for $\varepsilon_x = \bar{\varepsilon}$ we have $\bar{\varepsilon}(\varepsilon_x) = \bar{\varepsilon}_*(\Delta t)$, where $\bar{\varepsilon}^*$ and $\bar{\varepsilon}_*$ are defined from the relations:

[56] Note that the function $\bar{\varepsilon}(\varepsilon_x)$ may not exist for the upper limit of the range of ε_x: (4.17b) may have no solution $\bar{\varepsilon}_* \in [\varepsilon_{\min}, \varepsilon_{\max}]$ for sufficiently small Δt. By contrast, (4.17a) has a solution and $\bar{\varepsilon}^*(\Delta t)$ always exists for Δt in an open neighborhood of 0^+.

$$1+\mu\Delta t+\sigma E\left[\varepsilon \middle| \varepsilon \le \overline{\varepsilon}^{*}(\Delta t)\right]\sqrt{\Delta t}=1+r\Delta t \ , \tag{4.17a}$$

$$1+\mu\Delta t+\sigma E\left[\varepsilon \middle| \varepsilon \le \overline{\varepsilon}_{*}(\Delta t)\right]\sqrt{\Delta t}=\varphi(k)(1+r\Delta t). \tag{4.17b}$$

From (4.16) we also get that

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \frac{1}{\Delta t}\left[\frac{1+\mu\Delta t+\sigma E\left[\varepsilon \middle| \varepsilon \le \overline{\varepsilon}\right]\sqrt{\Delta t}}{(1-k)E(I)}-1\right]=r\,, \tag{4.18}$$

implying that the process governing the stock return $X_{t+\Delta t}$ given by (4.13a) that enters into the recursive relation defining the lower bound (4.15) is risk neutral by construction for any $t \le T-2$. At *T-1* we have

$$\underline{C}_5(Z_{T-1},T-1)=\frac{E\left[(Z_{T-1}z-K)^{+} \middle| \varepsilon \le \overline{\varepsilon}\right]}{R}, \tag{4.19}$$

where now the truncation of the return is defined by the equation

$$E\left[z \middle| \varepsilon \le Z_{T-1}z-K\right]=1+\mu\Delta t+\sigma E\left[\varepsilon \middle| \varepsilon \le Z_{T-1}z-K\right]\sqrt{\Delta t}=\varphi(k)R\,, \tag{4.20}$$

which is the same as (4.17b). Observe that by Jensen's inequality we have

$$\underline{C}_5(Z_{T-1},T-1)\ge\left(\varphi(k)\,Z_{T-1}-\frac{K}{R}\right)^{+}. \tag{4.21}$$

This last result is important for numerical derivation of the bound since using it eliminates the necessity of solving for (4.20). Also, using the RHS of (4.21) as the lower bound at

$T-1$ implies $g_{T-1}(Z_{T-1})$ equal to 1 for the RHS of (4.21) greater than 0, 0 otherwise. This result may be demonstrated very simply by substituting the RHS of (4.21) for $\underline{C}_5(Z_{T-1}, T-1)$ in (3.5) and taking the resulting expression to the limit.

For $X_{t+\Delta t}$ given by (4.13abc) and for the distribution of ε given by (4.4a) it can be shown by applying the Lindeberg condition as in Proposition 3 that X_t converges weakly to a diffusion as $\Delta t \to 0$. The diffusion is of the form (4.12), with time- and state-dependent parameters. We wish to establish bounds on these parameters of the limiting process by applying (4.6) and (4.7). Since by (4.18) the mean of the process is the riskless rate by construction, it suffices to apply (4.7) to the process given by (4.13abc) to find the variance. We have, if $E(X_{t+\Delta t})$ denotes the expectation given by (4.16) and we neglect the terms $o(\Delta t)$,

$$\begin{aligned}&\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|Y-X\|<\delta} (y_i - E(X_{t+\Delta t}) + E(X_{t+\Delta t}) - x_i)^2 Q_{\Delta t}(X, dY) \\ &= \sigma^2 \lim_{\substack{\Delta t \to 0 \\ \|Y-X\|<\delta}} \frac{E\left[\varepsilon^2 \middle| \varepsilon \le \bar{\varepsilon}\right] - \left(E\left[\varepsilon \middle| \varepsilon \le \bar{\varepsilon}\right]\right)^2}{[(1-k)E(I)]^2}\end{aligned} \quad . \tag{4.22}$$

To evaluate the limit in the RHS of (4.22) we first prove the following result.

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

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t \to 0}} \frac{\sigma E\left[\varepsilon \middle| \varepsilon \le \bar{\varepsilon}\right]}{(1-k)E(I)} = \lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t \to 0}} \sigma E\left[\varepsilon \middle| \varepsilon \le \bar{\varepsilon}^*(\Delta t)\right]. \tag{4.23}$$

<u>Proof</u>: Since (4.16a) shows that $A(\Delta t)$ is at most $O(\sqrt{\Delta t})$, we have, using the definition of $A(\Delta t)$

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} A(\Delta t) = \lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \frac{1}{(1-k)E(I)} \frac{2k}{1+k} \frac{G(\varepsilon_x)}{G(\bar{\varepsilon})} = 0 . \tag{4.24}$$

Since $\bar{\varepsilon}(\varepsilon_x)$ is a decreasing function (4.24) implies that $\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \frac{G(\varepsilon_x)}{G\left(\bar{\varepsilon}(\varepsilon_x)\right)} = 0$ is possible only if

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} (\varepsilon_x) = \varepsilon_{\min} , \tag{4.25a}$$

in which case

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \bar{\varepsilon}(\varepsilon_x) = \lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \bar{\varepsilon}^*(\Delta t) . \tag{4.25b}$$

Dividing now both sides of (4.16a) by $\sqrt{\Delta t}$ and passing to the limit, we observe that $\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \left\{ \frac{A(\Delta t)}{\sqrt{\Delta t}} + \frac{\sigma E\left[\varepsilon \middle| \varepsilon \le \bar{\varepsilon}\right]}{(1-k)E(I)} \right\} = 0$. 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 $\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} \frac{A(\Delta t)}{\sqrt{\Delta t}} = \Lambda \ge 0$ and by (4.24) and (4.25ab) the denominator in the LHS of (4.23) becomes equal to 1 and the limit satisfies (4.23), QED.

To find the limit in the RHS of (4.23) we consider the definition of $\bar{\varepsilon}^*(\Delta t)$ in (4.19). We have the following result, whose proof is an alternative to the one of Proposition 2 in Oancea-Perrakis (2007).

Lemma 2: We have

$$\lim_{\substack{|X_{t+\Delta t}-X_t|<\delta \\ \Delta t\to 0}} E\left[\varepsilon \middle| \varepsilon \le \bar{\varepsilon}^*(\Delta t)\right] = E[\varepsilon] = 0 . \tag{4.26}$$

Proof: From (4.17a) we have $E\left[\varepsilon \middle| \varepsilon \le \bar{\varepsilon}^*(\Delta t)\right] = -\frac{(\mu - r)}{\sigma}\sqrt{\Delta t}$, implying that for any $\Delta t > 0$ we have

$$\frac{d\left(E\left[\varepsilon \middle| \varepsilon \le \bar{\varepsilon}^*(\Delta t)\right]\right)}{d(\Delta t)} = \frac{d\left(E\left[\varepsilon \middle| \varepsilon \le \bar{\varepsilon}^*(\Delta t)\right]\right)}{d\left(\bar{\varepsilon}^*(\Delta t)\right)} \frac{d\left(\bar{\varepsilon}^*(\Delta t)\right)}{d(\Delta t)} < 0. \tag{4.27}$$

The first derivative in the RHS of (4.27) is clearly positive, which implies that the second one must be negative, in which case

$$\lim_{\substack{|X_{t+\Delta t} - X_t| < \delta \\ \Delta t \to 0}} \bar{\varepsilon}^*(\Delta t) = \varepsilon_{\max}, \tag{4.28}$$

thus proving (4.26), QED.

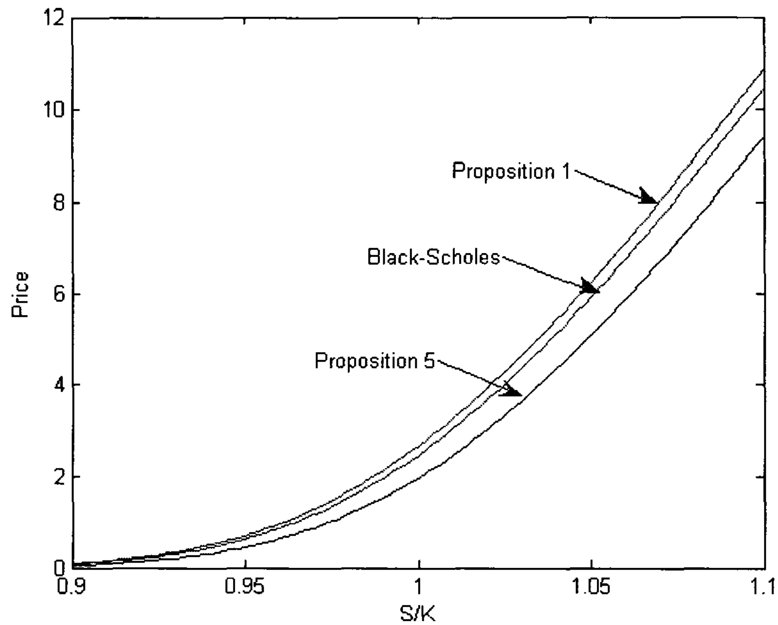
Applying now (4.23) and (4.28) to the RHS of (4.22) we see that the limit becomes equal to 1, implying that the variance of the limiting process of $X_{t+\Delta t}$ tends to σ^2, the variance of the limit of the original process (4.1). We have thus shown that the process X_t entering into the option lower bound (4.15) converges weakly to a diffusion as $\Delta t \to 0$ with time- and state-independent parameters for all $t \le T-1$. Applying (4.15) and using (4.21) as well as the law of iterated expectations, we get

$$\underline{C}_5(S_t, t) \ge \frac{1}{R^{T-1}} E\left[\left(\varphi(k) S_{T-1} - K\right)^+ \middle| S_t\right]. \tag{4.29}$$

Since we showed that the paths generated by the transformed stock returns $X_{t+\Delta t}$ given by (4.13a) tend to a risk neutral diffusion with volatility σ, it follows that by the definition of weak convergence the expectation in the RHS of (33), $\lim_{\Delta t \to 0} \underline{C}_5(S_t, t)$, becomes a Black-Scholes expression, with volatility equal to that of the original process, but with the stock price multiplied by the factor $\frac{1-k}{1+k}$.

Figure 3 displays the limiting values of the Proposition 5 lower bound for the following parameters: $K=100$, $\sigma=20\%$, $\mu=8\%$, $r=4\%$, $T=30$, $k=0.5\%$, stock price range 90-110. Combined with a CP (2002) Proposition 1 upper bound[57], this figure presents as tight a spread as it may be feasible to achieve. In the following section we present our numerical approach which will be used to demonstrate the convergence of the bound to its continuous-time limit.

Figure 3: Continuous-time Limit Results for Proposition 5



The figure displays the continuous-time limit of the CP (2002) Proposition 5 call lower bound (4.19). These results are compared to the CP (2002) Proposition 1 call upper bound and to the Black-Scholes price. The parameter are as follows: $K=100$, $\sigma=20\%$, $\mu=8\%$, $r=4\%$, $T=30$ days, $k=0.5\%$.

[57] This upper bound is derived by taking expectations of the terminal payoff under the physical measure and multiplying them by a factor $\dfrac{1+k}{1-k}$.

5 Numerical Approach

Before presenting our numerical approach, we establish several preliminary results. In this section, we also specialize the distribution of stock returns (3.2) to the uniform one. Last, we present a special case of the bound with the transaction costs rate k set equal to 0, in which case the bound converges to the Black-Scholes price.

The generic numerical solution for the bound presented in Section 3 may be simplified by noting that (3.9) is the first order condition (FOC) for maximizing the lower bound (3.6), which we demonstrate below. Therefore, in the numerical work we may skip altogether the step of deriving (3.8) and (3.9) for candidate values of x and search instead for the maximum of (3.6) in x. Once this maximum is found, we have all the ingredients for the determination of the g-function by (3.8). It is apparent that we need to discretize the candidate x's even for this simplified program, since $\underline{C}_5(Z_t,t)$ can only be derived numerically. We detail this discretization later on in this section.

It will convenient to represent (3.6)-(3.7) in integral form. By applying the definition of the I-function (3.1) to (3.7), taking expectations in integral form and simplifying we have:

$$\frac{(1+k)\int_{\underline{z}}^{\hat{z}} zf(z)\,dz}{(1+k)\int_{\underline{z}}^{\hat{z}} f(z)\,dz-2k\int_{\underline{z}}^{x} f(z)\,dz}=R, \tag{5.1}$$

where $\underline{z}$ is the lower support of the one-period distribution and $f(z)$ is the density of the one-period stock return distribution. By differentiating (5.1) with respect to x, we have:

$$\hat{z}'=-\frac{2k}{1+k}\frac{f(x)}{f(\hat{z})}\frac{R}{\hat{z}-R}. \tag{5.2}$$

Notice that since $\hat{z}>R$ the sign of the derivative in (5.2) is strictly negative, which is consistent with the results of Section 4.

We may now prove the following result:

Lemma 3: Equation (3.9) denotes the FOC for the maximization of (3.6) with respect to x.

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

$$\underline{C}_5(Z_t,t)=\frac{(1+k)\int_{\underline{z}}^{\hat{z}}\underline{C}_5(Z_t z)f(z)dz-2k\int_{\underline{z}}^{x}\underline{C}_5(Z_t z)f(z)dz+2k\varphi(k)Z_t\int_{\underline{z}}^{x}g(z)zf(z)dz}{R\left[(1+k)\int_{\underline{z}}^{\hat{z}}f(z)dz-2k\int_{\underline{z}}^{x}f(z)dz\right]},\tag{5.3}$$

where we suppressed the time arguments in the RHS of (5.3) to simplify the notation. With the use of (5.1), the denominator of (5.3) may be simplified to $(1+k)\int_{\underline{z}}^{\hat{z}}zf(z)dz$. By denoting by N and D respectively the numerator and denominator of (5.3), it follows:

$$\frac{d\underline{C}_5(Z_t,t)}{dx}=\frac{N'D-ND'}{D^2}=\frac{N'}{D}-\frac{\underline{C}_5(Z_t,t)D'}{D}.\tag{5.4}$$

By equating (5.4) to zero and rearranging, we have the FOC as $\underline{C}_5(Z_t,t)=N'/D'$. From (5.3) we get:

$$N'=(1+k)\hat{z}'\underline{C}_5(Z_t\hat{z})f(\hat{z})-2k\underline{C}_5(Z_t x)f(x)+2k\varphi(k)Z_t xg(x)f(x),\tag{5.5}$$

and

$$D'=(1+k)\hat{z}'\hat{z}f(\hat{z}).\tag{5.6}$$

By substituting for $\hat{z}'$ from (5.2) and simplifying, we arrive at the following FOC:

$$\underline{C}_5(Z_t,t)=\frac{\underline{C}_5(Z_t\hat{z})}{\hat{z}}-\left(\frac{1}{R}-\frac{1}{\hat{z}}\right)\left[\varphi(k)g(x)Z_tx-\underline{C}_5(Z_tx)\right]. \tag{5.7}$$

The same condition as (5.7) may be derived by substituting for $g_t(Z_t)$ from (3.8) into (3.9) and rearranging, which demonstrates that (3.9) is the FOC for maximizing (3.6) or (5.3), *QED*.

Now we present the limiting result for the g-function (3.8). The proof can be found directly from (3.8). Replacing for $\hat{z}$ and R from (3.2) we have

$$\lim_{\Delta t\to 0} g_t(Z_t)=\lim_{\Delta t\to 0}\frac{\underline{C}_5\left(Z_t\left(1+\mu\Delta t+\sigma\overline{\varepsilon}\sqrt{\Delta t}\right),t+\Delta t\right)-\left(1+r\Delta t+o(\Delta t)\right)\underline{C}_5(Z_t,t)}{\varphi(k)\left[(\mu-r)\Delta t+\sigma\overline{\varepsilon}\sqrt{\Delta t}+o(\Delta t)\right]Z_t} \tag{5.8}$$

Since both numerator and denominator tend to 0, we take their derivative with respect to $\sqrt{\Delta t}$. It is then easy to see that (5.8) tends to $N\left(d_1^*\right)$, where $d_1^*=d_1\left(\varphi(k)S,.\right)$ since the stochastic process for the underlying tends to the risk-neutral diffusion with $\varphi(k)S$ replacing the stock price by the results of Section 4. Recall that in our notation Z_t represents time-t price of the underlying.

The numerical evaluation of equation (3.6) or (5.3) presents major challenges as the time step Δt becomes progressively smaller. Although the limit of (3.6) as $\Delta t\to 0$ is, of course, trivial to find as demonstrated in Section 4, the values of the option and $g_t(Z_t)$ for any discrete partition of the time interval to option expiration can only be found by recursive numerical methods. These values are of interest, not only for the evaluation of European options, but also for the verification of the speed of convergence, given that continuous trading may in fact be infeasible in practice. The numerical methods that we develop in the following section address the problems of accuracy, in the integration in evaluating (5.3) and in the solution of equations (3.8) and (3.9) or (5.7), as the number of time steps increases and the sizes of the returns for integration and numerical solutions become progressively smaller.

Having shown that the condition (4) is the FOC maximizing (5.3), we may describe in general terms how to derive the bound in numerical work. Our general setup implied by the compact-support process for the underlying return is a discrete-time continuous-state framework. For notational convenience we use the symbol t to denote 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. It follows that the underlying return Z_t spans $[\underline{z}^t, \overline{z}^t]$ at the epoch t. A natural method to work our computations backward in time is to use recursive numerical integration[58]. 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 a difficult task. We solve this problem by a log-transformation of the state variable. We elaborate below on this transformation once we specialize the distribution for the error term ε.

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 $I_0\left(h_{t+1}(y)\right)$ the time-t integral of the $t+1$ function h of a random variable y with the one-period support $[\underline{y}, \overline{y}]$. It follows:

$$I_0\left(h_{t+1}(y)\right) \cong \sum_{j=0}^{L} \tilde{w}_j \Delta y f\left(\underline{y} + j\Delta y\right) h_{t+1}\left(\underline{y} + j\Delta y\right) \equiv \sum_{j=0}^{L} w_j h_{t+1}\left(\underline{y} + j\Delta y\right), \tag{5.9}$$

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 $L\Delta y + \underline{y} \leq \overline{y}$, and w_j is the redefined

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

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 (5.3). Notice that for $L_{\max}$ defined by $L_{\max}\Delta y + \underline{y} = \overline{y}$ in (5.9) 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 $\underline{y} + j\Delta y$ 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[59], with the lowest (highest) increment $\underline{y}$ ($\overline{y}$) satisfying $\underline{y} = \log(\underline{z})$ ($\overline{y} = \log(\overline{z})$). It follows that at the epoch t the log-return is spaced by Δy over a segment $[t\underline{y},\ t\overline{y}]$ 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[60] in the lattice, m states (nodes) spaced by Δy over $[Y_t + \underline{y},\ Y_t + \overline{y}]$ may be reached in the subsequent epoch, and while going backward in the lattice we may easily compute integrals as in (5.9).

We are not ready yet to numerically derive the lower bound (5.3). Even if we limit ourselves to maximizing (5.3) 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. Our preliminary work indicates that approximating the $\hat{z}$'s with the points in the grid closest to the true value of $\hat{z}$ would not yield satisfactory results. To circumvent this problem, we use a non-linear interpolating function, the piece-wise Hermite polynomials.[61] This function, as opposed to the perhaps more widely used splines, has the desirable property of preserving the monotonicity of the data. Our preliminary work shows that the Hermite polynomials yield indeed excellent results when applied to the smooth functions we expect in our work.

---

[59] 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.

[60] 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.

[61] See Fritsch and Carlson (1980).

To delineate the search domain, let $x_{\max}$ denote the maximum feasible x, i.e. the one corresponding to $\hat{z} = R$ in (5.1) and let $\hat{z}_{\max}$ denote the maximum feasible $\hat{z}$, i.e. the one corresponding to $x = \underline{z}$ in (5.1). Before presenting the critical steps in our algorithm, we define two numerical integrals for a given node characterized by the log-return Y_t:

$$I_1(L_i) = \sum_{j=0}^{L_i} w_j \underline{C}_5(\underline{y} + j\Delta y),\ L_i = 1 \ldots L_{\hat{z}}$$

and , (5.10)

$$I_2(L_i) = \sum_{j=0}^{L_i} w_j \left[\varphi(k) \exp(Y_t + \underline{y} + j\Delta y) g(\underline{y} + j\Delta y) - \underline{C}_5(\underline{y} + j\Delta y) \right],\ L_i = 1 \ldots L_x$$

where the weights w_j are like in (5.9), we suppressed the time-t state argument Y_t from $\underline{C}_5(.)$ and $g(.)$ to simplify the notation, $L_{\hat{z}}\Delta y + \underline{y} > \log(\hat{z}_{\max})$ and $L_x \Delta y + \underline{y} > \log(x_{\max})$. 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 $\hat{z}$ is greater than or equal to the maximum value of x, we need to use $L_{\hat{z}} + 1$ weights w_j in our numerical work.

The following are the critical steps of our algorithm:

1) Fix a set of pairs $(x_s,\ \hat{z}_s)$, $s = 1 \ldots n$ linked by (5.1) and spanning the feasible region for x, which is $[\underline{z},\ x_{\max}]$. 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 (5.3) 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 (5.10) 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 $\underline{C}_5(Z_t\hat{z},t+1)$ interpolated in log-scale as $\underline{C}_5\left(Y_t+\log(\hat{z}_{s*}),t+1\right)$.
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 (4.21) on $\underline{C}_5(Y_{T-1},T-1)$ at $t=T-1$ with the corresponding result at this time epoch for the g-function, i.e. we set $g_{T-1}=1$ for the nodes with $\underline{C}_5(Y_{T-1},T-1)>0$, 0 otherwise.

In our numerical work we use the tree size $m=251$ with $n=250$ candidate values for x located in $[\underline{z},\ x_{\max}]$. 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).[62] Here we exemplify 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:

$$I_1^M(L_i)=\mathrm{IFFT}\left[\mathrm{FFT}\left(\underline{C}_5^M(t+1)\right)\times\mathrm{FFT}\left(MW_{L_i}^M\right)\right], \tag{5.11}$$

[62] Appendix B of the first essay of this thesis details applications of FFT to recursive computations applicable to option pricing.

where M denotes the number of nodes at the epoch $t+1$, $M=(t+1)(m-1)+1$, FFT denotes a Fast Fourier Transform, IFFT denotes the inverse of FFT, $I_1^M(L_i)$ and $\underline{C}_5^M(t+1)$ are vectors of length M with the latter vector representing the lower bound values at all nodes at the epoch $t+1$, $\times$ denotes a vector element-by-element multiplication, $W_{L_i}^M$ is a vector $[w_0 \ldots w_{L_i}]'$ padded down with zeros to the length M. The first and last $(m-1)/2$ entries to the vector $I_1^M(L_i)$ should be discarded since FFT (IFFT) applied to a vector yields a vector of the same length. $I_2^M(L_i)$ easily follows by an appropriate substitution in (5.11).

To derive the integration weights $\tilde{w}_j$ as in equation (5.9), we apply the Newton-Cotes composite rules[63]. 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.

Setting abscissas at equally spaced points without regard to the true zero of the integrand may cause an integration error.[64] On the other hand, relating the abscissas to this zero point may destroy our integration scheme. Consider a state for which the zero of the integrand may be reached in one period. To avoid the integration error, we might set equidistant abscissas from this zero point till the upper integration limit. Unfortunately, the implied integration step in general will not result in an integer number of abscissas whenever the integrand does not cross zero in one period. In a later section we demonstrate that disregarding the above problem results in only insignificant numerical errors.

In our numerical work, we use the uniform distribution as the compact support process for the return of the underlying. For the one-period stock return process z $(\equiv Z_t / Z_{t-1})$ we set:

$$z = 1 + \mu\Delta t + \varepsilon\sigma\sqrt{\Delta t}\,, \tag{5.12}$$

[63] For Newton-Cotes integration see, for instance, Davis and Rabinowitz (1966).

[64] This error may arise since in general the true zero remains between abscissas while the Newton-Cotes integration assumes that the zero is placed exactly one abscissa below the first non-zero value of the integrand.

where the error terms ε are iid uniformly distributed with zero mean and unit variance. These last two conditions imply the following error density:

$$f(\varepsilon)=\begin{cases}1/2\sqrt{3}, & \varepsilon\in\left[-\sqrt{3},\sqrt{3}\right]\\ 0 \text{ otherwise}\end{cases}, \tag{5.13}$$

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

$$f(z)=\begin{cases}1/2\sqrt{3}\sigma\sqrt{\Delta t}, & z\in\left[\underline{z},\,\overline{z}\right]\\ 0 \text{ otherwise}\end{cases}, \tag{5.14}$$

where $\underline{z},\ (\overline{z})=1+\mu\Delta t-(+)\sqrt{3}\sigma\sqrt{\Delta t}$.

For the uniformly distributed disturbances, there exists a closed-form solution for $\hat{z}$ in the equation (5.1) for a given x. Integrating (5.1) under the uniform density and rearranging yields the following second-order polynomial in $\hat{z}$:

$$\hat{z}^2-2R\hat{z}+c(x)=0, \tag{5.15}$$

where $c(x)=2R\left(\varphi(k)\underline{z}+\alpha(k)x\right)-\underline{z}^2$. The solution for $\hat{z}$ is given by the higher of the two roots of (5.15):

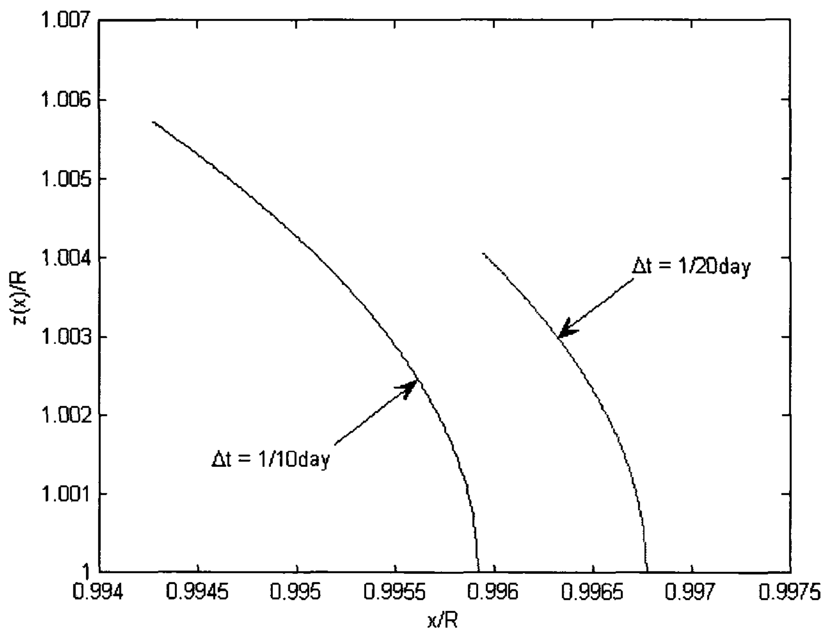
$$\hat{z}=R+\sqrt{R^2-c(x)}. \tag{5.16}$$

Another quantity of interest is the value $x_{\max}$ for which $\hat{z}$ attains its minimum in the search grid, which is R. Inverting (5.15) for this minimum $\hat{z}$ yields:

$$x_{\max}=\left[R-\left(2\varphi(k)-\underline{z}/R\right)\underline{z}\right]/2\alpha(k). \tag{5.17}$$

To demonstrate the dependence of $\hat{z}$ on x we display (5.16) in Figure 4 for Δt of 1/10 and 1/20 days. To have the two graphs comparable, we scale both dependent and independent variable by dividing them by R. Note a lower range of both x and $\hat{z}$ in terms of R for the greater time partition.

Figure 4: Behaviour of x and $\hat{z}$ in Time Partition



The figure displays the time-partition behaviour of the quantities x and $\hat{z}$ as defined by equation (5.1) and derived by equation (5.15) for the uniformly distributed disturbances in (5.12). The displayed quantities were normalized by the riskless return respective to each time partition. The parameter are as follows: $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $k = 0.5\%$.

As noted above, for our generalized lattice we need to perform a log-transformation of the state variable, which necessitates adjusting the form of the density function. We derive the density of the logarithm of the one-period return by standard statistical arguments for a monotonic transformation of a random variable. For $y \equiv \log(z)$, with z distributed uniformly as in (5.14), we have the following density function:

$$f(y) = \begin{cases} \exp(y)/2\sqrt{3}\sigma\sqrt{\Delta t}, \; y \in \left[\log(\underline{z}), \log(\overline{z})\right] \\ 0 \text{ otherwise} \end{cases} . \tag{5.18}$$

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:

$$w_j = \tilde{w}_j \Delta y \exp\left(\underline{y} + j\Delta y\right) / 2\sqrt{3}\sigma\sqrt{\Delta t},\ j = 0 \ldots L_{\hat{z}}. \tag{5.19}$$

Special case: k = 0. By the definition of the *I*-function (3.1), it may be easily shown that for the transaction costs parameter *k* set equal to 0 equations (3.6) and (3.7) for $t \leq T-1$ respectively collapse to:

$$\underline{C}_5(Z_t,t) = E\left[\underline{C}_5(Z_t z, t+1) \mid Z_t, z \leq \hat{z}\right] / R$$

and $\qquad$. $\qquad$ (5.20)

$$E\left[z \mid z \leq \hat{z}\right] = R$$

It may be shown that for the compact support for z the option value in (5.20) weakly converges to the Black-Scholes price since the process for the underlying converges to the risk-neutral diffusion, with the usual terminal condition at $t = T$.[65] For the uniform distribution, the solution to the second line in (5.20) is:

$$\hat{z}^* = 1 + (2r - \mu)\Delta t + \sqrt{3}\sigma\sqrt{\Delta t}\,. \tag{5.21}$$

This quantity falls close and immediately below the upper support for the distribution $\bar{z}$. To derive the price at any node at any epoch *t*, we apply (5.9) to the first line in (5.20) for several $t+1$ nodes below $\hat{z}^*$ up to the upper support of the distribution and interpolate for the value of $\underline{C}_5(Y_t,t)$ at $Y_t + \log\left(\hat{z}^*\right)$ as explained above.

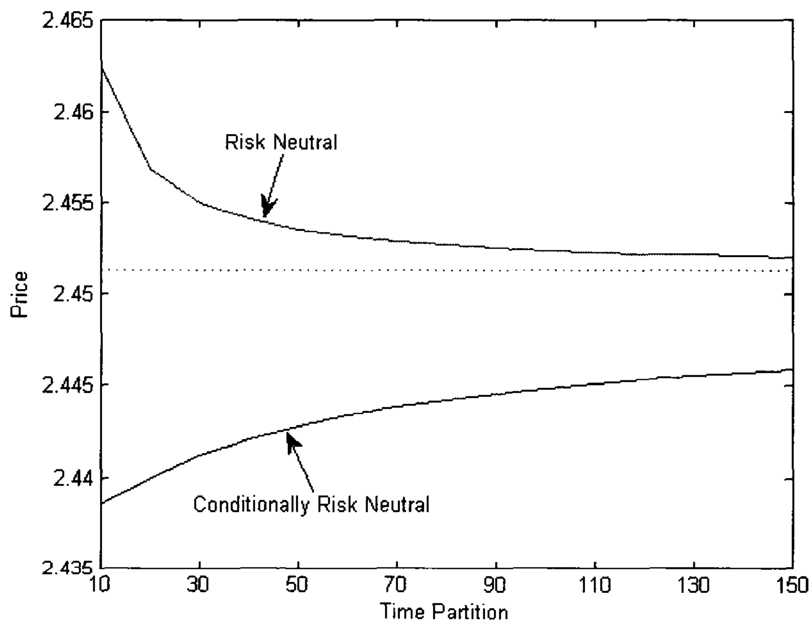
---

[65] See Proposition 2 in Oancea-Perrakis (2007).

6 Results

First, we compare the numerical results for the convergence of the expression in (5.20) with the similar results for the Black-Scholes price derived recursively under the risk-neutral distribution for the uniform support, i.e. $z = R + \varepsilon\sigma\sqrt{\Delta t}$. For clarity's sake, in this section we refer to the distribution implied by the second line of (5.20) as 'conditionally risk-neutral'. Our base case uses $S = 100$, $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$ and $T = 30$ days and the one-period lattice size of 251. Figure 5 displays the convergence under the risk-neutral and conditionally risk-neutral distributions. It is clear from the graph that for the latter distribution the convergence occurs uniformly from below and at a slower pace than for the former distribution. These findings indicate that for the lower bound the convergence will occur from below and will be substantially slower than the convergence to the Black-Scholes price under the risk-neutral distribution.

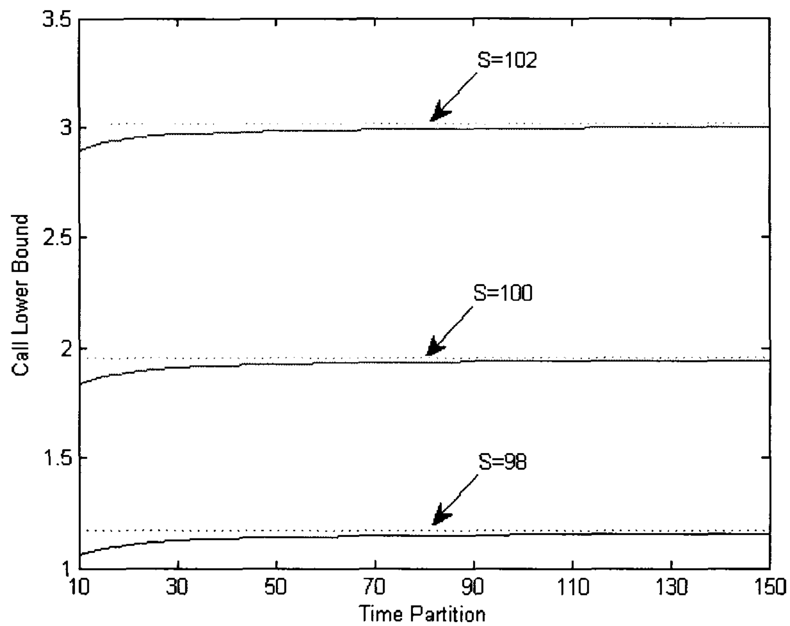
Figure 5: Convergence of the Proposition 5 Lower Bound to the Black-Scholes Price for $k = 0$



The figure compares the convergence behaviour of the Proposition 5 lower bound (3.6) (Conditionally Risk Neutral) to the convergence of recursive expectations of the terminal payoff under risk-neutral measure (Risk Neutral). Both quantities were derived for the uniform distribution of the stock returns (5.12). The parameter are as follows: $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $T = 30$ days, $k = 0$.

Now we turn to the presentation of our results for the convergence of the lower bound to its continuous time limit, which is the Black-Scholes price with $\varphi(k)S$ replacing the underlying price S. We use a one-way proportional transaction cost rate $k = 0.5\%$. Other parameters are as above except the stock price S. Figure 6 shows the convergence behaviour for three different stock prices 98, 100 and 102 with the time partition ranging from 10 to 150. In Figure 6 the numerically derived bounds clearly approach the limiting price.

Figure 6: Convergence of the Proposition 5 Lower Bound to Continuous-time Limit

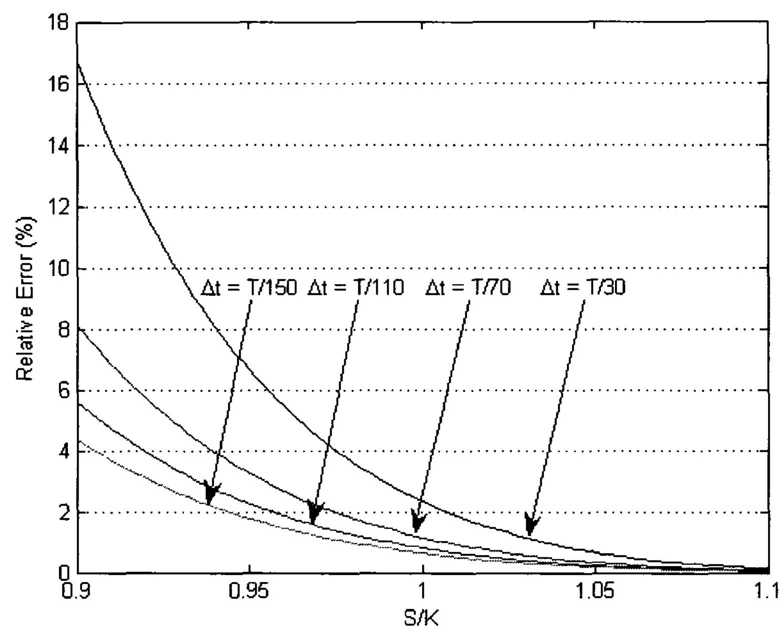


The figure displays the convergence behaviour of the Proposition 5 lower bound (3.6) to its continuous-time limit (4.19) derived for the uniform distribution of the stock returns (5.12). The parameter are as follows: $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $T = 30$ days, $k = 0.5\%$.

Given that continuous trading may in fact be infeasible in practice, it is of interest how close the lower bound falls to its limit for a 'realistic' time partition. For instance, for daily trading, i.e. 30 subdivisions for the stock prices 98, 100 and 102 we have the respective lower bounds 1.127, 1.909 and 2.967. The corresponding continuous-time limits respectively are 1.169, 1.954 and 3.011, with differences from the discrete-time values approximately equal to five cents.

We also derive relative errors of the convergence to the limit, defined as $1-\underline{C}_5/BS(\varphi(k)S,.)$. In Figure 7, we display these errors for the stock price range from 90 to 110 and for the time partitions of 30, 70, 110 and 150. It is apparent from Figure 7 that the relative errors tend to zero as the time partition becomes more dense 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.

Figure 7: Relative Convergence Errors of the Proposition 5 Lower Bound from Continuous-time Limit



The figure displays the relative convergence errors $1-\underline{C}_5/BS(\varphi(k)S,.)$ of the Proposition 5 lower bound (3.6) from its continuous-time limit (4.19) derived for the uniform distribution of the stock returns (5.12). The parameter are as follows: $K=100$, $\sigma=20\%$, $\mu=8\%$, $r=4\%$, $T=30$ days, $k=0.5\%$.

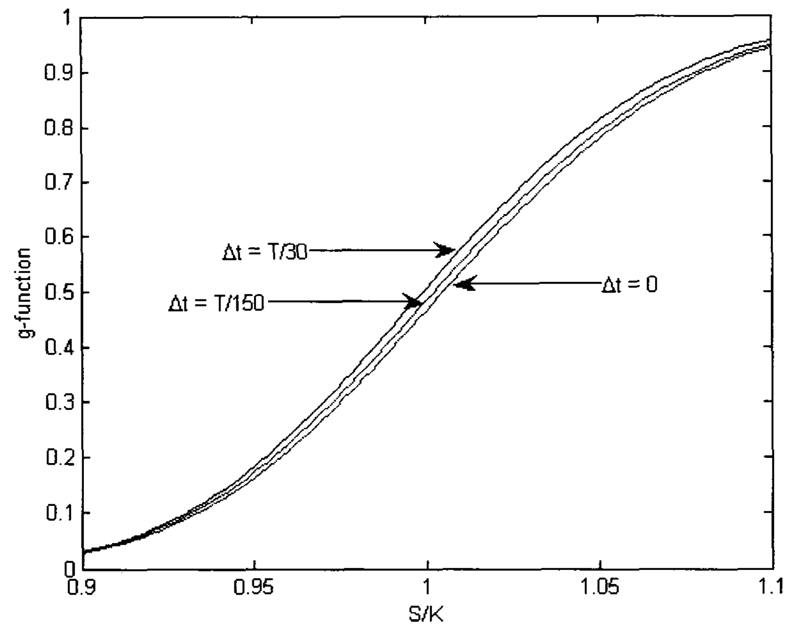
Although systematic results on dollar errors are not shown, we note that the dollar errors decrease as the density of time partition increases, as we may expect. These errors peak approximately for at-the-money options. For instance, for the time partition 150 and S = 90, 100 and 110, we find respective 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. For comparison's sake, we provide the respective Black-Scholes prices for $k=0$, which are: 0.081, 2.451 and 10.433.

Besides the convergence in time partition for a fixed size of the lattice, the convergence in the lattice size itself is of interest. It appears that the lattice size does not significantly influence the convergence of the bound. For instance, for the lattice size of 11 we noted the following results for 150 time subdivisions: 1.1581, 1.9421 and 2.9991 respectively for the stock prices of 98, 100 and 102. For the lattice size of 251, the equivalent results were 1.1577, 1.9417 and 2.9987. There is a theoretical reason for the proximity of the results for highly different lattice sizes. Recall the result (4.25a) stipulating that the value of x tends to its lower bound as the time partition gets more dense. If this is the case, the results across different lattice sizes should be similar given that the value of x is constant at $1+\mu\Delta t-\sqrt{3}\sigma\sqrt{\Delta t}$ for the majority of nodes.[66] Recall also that our algorithm relatively precisely derives integrals by interpolating for the value of $\hat{z}$. Finding that even that small lattices as the one with 11 nodes produces the correct results for the bound, strengthens the applicability of the bound in empirical research.

The last problem is the numerical verification of our result about the convergence of the g-function converges to $N\left(d_1^*\right)$, where $d_1^* = d_1\left(\varphi(k)S,.\right)$. Recall that the bound is derived by adopting the policy, whenever the call price is below the lower limit, of selling $g<1$ shares, purchasing the call option and investing the remainder of the proceeds in the riskless asset, which leads to an increase in the investor's expected utility. Figure 8 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 conjectured 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, $1-g/N\left(d_1^*\right)$ for the time partitions of 30, 70, 110 and 150 in Figure 9. These errors clearly decrease at the partition increases, with the convergence speed increasing in the S/K ratio.

[66] We verified this fact by checking which value of x maximizes the bound at each time subdivision.

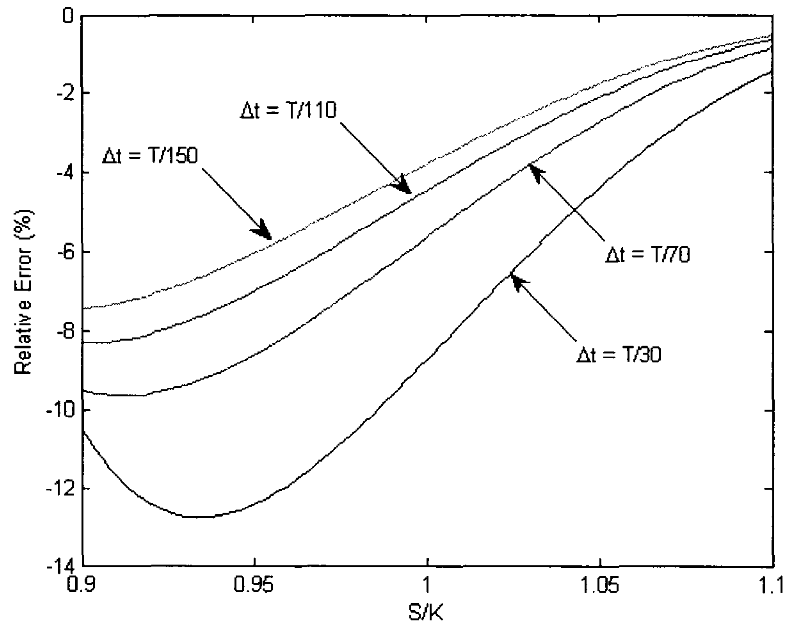
Figure 8: Convergence of the g-function to Continuous-time Limit



The figure displays the convergence behaviour of the g-function (3.8) to its continuous-time limit $N\left(d_1^*\right)$, where $d_1^* = d_1\left(\varphi(k)S,.\right)$ derived for the uniform distribution of the stock returns (5.12). The parameter are as follows: $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $T = 30$ days, $k = 0.5\%$.

Last in this section, we demonstrate that possible integration errors due to setting the lower integration limit not at the true zero of an integrand are small. All the results presented in this section were derived by setting the lower integration limit at the first non-zero value in our grid of the integrands in (5.10). It is clear that this setting results in undervaluing of the bound. It is also clear that setting the lower integration limit one integration step below the first non-zero value of integrands could overvalue the bound. The difference between the two approaches will tell the size of the integration error due to the absence of the true zero of integrands in our integration scheme. For several time partitions and lattice sizes we found this difference to be no greater than of the order of 1e-14 in dollar terms for the range of moneyness 0.9-1.1, which lets us conclude that this integration error does not affect significantly our numerical results.

Figure 9: 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 $d_1^* = d_1\left(\varphi(k)S,.\right)$ derived for the uniform distribution of the stock returns (5.12). The parameter are as follows: $K = 100$, $\sigma = 20\%$, $\mu = 8\%$, $r = 4\%$, $T = 30$ days, $k = 0.5\%$.

7 Concluding Remarks

In this essay we have derived a non-trivial continuous-time limit for a call lower bound in the presence of transaction costs, the only such convergence result available in the literature. We also showed that the discrete-time counterpart of this result converges relatively fast, which stipulates that the Proposition 5 lower bound may be applied in realistic trading conditions, i.e. when the trading frequency is relatively low. To show the convergence, we applied novel numerical methods which deal with the problem of the deriving recursive expectations under a Markovian but non-iid distribution.

Since this essay clearly demonstrated that the Proposition 5 lower bound may be applied to market data, future research may use the bound to test for the stochastic dominance efficiency of index options markets in the spirit of the empirical methodology presented in the first essay of this thesis. Future research may also extend the results of this essay to American options and options on futures contracts.

References

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

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., J. C. Jackwerth and S. Perrakis, 2007, "Mispricing of S&P 500 Index Options," *Review of Financial Studies,* forthcoming.

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., 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, 2007, "Stochastic Dominance and Option Pricing in Discrete and Continuous Time: an Alternative Paradigm," Concordia University working paper.

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.

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, 1-25.

Essay III: Portfolio Selection with Transaction Costs and Jump-Diffusion Asset Dynamics

Abstract

We derive the boundaries of the region of no transaction when the risky asset follows a mixed jump-diffusion process in the presence of proportional transaction costs. These boundaries are shown to differ from their diffusion counterparts in relation to the jump intensity for lognormally distributed jump size. A general numerical approach is presented for iid risky asset returns in discrete time. An error in earlier work on the region of no transaction for discretized diffusions is demonstrated and corrected results are presented. Comparative results with a recent study on the same topic are presented and it is shown that the numerical algorithm has equally attractive approximation properties to the unknown continuous time limit.

1 Introduction

In this essay we seek a twofold objective: first, we extend the portfolio selection model under transaction costs to a jump-diffusion process with a lognormally distributed jump component; second, we correct an error in earlier work and present corrected results for the no transaction (NT) region for diffusion in a discrete time-finite horizon framework. The derivation of the NT region obtained for diffusive processes was studied extensively in the literature. On the other hand, numerical results for jump-diffusion were only derived recently and concurrently with this work in Liu and Loewenstein (2008). This essay complements the Liu and Loewenstein (2008) continuous time results with a discrete time equivalent. We present evidence that our numerical algorithm reaches almost identical results with this latter study.

First, we introduce our problem in general terms. The investor maximizes her derived utility of consumption, be it the consumption of the entire wealth at the terminal finite date T or the consumption at all dates including the terminal date. The investor is constrained to hold two assets, a riskless bond and a risky stock, with the natural interpretation of an index. We denote the dollar holdings in the riskless bond as x and the dollar holdings in the stock as y. The investor faces proportional transaction costs at the rate k on transferring money from the stock account to the bound account and vice versa but not on liquidating her bond holdings. The choice variable of the investor at each discrete date t is the proportion of risky to riskless asset ($\lambda_t \equiv y_t/x_t$), which is a control maximizing the derived utility of consumption. When we allow for consumption at intermediate dates, the investor's problem necessitates in additional control ($\equiv c_t$), which is the optimal consumption at each discrete date t. In the discrete time setup we consider it a natural extension of our model to allow for time but not for state dependent consumption, as we explain in a later section. Such an extension will not be attempted in this paper.

Section 2 introduces our model and corrects an error in an earlier work of related scope. Section 3 presents the numerical algorithm; Section 4 presents numerical results. Section 5 summarizes and closes. In the remainder of this section we complete a literature review of the portfolio selection rules in the presence of proportional transaction costs.

Constantinides (1979) proved an earlier conjecture in Magill and Constantinides (1976) that the NT region is a cone composed of two boundaries and the optimal investment policy is simple, i.e. it consists of trading to the closest boundary of this region if the risky to riskless asset proportion falls outside the cone formed by the two boundaries.[67] Constantinides (1986) was the first to present numerical results for the NT region. This work considered an infinite horizon problem for a diffusion process. Under a simplifying assumption of state and time independent consumption, Constantinides (1986) derived the value function in a closed form; however, as the solution was composed of the value function and two first order conditions, the derivation of the NT region required numerical methods. Norman and Davies (1990) relaxed the simplifying assumption on consumption policy by considering both time and state dependent consumption and obtained a closed form solution composed of two ordinary differential equations. Their numerical results were not qualitatively different from the ones in Constantinides (1986). As opposed to Constantinides (1986) and Norman and Davies (1990), Dumas and Luciano (1991) considered a portfolio choice of an investor who maximizes the derived utility of consumption taking place upon the liquidation of the portfolio holdings at some future time *T*. Dumas and Luciano (1991) considered a limiting case as the liquidation time *T* tends to infinity. They assumed the discount factor to be endogenous to the problem, i.e. they solved for the discount factor for which the partial derivative of the value function with respect to time is zero. The results in this work differed from those of Constantinides (1986) first, in that the NT region was found to be considerably wider; second, no shift towards the riskless asset was found for increases in the transaction cost rate.

For frictionless markets, Liu *at al* (2003) considered portfolio rules for a wide class of jump processes. This work provided numerical results in a special case of jump-diffusion, for a fixed jump size. The portfolio rule was far apart from its diffusion counterpart, which is the Merton (1971) line.[68] However, this result was obtained under the

---

[67] This result was proven in fairly general settings: not necessarily Markovian risky asset returns, additively or multiplicatively separable utility, transaction costs function positively homogenous of degree one in the investment decision, possibly adapted process for the bond account, the presence of dividends, finite or infinite investment horizon. See Propositions 5 and 7 in Constantinides (1979).

[68] The Merton line is an optimal risky to riskless asset proportion equal to $a^*/(1-a^*)$, with $a^* = (\mu - r)/(1-\delta)\sigma^2$, the ratio of the risk premium to variance of the risky asset, for an infinite horizon and frictionless trading for an agent with power utility and relative risk aversion δ.

condition that in the presence of a jump component the total volatility of the risky asset increases by the volatility of this component. This is a less realistic setup than when the total volatility is decomposed into the diffusion and jump components, the case that we consider in our work. Similarly to Liu *at al* (2003), we also compensate the mean of the diffusion component by the negative of the mean of the jump counterpart.

Liu and Loewenstein (2002, 2008) considered a finite horizon problem in continuous time in the presence of transaction costs with a random terminal date, which occurs with the n-th passage of a Poisson process.[69] Since their later work provided a solution for a jump-diffusion process, we present the Liu-Loewenstein methodology in some detail. There are two serious technical problems with solving the Bellman equation in the presence of transaction costs:[70] first, there are two free boundaries varying through time; second, the time partial of the value function remains part of the equation for a fixed investment horizon. The Liu-Loewenstein methodology explores a randomization idea originally presented by Carr (1998) to produce an iterative sequence of ordinary differential equations whose successive solutions converge to yield the solutions for the value function and the boundaries of the no transaction region. Moreover, it was shown that the solution for a random terminal date converges to the solution for a fixed horizon equal to the expectation of this random quantity. In a later section we demonstrate that the numerical results derived by the methods applied in this essay mirror the numerical results in Liu and Loewenstein (2008).

[69] Under this stipulation, the horizon is Erlang distributed. Liu and Loewenstein (2002, 2008) also considered an exponentially distributed finite horizon, a case that is not relevant for the results of this essay.
[70] The Bellman equation is a natural representation for the partial differential equation of the value function for the optimal investment problem.

2 Optimal Portfolio Policy under Proportional Transaction Costs

We present the dynamics of the assets that we consider. This allows us to formulate the dynamic problem that an agent faces while undertaking investment decisions in the presence of proportional transaction costs. Last, we demonstrate an error in earlier work in the dynamic formulation for a discretized diffusion process and present correct results.

2.1

First, we present the continuous time counterparts whose discretization we consider. The bond holdings x_t follow:

$$dx_t = rx_t , \tag{2.1}$$

where r is the continuously compounded riskless rate. Our first case is the diffusion process for the stock holdings y_t:

$$dy_t = \mu y_t dt + \sigma y_t dW_t \ , \tag{2.2}$$

where W_t is a standard Gauss-Wiener process and μ, σ are its instantaneous mean and volatility parameters. In the second case we consider a mixed jump-diffusion process:

$$dy_t = \left(\mu - \eta\mu_K\right) y_t dt + \tilde{\sigma} y_t dW_t + K y_t dN_t \ , \tag{2.3}$$

where the last term is the jump component added to the diffusion. It is assumed that the jump and diffusion components are independent. The variable K represents the logarithm of the jump size and N_t is a Poisson counting process with intensity η. The volatility of the diffusion component of the stock process $\tilde{\sigma}$ is set so that the total volatility of the stock process is equal to the volatility in the pure diffusion case, which implies

$\tilde{\sigma}=\sqrt{\sigma^2-\eta\left(\sigma_K^2+\mu_K^2\right)}$. In our numerical work we assume, in line with earlier work and without loss of generality, that the jump size is distributed as $N\left(\mu_K,\sigma_K^2\right)$.

In our work we consider a discrete approximation that converges weakly to (2.3). Lemma 2 in Oancea and Perrakis (2007) demonstrates that the following return process ($\frac{y_{t+\Delta t}}{y_t}\equiv z_{t+\Delta t}$) is a valid approximation of (2.3):[71]

$$z_{t+\Delta t}=1+\left(\mu-\eta\mu_K\right)\Delta t+\tilde{\sigma}\varepsilon\sqrt{\Delta t}+K\Delta N\,, \tag{2.4}$$

where ε is a random variable with a given distribution of mean 0 and variance 1 that can be anything; for instance, the trinomial distribution that we consider in a later section. The return process can be conveniently represented as a mixture of the diffusion and jump components with corresponding probabilities $1-\eta\Delta t$ and $\eta\Delta t$:

$$z_{t+\Delta t}=\begin{cases}1+\left(\mu-\eta\mu_K\right)\Delta t+\tilde{\sigma}\varepsilon\sqrt{\Delta t} & \text{with probability } 1-\eta\Delta t\\ 1+K & \text{with probability } \eta\Delta t\end{cases}. \tag{2.5}$$

We lay out the details of our discretization in Section 3.

We present our problem for the case where there is no consumption at intermediate dates. Up to a certain point, our exposition mirrors the one in Genotte and Jung (GJ, 1994). We find an error in GJ and develop our own dynamic formulation of the problem, leading to an exact numerical solution given the approximation of the discretization of the state dynamics.

Under proportional transaction cost, the bond and stock accounts dynamics are:

[71] The approximation (2.4) converges weakly to (2.3), in the sense that the expectation of any continuous function of the random stock return at some future time taken with respect to the discrete process converges to the expectation of that same function taken with the continuous time limit of the process. This is the appropriate convergence criterion for our problem.

$$x_{t+1} = \left(x_t - v_t - k\left|v_t\right|\right) R \quad , \qquad (2.6)$$

$$y_{t+1} = \left(y_t + v_t\right) z_{t+1}$$

where v_t denotes the time-t investment decision, the dollar amount net of transaction costs by which the investor changes her risky asset account. The investor solves the following problem of maximizing the expected utility of the terminal wealth net of transaction costs:

$$\max_{v_\tau \ \forall \tau \in \{t, \ldots, T-1\}} E_t\left[U\left(x_T + (1-k)y_T,\ T\right)\right],$$

(2.7)

s. t. $x_t + y_t(1-k) > 0$ and $x_t + y_t(1+k) > 0$ (solvency constraints), where v_τ is the time-τ investment decision. The solution to the investor's problem is a pair of boundaries of the NT region. We denote the lower and upper boundaries by $\underline{\lambda}_t$ and $\overline{\lambda}_t$, respectively and by λ_t the time-t risky to riskless asset proportion. Note that the NT region is a convex subset of the solvency region characterized by the above two boundaries.[72] This effectively precludes borrowing in the case of the mixed process that we use in our numerical work, since under lognormally distributed jumps the investor will face a positive likelihood of ruin.

The most frequently used approach to solve for (2.7) is the dynamic programming formulation:

$$V\left(x_t, y_t, t\right) = \operatorname*{Max}_{v_t} \ E_t\left[V\left(x_{t+1}, y_{t+1}, t+1\right)\right] \qquad (2.8)$$

with the boundary conditions:

$$V\left(x_T, y_T, T\right) = U\left(x_T + (1-k)y_T, T\right). \qquad (2.9)$$

[72] When both x and y are positive, the solvency constraints are trivially satisfied due to limited liability. The first constraint ensures the positive net worth for borrowing, the second for selling the stock short. Under a positive risk premium and risk aversion, it is never optimal to sell short the risky asset.

The isoelastic utility function $(W_T)^\alpha/\alpha$, where $\alpha = 1-\delta$, with δ denoting the relative risk aversion (RRA) coefficient, results in a concave and homogenous of degree α in its arguments value function (2.8)-(2.9), as was shown in Constantinides (1979).

We use the dynamic programming approach to formulate the problem at hand. However, as we argue later, for the applied risky asset discrete-time dynamics and the resulting state dynamics for the problem (2.8), applying the formulation (2.6) will yield an easy to apply and precise numerical solution.

A central role in our analysis will be played by two functions deriving the indirect (not necessarily maximized) utility for purchase and sale of the risky asset, respectively $\underline{J}(.)$ and $\bar{J}(.)$, which we define as:

$$\underline{J}(x_t, y_t, v_t, t) = E_t\left[V\{(x_t-(1+k)v_t)R, (y_t+v_t)z_{t+1}, t+1\}\right]$$

and $\qquad\qquad\qquad\qquad$. $\qquad\qquad$ (2.10)

$$\bar{J}(x_t, y_t, v_t, t) = E_t\left[V\{(x_t-(1-k)v_t)R, (y_t+v_t)z_{t+1}, t+1\}\right]$$

To increase the proportion of the risky to riskless asset to some new proportion λ_t $(v_t > 0)$, the investment decision is:

$$v_t = \frac{\lambda_t x_t - y_t}{\lambda_t(1+k)+1}. \qquad (2.11)$$

Substituting this last quantity into the first line of (2.10) yields:

$$\underline{J}(x_t, y_t, \lambda_t, t) = (x_t+(1+k)y_t)^\alpha E_t\left[V\left\{\frac{R}{\lambda_t(1+k)+1}, \frac{\lambda_t z_{t+1}}{\lambda_t(1+k)+1}, t+1\right\}\right], \qquad (2.12)$$

where we used the homothetic property of the value function to take the term outside the expectations operator. A similar argument for the stock sale yields:

$$\bar{J}(x_t, y_t, \lambda_t, t) = (x_t + (1-k)y_t)^{\alpha} E_t\left[V\left\{\frac{R}{\lambda_t(1-k)+1}, \frac{\lambda_t z_{t+1}}{\lambda_t(1-k)+1}, t+1\right\}\right]. \quad (2.13)$$

It can be shown that maximizing (2.12) and (2.13) with respect to λ_t yields, respectively the lower and the upper boundary of the NT region $\underline{\lambda}_t$ and $\bar{\lambda}_t$.[73] Since it is apparent that the terms in powers α are inconsequential positive quantities, we have the following program solving for $\underline{\lambda}_t$ and $\bar{\lambda}_t$:

$$\underline{\lambda}_t = \arg\max_{\lambda_t}\left[\underline{V}(\lambda_t, t)\right]$$

and , (2.14)

$$\bar{\lambda}_t = \arg\max_{\lambda_t}\left[\bar{V}(\lambda_t, t)\right]$$

where:

$$\underline{V}(\lambda_t, t) = E_t\left\{V\left(\frac{R}{\lambda_t(1+k)+1}, \frac{\lambda_t z_{t+1}}{\lambda_t(1+k)+1}, t+1\right)\right\}$$

and . (2.15)

$$\bar{V}(\lambda_t, t) = E_t\left\{V\left(\frac{R}{\lambda_t(1-k)+1}, \frac{\lambda_t z_{t+1}}{\lambda_t(1-k)+1}, t+1\right)\right\}$$

If the NT region exists, the program (2.14)-(2.15) will always yield a solution since the value function is strictly concave.

[73]An induction proof is in the Appendix to GJ, itself an application of the general result in Constantinides (1979) to the CRRA utility function.

We may now formulate compactly the investor's dynamic problem with the inclusion of the optimal investment policy:

$$V(x_t, y_t, t) = \begin{cases} (x_t + (1+k)y_t)^\alpha \underline{V}(\underline{\lambda}_t, t) & \text{for } \lambda_t < \underline{\lambda}_t \\ E_t\left[V(x_t R, y_t z_{t+1}, t+1)\right] & \text{for } \underline{\lambda}_t \le \lambda_t \le \overline{\lambda}_t \\ (x_t + (1-k)y_t)^\alpha \overline{V}(\overline{\lambda}_t, t) & \text{for } \lambda_t > \overline{\lambda}_t \end{cases} . \tag{2.16}$$

2.2

Before turning to the solution of the problem, we demonstrate that the formulation for the state dynamics outside the NT region in GJ (Genotte and Jung, 1994) contains an error. First, define the value function for \$1 worth of portfolio:[74]

$$I_t(\lambda_t) \equiv V\left(\frac{1}{1+\lambda_t}, \frac{\lambda_t}{1+\lambda_t}, t\right). \tag{2.17}$$

Here we reproduce the GJ dynamic formulation in our notation, originally equation (6) in GJ:

$$I_t(\lambda_t) = \begin{cases} \left(\dfrac{\dfrac{\lambda_t(1+k)+1}{\lambda_t+1}}{\dfrac{\underline{\lambda}_t(1+k)+1}{\underline{\lambda}_t+1}}\right)^\alpha \underline{V}(\underline{\lambda}_t, t) & \text{for } \lambda_t < \underline{\lambda}_t \\ \\ \left(\dfrac{\dfrac{\lambda_t(1-k)+1}{\lambda_t+1}}{\dfrac{\overline{\lambda}_t(1-k)+1}{\overline{\lambda}_t+1}}\right)^\alpha \overline{V}(\overline{\lambda}_t, t) & \text{for } \lambda_t > \overline{\lambda}_t \end{cases} , \tag{2.18}$$

[74] p. 388 in GJ.

where $\underline{V}(\underline{\lambda}_t,t)$ and $\overline{V}(\overline{\lambda}_t,t)$ through which the maximized expectations enter the dynamic formulation are as in (2.15).[75] We will show that (2.18) is incorrect; specifically, both factors within the braces are incorrectly specified. We demonstrate the error for the upper factor multiplying $\underline{V}(\underline{\lambda}_t,t)$, the proof for the other being symmetric.

By the definition of $I_t(\lambda_t)$ we may dynamically formulate $\underline{V}(\underline{\lambda}_t,t)$ in the following way:[76]

$$\underline{V}(\underline{\lambda}_t,t)=\max_{\lambda_t}\left\{R^{\alpha}\left(\frac{\lambda_{t+1}+1}{\lambda_t(1+k)+1}\right)^{\alpha}I_{t+1}(\lambda_{t+1})\right\}. \tag{2.19}$$

Now assume that we have maximized the RHS of (2.19), which implies $\lambda_{t+1}=\underline{\lambda}_t\frac{z_{t+1}}{R}$.[77] Then

$$\begin{aligned}\underline{V}(\underline{\lambda}_t,t)&=R^{\alpha}\left(\frac{\underline{\lambda}_t\frac{z_{t+1}}{R}+1}{\underline{\lambda}_t(1+k)+1}\right)^{\alpha}I_{t+1}\left(\frac{1}{\underline{\lambda}_t\frac{z_{t+1}}{R}+1},\frac{\underline{\lambda}_t\frac{z_{t+1}}{R}}{\underline{\lambda}_t\frac{z_{t+1}}{R}+1},t+1\right)\\&=E_t\left\{V\left(\frac{R}{\underline{\lambda}_t(1+k)+1},\frac{\underline{\lambda}_t z_{t+1}}{\underline{\lambda}_t(1+k)+1},t+1\right)\right\}\end{aligned}, \tag{2.20}$$

where in the first line we used the definition (2.17) and the assumption that the value function is maximized; in the second line we used the homogeneity and definition of the value function. The final result demonstrates that the formulation in the first line is consistent with the formulation (2.15).

---

[75]Equation (2.15) which defines $\underline{V}(.)$ corresponds to equation (2) in GJ.

[76]Equation (5) in GJ.

[77]The maximizing asset ratio is $\underline{\lambda}_t$, which under the asset dynamics in equation (2.6) becomes in one period equal to the above quantity.

By substituting the second line of (2.20) into the first line of (2.18), using the homogeneity of the value function and simplifying we get:

$$\text{RHS of (2.18)} = E_t\left\{V\left(\frac{(\underline{\lambda}_t+1)(\lambda_t(1+k)+1)R}{(\lambda_t+1)(\underline{\lambda}_t(1+k)+1)^2}, \frac{(\underline{\lambda}_t+1)(\lambda_t(1+k)+1)\underline{\lambda}_t z_{t+1}}{(\lambda_t+1)(\underline{\lambda}_t(1+k)+1)^2}, t+1\right)\right\}. \quad (2.21)$$

By inspection, it is apparent that (2.21) cannot yield proper dynamics for the problem at hand. However, suppose there is a typographical error in (2.18) of reversing the positions of the numerator and denominator in the first term in the first line of (2.18), which error was perhaps rectified in the numerical work in GJ. Under this supposition we have

$$\text{RHS of (2.18)} = E_t\left\{V\left(\frac{(\lambda_t+1)R}{(\underline{\lambda}_t+1)(\lambda_t(1+k)+1)}, \frac{(\lambda_t+1)\underline{\lambda}_t z_{t+1}}{(\underline{\lambda}_t+1)(\lambda_t(1+k)+1)}, t+1\right)\right\}. \quad (2.21')$$

To verify whether (2.21') is correct, we must analyze the LHS of (2.18). Since we have maximized the RHS of (2.18), the same must hold for the LHS of this expression. Under this line of reasoning, the same arguments in the LHS and RHS of (2.18) except for the time transition factors R and z_{t+1} would demonstrate the correctness of (2.21').

To maximize the LHS of (2.18) we must transact from the positions implied by the definition of $I_t(\lambda_t)$ (2.17), which implies setting the respective bond and stock arguments equal to $1/(\lambda_t+1)$ and $\lambda_t/(\lambda_t+1)$. For these asset holdings, the optimal transaction v_t solves

$$\frac{\lambda_t/(\lambda_t+1)+v_t}{1/(\lambda_t+1)-(1+k)v_t} = \underline{\lambda}_t, \quad (2.22)$$

which implies the following optimal holdings:

$$x_t = \frac{\lambda_t(1+k)+1}{(\lambda_t+1)[\underline{\lambda}_t(1+k)+1]}, \; y_t = \frac{[\lambda_t(1+k)+1]\underline{\lambda}_t}{(\lambda_t+1)[\underline{\lambda}_t(1+k)+1]} \; . \tag{2.23}$$

Since $x_t R$ is not equal to the first argument in (2.21') and $y_t z_{t+1}$ is not equal to the second one, we conclude that the dynamic formulation in GJ is incorrect even with the benefit of doubt for a typographical error.

Apart from the algebraic errors in (2.18) shown above, another substantive error lies in taking outside the maximands $\underline{V}(\underline{\lambda}_t, t)$ and $\overline{V}(\overline{\lambda}_t, t)$ the multiplicative factors that respectively contain $\underline{\lambda}_t$ and $\overline{\lambda}_t$, the arguments in which these two expressions are maximized. In case of the lower boundary $\underline{\lambda}_t$, the numerical algorithm in GJ searches for a given λ_t for the value $\underline{\lambda}_t$ which maximizes (2.18). Note that for λ_t close to the solution $\underline{\lambda}_t$, the factor in front of the maximand in (2.18) is close to 1, in which case we obtain an approximately correct dynamic formulation since $\underline{V}(\underline{\lambda}_t, t)$ is itself the correct dynamic formulation. The same line of reasoning holds for the upper boundary $\overline{\lambda}_t$. This implies that, in principle, the GJ approach may lead to approximately correct results; however, this approach is not generally acceptable since it is based on an incorrect formulation, which may lead to spurious results should a guess for $\underline{\lambda}_t$ or $\overline{\lambda}_t$ prove to be a difficult one.[78]

Notwithstanding the apparent errors in the dynamic formulation in GJ, their numerical results in some cases appear to be qualitatively consistent with ours and with those of Constantinides (1986). This may be the result of their choice of parameter values. We formulate below an alternative approach that leads to exact results.

[78] Only if one guesses a value of λ_t close to the solution $\underline{\lambda}_t$, the factor in front of the maximand becomes close to 1, thus making the GJ dynamic formulation approximately correct.

Table 12

Comparison with Genotte-Jung (GJ, 1994) results

T	$\underline{\lambda}$ (GJ)	$\underline{\lambda}$	$\bar{\lambda}$ (GJ)	$\bar{\lambda}$	$\underline{\lambda}$ (GJ)	$\underline{\lambda}$	$\bar{\lambda}$ (GJ)	$\bar{\lambda}$
	Panel A: $\delta = 2, k = 0.5\%, \sigma = 15\%$				Panel B: $\delta = 2, k = 1\%, \sigma = 20\%$			
0.025	N/F	N/F	-4.05	-3.97	N/F	N/F	3.00	3.06
0.125	N/F	N/F	-3.95	-3.89	N/F	N/F	3.10	3.14
0.25	N/F	N/F	-3.86	-3.84	N/F	N/F	3.20	3.20
0.375	N/F	N/F	-3.86	-3.81	N/F	0.09	3.20	3.25
0.5	N/F	N/F	-3.77	-3.78	0.33	0.33	3.30	3.29
0.75	-27.60	-25.35	-3.77	-3.73	0.70	0.71	3.30	3.36
1	-9.96	-9.74	-3.68	-3.69	0.99	0.99	3.40	3.43
1.25	-7.50	-7.38	-3.68	-3.66	1.20	1.20	3.51	3.48
1.5	-6.50	-6.47	-3.68	-3.63	1.36	1.36	3.51	3.53
2	-5.77	-5.72	-3.59	-3.58	1.59	1.60	3.63	3.62
3	-5.38	-5.25	-3.51	-3.51	1.87	1.86	3.74	3.79
4	-5.13	-5.09	-3.51	-3.45	1.99	2.00	3.99	3.94
	Panel C: $\delta = 3, k = 1\%, \sigma = 20\%$				Panel D: $\delta = 3, k = 1\%, \sigma = 35\%$			
0.025	N/F	N/F	N/F	1.02	N/F	0.01	N/F	0.20
0.125	N/F	N/F	N/F	1.04	N/F	0.14	0.62	0.20
0.25	0.20	N/F	5.18	1.06	0.06	0.16	0.38	0.21
0.375	0.39	0.06	2.67	1.08	0.10	0.17	0.32	0.22
0.5	0.50	0.20	2.01	1.09	0.12	0.17	0.29	0.22
0.75	0.62	0.38	1.61	1.11	0.13	0.17	0.27	0.22
1	0.68	0.49	1.46	1.13	0.13	0.17	0.26	0.22
1.25	0.73	0.56	1.37	1.15	0.14	0.17	0.26	0.22
1.5	0.75	0.61	1.37	1.16	0.14	0.17	0.26	0.22
2	0.75	0.67	1.33	1.19	0.14	0.17	0.26	0.22
3	0.78	0.72	1.29	1.23	0.14	0.17	0.26	0.22
4	0.78	0.75	1.29	1.27	0.14	0.17	0.26	0.22
	Panel E: $\delta = 3, k = 0.1\%, \sigma = 15\%$				Panel F: $\delta = 3, k = 1\%, \sigma = 15\%$			
0.025	N/F	0.02	N/F	8.33	N/F	N/F	N/F	8.40
0.125	3.41	1.89	N/F	8.67	N/F	N/F	N/F	8.57
0.25	4.86	3.42	18.79	8.61	N/F	N/F	N/F	8.70
0.375	5.62	4.36	13.16	8.72	N/F	0.11	N/F	8.77
0.5	6.16	4.98	11.42	8.76	1.45	0.42	N/F	8.87
0.75	6.62	5.75	10.38	8.95	2.22	0.97	N/F	9.02
1	6.78	6.18	9.90	9.07	2.89	1.46	N/F	9.15
1.25	6.78	6.44	9.67	9.19	3.33	1.88	N/F	9.25
1.5	6.94	6.61	9.67	9.29	3.75	2.24	N/F	9.37
2	6.94	6.84	9.67	9.47	4.32	2.86	N/F	9.57
3	6.94	7.03	9.67	9.71	5.10	3.73	27.46	9.93
4	6.94	7.13	9.67	9.82	N/F	4.31	N/F	10.25

The table displays the results for the NT region for diffusion derived under the methodology developed in this essay compared to the results in Table 12b in GJ. The risk premium is 6%. The symbol 'N/F' corresponds to the cases for which a given boundary of the NT region could not be found.

We present in Table 12 similarities and differences between the GJ and our results.[79] It appears from panels A and B that for the RRA δ set equal to 2 the results are consistent apart from minor differences in the quantities and in the convergence pattern. However, for δ set equal to 3 in panels C-F the results diverge both qualitatively and numerically.[80] First, we observe an apparent convergence from above of the upper boundary $\bar{\lambda}$ in GJ. We did not find this type of convergence in any of different cases we considered, while in all cases we observed $\bar{\lambda}$ converging uniformly from below, with the values close to but above the Merton line for short time horizons. In panels C-E we observe relatively consistent results for longer time horizons with apparent differences for shorter time horizons for either boundary. However, the results in panel D for the longest time horizon, which is 4 years, differ by more than 10% for either boundary. In panel F, which differs from panel E by a higher transaction costs rate, the GJ results are either highly inconsistent with ours or non existing.

3 Numerical Analysis

We describe our numerical approach that is based on forward induction. A critical step in this approach consists of an efficient incorporation of already known solutions to the problem at all future dates. Further, we present details of the discretization of the risky asset dynamics. We finish this section by discussing the introduction of intermediate consumption to our problem, which is left for future research.

3.1

To solve for the program (2.14)-(2.15) we apply the direct approach (2.6). This forward-inductive approach yields $\underline{V}(\lambda_t,t)$ and $\overline{V}(\lambda_t,t)$ as *continuous* functions of λ_t, resulting in

[79] The GJ results presented in Table 12 are excerpts from Table 12b in GJ.
[80] RRA of 3 was the highest value for this parameter in the GJ results.

flexible modeling. This admits the derivation of the NT region by reliable optimization routines resulting in highly accurate estimates. In general, for the current setting, we may write the value function as the expectation of the terminal utility of wealth with respect to the underlying probability space:[81]

$$V(x_t, y_t, t) = \max_{v_\tau \ \forall \tau \in [t,\, T-1]} \sum_j U\left(x_{T,j} + (1-k) y_{T,j}\right) \Pr(\omega_j), \tag{3.1}$$

where ω_j represents a path j for a given discrete-time probability space, and $x_{T,j}$ and $y_{T,j}$ have been derived for the filtration of a given probability space and the optimal investment decisions at all dates $\tau = t, t+1, \ldots, T-1$. In particular, (3.1) applies to the program (2.14)-(2.15) with appropriate quantities substituted for x_t, y_t.

Since we use a (multidimensional) tree to represent the process of the risky asset, equation (3.1) appears to be difficult to solve in numerical work for more than a few time periods. The difficulty stems from the fact that the number of paths grows exponentially in time partition. To deal with this difficulty, in the following paragraphs we present a recursive model which efficiently aggregates the paths of the state variable λ_τ, $\tau = t+1, \ldots, T-1$. As we will show, exploiting the fact that the ratio y_t/x_t is the sole state variable, the homothetic property of the value function and the recombination property of the assumed discretization of the risky asset dynamics will allow us to aggregate the states at each forward step.

Assume that the boundaries of the NT region $\underline{\lambda}_\tau$ and $\overline{\lambda}_\tau$ were found for all times $\tau = t+1, \ldots, T-1$. Given this information set, we define the indirect but not necessarily maximized utility for a portfolio consisting of \$1 in the riskless and \$$\lambda_t$ in the risky asset at time t for the probability space as defined in (3.1) *and* for the future optimal portfolio restructuring:

[81] The idea of solving for the NT region by forward induction in numerical work was first presented in Boyle and Lin (1997). However, their numerical approach considered the first order condition on the terminal utility, which was suitable to solve the problem only for a limited number of periods.

$$J(1,\lambda_t,t) \equiv \sum_j U\left(x_{T,j}+(1-k)y_{T,j}\right)\Pr(\omega_j) \mid \left\{\underline{\lambda}_\tau, \overline{\lambda}_\tau\right\}_{\tau=t+1,\dots,T-1}. \tag{3.2}$$

By (2.14)-(2.15) and (3.2) it is clear that we have:

$$\underline{\lambda}_t = \arg\max_{\lambda_t}\left\{\left[(1+k)\lambda_t+1\right]^{-\alpha} J(1,\lambda_t,t)\right\}$$

$$\text{and} \qquad . \tag{3.3}$$

$$\overline{\lambda}_t = \arg\max_{\lambda_t}\left\{\left[(1-k)\lambda_t+1\right]^{-\alpha} J(1,\lambda_t,t)\right\}$$

We partition the time-τ paths of the state variable λ_τ into two types: the first type includes those paths which remain inside the NT region, and the second type includes the paths that fall outside this region and are traded to the nearest boundary $\underline{\lambda}_\tau$ or $\overline{\lambda}_\tau$ by virtue of the simple investment policy. The first type of paths presents no particular problem since inside the NT region the state variable will follow the recombination pattern of a given lattice. To see that, consider that in this case the time-τ portfolio holdings are $\left(R^{\tau-t}, \lambda_t \hat{Z}_\tau\right)$, with the cumulative stock return up to time τ $\hat{Z}_\tau \equiv \Pi_{i=1}^{\tau-t} z_{t+i}$ which implies $\lambda_\tau = \lambda_t \dfrac{\hat{Z}_\tau}{R^{\tau-t}}$, with the associated probabilities resulting from the ($\tau-t$)-period convolution of the one-period distribution of the risky asset with itself.

For the second type of paths, namely those for which the simple investment policy stipulates a trade to the nearest boundary of the NT region, at each time τ we derive a single number, the contribution of these paths to the terminal derived utility ($\equiv \Delta J_\tau$). As we will demonstrate, the quantity ΔJ_τ will subsume all the relevant past and future path information as of time-τ. We elaborate below on the derivation of ΔJ_τ; here we define it implicitly by

$$J(1,\lambda_t,t) = \sum_{\tau=t+1}^{T-1} \Delta J_\tau + \alpha^{-1} \sum_{k=1}^{N_T} \Pr\left(\hat{Z}_{T,k}\right)\left(R^{T-t}+(1-k)\lambda_t \hat{Z}_{T,k}\right)^\alpha, \tag{3.4}$$

where the second summation is over the N_T paths which remained inside the NT region at *each and every* time τ, $\tau = t+1,\dots,T-1$, with $\Pr(.)$ denoting time-t probabilities of the terminal states of the risky asset.[82]

To aggregate the time-τ paths outside the NT region to ΔJ_τ defined above, we use the homothetic property of the value function and the fact that we already solved for the value function $V(1,\underline{\lambda}_\tau,\tau)$ and $V(1,\overline{\lambda}_\tau,\tau)$, $\tau = t+1,\dots,T-1$ by maximizing (3.3). To demonstrate that there exists an aggregation that yields (3.2), the derived utility that subsumes all the relevant path information, we use the following result.

<u>Lemma 1</u>: The contribution ΔJ_τ of all time-τ paths outside the NT region to the time-t derived utility of terminal wealth as defined by (3.4) is the following:

$$\Delta J_\tau = \left[\underline{X}_\tau V(1,\underline{\lambda}_\tau,\tau) + \overline{X}_\tau V(1,\overline{\lambda}_\tau,\tau)\right], \tag{3.5}$$

where $\underline{X}_\tau = \sum_{i=1}^{\underline{n}_\tau} \Pr(\hat{Z}_{\tau,i}) \left(\dfrac{R^{\tau-t} + (1+k)\lambda_t \hat{Z}_{\tau,i}}{\underline{\lambda}_\tau (1+k)+1} \right)^\alpha$ and

$\overline{X}_\tau = \sum_{j=1}^{\overline{n}_\tau} \Pr(\hat{Z}_{\tau,j}) \left(\dfrac{R^{\tau-t} + (1-k)\lambda_t \hat{Z}_{\tau,j}}{\overline{\lambda}_\tau (1-k)+1} \right)^\alpha$, where $\hat{Z}_{\tau,i}$ and $\hat{Z}_{\tau,j}$ are respectively the stock returns resulting in the portfolio proportion λ_τ below or above the NT region, $\Pr(.)$ are the time-t probabilities of these returns, and we denote the time-τ number of $\hat{Z}_{\tau,i}$'s ($\hat{Z}_{\tau,j}$'s) by $\underline{n}_\tau$ ($\overline{n}_\tau$). Note that these probabilities may be derived by simply following the recombination pattern of a given lattice. The terms under the power of α are time-τ dollar values of the bond account after transacting to a given boundary of the NT region.

<u>Proof</u>: The lemma may be demonstrated very simply by induction. Without loss of generality, in our proof we consider only the portfolio adjustments to the lower boundary of

[82] In our numerical work, we use the fact the lattice is recombining inside the NT region, which implies that at the terminal date we have nodes rather than paths.

the NT region $\underline{\lambda}_\tau$ and the resulting quantity $\underline{X}_\tau$. The inclusion of the adjustment to the other boundary $\overline{\lambda}_\tau$ follows easily by extension. Consider $t = T-2$. We have $V(1,\underline{\lambda}_{T-1},T-1) = \alpha^{-1}\sum_{i=1}^{m} p_s\left(R+(1-k)\underline{\lambda}_{T-1}z_s\right)^\alpha$, where p_i is the probability of a one-period stock return z_i, $i = 1...m$. With the use of the lemma, we have:

$$\Delta J_{T-1} = \alpha^{-1}\sum_{i=1}^{\underline{n}_{T-1}} p_i\left(\frac{R+(1+k)\lambda_{T-2}z_i}{\underline{\lambda}_{T-1}(1+k)+1}\right)^\alpha \sum_{s=1}^{m} p_s\left(R+(1+k)\underline{\lambda}_{T-1}z_s\right)^\alpha, \tag{3.6}$$

where the first summation results from the definition of $\underline{X}_\tau$, $\tau = T-1$, and $m \geq \underline{n}_{T-1} \geq 0$. By the homothetic property which in this simple case collapses to multiplying terms under the same power, (3.6) yields the same result we would get from equation (3.2) by considering each path separately.

Consider any time *t*. Assume that the lemma holds at $\tau+1$. At time τ we have:

$$\Delta J_\tau = \sum_{i=1}^{\underline{n}_\tau} \Pr\left(\hat{Z}_{\tau,i}\right) V\left(\frac{R+(1+k)\lambda_t\hat{Z}_i}{\underline{\lambda}_\tau(1+k)+1}, \frac{R+(1+k)\lambda_t\hat{Z}_i}{\underline{\lambda}_\tau(1+k)+1}\underline{\lambda}_\tau, \tau\right) \equiv \sum_{i=1}^{\underline{n}_\tau} \underline{X}_\tau V(1,\underline{\lambda}_\tau,\tau). \tag{3.7}$$

When we apply one forward inductive step to the quantity above we get:

$$\Delta J_\tau = \sum_{i=1}^{\underline{n}_\tau} \underline{X}_\tau \begin{bmatrix} \sum_{k=1}^{\underline{n}_{\tau+1}} p_k \left(\frac{R+(1+k)\underline{\lambda}_\tau z_k}{\underline{\lambda}_{\tau+1}(1+k)+1}\right)^\alpha V(1,\underline{\lambda}_{\tau+1},\tau+1) + \\ \sum_{j=1}^{\overline{n}_{\tau+1}} p_j \left(\frac{R+(1-k)\underline{\lambda}_\tau z_j}{\overline{\lambda}_{\tau+1}(1-k)+1}\right)^\alpha V(1,\overline{\lambda}_{\tau+1},\tau+1) + \\ \sum_{s=1}^{m-\underline{n}_{\tau+1}-\overline{n}_{\tau+1}} p_s V(R,\underline{\lambda}_\tau z_s,\tau+1) \end{bmatrix}, \tag{3.8}$$

where the first two summation in square brackets consider these one-period paths that are outside the NT region by using the induction hypothesis and the third one considers these paths which remain inside this region, with z_k's, z_j's and z_s's denoting appropriate one-period returns and m denoting the number of one-period returns in a given lattice. Since, by the definition of $\underline{X}_\tau$ it is apparent that equation (3.8) considers all the relevant path information, continuing forward until the terminal date T is reached will reproduce the definition (3.2) for the time-τ paths outside the NT region. This ends the proof, QED.

With the use of Lemma 1, we rewrite the maximization problem (3.3) as follows:

$$\underline{\lambda}_t = \arg\max\left\{\left[(1+k)\lambda_t+1\right]^{-\alpha}\left[\sum_{\tau=t+1}^{T-1}\Delta J_\tau+\alpha^{-1}\sum_{k=1}^{N_T}\Pr\left(\hat{Z}_{T,k}\right)\left(R^{T-t}+(1-k)\lambda_t\hat{Z}_{T,k}\right)^{\alpha}\right]\right\}$$

and , (3.9)

$$\overline{\lambda}_t = \arg\max\left\{\left[(1-k)\lambda_t+1\right]^{-\alpha}\left[\sum_{\tau=t+1}^{T-1}\Delta J_\tau+\alpha^{-1}\sum_{k=1}^{N_T}\Pr\left(\hat{Z}_{T,k}\right)\left(R^{T-t}+(1-k)\lambda_t\hat{Z}_{T,k}\right)^{\alpha}\right]\right\}$$

where the second summation is over the N_T recombined paths which remained inside the NT region at *each and every* time τ, $\tau = t+1,...,T-1$.

Now we can describe the major steps of our numerical solution. Take a candidate solution λ_t for either maximization problem in (3.3) and proceed forward with the lattice. At each time τ derive the terminal contribution ΔJ_τ of all paths outside the NT region by (3.5). Delete these paths from the lattice and repeat the process till $\tau = T-1$ is reached. Since it is apparent that equation (3.9) yields the derived utility as a continuous function of λ_t, the maximization problem can be passed to optimization routines present in many software packages such as Matlab. Except for possible numerical errors, the formulation (3.9) yields the exact solution for the NT region for a given discretization approach. This algorithm executes in short time even for the large number of nodes in a one-period lattice that we use to approximate a jump-diffusion process, since a limited number of nodes remains inside the NT region at any time τ. For instance, for a 10-year span with 50 subdivisions a year we derive the results for all subdivisions in app. 1000 seconds for either diffusion (trinomial lattice) or mixed process (19-nomial lattice) by using 1.83 GHz Intel dual processor with 1 GB of RAM.

3.2

To solve for the NT region for the diffusion case, we approximate the stock dynamics by the Kamrad and Ritchken (1991) trinomial model. For the jump-diffusion case we use a multinomial approximation. First, we approximate the diffusion component by the trinomial model with the mean and variance implied by (2.3) with the probabilities adjusted by the factor $1-\eta_K \Delta t$ implied by (2.5). To approximate the jump component, we space $m > 3$ states by the same distance in the log scale as in the trinomial model and derive the probabilities as the normalized to 1 densities implied by the distribution of J. These probabilities are adjusted by the factor $\eta_K \Delta t$ implied by (2.5). In the final step, the adjusted trinomial probabilities are added to the adjusted three central probabilities of the jump component.

To derive the NT region, we start at $t = T-1$ and move recursively backward while using the forward induction (3.1)-(3.2) to solve the problem at each time t. The function (3.9) is derived and passed to an optimization routine, which derives its maximizing arguments. We use 50 time divisions for one calendar year and 19-nomial tree for the mixed process.[83]

3.3

The introduction of intermediate consumption presents a significant challenge for the numerical work. The simplifying assumption in Constantinides (1986) of state and time independent consumption is of little use here because of the recursive derivation in the discrete time discrete state setup for the risky asset. It will also be difficult to deal with the time and state dependent consumption since it would break down the recombining pattern for the state variable λ_t inside the NT region. The remaining option is to use time- but not state-dependent consumption, i.e. to assume that for every state at given time t the agent

[83] Numerical simulations showed that increasing these numbers beyond these values did not affect significantly our numerical results, at least for the chosen parameter values. As a rule of thumb determining the lattice size, we keep branches with one-period probabilities greater than 1e-10.

consumes the same proportion of her bond account. It can be easily shown that under this assumption λ_t will recombine inside the NT according to the recombination pattern of the risky asset. It is in principle possible to introduce a second auxiliary state variable similar to X_τ which will subsume the intermediate consumption. However, adding another control in the optimization problem may create numerical problems not easy to handle. This is a topic for future research.[84]

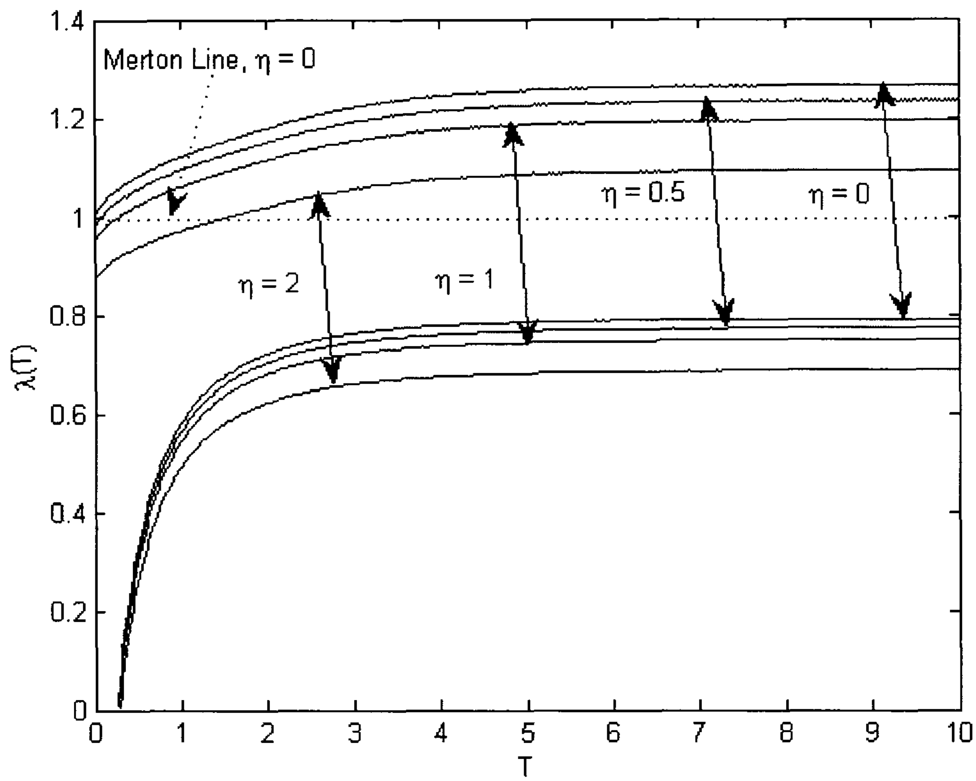
[84] As Davis and Norman (1990) showed, the Constantinides (1986) simplifying assumption of a time- and state-dependent consumption does not hold in a more general model. Nonetheless, the value function, the NT region and the frequency of trading remain approximately equal in the simplified setup.

4 Results

We focus the presentation of our results on differences between the diffusion and jump-diffusion cases. Our base case uses the following parameters: the transaction cost rate k=0.5%, the RRA δ=2, the total volatility σ=20%, the risk premium is 4%, the logarithm of the expected jump size μ_K=-1%, the jump volatility σ_K=7%. Figure 10 displays the NT region for the parameters above while the jump intensity η varies from 0 (diffusion) to 2. It is apparent that increasing the jump intensity shifts the NT region towards the riskless asset with significant changes where this intensity is large. Note also that the upper boundary of the NT region for the mixed process falls below the Merton line for short time horizons, which is never the case for diffusion. In Figure 10, we may also clearly observe the convergence of the NT region boundaries to constant levels as the horizon length increases for a given set of parameters.

In Table 13 we present the shift towards the riskless asset more systematically by displaying the proportional change from the diffusion case for the same jump intensities as in Figure 10 but for four different levels of RRA δ: 2, 3, 5 and 10. We define this change as $1-\lambda_{jd}/\lambda_d$, with the subscripts jd and d denoting respectively jump-diffusion and diffusion. We can clearly see in Table 13 that the proportional shift towards the riskless asset decreases in risk aversion, while it more than doubles for each RRA when the jump intensity doubles. We can also see that for low risk aversion the proportional change in the lower boundary of the region from the diffusion case to the mixed process case is lower than the equivalent change in the upper boundary, with reversal of this effect for high risk aversion.

Figure 10: No Transaction Region for Diffusion and Mixed Process



The figure displays the NT region for mixed jump-diffusion process for varying jump intensity η. Other parameters are as follows: the transaction costs rate k= 0.5%, the RRA δ= 2, the total volatility σ = 20%, the risk premium is 4%, the logarithm of the expected jump size μ_K = -1%, the jump volatility σ_K = 7%. The Merton line corresponds to diffusion with the parameters as above except for the transaction cost rate, which is 0.

In the second part of our results we present the sensitivity of the boundaries of the NT region to the total volatility of the stochastic process of the risky asset, to the transaction costs rate and to the RRA. For the former, when we change the total volatility σ from the base case of 20% to a new value, we also change the base volatility σ_K 7% of the jump component by a factor. For instance, if we change the total volatility from 20% to 30%, a new value for the volatility of the jump component becomes 1.5x7%. We display the results in Table 14. In panel A we observe an expected shift towards the riskless asset as the total volatility increases, with the proportional changes in the boundaries from the diffusion case to the mixed process case decreasing in total volatility. In panel B we observe the expected widening of the boundaries as the transaction costs rate increases, with similar proportional changes in the boundaries from the diffusion case to the mixed process case for every transaction costs rate level. In panel C we observe an expected shift towards the riskless asset as the risk aversion increases.

Table 13

Proportional Differences in the Boundaries of the NT Region between Diffusion and Jump-Diffusion

RRA	$1-\underline{\lambda}_{jd}/\underline{\lambda}_{d}$ (%)			$1-\overline{\lambda}_{jd}/\overline{\lambda}_{d}$ (%)		
	$\eta=0.5$	$\eta=1$	$\eta=2$	$\eta=0.5$	$\eta=1$	$\eta=2$
2	2.28	5.27	13.12	2.47	5.57	13.56
3	1.73	4.05	10.31	1.75	4.07	10.22
5	1.47	3.52	9.60	1.45	3.45	9.27
10	1.29	3.04	7.90	1.25	2.93	7.52

The table displays the results for the NT region for a 10-year investment horizon. Other parameters are as follows: the transaction costs rate k = 0.5%, the RRA δ = 2, the total volatility σ = 20%, the risk premium is 4%, the logarithm of the expected jump size μ_K = -1%, the jump volatility σ_K = 7%.

We may expect a low sensitivity of the boundaries of the NT region to the mean of the jump component μ_K since the mean of the diffusion component is adjusted in our formulation by this quantity times the jump intensity η. For instance, if we change μ_J form -1% to -3% in our base case, the boundaries of the NT region change from 0.7745 and 1.2355 to 0.7777 and 1.2420. We may expect a similar low sensitivity of the boundaries of the NT region to changes in σ_K since a change in this parameter is compensated in the volatility of the diffusion component in (2.4). For instance, if we change σ_K form 7% to 10.5% in our base case, the boundaries of the NT region change from 0.7745 and 1.2355 to 0.7839 and 1.2476. These numbers are typical of our approach that keeps the total volatility constant and singles out the jump intensity as the most important factor governing the relation of the NT region in the diffusion case to this region in the mixed process case.[85]

[85] The NT region changes very significantly if the jump component is added on to the diffusion without keeping a constant volatility; see Liu *et al* (2003).

Table 14

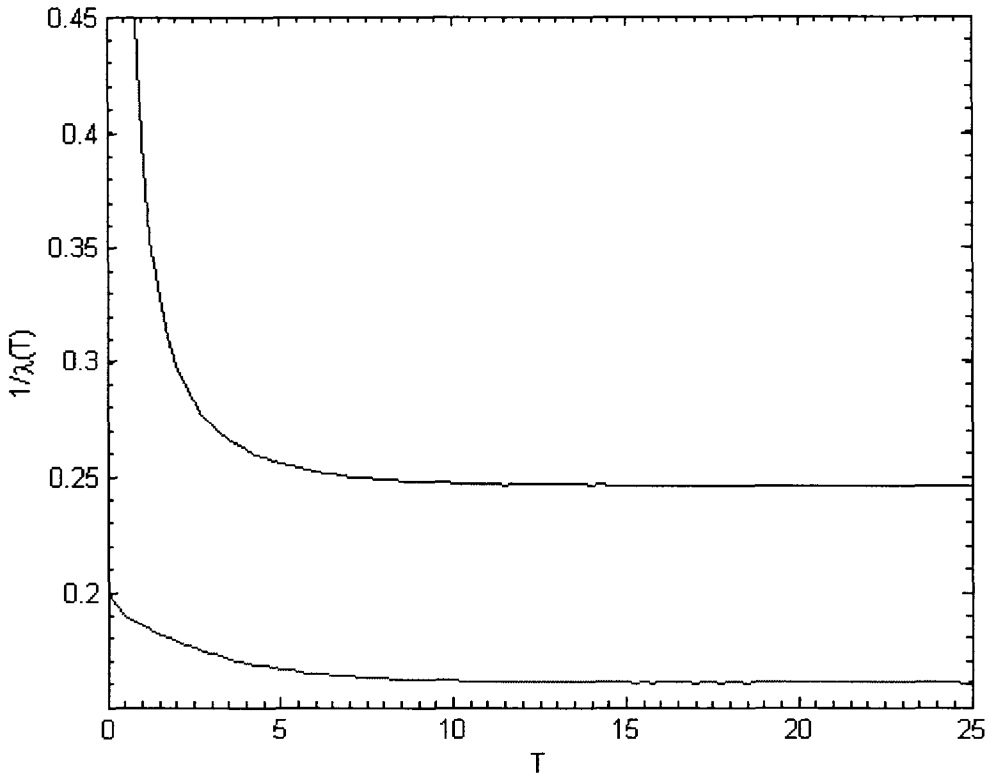
Sensitivity of the Boundaries of the NT Region

Variable	$\underline{\lambda}_{jd}$	$\overline{\lambda}_{jd}$	$1-\underline{\lambda}_{jd}/\underline{\lambda}_d$ (%)	$1-\overline{\lambda}_{jd}/\overline{\lambda}_d$ (%)
σ	Panel A: Sensitivity to Total Volatility			
20%	0.7745	1.2355	2.28	2.47
25%	0.3591	0.5822	1.64	1.68
30%	0.2131	0.3570	1.25	1.44
35%	0.1414	0.2466	1.33	1.28
k	Panel B: Sensitivity to Transaction Costs Rate			
0.1%	0.8609	1.1124	2.26	2.36
0.5%	0.7745	1.2355	2.28	2.47
1%	0.7210	1.3175	2.32	2.53
2%	0.6427	1.4037	2.42	2.55
RRA	Panel C: Sensitivity to Risk Aversion			
2	0.7745	1.2355	2.28	2.47
3	0.3950	0.6027	1.73	1.75
5	0.1997	0.2969	1.45	1.43
10	0.0893	0.1307	1.29	1.25

The table displays the results for the NT region for a 10-year investment horizon. Other parameters are as follows: the transaction costs rate k=0.5%, the RRA δ=2, the total volatility σ=20%, the risk premium is 4%, the logarithm of the expected jump size μ_K = -1%, the jump volatility σ_K = 7%, the jump intensity η = 0.5 except that in each panel the name and values for a varying parameter are provided in the first column.

An open question of our work is the accuracy of our discrete time numerical algorithm in approximating the continuous time solution. Unfortunately there are no closed form expressions for the NT region for the jump-diffusion problem (or, for that matter, for simple diffusion) when the investment horizon is finite. Liu and Lowenstein (2008) have provided an approximation to the continuous time solution in the form of an Erlang-distributed horizon that produces a sequence of ordinary differential equations, whose successive solutions converge to the value function and NT region of the fixed-horizon jump-diffusion case.

Figure 11: No Transaction Region for Liu-Loewenstein (2008) Parameters



The figure displays the results for the NT region for mixed jump-diffusion process for the Liu-Loewenstein (2008) parameters: the RRA δ = 5, the total volatility σ = 12.86%, the risk premium of 7%, the jump intensity η = 0.1, the logarithm of the expected jump size μ_K = -6.75%, the jump volatility σ_K = 8.53%, the transaction cost rate for stock purchase (sale) of 1% (0).

Figure 11 and Table 15 show the NT region for the jump-diffusion case evaluated with our algorithm for the parameter values used by Liu and Lowenstein (2008). We follow the presentation in that latter study, i.e. we display the reciprocals of the NT region boundaries as defined in this essay with the following set of parameters: the RRA δ = 5, the total volatility σ = 12.86%,[86] the risk premium of 7%, the jump intensity η = 0.1, the logarithm of the expected jump size μ_K = -6.75%, the jump volatility σ_K = 8.53%, the transaction cost rate for stock purchase (sale) of 1% (0). Following our rule of thumb determining the lattice size, we used a tree with 81 branches for this parameter set.

Although the exact numerical values of Liu and Lowenstein (2008) are not available, it is clear that our diagram is virtually identical to the corresponding Liu-Lowenstein Figure

[86] This value of the total volatility corresponds to the value of the diffusion component of 12.39%.

6. Hence, there is reason to believe that our numerical algorithm has equally good convergence and approximation properties to the "true" continuous time solution as the alternative approximation through the Erlang-distributed horizon of the Liu-Lowenstein (2002, 2008) approach.

Table 15

Results for Liu-Loewenstein (2008) Parameters

T	$1/\overline{\lambda}_{jd}$	$1/\underline{\lambda}_{jd}$
0.25	0.194	1.760
0.5	0.190	0.665
1	0.185	0.397
2	0.179	0.299
5	0.167	0.256
10	0.161	0.247
15	0.161	0.246
20	0.161	0.246
25	0.161	0.246

The table displays the results for the NT region for mixed jump-diffusion process for the Liu-Loewenstein (2008) parameters: the RRA δ = 5, the total volatility σ = 12.86%, the risk premium of 7%, the jump intensity $\eta = 0.1$, the logarithm of the expected jump size μ_K = -6.75%, the jump volatility σ_K = 8.53%, the transaction cost rate for stock purchase (sale) of 1% (0).

5 Concluding Remarks

We presented an efficient numerical solution to the problem of deriving the NT region in discrete-time finite-horizon case for iid risky asset returns. We also corrected errors in an earlier study. The solution to our main research question indicates that the major factor driving the NT region for the mixed process apart from its diffusion counterpart is the jump intensity. It remains an empirical question whether the jump intensity estimated from market data would lead to major changes in portfolio rules compared to the simple diffusion case. Further, an empirical study may examine relative gains or losses in derived utility resulting from the adoption of either investment policy and tested with the observed paths of an index.

A factor that may modify the influence of the jump intensity on the NT region is intermediate consumption. We hypothesize that with intermediate consumption even relatively low jump intensity may lead to the portfolio rules for the mixed process relatively far apart from its diffusion counterpart. The reason for this conjecture is the plausibility that the risk aversion will sway an agent from holding the risky asset given that a large proportion of the risky asset in the agent's portfolio may lead to low consumption states at intermediary dates in the presence of jumps. The verification of this conjecture should be a topic of future research.

References

Boyle, P. P. and X. Lin, 1997, "Optimal Portfolio Selection with Transaction Costs," *North American Actuarial Journal* 1, 27-39.

Carr, P., 1998, "Randomization and the American Put," *Review of Financial Studies* 11, 597-626.

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.

Davis, M. H. A. and A. R. Norman, 1990, "Portfolio Selection with Transaction Costs," *Mathematics of Operations Research* 15, 676-713.

Dumas, B. and E. Luciano, 1991, "An Exact Solution to a Dynamic Portfolio Choice Problem under Transactions Costs," *Journal of Finance* 46, 577-596.

Genotte, G. and A. Jung, 1994, "Investment Strategies under Transaction Costs: The Finite Horizon Case," *Management Science* 40, 385-404.

Kamrad, B. and P. Ritchken, 1991, "Multinomial Approximating Models for Options with k State Variables," *Management Science* 37, 1640-1652.

Liu, H. and M. Loewenstein, 2002, "Optimal Portfolio Selection with Transaction Costs and Finite Horizons," *Review of Financial Studies* 15, 805-835.

Liu, H. and M. Loewenstein, 2008, "Optimal Portfolio Selection with Transaction Costs and 'Event Risk'," working paper.

Liu, J., F., A. Longstaff and J. Pan, 2003, "Dynamic Asset Allocation with Event Risk," *Journal of Finance* 58, 231-259.

Magill, M. J. P., and G. M. Constantinides, 1976, "Portfolio Selection with Transactions Costs," *Journal of Economic Theory* 13, 245-263.

Merton, R. C., 1969, "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *Review of Economics and Statistics* 51, 247-257.

Merton, R. C., 1971, "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory* 3, 373-413.

Oancea I. M. and S. Perrakis, 2007, "Stochastic Dominance and Option Pricing in Discrete and Continuous Time: an Alternative Paradigm," working paper.