

Electronic Thesis and Dissertation Repository

12-15-2021 2:00 PM

Essays in financial asset pricing

Dillon Ross Huddleston, *The University of Western Ontario*

Supervisor: Chung, Tai-Yeong, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Economics

© Dillon Ross Huddleston 2021

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Finance Commons](#)

Recommended Citation

Huddleston, Dillon Ross, "Essays in financial asset pricing" (2021). *Electronic Thesis and Dissertation Repository*. 8292.

<https://ir.lib.uwo.ca/etd/8292>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Three essays in financial asset pricing are given; one concerning the partial differential equation (PDE) pricing and hedging of a class of continuous/generalized power mean Asian options, via their (optimal) Lie point symmetry groups, leading to practical pricing formulas. The second presents high-frequency predictions of S&P 500 returns via several machine learning models, statistically significantly demonstrating short-horizon market predictability and economically significantly profitable (beyond transaction costs) trading strategies. The third compares profitability between these [(mean) ensemble] strategies and Asian option Δ -hedging, using results of the first. Interpreting bounds on arithmetic Asian option prices as ask and bid values, hedging profitability depends largely on securing prices closer to the bid, and settling midway between the bid and ask, significant profits are consistently accumulated during the years 2004-2016. Ensemble predictive trading the S&P 500 yields comparatively very small returns, despite trading much more frequently.

The pricing and hedging of (arithmetic) Asian options are difficult and have spurred several solution approaches, differing in theoretical insight and practicality. Multiple families of exact solutions to relaxed *power mean* Asian option pricing boundary-value problems are explicitly established, which approximately satisfy the full pricing problem, and in one case, converge to exact solutions under certain parametric restrictions. Corresponding hedging parameters/Greeks are derived. This family consists of (optimal) invariant solutions, constructed for the corresponding pricing PDEs. Numerical experiments explore this family behaviorally, achieving reliably accurate pricing.

The second chapter studies intraday market return predictability. Regularized linear and non-linear tree-based models enjoy significant predictability. Ensemble models perform best across time and their return predictability realizes economically significant profits with Sharpe ratios after transaction costs of 0.98. These results strongly evidence that intraday market returns are predictable during short time horizons, beyond that explainable by transaction costs. The lagged constituent returns are shown to hold significant predictive information not contained in lagged market returns or price trend and liquidity characteristics. Consistent with the hypothesis that predictability is driven by slow-moving trader capital, predictability decreased post-decimalization, and market returns are more predictable midday, on days with high volatility or illiquidity, and during financial crises.

Keywords: Partial differential equations, Asian option pricing and hedging, power means, mathematical finance, quantile options, (optimal) Lie point symmetry analysis and invariant solutions, separation of variables, Fourier-Bessel series, Hankel and Kontorovich-Lebedev transforms, special functions and integral transforms, Gaussian process regression; machine learning, statistically significant S&P 500 high-frequency return prediction, economically significantly profitable trading strategies, short-horizon market inefficiency, financial econometrics, fintech

Summary for Lay Audience

The pricing of two financial goods (options and stocks) is studied in three essays: One examines both the pricing and portfolio/risk management of a large family of options which encapsulates others previously proposed and studied. The second predicts five-minute samples of S&P 500 returns via several machine learning models: Statistically and economically significant results follow through trading strategies that are profitable beyond transaction costs. The third compares profitability between these strategies and others which follow from the portfolio/risk management results for options. In a framework with a trader deciding to engage in such portfolio management/optimization from day to day, given ask and bid option prices, profitability depends largely on securing prices closer to the bid, and settling midway between the bid and ask, significant profits are consistently accumulated during the years 2004-2016. Trading the S&P 500 using the machine learning results yields comparatively very small returns, despite trading much more frequently.

Option pricing and portfolio/risk management are difficult and have spurred several solution approaches, differing in theoretical insight and practicality. Multiple families of option pricing formulas are explicitly established, which approximately satisfy the full pricing problem, and in one case, are exact under certain restrictions. Corresponding risk management parameters are derived, and numerical experiments explore these families behaviorally, achieving practical, reliably accurate pricing.

The second chapter predicts five-minute market returns using models of varying sophistication and complexity. Models that average results of others perform best and yield economically significant and reliable profits after transaction costs which nearly equal their standard deviation. These results strongly evidence that five-minute market returns are predictable, beyond what can be explained by (trading behaviour due to) transaction costs. Returns of the (500) market constituents are shown to hold significant predictive information not contained in those of the S&P 500 alone; or statistics that characterize market solvency. Consistent with the hypothesis that slow-moving trader capital drives predictability, the latter decreased in the early millennium when various technological and regulatory factors increased market solvency. Similarly, market returns are more predictable on highly volatile or insolvent days/times and particularly during financial crises.

Co-Authorship Statement

I acknowledge my coauthors, Fred Liu and Lars Stentoft, for contributing equally to chapter 2.

Acknowledgments

This work was supported by Mitacs through the Mitacs Research Training Award.

Contents

Abstract	ii
Summary for Lay Audience	iii
Co-Authorship Statement	iv
Acknowledgments	v
List of Figures	ix
List of Tables	xi
List of Appendices	xii
Introduction	1
1 Asian option pricing and hedging with generalized averaging	3
1.1 Introduction	3
1.1.1 Mathematical and market environment. Problem statement	3
1.1.2 Financial and economic motivations and applications	10
1.1.3 Relation to β -quantile options	11
1.1.4 Existing approaches to (arithmetic) Asian option pricing	13
Monte Carlo methods	14
Series expansions	15
Integral transforms	16
Numerical PDE solvers	16
Path integrals	18
Price bounds	19
Low volatility, σ , and/or short maturity, T	19
1.1.5 Broad methodology and results	20
A primer on Lie point symmetry analysis	20
Further details and related literature	21
A primer on Gaussian process regression	24
1.1.6 List of contributions	26
1.2 Methods	27
1.2.1 Lie point symmetry analysis and consequent results	27
(Optimal) Lie point symmetries	28

	Invariant solutions	33
	Optimal invariant solutions	43
	Separation of variables for reduced pricing PDEs (1.30) and (1.51)	45
	Fourier-Bessel series solutions of pricing PDE (1.7)	50
	Hankel transform solutions of pricing PDE (1.7)	56
	Kontorovich–Lebedev transform solutions of pricing PDE (1.7)	58
1.2.2	Novel numerical problems and experiments	63
1.2.3	Results for $t = 0$ ($\Rightarrow A = t\mathcal{A}^\alpha = 0$)	63
	Preliminary studies of the function $\lambda(S_0, 0, 0; \sigma, r, T, K)$	64
	Design of experiments (1.111) and solutions of problem 3	64
	Predicting the function $\lambda(S_0, 0, 0; \sigma, r, T, K)$ in a general variable/parameter space (1.111)	69
	Predicting the function $\lambda(S_0, 0, 0; \sigma, r, T, K)$ in case $r \approx \sigma^2$	74
	Direct GPR predictions of the function $V(S_0, 0, 0; \sigma, r, T, K, S_0)$	80
1.2.4	Results for $t > 0$	83
1.3	Results	85
1.4	Conclusions	88
2	Intraday market predictability: A machine learning approach	89
2.1	Introduction	89
2.2	Methodology	94
2.2.1	Linear models	94
	Regularization	95
	Principal component regression	96
2.2.2	Nonlinear models	97
	Tree-based models	97
	Artificial neural networks	99
2.2.3	Ensemble methods	101
2.2.4	Evaluation criteria	101
	Statistical significance	102
	Economic significance	102
2.3	Empirical results	103
2.3.1	Market predictability	104
2.3.2	Economic significance	106
2.3.3	Economic significance after transaction costs	107
2.3.4	Robustness to training sample size	111
2.4	Additional analysis	111
2.4.1	Time-of-day patterns	112
2.4.2	Volatility and illiquidity effects	112
2.4.3	Impact of financial crisis	114
2.4.4	Comparison to autoregressive models	115
2.4.5	Effect of additional variables	115
2.5	Conclusion	116
3	Asian Δ-hedging and S&P 500 ensemble predictive trading profitability	127

3.1	Introduction	127
3.2	Methods	128
3.3	Results	129
3.4	Conclusions	137
	Conclusion	139
	Bibliography	141
A	Exponential operator formulas	146
B	Data	149
	B.1 Trade and quotes cleaning	149
	B.2 Additional firm characteristics	150
C	Hyperparameters	152
	Curriculum Vitae	155

List of Figures

1.1	Behavior of power means (1.3)	5
1.2	Behavior of the function $C^*(\alpha, r, \sigma, 0.25, t)$ defined by the inequality (1.92)	55
1.3	Behavior of the function $f(\zeta, t; 1, 0.1, 0.01, 0.25, 1, {}^7/_{640})$	56
1.4	Behavior of the Whittaker $W_{0, \nu_c}(z)$ -function of theorem 1.2.14	62
1.5	Behavior of the function $V(\sigma, r, T, 0, S_0; 0, 0, \lambda)$	65
1.6	Behavior of the function $V(0.1, 0.1, 0.25, K, 100; 0, 0, \lambda)$ with respect to K	66
1.7	Behavior of the function $V(\sigma, r, T, 100, S_0; 0, 0, \lambda)$	67
1.8	Experiments with suboptimal λ_0 and without \$0.04 tolerance	68
1.9	Experiments with optimal λ_0 and within \$0.04 tolerance	69
1.10	All experiments, and those unsolved	69
1.11	Quality of λ predictions, training on 1% of a general parameter space	71
1.12	Quality of λ predictions, training on 10% of a general parameter space	71
1.13	Quality of λ predictions, training on 25% of a general parameter space	72
1.14	Quality of λ predictions, training on 50% of a general parameter space	72
1.15	Quality of V predictions, training on 1% of a general parameter space	73
1.16	Quality of V predictions, training on 10% of a general parameter space	74
1.17	Quality of V predictions, training on 25% of a general parameter space	74
1.18	Quality of V predictions, training on 50% of a general parameter space	75
1.19	Quality of λ predictions, training on 1% of a parameter space satisfying $r \approx \sigma^2$	75
1.20	Quality of λ predictions, training on 10% of a parameter space satisfying $r \approx \sigma^2$	76
1.21	Quality of λ predictions, training on 25% of a parameter space satisfying $r \approx \sigma^2$	76
1.22	Quality of λ predictions, training on 50% of a parameter space satisfying $r \approx \sigma^2$	77
1.23	Quality of λ predictions, training on 75% of a parameter space satisfying $r \approx \sigma^2$	77
1.24	Quality of V predictions, training on 1% of a parameter space satisfying $r \approx \sigma^2$	78
1.25	Quality of V predictions, training on 10% of a parameter space satisfying $r \approx \sigma^2$	78
1.26	Quality of V predictions, training on 25% of a parameter space satisfying $r \approx \sigma^2$	79
1.27	Quality of V predictions, training on 50% of a parameter space satisfying $r \approx \sigma^2$	79
1.28	Quality of V predictions, training on 75% of a parameter space satisfying $r \approx \sigma^2$	80
1.29	Quality of direct V predictions, training on 1% of a parameter space with $r \approx \sigma^2$	81
1.30	Quality of direct V predictions, training on 10% of a parameter space with $r \approx \sigma^2$	82
1.31	Quality of direct V predictions, training on 25% of a parameter space with $r \approx \sigma^2$	82
1.32	Quality of direct V predictions, training on 50% of a parameter space with $r \approx \sigma^2$	83
1.33	Quality of direct V predictions, training on 75% of a parameter space with $r \approx \sigma^2$	83
1.34	Prices matching cubic spline-extrapolations of $\lambda(100, A, t; 0.1, 0.1, 0.25, 100)$	84
2.1	R_{OOS}^2 and trading costs	105

2.2	Cumulative returns by model	108
2.3	Cumulative returns after transaction costs by model	110
2.4	R_{OOS}^2 and trading volume	113
3.1	Profitability comparison for December 30, 2016	130
3.2	Further profitability comparison for December 30, 2016	131
3.3	Yet further profitability comparison for December 30, 2016	131
3.4	Charm prior to certain exercise on December 30, 2016	132
3.5	Profitability comparison with ensemble trading and multiples of the S&P 500 .	133
3.6	Profitability by maximum allowable balancing cost	134
3.7	Mean strategy and realized monthly illiquidity and kurtosis	136
3.8	Realized monthly extrema, momentum and skewness	136
3.9	Mean strategy profitability and realized monthly volatility and volume	137
3.10	Mean strategy net holding and waiting times	137
3.11	Mean trading strategy and realized monthly statistical sample cross correlations	138

List of Tables

1.1	Commutators, $[X_i, X_j] \equiv X_i X_j - X_j X_i$, of symmetries (1.22) with $A = t\mathcal{A}^\alpha$	29
1.2	Commutators, $[e_i, e_j] \equiv e_i e_j - e_j e_i$, of symmetries (1.23) with $A = t\mathcal{A}^\alpha$	30
1.3	Lagrange systems and characteristics of symmetries (1.25) [excluding Y_2]	32
1.4	Reduced pricing PDEs (1.7) of symmetries (1.25) [excluding Y_2]	32
1.5	Reduced pricing PDE (1.51) Bessel function solutions by constraints imposed. .	47
1.6	Pricing PDE (1.7) Bessel function solutions by constraints imposed.	49
1.7	Performance measures of V and λ predictions in a general parameter space . . .	73
1.8	Performance measures of V and λ predictions in a parameter space with $r \approx \sigma^2$	81
1.9	Performance measures of direct V predictions in a parameter space with $r \approx \sigma^2$	84
2.1	Market predictability	118
2.2	Excess Returns and Sharpe ratios	119
2.3	Excess Returns and Sharpe ratios with transaction costs	120
2.4	Market predictability (percentage R_{OOS}^2) and profitability (Sharpe ratio) by training duration post-decimalization	121
2.5	Market predictability (percentage R_{OOS}^2) and profitability (Sharpe ratio) by time post-decimalization	122
2.6	Market predictability (percentage R_{OOS}^2) and profitability (Sharpe ratio) by volatility in the post-decimalization period	123
2.7	Market predictability (percentage R_{OOS}^2) and profitability (Sharpe ratio) by illiquidity in the post-decimalization period	123
2.8	Market predictability (percentage R_{OOS}^2) and profitability (Sharpe ratio) by crisis period	124
2.9	Market predictability (percentage R_{OOS}^2) and profitability (Sharpe ratio) for autoregressive models	125
2.10	Market predictability (percentage R_{OOS}^2) and profitability (Sharpe ratio) with additional characteristics in the post-decimalization period	126
C.1	Model hyperparameters	154

List of Appendices

Appendix A Exponential operator formulas	148
Appendix B Data	151
Appendix C Hyperparameters	153

Introduction

The 2007-2008 financial crisis severely undermined public trust in the world financial system. It was facilitated by investment banks' trading of esoteric financial products like mortgage-backed securities (MBSs) and collateralized debt obligations (CDOs): In lay terms, 'slicing and dicing' Americans' mortgages, and recombining them such that products containing primarily junk debt were incorrectly prime rated. Short-selling by investment banks of these products, containing subprime mortgages, constituted the banks betting against the corresponding mortgage holders. One problem identified following the crisis was the opacity of financial derivatives, in particular, to the financial professionals trading them, with an only superficial knowledge of how the products work. A disconnect exists between the traders and those who design and analyze such products, often math and physics doctorates with little knowledge of their real-world application.

Closing this gap is one way of restoring public faith in the financial system, by legitimizing its activities. This happened in 1973, following the publication of the option-pricing formula of Black and Scholes [38], regarding the goings-on of the Chicago Board Options Exchange. Furthermore, the volume of options traded increased significantly as a result of that paper. However, many derivatives since introduced are much more complicated, including MBSs and CDOs. So it is important to have financial professionals competent in the analysis and pricing of these products, who are also cognizant of practicalities concerning their implementation. My economics Ph.D. research addresses this goal.

Financial options are contracts whose payoffs depend on an underlying asset. A 'strike price' is specified, to which some function of the underlying asset's price path is compared upon the contract's expiry, in determining the payoff. The contract allows the holder to 'exercise' the option by buying or selling the underlying asset at the strike price. Various options allow for different exercise schedules. European options, exercisable only at expiry, and the American kind, exercisable any time preceding expiry, are classified as 'vanilla.' Their payoffs depend only on the terminal price of the underlying asset, typically a stock. While these options comprise the greatest volume in real-world trading, other 'exotic' options exist. One popular example, which chapter 1 aims to price, is the Asian option, whose payoff depends on the average underlying asset price over the option's lifetime. This is an important problem, and many numerical, approximate, and hybrid approaches have been proposed, but an exact pricing formula, like that of Black and Scholes [38], remains elusive. My research has elicited an approximate formula that is exact under certain restrictions and a good approximation when these restrictions are 'nearly' satisfied. Numerical experiments illustrating this result indicate suit-

ability for both pricing and hedging. In fact, despite trading far less frequently, some hedging portfolios in chapter 3 consistently and vastly outperform the profitability of trading strategies derived from predictive machine learning models. For derivatives generally and Asian options, in particular, proper pricing encourages their trade, promoting investment, economic growth, financial institutions' reputations, and social welfare.

Chapter 2 studies market return intraday predictability. Numerous models were run on S&P 500 return data, for prices sampled at the half-hour, five-minute, and one-minute frequencies, using the Compute Canada high-performance computing cluster. As such, this constitutes a big-data, high-frequency trading analysis (the trading strategy being based upon price predictions). Beyond establishing benchmarks via standard linear machine learning models, a wide selection of nonlinear ones are considered, attempting to optimally construct such architectures to provide the greatest predictability of S&P 500 returns. The resulting trading strategies are shown to be statistically and economically significantly profitable, even after accounting for transaction costs.

In practice, this research offers market participants greater predictability, which is highly sought-after. Long-term market predictability has been tentatively established via various financial ratios, which are however not applicable to short-term, and in particular intraday, price moves. The long-term market predictability results from one form of the efficient markets hypothesis (EMH) (in short, that prices account for 'all' information available to the market, with various forms defined according to the scope of 'all'). This also implies that over short, near-instantaneous intervals, prices should remain constant, so long as there are no so-called 'news shocks' which constitute 'pulses' of information injected into the market at different moments. But news shocks also demand a transient adjustment period, during which increased trading (and pursuant transaction costs) may exacerbate market illiquidity, which offers potential predictability. In other words, while the information from a news shock is reconciled, the EMH is temporarily violated, as the information of the news shock has not been factored into prices. And so it may be possible to predict (better than a fair coin toss, at least) price movements during such transient periods, using machine learning models trained on high-frequency data, which is exactly the research objective. Such predictability also yields effective statistical arbitrage trading strategies, providing compelling financial incentive. The noted market illiquidity encompasses inefficiencies that violate the EMH so that the determined predictabilities gauge such violations' extent. Prior research has found some predictability using low-frequency data and linear models, but the non-linear machine learning models, trained on high-frequency data, suss out undiscovered predictabilities.

Both primary chapters 1 and 2 thus concern *illiquid* markets (in which Asian options are traded most frequently, e.g., those of commodities and foreign exchange); the brief, empirical/numerical 'bridge' chapter 3 compares the profitability of trading strategies derived from each, finding that Δ -hedging Asian options may significantly outperform statistical arbitrage achieved via machine learning predictions of intraday S&P 500 returns.

Chapter 1

Asian option pricing and hedging with generalized averaging

Abstract

The pricing and hedging of (arithmetic) Asian options poses a difficult problem, which has prompted the development of several solution approaches, differing in theoretical informativeness and practicality. Multiple families of exact solutions to relaxed *power mean* Asian option pricing boundary-value problems are explicitly established, which approximately satisfy the full pricing problem, and in one case, converge to exact solutions under certain parametric restrictions. Corresponding hedging parameters/Greeks are derived. This family consists of (optimal) invariant solutions, constructed for the pricing partial differential equations of a class of continuous/generalized power mean Asian options, via their (optimal) Lie point symmetry groups. Numerical experiments explore the behavior of this family, and lead to accurate, reliable, consistent pricing models suitable for use in practice.

1.1 Introduction

The mathematical and market structure of the problem to be solved are first presented, along with relevant details on analytical option pricing, in particular for European vanilla and power mean Asian options, illustrating why certain among the latter are (not) easily priced as in the vanilla case. A related exotic option, the so-called β -quantile variety, is then discussed. Existing approaches to (arithmetic) Asian option pricing are then reviewed. Finally, the introduction concludes with a discussion of the employed methodology and broad results.

1.1.1 Mathematical and market environment. Problem statement

Let a *maturity*, $T > 0$, be given, and consider a complete, filtered probability space, $(\Omega, \mathcal{F}, \mathbb{F}, P)$, with the natural filtration of \mathcal{F} with respect to a standard (real-valued) Brownian motion, $(W_t)_{t \in [0, T]}$, $\mathbb{F} \equiv (\mathcal{F}_t)_{t \in [0, T]}$. Let the spot price of a risky, underlying asset (first factor), $(S_t)_{t \in [0, T]}$, evolve from initial value, $S_0 \geq 0$, according to a geometric Brownian motion (GBM); i.e., to

solve the following stochastic differential equation (SDE):

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \quad (1.1)$$

Here, $\mu \in \mathbb{R}$ and $\sigma > 0$ are given. Letting P be the unique risk neutral measure in this framework,¹ it follows that $\mu = r$, the risk-free interest rate according to which a numeraire bond, $(B_t)_{t \in [0, T]}$, is defined to evolve from initial value, $B_0 = 1$:

$$\frac{dB_t}{B_t} = r dt \quad (1.2)$$

A *European call (put)* option is a contract which affords the holder the right to buy (sell) one unit of the underlying asset, S_T , at a *strike* price, $K \geq 0$, upon maturity of the contract at time T . The *payoff*, $F(S_T, T)$, of such a contract at maturity is $\max\{S_T - K, 0\}$ ($\max\{K - S_T, 0\}$). Thus, denoting by $V(S, t)$ the no-arbitrage price of the option at time t (dropping the subscript on S_t), the problem of pricing the option contract when it is written at time $t = 0$, is to solve for $V(S_0, 0)$, given the asset dynamics (1.1) and (1.2), and subject to the terminal condition, $V(S_T, T) = F(S_T, T)$. Assuming both asset markets to be liquid and frictionless, i.e. that arbitrary transactions of each are not subject to costs, Black and Scholes [38] demonstrated that a closed-form solution to this problem exists, and derived their celebrated pricing formulae.

Note that the European payoff is highly sensitive to the terminal value of the underlying asset, S_T : This payoff is largely subject to vagaries in the market price of the underlying asset near expiry. This is understandably not an attractive feature of such contracts for many (would-be) holders, to the extent of such vagaries: There is motivation to reduce payoff sensitivity to terminal underlying asset price, which may be accomplished, for instance, by *averaging* the underlying asset price over the lifetime of the option, i.e., the interval $[0, T]$.² Consider the following family of *power means*:³

$$(\forall \alpha \in \overline{\mathbb{R}}) \quad \mathcal{A}_\alpha[S_t] \equiv \mathcal{A}_\alpha[(S_\tau)_{\tau \in [0, T]}] \equiv \left[\frac{1}{t} \int_0^t S_\tau^\alpha d\tau \right]^{\frac{1}{\alpha}} \quad (1.3)$$

(Note that these functionals are evaluated for a sample path of the GBM, $(S_\tau)_{\tau \in [0, T]}$, which is known to be continuous almost everywhere; see Pascucci [245].) For only two variables (say asset prices S_{t_i} at times $i \in \{1, 2\}$), some level curves of (discrete) power means are plotted in figure 1.1a. The linear plot corresponds to $\alpha = 1$, and α then decreases (increases) moving

¹The uniqueness of the risk neutral measure for assets evolving according to equation (1.1), follows from the fundamental theorem of asset pricing; see the extensive account of analytical option pricing by Pascucci [245].

²More generally, an average over some interval, say $[t, \bar{t}]$, with $0 \leq t \leq \bar{t} \leq T$, or a union of such intervals, may be considered. The solution process then becomes piecewise, treating the option as European over times when the underlying asset is not averaged, and otherwise as in this paper, i.e., for individual sub-intervals treated as option lifetimes of their own. This process is best implemented numerically; see Wilmott, Dewynne, and Howison [315].

³The (affinely) *extended* real line, $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$, includes the real line, \mathbb{R} , and two ends; the positive and negative points at infinity: It is the two-point compactification of \mathbb{R} , in contrast to the one-point compactification (Alexandroff extension) obtained by projectively extending \mathbb{R} with a single end/point at infinity; $\widehat{\mathbb{R}} \equiv \mathbb{R} \cup \{\infty\} \cong \mathbb{RP}^1$, (homeomorphic to) the real projective line.

rightward (leftward) across each plot. Various power means are plotted in figure 1.1b for the spot price of West Texas Intermediate crude oil, from January 3 through November 20, 2017; data which was obtained from the United States Energy Information Administration.

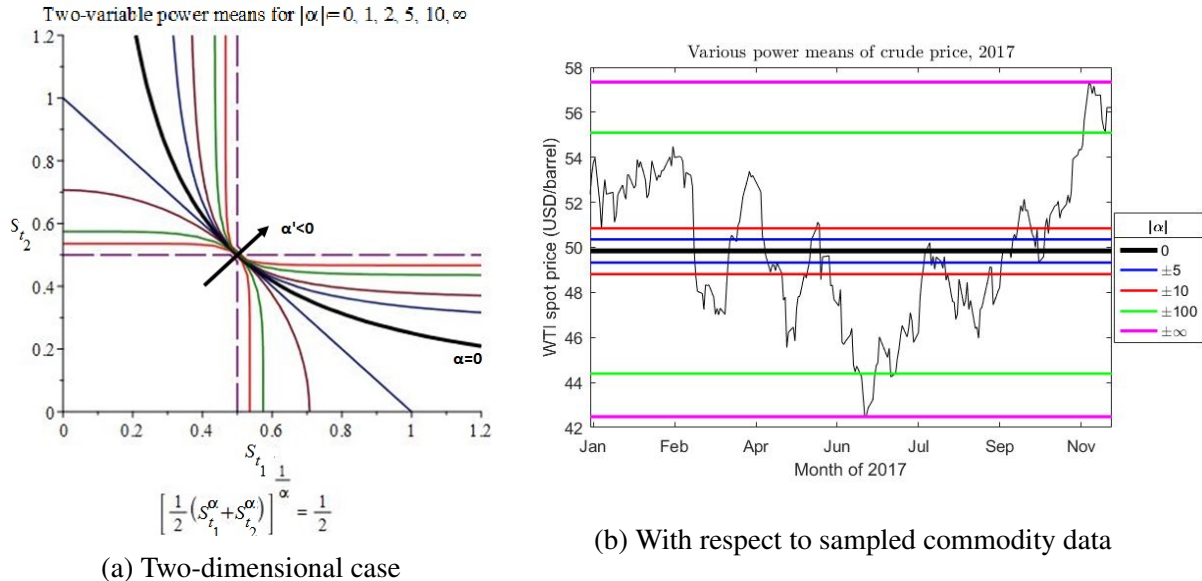


Figure 1.1: Behavior of power means (1.3)

A *floating-strike* European Asian call (put) option is a contract which affords the holder the right to buy (sell) one unit of the underlying asset, S_T , at a *floating* strike price, $\mathcal{A}_T \equiv \mathcal{A}_\alpha[S_T]$,⁴ upon maturity of the contract at time T . The payoff, $F(S_T, \mathcal{A}_T, T)$, of such a contract at maturity is $\max\{S_T - \mathcal{A}_T, 0\}$ ($\max\{\mathcal{A}_T - S_T, 0\}$). Note that this is as for a (non-Asian) European option, replacing the fixed strike price, K , with the floating one, \mathcal{A}_T , determined as an average price path of the underlying asset over the lifetime of the option: Similarly, a *fixed-strike* European Asian call (put) option is a contract which affords the holder the right to buy (sell) one unit of the *second factor*, \mathcal{A}_T , at a fixed strike price, K , upon maturity of the contract at time T . The payoff, $F(S_T, \mathcal{A}_T, T)$, of such a contract at maturity is $\max\{\mathcal{A}_T - K, 0\}$ ($\max\{K - \mathcal{A}_T, 0\}$). Thus, denoting by $V(S, \mathcal{A}, t)$ the no-arbitrage price of the option at time t (dropping the subscripts on S_t and \mathcal{A}_t), the problem of pricing the option contract when it is written at time $t = 0$, is to solve for $V(S_0, \mathcal{A}_0, 0)$, given the asset dynamics (1.1)-(1.3), and subject to the terminal condition, $V(S_T, \mathcal{A}_T, T) = F(S_T, \mathcal{A}_T, T)$. Assuming both bond and underlying (first factor) asset markets to be liquid and frictionless, the difficulty of this problem depends greatly on the value of the power mean index, $\alpha \in \bar{\mathbb{R}}$: When $\alpha = 0$, the power mean (1.3) is *geometric*, and so the resulting Asian options are also qualified as such. Recall that, for any $t \in [0, T]$, the random variable, S_t , from the GBM, $(S_t)_{t \in [0, T]}$, is lognormally distributed, and that the geometric mean of lognormally distributed random variables is likewise lognormally distributed: It is then a straightforward exercise with the resulting correlated GBMs, $(S_t)_{t \in [0, T]}$ and $(\mathcal{A}_t)_{t \in [0, T]}$, to derive a closed-form solution for the geometric Asian option pricing problem, using that of Black and

⁴The subscript, $\alpha \in \bar{\mathbb{R}}$, is often suppressed, but understood as an implicit parameter until explicitly specified.

Scholes [38]; see Zhang [327].

When $\alpha = (-)\infty$, the power mean (1.3) yields the *supremum (infimum)*, and the resulting (limiting) Asian options are known as *lookback call (put) options*. It is also a straightforward exercise to derive a closed-form solution for the lookback option pricing problem, leveraging that of Black and Scholes [38]; see Goldman, Sosin, and Gatto [149] and Wilmott, Dewynne, and Howison [315].⁵ When $\alpha = (-)1$, the power mean (1.3) is *arithmetic (harmonic)*, and so the resulting Asian options are (may be) also qualified as such. The cases discussed, $|\alpha| \in \{0, 1, \infty\}$, are all those whose corresponding Asian options have apparently been studied before, and one contribution of the present paper is to propose, and argue for the potential benefits of, the entire family, indexed by $\alpha \in \overline{\mathbb{R}}$. The analysis for such α is carried up to a point, deriving generalizations, corrections, and justifications of results of Antoniou [15] for arithmetic Asian options, prior to focusing on the arithmetic case for the remainder of the paper. Indeed, the arithmetic case has been most thoroughly studied, partly because it is the most important and widely traded (including on the Chicago Board Options Exchange), but also because its analysis is much more difficult than the cases of geometric Asian and lookback options.⁶ Regarding the harmonic case, it is shown in corollary 1.2.4 that the options with $\alpha < 0$ may be priced in terms of their $-\alpha$ counterparts; a sort of duality (parity) result. But it is clearly seen in the analysis that, among the proposed family of power mean Asian options, the arithmetic (and so harmonic) case stands out as a particularly simple one, and the case of $|\alpha| \notin \{0, 1, \infty\}$ is yet more difficult than the already challenging arithmetic case.

It is natural that a fixed-strike Asian call should be more expensive for higher α , as is its payoff. Similar monotonicity arguments follow immediately for floating/average-strike puts and calls. The ensuing analysis can not only verify that such monotonicity constraints are satisfied, but by precisely how much, depending on the values of α concerned. Particularly, this could inform empirical work concerning price differentials of arithmetic and geometric Asian options.

There are two primary approaches to analytical (as opposed to numerical) option pricing; martingale- and partial differential equation (PDE)-based. Both are treated extensively in Pascucci [245]. Martingale-based methods exploit the fundamental theorem of asset pricing: With respect to the unique risk-neutral measure, say Q , any financial product derived from a GBM is a martingale, and therefore its value may be computed as the (risk-free rate, r) discounted, expected value of its payoff. For example, if such a derivative contract has payoff at maturity, T , equal to $F(S_T, T)$, then its value, $V(S_t, t)$, for any time, $0 \leq t \leq T$, is computed as follows:

$$V(S_t, t) = e^{-r(T-t)} \mathbb{E}_Q[F(S_T, T) | \mathcal{F}_t] \quad (1.4)$$

In the case of European (geometric Asian) options, the log-normality of (the geometric mean of) the underlying asset, S_t , allows explicit computation of this expectation. However, the dis-

⁵Note that the power mean functional (1.3) has not apparently before been used to define payoffs for a corresponding family of Asian options, as this paper proposes, other than instrumentally as a means to derive the pricing formula for lookbacks in the limit $|\alpha| \rightarrow \infty$, by Wilmott, Dewynne, and Howison [315].

⁶Since lookbacks perfectly solve the market timing problem, by allowing the holder access to the optimal (supremal or infimal) price of the underlying asset over the lifetime of the option, they are always exercised. This significantly simplifies their analysis, since the payoff is smooth, and in this sense they are not really ‘options.’

tribution of the *arithmetic* (and more generally, other power) mean of the underlying asset is unknown, which is one reason for the much greater difficulty of pricing Asian options contingent on such means. Martingale and PDE-based methods of option pricing are intimately related through the Feynman-Kac formula, which identifies the martingale (1.4) as the solution of a corresponding parabolic, *pricing* PDE [its associated Fokker-Planck-Kolmogorov (FPK) equation]: In the case of a European (geometric Asian) option, this is the Black-Scholes [38] equation (a similar one listed by Zhang [327]). But this is not useful for other power mean Asian options, since such means of the underlying asset are unknown, and in particular are known *not* to be lognormal, so that they do not constitute a GBM correlated with that of the underlying asset, as is the case for geometric Asian options.

Fortunately, the PDE-based approach to option pricing also directly obtains via financial arguments: By constructing a portfolio consisting of one power mean Asian option and a number, Δ , of the underlying asset, it is possible to perfectly hedge, i.e., to remove the dW_t term from the SDE of, the portfolio. No-arbitrage implies that such a portfolio cannot evolve faster than the risk-free rate, and in the case of European contracts which only allow exercise at maturity, T , the portfolio must evolve *at* the risk-free rate. Expressed mathematically, the result is a PDE, the solution of which is the option price, $V(S, \mathcal{A}, t)$; see Antoniou [15] for the (standard) argument applied to arithmetic Asian options. At this point it is important to observe that the precise PDE obtained, which may strongly impact solution difficulty using a computer algebra system (CAS), depends entirely upon how the second factor, to this point taken as the power mean, \mathcal{A}_t , of the underlying asset, S_t , is defined in terms of the former factor and the time variable, t . Specifically, defining an alternative second factor, $A \equiv A(\mathcal{A}, t)$ (again dropping the subscripts on \mathcal{A}_t and A_t), the following pricing PDE form obtains via the standard Δ -hedging argument for the price, $V(S, A, t)$, of a power mean Asian option:

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + g(S, A, t)V_A + rS V_S - rV = 0 \quad (1.5)$$

(Subscripts denote partial differentiation.) The coefficient of V_A , corresponding to three candidate definitions of A , are as follows:

$$g(S, A, t) = \begin{cases} \frac{A}{\alpha t} \left[\left(\frac{S}{A} \right)^\alpha - 1 \right] & A = \mathcal{A} \\ \frac{S^\alpha - A}{t} & A = \mathcal{A}^\alpha \\ S^\alpha & A = t\mathcal{A}^\alpha \end{cases} \quad (1.6)$$

The bottom case is simplest in that it depends only on S , whereas the other two depend also on A and t . This simplicity is borne out in the derivation of lemma 1.2.1 using a CAS, which required approximately seconds, six hours, and 24 hours, for respectively the bottom, middle, and top definitions of A in formula (1.6). But this simplicity is not without cost, which can be seen by considering $\lim_{t \rightarrow 0^+} A_t$:

A_t	$\lim_{t \rightarrow 0^+} A_t$
\mathcal{A}_t	S_0
\mathcal{A}_t^α	S_0^α
$t\mathcal{A}_t^\alpha$	0

The ‘simplest’ case, viewed this way, is the most complex: Practically, it is paramount to price the option as $t \rightarrow 0^+$, when the contract is written. The limiting value of zero in this case becomes problematic at later stages of the analysis, when it results in complicated indeterminate expressions which typically vanish or diverge; neither is useful from the standpoint of obtaining an actual option price. So it becomes worthwhile to consider alternate definitions of A , and in particular the middle one in formula (1.6), which is ‘simpler’ than the top one as described, and has a nonzero limit as $t \rightarrow 0^+$, so long as $|\alpha| < \infty$ and $S_0 > 0$; the case of $S_0 = 0$ is trivial, since all options will (almost always) have zero payoff and therefore zero price. However, preliminary analyses of the two upper definitions of A in formula (1.6), exposed no major advantage to the (algebraically) simpler, bottom case, and so shortly after lemma 1.2.1, the definition $A \equiv t\mathcal{A}_t^\alpha$ is assumed. For that reason and clarity of exposition, the same assumption is made for the remainder of the current section. In this case, the pricing PDE (1.5) becomes

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + S^\alpha V_A + rS V_S - rV = 0 \quad (1.7)$$

This PDE is subject to the terminal condition that the option price match its payoff at maturity:

$$V(S_T, A_T, T) = F(S_T, A_T, T) \equiv \max \left\{ \left(\frac{A_T}{T} \right)^{\frac{1}{\alpha}} - K, 0 \right\} \quad (1.8)$$

(This is the payoff for a fixed-strike call option, assumed without loss of generality as discussed in section 1.3.) The primary theoretical contribution of this paper, in the case with $\alpha = 1$ of an arithmetic Asian option (and so $\alpha = -1$ by corollary 1.2.4), is a family of exact solutions to the boundary-value problem (BVP) consisting of PDE (1.7) and terminal condition (1.8), indexed by the nonnegative real numbers, over the domain $(S, A, t) \in (0, \infty)^2 \times (0, T)$. Further, this family converges, under parametric restrictions on r , σ , and T , discussed in the preamble to theorem 1.2.6, to a (single) exact solution to the BVP consisting of PDE (1.7) and terminal condition (1.8), *and* the following boundary condition (BC) in the case $S = 0$:

$$V(0, A_t, t) = e^{-r(T-t)} \max \left\{ \frac{A_t}{T} - K, 0 \right\} \quad (1.9)$$

This BC is imposed since, if $S = 0$ at any time, $0 \leq t < T$, then (almost always) $S \equiv 0$ for all times, $t \leq \tau \leq T$, and $A_\tau \equiv A_t$. The payoff is thus a constant, which is its own expectation with respect to the risk-neutral measure, and BC (1.9) follows from the martingale pricing formula (1.4). Note also that $t \rightarrow 0^+ \Rightarrow A \rightarrow 0^+$, so that the boundary $A = 0$ is approached when the pricing problem is solved when the contract is written at time, $t = 0$. Two additional asymptotic BCs/transversality conditions, as $S \rightarrow \infty$ and $A \rightarrow \infty$, are required to specify the full BVP for PDE (1.7) in all three variables, S , A , and t , and are discussed by Shreve [279]. However, the dimension reduction that is achieved in reducing this PDE to the heat equation below, requires only the terminal condition (1.8) to be fully specified. As such, the BVP solved exactly by the family derived below, is a *relaxation* of the full BVP for PDE (1.7), including BC (1.9) and the asymptotic BCs noted. To summarize and consolidate this discussion, *the primary theoretical contribution of this paper, is a family of exact solutions to the arithmetic Asian option BVP consisting of PDE (1.7) and terminal condition (1.8), over the domain $(S, A, t) \in (0, \infty)^2 \times (0, T)$. This problem is a relaxation of the full BVP for PDE (1.7), including additionally BC (1.9) and*

the asymptotic BCs discussed by Shreve [279]. Under parametric restrictions on r , σ , and T , all members of the exact family of solutions to the relaxed BVP converge to an exact solution of the full BVP.

Prior to discussing the relationship between the power mean Asian options proposed and β -quantile options, in which the power mean index $\alpha \in \overline{\mathbb{R}}$ is crucial, the BVP for this general case is presented, and will also be referred to in the general development of section 1.2.1. The general form of BC (1.9) is

$$V(0, A_t, t) = e^{-r(T-t)} \max \left\{ \left(\frac{A_t}{T} \right)^{\frac{1}{\alpha}} - K, 0 \right\} \quad (1.10)$$

One of two asymptotic BCs discussed by Shreve [279], as well as by Kemna and Vorst [177], in the case of an arithmetic Asian option with $\alpha = 1$, relates to the fact that an (arithmetic) Asian call option is certainly exercised once $A_t \geq TK^\alpha$ (with $\alpha = 1$) which, similarly to the case of lookbacks, permits a straightforward exact solution. Furthermore, it provides a BC in case $A = KT^\alpha$, which features prominently in sections 1.2.1 and 1.2.4. The derivation follows à la Kemna and Vorst [177]: At time $t \leq T$, given $A_t \geq TK^\alpha$, a power mean call option is sure to be exercised, yielding the following payoff at maturity, T :

$$\left(\frac{A_t}{T} \right)^{\frac{1}{\alpha}} - K + \left(\frac{1}{T} \int_t^T S_\tau^\alpha d\tau \right)^{\frac{1}{\alpha}} \quad (1.11)$$

In order to replicate the latter (S_t -dependent) term of this payoff, an investor may sell a discounted proportion of ‘composite’ underlying S_τ^α , $e^{-r(T-\tau)}/T$, and use the proceeds to purchase bonds with the risk-free rate r , during an increment $[\tau, \tau + d\tau]$. Over the duration $[t, T]$, this accumulates $[1 - e^{-r(T-t)}]/rT$ units of underlying, and payoff (1.11) discounted to time t :

$$V(S_t, A_t, t)|_{A_t \geq TK^\alpha} = \left[\left(\frac{A_t}{T} \right)^{\frac{1}{\alpha}} - K \right] e^{-r(T-t)} + \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} S_t$$

Of course, this reduces to the formula (1.45) of Kemna and Vorst [177] in case $\alpha = 1$, and substituting $A = TK^\alpha$ yields the corresponding BC:

$$V(S_t, TK^\alpha, t) = \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} S_t \quad (1.12)$$

In contrast to the second asymptotic BC presented by Shreve [279] in case $\alpha = 1$, a BC similar in spirit to BC (1.12) is derived in the same case by Kemna and Vorst [177]. Namely, it concerns the behavior of V for large S_t , and relies upon Jensen’s inequalities:

$$\lim_{S_t \rightarrow \infty} V_S(S_t, A_t, t) = \lim_{S_t \rightarrow \infty} e^{-r(T-t)} \frac{\partial}{\partial S} \left(\frac{1}{T} \int_t^T S_\tau^\alpha d\tau \right)^{\frac{1}{\alpha}} = \lim_{S_t \rightarrow \infty} e^{-r(T-t)} \frac{A_T^{\frac{1-\alpha}{\alpha}}}{T^{\frac{1}{\alpha}}} \int_t^T S_\tau^{\alpha-1} d\tau = \begin{cases} \infty & \alpha > 1 \\ e^{-r(T-t)} \frac{T-t}{T} & \alpha = 1 \\ 0 & \alpha < 1 \end{cases} \quad (1.13)$$

Of course, this reduces to the asymptotic BC of Kemna and Vorst [177] and Shreve [279] in case $\alpha = 1$. It reflects the fact that, as $S_t \rightarrow \infty$ (and so $A_t/A_T \rightarrow 0$) for $t < T$, it becomes certain that $A_T \geq TK^\alpha$ and the power mean Asian call option will be exercised, yielding the BC as a derivative (with respect to S_t) of the discounted payoff (1.11) at time t . It is also seen that the arithmetic case with $\alpha = 1$ chaotically affects this BC: The contribution of remaining realizations, $(\forall t \leq \tau \leq T) S_\tau$, toward A_T , increases V without bound (intuitively for a lookback call), proportionally to the time to maturity, or does not impact V , respectively as $\alpha \gtrless 1$.

Prior to discussing the (statistical) relationship between the proposed power mean Asian options and another class of path-dependent, known as *quantile*, options, some remarks are collected regarding the financial and economic motivations, applications, and complications of pricing and hedging power mean Asian options.

1.1.2 Financial and economic motivations and applications

Recall that Asian options are motivated by a desire to mitigate the payoff sensitivity to terminal underlying asset price. This is particularly acute in illiquid/thin(ly-traded) markets, in which trading occurs infrequently and may be subject to transaction costs and restrictions on short-selling, trade quantity and transaction time and date. In such circumstances, with fewer market participants, there is potential for the actions of a ‘large’ trader to substantively alter the price of the underlying asset. Whether such market manipulation is incidental or intentional, and in the latter case whether or not it is directed toward a particular holder of a European (vanilla) option on the underlying asset, such an holder may prefer not to be subject to such uncertainty. Asian options address just this concern, and so are popular in illiquid markets, including those for commodities, currency/foreign exchange, and interest rates; see Kemna and Vorst [177].

Suppose a firm acquires a production input periodically, exposing itself to volatility in the underlying (input) price on transaction dates. It may want to forgo the benefit of low input prices on these dates, in exchange for protection against high ones. Taking an Asian option on the input accomplishes this: The firm guarantees that the payoff is based on a mean of the input price, rather than that on transaction dates. This scenario accounts for the volume of Asian options traded in practice. For the same reason, Asian options naturally insure against the risk of ‘large’ traders, i.e., ones trading in sufficiently large quantities of the underlying asset, so as to substantively alter its price. The ability to self-insure fosters investment, promoting economic growth and social welfare. And proper option pricing encourages their trade.

Given that Asian options are thus largely motivated by their advantages in illiquid markets, some comments are in order regarding modeling the dynamics of the underlying asset: As in formula (1.1), the assumption is made that the latter evolve according to a GBM, the canonical process in stochastic calculus models of mathematical finance. Though this is known to have undesirable empirical implications, e.g., thin tails due to each underlying price, S_t , being normally distributed, it is chosen for tractability purposes. And so it is taken here, to develop novel results on GBM-driven power mean Asian options, using classical Lie point symmetry analysis, which may constitute a foundation for subsequent investigations driven by other pro-

cesses, e.g., jump-diffusion [by adding a compound Poisson process (CPP) to the GBM (1.1)]. In particular, imagining an illiquid market with finitely many traders transacting at distinct times, a more appropriate driving model could be a pure jump process without diffusion, i.e., a CPP without GBM (1.1). In either case of pure or jump-diffusion processes, the inclusion of jumps via an added CPP, results in a partial *integrodifferential* equation in place of pricing PDE (1.7). This naturally adds many complications, though (non-local/point) Lie symmetry concepts exist for such equations; see Meleshko [221] and Grigoriev et al. [151]. Thus it may be a viable, and particularly suitable, course for extending the following GBM-driven results on power mean Asian options.

Cummins and Geman [86], Geman [136], [137] and [140], Eydeland and Geman [114], and P ezier and Scheller [249] discuss some basic applications of Asian options as weather, electricity/energy/power and (portfolio) insurance derivatives; the latter particularly note that weather are typically modeled as Asian options, as are those on oil. Other discussions of Asian electricity options are due to Weron [312] and Fanelli, Maddalena and Musti [116]. Studies of Asian options on general commodities markets include those of Cortazar and Schwartz [83], Fusai, Marena and Roncoroni [131], and Shen et al. [277]. Majed [213] surveys applications of Asian options to the petroleum industry, Ph elipp e-Guinvarch and Cordier [251] consider Asian options as risk management tools for European hog producers, and Koekebakker, Adland and S odal [180] develop approximate closed-form prices for Asian options on freight rates. Ritchken and Sankarasubramanian [259], Cheuk and Vorst [73], and Chacko and Das [64] and [63] study Asian interest-rate derivatives; Nielsen and Sandmann [234] address Asian exchange-rate options. Easton [107] prices bonds with embedded Asian options, and J orgensen [208] considers Asian options in the context of pension schemes.

Australian options are similar to Asian ones, in that they have payoffs contingent on the arithmetic average with the underlying asset as numeraire, e.g., $\max\{A_\tau/T_{S_\tau} - K\}$ for a call, or the reciprocal quantity, e.g., with call payoff $\max\{T_{S_\tau}/A_\tau - K\}$. Such options have been traded on the Australian Securities Exchange since 1992; see Moreno and Navas [225] and Ewald, Menkens and Ting [113]. Funahashi and Kijima [129] present some further results for Australian, as well as several other kinds of Asian options.

1.1.3 Relation to β -quantile options

Miura [223] proposed novel, so-called ' β^7 -quantile' options, which are identical to the power mean Asian options proposed presently, with the exception that they use, rather than power means (1.3), a similar measure of central tendency; *quantiles*:

$$(\forall \beta \in [0, 1]) \quad \inf \left\{ x \geq 0 \left| \frac{1}{t} \int_0^t \mathbb{I}_{\{S_\tau < x\}} d\tau \geq \beta \right. \right\} \quad (1.14)$$

(Here, $\mathbb{I}_{\{\cdot\}}$ denotes an indicator function.) Given $\beta \in [0, 1]$, the corresponding quantile (1.14) represents a barrier value of the underlying asset, S_t , under which the latter is occupied 100 β %

⁷Miura [223] used α -quantile, but here α has already been used in defining the power means (1.3).

of the duration, $[0, t]$. As such, β -quantile options are occupation time derivatives; see Pechtl [248], in which the similarity of (arithmetic) Asian and β -quantile options is also noted. On one hand, the quantiles (1.14) are preferable to the power means (1.3), as Miura [223] demonstrated in computing their distribution, whereas that of the power means (1.3), and in particular the arithmetic case with $\alpha = 1$, remains unknown; see Dassios [93] and Bertoin, Chaumont, and Yor [32]. On the other hand, the power means (1.3) are preferable to the quantiles (1.14), as they imply the continuous pricing PDE (1.5), whereas a pricing PDE implied by the quantiles (1.14) would, at best, retain a Dirac delta functional evaluated at the quantile itself; i.e., a non-classical/distributional⁸ term, as opposed to the classical (and continuous) functions (1.6).

The computation by Miura [223] of the distribution of the quantiles (1.14), enabled Akahori [7], Dassios [92], Yor [320], and Embrechts et al. [111], to obtain further results concerning the same distributions, as well as pricing formulas for β -quantile options. However, these martingale-based approaches achieve final integral expressions which are not particularly amenable to numerical implementation; e.g., they require Monte Carlo simulations of entire price paths of the underlying asset, $(S_t)_{t \in [0, T]}$; or numerical integration otherwise involving piecewise and/or (numerically) pathological functions; or numerical inversion of Laplace transforms; see Ballotta and Kyprianou [25]. More recently, the (still inefficient) approach of the latter paper has been superseded by Cai [53], which itself nonetheless remains suboptimal. Overall, the persistent numerical intractability of pricing (and hedging) β -quantile options is a difficulty which may account for their apparent lack of trading either through exchanges or over-the-counter markets. The numerical results presented below demonstrate that many similar (though perhaps not as difficult) problems related to numerically pricing arithmetic Asian options, are avoided via the analytical approximation derived in theorem 1.2.8, the accuracy of which is also studied.

The preceding hypothesis concerning a lack of trading of β -quantile options due to their difficult (and presently necessarily) numerical pricing, is bolstered by the fact that they serve unmet risk management needs: As discussed by Cai [53], Cai, Chen, and Wan [54], and Phelan, Marazzina and Germano [250], one motivation of β -quantile options is that they introduce a barrier-like feature [the quantiles (1.14)] determined precisely by the occupation of the price path to either side, without the ‘harsh’ feature of classical barrier options (otherwise identical to the vanilla European options discussed above), that the option loses all value as soon as (or unless, depending upon the barrier being an ‘out’ or ‘in’ type, respectively) the barrier is crossed. In particular, this introduces again the specter of (particularly, thin/illiquid) market manipulation, since a transient price jump, which is effectively ‘smoothed out’ by a β -quantile option, given that it contributes very little occupation time, can instantly ‘make or break’ the eventual option value and exercise; a potential concern for both option writers and holders.

Similarly, being based on extreme values, lookback options are at the mercy of transient price jumps; a short-lived jump to a large maximum or small minimum underlying price will irrevocably

⁸Although distributional terms may in some cases be simpler to handle than classical ones (see Zemanian [323] or Friedman [124]), this would certainly complicate, if not invalidate, the use of symmetry analytical methodology, which has been developed primarily along classical lines; see Meleshko [221].

cably alter the payoff. In the same way that power mean Asian options constitute generalized lookback options, i.e., that the particular case of $\alpha = (-)\infty$ corresponds to a lookback call (put), β -quantile options are an alternate generalization: The particular case of $\beta = 1$ (0) corresponds to a lookback call (put). Given that sample paths of the GBM (1.1) are continuous almost everywhere, it is clear that the quantiles (1.14) are continuous and nondecreasing in β ; the same follows in α for the power means (1.3) from the power mean inequality (see Bullen [52]). Ignoring the trivial case that $S_t = S_0$ almost everywhere in $[0, T]$, which itself occurs with probability zero, these monotonicity properties are strict, and it follows from the intermediate value theorem that there is a unique bijection between $\overline{\mathbb{R}}$ and $[0, 1]$, which associates to each $\alpha \in \overline{\mathbb{R}}$ the corresponding $\beta \in [0, 1]$, such that the α -power mean (1.3) equals the β -quantile (1.14). It then follows that the associated power mean Asian and β -quantile options have the same payoff, and therefore the same price. Note that this bijection is generally *specific to each price path of the GBM (1.1), $(S_t)_{t \in [0, T]}$* , meaning that such an equivalence between power mean Asian and β -quantile options is only *statistical* in nature. That is, the foregoing argument is not to say that either option can directly be used to price the other, through some price path-independent bijection, but, rather, to demonstrate and formalize the close relation between the two types of option, in order to argue that the same risk management benefits which motivate the proposal of β -quantile options, similarly motivate the proposal of power mean Asian options. And given the discussion regarding the state of the literature on β -quantile options,⁹ an investigation into PDE-based pricing of power mean Asian options may be worthwhile.

Power mean Asian options confer the reduction in payoff sensitivity to the underlying asset's terminal price achieved by a geometric Asian option, and some of the mitigation of the problem of market timing perfectly achieved by lookback options. As geometric Asian (lookback) options are therefore typically less (more) expensive than European contracts, both the power mean Asian and β -quantile options should have intermediate prices, which could be calibrated to be close to the European alternative, whilst retaining some of the benefits of both geometric Asian and lookback options.

1.1.4 Existing approaches to (arithmetic) Asian option pricing

The financial imperative and mathematical difficulty of pricing arithmetic versus geometric Asian options, principally due to the unknown distribution of the former average, has led the development of an expansive, rich, and diverse literature on the arithmetic Asian option pricing problem in particular, extending more than three decades. In the following, this literature is summarized as it pertains to the canonical Black-Scholes (BS) framework examined presently, and roughly categorized in a way that allows related methods to be described together. An extensive survey and comparison of methods for arithmetic Asian option pricing is given by Boyle and Potapchik [42], and so particular attention will be paid to subsequent developments.

⁹See Phelan, Marazzina and Germano [250] for an extensive, current overview, which demonstrates modest improvements in numerical pricing, e.g., via more efficient inversion of Fourier and other integral transforms preferable to that of Laplace.

Monte Carlo methods

One of the first major approaches to numerically pricing arithmetic Asian options was via Monte Carlo (MC) simulations, in order to estimate martingale (1.4). Boyle [43] first applied MC methods to vanilla option pricing, and Kemna and Vorst [177] in turn first applied them to arithmetic Asian option pricing, reducing the variance of such estimates by approximating the arithmetic with the geometric mean as a control variate (Zhang, Yang and Wang [324] estimate the corresponding Δ -hedging parameter in the same manner). Such (re)simulation methods however yield biased estimates of price *derivatives*, i.e., hedging parameters (also known as the option Greeks, price and risk sensitivities, and comparative statics), which are central to risk and portfolio management and optimization. Partly to address this shortcoming, Broadie and Glasserman [48] develop direct *pathwise* and *likelihood ratio* methods, which are much less computationally expensive and yield unbiased hedging parameter estimates. The likelihood ratio method refers to the Radon-Nikodym derivative which appears in Girsanov's theorem and so is a change of measure technique, which Vázquez-Abad and Dufresne [304] combine with a control variate to significantly reduce MC price estimate variance, particularly in the more difficult case when the option is significantly out of the money (i.e., $S_0 \ll K$). This is optimized at run-time via a perturbative (estimation variance-)gradient-based update rule. Glasserman, Heidelberger and Shahabuddin [147] combine importance and stratified sampling to (asymptotically) optimally reduce MC arithmetic Asian option price estimate variance. The importance sampling yields an optimal change of drift of the underlying diffusion, so that stratified sampling can be applied selectively (to high-probability, high underlying value, 'important' samples), yielding an efficient and effective procedure. Kebaier [173] achieves substantial variance reduction by implementing a statistical analogue of Romberg extrapolation. Han and Lai [155] interpret the geometric mean control variate of Kemna and Vorst [177] as a constant linear martingale control parameter, and generalize to control *processes* which yield greater variance reduction. Dinguç and Hörmann [98] reduce Asian option price estimate variance by combining a *quadratic* control variate with *conditional* MC simulation. Dinguç and Hörmann [99] improve and extend this approach to additionally include estimation, in a *single* simulation using 'pathwise derivatives,' of the Δ - and Γ -hedging parameters, by combining *randomized quasi*-MC (QMC) estimation with principal component analysis, in the latter case to achieve dimensional reduction of the (conditional) integrand in martingale (1.4). Similar improvements are realized by the 'conditional pathwise' QMC method of Zhang and Wang [325]. Zhang and Lai [328] use *multiple* control variates to construct an highly efficient MC algorithm, whereas Mehrdoust [220] complements and hybridizes such an approach with *antithetic* variates to similar effect. Rebentrost, Gupt and Bromley [257] develop a quantum computational MC pricing algorithm, which achieves quadratic reduction in (time-)complexity, compared to the standard, classical computational implementation.

Despite the many performance improvements achieved, in particular variance-reduction techniques, MC methods remain inherently inefficient, a problem which is significantly exacerbated by the need to simulate entire price *paths* of the underlying asset, $(S_t)_{t \in [0, T]}$, as opposed to only the terminal price, S_T , in the case of vanilla, European options. The result of theorem 1.2.6 below restores the need only to simulate S_T , which is however itself unnecessary, as theorem 1.2.8 explicitly evaluates the pertinent expectation with respect to S_T .

Series expansions

Another early numerical technique sought to approximate the unknown distribution of the arithmetic mean, by the known (and lognormal) distribution of the geometric mean, and then to leverage the explicit pricing formula for geometric Asian options, given early on by Kemna and Vorst [177]. Formal approximations of this kind often proceed via (generalized) Edgeworth series expansions of the unknown distribution of the arithmetic mean, about the known, *lognormal* distribution of the geometric mean, beginning with Turnbull and Wakeman [303], Levy [191], and Ritchken, Sankarasubramanian and Vijh [260], and more recently by Li and Chen [192]. This specification requires matching of two moments, which is the crux of the difference between the first, second and fourth of these papers (the second is essentially a simplified, more practical form of the first). Other moment-based methods have attempted to fit the unknown distribution of the arithmetic average by (orthogonal) Laguerre series expansions (Dufresne [102]) and (reciprocal) gamma distribution (Milevsky and Posner [222]). Fusai and Tagliani [132] compare these moment-based methods and propose their own, using the maximum entropy principle (Jaynes [166]) of selecting the least certain distribution which matches specified moments; a sort of Occam's razor requiring the weakest assumptions. Schröder [269], [271], [272], and [273] makes rigorous, extends, and modifies the original approach to pricing arithmetic Asian options by Laguerre series, due to Dufresne [102], developing two rapidly and stably convergent series expansions for integrals of theta functions, which solve the corresponding functional recursion required for its implementation.

Many, often moment-based, (asymptotic) series expansions have been proposed for pricing arithmetic Asian options. Ju [170] develops a Taylor expansion (about zero volatility, $\sigma = 0$, up to order σ^6) of the ratio of characteristic functions of the arithmetic mean and a lognormal variable with first two moments matching. The resulting pricing formula is competitive with its contemporaries in both speed and accuracy. Linetsky [199], [198] and [200] develops eigenfunction expansions for arithmetic Asian options which involve Whittaker functions, as well as an integral representation involving a single Whittaker (W) function, summed with finitely many other terms involving incomplete Gamma functions and Laguerre polynomials. Dufresne and Li [105] comment that hardly any mathematical properties have been established regarding generalized Edgeworth series, expanded about a lognormal distribution. In particular, no convergence results have been found, nor are they expected to be. Instead, Dufresne and Li [105] establish convergence of Gram-Charlier expansions of the unknown distribution of the *logarithm* of the arithmetic mean, considered earlier by Airoldi [5] and [6]. Numerical applications demonstrate that the quality of the approximation by truncating such expansions, is parameter-sensitive. A key component in Gram-Charlier series is the (classical orthogonal) family of Hermite polynomials. Aprahamian and Maddah [20] use orthogonal polynomials to extend the capability of their moment-matching approach to pricing arithmetic Asian options, from matching three to any greater number of moments of the arithmetic mean with those of compound Gamma distributions (Dubey [100]). Willems [313] develops a family of polynomials, orthogonal with respect to the lognormal density as weighting function, and constructs corresponding expansions of the arithmetic Asian option price. This presents an attractive alternative to analogous Laguerre expansions, as it does not incur the computational overhead of computing negative moments of the arithmetic mean. Edeki et al. [109] and Edeki, Akinlabi

and González-Gaxiola [108] respectively apply to the hedging and pricing of arithmetic Asian options, the Adomian [3] decomposition method. The latter is a special case of the homotopy analysis method (Liao [193]), an expansion technique not subject to the stringent validity restrictions of classical perturbation and asymptotic expansions. Edeki, Ugbebor and Ogundile [110] apply a related series expansion technique, the projected differential transform method, to price arithmetic Asian options.

Integral transforms

Much analytical research into arithmetic Asian option pricing has focused on various integral transforms, most commonly that of Laplace. Using Bessel processes and the Hartman-Watson (HW) distribution, Geman and Yor [141] develop [in their formula, (3.10)] an exact, closed-form expression for the Laplace transform (with respect to maturity, T) of the fixed-strike, arithmetic Asian call option price, given in terms of an integral of an elementary function and the Γ function. Carr and Schröder [58] subsequently correct, qualify, and extend some of these findings. Two principal drawbacks of this particular result and general approach, are that it is difficult to interpret the underlying pricing mechanism from the integral transform, and the latter typically requires numerical inversion, which can be computationally expensive. Noting, in particular, difficulties stemming from reliance on the HW distribution, Schröder [268] and [270] and Carr and Schröder [59] analytically study the Laplace inversion problem corresponding to Geman and Yor [141] and arrive at an expression involving integrals of Hermite functions. Schröder [274] and [275] refines these results and derives (asymptotic) series expansions to expedite their numerical implementation. Barrieu, Rouault and Yor [29] derive some asymptotic results for the HW distribution which compensate for some of its numerical shortcomings. Geman [139] and [138] extends in another direction the results of Geman and Yor [141] via time-changed (squared) Bessel processes, and also considers the (Δ -)hedging problem for arithmetic Asian options. Fu, Madan and Wang [128] extend the arithmetic Asian option price Laplace transform (in maturity, T) of Geman and Yor [141], to a *double* such Laplace transform, the second with respect to the strike price, K . An early comparison of MC simulations, with Richardson extrapolation or *biased* control variates chosen to counteract bias implicit in the (sample path) discretization, and numerical inversion of Laplace transforms, is also provided. Fusai [130] develops [in his Theorem 1] an exact, closed-form expression for the Laplace transform of the *Fourier transform (with respect to the logarithm of the strike price, $\log K$)* of the fixed-strike, arithmetic Asian call option price, given *purely* in terms of the Γ function, and proceeds to study both Δ and Γ hedging parameters. Nwozo and Fadugba [237] implement an accurate, computationally inexpensive numerical inversion of the double transform in order to price arithmetic Asian options.

Numerical PDE solvers

Direct analyses of pricing PDE (1.7) began with finite difference (FD) methods. Zvan, Forsyth and Vetzal [331] demonstrated that standard option pricing FD schemes result in spurious oscillations when applied to Asian options, which they are able to counter via implementation with so-called flux limiter functions in corresponding flux limiting schemes. But pricing PDE (1.7) is also *degenerate* by the absence of V_{AA} or V_{SA} terms, which hinders the applica-

tion of FD schemes. Barraquand and Pudet [28] proposed a forward shooting grid method which is able to overcome this particular difficulty. However, Forsyth, Vetzal and Zvan [121] demonstrate that this achievement is subject to a number of caveats and constraints on lattice structure. Tangman, Gopaul and Bhuruth [294] remove oscillatory arithmetic Asian option prices by combining Runge-Kutta time stepping with exponential time differences. Cen, Le and Xu [62] achieve both spatial and temporal second-order convergence, as well as spatial maximum-norm stability, by combining Rannacher time stepping with an optimally adaptive, *moving* spatial mesh. Rai and Chakrabarty [253] develop an implicit Crank-Nicolson scheme on a *nonuniform* lattice which is better able to address spurious oscillations which arise in the case of low volatility, σ . Bhatoo et al. [33] implement a central-upwind scheme which is able to efficiently price and hedge arithmetic Asian options without spurious oscillations, which is extended to achieve improved accuracy by Bhatoo et al. [34]. Kumar, Tripathi and Kadalbajoo [184] achieve similar accuracy to the former paper by incorporating radial basis functions into the finite difference scheme. Patel and Mehra [246] (and [247]) develop an unconditionally stable, compact finite difference scheme for pricing arithmetic Asian options, which is accurate to fourth order (and uses a moving mesh) spatially, and (up to) second order temporally. Wang et al. [306] incorporate operator splitting methods into a finite difference scheme, and demonstrate its ability to match MC simulations.

Alternatively, a more flexible (e.g., nonuniform mesh) and general finite element approach to pricing arithmetic Asian options is given by Zvan, Forsyth and Vetzal [330]. Fofas and Larson [122] develop a fast and stable adaptive method which adjusts mesh size locally according to price and hedging parameters of interest. Marcozzi [214] implements an accurate discontinuous Galerkin method, which combines finite elements with finite volume discretization, also introduced for the pricing of arithmetic Asian options by Zvan, Forsyth and Vetzal [332]. The finite volume approach is more naturally suited to the degeneracy of pricing PDE (1.7), and is also able to counter the spurious oscillations which result. Ramírez-Espinoza and Ehrhardt [255] apply rapidly convergent finite volume schemes to arithmetic Asian option pricing, as do Chernogorova and Vulkov [72] after first simplifying and approximating the pricing problem by splitting in two different ways the operator of pricing PDE (1.7) into two recursive sub-problems.

Both operating on lattices, there is a close relationship between finite difference and multinomial approaches to arithmetic Asian option pricing.¹⁰ The essential difficulty in the latter case is that the number of possible asset paths grows exponentially with the number of time steps, hindering the efficiency of standard implementations. Hull and White [164] provided an early improvement (subsequently further improved and effected by Costabile, Massabó and Russo [84], the ‘adaptive placement’ method of Dai, Wang and Wei [90] and [91], and the modified Edgeworth binomial model of Wang and Hsu [307]), more efficient than standard MC pricing, by computing path averages of select ‘representative’ nodes and interpolating as required for the remainder. Klassen [179] developed an unconditionally stable, rapid implementation of this method using Richardson extrapolation. Neave [231] (and Neave and Ye [232] further) reduced computational costs (and increased accuracy) using generating function

¹⁰Indeed, finite differences are lattice derivatives.

and ‘path bundling’ techniques. Chalasani, Jha and Saias [65] achieved in polynomial time the observance of certain error bounds with corresponding high probability, prior to the guaranteed observance secured by Aingworth, Motwani and Oldham [4]. Both accuracy and efficiency of the latter algorithm were improved in the randomized implementations of Akcoglu, Kao and Raghavan [8], Ohta et al. [238] and Shioura and Tokuyama [278]. Dai, Huang and Lyuu [87] and Dai and Lyuu [88] maintain efficiency (i.e., sub-exponential-time) in a multiresolution trinomial lattice, which necessarily converges to the continuous-time martingale (1.4) as the number of time steps diverges. This result was improved to quadratic-time by Hsu and Lyuu [163], and certain caveats and qualifications were removed by Dai and Lyuu [89]. Using ‘state conditional expected value’ *forward* induction, Abrahamson [1] achieves accuracy competitive with contemporary MC and numerical PDE solvers, but via a simpler and much faster implementation. Jiang and Dai [168] establish the equivalence (to first order in the time step, Δt) of certain binomial tree and explicit finite difference methods for pricing arithmetic Asian options, and using viscosity solutions, the uniform convergence of such prices to martingale (1.4) as $\Delta t \rightarrow 0$. The ‘singular points [binomial] method’ of Gaudenzi, Zanette and Lepellere [134] achieves very tight bounds which vanish as $\Delta t \rightarrow 0$, and for which estimates are established. Significant efficiency and accuracy improvements also follow from the cell averaging technique of Moon, Jeong and Kim [224], which reduces computational burden by replacing discrete nodes with continuous ‘cells’ (intervals) and computes (arithmetic) averages over the latter, in contrast to the selection of and interpolation from representative nodes/paths, adopted in most other multinomial arithmetic Asian option pricing approaches.

Path integrals

An approach to arithmetic Asian option pricing, which partially unifies all of the discussed, more established methods, is via (functional) *path integration*, described in a mathematical financial context by Linetsky [197]. This is natural, as the expectation computed in martingale (1.4), is indeed a functional integral over all Brownian paths (1.1) satisfying an initial condition, $S(t) = S_t$. Matacz [218] approximates the (arithmetic) mean (1.3) by a corresponding *partial average*, which significantly reduces numerical computational demands, by well-approximating the functional by a *single*, i.e., one- instead of uncountable-dimensional, integral. Lim [195] and [196] develops a ‘recursive integration’ method for pricing arithmetic Asian options, which backward-solves the pricing PDE (1.7), in effect by discretely approximating the path integral by a finite sequence of one-dimensional integrals. Baaquie, Coriano and Srikant [22] express pricing PDE (1.7) as a Schrödinger equation, and apply quantum mechanical methods to the corresponding Hamiltonian and potential, in order to numerically price arithmetic Asian options via both (Langevin) lattice simulations of the Brownian paths (1.1), as well as corresponding simulations via the Metropolis(-Hastings), Markov chain MC algorithm. Bormetti et al. [41] develop a competitive and accurate MC algorithm from the path integral formulation of martingale (1.4), which they term a ‘path integral with external integration.’ A second, ‘pure’ MC algorithm (without ‘external integration’) is similarly constructed. Zhang [326] manipulates the Dirac delta functional, to express the path integral formulation of martingale (1.4) as an action functional with exponential potential, which is evaluated using the Feynman-Kac formula and reduced to a spectral expansion for the arithmetic Asian option price, after further manipulation of Bessel and Whittaker functions. This derivation il-

illustrates the (Laplace) integral transform (inversion), as well as eigenfunction decomposition methods, examples respectively of the integral transforms and series expansions already discussed. Giles [144] (and [146] and Giles, Debrabant and Rössler [145]) introduces *multilevel* MC arithmetic Asian option pricing (with improved convergence), which utilizes the (higher-order) *stochastic* Euler–Maruyama (Milstein) finite difference scheme, to generate Brownian paths (1.1). Gerstner and Noll [143] implement a randomized QMC analogue and explore both schemes, resulting in some cases in significantly improved convergence rates.

Functional (Feynman path) integration is an ongoing area of mathematical research, and a common, rigorous definition of such integrals has yet to be defined; see, e.g., DeWitt-Morette and Folacci [96], in particular Cartier and DeWitt-Morette [61]. In any case, path integrals typically require numerical approximation.

Price bounds

Given the mathematical difficulties associated with all other arithmetic Asian option pricing methods discussed, various attempts have also been made to derive ever-sharper upper and lower bounds on such prices. Iwaki, Kijima and Yoshida [165] provided an early benchmark for subsequent refinements, including the improved lower (upper) bound of Rogers and Shi [262] (Thompson [299]). The former, restricted to unit maturities, $T = 1$ year, is improved on and economized by Chalasani, Jha and Varikooty [66], and generalized to arbitrary maturities by Chen and Lyuu [69]. Lasserre, Prieto-Rumeau and Zervos [186] obtain monotone sequences of upper and lower arithmetic Asian option price bounds, and establish related convergence results, corresponding to sequences of increasingly high-dimensional semidefinite programming problems. Furthermore, good bounds are found to follow from even low-dimensional programs. Novel, tighter upper bounds were subsequently derived by Lord [211] and Novikov and Kordzakhia [236] and [235]. Recently, Cibelli, Polidoro and Rossi [81] have derived yet sharper lower and upper bounds using techniques and results of optimal control.

Low volatility, σ , and/or short maturity, T

The approximation achieved in theorems 1.2.6 and 1.2.8 improves (and converges to the¹¹ exact price if $r = \sigma^2$), as the *scale* (i.e., non-dimensional variance of S_T), $\sigma^2 T \rightarrow 0$. Results related to this asymptotic circumstance have been found across several of the solution methodologies for pricing arithmetic Asian options. In many cases, such results indicate that existing approaches behave poorly in this circumstance: As noted recently by Cibelli, Polidoro, and Rossi [81], the difficulties posed without qualification by numerically inverting the Laplace transform of Geman and Yor [141], are substantially exacerbated in cases with low volatility and/or short maturity. Horvath and Medvedev [162] document significant computational improvements in numerical approximation of (functional) path integrals over the last two decades, but find that it still lags MC simulations when volatility or option lifetime are sufficiently small. Ewald, Menkens and Ting [113] acknowledge potential numerical instabilities when solving their reduced pricing PDE with small volatility values; the same claim is made regarding that

¹¹See, e.g., Wilmott, Dewynne, and Howison [315] and Shreve [279], for a discussion of the uniqueness of ('classical'/everywhere smooth) solutions to the full pricing problem.

of Rogers and Shi [262]. More broadly, Dufresne [103] and [104], Boyle and Potapchik [42], and Tangman et al. [295] note that the performance of many contemporary arithmetic Asian option pricing methods suffers in the case of low volatility and/or short maturity.

1.1.5 Broad methodology and results

The theoretical results presented below are principally derived using the Lie point symmetries of pricing PDE (1.7). Such *symmetry analysis* of differential equations was introduced by Lie [194], as a systematic, algorithmic method for manipulating such equations, which can be harnessed to reduce their order, derive novel solutions from existing ones, and construct novel solutions, *invariant* under corresponding *Lie point symmetries*, which themselves leave the equations of interest invariant under induced changes of variable. From one viewpoint, symmetry analysis is a systematic, generalized method of constructing similarity solutions of differential equations, which map to one another under such equations' point symmetries (and to themselves in the case of invariant solutions), and may yield equations of lower order when substituted into the originals. After providing a relatively general introduction to the basic tools and terminology of Lie point symmetry analysis, it and its generalizations are discussed in the broader context of analytical solutions of PDEs, and then a review is given of its existing applications to arithmetic Asian option pricing and related problems in mathematical finance.

A primer on Lie point symmetry analysis

Take an evolutionary system of PDEs of second order, in the unknown $y(x, t) : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^m$:

$$y_t = \Phi(t, x, y, \nabla y, \nabla \otimes \nabla y) \quad (1.15)$$

Here, $\nabla \otimes \nabla y$ denotes the (spatial) Hessian ($y_{x^i x^j}$) of y , via the tensor product, \otimes . Determination of the *Lie point symmetry group* of system (1.15) is equivalent to determining its (first-order) *infinitesimal symmetries*, namely, η and ξ^i for integers $0 \leq i \leq n$ in the following system:

$$\begin{aligned} \tilde{t} &= t + \epsilon \xi^0(t, x, y) + o(\epsilon) \\ \tilde{x}^i &= x^i + \epsilon \xi^i(t, x, y) + o(\epsilon) \\ \tilde{y} &= y + \epsilon \eta(t, x, y) + o(\epsilon) \end{aligned} \quad (1.16)$$

Infinitesimal symmetries for the derivatives y_t , ∇y , and $\nabla \otimes \nabla y$ must be computed, so that system (1.15) is invariant under (an augmented) system (1.16): For all integers $1 \leq i, j \leq n$,

$$\tilde{y}_{\tilde{t}} = y_t + \epsilon \zeta^0(t, x, y, y_t, \nabla y) + o(\epsilon) \quad (1.17)$$

$$\tilde{y}_{\tilde{x}^i} = y_{x^i} + \epsilon \zeta^i(t, x, y, y_t, \nabla y) + o(\epsilon) \quad (1.18)$$

$$\tilde{y}_{\tilde{x}^i \tilde{x}^j} = y_{x^i x^j} + \epsilon \zeta^{ij}(t, x, y, y_t, \nabla y, \nabla \otimes \nabla y) + o(\epsilon) \quad (1.19)$$

Systems (1.16) and (1.17-1.19) together define an extended symmetry group, with the infinitesimal symmetries of the latter being given by the so-called *prolongation formulas*:

$$\begin{aligned} \zeta^0 &= D_t \eta - y_t D_t \xi^0 - \nabla y \cdot D_t \xi \\ \zeta^i &= D_{x^i} \eta - y_t D_{x^i} \xi^0 - \nabla y \cdot D_{x^i} \xi \\ \zeta^{ij} &= D_{x^i} \zeta^j - y_{tx^j} D_{x^i} \xi^0 - \nabla y_{x^j} \cdot D_{x^i} \xi \end{aligned} \quad (1.20)$$

Here, ξ denotes the (spatial) vector (ξ^i) , and D total differentiation with respect to the subscript. It then obtains that solutions of the following, so-called *determining equation*, define all infinitesimal symmetries of system (1.15), i.e., all systems (1.16) under which it is invariant:¹²

$$\zeta^0 = \xi^0 \Phi_t + \xi \cdot \Phi_x + \eta \cdot \Phi_y + \zeta \cdot \Phi_{\nabla y} + Z : \Phi_{\nabla \otimes \nabla y} \quad (1.21)$$

Here, Z denotes the matrix (ζ^{ij}) and $:$ the corresponding ‘dot’ product, $1_n^T (Z \odot \Phi_{\nabla \otimes \nabla y}) 1_n$, where \odot is the (pointwise) Hadamard product and 1_n a column vector of n ones. Equation (1.21) can be heuristically seen to express the invariance of system (1.15) under systems (1.16) and (1.17-1.19): Relative to the parameter $\epsilon \downarrow 0$, ζ^0 expresses the deviation of y_t , the left-hand side of system (1.15). Conversely, each left multiplicand in each summand of the right-hand side of determining equation (1.21) expresses the deviation of the term with respect to which Φ is differentiated in the corresponding right multiplicand. Thus, the right-hand side of equation (1.21) can be interpreted as a total deviation, relative to $\epsilon \downarrow 0$, of Φ , the right-hand side of equation (1.15). Therefore, equation (1.21) expresses that, relative to $\epsilon \downarrow 0$, both sides of equation (1.15) deviate by the same amount, i.e., this equation is preserved/invariant, as claimed.

Further details and related literature

The rigorous development of the theory of Lie point symmetries alone is quite onerous, with one authoritative reference being due to Olver [239]. The space spanned by the variables $(t, x, y, y_t, \nabla y, \nabla \otimes \nabla y)$, *interpreted as being independent*, is known as a (prolonged) *jet space*. In this space, the system (1.15) *algebraically* specifies a manifold, which the point symmetries (1.16), as well as the (first-order) *contact* symmetries (1.17-1.18), and (higher-order) *Lie-Bäcklund* symmetries (1.19), map into itself, i.e., map *symmetry solutions* constituting points on this manifold into one another. *Invariant* solutions are fixed points of such symmetries. Beyond these three kinds of symmetries, several ‘non-classical’ and approximate symmetries of differential equations, and *non-local* symmetries applicable to integral equations, have also been proposed. Such symmetries and their analogues for more general functional equations continue to be actively researched. Primarily considered below are Lie point symmetries (1.16); a preliminary analysis of non-classical and Lie-Bäcklund symmetries yielded no apparent symmetries of pricing PDE (1.5), beyond those of lemma 1.2.1, which might lead to invariant solutions in addition to those of lemma 1.2.3. Many kinds of symmetries are discussed by Meleshko [221], in the broader context of analytical solutions of differential equations. Also discussed are two major alternatives to symmetry analysis: One, the *degenerate hodograph method*, generalizes the hodograph method, which attempts to solve a system of differential equations, with equally many (in)dependent variables, by swapping the roles of such variables. In the degenerate case, the requirement of there being *equally* many (in)dependent variables is dropped, but in the case of pricing PDE (1.5), with only one dependent variable, V , and three independent variables, S, A, t , the method appears unproductive. The second approach to constructing analytical solutions of PDEs discussed by Meleshko [221], is the *differential constraints method*: It complements system (1.15) [or, e.g., pricing PDE (1.5)] with an ancillary

¹²Here, *invariant* means that system (1.15) is satisfied exactly when $\tilde{y}_t = \Phi(\tilde{t}, \tilde{x}, \tilde{y}, \tilde{\nabla} \tilde{y}, \tilde{\nabla} \otimes \tilde{\nabla} \tilde{y})$. That is, substituting for the original variables $(t, x, y, y_t, \nabla y, \nabla \otimes \nabla y)$ their transformed analogues, given by systems (1.16) and (1.17-1.19), the original functional form specified by Φ continues to hold. This function is assumed sufficiently differentiable for the determining equation (1.21) to be interpreted classically.

system of *differential constraints*. Solving the systems so combined, it is possible to generate solutions of the original system which satisfy the supplemental constraints. The primary question is which constraints to impose: This is often determined by the physical or other nature of a problem, and in fact, the *Lie equations* (1.16) are one popular such choice, so that Lie point symmetry analysis is a(n important) special case of the method of differential constraints. In the case of pricing PDE (1.5), an alternative system of differential constraints is not immediately apparent from a mathematical or financial perspective, but may present one promising direction of future research.

Gazizov and Ibragimov [135] apparently first applied Lie point symmetry analysis in mathematical finance, in part to determine the (six-dimensional/parameter¹³) Lie point symmetry group of the Black-Scholes [38] equation. This provides a systematic, alternative derivation of the results of the latter, which elucidates why the transformation approach originally taken was available and successful. Invariant solutions are also constructed, and similar analysis is applied to a two-factor model. Similar, generalized studies have since been undertaken by Silberberg [280], Leach and Andriopoulos [187], Yang, Zhang and Qu [316], Singh and Prabakaran [281], Sophocleous and Leach [287], Polat [252], Liu and Wang [202], Cimpoiasu and Constantinescu [82], Masebe and Manale [217], Tamizhmani, Krishnakumar and Leach [293], Hashemi [156], Sinkala [284], Davison and Mamba [94], and Chong and O'Hara [77]. Several Lie point symmetry analyses have also been applied to bond and other interest rate derivative pricing problems, including those of Goard [148], Sinkala, Leach and O'Hara [285], [283] and [282], Dimas et al. [97], Motsepa, Khalique and Molati [226], Charalambous and Leach [67], Antoniou [16], Bakkaloglu, Aziz and Mahomed [23], Aziz, Fatima and Khalique [21], Khalique and Motsepa [178], and Kaibe and O'Hara [171]. Additional Lie point symmetry analyses in mathematical finance have concerned commodity pricing (Leach and Andriopoulos [188], Sophocleous, Leach and Andriopoulos [288], Paliathanasis, Morris and Leach [243], and Leach, Morris and Paliathanasis [189]); pricing options with barrier features, i.e., which pay off contingent on certain underlying price barrier(s) (not) being crossed (O'Hara, Sophocleous and Leach [241] and Bozhkov and Dimas [44]); option pricing in the case that the volatility, σ , is specified as a stochastic process (Sophocleous, O'Hara and Leach [289], Paliathanasis et al. [242], and Antoniou [17]); and portfolio optimization and selection (Naicker, Andriopoulos and Leach [227], Leach, O'Hara and Sinkala [190], and Antoniou [18]). Hernández et al. [159] discuss some of these, as well as other Lie point symmetry analyses in mathematical finance and economics.

To date, there have only been a handful of Lie point symmetry analyses of arithmetic Asian option pricing problems. Apparently the first was due to Taylor and Glasgow [296], who provide an expedient derivation of the full, six-dimensional/parameter Lie point symmetry group of pricing PDE (1.7), in the arithmetic case with $\alpha = 1$. They further construct the general symmetry solution, invariant under all but one of the six symmetries (in particular, that permitting arbitrary translation in time, t), and use the resulting terminal value, at maturity $t = T$, to define a payoff and propose a corresponding 'put-type' Asian option to optimally approximate

¹³The parameter, ϵ , in the Lie equations (1.16) defining the point symmetries, may vary independently across alternate symmetries (i.e., solutions to the Lie equations).

a true (fixed-strike) put payoff, arguing that (asymptotic) smoothness of the solution and its comparative statics (namely, strictly negative Δ - and positive Γ -hedging parameters) are desirable analytical properties which favor the consideration of such contracts. (The restriction $rT < 1$ is also imposed, in order that the solutions remain nonnegative.) Caister, Govinder and O'Hara [55] and [56] undertake an exhaustive Lie point symmetry group classification of pricing PDE (1.7), in the arithmetic case with $\alpha = 1$: After also finding the corresponding six-dimensional/parameter Lie point symmetry group, they consider the discrete, five-dimensional subalgebra spanned by the Lie point symmetries, discarding that corresponding to the fact that pricing PDE (1.7) is linear, i.e., the so-called 'solution symmetry,' $\phi\partial_y$, for $\phi(x, t)$ an arbitrary solution of system (1.15). Linearly transforming this subalgebra, they optimally decompose it such that a minimal number of dissimilar [through conjugation by elements, i.e. symmetries, of the subalgebra] symmetries, as it happens seven, are uniquely similar to each symmetry of the subalgebra, i.e., partition it into conjugacy classes. Reductions of pricing PDE (1.7) can then be analyzed with respect to each of these (seven) symmetries, to separately consider all possible invariant solutions of pricing PDE (1.7) in a systematic manner. In six of these cases, some further symmetry reductions are performed for the corresponding reductions of pricing PDE (1.7), and resulting particular solutions, expressed in terms of various [mostly (confluent) hypergeometric and Bessel] special functions, are reported. For each case, the feasibility of respecting the terminal BC, i.e., payoff (1.8), is then considered, as well as the reductions' implications for applying finite difference and Laplace transform methods. Rather than relying on a CAS, Antoniou [15] provides a manual derivation of the six-dimensional/parameter Lie point symmetry group of pricing PDE (1.7), in the arithmetic case with $\alpha = 1$. An invariant solution corresponding to one symmetry is also partially developed, including fulfillment of the payoff (1.8). The present analysis principally expands upon the latter paper, generalizing and correcting its results in the arithmetic case with $\alpha = 1$, in addition to proposing use of the general power mean, for arbitrary $\alpha \in \overline{\mathbb{R}}$.

Separately from Lie point symmetry group analysis, another Lie algebraic approach to solving evolution PDEs is predicated upon the theorem of Wei and Norman [309], which permits the decomposition of solutions of linear, first-order ordinary operator differential equations, via an analogue of the scalar multiplication law for exponential functions; ($\forall a, b \in \mathbb{C}$) $e^{a+b} = e^a e^b$. By rewriting the original PDE in terms of an operator, $U(t)$, the decomposition holds so long as the coefficient in the operator equation, $U'(t) = A(t)U(t)$, is generated by a solvable Lie algebra of operators. This method was first applied to mathematical finance in the context of bond and other interest-rate derivative pricing problems by Lo, Yuen and Hui [207] and [203], Lo and Hui [205], and Lo [204], and the pricing of barrier options by Lo and Hui [206]. Parmar [244] has applied the technique to arithmetic Asian option pricing, and claims to have derived corresponding prices [satisfying the appropriate pricing PDE (1.7) and terminal payoff BC (1.8)] and risk sensitivities/hedging parameters, for a variety of arithmetic Asian options, differing according to the subset of $[0, T]$ over which the arithmetic average is computed. However, the gravitas of this claim, combined with the failure of a careful attempt via a CAS to verify Proposition 2 of Parmar [244], i.e., satisfaction by the stated solution, $C(S_t, I_t, t)$, of the corresponding pricing FPK PDE stated in Proposition 1, even in case of time-independent parameters r and σ , casts doubt on at least the tested result. (The subsequent results have not been tested, but as they are generally more complicated than, predicated upon, or 'proved' in

the same manner as Proposition 2, their correctness may also be doubtful. Close scrutiny revealed no major missteps in the proof provided for Proposition 2, *assuming the perhaps not so ‘trivial’ results of Proposition 14 (established in appendix A), upon which they are contingent, to be valid and correctly applied.* One apparent issue is the application of Taylor series expansions of solutions of pricing PDE 1.7, not knowing in advance the radii of convergence of such expansions, if they exist at all. This line of inquiry, with the hope of verifying even some form of Proposition 2, could be a worthwhile future endeavor. Similar efforts could be made toward the invariant solutions determined in theorem 1.2.9 or the more general ansätze determined via invariants (1.27)-(1.29) and (1.48)-(1.50), satisfying respectively the reduced pricing PDEs (1.30) and (1.51), significantly generalizing the rudimentary type I and II similarity reductions considered by Parmar [244].)

A primer on Gaussian process regression

The numerical results presented below are principally derived using Gaussian process regression (GPR) to construct predictors of a certain function, denoted $\lambda(S, A, t; \sigma, r, T, K)$, which allows pricing of *arithmetic* Asian options via theorem 1.2.8. Essentially, this function (predictor) is indicating which among a family of invariant solutions derived in theorem 1.2.5 to choose, given values of the variables, S , A , and t , and parameters, σ , r , T , and K . This becomes necessary as formula (1.43) does *not* satisfy pricing PDE (1.7) with $\alpha = 1$, in contrast to that of theorem 1.2.6, the expectation of which with respect to S_T (which depends on S and t) yields the result of theorem 1.2.8. An introduction to GPR is provided, which was chosen primarily because the resulting function predictors are *smooth*, yielding smooth prices via formula (1.43), which may sensibly be substituted into pricing PDE (1.7) with $\alpha = 1$, to examine to what extent the function predictors, say $\tilde{\lambda} \equiv \tilde{\lambda}(S, A, t; \sigma, r, T, K)$, correct the ‘bias’ introduced in computing formula (1.43) as an expectation with respect to S_T of formula (1.40), by yielding prices, say $V(S, A, t; \sigma, r, T, K, \tilde{\lambda})$, which (better) satisfy pricing PDE (1.7) with $\alpha = 1$. (This task itself is a large one which may best be left for a subsequent paper.) A number of alternate (machine learning) approaches to function prediction are examined in section 2.2, which are either linear or (generally) nonsmooth: Linearity (in S , A , and t) is a strong ansatz to impose on candidate forms for $\tilde{\lambda}$,¹⁴ and as noted, only (almost everywhere, at least) smooth predictors are sensible for substitution into formula (1.43) and pricing PDE (1.7) with $\alpha = 1$.

One canonical reference for GPR is the book by Williams and Rasmussen [314]. GPR is a nonparametric method of function estimation, which assigns to each point in some variable/parameter set, e.g., tuples $\mathbf{x} \equiv (S, A, t; \sigma, r, T, K) \in [0, \infty)^7$ such that $t \leq T$ (and assuming nonnegative interest rates, r), some distribution, $F_{\mathbf{x}}$, such that *any* finite collection of points, say $\mathbf{X} \equiv \{\mathbf{x}_i | 1 \leq i \leq p\}$, is *jointly* normally distributed:

$$\{F_{\mathbf{x}} | \mathbf{x} \in \mathbf{X}\} \sim N_p(\mu_{\mathbf{X}}, \Sigma_{\mathbf{X}})$$

¹⁴Investigations similar to those reported in section 1.2.4 indicated that $\tilde{\lambda}$ *may* be linear in S , but a pertinent ansatz, $\tilde{\lambda} \equiv S f(A, t) + g(A, t)$, substituted into formula (1.43) and pricing PDE (1.7) with $\alpha = 1$, yielded an apparently still unwieldy system of PDEs for $\tilde{\lambda}$; perhaps a topic for a subsequent paper.

This particularly implies that the distribution of any single point is normal, say, $\mathbf{x}_i \sim N(\mu_i, \sigma_{ii})$. By computing the $p(p-1)/2$ pairwise covariance values, say σ_{ij} , the covariance matrix obtains:

$$\Sigma_{\mathbf{X}} \equiv (\sigma_{ij}) \equiv (\kappa(\mathbf{x}_i, \mathbf{x}_j)) \equiv K_{\mathbf{X}}$$

Here, denoting by \mathcal{X} the pertinent variable/parameter set, $\kappa : \mathcal{X}^2 \rightarrow \mathbb{R}$ is a positive (semi)definite *kernel* mapping points to their (co)variance values; there are several popular kernel functions, some of which are discussed in the numerical implementation below, and the choice of kernel is one *hyperparameter* to be optimized in the fitting process. (Another secondary reason for choosing GPR for the numerical applications of the theoretical results, is that hyperparameter optimization is relatively straightforward for GPR; indeed, Gaussian processes are often utilized in coordination with Bayesian optimization for optimizing hyperparameters of *other* machine learning models. This is partly because the normal distributions involved in GPR permit explicit computation of confidence intervals and other salient quantities, without additional Monte Carlo simulations, for example.) The training of GPR models for fixed hyperparameters (others also being discussed in the applications below) proceeds via iterated Bayesian inference: Typically beginning with a uniform prior on the mean, $\mathbf{m}_{\mathbf{X}} \equiv \mu_{\mathbf{X}}$ (which is often demeaned to vanish identically), and covariance, $\mathbf{K}_{\mathbf{X}}$, functions, the likelihood of observed *training* points is assessed with respect to the prior, and their product normalized to yield the posterior via Bayes's theorem. This can be iterated, regarding the old posterior as the new prior, up to a specified limited number of times or until some acceptable convergence of the sequence of posteriors is satisfied. Once this is the case, the final posterior yields predictions on unobserved *testing* points via the posterior *predictive* (normal) distribution, as well as confidence intervals.¹⁵

The mathematical derivation and description of this procedure is standard and in terms of basic facts of linear algebra and multivariate normal distributions; one good reference is again Williams and Rasmussen [314]. But without addressing nuances concerning different kernels and other hyperparameters, or exact algorithmic implementation details, it is crucial to note that the GPR is completely specified by $\mathbf{m}_{\mathbf{X}}$ (again, often taken to vanish) and $\mathbf{K}_{\mathbf{X}}$: Given training points \mathbf{X} with $|\mathbf{X}| = p$ as above, the process of finding an optimal posterior results in the estimation of $p(p+3)/2$ parameters [or $p(p+1)/2$ with $\mathbf{m}_{\mathbf{X}} \equiv \mathbf{0}$]. Testing or otherwise predicting r unobserved points, up to $r(r+2p+3)/2$ new parameter estimates are required to specify the posterior predictive (joint normal) distribution of the (un)observed points. So in practice, the nonparametric GPR method is finite-dimensional, as are the (un)observed data, despite the method often being marketed as one of *functional* estimation: The fact that any finite set of points (always the case in practice) are individually normally distributed, reduces the 'functional' aspect to specification of finitely many parameters, similarly as a one-dimensional normal distribution function is completely specified by two parameters. But generally speaking, the method is functional both in the sense that the sample space of data points is typically uncountable, and that even countably infinitely many points are no longer sensibly multivariate-normally distributed, and specification by even countably infinitely many parameters fails. Finally, note that GPR is fundamentally motivated (and successful) by the well-known and intuitive fact that

¹⁵Confidence intervals of GPR models diverge outside of the training data, making them potentially unsuitable for forecasting/extrapolation purposes, including that for $\tilde{\lambda}$ in section 1.2.4.

some multivariate normal distribution exists to interpolate *any* finite set of points. (Practically, some implementations do not interpolate for sufficiently large numbers of data points, or when the latter are understood to include noise.)

1.1.6 List of contributions

Listed are the major contributions which follow and provide a guide for the succeeding section:

1. Lemma 1.2.1 gives the Lie point symmetries of pricing PDE (1.5), generalizing results of Taylor and Glasgow [296], Caister, Govinder and O'Hara [55] and [56] and Antoniou [15].
2. Theorem 1.2.2 gives an optimal system of Lie point symmetries of pricing PDE (1.7), generalizing results of Caister, Govinder and O'Hara [55] and [56].
3. Lemma 1.2.3 [theorem 1.2.5] determines *all* invariant solutions of pricing PDE (1.7), corresponding to the Lie point symmetries of contribution 1., which reduce this PDE to that of Black and Scholes [38] [the heat equation¹⁶ (1.35)], generalizing Antoniou [15].
4. Corollary 1.2.4 provides a parity relation for pricing power mean Asian options with positive (negative) index $\alpha \in \overline{\mathbb{R}}$ in terms of those with negative (positive) index, using the invariant solutions of contribution 3.
5. Theorem 1.2.6 develops the invariant solutions of contribution 3. in order to satisfy the terminal payoff BC (1.8), applying Poisson's formula (1.38) to generally solve the corresponding Cauchy problem of the heat equation (1.35), *in the arithmetic case with $\alpha = 1$ and the terminal underlying asset price, S_T , as a parameter*. These solutions approximately satisfy BC (1.9) with $S = 0$, and do so exactly as the scale $\sigma^2(T - t) \rightarrow 0$ and $|r - \sigma^2| \rightarrow 0$.
6. Corollary 1.2.7 provides several price sensitivities/hedging parameters/comparative statics/'Greeks' for the invariant solutions of contribution 5., in part as proof of the latter.
7. Theorem 1.2.8 computes the expectation with respect to the terminal underlying asset price *parameter*, S_T , of the invariant solution of contribution 5. Given that S_T has a (geometric Brownian motion) distribution dependent on the variables S and t , the resulting expression no longer satisfies pricing PDE (1.7), but does suggest the problem of determining the parameter, $\lambda(S, A, t; \sigma, r, T, K)$, such that the composition of the pertinent (S_T -mean) invariant solution (1.43) with the latter *does* satisfy pricing PDE (1.7).
8. Theorem 1.2.9 determines *all* invariant solutions of pricing PDE (1.7), corresponding to the *optimal* Lie point symmetries of contribution 2., which reduce this PDE to that of Black and Scholes [38] and thence the heat equation (1.35). The additional parametric flexibility afforded by the optimal symmetries is seen only to be apparent; the resulting invariant solutions are equivalent to those of contribution 3.

¹⁶A canonical PDE of mathematical physics, the heat equation is much studied, non-degenerate, and more amenable to finite difference methods.

9. Theorem 1.2.10 provides *all* (multiplicative) separation of variables invariant solutions of PDEs (1.30) and (1.51), corresponding respectively to the Lie point symmetries of contributions 1. and 2. These are the *most general* PDEs reduced from the pricing PDE (1.7) using the corresponding symmetries and, in particular, are more so than the respective BS PDEs (1.34) and (1.56).
10. Corollary 1.2.11 develops invariant solutions of pricing PDE (1.7) from contribution 9.
11. Theorems 1.2.12-1.2.14 respectively develop Fourier-Bessel series, Hankel and Kontorovich-Lebedev transform solutions of pricing PDE (1.7) satisfying BC (1.12) in case $A = TK^\alpha$.
12. Corollary 1.2.15 gives parametric properties of the Kontorovich-Lebedev transform solutions of contribution 11., on divergence of the exponential function in formula (1.109).
13. Section 1.2.2 proposes problems 1 and 2, simplified from that noted in contribution 7.
14. Section 1.2.3 undertakes exploratory analysis of the behavior of the parameter, $\lambda(S_0, 0, 0; \sigma, r, T, K)$; proposes a further simplified problem 3; and addresses the latter in two separate (smooth) GPRs with training and testing points drawn in varying proportions from a large design of experiments (1.111), the second (first) (not) imposing a constraint $r \approx \sigma^2$ related to the approximate nature of the invariant solutions of contribution 5. These are compared with (apparently the first) direct GPR predictions of prices satisfying PDE (1.7), based on testing and training points simulated by MC à la Longstaff and Schwartz [210].
15. Section 1.2.4 presents much-compacted results similar to those of contribution 14. in case $t > 0$, which do not require GPR predictions of a design of experiments, relying rather on curve or surface estimation of the parameter, $\lambda(S, A, t; \sigma, r, T, K)$, exploiting formula (1.45), valid on the *region* $A \geq TK$ and encompassing BC (1.44) in case of equality. Results shown in figure 1.34b vastly out- (under)perform the MC simulations noted in contribution 14. for near- (deeply-out-of-)the-money arithmetic Asian option pricing, i.e., when $0 \ll A \lesssim TK$ ($0 \lesssim A \ll TK$), yielding complementary contributions.

1.2 Methods

This section develops the majority of results obtained, along with further contextualization in relation to pertinent literature. First, the theoretical contributions are presented. These are then applied to the numerical experiments reported afterwards.

1.2.1 Lie point symmetry analysis and consequent results

This subsection begins with the derivation of (optimal) Lie point symmetries of pricing PDE (1.7), followed by a (naive) derivation of a family of invariant solutions whose reduced (BS) pricing PDE can be further reduced to the heat equation by standard transformations, and finally the implementation of the terminal (payoff) BC (1.8). The latter stage yields a family of invariant solutions satisfying pricing PDE (1.7) and BC (1.8) exactly, and BC (1.9) when $S = 0$ approximately, with the approximation becoming exact under certain parametric restrictions.

These solutions depend on the terminal underlying asset price, S_T , as a parameter; analytically computing the expectation of the former with respect to the latter yields a family which can be used for obtaining numerical prices in the following section.

(Optimal) Lie point symmetries

As may be appreciated from the primer in subsection 1.1.5, Lie point symmetry analysis of differential equations is very computationally intensive; so much so that, with the *linear* system (1.21) typically involving tens of coupled equations, ‘hand’ computations are practically infeasible. As such, many mathematical software packages introduced routines for finding symmetries of differential equations, as early as the eighties. Papers have been devoted solely to surveying the available software, and one popular program is `Maple`. The `Maple` symmetry package SADE, written by Rocha Filho and Figueiredo [261], was used to compute the following Lie point symmetries of equation (1.5), as the built-in tools of `Maple` for Lie point symmetry analysis often omit point symmetries (e.g., some of those found by Antoniou [19]):

Lemma 1.2.1 *The Lie point symmetries of equation (1.5) are as follows:*

$$\begin{aligned}
X_1 &\equiv \xi_1^0 \partial_t + \xi_1^A \partial_A \\
X_2 &\equiv \xi_2^A \partial_A \\
X_3 &\equiv \xi_3^A \partial_A + \xi_3^S \partial_S + \eta_3 V \partial_V \\
X_4 &\equiv \xi_4^A \partial_A + S \partial_S \\
X_5 &\equiv \xi_5^A \partial_A + \xi_5^S \partial_S + \eta_5 V \partial_V
\end{aligned} \tag{1.22}$$

For any solution ϕ of equation (1.5), the operators $X_\phi \equiv \phi(S, A, t) \partial_V$ generate a subalgebra.

Here the infinitesimal symmetries ξ_1^0 , ξ_1^A , ξ_i^S and $(\forall 1 \leq i \leq 5) \eta_i$, depend on the form of A :

A	\mathcal{A}	\mathcal{A}^α	$t\mathcal{A}^\alpha$
ξ_1^0	α	1	1
ξ_1^A	$-A/t$	$-A/t$	0
ξ_2^A	$A^{1-\alpha}/t$	t^{-1}	1
ξ_3^A	A	0	0
ξ_4^A	A	αA	αA
ξ_5^A	$\alpha \sigma^2 A^{\alpha+1} t$	$\alpha^2 \sigma^2 A^2 t$	$\alpha^2 \sigma^2 A^2$
ξ_3^S	S	0	0
ξ_5^S	$2\alpha \sigma^2 A^\alpha S t$	$2\alpha \sigma^2 A S t$	$2\alpha \sigma^2 A S$
η_3	$1/2 - r/\sigma^2$	1	1
η_5	$2S^\alpha + t[\alpha(1-\alpha)\sigma^2 A^\alpha - 2\alpha r A^\alpha]$	$2S^\alpha + t[\alpha(1-\alpha)\sigma^2 A - 2\alpha r A]$	$2S^\alpha + \alpha(1-\alpha)\sigma^2 A - 2\alpha r A$

The operators X_ϕ reflect the superposition principle, due to the linearity of pricing PDE (1.5). The symmetries are also apparently simpler, in terms of their dependence on S , A , and t , in case $A = t\mathcal{A}^\alpha$: In that case, the translations X_1 and X_2 respectively reflect the autonomy of pricing PDE (1.5) with respect to t and A , i.e., the fact that the coefficients of V and its derivatives

in pricing PDE (1.5) do not depend on t or A . For the latter two definitions of A , dilation X_3 reflects the homogeneity of pricing PDE (1.5). Despite reflecting superficial properties of pricing PDE (1.5), these symmetries may nonetheless be of value; in particular, they may still lead to novel invariant solutions of pricing PDE (1.5). With $\alpha = 1$, the symmetries of system (1.22) in case $A = t\mathcal{A}^\alpha$ correspond to those found by Caister, Govinder and O'Hara [55] and [56] and Antoniou [15], so that the symmetry structure of pricing PDE (1.5) with arbitrary $\alpha \in \overline{\mathbb{R}}$ is as rich as that with $\alpha = 1$, but not as rich as with $\alpha = 0$; see Antoniou [19].

Remark Note that, in system (1.22), the Lie point symmetries have been expressed in the standard operator form; they are of the form $X = \xi^0 \partial_t + \xi^A \partial_A + \xi^S \partial_S + \eta \partial_V$, with the coefficients being as defined in system (1.16). Under the second prolongation of this operator, $X^{(2)} \equiv X - \zeta^0 \partial_{V_t} + \zeta^A \partial_{V_A} + \zeta^S \partial_{V_S} + \zeta^{SS} \partial_{V_{SS}} \equiv X^{(1)} + \zeta^{SS} \partial_{V_{SS}}$, the determining equation (1.21) reads simply $X^{(2)}\Phi = 0$. That is, the equation Φ [in this case, equation (1.5) solved for V_t] is invariant under $X^{(2)}$, which derives from the initial perturbations of the (in)dependent variables (1.16), as expected; the character of $X^{(2)}$ as a total differentiation operator is also clear. The prolongations $X^{(n)}$ for n up to the order of Φ , define [first-order, $X^{(1)}$] contact and (higher-order) Lie-Bäcklund symmetries; like the point symmetries X , they constitute (covariant) vector fields in the *jet space* parameterized by the (in)dependent variables and derivatives of the latter with respect to the former, with respect to t and as they appear in Φ . The perturbations, e.g., those of equations (1.16)-(1.20), which are the coefficients of these symmetries, can then be interpreted as the corresponding (one-)form/contravariant vector fields. The pricing PDE then constitutes a submanifold of the pertinent jet space, the tangent bundle of which consists of all general symmetries $X^{(n)}$, and in particular those of lower order $X^{(i)}$ with $0 \leq i < n$, including the point symmetries (with $i = 0$). Solutions of the pricing PDE are realized as its integral curves in the jet space, and *invariant solutions* are those tangent to elements of the tangent bundle (prototypically, point symmetries), at *all* points, whereas elements of the tangent bundle need be in the tangent space of only one point on the pricing PDE manifold. This remark is intended as an heuristic description of the differential geometry underlying Lie symmetry analysis, to motivate and contextualize the search for invariant solutions of pricing PDE (1.5); comprehensive texts include those of Olver [239], Ovsiannikov [240], and Bluman and Kumei [39].

Proceeding in case $A = t\mathcal{A}^\alpha$, the commutators of symmetries (1.22) are given in table 1.1:

	X_1	X_2	X_3	X_4	X_5
X_1	0	0	0	0	0
X_2	0	0	0	αX_2	$2\alpha\sigma^2[X_4 - (\alpha^{-1/2} + r/\sigma^2)X_3]$
X_3	0	0	0	0	0
X_4	0	$-\alpha X_2$	0	0	αX_5
X_5	0	$-2\alpha\sigma^2[X_4 - (\alpha^{-1/2} + r/\sigma^2)X_3]$	0	$-\alpha X_5$	0

Table 1.1: Commutators, $[X_i, X_j] \equiv X_i X_j - X_j X_i$, of symmetries (1.22) with $A = t\mathcal{A}^\alpha$.

Like the symmetries themselves, this table reduces to Table I of Caister, Govinder and O'Hara [55] in the case $\alpha = 1$, except for multiplication by $2\sigma^2$ of the commutators $\pm[X_2, X_5]$. As

such, slight modifications of formula (8) of Caister, Govinder and O'Hara [55] may be found by inspection, which reveal that pricing PDE (1.7), for arbitrary $\alpha \in \overline{\mathbb{R}}$, has the *same* Lie point symmetry algebra as the arithmetic case:

$$\begin{aligned}
e_1 &\equiv 2X_2 \\
e_2 &\equiv \frac{1}{\alpha} \left[X_4 - \left(\frac{\alpha-1}{2} + \frac{r}{\sigma^2} \right) X_3 \right] \\
e_3 &\equiv -\frac{X_5}{2\alpha^2\sigma^2} \\
e_4 &\equiv 2X_1 \\
e_5 &\equiv -X_3
\end{aligned} \tag{1.23}$$

Indeed, the commutators in table 1.2 match table II of Caister, Govinder and O'Hara [55]:

	e_1	e_2	e_3	e_4	e_5
e_1	0	e_1	$-2e_2$	0	0
e_2	$-e_1$	0	e_3	0	0
e_3	$2e_2$	$-e_3$	0	0	0
e_4	0	0	0	0	0
e_5	0	0	0	0	0

Table 1.2: Commutators, $[e_i, e_j] \equiv e_i e_j - e_j e_i$, of symmetries (1.23) with $A = t\mathcal{A}^\alpha$.

As remarked by Caister, Govinder and O'Hara [55], the (three-dimensional split simple) Lie algebra spanned by generators $\{e_1, e_2, e_3\}$ is $\mathfrak{sl}(2, \mathbb{R})$, represented by the algebra of 2×2 real matrices with zero trace. Each of e_4 and e_5 span an abelian algebra isomorphic to \mathbb{R} , denoted A_1 , and the block-diagonal structure of the commutators in table 1.2 implies a (direct sum) complete reduction of the Lie point symmetry algebra of pricing PDE (1.7), for arbitrary $\alpha \in \overline{\mathbb{R}}$; as in Caister, Govinder and O'Hara [55], $\mathfrak{sl}(2, \mathbb{R}) \oplus 2A_1$. As such, the adjoint ($[\exp(\delta e_i), e_j]$) table (III) of Caister, Govinder and O'Hara [55] remains unchanged, in terms of the new symmetries (1.23), as does the optimal [not necessarily unique, minimal set of algebras dissimilar to one another, such that all symmetry subalgebras are similar (conjugate) to exactly one member of the set] system of one-dimensional subalgebras: Given ($\forall 1 \leq i \leq 6$) $a_i \in \mathbb{C}$ and $\epsilon \in \{\pm 1\}$,

$$\begin{aligned}
&(e_1) \\
&(e_5) \\
&(e_4 + a_1 e_5) \\
&(e_1 + \epsilon e_5) \\
&(e_1 + \epsilon e_4 + a_2 e_5) \\
&(e_2 + a_3 e_4 + a_4 e_5) \\
&(e_1 + e_3 + a_5 e_4 + a_6 e_5)
\end{aligned} \tag{1.24}$$

Expressing each of these algebras/generators in terms of symmetries (1.23) and (1.22) in the case $A = t\mathcal{A}^\alpha$, the optimal symmetries follow:

Theorem 1.2.2 *If $a, b \in \mathbb{C}$, then equation (1.7) has optimal Lie point symmetries;*

$$\begin{aligned}
Y_1 &\equiv \partial_A \\
Y_2 &\equiv V\partial_V \\
Y_3 &\equiv \partial_t - a V\partial_V \\
Y_4 &\equiv 2\partial_A - V\partial_V \\
Y_5 &\equiv \partial_t + \partial_A - a V\partial_V \\
Y_6 &\equiv a\partial_t + A\partial_A + \frac{S}{\alpha}\partial_S - \frac{1}{\alpha}\left(b + \frac{\alpha-1}{2} + \frac{r}{\sigma^2}\right)V\partial_V \\
Y_7 &\equiv a\partial_t + \frac{4-A^2}{2}\partial_A - \frac{AS}{\alpha}\partial_S + \frac{1}{\alpha}\left(b + \frac{\alpha-1}{2}A + \frac{r}{\sigma^2}A - \frac{S^\alpha}{\alpha\sigma^2}\right)V\partial_V
\end{aligned} \tag{1.25}$$

When $\alpha = 1$, symmetries (1.25) reduce to those given in formula (13) of Caister, Govinder and O'Hara [55], recalling these authors take $A = t\mathcal{A}^\alpha T^{-1}$, and differently defining the arbitrary constant $b \in \mathbb{C}$.

For $1 \leq i \leq 7$, invariant solutions are determined for each Lie point symmetry (1.25). As for Caister, Govinder and O'Hara [55], the symmetry Y_2 yields no simplification: Given any solution, $V(S, A, t)$, of pricing PDE (1.7), the invariance condition, $0 \equiv Y_2 V = V\partial_V V = V$, yields a Lagrange system:

$$\frac{dV}{V} = \frac{dS}{0} = \frac{dA}{0} = \frac{dt}{0}$$

Integrating these equations yields characteristics S, A, t , and thus also $V(S, A, t)$, the latter being the 'reduced' solution. As such, as done by Caister, Govinder and O'Hara [55], symmetry Y_2 is ignored from this point onward. The other six symmetries (1.25) similarly yield the Lagrange systems shown in the leftmost column of the body of table 1.3:

All characteristics in table 1.3 follow directly by integrating the corresponding Lagrange systems, which in turn arise immediately from the symmetries (1.25), except for the two (common) bottommost values of the transformed dependent variable, $u \equiv u(x, \tau)$, itself a function of the transformed independent variables, x and τ . In this case, the value of u is determined via

$$\begin{aligned}
\frac{\alpha dV}{V} &= \left(b + \frac{\alpha-1}{2}A + \frac{r}{\sigma^2}A\right) \frac{2dA}{4-A^2} + \frac{S^{\alpha-1}dS}{\sigma^2 A} \\
&= \frac{b}{2}d\left(\ln \frac{2+A}{2-A}\right) - \left(\frac{\alpha-1}{2} + \frac{r}{\sigma^2}\right)d[\ln(4-A^2)] + \frac{4-A^2}{\alpha\sigma^2 A} \left[d\left(\frac{S^\alpha}{4-A^2}\right) - \frac{2AS^\alpha dA}{(4-A^2)^2}\right] \\
&= \frac{b}{2}d\left(\ln \frac{2+A}{2-A}\right) - \left(\frac{\alpha-1}{2} + \frac{r}{\sigma^2}\right)d[\ln(4-A^2)] + \frac{4-A^2}{\alpha\sigma^2 A} dx - \frac{2xdA}{\alpha\sigma^2}
\end{aligned}$$

The result now follows by direct integration, since the characteristics, in particular x , are *constants* of integration (in particular, with respect to A), so that $dx = 0$. As before, the characteristics of table 1.3 reduce to their counterparts in table IV of Caister, Govinder and O'Hara [55], in the arithmetic Asian option case that $\alpha = 1$, similarly accounting for differences in

i	Lagrange system	x	τ	u
1	$\frac{dV}{0} = \frac{dS}{0} = \frac{dA}{1} = \frac{dt}{0}$	S	t	V
3	$-\frac{dV}{aV} = \frac{dS}{0} = \frac{dA}{0} = \frac{dt}{1}$	S	A	$e^{at}V$
4	$-\frac{dV}{V} = \frac{dS}{0} = \frac{dA}{2} = \frac{dt}{0}$	S	t	$e^{\frac{A}{2}}V$
5	$-\frac{dV}{aV} = \frac{dS}{0} = \frac{dA}{1} = \frac{dt}{1}$	S	$A - t$	$e^{at}V$
6 ($a \neq 0$)	$-\frac{adV}{\left(b + \frac{\alpha-1}{2} + \frac{r}{\sigma^2}\right)V} = \frac{\alpha dS}{S} = \frac{dA}{A} = \frac{dt}{a}$	$e^{-\frac{t}{a}}S^\alpha$	$e^{-\frac{t}{a}}A$	$Ve^{\frac{t}{a}\left(b + \frac{\alpha-1}{2} + \frac{r}{\sigma^2}\right)}$
($a = 0$)	$-\frac{adV}{\left(b + \frac{\alpha-1}{2} + \frac{r}{\sigma^2}\right)V} = \frac{\alpha dS}{S} = \frac{dA}{A} = \frac{dt}{0}$	$\frac{A}{S^\alpha}$	t	$VS^{b + \frac{\alpha-1}{2} + \frac{r}{\sigma^2}}$
7 ($a \neq 0$)	$\frac{adV}{\left(b + \frac{\alpha-1}{2}A + \frac{r}{\sigma^2}A - \frac{S^\alpha}{a\sigma^2}\right)V} = -\frac{\alpha dS}{AS} = \frac{2dA}{4-A^2} = \frac{dt}{a}$	$\frac{S^\alpha}{4-A^2}$	$e^{-\frac{2t}{a}\frac{2+A}{2-A}}$	$V\left(\frac{2+A}{2-A}\right)^{-\frac{b}{2\alpha}}(4-A^2)^{\frac{\alpha-1}{2\alpha} + \frac{r}{\alpha\sigma^2}}e^{\frac{2AS^\alpha}{\alpha^2\sigma^2(4-A^2)}}$
($a = 0$)	$\frac{adV}{\left(b + \frac{\alpha-1}{2}A + \frac{r}{\sigma^2}A - \frac{S^\alpha}{a\sigma^2}\right)V} = -\frac{\alpha dS}{AS} = \frac{2dA}{4-A^2} = \frac{dt}{0}$	$\frac{S^\alpha}{4-A^2}$	t	$V\left(\frac{2+A}{2-A}\right)^{-\frac{b}{2\alpha}}(4-A^2)^{\frac{\alpha-1}{2\alpha} + \frac{r}{\alpha\sigma^2}}e^{\frac{2AS^\alpha}{\alpha^2\sigma^2(4-A^2)}}$

Table 1.3: Lagrange systems and characteristics of symmetries (1.25) [excluding Y_2]

i	Reduced pricing PDE (1.7)
1	$0 = u_\tau + \frac{1}{2}\sigma^2x^2u_{xx} + r - xu_x - u$
3	$0 = x^\alpha u_\tau + \frac{1}{2}\sigma^2x^2u_{xx} + r - xu_x - (a + r)u$
4	$0 = u_\tau + \frac{1}{2}\sigma^2x^2u_{xx} + r - xu_x - \left(\frac{x^\alpha}{2} + r\right)u$
5	$0 = (x^\alpha - 1)u_\tau + \frac{1}{2}\sigma^2x^2u_{xx} + r - xu_x - (a + r)u$
6 ($a \neq 0$)	$0 = (ax - \tau)u_\tau + \frac{a}{2}\alpha^2\sigma^2x^2u_{xx} + [a\alpha r - 1 + \frac{a}{2}\alpha(\alpha - 1)\sigma^2]xu_x - (ar + \frac{b}{\alpha} + \frac{\alpha-1}{2a} + \frac{r}{\alpha\sigma^2})u$
($a = 0$)	$0 = u_\tau + \frac{1}{2}\alpha^2\sigma^2x^2u_{xx} + [1 + \alpha(\alpha + b)\sigma^2]xu_x + \frac{1}{2}[b(b + \alpha)\sigma^2 - r + \frac{(\alpha^2-1)\sigma^2}{4} - \frac{r^2}{\sigma^2}]u$
7 ($a \neq 0$)	$0 = (4x - \frac{2}{a})\tau u_\tau + \frac{1}{2}\alpha^2\sigma^2x^2u_{xx} + [ar + \frac{1}{2}\alpha(\alpha - 1)\sigma^2]xu_x - (r - \frac{2bx}{\alpha} + \frac{8x^2}{\alpha^2\sigma^2})u$
($a = 0$)	$0 = u_\tau + \frac{1}{2}\alpha^2\sigma^2x^2u_{xx} + [ar + \frac{1}{2}\alpha(\alpha - 1)\sigma^2]xu_x - (r - \frac{2bx}{\alpha} + \frac{8x^2}{\alpha^2\sigma^2})u$

Table 1.4: Reduced pricing PDEs (1.7) of symmetries (1.25) [excluding Y_2]

defining A and b .

As in Caister, Govinder and O'Hara [55], the case $i = 1$ amounts to ignoring dependence of V on A (and α), resulting in the BS equation for a vanilla European option, the Lie point symmetry group and algebra of which are provided by Gazizov and Ibragimov [135]. As such, this case is ignored from this point onward. The remaining five cases (for i) in table 1.4 all involve α , though this is only so for the latter two symmetries (1.25) and Lagrange systems and characteristics of table 1.3. Thus, there is potentially much deviation from (invariant) solutions to the reduced equations listed in table IV of Caister, Govinder and O'Hara [55] (to which all of those in table 1.4 reduce in the arithmetic Asian option case with $\alpha = 1$), due to the power mean index, $\alpha \in \overline{\mathbb{R}}$.

However, Lie point symmetry analyses for the reduced pricing PDEs listed in table 1.4 have

so far yielded results identical to those of Caister, Govinder and O'Hara [55]; i.e., no new or alternate symmetries have been found. As such, barring as these authors do further *optimal* symmetry determination for each reduced pricing PDE listed in table 1.4 (perhaps a topic for a subsequent paper), it appears best to proceed directly, e.g., with separation of variables as indicated by the symmetry ∂_τ : This symmetry reduces the dependence of solutions to be on x only, yielding an ordinary differential equation in that variable, and a solution of which may then be multiplied by a suitable exponential function of (reversed) time, τ . This is generally unsatisfactory, as the maturity, T , does not enter the terminal payoff (1.8) exponentially, so at best a (generalized Fourier) series solution of the BVP obtains. But proceeding with separation of variables in Maple, special, e.g., Bessel, Kummer, and Whittaker, function solutions follow, similar to those reported by Caister, Govinder and O'Hara [55]. These do depend on the power mean index, $\alpha \in \overline{\mathbb{R}}$, but in fairly complicated manners which obscure the effect without cross-checking several known properties of the pertinent special functions. (This might also, however, be a suitable topic for a subsequent paper.) Thus, given the apparent lack of tractability beyond Caister, Govinder and O'Hara [55], and the inconclusive nature of their most developed results [e.g., the satisfaction of BC (1.9) and resulting accuracy concerns, despite modest computational performance depicted in their Figure 1], a naive approach is applied instead using symmetries (1.22), by which it is seen that one family of invariant solutions arises naturally by reducing pricing PDE (1.7) to a BS equation, and thence by standard transformations to the heat equation, for which a Cauchy problem is solved by applying Poisson's formula (1.38) to the payoff (1.8).

Invariant solutions

As its name suggests, an invariant solution is one invariant under the Lie symmetries of a given PDE; these symmetries map solutions of the PDE into others (i.e., all solutions are *symmetry solutions*), and the invariant ones are those that map to themselves under some symmetry. An invariant solution is determined by the symmetry under which it is invariant, with a general (point) symmetry being given by a linear combination of the generators X_1 through X_5 , as well as the subalgebra generated by the X_ϕ . Setting the latter aside, such a linear combination is

$$X \equiv \lambda_1 \xi_1^0 \partial_t + (\lambda_1 \xi_1^A + \lambda_2 \xi_2^A + \lambda_3 \xi_3^A + \lambda_4 \xi_4^A + \lambda_5 \xi_5^A) \partial_A + (\lambda_3 \xi_3^S + \lambda_4 S + \lambda_5 \xi_5^S) \partial_S + (\lambda_3 \eta_3 + \lambda_5 \eta_5) V \partial_V$$

A solution V of equation (1.5), invariant with respect to X , satisfies $XV = 0$, which is a first-order PDE to which the method of characteristics may be applied. The Lagrange system is

$$\frac{dt}{\lambda_1 \xi_1^0} = \frac{dA}{\lambda_1 \xi_1^A + \lambda_2 \xi_2^A + \lambda_3 \xi_3^A + \lambda_4 \xi_4^A + \lambda_5 \xi_5^A} = \frac{dS}{\lambda_3 \xi_3^S + \lambda_4 S + \lambda_5 \xi_5^S} = \frac{dV}{(\lambda_3 \eta_3 + \lambda_5 \eta_5) V}$$

Each definition of A needs to be analyzed separately. The case $A = t\mathcal{A}^\alpha$ has been most thoroughly studied and is now assumed.¹⁷ The appropriate Lagrange system is

$$\begin{aligned} \frac{dt}{\lambda_1} &= \frac{dA}{\lambda_2 + \lambda_4 \alpha A + \lambda_5 \alpha^2 \sigma^2 A^2} = \frac{dS}{[\lambda_4 + 2\lambda_5 \alpha \sigma^2 A] S} \\ &= \frac{dV}{(\lambda_3 + \lambda_5 [\alpha(1 - \alpha)\sigma^2 A - 2\alpha r A + 2S^\alpha]) V} \end{aligned} \quad (1.26)$$

¹⁷Preliminary investigations of the two remaining cases have yielded results no more general than those given. Though their analysis is more intricate, such could yield more general and/or flexible solutions.

Integration of the first equality leads to the following invariant:

$$\tau \equiv t - \frac{2\lambda_1}{\alpha \sqrt{4\lambda_2\lambda_5\sigma^2 - \lambda_4^2}} \arctan \frac{\lambda_4 + 2\lambda_5\alpha\sigma^2 A}{\sqrt{4\lambda_2\lambda_5\sigma^2 - \lambda_4^2}} \quad (1.27)$$

Integrating the second equality in equation (1.26) yields another invariant:

$$x \equiv \frac{\lambda_2 + \lambda_4\alpha A + \lambda_5\alpha^2\sigma^2 A^2}{S^\alpha} \quad (1.28)$$

Solving for S^α , and substituting the result into the right-hand side of equation (1.26),

$$\frac{\lambda_3 + \lambda_5[\alpha(1 - \alpha)\sigma^2 A - 2\alpha r A + \frac{2}{x}(\lambda_2 + \lambda_4\alpha A + \lambda_5\alpha^2\sigma^2 A^2)]}{\lambda_2 + \lambda_4\alpha A + \lambda_5\alpha^2\sigma^2 A^2} dA = \frac{dV}{V}$$

Integrating this equation yields an expression for V in terms of a function of x and τ , say u :

$$V = \exp\left[\frac{2\lambda_5 A + \frac{\lambda_4}{\alpha\sigma^2}}{x}\right] S^{\frac{1-\alpha}{2} - \frac{r}{\sigma^2}} \exp\left[\frac{2\lambda_3 + \lambda_4(\alpha - 1) + \frac{2\lambda_4 r}{\sigma^2}}{\alpha \sqrt{4\lambda_2\lambda_5\sigma^2 - \lambda_4^2}} \arctan \frac{\lambda_4 + 2\lambda_5\alpha\sigma^2 A}{\sqrt{4\lambda_2\lambda_5\sigma^2 - \lambda_4^2}}\right] u(x, \tau) \quad (1.29)$$

Under the change of variables from $V(S, A, t)$ to $u(x, \tau)$, equation (1.7) transforms into the following PDE in one fewer independent variable, satisfied by u :

$$0 = u_\tau + \frac{1}{2}(\alpha\sigma)^2 x^2 u_{xx} + (\alpha\sigma)^2 x u_x - \left(\frac{1 - \alpha^2}{4}\sigma^2 + r + \frac{r^2}{\sigma^2}\right) \frac{u}{2} - \frac{1}{x} \left[\lambda_1 u_\tau - \left(\lambda_3 + \left[\frac{\alpha - 1}{2} + \frac{r}{\sigma^2} \right] \lambda_4 \right) u + \left(\frac{\lambda_4^2}{2\sigma^2} - 2\lambda_2\lambda_5 \right) \frac{u}{x} \right] \quad (1.30)$$

Setting $\lambda_1 = 0$, a BS equation is recovered in two ways: First, the trivial case $\lambda_3 = \lambda_4 = \lambda_5 = 0$, which removes dependence of u on A . Second, for arbitrary λ_4 and λ_5 ,

$$\lambda_2 = \frac{\lambda_4^2}{4\sigma^2\lambda_5} \quad (1.31)$$

$$\lambda_3 = \left(\frac{1 - \alpha}{2} - \frac{r}{\sigma^2} \right) \lambda_4 \quad (1.32)$$

Continuing, the parameters λ_4 and λ_5 enter only as $\lambda \equiv \frac{\lambda_4}{2\lambda_5}$. Normalizing by λ_5 ,

$$\lambda_2 = \frac{\lambda^2}{\sigma^2} \quad (1.33)$$

$$\lambda_3 = \left(1 - \alpha - \frac{2r}{\sigma^2} \right) \lambda$$

This result may be collected in terms of the operator $\mathcal{X}_\lambda \equiv \frac{\lambda^2}{\sigma^2} X_2 + \left(1 - \alpha - \frac{2r}{\sigma^2} \right) \lambda X_3 + 2\lambda X_4 + X_5$:

Lemma 1.2.3 *Invariant (under operator \mathcal{X}_λ) solutions V of equation (1.7), may be written*

$$V(S, A, t) = S^{\frac{1-\alpha}{2} - \frac{r}{\sigma^2}} u \left[\frac{\left(A + \frac{\lambda}{\alpha\sigma^2}\right)^2}{S^\alpha}, t \right] e^{\frac{2S^\alpha}{\alpha^2\sigma^2 A + \alpha\lambda}}$$

Denoting $x \equiv \frac{\left(A + \frac{\lambda}{\alpha\sigma^2}\right)^2}{S^\alpha}$, $u(x, t)$ is a solution of the following BS PDE:

$$0 = u_\tau + \frac{1}{2}(\alpha\sigma)^2 x^2 u_{xx} + (\alpha\sigma)^2 x u_x - \left(\frac{1-\alpha^2}{4} \sigma^2 + r + \frac{r^2}{\sigma^2} \right) \frac{u}{2} \quad (1.34)$$

Aside from generalizing the intermediate result (Theorem 1) of Antoniou [15], the foregoing considers *all* invariant solutions of equation (1.7) corresponding to its point symmetries (1.22). The case with $\lambda_4 = 0 \neq \lambda_5$, i.e., of invariant solutions corresponding to generator X_5 , is analogous to that considered by Antoniou [15]. However, while he states that such a solution is the best candidate without providing any pursuant justification, the above analysis demonstrates that such a statement is not necessarily justified; namely, because recovery of the BS equation (1.34) follows from equation (1.7) for all operators \mathcal{X}_λ , not only \mathcal{X}_0 .

Equation (1.34) depends on α only through α^2 , i.e., is invariant with respect to its sign:

Corollary 1.2.4 *Given factors S and A_α respectively defined by equations (1.1) and (1.3), the price of an arithmetic Asian option defined with respect to the ‘composite’ security $S^{-1}(A_1 + \lambda\sigma^{-2})^2$ equals that of an harmonic option defined on $S(A_{-1} - \lambda\sigma^{-2})^2$. Thus, the pricing problem of an harmonic reduces to that of an arithmetic Asian option. The same applies for any $\alpha > 0$, with the $(-)\alpha$ option defined on $S^{(-)\alpha}[A_{(-)\alpha} + (-)\alpha^{-1}\lambda\sigma^{-2}]^2$.*

Though they appear to have received significantly less attention than the geometric and arithmetic cases, harmonic Asian options *do* have a financial role: As noted by Parmar [244], given a set of domestic-foreign exchange rates, the inverse of each is a foreign-domestic rate: The arithmetic average of the latter is likewise, and finally, the inverse is an average domestic-foreign rate. That is, the harmonic is a natural way of averaging foreign and other exchange rates. Further examination of such applications and the pricing of harmonic Asian options has been undertaken by Vecer [305] and Al-Azemi and Calin [9].

Transformation of equation (1.34) to the heat equation is standard (e.g., see section 5. of Antoniou [15]). In terms of $z \equiv \ln \sqrt{x}$, $4\tau \equiv T - t$, and $v(z, \tau) \equiv u(x, t)$, equation (1.34) is

$$v_\tau = \frac{1}{2}(\alpha\sigma)^2 v_{zz} + (\alpha\sigma)^2 v_z - \left(\frac{1-\alpha^2}{2} \sigma^2 + 2r + \frac{2r^2}{\sigma^2} \right) v$$

In terms of $v(\zeta, \tau) \equiv \exp \left[- \left(\frac{1-\alpha^2}{2} \sigma^2 + 2r + \frac{2r^2}{\sigma^2} \right) \tau \right] v \left(\frac{\alpha\sigma}{\sqrt{2}} \zeta - \alpha^2 \sigma^2 \tau, \tau \right)$,

$$v_{\zeta\zeta} = v_\tau \quad (1.35)$$

That is, v so defined satisfies the heat equation with unit thermal conductivity. Combined with lemma 1.2.3, the following result is apparent:

Theorem 1.2.5 *Invariant (under operator \mathcal{X}_λ) solutions V of equation (1.7), may be written*

$$V(S, A, t) = S^{\frac{1-\alpha}{2} - \frac{r}{\sigma^2}} v \left[\sqrt{2} \left(\frac{\ln \frac{A + \frac{\lambda}{\alpha\sigma^2}}{\sqrt{S^\alpha}}}{\alpha\sigma} + \alpha\sigma \frac{T-t}{4} \right), \frac{T-t}{4} \right] e^{\frac{2S^\alpha}{\alpha^2\sigma^2 A + \alpha\lambda} - \left(\frac{1-\alpha^2}{2}\sigma^2 + 2r + \frac{2r^2}{\sigma^2} \right) \frac{T-t}{4}} \quad (1.36)$$

Denoting $\zeta \equiv \sqrt{2} \left(\frac{\ln \frac{A + \frac{\lambda}{\alpha\sigma^2}}{\sqrt{S^\alpha}}}{\alpha\sigma} + \alpha\sigma \frac{T-t}{4} \right)$ and $\tau \equiv \frac{T-t}{4}$, then $v(\zeta, \tau)$ solves the heat equation (1.35).

The next step is to solve equation (1.7) subject to the *terminal* payoff BC (1.8). [Recall that the fixed-strike call is addressed without loss of generality; see section 1.3 for discussion regarding addressing the average- (put) via the fixed-strike (call) solution using relevant parity results from the literature.] Normalizing equation (1.28) by λ_5 , with λ_4 and so λ free and λ_2 from (1.33), A_T is given by S_T and $\chi \equiv \sqrt{x}$:

$$A_T = \chi \sqrt{S_T^\alpha} - \frac{\lambda}{\alpha\sigma^2}$$

Substituting this expression into BC (1.8), and evaluating equation (1.36) at $t = T$,

$$v(\zeta, 0) = \max \left\{ \left(\frac{\chi \sqrt{S_T^\alpha} - \frac{\lambda}{\alpha\sigma^2}}{T} \right)^{\frac{1}{\alpha}} - K, 0 \right\} S_T^{\frac{\alpha-1}{2} + \frac{r}{\sigma^2}} e^{-\frac{2}{\alpha^2\sigma^2} \frac{\sqrt{S_T^\alpha}}{\chi}} \quad (1.37)$$

Poisson's formula specifies the heat equation general solution; see Carslaw and Jaeger [60]:

$$v(\zeta, \tau) = \frac{1}{2\sqrt{\pi\tau}} \int_{-\infty}^{\infty} v(\eta, 0) e^{-\frac{(\zeta-\eta)^2}{4\tau}} d\eta \quad (1.38)$$

Since, for $\tau = 0$, $\eta = \frac{\sqrt{2}}{\alpha\sigma} \ln \chi$, then assuming $\alpha > 0$,¹⁸ and substituting equation (1.37) into Poisson's formula (1.38),¹⁹

$$v(\zeta, \tau) = \frac{S_T^{\frac{\alpha-1}{2} + \frac{r}{\sigma^2}}}{2\sqrt{\pi\tau}} \int_0^{\infty} \max \left\{ \left(\frac{\chi \sqrt{S_T^\alpha} - \frac{\lambda}{\alpha\sigma^2}}{T} \right)^{\frac{1}{\alpha}} - K, 0 \right\} e^{-\frac{2}{\alpha^2\sigma^2} \frac{\sqrt{S_T^\alpha}}{\chi}} e^{-\frac{\left(\zeta - \frac{\sqrt{2}}{\alpha\sigma} \ln \chi \right)^2}{4\tau}} \frac{\sqrt{2}}{\alpha\sigma\chi} d\chi$$

Denoting $\bar{\chi}_T \equiv \frac{TK^\alpha + \frac{\lambda}{\alpha\sigma^2}}{\sqrt{S_T^\alpha}} \geq 0$, since $A \geq 0$ and $\chi \geq 0$ so that ζ is real, this expression equals

$$\begin{aligned} v(\zeta, \tau) &= \frac{T^{-\frac{1}{\alpha}} S_T^{\frac{\alpha-1}{2} + \frac{r}{\sigma^2}}}{\alpha\sigma \sqrt{2\pi\tau}} \int_{\bar{\chi}_T}^{\infty} e^{-\frac{2}{\alpha^2\sigma^2} \frac{\sqrt{S_T^\alpha}}{\chi} - \frac{\left(\zeta - \frac{\sqrt{2}}{\alpha\sigma} \ln \chi \right)^2}{4\tau}} \frac{d\chi}{\chi} \left(\chi \sqrt{S_T^\alpha} - \frac{\lambda}{\alpha\sigma^2} \right)^{\frac{1}{\alpha}} \\ &\quad - \frac{K}{\alpha\sigma \sqrt{2\pi\tau}} \int_{\bar{\chi}_T}^{\infty} e^{-\frac{2}{\alpha^2\sigma^2} \frac{\sqrt{S_T^\alpha}}{\chi} - \frac{\left(\zeta - \frac{\sqrt{2}}{\alpha\sigma} \ln \chi \right)^2}{4\tau}} \frac{d\chi}{\chi} \end{aligned} \quad (1.39)$$

¹⁸Recall that the duality of corollary 1.2.4 then permits computation for $\alpha < 0$.

¹⁹Antoniou [15] does not take the relationship between ζ and x into account in equation (6.6), which appears to render the corresponding section's subsequent calculations incorrect.

Due to the α^{th} root in the first integral, it is not clear how to proceed for general $\alpha > 0$; e.g., whether the (generalized) binomial theorem would prove effective. The integral can be evaluated via the binomial theorem for $\alpha^{-1} \in \mathbb{N}$, which may be useful for evaluating the limit $\alpha \rightarrow 0^+$. But recall that the arithmetic case, with $\alpha = 1$, is of the most practical interest, and consider it going forward. Substituting the appropriately simplified expressions for ζ and τ from theorem 1.2.5, into the similarly simplified preceding formula for v , the following obtains for V under the change of variable $\sqrt{S_T \chi} \equiv y^{-1} \left(TK + \frac{\lambda}{\sigma^2} \right)$:

$$\begin{aligned}
V(S, A, t) &= \frac{1}{\sigma} \sqrt{\frac{2}{\pi(T-t)}} \left(\frac{S_T}{S} \right)^{\frac{r}{\sigma^2}} e^{\frac{2S}{A\sigma^2+\lambda} - r\left(1+\frac{r}{\sigma^2}\right)\frac{T-t}{2}} \\
&\cdot \left[\left(\frac{A}{T} + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2S_T y}{TK\sigma^2+\lambda} - \frac{2\left[\ln\left(y\frac{A\sigma^2+\lambda}{TK\sigma^2+\lambda}\sqrt{\frac{S_T}{S}}\right) + \sigma^2\frac{T-t}{2}\right]^2}{\sigma^2(T-t)}} \frac{dy}{y} \sqrt{\frac{S_T}{S}} e^{\frac{3}{8}\sigma^2(T-t)} \right. \\
&\quad \left. - \left(K + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2S_T y}{TK\sigma^2+\lambda} - \frac{2\left[\ln\left(y\frac{A\sigma^2+\lambda}{TK\sigma^2+\lambda}\sqrt{\frac{S_T}{S}}\right) + \sigma^2\frac{T-t}{4}\right]^2}{\sigma^2(T-t)}} \frac{dy}{y} \right] \quad (1.40)
\end{aligned}$$

It is clear that such V satisfies the terminal payoff BC (1.8), since $S \rightarrow S_T$ as $t \rightarrow T$, (geometric) Brownian motion having almost surely continuous sample paths, and each integrand converges to a point mass at $(TK\sigma^2 + \lambda)/(A\sigma^2 + \lambda)$, when divided by $\sigma \sqrt{\pi(T-t)}/2$. The corresponding integral then converges to $\exp[-2S_T/(A_T\sigma^2 + \lambda)]$ or 0, according as the point mass is located at a value less than or greater than one, respectively, yielding the result.

Consider next the case $S = 0$, so that $S_T = S e^{(r-\frac{\sigma^2}{2})(T-t) + \sigma\sqrt{T-t}Z} = 0$ [where $Z \sim N(0, 1)$]:

$$\begin{aligned}
V(0, A, t) &= \frac{1}{\sigma} \sqrt{\frac{2}{\pi(T-t)}} e^{-r\left(1+\frac{r}{\sigma^2}\right)\frac{T-t}{2}} \\
&\cdot \left[\left(\frac{A}{T} + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2\left[\ln\left(y\frac{A\sigma^2+\lambda}{TK\sigma^2+\lambda}\right) + \sigma^2\frac{T-t}{2}\right]^2}{\sigma^2(T-t)}} \frac{dy}{y} e^{\frac{3}{8}\sigma^2(T-t)} \right. \\
&\quad \left. - \left(K + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2\left[\ln\left(y\frac{A\sigma^2+\lambda}{TK\sigma^2+\lambda}\right) + \sigma^2\frac{T-t}{4}\right]^2}{\sigma^2(T-t)}} \frac{dy}{y} \right] \quad (1.41)
\end{aligned}$$

By similar arguments as for the terminal payoff BC (1.8),

$$\lim_{\sigma^2(T-t) \rightarrow 0^+} V(0, A, t) = e^{-r\left(1+\frac{r}{\sigma^2}\right)\frac{T-t}{2}} \max \left\{ \frac{A}{T} - K, 0 \right\} \quad (1.42)$$

If $r = \sigma^2$, then this limit equals $e^{-r(T-t)} \max \{A/T - K, 0\}$, i.e., BC (1.9) is satisfied. It is desirable that the option price, V , should also satisfy BC (1.9), which accounts for the fact that the GBM factor, S , almost surely remains zero once it reaches that value. I.e., at that point the terminal average, $A_T/T = A/T$, is determined almost surely, and the corresponding payoff is discounted to time t . Solution (1.40) can therefore be understood as an exact solution to the pricing PDE

(1.7) with $\alpha = 1$ and the terminal payoff BC (1.8), augmented with BC (1.9) for $S = 0$. The approximation is better with smaller scale $\sigma^2(T - t)$, but violates BC (1.9) by a constant factor, $\exp[(1 - r/\sigma^2)^{(T-t)/2}]$, which of course is smaller with close r and σ^2 . The preceding observations may be collected as follows:

Theorem 1.2.6 *Solutions V of equation (1.7), invariant under the operator \mathcal{X}_λ , and satisfying the terminal payoff BC (1.8), may be expressed in the form (1.40). In the case $S = 0$, the value (1.41) of $V(0, A, t)$ better approximates BC (1.42) with smaller scale $\sigma^2(T - t)$, and better satisfies BC (1.9) for closer r and σ^2 .*

Remark The BVP consisting of equation (1.7), subject to BCs (1.8) and (1.9), is that addressed by theorem 7.5.1 of Shreve [279], save that the latter is defined for $A \in \mathbb{R}$ and subject to the corresponding asymptotic BC/transversality condition (7.5.10):

$$\lim_{A \downarrow -\infty} V(S, A, t) = 0$$

This is taken as a natural boundary (condition) for A , for the purpose of numerical solution of the pricing PDE (1.7), and so is not needed presently. And indeed, here it is assumed that $A \geq 0$, since firstly, this follows from equation (1.3) and the fact that (almost always) $S \geq 0$, and second, the argument of \ln in ζ must be nonnegative in order to yield real values for ζ .

It is straightforward but tedious to verify that solution (1.40) indeed satisfies equation (1.7). However, doing so yields hedging parameters, often as important as the price itself, which are approximate in the same sense as discussed for V . They may be listed as follows:

Corollary 1.2.7 *Price sensitivities corresponding to solution (1.40) include the following:*

$$\begin{aligned}
\Delta \equiv V_S &= \left[\frac{2}{A\sigma^2 + \lambda} + \frac{1}{2S} - \frac{r}{\sigma^2 S} + \frac{2}{\sigma^2 S(T-t)} \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) \right] V + \Delta^* \\
&\equiv \left[\frac{2}{A\sigma^2 + \lambda} + \frac{1}{2S} - \frac{r}{\sigma^2 S} + \frac{2}{\sigma^2 S(T-t)} \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) \right] V \\
&\quad + \frac{2}{\sigma^3 S} \sqrt{\frac{2}{\pi(T-t)^3}} \left(\frac{S_T}{S} \right)^{\frac{r}{\sigma^2}} e^{\frac{2S}{A\sigma^2 + \lambda} - r \left(1 + \frac{r}{\sigma^2}\right) \frac{T-t}{2}} \\
&\quad \cdot \left[\left(\frac{A}{T} + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2STy}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{2} \right]^2}{\sigma^2(T-t)}} \frac{\ln(y)}{y} dy \sqrt{\frac{S_T}{S}} e^{\frac{3}{8}\sigma^2(T-t)} \right. \\
&\quad \left. - \left(K + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2STy}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{4} \right]^2}{\sigma^2(T-t)}} \frac{\ln(y)}{y} dy \right] \\
\Omega \equiv V_S \frac{S}{V} &= \left[\frac{2}{A\sigma^2 + \lambda} + \frac{1}{2S} - \frac{r}{\sigma^2 S} + \frac{2}{\sigma^2 S(T-t)} \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) \right] S + \Omega^* \\
&\equiv \left[\frac{2}{A\sigma^2 + \lambda} + \frac{1}{2S} - \frac{r}{\sigma^2 S} + \frac{2}{\sigma^2 S(T-t)} \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) \right] S \\
&\quad + \frac{2}{\sigma^3 V} \sqrt{\frac{2}{\pi(T-t)^3}} \left(\frac{S_T}{S} \right)^{\frac{r}{\sigma^2}} e^{\frac{2S}{A\sigma^2 + \lambda} - r \left(1 + \frac{r}{\sigma^2}\right) \frac{T-t}{2}} \\
&\quad \cdot \left[\left(\frac{A}{T} + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2STy}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{2} \right]^2}{\sigma^2(T-t)}} \frac{\ln(y)}{y} dy \sqrt{\frac{S_T}{S}} e^{\frac{3}{8}\sigma^2(T-t)} \right. \\
&\quad \left. - \left(K + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2STy}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{4} \right]^2}{\sigma^2(T-t)}} \frac{\ln(y)}{y} dy \right]
\end{aligned}$$

$$\begin{aligned}
\Gamma \equiv V_{SS} &= \left[\frac{r}{\sigma^2 S^2} - \frac{1}{2S^2} - \frac{1 + 2 \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right)}{\sigma^2 S^2 (T-t)} \right] V \\
&+ \left[\frac{2}{A\sigma^2 + \lambda} + \frac{1}{2S} - \frac{r}{\sigma^2 S} + \frac{2}{\sigma^2 S (T-t)} \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) \right] \Delta \\
&+ \left[\frac{2}{A\sigma^2 + \lambda} - \frac{1}{2S} - \frac{r}{\sigma^2 S} + \frac{2}{\sigma^2 S (T-t)} \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) \right] \Delta^* \\
&+ \frac{4}{\sigma^5 S^2} \sqrt{\frac{2}{\pi(T-t)^5}} \left(\frac{S_T}{S} \right)^{\frac{r}{\sigma^2}} e^{\frac{2S}{A\sigma^2 + \lambda} - r \left(1 + \frac{r}{\sigma^2}\right) \frac{T-t}{2}} \\
&\cdot \left[\left(\frac{A}{T} + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2S_T y}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{2} \right]^2}{\sigma^2 (T-t)}} \frac{\ln(y)^2}{y} dy \sqrt{\frac{S_T}{S}} e^{\frac{3}{8}\sigma^2(T-t)} \right. \\
&\quad \left. - \left(K + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2S_T y}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{4} \right]^2}{\sigma^2 (T-t)}} \frac{\ln(y)^2}{y} dy \right] \\
-\Theta \equiv V_t &= \left[\frac{1}{2(T-t)} + \frac{r}{2} \left(1 + \frac{r}{\sigma^2} \right) + \frac{\sigma^2}{8} \right] V \\
&- \frac{2}{\sigma^3} \sqrt{\frac{2}{\pi(T-t)^5}} \left(\frac{S_T}{S} \right)^{\frac{r}{\sigma^2}} e^{\frac{2S}{A\sigma^2 + \lambda} - r \left(1 + \frac{r}{\sigma^2}\right) \frac{T-t}{2}} \\
&\cdot \left[\left(\frac{A}{T} + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2S_T y}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{2} \right]^2}{\sigma^2 (T-t)}} \frac{\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right)^2}{y} dy \sqrt{\frac{S_T}{S}} e^{\frac{3}{8}\sigma^2(T-t)} \right. \\
&\quad \left. - \left(K + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2S_T y}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{4} \right]^2}{\sigma^2 (T-t)}} \frac{\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right)^2}{y} dy \right] \\
-V_A &= \frac{\sigma^2}{A\sigma^2 + \lambda} \left[1 + \frac{2S}{A\sigma^2 + \lambda} + \frac{4}{\sigma^2 (T-t)} \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) \right] V \\
&+ \frac{4}{\sigma(A\sigma^2 + \lambda)} \sqrt{\frac{2}{\pi(T-t)^3}} \left(\frac{S_T}{S} \right)^{\frac{r}{\sigma^2}} e^{\frac{2S}{A\sigma^2 + \lambda} - r \left(1 + \frac{r}{\sigma^2}\right) \frac{T-t}{2}} \\
&\cdot \left[\left(\frac{A}{T} + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2S_T y}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{2} \right]^2}{\sigma^2 (T-t)}} \frac{\ln(y)}{y} dy \sqrt{\frac{S_T}{S}} e^{\frac{3}{8}\sigma^2(T-t)} \right. \\
&\quad \left. - \left(K + \frac{\lambda}{T\sigma^2} \right) \int_0^1 e^{-\frac{2S_T y}{TK\sigma^2 + \lambda} - \frac{2 \left[\ln \left(y \frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \sqrt{\frac{S_T}{S}} \right) + \sigma^2 \frac{T-t}{4} \right]^2}{\sigma^2 (T-t)}} \frac{\ln(y)}{y} dy \right]
\end{aligned}$$

Solution (1.40) is confirmed by substituting these expressions into equation (1.7) with $\alpha = 1$:

$$\frac{1}{2}\sigma^2 S^2 \Gamma + rS \Delta + S V_A - rV - \Theta = 0$$

Clearly, via the invariant $\zeta \equiv \zeta(S, A, t)$, symmetry analysis allows for an analytically tractable family of solutions (1.40) to the relaxed Asian option pricing problem consisting only of pricing PDE (1.7) and terminal payoff BC (1.8). It has also been shown that the supplementary BC (1.42) is approximately satisfied, more accurately for smaller scale $\sigma^2(T-t)$ and closer r and σ^2 . One interesting possibility is to consider the formulas following solution (1.40), subject to these conditions, as exact solutions and hedging parameters in such circumstances. Also note that the small scale case is particularly difficult for finite-difference approaches to Asian option pricing, as observed by Klassen [179]. However, solution (1.40) and its hedging parameters depend on the terminal value of the first factor, S_T , a (lognormally distributed) random variable.²⁰ Therefore, to apply these formulas, their expectation over S_T must be taken. This can be done via MC simulations, yielding another massive benefit over other MC approaches to Asian option pricing, e.g. Kemna and Vorst [177] and Longstaff and Schwartz [210]: Only the *terminal* factor price, S_T , needs to be simulated, as for vanilla options, rather than the entire factor price path, yielding exponential savings of computational resources; see Klassen [179], e.g., for some discussion of this issue in the context of binomial pricing. However, taking the expectation of solution (1.40) with respect to S_T is analytically tractable, and a sequence of changes-of-variable yields the following:

Theorem 1.2.8 *The expectation of solution (1.40) over $S_T = S e^{(r-\frac{\sigma^2}{2})(T-t)+\sigma\sqrt{T-t}Z}$, is*

$$\begin{aligned} \mathbb{E}_{S_T} V(S, A, t) &\equiv \frac{1}{\sigma \sqrt{2\pi(T-t)}} \int_0^\infty V(S, A, t) e^{-\frac{[\ln(\frac{y}{S}) - (r-\frac{\sigma^2}{2})(T-t)]^2}{2\sigma^2(T-t)}} \frac{dy}{y} \\ &= \frac{e^{\frac{2}{\sigma^2}(T-t)}}{\sigma \sqrt{\pi(T-t)}} \int_0^\infty e^{-\frac{2(y-S)}{A\sigma^2+\lambda} - \frac{[\ln(\frac{y}{S}) - (r-\frac{\sigma^2}{2})(T-t)]^2}{\sigma^2(T-t)}} \frac{dy}{y} \\ &\cdot \left[\left(\frac{A}{T} + \frac{\lambda}{T\sigma^2} \right) \Phi \left[\sqrt{\frac{T-t}{2}} \sigma + \frac{1}{\sigma} \sqrt{\frac{2}{T-t}} \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \right) + \frac{\sqrt{2(T-t)}r}{\sigma} \right] e^{\frac{\sigma^2}{4}(T-t)} \right. \\ &\left. - \left(K + \frac{\lambda}{T\sigma^2} \right) \Phi \left[\frac{1}{\sigma} \sqrt{\frac{2}{T-t}} \ln \left(\frac{A\sigma^2 + \lambda}{TK\sigma^2 + \lambda} \right) + \frac{\sqrt{2(T-t)}r}{\sigma} \right] e^{-r(T-t)} \right] \end{aligned} \quad (1.43)$$

Here, Φ is the cumulative distribution function of a standard normal variable, $Z \sim N(0, 1)$.

Thus, the expectation with respect to S_T of solution (1.40) yields a formula similar to that of Black and Scholes [38], with the minimal terminal average, A/r , in place of S , and multiplied

²⁰Note that implementations of the terminal payoff BC (1.8) in Antoniou [15] and [19] do not take into account the distinction between S and S_T ; only the former is used, which may partly account for small differences in the final pricing formula for geometric Asian options in Antoniou [19] and Zhang [327].

by the moment generating function (MGF) of a lognormal random variable,²¹ which always exists since its argument, $-2/(A\sigma^2 + \lambda) < 0$. Though it no longer depends on S_T , in contrast to solution (1.40), it no longer satisfies the pricing PDE (1.7), since the lognormal distribution of S_T , by which it is obtained as the expectation of solution (1.40), depends on both S and t .

In case $A = KT$ ('at-the-money'²²), expression (1.43) for $\mathbb{E}_{S_T} V(S, A, t) \equiv \bar{V}(S, A, t; \lambda)$ is

$$\begin{aligned} \bar{V}(S, KT, t; \lambda) &= \frac{e^{\frac{r}{\sigma^2}(T-t)}}{\sigma \sqrt{\pi(T-t)}} \left(K + \frac{\lambda}{T\sigma^2} \right) \int_0^\infty e^{-\frac{2(y-S)}{KT\sigma^2+\lambda} - \frac{\left[\ln\left(\frac{y}{S}\right) - \left(r - \frac{\sigma^2}{2}\right)(T-t) \right]^2}{\sigma^2(T-t)}} \frac{dy}{y} \\ &\cdot \left[\Phi \left(\sqrt{\frac{T-t}{2}} \sigma + \frac{\sqrt{2(T-t)}r}{\sigma} \right) e^{\frac{\sigma^2}{4}(T-t)} - \Phi \left(\frac{\sqrt{2(T-t)}r}{\sigma} \right) e^{-r(T-t)} \right] \end{aligned} \quad (1.44)$$

In the in-the-money case that $A \geq KT$, it is known that²³

$$V(S, A, t) = \left(\frac{A}{T} - K \right) e^{-r(T-t)} + S \frac{1 - e^{-r(T-t)}}{rT} \quad (1.45)$$

Equating expressions (1.44) and (1.45) in the case that $A = KT$,

$$\begin{aligned} S \frac{1 - e^{-r(T-t)}}{rT} &= \frac{e^{\frac{r}{\sigma^2}(T-t)}}{\sigma \sqrt{\pi(T-t)}} \left(K + \frac{\lambda}{T\sigma^2} \right) \int_0^\infty e^{-\frac{2(y-S)}{KT\sigma^2+\lambda} - \frac{\left[\ln\left(\frac{y}{S}\right) - \left(r - \frac{\sigma^2}{2}\right)(T-t) \right]^2}{\sigma^2(T-t)}} \frac{dy}{y} \\ &\cdot \left[\Phi \left(\sqrt{\frac{T-t}{2}} \sigma + \frac{\sqrt{2(T-t)}r}{\sigma} \right) e^{\frac{\sigma^2}{4}(T-t)} - \Phi \left(\frac{\sqrt{2(T-t)}r}{\sigma} \right) e^{-r(T-t)} \right] \end{aligned}$$

In terms of the variables $S_\lambda \equiv \frac{2S}{KT\sigma^2+\lambda}$ and $y_S \equiv \frac{y}{S}$, this expression simplifies to

$$\begin{aligned} S_\lambda e^{-S_\lambda} \frac{\sigma^2}{2r} \left[1 - e^{-r(T-t)} \right] &= \frac{e^{\frac{r}{\sigma^2}(T-t)}}{\sigma \sqrt{\pi(T-t)}} \int_0^\infty e^{-S_\lambda y_S - \frac{\left[\ln y_S - \left(r - \frac{\sigma^2}{2}\right)(T-t) \right]^2}{\sigma^2(T-t)}} \frac{dy_S}{y_S} \\ &\cdot \left[\Phi \left(\sqrt{\frac{T-t}{2}} \sigma + \frac{\sqrt{2(T-t)}r}{\sigma} \right) e^{\frac{\sigma^2}{4}(T-t)} - \Phi \left(\frac{\sqrt{2(T-t)}r}{\sigma} \right) e^{-r(T-t)} \right] \end{aligned}$$

Given variables S and t and parameters σ , r , T , and K , this expression may be numerically solved for S_λ and λ itself. In fact, the same may be done in the in-the-money case, substituting formula (1.45) into expression (1.43). This permits to generate arbitrarily fine, exact realizations of $\tilde{\lambda}(S, A, t; \sigma, r, T, K)$ satisfying the resulting equations. Such realizations may then be extrapolated to out-of-the-money cases, and substituted into formula (1.43) to yield price predictions, $\bar{V}(S, A, t; \sigma, r, T, K; \tilde{\lambda})$: This is pursued in the following subsection on numerical experiments, specifically in subsection 1.2.4, where it is found to significantly outperform the MatLab built-in MC routines for arithmetic Asian option pricing, in sufficiently near-the-money cases which yield useful extrapolations, $\tilde{\lambda}(S, A, t; \sigma, r, T, K)$.

²¹This has been studied, e.g., by Tellambura and Senaratne [297].

²²Once $A = KT$, the arithmetic Asian option will be exercised, since A is nondecreasing (since $S \geq 0$). For similar reasons, the cases $A \geq KT$ are respectively referred to as being 'in-[out-of-]the-money.'

²³See equation (16) of Kemna and Vorst [177].

Optimal invariant solutions

For completeness the naive procedure used with the Lie point symmetries (1.22), is also applied to the optimal symmetries (1.25). Eventually, the outcome of the chosen approach [choosing linear combinations of the Lie point symmetries (1.22) or (1.25) which reduce pricing PDE (1.7) to a BS equation, and thence to an heat equation to which Poisson's formula (1.38) applies] reduces to that of lemma 1.2.3, despite the (only) apparently greater parametric flexibility: Assign to each optimal Lie point symmetry (1.25) a coefficient, ($\forall 1 \leq i \leq 7$) $\mu_i \in \mathbb{C}$, and independent parameters, a_i and b_i , with such parameters being zero when the corresponding symmetry (1.25) has no a or b , respectively. [In principle, these can be varied independently across the symmetries (1.25), with the new parameters, a_3, a_5, a_6, a_7, b_6 , and b_7 , given as functions of the parameters, ($\forall 1 \leq i \leq 6$) $a_i \in \mathbb{C}$, defined in the optimal system of one-dimensional Lie subalgebras (1.24).] The most general (optimal) Lie point symmetry is then

$$Y \equiv (\mu_3 + \mu_5 + a_6\mu_6 + a_7\mu_7)\partial_t + \left(\mu_1 + 2\mu_4 + \mu_5 + A\mu_6 + \frac{4-A^2}{2}\mu_7 \right) \partial_A + (\mu_6 - \mu_7A) \frac{S\partial_S}{\alpha} + \left[\mu_2 - a_3\mu_3 - \mu_4 - a_5\mu_5 - \frac{\mu_6}{\alpha} \left(b_6 + \frac{\alpha-1}{2} + \frac{r}{\sigma^2} \right) + \frac{\mu_7}{\alpha} \left(b_7 + \frac{\alpha-1}{2}A + \frac{rA}{\sigma^2} - \frac{S^\alpha}{\alpha\sigma^2} \right) \right] V\partial_V \quad (1.46)$$

An (optimal Y -)invariant solution V of equation (1.5) with $YV = 0$ has the Lagrange system

$$\frac{dt}{\mu_3 + \mu_5 + a_6\mu_6 + a_7\mu_7} = \frac{dA}{\mu_1 + 2\mu_4 + \mu_5 + A\mu_6 + \frac{4-A^2}{2}\mu_7} = \frac{\alpha dS}{(\mu_6 - \mu_7A)S} = \frac{dV}{\left[\mu_2 - a_3\mu_3 - \mu_4 - a_5\mu_5 - \frac{\mu_6}{\alpha} \left(b_6 + \frac{\alpha-1}{2} + \frac{r}{\sigma^2} \right) + \frac{\mu_7}{\alpha} \left(b_7 + \frac{\alpha-1}{2}A + \frac{rA}{\sigma^2} - \frac{S^\alpha}{\alpha\sigma^2} \right) \right] V} \quad (1.47)$$

Integration of the first equality leads to the following invariant:

$$\tau \equiv t - \frac{2(\mu_3 + \mu_5 + a_6\mu_6 + a_7\mu_7)}{\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \operatorname{arctanh} \frac{\mu_7A - \mu_6}{\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \quad (1.48)$$

Integrating the second equality in equation (1.47) yields another invariant:

$$x \equiv \frac{A^2\mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7}{S^\alpha} \quad (1.49)$$

Solving for S^α , and substituting the result into the right-hand side of equation (1.47), $\frac{dV}{VdA}$ equals

$$\frac{\mu_2 - a_3\mu_3 - \mu_4 - a_5\mu_5 - \frac{\mu_6}{\alpha} \left(b_6 + \frac{\alpha-1}{2} + \frac{r}{\sigma^2} \right) + \frac{\mu_7}{\alpha} \left[b_7 + \frac{\alpha-1}{2}A + \frac{rA}{\sigma^2} - \frac{1}{\alpha\sigma^2x} (A^2\mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7) \right]}{\lambda_2 + \lambda_4\alpha A + \lambda_5\alpha^2\sigma^2A^2}$$

Integrating this equation yields an expression for V in terms of a function of x and τ , say u :

$$V = \exp \left[\frac{2(\mu_7 A - \mu_6)}{\alpha^2 \sigma^2 x} \right] S^{\frac{1-\alpha}{2} - \frac{r}{\sigma^2}} u(x, \tau) \cdot \exp \left[\frac{2[\alpha(\mu_2 - a_3 \mu_3 - \mu_4 - a_5 \mu_5) - b_6 \mu_6 + b_7 \mu_7]}{\alpha \sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \operatorname{arctanh} \frac{\mu_7 A - \mu_6}{\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \right] \quad (1.50)$$

Under the change of variables from $V(S, A, t)$ to $u(x, \tau)$, equation (1.7) transforms into the following PDE in one fewer independent variable, satisfied by u :

$$0 = u_\tau + \frac{1}{2}(\alpha\sigma)^2 x^2 u_{xx} + (\alpha\sigma)^2 x u_x - \left(\frac{1-\alpha^2}{4} \sigma^2 + r + \frac{r^2}{\sigma^2} \right) \frac{u}{2} + \frac{2}{\alpha x} (\alpha[\mu_3 + \mu_5 + a_6 \mu_6 + a_7 \mu_7] u_\tau - [\alpha(\mu_2 - a_3 \mu_3 - \mu_4 - a_5 \mu_5) - b_6 \mu_6 + b_7 \mu_7] u) - \frac{2}{\alpha^2 \sigma^2 x^2} [\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)] u \quad (1.51)$$

Consider the following system:

$$\begin{aligned} 0 &= \mu_3 + \mu_5 + a_6 \mu_6 + a_7 \mu_7 \\ 0 &= \alpha(\mu_2 - a_3 \mu_3 - \mu_4 - a_5 \mu_5) - b_6 \mu_6 + b_7 \mu_7 \\ -\frac{\mu_6^2}{2\mu_7} &= \mu_1 + 2\mu_4 + \mu_5 + 2\mu_7 \end{aligned}$$

One solution to this system is the following:

$$-\mu_1 = 2\mu_4 + \mu_5 + \frac{\mu_6^2}{2\mu_7} + 2\mu_7 \quad (1.52)$$

$$\alpha\mu_2 = \alpha\mu_4 - \alpha(a_3 - a_5)\mu_5 - (\alpha a_3 a_6 - b_6)\mu_6 - (\alpha a_3 a_7 + b_7)\mu_7 \quad (1.53)$$

$$-\mu_3 = \mu_5 + a_6 \mu_6 + a_7 \mu_7 \quad (1.54)$$

[Another solution consists of $\mu_6 = 0 = \mu_7$ and the corresponding simplifications of equations (1.53) and (1.54), but is ignored as this removes all dependence of V on A .] Substitution of expressions (1.52)-(1.54) yields the expected BS and heat equations in the following analogue of lemma 1.2.3 and theorem 1.2.5:

Theorem 1.2.9 Consider the ten-dimensional family of optimal Lie point symmetries (1.46), parameterized by $\gamma \equiv (a_3, a_5, a_6, a_7, b_6, b_7, \mu_4, \mu_5, \mu_6, \mu_7) \in \mathbb{C}^{10}$, and subject to constraints (1.52)-(1.54). Then solutions V of equation (1.7), invariant under a symmetry of the family, say \mathcal{Y}_μ , may be written only in terms of $\mu \equiv -\alpha\sigma^2 \frac{\mu_6}{\mu_7}$:

$$V(S, A, t) = S^{\frac{1-\alpha}{2} - \frac{r}{\sigma^2}} u \left[\frac{A + \frac{\mu}{\alpha\sigma^2}}{S^{\frac{\alpha}{2}}}, t \right] e^{\frac{2S^\alpha}{\alpha^2 \sigma^2 A + \alpha\mu}} \quad (1.55)$$

Denoting $x \equiv \frac{A + \frac{\mu}{\alpha\sigma^2}}{\frac{\sigma^2}{2}}$, $u(x, t)$ is a solution of the following BS PDE:

$$0 = 4u_t + \frac{1}{2}(\alpha\sigma)^2 x^2 u_{xx} + \frac{3}{2}(\alpha\sigma)^2 x u_x - 2 \left(\frac{1 - \alpha^2}{4} \sigma^2 + r + \frac{r^2}{\sigma^2} \right) u \quad (1.56)$$

Denoting $\zeta \equiv \sqrt{2} \left(\frac{\ln x}{\alpha\sigma} + \alpha\sigma \frac{T-t}{4} \right)$ and $\tau \equiv \frac{T-t}{4}$, then given a solution $v(\zeta, \tau)$ of the heat equation (1.35) with unit thermal conductivity, the following is a \mathcal{Y}_μ -invariant solution of equation (1.7):

$$S^{\frac{1-\alpha}{2} - \frac{r}{\sigma^2}} v(\zeta, \tau) e^{\frac{25\alpha}{\alpha^2\sigma^2 A + \alpha\mu} - \left(\frac{1-\alpha^2}{2} \sigma^2 + 2r + \frac{2r^2}{\sigma^2} \right) \frac{T-t}{4}}$$

Thus, identifying $\lambda \equiv \mu$, and noting that in this case solutions (1.36) and (1.55), including their respective definitions of ζ , are identical, it is clear that the greater parametric flexibility obtained using the optimal Lie point symmetries (1.25) instead of symmetries (1.22), is only apparent: The technique of choosing invariant solutions which permit a reduction of pricing PDE (1.7) to a BS equation, and thence the heat equation (1.35), yields the same result [i.e., \mathcal{X}_λ -, or equivalently, \mathcal{Y}_μ -invariant solution of pricing PDE (1.7)] regardless of whether optimal Lie point symmetries are utilized.

Separation of variables for reduced pricing PDEs (1.30) and (1.51)

Some more comments regarding the solution technique [of choosing invariant solutions which permit a reduction of pricing PDE (1.7) to a BS equation, and thence the heat equation (1.35)] are in order: Considering only finite/non-‘solution’ (i.e., X_ϕ) Lie point symmetries, the reduced pricing PDE (1.30) [(1.51)] remains fully general, using the [optimal] Lie point symmetries (1.22) [(1.25)]. However, in proceeding in either case to choose invariant solutions which permit a further reduction to a BS equation, there *is* a loss of generality: This approach is computationally expedient, as further reduction of BS to heat equations, and solution of the corresponding Cauchy problem via Poisson’s formula (1.38), is standard. It is also arguably financially sensible, given the cornerstone position of BS equations in PDE approaches to option pricing. But progress *can* be made in more general cases: By imposing none or some (but not all²⁴) of the constraints (1.31)-(1.32) [(1.52)-(1.54)], two linearly independent Bessel or Whittaker function solutions to reduced pricing PDE (1.30) [(1.51)] are obtained.

Theorem 1.2.10 *For any $c \in \mathbb{C}$, two linearly independent solutions of the reduced pricing PDE (1.30), are the product of $e^{c\tau}$ and either the Whittaker $M_{\kappa,\nu}(z)$ - or $W_{\kappa,\nu}(z)$ -function (see section 7.3.1. of Temme [298]), each of the latter with the following parameters and argument:*

$$\kappa \equiv - \frac{c\lambda_1 - \lambda_3 + \lambda_4 \left(\frac{1-\alpha}{2} - \frac{r}{\sigma^2} \right)}{\alpha \sqrt{\lambda_4^2 - 4\sigma^2 \lambda_2 \lambda_5}} \quad (1.57)$$

$$\nu \equiv \frac{1}{\alpha\sigma} \sqrt{\frac{\sigma^2}{4} + r - 2c + \frac{r^2}{\sigma^2}} \quad (1.58)$$

$$z \equiv \frac{2 \sqrt{\lambda_4^2 - 4\sigma^2 \lambda_2 \lambda_5}}{\alpha\sigma^2 x} \quad (1.59)$$

²⁴Lemma 1.2.3 [theorem 1.2.9] yields BS equation (1.34) [(1.56)] via constraints (1.31)-(1.32) [(1.52)-(1.54)].

Imposition of constraint (1.32) yields the same solution but for a simplified parameter:

$$\kappa \equiv -\frac{c\lambda_1}{\alpha\sqrt{\lambda_4^2 - 4\sigma^2\lambda_2\lambda_5}} \quad (1.60)$$

Imposition of constraint (1.31) yields two linearly independent solutions of the reduced pricing PDE (1.30), given by the product of $e^{c\tau}/x$ and either the Bessel $J_\nu(z)$ - or $Y_\nu(z)$ -function (see section 9.2. of Temme [298]), each of the latter with the following parameter and argument:

$$\nu \equiv -\frac{1}{\alpha\sigma}\sqrt{\sigma^2 + 4r - 8c + 4\frac{r^2}{\sigma^2}} \quad (1.61)$$

$$z \equiv \frac{2\sqrt{2\lambda_3 + \lambda_4\left(\alpha - 1 + \frac{2r}{\sigma^2}\right) - 2\lambda_1c}}{\alpha^2\sigma^2\sqrt{\lambda_5}x} \quad (1.62)$$

Two linearly independent solutions of the reduced pricing PDE (1.51), are the product of $e^{c\tau}$ and either the Whittaker $M_{\kappa,\nu}(z)$ - or $W_{\kappa,\nu}(z)$ -function, each of the latter with the following parameter and argument, as well as parameter (1.58):

$$\kappa \equiv \frac{\alpha [c(\mu_3 + \mu_5 + a_6\mu_6 + a_7\mu_7) - \mu_2 + a_3\mu_3 + \mu_4 + a_5\mu_5] + b_6\mu_6 - b_7\mu_7}{\alpha\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \quad (1.63)$$

$$z \equiv \frac{4\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}}{\alpha^2\sigma^2x} \quad (1.64)$$

Imposition of constraint (1.54) yields an analogous solution with simplified parameters:

$$\kappa \equiv -\frac{\alpha [\mu_2 - \mu_4 + (a_3 - a_5)\mu_5 + a_3(a_6\mu_6 + a_7\mu_7)] - b_6\mu_6 + b_7\mu_7}{\alpha\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \quad (1.65)$$

$$\nu \equiv \sqrt{\frac{1}{4} - \frac{c}{\alpha^4\sigma^4}} \quad (1.66)$$

Here, the product of each Whittaker function is taken with $e^{c\tau}$, denoting

$$C \equiv \frac{c}{2\alpha^2\sigma^2} + \frac{1}{2}\left(\frac{r}{\sigma} + \frac{\sigma}{2}\right)^2 - \frac{\alpha^2\sigma^2}{8} \quad (1.67)$$

Imposition of other combinations of constraints (1.52)-(1.54) yields two linearly independent solutions of the reduced pricing PDE (1.51), given by the product of $e^{c\tau}/x$ and either the Bessel $J_\nu(z)$ - or $Y_\nu(z)$ -function, each of the latter with parameter and argument listed in table 1.5.

Imposition of constraint (1.53) alone, or together with constraint (1.54), does not yield analogous special function solutions; see Temme [298].

Implicitly this indicates that the reduced pricing PDEs (1.30) and (1.51) may be transformed into the Whittaker (a form of the confluent hypergeometric) or Bessel differential equations,

Constraints imposed	c^*	ν	z
(1.52)	c	(1.61)	$\frac{4\sqrt{\alpha[c(\mu_3+\mu_5+a_6\mu_6+a_7\mu_7)-\mu_2+a_3\mu_3+\mu_4+a_5\mu_5]+b_6\mu_6-b_7\mu_7}}{\alpha\sigma x\sqrt{\alpha\mu_7}}$
(1.52), (1.53)	c	(1.61)	$\frac{4\sqrt{(c+a_3)(\mu_3+\mu_5+a_6\mu_6+a_7\mu_7)}}{\alpha\sigma x\sqrt{\mu_7}}$
(1.52), (1.54)	$C - \frac{c}{2\alpha}\left(\frac{1}{\alpha\sigma^2} + \frac{1}{4\mu_7}\right)$	$-\sqrt{1 + \frac{c}{\alpha^3\sigma^2\mu_7}}$	$\frac{4i\sqrt{\alpha[\mu_2-\mu_4+(a_3-a_5)\mu_5+a_3(a_6\mu_6+a_7\mu_7)]-b_6\mu_6+b_7\mu_7}}{\alpha\sigma x\sqrt{\alpha\mu_7}}$

Table 1.5: Reduced pricing PDE (1.51) Bessel function solutions by constraints imposed.

which are also related to one another in myriad ways; see chapters seven and nine of Temme [298]. These solutions depend on the constraints imposed among (1.31)-(1.32) and (1.52)-(1.54), follow via separation of variables, and so consist of products of Whittaker or Bessel functions dependent on x , and exponential functions of τ . Substituting the appropriate invariants into each result of theorem 1.2.10 yields corresponding solutions of pricing PDE (1.7).

Corollary 1.2.11 *For any $c \in \mathbb{C}$, two linearly independent solutions of the pricing PDE (1.7), are the product of the Whittaker $M_{\kappa,\nu}(z)$ - or $W_{\kappa,\nu}(z)$ -function with the following expression:*

$$\begin{aligned}
& S^{\frac{1-\alpha}{2}-\frac{r}{\sigma^2}} \exp \left[ct + \frac{\left(2\lambda_5 A + \frac{\lambda_4}{\alpha\sigma^2}\right) S^\alpha}{\lambda_2 + \lambda_4 \alpha A + \lambda_5 \alpha^2 \sigma^2 A^2} + \frac{2\lambda_3 + \lambda_4(\alpha - 1) + \frac{2\lambda_4 r}{\sigma^2} - 2\lambda_1 c}{\alpha \sqrt{4\lambda_2 \lambda_5 \sigma^2 - \lambda_4^2}} \arctan \frac{\lambda_4 + 2\lambda_5 \alpha \sigma^2 A}{\sqrt{4\lambda_2 \lambda_5 \sigma^2 - \lambda_4^2}} \right] \\
& \equiv S^{\frac{1-\alpha}{2}-\frac{r}{\sigma^2}} \exp \left[ct + \frac{\left(2\lambda_5 A + \frac{\lambda_4}{\alpha\sigma^2}\right) S^\alpha}{\lambda_2 + \lambda_4 \alpha A + \lambda_5 \alpha^2 \sigma^2 A^2} - 2i\kappa \arctan \frac{\lambda_4 + 2\lambda_5 \alpha \sigma^2 A}{\sqrt{4\lambda_2 \lambda_5 \sigma^2 - \lambda_4^2}} \right]
\end{aligned} \tag{1.68}$$

The parameters κ and ν are respectively given by formulas (1.57) and (1.58), and the argument,

$$z \equiv \frac{2\sqrt{\lambda_4^2 - 4\sigma^2 \lambda_2 \lambda_5} S^\alpha}{\alpha\sigma^2 (\lambda_2 + \lambda_4 \alpha A + \lambda_5 \alpha^2 \sigma^2 A^2)} \tag{1.69}$$

Imposition of constraint (1.32) yields the same solution but for a simplified parameter (1.60), and multiple (1.68) given in terms of the latter. Imposition of constraint (1.31) yields two linearly independent solutions of the pricing PDE (1.7), given by the product of the Bessel $J_\nu(z)$ - or $Y_\nu(z)$ -function with the following expression:

$$\begin{aligned}
& \frac{S^{\frac{1}{2}-\frac{r}{\sigma^2}}}{A + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5}} \exp \left[ct + \frac{4\lambda_5 S^\alpha - 2\lambda_3 - \lambda_4(\alpha - 1) - \frac{2\lambda_4 r}{\sigma^2} + 2\lambda_1 c}{\alpha (\lambda_4 + 2\alpha\sigma^2\lambda_5 A)} \right] \\
& \equiv \frac{S^{\frac{1}{2}-\frac{r}{\sigma^2}}}{A + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5}} \exp \left[ct + \frac{4\lambda_5 S^\alpha}{\alpha (\lambda_4 + 2\alpha\sigma^2\lambda_5 A)} - \frac{\alpha (\lambda_4 + 2\alpha\sigma^2\lambda_5 A) z^2}{4\lambda_5 S^\alpha} \right]
\end{aligned} \tag{1.70}$$

The parameter ν is given by formula (1.61), and the argument,

$$z \equiv \frac{2\sqrt{2\lambda_3 + \lambda_4\left(\alpha - 1 + \frac{2r}{\sigma^2}\right) - 2\lambda_1 c} S^{\frac{\alpha}{2}}}{\alpha^2 \sigma^2 \sqrt{\lambda_5} \left(A + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5}\right)} \tag{1.71}$$

Two linearly independent solutions of the pricing PDE (1.7), are the product of the Whittaker $M_{\kappa,\nu}(z)$ - or $W_{\kappa,\nu}(z)$ -function with the following expression:

$$\begin{aligned}
& S^{\frac{1-\alpha}{2}-\frac{r}{\sigma^2}} \exp \left[ct + \frac{2(\mu_7 A - \mu_6) S^\alpha}{\alpha^2 \sigma^2 (A^2 \mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7)} \right] \\
& \cdot \exp \left[\frac{2(\alpha [\mu_2 - (c + a_3)\mu_3 - \mu_4 - (c + a_5)\mu_5] - [\alpha c a_6 + b_6]\mu_6 - [\alpha c a_7 - b_7]\mu_7)}{\alpha \sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \right] \\
& \cdot \operatorname{arctanh} \left[\frac{\mu_7 A - \mu_6}{\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \right] \\
& \equiv S^{\frac{1-\alpha}{2}-\frac{r}{\sigma^2}} \exp \left[ct - 2\kappa \operatorname{arctanh} \frac{\mu_7 A - \mu_6}{\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \right. \\
& \quad \left. + \frac{2(\mu_7 A - \mu_6) S^\alpha}{\alpha^2 \sigma^2 (A^2 \mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7)} \right]
\end{aligned} \tag{1.72}$$

The parameters κ and ν are respectively given by formulas (1.63) and (1.58), and the argument,

$$z \equiv \frac{4 \sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)} S^\alpha}{\alpha^2 \sigma^2 (A^2 \mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7)} \tag{1.73}$$

Imposition of constraint (1.54) yields an analogous solution with the simplified parameter ν (1.66), and the product of each Whittaker function is taken with the following expression:

$$\begin{aligned}
& S^{\frac{1-\alpha}{2}-\frac{r}{\sigma^2}} \exp \left[Ct - 2\kappa \operatorname{arctanh} \frac{\mu_7 A - \mu_6}{\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \right. \\
& \quad \left. + \frac{2(\mu_7 A - \mu_6) S^\alpha}{\alpha^2 \sigma^2 (A^2 \mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7)} \right]
\end{aligned} \tag{1.74}$$

Here, $C[\kappa]$ is given by [the simplified] formula (1.67) [(1.65)]. Imposition of other combinations of constraints (1.52)-(1.54) yields two linearly independent solutions of the pricing PDE (1.7), given by the product of either the Bessel $J_\nu(z)$ - or $Y_\nu(z)$ -function, each of the latter with parameters and arguments as in tables 1.5 and 1.6, with the following expression:

$$\frac{\mu_7 S^{\frac{1-\alpha}{2}-\frac{r}{\sigma^2}}}{\mu_7 A - \mu_6} \exp \left[c^* \tau^* + \frac{2}{\alpha} \cdot \frac{\psi^* + \frac{\mu_7}{\alpha \sigma^2} S^\alpha}{\mu_7 A - \mu_6} \right] \tag{1.75}$$

Imposition of constraint (1.53) alone, or together with constraint (1.54), does not yield analogous special function solutions.

Constraints	τ^*
(1.52)	$\frac{\mu_7(At-2a_7)-\mu_6(t+2a_6)-2(\mu_3+\mu_5)}{\mu_7A-\mu_6}$
(1.52), (1.53)	$\frac{\mu_7(At-2a_7)-\mu_6(t+2a_6)-2(\mu_3+\mu_5)}{\mu_7A-\mu_6}$
(1.52), (1.54)	t
Constraints	ψ^*
(1.52)	$\alpha(\mu_2 - a_3 \mu_3 - \mu_4 - a_5 \mu_5) - b_6 \mu_6 + b_7 \mu_7$
(1.52), (1.53)	$-\alpha a_3(\mu_3 + \mu_5) + a_6 \mu_6 + a_7 \mu_7$
(1.52), (1.54)	$\alpha[\mu_2 - \mu_4 + (a_3 - a_5)\mu_5] + (\alpha a_3 a_6 - b_6)\mu_6 + (\alpha a_3 a_7 + b_7)\mu_7$
Constraints	z
(1.52)	$\frac{4\sqrt{\alpha[c(\mu_3+\mu_5+a_6\mu_6+a_7\mu_7)-\mu_2+a_3\mu_3+\mu_4+a_5\mu_5]+b_6\mu_6-b_7\mu_7}S^{\frac{\alpha}{2}}}{\alpha\sigma(\mu_7A-\mu_6)\sqrt{\alpha\mu_7^{-1}}}$
(1.52), (1.53)	$\frac{4\sqrt{(c+a_3)(\mu_3+\mu_5+a_6\mu_6+a_7\mu_7)}S^{\frac{\alpha}{2}}}{\alpha\sigma(\mu_7A-\mu_6)\sqrt{\mu_7^{-1}}}$
(1.52), (1.54)	$\frac{4i\sqrt{\alpha[\mu_2-\mu_4+(a_3-a_5)\mu_5+a_3(a_6\mu_6+a_7\mu_7)]-b_6\mu_6+b_7\mu_7}S^{\frac{\alpha}{2}}}{\alpha\sigma(\mu_7A-\mu_6)\sqrt{\alpha\mu_7^{-1}}}$

Table 1.6: Pricing PDE (1.7) Bessel function solutions by constraints imposed.

Consider the following system:

$$\begin{aligned}
\lambda_2 &= -2\alpha^2\sigma^2(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7) \\
\lambda_4 &= -2\alpha\sigma^2\mu_6 \\
\lambda_5 &= \mu_7
\end{aligned} \tag{1.76}$$

Recalling that $i \arctan \frac{x}{i} \equiv \operatorname{arctanh}$, it follows that solutions (1.68)-(1.69) and (1.72)-(1.73) coincide when system (1.76) is satisfied and the corresponding κ parameters match:

$$\begin{aligned}
0 &= 2\alpha\sigma^2(\alpha[c(\mu_3 + \mu_5 + a_6\mu_6 + a_7\mu_7) - \mu_2 + a_3\mu_3 + \mu_4 + a_5\mu_5] + b_6\mu_6 - b_7\mu_7) \\
&+ c\lambda_1 + \lambda_3 - \lambda_4 \left(\frac{1-\alpha}{2} - \frac{r}{\sigma^2} \right)
\end{aligned} \tag{1.77}$$

Thus, by imposing three constraints (1.76) on five parameters, μ_1 and $(\forall 4 \leq i \leq 7) \mu_i$, and up to two further constraints (1.77) on twelve parameters, μ_2 , μ_3 , and γ , solutions (1.68)-(1.69) are indeed realized as special cases of (1.72)-(1.73). This is *not* preserved in case of respectively imposing constraints (1.32) and (1.54), since the respective ν parameters (1.58) and (1.66) differ; or $c \neq C$ (1.67), respectively in multiples (1.68) and (1.74). Similarly, the solutions (1.70)-(1.71) correspond to *none* of those with multiple (1.75) and parameters and argument as in tables 1.5 and 1.6, since for none of three cases in the latter is $c^* = c$ (first two rows) and $\tau^* = t$ (third row), so that $c^*\tau^* \neq ct$, respectively in multiples (1.75) and (1.70). So the ‘optimal’ Lie point symmetries (1.25) are unable to yield as special cases most of the solutions obtained by separation of variables from the symmetries (1.22), and vice versa, by imposing different combinations of the respectively corresponding constraints (1.31)-(1.32) and (1.52)-(1.54).

To construct solutions to pricing PDE (1.7) which satisfy the terminal payoff BC (1.8), BC

(1.9) with $S = 0$ or BC (1.12) with $A = TK^\alpha$, using those from corollary 1.2.11, it is desirable that the solutions form an orthogonal basis. Such results for Whittaker functions remain scant, but Szmytkowski and Bielski [292] have established an orthogonality relation [their formula (1.3)] for the Whittaker $W_{\kappa,\nu}(z)$ -function in case $(\exists v \in \mathbb{R}) \nu = iv$ is imaginary. For the corresponding solutions of corollary 1.2.11, this implies a constraint $c \geq (\sigma/2 + \sigma^2/2) [c \geq \alpha^4 \sigma^4/4]$ in case of the parameter ν (1.58) [(1.66)]. So restricting c , a corresponding (generalized) ‘Fourier-Whittaker’ series could be constructed, attempting to sum the corresponding solutions of corollary 1.2.11 in order to satisfy the terminal payoff BC (1.8), BC (1.9) with $S = 0$ or BC (1.12) with $A = TK^\alpha$. A similar approach could be pursued in case the parameter $\kappa = 0$ and $W_{0,iv}(z) \equiv K_{iv}(z/2) \sqrt{z/\pi}$ reduces to a MacDonald function, via the Kontorovich–Lebedev transform, as noted by Szmytkowski and Bielski [292]: One advantage in this case is that the inverse transform, used to construct a solution satisfying the terminal payoff BC (1.8), BC (1.9) with $S = 0$ or BC (1.12) with $A = TK^\alpha$, from those of corollary 1.2.11, reduces to an *index* transform over c , which enters in a comparatively straightforward manner in such solutions; see Yakubovich [318]. Two possibilities for combining the Bessel $J_\nu(z)$ -functions of corollary 1.2.11, overlapping respectively with the series expansion and integral transform approaches discussed in section 1.1.4, are the use of Fourier-Bessel series²⁵ and the Hankel transform.²⁶

Fourier-Bessel series solutions of pricing PDE (1.7)

Consider the following Fourier-Bessel series:

$$\sum_{n=1}^{\infty} d_n J_\nu \left(\frac{j_{\nu,n}}{\bar{\zeta}} \zeta \right) \quad (1.78)$$

Here, $(j_{\nu,n})_{n \in \mathbb{N}}$ are the zeros of $J_\nu(\zeta)$; $(\forall n \in \mathbb{N}) J_\nu(j_{\nu,n}) = 0$. To optimally approximate a mapping $f : [0, \bar{\zeta}] \rightarrow \mathbb{R}$ via a Fourier-Bessel series (1.78),²⁷ the optimal coefficients are

$$(\forall n \in \mathbb{N}) \quad d_n = \frac{2 \int_0^{\bar{\zeta}} \zeta f(\zeta) J_\nu \left(\frac{j_{\nu,n}}{\bar{\zeta}} \zeta \right) d\zeta}{\bar{\zeta}^2 J_{\nu+1}^2(j_{\nu,n})} \quad (1.79)$$

This requires $\nu > -1$, in order that $\{J_\nu(j_{\nu,n}\zeta/\bar{\zeta})\}$ constitutes an orthogonal basis with weighting function, $\zeta \in [0, \bar{\zeta}]$; see section 9.9. of Temme [298]. The principal difficulty is to determine the mapping f so that the terminal payoff BC (1.8), BC (1.9) with $S = 0$ or BC (1.12) with $A = TK^\alpha$, are satisfied. The first possibility necessarily renders the terminal underlying asset

²⁵The (Neumann) Bessel $Y_\nu(z)$ -functions are not used for such purposes, since they are unbounded as $z \rightarrow 0$; see section 9.3. of Temme [298].

²⁶The Y -transform is of the same form as the Hankel transform (1.93), with Neumann $Y_\nu(z)$ -function kernel in place of the Bessel $J_\nu(z)$ -function; see chapter IX of Bateman [30]. However, the inverse \mathbf{H} -transforms have *Struve* $\mathbf{H}_\nu(z)$ -function kernels, rather than Neumann $Y_\nu(z)$ -functions; see Rooney [264], chapter XI of Bateman [30], and chapter 12 of Abramowitz and Stegun [2]. The inverse Y - and \mathbf{H} -transform pair is thus less apparently suited to combining the Neumann $Y_\nu(z)$ -function solutions of corollary 1.2.11, compared to application of the (self-inverse) Hankel transform (1.93) to the corresponding Bessel $J_\nu(z)$ -function solutions, but may be worth exploring in future investigations.

²⁷In case of choosing f to satisfy the terminal payoff BC (1.8) for a fixed-strike (K) Asian *put* option, the upper bound, $\bar{\zeta} \equiv K$, since f vanishes for greater values of its argument.

price, S_T , as a parameter, the expectation over which, evaluated via MC simulations or analytically as in theorem 1.2.8, is then required for computing numerical prices; the second presents asymptotic complications as $S \rightarrow 0$ [as does BC (1.13) as $S \rightarrow \infty$], due to the multiplicative dependence on S of the multiples (1.70) and (1.75) and the arguments of table 1.6 and (1.71). As such, only BC (1.12) with $A = TK^\alpha$ is considered: In the case of solutions (1.70)-(1.71),

$$\left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} S = V(S, TK^\alpha, t) \equiv S^{\frac{1}{2} - \frac{r}{\sigma^2}} e^{ct} \sum_{n=1}^{\infty} \frac{d_n}{TK^\alpha + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5}} J_\nu(z) \cdot \exp \left[\frac{4\lambda_5 S^\alpha - 2\lambda_3 - \lambda_4(\alpha - 1) - \frac{2\lambda_4 r}{\sigma^2} + 2\lambda_1 c}{\alpha(\lambda_4 + 2\alpha\sigma^2\lambda_5 TK^\alpha)} \right] \quad (1.80)$$

[Here it is understood that the free parameters, $\lambda_1, \lambda_3, \lambda_4$, and λ_5 , may be chosen differently for each solution n ; c may *not*, since it determines ν (1.61), understood to be fixed in the Fourier-Bessel series (1.78).] The Bessel functions $J_\nu\left(\frac{j_{\nu,n}}{\bar{\zeta}}\zeta\right)$ in the Fourier-Bessel series (1.78) are chosen to have the same form as in solutions (1.70)-(1.71), whence the following is required:

$$\frac{j_{\nu,n}}{\bar{\zeta}}\zeta \equiv \frac{2\sqrt{2\lambda_3 + \lambda_4\left(\alpha - 1 + \frac{2r}{\sigma^2}\right) - 2\lambda_1 c S^{\frac{\alpha}{2}}}}{\alpha^2\sigma^2\sqrt{\lambda_5}\left(TK^\alpha + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5}\right)} \quad (1.81)$$

Enforcing this (without loss of generality) by specifying $\zeta \equiv S^{\frac{\alpha}{2}}$, expression (1.80) reads

$$\left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} \zeta^{\frac{1}{\alpha} + \frac{2r}{\alpha\sigma^2}} = e^{ct} \sum_{n=1}^{\infty} \frac{d_n}{TK^\alpha + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5}} J_\nu\left(\frac{j_{\nu,n}}{\bar{\zeta}}\zeta\right) \cdot \exp \left[\frac{4\lambda_5\zeta^2 - 2\lambda_3 - \lambda_4(\alpha - 1) - \frac{2\lambda_4 r}{\sigma^2} + 2\lambda_1 c}{\alpha(\lambda_4 + 2\alpha\sigma^2\lambda_5 TK^\alpha)} \right] \quad (1.82)$$

In order to be interpreted as a Fourier-Bessel series (1.78), the exponential function of ζ must factor out of the sum and into f , i.e., it must not depend on n . Two constraints on the free parameters, $\lambda_1, \lambda_3, \lambda_4$, and λ_5 , recalling $\zeta \equiv S^{\frac{\alpha}{2}}$, are thus as follows:

$$\frac{j_{\nu,n}}{\bar{\zeta}} \equiv \frac{2\sqrt{2\lambda_3 + \lambda_4\left(\alpha - 1 + \frac{2r}{\sigma^2}\right) - 2\lambda_1 c}}{\alpha^2\sigma^2\sqrt{\lambda_5}\left(TK^\alpha + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5}\right)} \quad (1.83)$$

$$I \equiv \frac{4\lambda_5}{\alpha(\lambda_4 + 2\alpha\sigma^2\lambda_5 TK^\alpha)} \quad (1.84)$$

(Here, I denotes an expression independent of the summation index n .) Formula (1.82) reads

$$f \equiv \frac{2}{\alpha^2\sigma^2 I} \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} \zeta^{\frac{1}{\alpha} + \frac{2r}{\alpha\sigma^2}} e^{-ct - I\zeta^2} = \sum_{n=1}^{\infty} d_n e^{-\frac{j_{\nu,n}^2}{4I\bar{\zeta}^2}} J_\nu\left(\frac{j_{\nu,n}}{\bar{\zeta}}\zeta\right) \equiv \sum_{n=1}^{\infty} \mathcal{D}_n J_\nu\left(\frac{j_{\nu,n}}{\bar{\zeta}}\zeta\right) \quad (1.85)$$

Thus $f(\zeta; t)$ is represented by a Fourier-Bessel series (1.78) determining via formula (1.79) [for \mathcal{D}_n] the coefficients in expression (1.80):

$$(\forall n \in \mathbb{N}) \quad d_n = e^{\frac{j_{v,n}^2}{4I\bar{\zeta}^2}} \frac{2 \int_0^{\bar{\zeta}} \zeta f(\zeta; t) J_\nu \left(\frac{j_{v,n}}{\bar{\zeta}} \zeta \right) d\zeta}{\bar{\zeta}^2 J_{\nu+1}^2(j_{v,n})} \quad (1.86)$$

Linearly combining these coefficients with solutions (1.70)-(1.71) as in formula (1.80) yields an expression satisfying BC (1.12) with $A = TK^\alpha$ which, *multiplied* by $e^{ct(rT)^{1/\alpha}} / [1 - e^{-r(T-t)}]^{1/\alpha}$, gives a further solution to (the linear) pricing PDE (1.7), as is readily seen by direct substitution: (This multiplication is made necessary by the corresponding dependence on t of f (1.85), which is however fortunately separable from ζ .)

$$S^{\frac{1}{2} - \frac{r}{\sigma^2}} e^{ct} \sum_{n=1}^{\infty} \frac{d_n}{A + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5}} J_\nu \left(S^{\frac{\alpha}{2}} \frac{j_{v,n}}{\bar{\zeta}} \frac{\lambda_4 + 2\alpha\sigma^2\lambda_5 TK^\alpha}{\lambda_4 + 2\alpha\sigma^2\lambda_5 A} \right) \cdot \exp \left[\left(IS^\alpha - \frac{j_{v,n}^2}{4I\bar{\zeta}^2} \right) \frac{\lambda_4 + 2\alpha\sigma^2\lambda_5 TK^\alpha}{\lambda_4 + 2\alpha\sigma^2\lambda_5 A} \right] \quad (1.87)$$

Recall that the free parameters, λ_1 , λ_3 , λ_4 , and λ_5 , are required to satisfy constraints (1.83)-(1.84), so that one simple possibility is to fix both λ_4 and λ_5 to be independent of n , so that I is also, and choose λ_1 and λ_3 such that constraint (1.83) is satisfied. Substituting formulas (1.85)-(1.86) into (1.87), and denoting $\Lambda \equiv \lambda_4/2\alpha\sigma^2\lambda_5$ and $h(A; \Lambda) \equiv (\Lambda + TK^\alpha)(\Lambda + A)^{-1}$,

$$\frac{2h(A; \Lambda)}{\bar{\zeta}^2} \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} S \sum_{n=1}^{\infty} \frac{e^{\frac{j_{v,n}^2}{4I\bar{\zeta}^2} [1-h(A; \Lambda)]}}{J_{\nu+1}^2(j_{v,n})} \cdot \int_0^{\bar{\zeta}} \zeta \left(\frac{\zeta}{S^{\frac{\alpha}{2}}} \right)^{\frac{1}{\alpha} + \frac{2r}{\alpha\sigma^2}} e^{-I[\zeta^2 - S^\alpha h(A; \Lambda)]} J_\nu \left[\frac{j_{v,n}}{\bar{\zeta}} \zeta \right] J_\nu \left[\frac{j_{v,n}}{\bar{\zeta}} S^{\frac{\alpha}{2}} h(A; \Lambda) \right] d\zeta \quad (1.88)$$

In the case of solutions with multiples (1.75) and parameters and arguments as in tables 1.5 and 1.6, the following is analogous to expression (1.80):

$$\left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} S = V(S, TK^\alpha, t) \equiv S^{\frac{1}{2} - \frac{r}{\sigma^2}} \sum_{n=1}^{\infty} \frac{\mu_7 d_n}{\mu_7 TK^\alpha - \mu_6} J_\nu(z) \exp \left[c^* \tau^* + \frac{2}{\alpha} \frac{\psi^* + \frac{\mu_7}{\alpha\sigma^2} S^\alpha}{\mu_7 TK^\alpha - \mu_6} \right] \quad (1.89)$$

[Again it is understood that the free parameters, μ_2 , μ_3 , and γ , as applicable, may be chosen differently for each solution n ; and again c may *not*, since it determines each parameter ν in table 1.5.] The Bessel functions $J_\nu \left(\frac{j_{v,n}}{\bar{\zeta}} \zeta \right)$ in the Fourier-Bessel series (1.78) are chosen to satisfy condition (1.81) for each argument z in table 1.6, again enforced (without loss of

generality) by specifying $\zeta \equiv S^{\frac{\alpha}{2}}$, so that expression (1.89) reads

$$\left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} \zeta^{\frac{1}{\alpha} + \frac{2r}{\alpha\sigma^2}} = \sum_{n=1}^{\infty} \frac{\mu_7 d_n}{\mu_7 T K^\alpha - \mu_6} J_\nu(z) \exp \left[c^* \tau^* + \frac{2}{\alpha} \cdot \frac{\psi^* + \frac{\mu_7}{\alpha\sigma^2} \zeta^2}{\mu_7 T K^\alpha - \mu_6} \right]$$

Taking $I \equiv \frac{2}{\alpha^2 \sigma^2} \left(T K^\alpha - \frac{\mu_6}{\mu_7} \right)^{-1}$, an expression independent of the summation index n , this reads

$$f \equiv \frac{2}{\alpha^2 \sigma^2 I} \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} \zeta^{\frac{1}{\alpha} + \frac{2r}{\alpha\sigma^2}} e^{-I\zeta^2} = \sum_{n=1}^{\infty} d_n e^{c^* t - \frac{j_{\nu,n}^2}{4I\zeta^2}} J_\nu \left(\frac{j_{\nu,n}}{\bar{\zeta}} \zeta \right)$$

Here, c^* is as in table 1.5; i.e., in the cases of the first two rows of the latter, the solution (1.85) results. In case of the third row, c^* depends on only one free parameter, μ_7 : If the latter varies with n , then the coefficients (1.86) are not separable in t , and extraction of a solution of pricing PDE (1.7) is not immediate as in the other cases. If μ_7 is taken independent of the summation index n , then so too must be μ_6 since $I \equiv I(\mu_6, \mu_7)$ is similarly invariant and depends on only the two free parameters, μ_6 and μ_7 . But given that there remain nine free parameters in the corresponding constraint (1.83), this may be imposed and a solution of the form (1.85) again follows, with $\exp([C - c^{(1/\alpha\sigma^2 + 1/4\mu)}] / 2\alpha] t)$ independent of the summation index n and in place of $e^{c^* t}$. Thus, imposing only two constraints (1.83)-(1.84) in all but the latter case, in which three are imposed (additionally requiring that μ_6 and so μ_7 be independent of the summation index n) and nine are left ‘free,’ the remaining parametric flexibility is only apparent, i.e., solutions of the form (1.85) necessarily follow and permit only three degrees of freedom, independent of the summation index n : The invariant $I \in \mathbb{C}$, the upper bound $\bar{\zeta} \geq 0$, and the parameter $c \in \mathbb{C}$, chosen such that the corresponding parameter $\nu > -1$ of table 1.5. In the case with three constraints imposed, the further degree of freedom, $\mu_7 \in \mathbb{C}$ independent of the summation index n , is also permitted. These results may be gathered as follows.

Theorem 1.2.12 *Linearly combining the Fourier-Bessel coefficients (1.86) with Bessel $J_\nu(z)$ -function solutions of corollary 1.2.11, the free parameters of whose arguments z must satisfy constraints (1.83)-(1.84), yields expression (1.87) which satisfies BC (1.12) with $A = T K^\alpha$. Division of this expression by the following function yields a solution of pricing PDE (1.7):*

$$g(t) \equiv e^{-c^* t} \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} \quad (1.90)$$

Here, c^* is c in the case of solutions (1.70)-(1.71) or as in table 1.5; in the case of the third row of the latter, $\mu_7 \in \mathbb{C}$ (and so $\mu_6 \in \mathbb{C}$) is required to be independent of the summation index n . The parameters $c \in \mathbb{C}$ must be such that $\nu > -1$ of table 1.5.

Given a solution prescribed by theorem 1.2.12, say, $V \equiv V(S, A, t)$, it follows that

$$(gV)_t + \frac{1}{2} \sigma^2 S^2 (gV)_{SS} + S^\alpha (gV)_A + rS (gV)_S - r(gV) = g'V = \frac{r}{\alpha[1 - e^{r(T-t)}]} gV$$

Thus, as $|\alpha| \rightarrow \infty$ in the case of lookbacks, and/or $T \rightarrow \infty$ in the case of perpetual options, the expression gV with $g \equiv g(t)$ given by formula (1.90), both satisfies BC (1.12) with $A = T K^\alpha$

and pricing PDE (1.7). Consideration of the remaining BCs, however, presents difficulties, again due to the asymptotic nature of BC (1.9) with $S = 0$ [and BC (1.13) as $S \rightarrow \infty$], together with the multiplicative dependence upon S of solutions prescribed by theorem 1.2.12. Regarding the terminal payoff BC (1.8), the primary difficulty, which may prohibit its satisfaction, is that $g \rightarrow 0$ as $t \rightarrow T$. Nonetheless, the solutions prescribed by theorem 1.2.12 may be useful, properly combined with formula (1.90), for pricing in cases with $S \gg 0$ or $t \ll T$.

Note also that the solutions prescribed by theorem 1.2.12 are assumed to satisfy $S \equiv \zeta^{2/\alpha} \in [0, \bar{\zeta}^{2/\alpha}]$, so that $\bar{\zeta} \rightarrow \infty$ should be taken for $\alpha > 0$; but some sufficiently large value, e.g., $1/\epsilon$ for the machine epsilon, ϵ , may yield effectively ‘exact’ numerical computations. For similar purposes, $c^*(t)$ may also be chosen such that $g(t) = 1$:

$$c^*(t) \equiv \frac{1}{\alpha t} \ln \frac{1 - e^{-r(T-t)}}{rT} \quad (1.91)$$

The expression $gV \equiv V$ then both satisfies BC (1.12) with $A = TK^\alpha$ and pricing PDE (1.7), when the dependence of c^* on t in formula (1.91) is ignored in computing V_t . Effectively, this yields solutions of pricing PDE (1.7) along slices of fixed time t , which collectively satisfy BC (1.12) with $A = TK^\alpha$; the mapping (1.91) indicates how the solutions prescribed by theorem 1.2.12 are selected for each time slice t . Only values of c^* (1.91) may be accepted which also satisfy $\nu > -1$; in the case of $c^* = c$ and ν (1.61), for solutions (1.70)-(1.71) or those with multiple (1.75) corresponding to the first two rows of tables 1.5 and 1.6,

$$\begin{aligned} -1 &< -\frac{1}{\alpha\sigma} \sqrt{\sigma^2 + 4r - \frac{8}{\alpha t} \ln \frac{1 - e^{-r(T-t)}}{rT} + 4\frac{r^2}{\sigma^2}} \\ &\equiv \\ \frac{1}{\alpha t} \ln \frac{1 - e^{-r(T-t)}}{rT} &> \frac{1 - \alpha^2}{8}\sigma^2 + \frac{r}{2} + \frac{r^2}{2\sigma^2} \end{aligned} \quad (1.92)$$

For those with multiple (1.75) corresponding to the third rows of tables 1.5 and 1.6,

$$-1 < -\sqrt{1 + \frac{8}{\alpha^2\sigma^2} \left[\frac{1}{2} \left(\frac{r}{\sigma} + \frac{\sigma}{2} \right)^2 - \frac{\alpha^2\sigma^2}{8} - \frac{1}{\alpha t} \ln \frac{1 - e^{-r(T-t)}}{rT} \right]}$$

This immediately simplifies also to condition (1.92). The behavior of mapping (1.91) is not immediately clear, e.g., whether it always has an interior extremal point $t \in (0, T)$, and if so, whether that point is unique. Some preliminary investigations, plotted in figure 1.2, suggest that the mapping C^* , defined by moving all terms to the left of the second inequality (1.92), may be convex or concave for relevant parameters considered elsewhere, but, e.g. in the case $r \rightarrow 0$, may only have a boundary extremal point at $t = 0$ (shown in blue in figure 1.2). Both the plots of figure 1.2 and the form of mapping (1.91) exhibit some similarity to the Lambert W function, but whilst mapping (1.91) is smooth for $t \in (0, T)$, both the first- and second-derivative tests yield similarly nonlinear equations, whose further consideration might be sensibly pursued in a subsequent paper. Note also that scaling T away from 0.25 proportionally scales the plots of figure 1.2, so alternate values of this parameter were not pursued/presented.

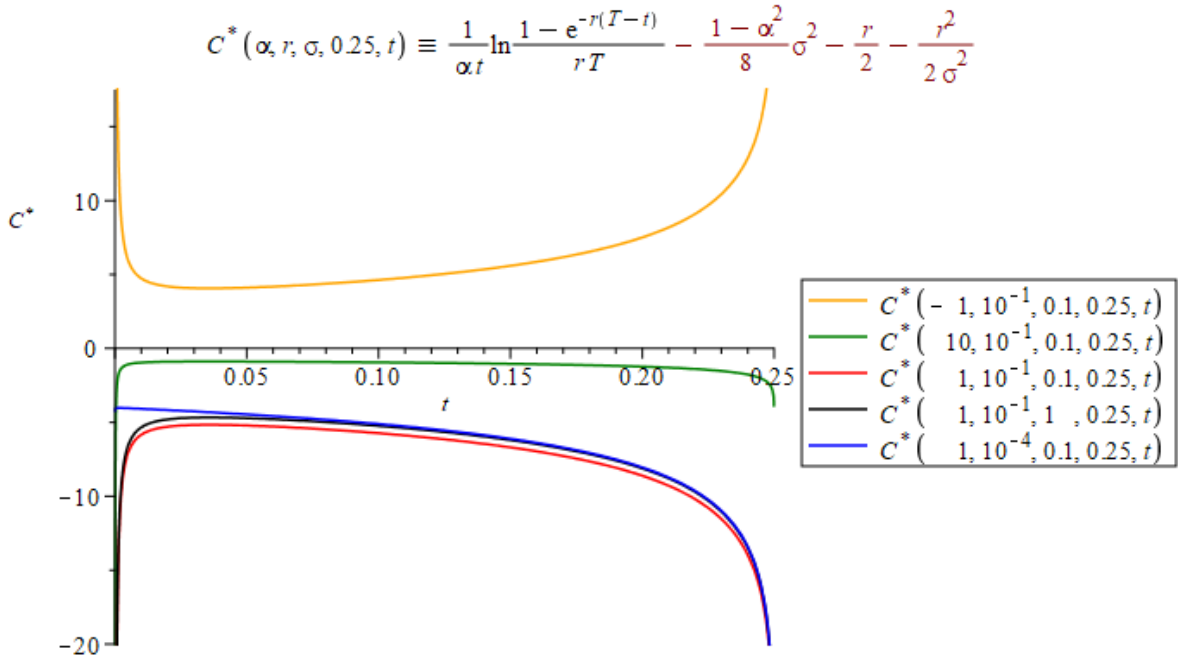


Figure 1.2: Behavior of the function $C^*(\alpha, r, \sigma, 0.25, t)$ defined by the inequality (1.92)

Added to this difficulty, is sensibly choosing I (1.84) from a financial perspective, in a manner suitable from a numerical one: Clearly, $I > 0$ so that $0 \leq f$ (1.85), which however results in the multiples of the Fourier-Bessel coefficients (1.86), $\exp[\frac{j_{\nu n}^2}{4I\zeta^2}] \rightarrow \infty$ as $n \rightarrow \infty$ for any finite bound $\zeta > 0$. This, as well as the sensitivity of f (1.85) to both I and ζ , complicates numerical implementation of the solutions prescribed by theorem 1.2.12, to the extent that such analysis may best be left to another study. Choosing similarly ‘representative’ parameters as in figure 1.2, but with $r = 0.01 = \sigma^2$ to address some of the noted sensitivity of f (1.85), figure 1.3 depicts the form of the latter.

Here, the value $c = \sqrt[7]{640}$ is such that $-1/2 = \nu$ (1.61), so that $\sqrt{\pi z} J_{-1/2}(z) \equiv \sqrt{2} \cos z$ and ($\forall n \in \mathbb{N}$) $j_{-1/2, n} = (n - 1/2)\pi$, which expedites the nonetheless difficult numerical implementation of the solutions prescribed by theorem 1.2.12, which of course in this case reduce to Fourier cosine series; see section 9.11. of Temme [298]. Also, the Fourier-Bessel series (1.78) approximation for f (1.85) additionally assumes that $f(\zeta) = 0$, which also implies $I > 0$; $I = 1$ was chosen in figure 1.3 as it yielded values of a single order of magnitude. The combined formula (1.88) sacrifices the ability to analyze separately the impact of f (1.85) and d_n (1.86) in formula (1.87), but is somewhat more amenable to numerical computation than the latter, as it simultaneously removes some asymptotic hindrances from the separate formulas, as is clear from its form. And it also reduces the influence of the remaining free parameters, λ_4 and λ_5 , to the single $\Lambda \equiv \lambda_4/2\alpha\sigma^2\lambda_5$. Clearly this latter should lie outside of $[-TK^\alpha, 0]$, to avoid nonnegative values of $h(A; \Lambda) \equiv (\Lambda + TK^\alpha)(\Lambda + A)^{-1}$.

Recall that, by imposing three constraints (1.76) on five parameters, μ_1 and ($\forall 4 \leq i \leq 7$) μ_i , and up to two further constraints (1.77) on twelve parameters, μ_2, μ_3 , and γ , solutions (1.68)-

$$f(\zeta, t, \alpha, \sigma, r, T, I, c) \equiv \frac{2}{\alpha^2 \sigma^2 I} \left(\frac{1 - e^{-r(T-t)}}{rT} \right)^{\frac{1}{\alpha}} \zeta^{\frac{1}{\alpha} + \frac{2r}{\alpha \sigma^2}} e^{-ct - I\zeta^2}$$

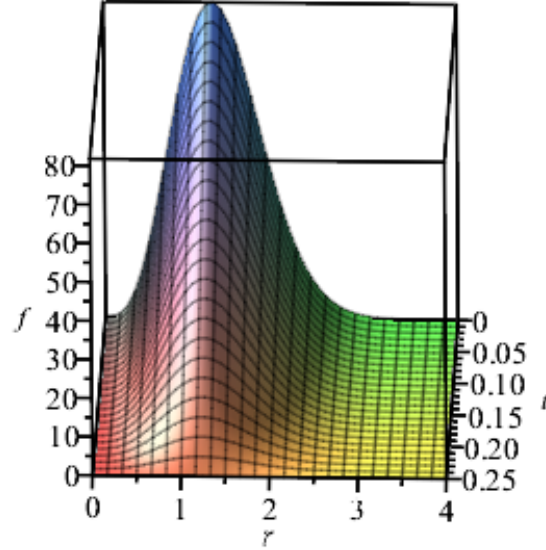


Figure 1.3: Behavior of the function $f(\zeta, t; 1, 0.1, 0.01, 0.25, 1, 7/640)$

(1.69) are realized as special cases of (1.72)-(1.73). Thus, a ‘Fourier-Whittaker’ analogue of the Fourier-Bessel series (1.85) need only be considered for solutions (1.72)-(1.73). As noted, Szmytkowski and Bielski [292] have established an *orthogonality* relation for the Whittaker $W_{\kappa, i\nu}(z)$ -function in case the second parameter is *imaginary*. However, such a Fourier-Whittaker series expansion would additionally require that the collection of such functions not only be orthogonal, but also be *complete*, i.e., a basis, e.g., for $L^2(0, \infty)$. This rests on the zeros (eigenvalues) of such (eigen)functions, as for $j_{\nu, n}$ in the Fourier-Bessel series (1.78), and more specifically, $[n(-1/2)]\pi$ in the case of Fourier (co)sine series. There has apparently been little development on this topic, and such has focused on the case of *real* second parameter; see Esparza, Lopez and Sesma [112] and section 17.3 of Buchholz [51]. Thus this avenue is not pursued.

Hankel transform solutions of pricing PDE (1.7)

Let $(\forall 1 \leq i \leq 7) \{\mu_{i,\beta}\}, \{a_{3,\beta}\}, (\forall 5 \leq j \leq 7) \{a_{j,\beta}\}, \{b_{6,\beta}\}, \{b_{7,\beta}\},$ and $\{c_\beta\}$, denote directed sets (in \mathbb{C}), which may be chosen arbitrarily to define a corresponding family of solutions of pricing PDE (1.7), say, $\{m_\beta J_{\nu_\beta}(z_\beta)\}$, with each multiple m_β , parameter ν_β , and argument z_β computed according to such choices from formulas (1.70)-(1.71) and (1.75), and those of tables 1.5 and 1.6. Taking $(\forall 0 < \beta, \kappa) \nu_\beta \equiv \nu \geq -1/2, z_\beta \equiv \kappa\beta$, consider the Hankel transform,

$$\mathfrak{H}_\nu[g(\beta)](\kappa) \equiv \int_0^\infty \sqrt{\kappa\beta} J_\nu(\kappa\beta) g(\beta) d\beta \quad (1.93)$$

(See section 9.4. of Temme [298].) Given any mapping $g \in L^2(0, \infty)$, the Hankel transform (1.93) exists and has (self-)inverse g . In the case of solutions (1.70)-(1.71), the requirement that $z_\beta \equiv \kappa\beta$ can be addressed without loss of generality by taking $\kappa \equiv S^{\alpha/2}$ and

$$\beta \equiv \frac{2 \sqrt{2\lambda_3 + \lambda_4 \left(\alpha - 1 + \frac{2r}{\sigma^2} \right) - 2\lambda_1 c}}{\alpha^2 \sigma^2 \sqrt{\lambda_5} \left(TK^\alpha + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5} \right)} \quad (1.94)$$

As in the case of, and for the same reason as in, the preceding section on Fourier-Bessel series solutions of pricing PDE (1.7), only BC (1.12) with $A = TK^\alpha$ is considered. Similarly,

$$f \equiv \frac{2}{\alpha^2 \sigma^2 I} \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} \kappa^{\frac{1}{2} + \frac{1}{\alpha} + \frac{2r}{\alpha\sigma^2}} e^{-ct - I\kappa^2} = \int_0^\infty \sqrt{\kappa\beta} \frac{e^{-\frac{\beta^2}{4I}}}{\sqrt{\beta}} J_\nu(\kappa\beta) g(\beta) d\beta \equiv \int_0^\infty \sqrt{\kappa\beta} J_\nu(\kappa\beta) \mathcal{G}(\beta) d\beta \quad (1.95)$$

[Again it is understood that the free parameters, λ_1 , λ_3 , λ_4 , and λ_5 , may be chosen differently for each solution β ; c may *not*, since it determines ν (1.61), understood to be fixed in the Hankel transform (1.93).] Thus $f(\kappa; t)$ is represented by a Hankel transform (1.93) determining via the corresponding self-inverse transform (for \mathcal{G}) the weighting function for solutions (1.70)-(1.71):

$$g(\beta; t) \equiv \sqrt{\beta} e^{\frac{\beta^2}{4I}} \int_0^\infty \sqrt{\kappa\beta} J_\nu(\kappa\beta) f(\kappa; t) d\kappa \quad (1.96)$$

Integrating with respect to this weight function the solutions (1.70)-(1.71) yields the following expression satisfying BC (1.12) with $A = TK^\alpha$ which, *multiplied by* $e^{ct(rT)^\nu} / [1 - e^{-r(T-t)}]^\nu$, gives a further solution to (the linear) pricing PDE (1.7), as is readily seen by direct substitution:

$$S^{\frac{1}{2} - \frac{r}{\sigma^2}} e^{ct} \int_0^\infty \frac{g(\beta; t)}{A + \frac{\lambda_4}{2\alpha\sigma^2\lambda_5}} J_\nu \left(S^{\frac{\alpha}{2}} \beta \frac{\lambda_4 + 2\alpha\sigma^2\lambda_5 TK^\alpha}{\lambda_4 + 2\alpha\sigma^2\lambda_5 A} \right) \cdot \exp \left[\left(IS^\alpha - \frac{\beta^2}{4I} \right) \frac{\lambda_4 + 2\alpha\sigma^2\lambda_5 TK^\alpha}{\lambda_4 + 2\alpha\sigma^2\lambda_5 A} \right] d\beta \quad (1.97)$$

Recall that the free parameters, λ_1 , λ_3 , λ_4 , and λ_5 , are required to satisfy constraints (1.84) and (1.94), so that one simple possibility is again to fix both λ_4 and λ_5 to be independent of β , so that I is also (instead of n as in the case of Fourier-Bessel series), and choose λ_1 and λ_3 such that constraint (1.94) is satisfied. Substituting formulas (1.95)-(1.96) into (1.97), and denoting $\Lambda \equiv \lambda_4 / 2\alpha\sigma^2\lambda_5$ and $h(A; \Lambda) \equiv (\Lambda + TK^\alpha)(\Lambda + A)^{-1}$,

$$h(A; \Lambda) \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} S \int_0^\infty \int_0^\infty \kappa\beta \left(\frac{\kappa}{S^{\frac{\alpha}{2}}} \right)^{\frac{1}{\alpha} + \frac{2r}{\alpha\sigma^2}} e^{\frac{\beta^2}{4I} [1 - h(A; \Lambda)] - I [\kappa^2 - S^\alpha h(A; \Lambda)]} J_\nu[\kappa\beta] J_\nu \left[S^{\frac{\alpha}{2}} h(A; \Lambda) \beta \right] d\kappa d\beta \quad (1.98)$$

Precisely in the same manner, and for precisely the same reasons, as in the case of Fourier-Bessel series solutions of pricing PDE (1.7), imposing only two constraints (1.84) and (1.94) in

all Bessel $J_\nu(z)$ -function solutions of corollary 1.2.11 but that corresponding to multiple (1.75) and the third row of tables 1.5 and 1.6, in which three are imposed (additionally requiring that μ_6 and so μ_7 be independent of the integration index β) and nine are left ‘free,’ the remaining parametric flexibility is only apparent, i.e., solutions of the form (1.97) necessarily follow and permit only four degrees of freedom, independent of the integration index β : The invariant $I \in \mathbb{C}$, the parameter $c \in \mathbb{C}$, chosen such that the corresponding parameter $\nu \geq -1/2$ of table 1.5, and the two parameters, $\lambda_4, \lambda_5 \in \mathbb{C}$.²⁸ In the case with three constraints imposed, the further degree of freedom, $\mu_7 \in \mathbb{C}$ independent of the integration index β , is also permitted.

Theorem 1.2.13 *Integrating with respect to weight function (1.96) the Bessel $J_\nu(z)$ -function solutions of corollary 1.2.11, the free parameters of whose arguments z must satisfy constraints (1.84) and (1.94), yields expression (1.97) which satisfies BC (1.12) with $A = TK^\alpha$. Division of this expression by function (1.90) yields a solution of pricing PDE (1.7). In the case of the third row of table 1.5, $\mu_7 \in \mathbb{C}$ (and so $\mu_6 \in \mathbb{C}$) is required to be independent of the integration index β . The parameters $c \in \mathbb{C}$ must be such that $\nu \geq -1/2$ of table 1.5.*

Similar comments to those following theorem 1.2.12 in the preceding section, apply to theorem 1.2.13, which are analogous nearly to the point of being verbatim, and so will not be explicated.

Kontorovich–Lebedev transform solutions of pricing PDE (1.7)

Unlike the undeveloped ‘Fourier-Whittaker’ discussed in the section on Fourier-Bessel series solutions of pricing PDE (1.7), some progress may be directly made via the Kontorovich–Lebedev transform. Recalling from the discussion following corollary 1.2.11, that solutions (1.68)-(1.69) follow as special cases of (1.72)-(1.73) by imposing formulas (1.76)-(1.77), consider

$$\underline{c} \equiv \begin{cases} \left(\frac{\sigma + \frac{t}{\sigma}}{2}\right)^2 & \text{Unconstrained} \\ \frac{\alpha^4 \sigma^4}{4} & \text{(1.54) constrained} \end{cases}$$

Take $(\forall \underline{c} \leq c) \nu_c$ given by formula (1.66) [(1.58)] in case constraint (1.54) is [un]imposed, and

$$\mu_{2,c} \equiv \begin{cases} \frac{b_6 \mu_6 - b_7 \mu_7}{\alpha} + \mu_4 + a_3 \mu_3 + a_5 \mu_5 + c (\mu_3 + \mu_5 + a_6 \mu_6 + a_7 \mu_7) & \text{Unconstrained} \\ \frac{b_6 \mu_6 - b_7 \mu_7}{\alpha} + \mu_4 - (a_3 - a_5) \mu_5 - a_3 (a_6 \mu_6 + a_7 \mu_7) & \text{(1.54) constrained} \end{cases} \quad (1.99)$$

It then follows that $0 = \kappa$ (1.65) [(1.63)] in case constraint (1.54) is [un]imposed, and $W_{0,\nu_c}(z) \equiv K_{\nu_c}(z/2) \sqrt{z/\pi}$ reduces to a MacDonald function, since $(\forall \underline{c} \leq c) \nu_c \equiv ik$ is imaginary. Consider

$$G(\kappa) \equiv \mathfrak{R}_\kappa[g(\beta)] \equiv \int_0^\infty g(\beta) K_{ik}(\beta) d\beta \quad (1.100)$$

²⁸The latter two equations of system (1.76) map between solutions of the form (1.97), corresponding either to the [optimal] Lie point symmetries (1.22) [(1.25)].

Yakubovich [318] defines this *Kontorovich–Lebedev transform*²⁹ on a weighted L^p -space denoted $L^p_\nu(0, \infty)$, with norm $\|\cdot\|_{L^p_\nu(0, \infty)} : L^p_\nu(0, \infty) \rightarrow [0, \infty)$ given by his formula (1.19):

$$g \mapsto \left(\int_0^\infty |\beta^\nu g(\beta)|^p \frac{d\beta}{\beta} \right)^{\frac{1}{p}}$$

In particular, given any mapping $g \in L^2_1(0, \infty)$, the Kontorovich–Lebedev transform (1.100) exists in $L^2(0, \infty)$ and has the following inverse with kernel ${}^{2\kappa \sinh(\pi\kappa)}/\pi^2$:

$$g(\beta) \equiv \frac{2}{\pi^2 \beta} \int_0^\infty G(\kappa) K_{ik}(\beta) \kappa \sinh(\pi\kappa) d\kappa \quad (1.101)$$

[See the Plancherel theorem 2.4 of Yakubovich [318]; the Parseval equality (2.48) also holds.] For the same reason as in the preceding sections on Fourier-Bessel series and Hankel transform solutions of pricing PDE (1.7), only BC (1.12) with $A = TK^\alpha$ is considered, setting

$$\beta \equiv \frac{2 \sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)} S^\alpha}{\alpha^2 \sigma^2 (T^2 K^{2\alpha} \mu_7 - 2TK^\alpha \mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7)} \equiv \beta_\mu S^\alpha \quad (1.102)$$

Solutions are sought such that the following holds:

$$\left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} S = V(S, TK^\alpha, t) \equiv S^{\frac{1-\alpha}{2} - \frac{r}{\sigma^2}} \int_0^\infty e^{c^* t + \frac{\mu_7 T K^\alpha - \mu_6}{\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \beta} G(\kappa) W_{0,ik}(2\beta) d\kappa \quad (1.103)$$

Here, $c^* = C(1.67)[c]$ in case constraint (1.54) is [un]imposed. Denote

$$I \equiv \frac{\mu_7 T K^\alpha - \mu_6}{\sqrt{\mu_6^2 + 2\mu_7(\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7)}} \quad (1.104)$$

(This expression is required to be independent of the index of integration κ .) Rewriting S [c^*] in terms of β (1.102) [κ] and recalling $W_{0,\nu_c}(z) \equiv K_{\nu_c}(z/2) \sqrt{z/\pi}$, formula (1.103) simplifies to

$$\begin{aligned} g(\beta; t) &\equiv \sqrt{\frac{2}{\pi^3 \beta^3}} \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} \left(\frac{\beta}{\beta_\mu} \right)^{\frac{1+\alpha}{2\alpha} + \frac{r}{\alpha\sigma^2}} e^{-I\beta - \frac{1}{2} \left(\frac{r}{\sigma} + \frac{\sigma}{2} \right)^2 t} = \frac{2}{\pi^2 \beta} \int_0^\infty \frac{e^{\frac{\alpha^2 \sigma^2}{2} \kappa^2 t} G(\kappa)}{\kappa \sinh(\pi\kappa)} K_{ik}(\beta) \kappa \sinh(\pi\kappa) d\kappa \\ &\equiv \frac{2}{\pi^2 \beta} \int_0^\infty \mathcal{G}(\kappa; t) K_{ik}(\beta) \kappa \sinh(\pi\kappa) d\kappa \end{aligned} \quad (1.105)$$

²⁹The K -transform is as the Kontorovich-Lebedev transform (1.100), but restricting ($\forall c \leq c$) $\nu_c \in \mathbb{C}$ to be such that $|\nu_c| \leq 1/2$, rather than to be purely imaginary; see chapter X of Bateman [30]. Inversion of the K -transform, however, is much more complicated than formula (1.101) for the Kontorovich-Lebedev transform, and so is not pursued presently; see Nasim [229].

Thus $g(\beta; t)$ is represented by an *inverse* Kontorovich-Lebedev transform (1.101) determining via the dual transform formula (1.100) [for $\mathcal{G}(\kappa; t)$] the weight function in expression (1.103):

$$G(\kappa; t) = \kappa \sinh(\pi\kappa) e^{-\frac{\alpha^2 \sigma^2}{2} \kappa^2 t} \int_0^\infty g(\beta; t) K_{i\kappa}(\beta) d\beta \quad (1.106)$$

Integrating with respect to weight function (1.106) *any* of the Whittaker W_{0, ν_c} -function solutions of corollary 1.2.11 yields the following expression satisfying BC (1.12) with $A = TK^\alpha$ which, *multiplied* by $e^{c^*t} / [1 - e^{-r(T-t)}]^\mu$, gives a further solution to (the linear) pricing PDE (1.7), as is readily seen by direct substitution:

$$S^{\frac{1-\alpha}{2} - \frac{r}{\sigma^2}} \int_0^\infty e^{c^*t + \frac{2(\mu_7 A - \mu_6) S^\alpha}{\alpha^2 \sigma^2 (A^2 \mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7)}} G(\kappa; t) W_{0, i\kappa} \left[\frac{4 \sqrt{\mu_6^2 + 2\mu_7 (\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7) S^\alpha}}{\alpha^2 \sigma^2 (A^2 \mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7)} \right] d\kappa \quad (1.107)$$

Denote $\mu \equiv \{\mu_1, \mu_4, \mu_5, \mu_6, \mu_7\}$ and

$$H(A; \mu) \equiv \frac{\mu_7 A - \mu_6}{\mu_7 T K^\alpha - \mu_6} \cdot \frac{T^2 K^{2\alpha} \mu_7 - 2TK^\alpha \mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7}{A^2 \mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7} \quad (1.108)$$

Substituting formulas (1.105)-(1.106) into (1.107),

$$\frac{2}{\pi^2} \left[\frac{1 - e^{-r(T-t)}}{rT} \right]^{\frac{1}{\alpha}} S \int_0^\infty \int_0^\infty \frac{\kappa \sinh(\pi\kappa)}{\beta} \left(\frac{\beta}{\beta_\mu S^\alpha} \right)^{\frac{1+\alpha}{2\alpha} + \frac{r}{\alpha\sigma^2}} e^{[H(A; \mu) - 1] \beta} W_{0, i\kappa}[2\beta] \cdot W_{0, i\kappa} \left[\frac{4 \sqrt{\mu_6^2 + 2\mu_7 (\mu_1 + 2\mu_4 + \mu_5 + 2\mu_7) S^\alpha}}{\alpha^2 \sigma^2 (A^2 \mu_7 - 2A\mu_6 - 2\mu_1 - 4\mu_4 - 2\mu_5 - 4\mu_7)} \right] d\beta d\kappa \quad (1.109)$$

Theorem 1.2.14 *Integrating with respect to weight function (1.106) the Whittaker $W_{0, \nu_c}(z)$ -function solutions of corollary 1.2.11, yields expression (1.107) which satisfies BC (1.12) with $A = TK^\alpha$. Division of this expression's integrand by function (1.90) yields a solution of pricing PDE (1.7). The parameters μ_1, μ_2, μ_3 and γ must satisfy constraints (1.99), (1.102) and (1.104).*

In this case, the term e^{c^*t} cannot be extricated from integration in formula (1.107), as $c^* = C$ (1.67) [c] depends on $\kappa \equiv -i\nu_c$ through formula (1.66) [(1.58)] in case constraint (1.54) is [un]imposed. This results in the required multiplication *within* the integrand of formula (1.107) to yield a solution to pricing PDE (1.7), whereas such multiplication was without the integrand in each of theorems 1.2.12 and 1.2.13; in this sense theorem 1.2.14 yields a comparatively weak result. As in the case following theorem 1.2.13, similar comments to those following theorem 1.2.12 apply to theorem 1.2.14, which will not be expounded.

However, some progress *can* be made due to the added parametric flexibility of the five-dimensional μ , as opposed to the one-dimensional Λ . Consider the following quantities:

$$\begin{aligned} m &\equiv \mu_1 + 2\mu_4 + \mu_5 \\ M &\equiv \{\mu_6, \mu_7, m\} \\ q(M) &\equiv \sqrt{\mu_6^2 + 2m\mu_7 + 4\mu_7^2} \geq 0 \\ p_1(x; M) &\equiv \mu_7 x - \mu_6 \\ p_2(x; M) &\equiv \mu_7 x^2 - 2\mu_6 x - 2m - 4\mu_7 \end{aligned}$$

In terms of these quantities, the expressions

$$\begin{aligned} \beta_\mu &\equiv \frac{2}{\alpha^2 \sigma^2} \frac{q(M)}{p_2(TK^\alpha; M)} \\ I &\equiv \frac{p_1(TK^\alpha; M)}{q(M)} \\ H(A; M) \equiv H(A; \mu) &\equiv \frac{p_1(A; M) p_2(TK^\alpha; M)}{p_1(TK^\alpha; M) p_2(A; M)} \end{aligned}$$

To yield real values, the argument of the second Whittaker $W_{0, \nu_c}(z)$ -function in formula (1.109) must be nonnegative; see figure 1.4. This argument being a nonnegative number divided by $p_2(A; M)$, so too must the latter be nonnegative. Thus the domain $A \in (0, TK^\alpha)$ *must* contain neither pole of $H(A; M)$, say, $A_\pm \equiv \mu_7^{-1}[\mu_6 \pm q(M)]$. In case $\mu_7 < 0$ (> 0), there is (are) one (two) case(s); $A_+ \leq 0 < TK^\alpha \leq A_-$ ($A_\pm \leq 0$ and $A_\pm \geq TK^\alpha$). In the first (second) case with $\mu_7 > 0$, $I > 0$ (< 0) and $[\forall A \in (0, TK^\alpha)] H(A; M) > 1$ (< 1), since $H(TK^\alpha; M) = 1$, $H(A; M) \rightarrow 0^+$ as $A \rightarrow (-)\infty$, $H(A; M)$ has only two stationary *nonreal* points $\mu_7^{-1}[\mu_6 \pm iq(M)]$, and $A > 0 \geq \mu_6/\mu_7$ ($\geq TK^\alpha > A$), the only zero of $H(A; M)$; so that $e^{[H(A; \mu)-1]I\beta} \rightarrow \infty$ as $\beta \rightarrow \infty$ in formula (1.109). In case $\mu_7 < 0$, the condition $A_+ \leq 0 < TK^\alpha \leq A_-$ implies that $-2q(M) \leq \mu_7 TK^\alpha$. Two cases arise according to whether $\mu_6 \leq \mu_7 TK^\alpha$ in which, respectively, $I \geq 0$ and $[\forall A \in (0, TK^\alpha)] H(A; m) \geq 1$ so that, in either case, $e^{[H(A; \mu)-1]I\beta} \rightarrow \infty$ as $\beta \rightarrow \infty$ in formula (1.109). In case $\mu_6 = \mu_7 TK^\alpha$, $I = 0$ and the corresponding exponential term is identically one. In case $\mu_7 = 0$, the following simplified expressions result, denoting $P_1(x; \mu_6, m) \equiv \mu_6 x + m$:

$$\begin{aligned} \beta_\mu &\equiv -\frac{|\mu_6|}{\alpha^2 \sigma^2 P_1(TK^\alpha; \mu_6, m)} \\ I &\equiv -\frac{\mu_6}{|\mu_6|} \\ H(A; \mu_6, m) &\equiv \frac{P_1(TK^\alpha; \mu_6, m)}{P_1(A; \mu_6, m)} \end{aligned}$$

The case $\mu_6 = 0$ removes all dependence on A and so is ignored. In case $\mu_6 < 0$ (> 0), $I = (-)1$ and $m(+\mu_6 TK^\alpha) \leq 0$ in order that $\beta_\mu \geq 0$, as well as the argument of the second Whittaker $W_{0, \nu_c}(z)$ -function in formula (1.109), and $[\forall A \in (0, TK^\alpha)] H(A; \mu_6, m) > 1$ (< 1) so that, in either case, $e^{[H(A; \mu_6, m)-1]I\beta} \rightarrow \infty$ as $\beta \rightarrow \infty$ in formula (1.109). These results are summarized.

Corollary 1.2.15 *In case $\mu_7 \geq 0$ or $\mu_7 < 0$ and $\mu_6 \neq \mu_7 TK^\alpha$, $e^{[H(A; \mu)-1]I\beta} \rightarrow \infty$ as $\beta \rightarrow \infty$ in formula (1.109); in case $\mu_6 = \mu_7 TK^\alpha < 0$, the latter expression is identically one.*

Thus unfortunately in no case does $e^{[H(A;\mu)-1]I\beta} \rightarrow 0$ as $\beta \rightarrow \infty$ in formula (1.109), but the particular case $\mu_6 = \mu_7 TK^\alpha < 0$ does *not* result in $e^{[H(A;\mu)-1]I\beta} \rightarrow \infty$ as $\beta \rightarrow \infty$. Though the Whittaker $W_{0,\nu_c}(z)$ -function arising in theorem 1.2.14 and plotted in figure 1.4 apparently decays rapidly in $x \equiv 2\beta$, it appears yet difficult to directly numerically evaluate expression (1.109) without referring to the (particularly, asymptotic) theory of this function; see Buchholz [51]. As such, that task is left for another study.

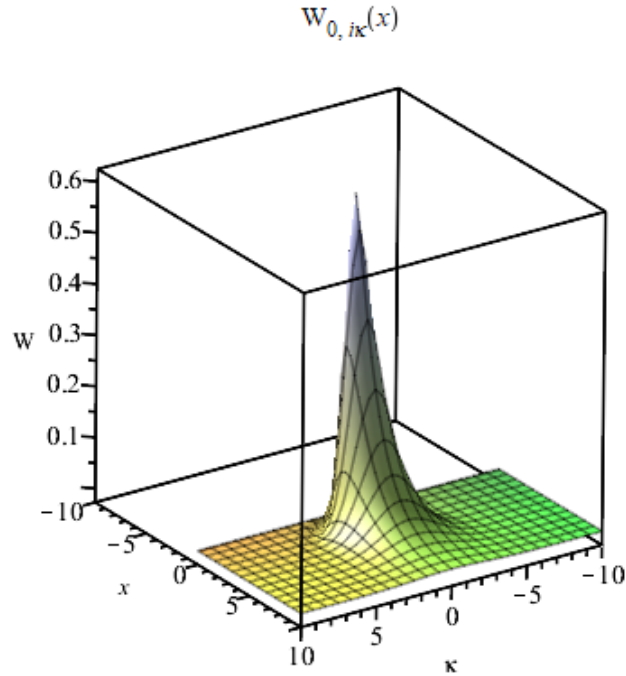


Figure 1.4: Behavior of the Whittaker $W_{0,\nu_c}(z)$ -function of theorem 1.2.14

Similarly, several less popular/well-studied integral transforms involving Bessel or Whittaker functions could be implemented, perhaps yielding more numerically straightforward pricing formulas, as well as satisfying more or other BCs than (1.12) with $A = TK^\alpha$. This too may be worthwhile/interesting for future study. Referring parenthetically to sections of Brychkov and Prudnikov [49], such transforms involving Whittaker functions include those of Varma (7.1) and Meijer (7.2), and the generalized transforms of Whittaker (7.3) and Stieltjes (7.4). (Transforms involving Whittaker functions potentially allow to combine the most general [optimal] invariant solutions determined, satisfying the reduced pricing PDE (1.30) [(1.51)].) In addition to the well-known Hankel and Kontorovich-Lebedev transforms addressed in the preceding two subsections, a number of others are detailed in chapter 6 of Brychkov and Prudnikov [49], and may also potentially be used to effectively combine the Bessel function separation of variables solutions of corollary 1.2.11. In particular, Lo and Hui [206] applied the Weber transform (see section 6.10 of Brychkov and Prudnikov [49]) to price barrier options, as well as Fourier-Bessel series expansions for the same purpose.

1.2.2 Novel numerical problems and experiments

One interesting possibility is to take formula (1.43) as an ansatz, and solve for $\lambda \equiv \lambda(S, A, t) \geq 0$ ³⁰ such that the pricing PDE (1.7) is satisfied. But perhaps a more pressing issue is as follows:

Problem 1 Denoting $\lambda \equiv \lambda(S, A, t; r, \sigma, T, K)$ and $\mathbb{E}_{S_T} V(S, A, t) \equiv \bar{V}(S, A, t; \lambda)$, solve

$$\min_{\lambda \geq 0} \|\bar{V}(S, A, t; \lambda) - \tilde{V}(S, A, t)\|$$

Here, $\|\cdot\|$ denotes a norm, e.g., the sup norm $\|\cdot\|_\infty$, and \tilde{V} an ‘exact’ reference value for \bar{V} , e.g., MC estimates via Longstaff and Schwartz [210], respecting acceptable confidence bounds.

One way to address this function estimation problem is via a nonparametric or machine learning regression model, e.g., a support vector machine, gradient-boosted regression tree, random forest, or neural network. The latter three models are discussed in section 2.2, and an authoritative reference for all four and other such techniques is provided by Hastie, Tibshirani, and Friedman [157]. Training, validation, and/or testing sets can be generated via \tilde{V} for some experimental design on the variables, S, A, t , and parameters, σ, r, T , and K , upon the latter of which V and \tilde{V} implicitly depend. One goal of such an approach is to train a surrogate model, \bar{V} , for the ‘exact’ model, \tilde{V} , given a comparatively small number of evaluations of the latter, much more computationally expensive, model. This approach is first pursued in the following numerical exercises, as a demonstration of some practical capabilities and limitations of formula (1.43) in pricing arithmetic Asian options. The following related problem is addressed:

Problem 2 Given a solution $\tilde{\lambda} \equiv \tilde{\lambda}(S, A, t; \sigma, r, T, K)$ to and notation of problem 1, solve

$$\min_{\lambda \geq 0} \|\lambda - \tilde{\lambda}\|$$

That is, for some experimental design on the variables, S, A, t , and the parameters, r, σ, T , and K , problem 1 is solved pointwise, i.e., individually for each scalar $\tilde{\lambda}(S, A, t; \sigma, r, T, K) \geq 0$. The estimated function, say $\hat{\lambda}$, for problem 2, is then substituted into \bar{V} to generate an estimate for \tilde{V} , namely, $\bar{V}(S, A, t; \hat{\lambda})$. Fortunately, to this end the `Matlab Financial Instruments Toolbox` has built-in methods for arithmetic Asian option pricing by the approaches of Cox, Ross, and Rubinstein [85], Kemna and Vorst [177], Haug [158], Levy [191], Turnbull and Wakeman [303], and Longstaff and Schwartz [210], the latter of which is used to generate $\tilde{\lambda}$ values with which to determine $\hat{\lambda}$ and assess estimates \bar{V} of the MC-simulated \tilde{V} values.

1.2.3 Results for $t = 0$ ($\Rightarrow A = t\mathcal{A}^\alpha = 0$)

The pricing of a fixed-strike, Asian call option is considered at the contract’s inception, i.e., when $t = 0$, which implies that $A = t\mathcal{A}^\alpha = 0$. In the case that $K = 0$, it is known that³¹

$$V(S_0, 0, 0) = S_0 e^{-rT} \frac{e^{rT} - 1}{rT} \quad (1.110)$$

³⁰The argument of \ln in ζ from theorem 1.2.5 must be nonnegative in order to avoid non-real values for ζ .

³¹See, e.g., equation (4.16) of Klassen [179].

[This also follows from formula (1.45) with $K = 0$.] The resulting experiments occupy much of the remaining (methods) section 1.2, as the surrogate model, $\bar{V}(S_0, 0, 0; \sigma, r, T, K; \lambda)$, is constructed over the large variable and parameter space (1.111). The analysis in case $t > 0$, described at the end of subsection 1.2.1, is carried out in subsection 1.2.4, and is comparatively (and advantageously) short, as only the variables, S and t , and parameters, σ , r , T , and K , of interest, need be considered. But this complements the analysis in case $t = 0 \Rightarrow A = 0$, since the results of subsection 1.2.4 are strongest near-the-money, i.e., when $0 \ll A \leq KT$. For case $t = 0$, exploratory analyses of the behavior of the (implied) function(s) $\tilde{\lambda}(S_0, 0, 0; \sigma, r, T, K)$, and solutions to the following problem 3, related to problem 2, are first developed.

Preliminary studies of the function $\lambda(S_0, 0, 0; \sigma, r, T, K)$

With $K = 0$, departures of each parameter from the values $r = \sigma = 0.1$, $T = 0.25$, and $S_0 = 100$, are considered in figures 1.5a through 1.5d. Values computed from equation (1.43) [(1.110)] are shown in dark [light] blue for values of $\lambda \in (0, 100]$. It is clear that there are two families of solutions, $\lambda(\sigma)$ and $\lambda(S_0)$, for all (but the smallest) values of S_0 (σ) considered. Similarly, there are two families of solutions, $\lambda(r)$ and $\lambda(T)$, for the considered values of r (T) up to about 0.3 (1.5).

Figure 1.6 presents departures from the same parameter values for positive K , with the ‘true’ values being computed from the `Matlab` built-in implementation of MC Asian option pricing à la Longstaff and Schwartz [210]. Again, there are two families of solutions, $\lambda(K)$, for up to slightly-out-of-the-money calls, when the true price vanishes. But as figure 1.6 illustrates, at this point all sufficiently small λ are solutions, i.e., for some $\bar{K} > 100$, for each $K \geq \bar{K}$ there is some $\lambda_K \geq 0$, such that all $\lambda \in [0, \lambda_K]$ are solutions.

Though exact solutions are only known in the case $K = 0$, that of at-the-money options (with $K = 100$) is a relevant reference point, and so figures 1.7a through 1.7d present analogous relations to 1.5a through 1.5d, using the same `Matlab` built-in MC Asian option pricing method for the ‘true’ reference values. The new set is clearly similar to the old, and the same comments apply regarding paired families of solutions, $\lambda(\cdot)$, subject to similar parametric restrictions.

Design of experiments (1.111) and solutions of problem 3

The preliminary investigations suggest that problems 1 and 2 may be solved exactly for parameters not ‘significantly’ deviant from the values $r = \sigma = 0.1$, $T = 0.25$, and $K = S_0 = 100$, as discussed. These parameters comfortably include most considered in the papers, e.g., of Kemna and Vorst [177], Klassen [179], Li and Chen [192], and Turnbull and Wakeman [303]. To assess the ability of formula (1.43) to price arithmetic Asian options, the following full-factorial experimental design was used to generate training and testing data:

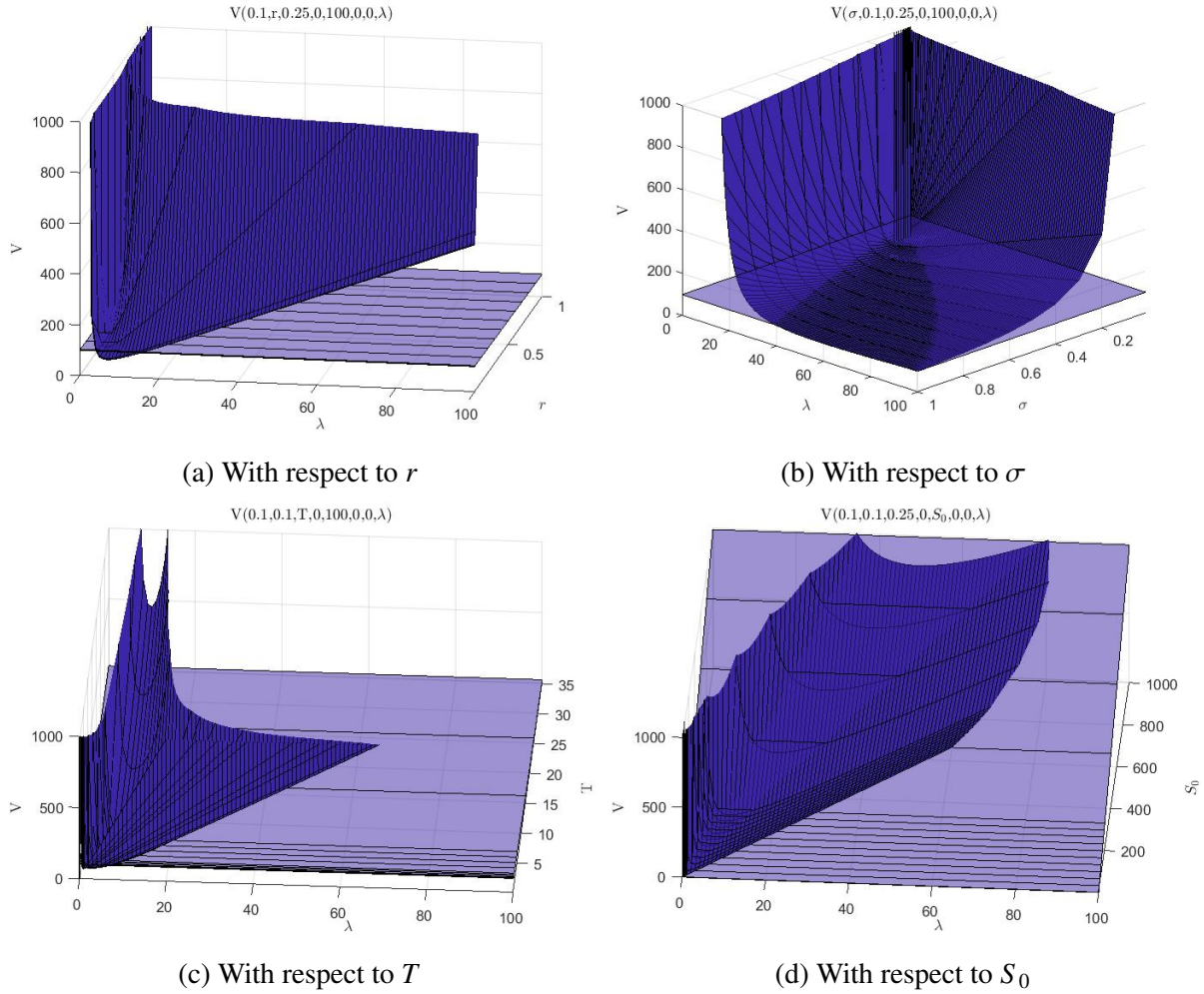


Figure 1.5: Behavior of the function $V(\sigma, r, T, 0, S_0; 0, 0, \lambda)$

$$\begin{aligned}
 \sigma & \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\} \\
 r & \{10^{-4}, 10^{-3}, 10^{-2}, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 1\} \\
 T & \in \{1\text{-}5 \text{ days}, 2\text{-}4 \text{ weeks}, 2\text{-}3 \text{ months}, 2\text{-}4 \text{ quarters}, 2\text{-}5 \text{ years}, \text{ and } 15, 25, \text{ and } 35 \text{ years}\} \\
 K & \{0, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 400, 600, 800, 1000, 2500, 5000, 7500, 10000\} \\
 S_0 & \{10^{-3}, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 400, 600, 800, 1000\}
 \end{aligned}
 \tag{1.111}$$

Note that T is converted to years, with 252 (trading/business/working) days per year, 63 days per quarter, 21 days per month, and five days per week. For each of the 627000 resulting experiments, the following is addressed in the notation of problem 1:

Problem 3 Given $\sigma, r, T, K,$ and S_0 , solve subject to $\lambda \geq 0$ and $\lambda_0 = 10^4$,

$$\bar{V}(S_0, 0, 0; \lambda) = \tilde{V}(S_0, 0, 0)$$

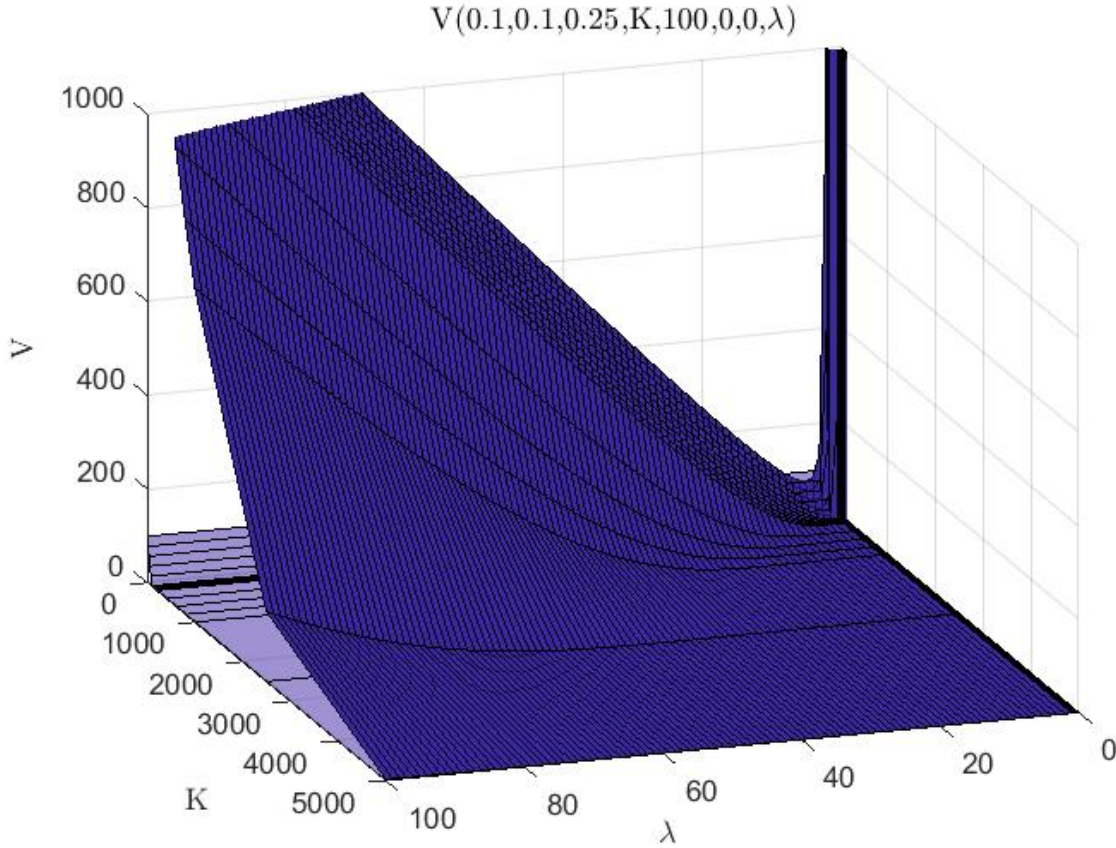
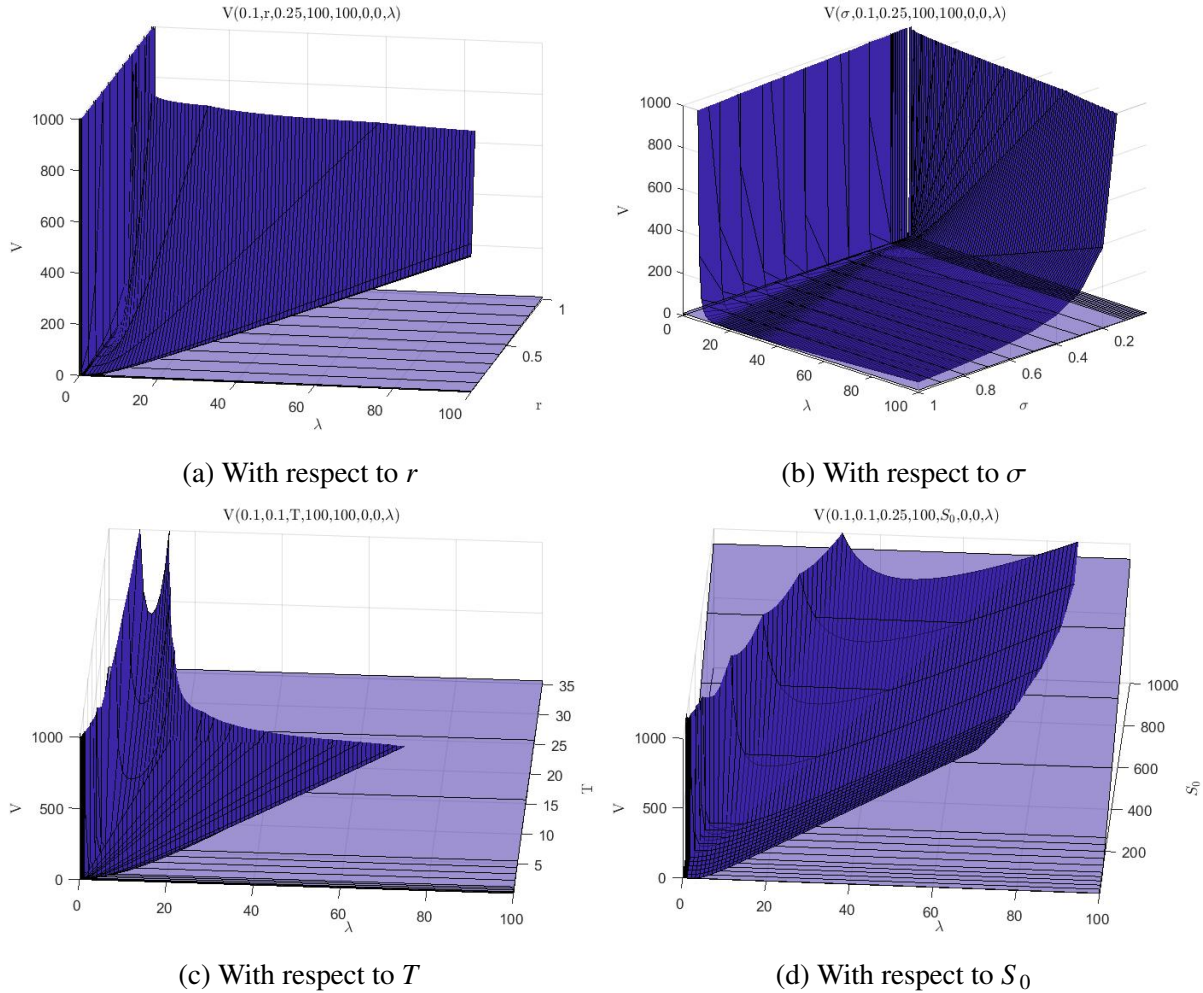


Figure 1.6: Behavior of the function $V(0.1, 0.1, 0.25, K, 100; 0, 0, \lambda)$ with respect to K

If a solution is not found, solve the relaxed problem, subject to the same initial condition,

$$\min_{\lambda \geq 0} |\bar{V}(S_0, 0, 0; \lambda) - \tilde{V}(S_0, 0, 0)|$$

The initial condition, $\lambda_0 = 10^4$, was chosen to encourage convergence to the larger of the two solutions λ as applicable and discussed above. Numerical solution of this problem is computationally difficult, largely due to evaluation of the lognormal MGF in equation (1.43): This is best done symbolically using variable-precision arithmetic (VPA), which cannot however presently be economized using the Matlab compiler, for example. Without compilation, evaluation times and parallelization complications are prohibitive for running 627000 experiments, even across tens of processors and hundreds of gigabytes' (GB) memory. So the regular, non-symbolic `integral()` method was used in Matlab, instead of the far-superior `vpaintegral()`. Compiling, all 627000 experiments completed in under two hours, parallelizing where possible and using 48 processors and 186 GB memory, whereas using `vpaintegral()` without compilation would take weeks and be subject to a number of possible failures/outages/interruptions/etc. Still, 440882 of the 627000 experiments, or about 70%, were solved exactly using `fsolve()`. Of these, 9546 were dropped for having solution at the initial condition, $\lambda_0 = 10^4$: As exemplified in figure 1.6, these most likely are cases in which there are 'interval' solutions, rather than two distinct families, as observed elsewhere. Including these

Figure 1.7: Behavior of the function $V(\sigma, r, T, 100, S_0; 0, 0, \lambda)$

‘initial solutions,’ without further knowledge of their containing intervals which are also solutions, may unnecessarily introduce discontinuities and/or inaccuracies into functional estimates of $\lambda(S_0, 0, 0; \sigma, r, T, K)$. This could be addressed, e.g., by lowering λ_0 until the initial condition is not a solution, indicating where one of the two distinct families merges with the interval solution, but as these cases represent only about 2% of the 440882 solved ‘exactly,’ they were simply dropped. Of the 186118 cases so unsolved, 176903 were solved using `fmincon()`, only 5662 of which were as accurate as the ‘exactly’ solved ones (up to \$0.04 resulting absolute deviation from the ‘true’ prices³²). Again, as these and the 9215 cases unsolved using either method together constitute only about 3% of the exactly solved experiments, both collections were simply dropped. This leaves 171241 of the 627000 experiments, or about 27%, which were solved inexactly using `fmincon()`, i.e., without the \$0.04 tolerance. The 431336 exactly solved cases without solution λ_0 were taken for training/testing data.

³²This allowance is likely not tolerable in practice, but was accepted in light of the numerical example’s illustrative purpose and the increased computational burden of a tighter tolerance. But this could be done, in particular to also improve the pursuant training/testing accuracies of the functional estimates of $\lambda(S_0, 0, 0; \sigma, r, T, K)$.

To further explore the parameter values across these five subsets of the 627000 experiments, parallel coordinate plots for each are shown in figures 1.8a through 1.10a. For reference, all experiments are included (undistinguished) in figure 1.10b. It is notable that, with the plot settings (line width and transparency) used, figures 1.8b and 1.10b are indistinguishable: This suggests that, in some sense, the 171241 (27%), solved inexactly without the \$0.04 tolerance, are ‘representative’ of the 627000 experiments. Some of these may be due to the difficulties of computing the lognormal MGF discussed above, but others may simply not be exactly solvable, due to the approximate nature of formula (1.43). Examples of this are apparent in figures 1.5c and 1.7c, where the two surfaces are disjoint for all $\lambda \geq 0$ and sufficiently large T , typically between 1-2 years.³³ This is consistent both with the notable absence of parameter combinations including both large r and T values in figure 1.8a and the worsening of the approximation by formula (1.43) for larger T values. By similar reasoning, it is sensible that figures 1.8a and 1.9b are much the same. Regarding the exactly solved cases with optimal λ_0 , it is clear from figure 1.9a that these mostly involve smaller σ and T values. As both of these factors reduce the option’s risk premium, the option price will fall to zero and therefore be insensitive to λ_0 for sufficiently large K , as seen in figure 1.6. It is also apparent that figure 1.10a is restricted to smaller values of σ , and so σ^2 , which is consistent with not being able to accurately evaluate the lognormal MGF without VPA, and therefore to solve problem 3, despite the theoretically improved approximation by formula (1.43) for smaller σ values. Note also that there is not much difference across the parallel coordinates plots in terms of K and S_0 , indicating little influence of those parameters on the (ease of) solubility of problem 3, in agreement with their lack of influence on the quality of approximation by formula (1.43). Finally, note that the training/testing data with parameter tuples in figure 1.8a comfortably include most of those analyzed in the papers noted at the end of section 1.2.2.

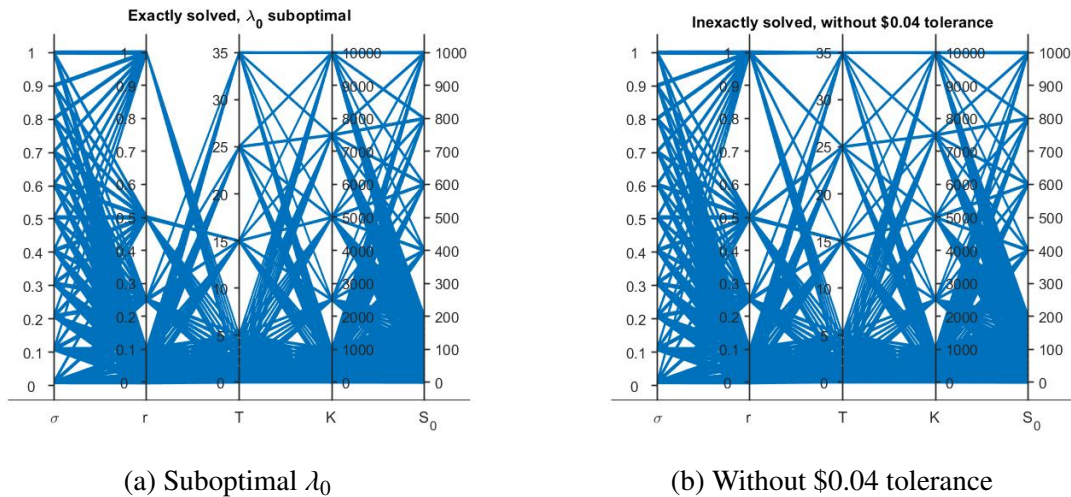


Figure 1.8: Experiments with suboptimal λ_0 and without \$0.04 tolerance

³³It is noted in the paragraph preceding section 14.2.1 of Wilmott, Dewynne, and Howison [315], that options typically mature within nine months, i.e., satisfy $T \lesssim 0.75$.

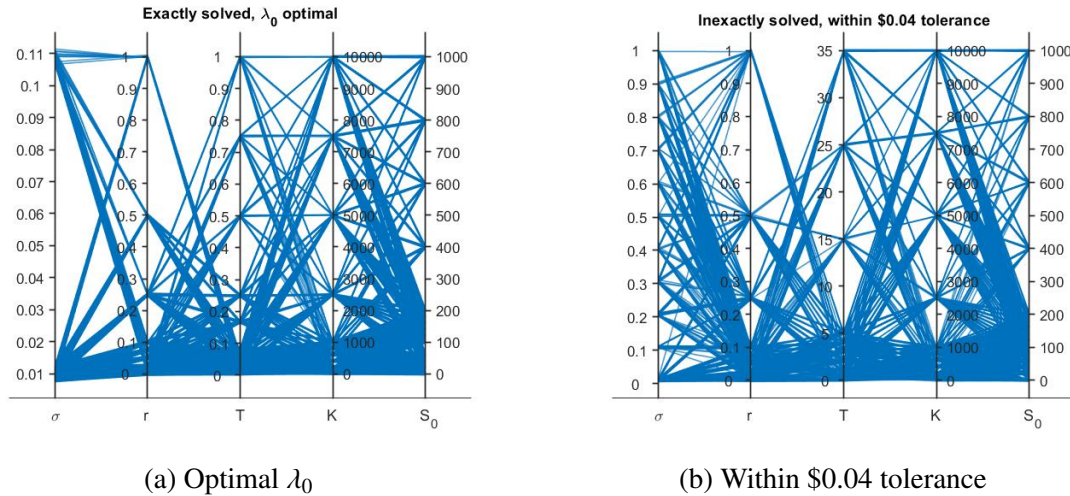
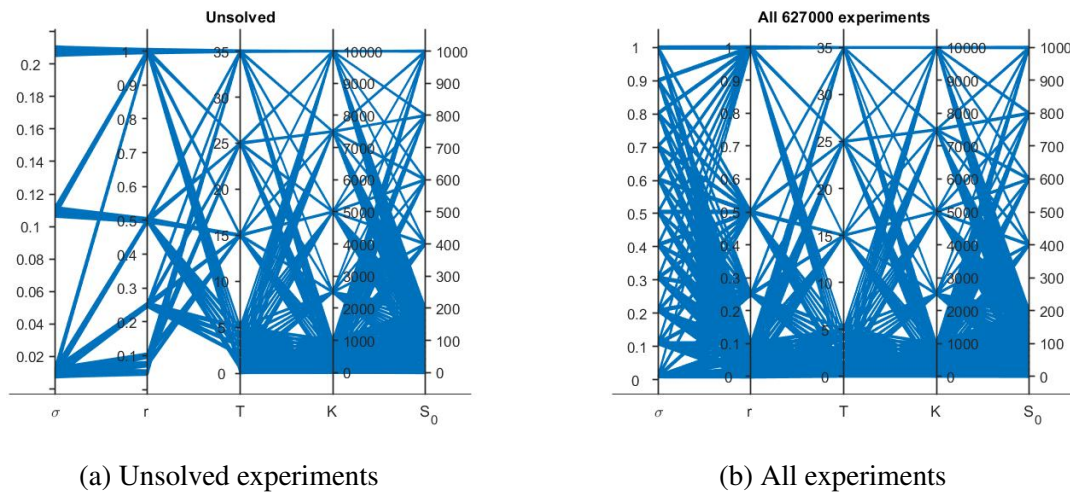
Figure 1.9: Experiments with optimal λ_0 and within \$0.04 tolerance

Figure 1.10: All experiments, and those unsolved

Predicting the function $\lambda(S_0, 0, 0; \sigma, r, T, K)$ in a general variable/parameter space (1.111)

Step-wise generalized linear models (GLMs) were fitted to the generated training data, for a variety of distributions [Gamma, (inverse) Gaussian, and Poisson] and sensible corresponding link functions [identity, logarithm, logit, probit, (complementary) double logarithm, and (squared) inverse], but the resulting mean absolute error (MAE) values were at best two orders of magnitude, whereas the Gaussian process regressions³⁴ (GPRs) achieve near-zero training MAE values (as specified further below), with hyperparameters tuned via Bayesian optimization (see, e.g., Snoek, Larochelle, and Adams [286]). Similar results were observed when constraining the predicted λ values to be nonnegative, which the GPRs do not do, but mostly respect for the testing data anyway (also specified further below). One reason for choosing GPRs for the following numerical demonstrations, not only over GLMs, but other alternatives

³⁴Schober et al. [267] and Raissi et al. [254], e.g., have directly solved differential equations using GPRs.

including various additional machine learning techniques listed following problem 1, is that they have only a handful of hyperparameters, the tuning of which by Bayesian optimization is itself described in terms of GPs. And despite the few number of hyperparameters (considering only, e.g., standard basis and kernel functions) and straightforward tuning of such, the resulting models are highly nonlinear and so are able to capture much of the observed variation in the training and testing data, including a low prediction rate of nonpositive λ values. A cubic spline interpolation, although it fits any training data ‘perfectly’ (with purely numerical errors), and yields a (very intricate) twice-differentiable approximation that may be substituted into the PDE (1.7), not surprisingly hardly generalizes to testing data: Training on half of the data with odd (sorted) maturity ranks, and testing on the other half, the associated testing MAE value was 1244.5, and systematic prediction of large negative λ values was observed in some cases. Other varieties of splines and wavelet techniques, best suited to curve- or surface-fitting, were not considered for the function λ of five arguments.

Authoritative references for GLMs and GPRs are respectively McCullagh and Nelder [219] and Williams and Rasmussen [314]. In an initial round of numerical experiments, the 431336 exactly solved cases without solution λ_0 were taken for training/testing data, as noted earlier. These were pseudorandomly partitioned into pairs consisting of 1%, 10%, 25%, and 50% training data. GPRs were fit in each case, using expected improvement acquisition functions modified to avoid any local over-exploitation of hyperparameter space, and re-partitioning of and 25% holdout from the training data for validation/to avoid over-fitting. Each regression was given between 21 and 28 days’ computation time, using 48 CPUs and 186 GB memory.

Some results of these numerical experiments are shown in figures 1.11-1.14. Naturally, the number of hyperparameter optimization iterations completed in the available computational time decreases with the number of training observations, apparently by somewhat more than a factor of proportionality, as shown in the left-hand column of figures 1.11-1.14. Each of these plots are similar, and the minimal observed and predicted losses naturally decrease with the number of training observations, with the exception of the bottom figure, in which only 30 hyperparameter optimization iterations were completed in the available computational time. Shown in the middle column are plots of predicted versus true λ values on the training data: These are clearly very good for the first three rows, although do naturally worsen with the number of training observations, which culminates in a notable loss in training accuracy on the bottom row. This is because the GPR is increasingly constrained, having to match (as best as possible) evermore training observations. Conversely, the predictions shown in the third column for the testing data, gradually improve with the number of training observations. This is natural, because the size of the testing set shrinks both absolutely (meaning fewer possible observations to test) and relative to the training set: The GPR is performed on increasingly many training observations, allowing a more thorough exploration of the predictor space, and also tested on fewer observations, which culminates in the performance on both sets being visually similar in the case on the bottom row of a 50-50 training/testing split. It is clear from the initial graphical analysis that the GPRs are able to faithfully mimic some features of the λ functions. But as observed also in earlier figures, the response of V to λ is quite pronounced, so small inaccuracies in estimating λ may not carry over to the same in estimating V . Thus, in addition to quantifying some of the graphical observations for estimating λ , it remains to

investigate how V itself is estimated, that being the ultimate goal of the numerical exercises.

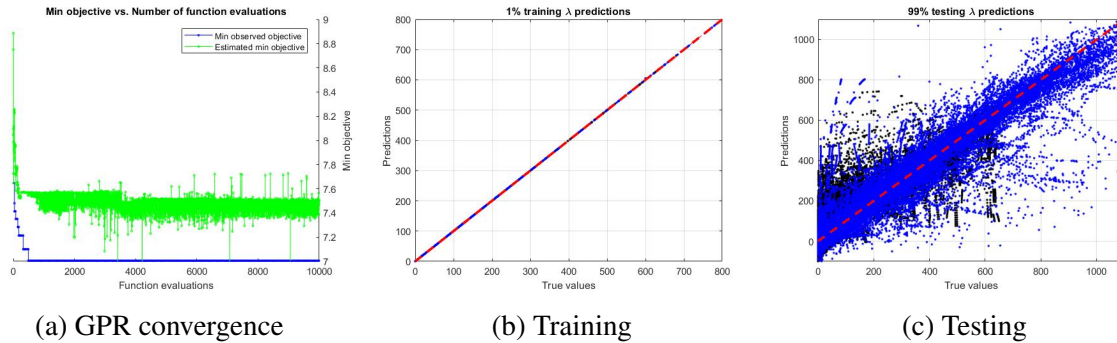


Figure 1.11: Quality of λ predictions, training on 1% of a general parameter space

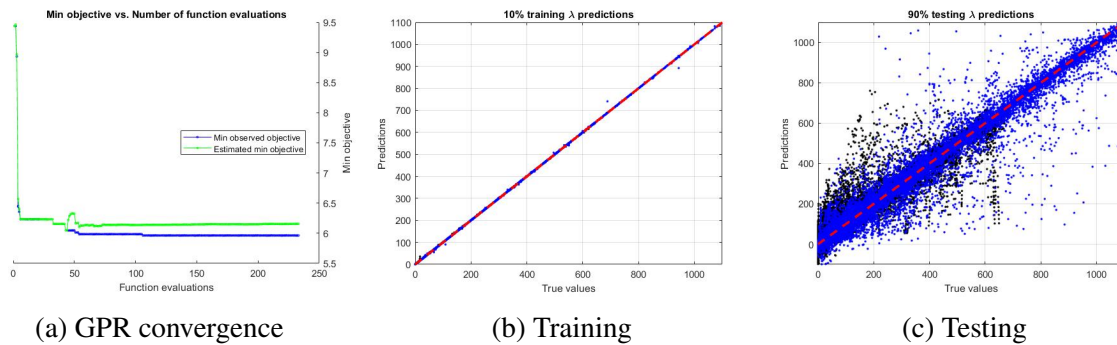


Figure 1.12: Quality of λ predictions, training on 10% of a general parameter space

Tabulated as follows are a number of results concerning the estimates of λ ($\tilde{\lambda}$) and V (\tilde{V}):

These numbers are mostly reflective of the visual observations from the figures noted above: MAE values, for both $\tilde{\lambda}$ and \tilde{V} , increase (decrease) with the training percentage on the training (testing) data. Similarly, negative $\tilde{\lambda}$ values (failed \tilde{V} evaluations) increase (decrease) with the training percentage on the training (testing) data, remaining (very) small, on the order of (a percent of) a percent.³⁵ The low prediction rate of negative λ values, noted above, validates the use of GPRs over alternate methods which permit enforcement of this constraint, and the near-zero rate of failure on predicting prices V , indicate a wide applicability of the preceding methodology. As argued earlier, it may prove a valuable preliminary tool for exploring prices over wide parameter intervals, prior to accurate estimation for specific parameter values chosen from the preliminary analysis, e.g., to optimize price values in some manner. While training the model first requires expensive MC simulations, or generation of accurate training/testing data

³⁵Aside from these two degenerate cases, in computing the MAE values for $\tilde{\lambda}$ and \tilde{V} , another class was excluded: All price predictions which exceeded by more than 2% the largest ‘true’ price in the entire data set. This tolerance was chosen, since the introduction to section 14.3 of Wilmott, Dewynne, and Howison [315] notes that the error of assuming a deterministic interest rate is typically 2%. Included in table 1.7, the incidence rates of this third case are generally low/inconsequential.

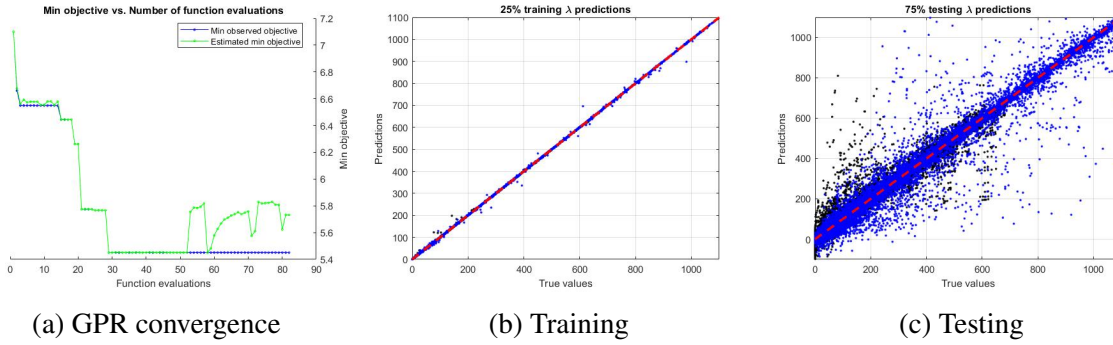


Figure 1.13: Quality of λ predictions, training on 25% of a general parameter space

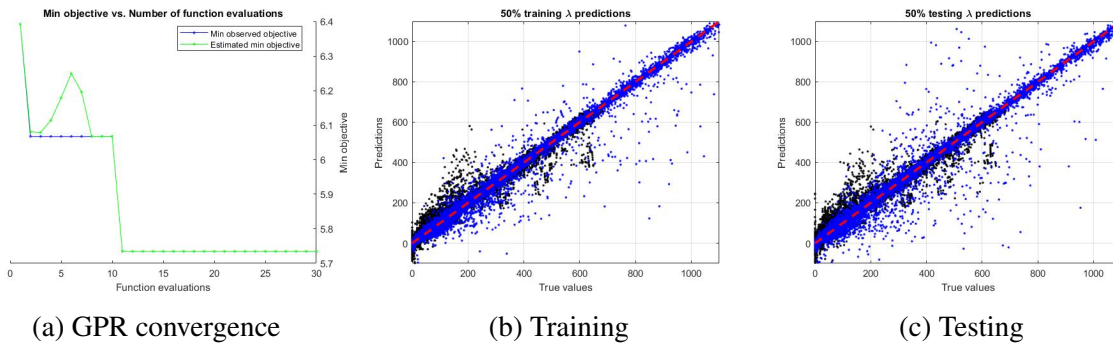


Figure 1.14: Quality of λ predictions, training on 50% of a general parameter space

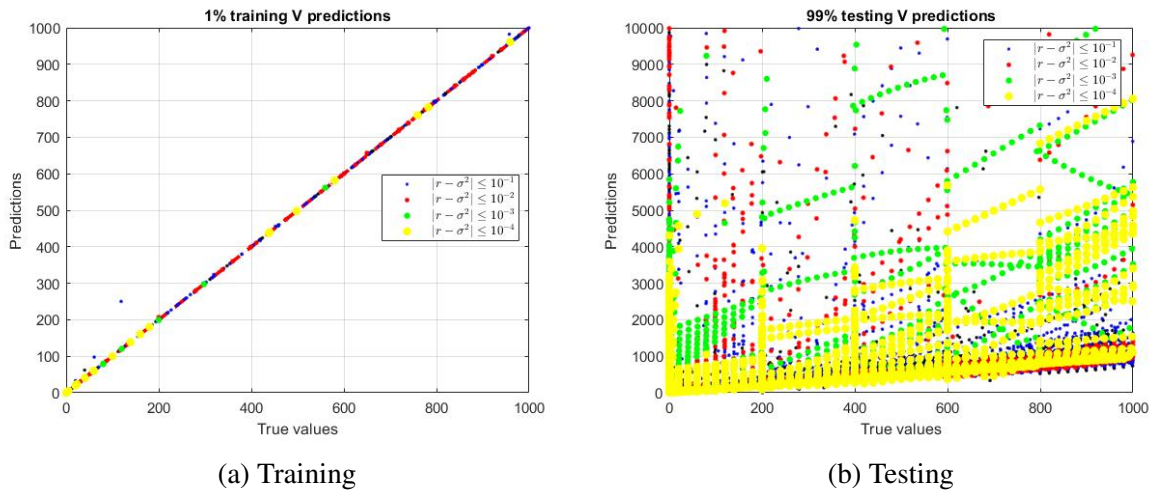
via other approaches, it can rapidly (approximately) price options for parameter combinations not ‘too’ without the parameter space of this data, evaluating hundreds of thousands of options in seconds or minutes. So a relatively small time can be spent generating training/testing data and building the model, which provides (limited) insight into novel option parameters, which can in turn be used to narrow the parameter values of interest down to a smaller range for further valuation via more accurate, but expensive methods. Also, it may well be that a more ‘targeted’ model built on a smaller parameter space will perform better as a direct pricing tool, but the present goal is to explore a reasonably large parameter space, partly as a means of qualitatively understanding the behavior of formula (1.43). And of course, the various machine learning techniques listed following problem 1 may also outperform the GPRs and prove more practical direct pricing tools, but there are many alternative models including these and others, whose hyperparameter optimization is, at least in the case of neural networks, much more nuanced and difficult than for GPRs, as mentioned earlier. An exploration along these lines could be productive, but may be best left for another study. Similarly, consideration of basis and kernel functions aside from the ‘standard’ built-ins of `Matlab`, could improve the present results, but would also be a much larger effort best addressed separately.

To understand better the nature of the numerical approximation achieved above, plots analogous to the two right-hand columns of figures 1.11-1.14, but for \tilde{V} , are provided below.

Shown in blue in the two right-hand columns of figures 1.11-1.14 are the cases with $V \neq 0$,

Train %	$\tilde{\lambda}$ train MAE	$\tilde{\lambda}$ test MAE	\tilde{V} train MAE	\tilde{V} test MAE	\tilde{V} train (%) $> 1.02V_{\max}$
1	0.0378	12.980	0.114	101.56	0
10	0.2057	4.818	18.270	46.85	0.1044
25	0.2235	3.395	13.086	25.65	0.0956
50	4.7723	5.091	61.875	62.16	0.3057

Train %	$\tilde{\lambda} < 0$ train (%)	$\tilde{\lambda} < 0$ test (%)	\tilde{V} train error (%)	\tilde{V} test error (%)	\tilde{V} test (%) $> 1.02V_{\max}$
1	0.2548	5.4618	0	0.0955	0.6021
10	0.6215	1.9937	0.0046	0.0410	0.2285
25	0.6532	1.3488	0.0139	0.0340	0.1573
50	2.4738	2.5449	0.0719	0.0709	0.3091

Table 1.7: Performance measures of V and λ predictions in a general parameter spaceFigure 1.15: Quality of V predictions, training on 1% of a general parameter space

which comprise 43.96% of all cases. Though it may visually appear that these are better fit, on average, than for all cases, this is an artifact of their comprising such a reduced percentage of all cases: The corresponding MAE values from table 1.7, restricted to cases with $V \neq 0$, are somewhat worse than those for all cases. But this, too, may be a similar artifact, resulting in prioritized training of the majority of cases with $V = 0$. Considering that such options are literally worthless, this may well not yield the most useful results for valuable options. Also, noting from figure 1.6 that there may be many solutions λ yielding $V = 0$, aside from the default values already eliminated ($\lambda_0 = 10^4$), rather than the two distinct solution curves noted in figures 1.5 and 1.7, sensible training to estimate such λ values may prove much more difficult, e.g., without a condition to further restrict attention to solutions with neighborhood $V > 0$ values, i.e., to where V ‘attaches’ itself to the plane $V = 0$. For these reasons, cases with $V = 0$ are eliminated from consideration in a second, much more restricted set of experiments pursued below, where it is eventually shown that an even more stringent test condition, say, $V \geq \underline{V} > 0$, for

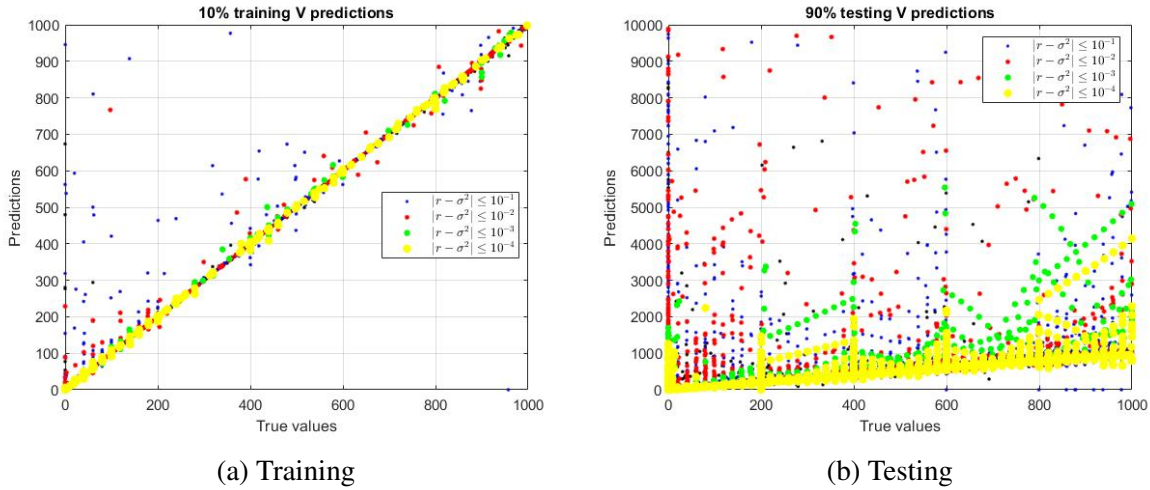


Figure 1.16: Quality of V predictions, training on 10% of a general parameter space

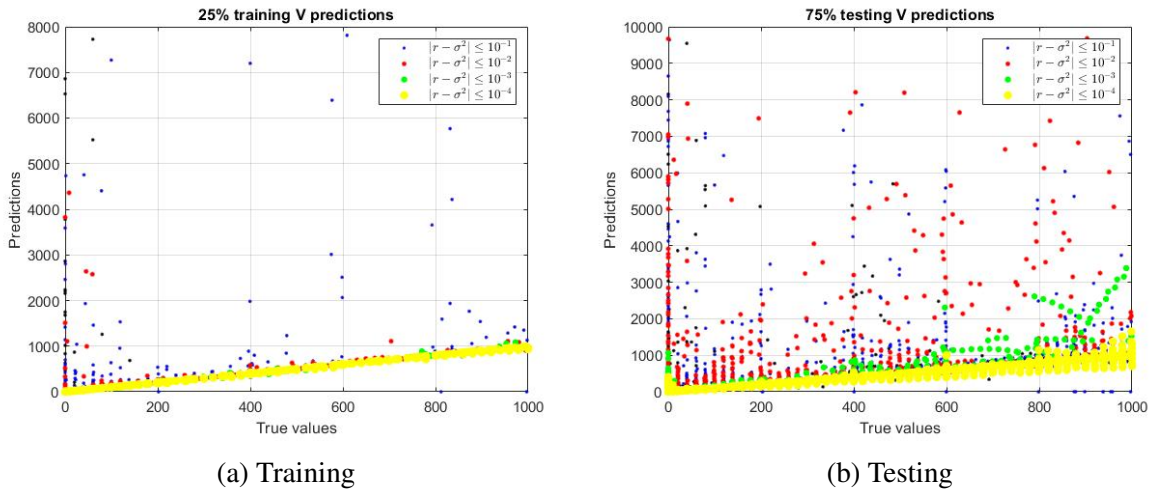


Figure 1.17: Quality of V predictions, training on 25% of a general parameter space

some judiciously chosen lower bound \underline{V} , may ultimately prove yet more effective.

Predicting the function $\lambda(S_0, 0, 0; \sigma, r, T, K)$ in case $r \approx \sigma^2$

It is visually apparent that option values are more closely fitted for smaller bounds on $|r - \sigma^2|$, shown in yellow in figures 1.15-1.18.³⁶ This corresponds to one condition regulating the quality of the approximation in theorem 1.2.8, and motivated a further set of numerical experiments similar to the preceding, but restricted to cases with $|r - \sigma^2| \leq 10^{-4}$. Also eliminating the small

³⁶Similar figures were generated with various bounds on $\sigma^2 T$ (since here $t = 0$), but a similar pattern was not apparent. There may not be sufficient variability across the generated cases for it to appear, but an alternate set of experiments could be designed to address this question specifically, perhaps in a companion paper.

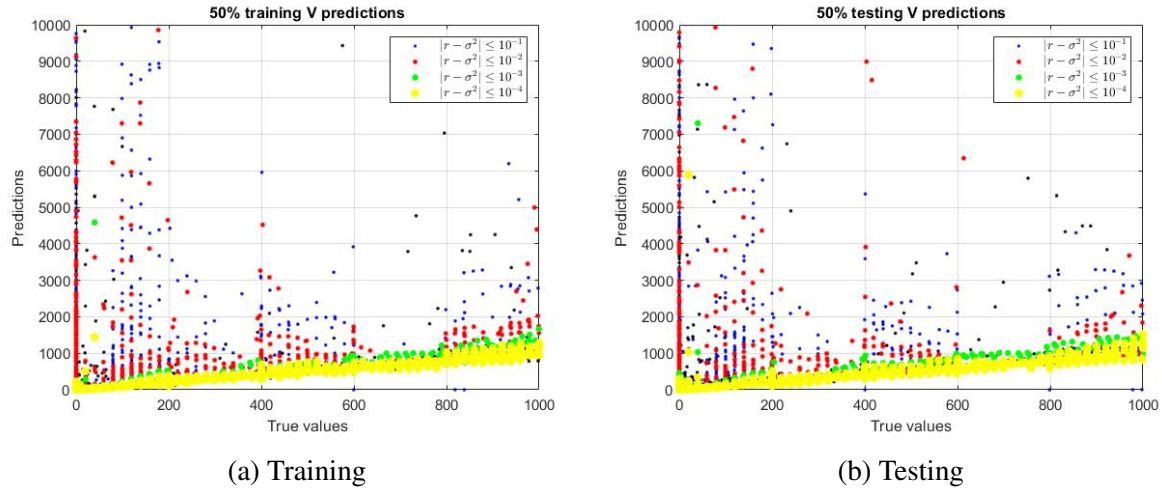


Figure 1.18: Quality of V predictions, training on 50% of a general parameter space

number of cases resulting in negative $\tilde{\lambda}$ values, failed computations of \tilde{V} , and/or predicted prices exceeding by more than 2% the largest ‘true’ price in the data set, with corresponding frequencies listed in table 1.8, this resulted in 4508 cases for training and testing. As a result, with only six days’ computational time limit and the same computational power noted above, training on each of 1, 10, 25, 50, and 75% of the cases, GPRs were fit with thousands of hyperparameter optimizations, apparently converging quite quickly. Plots analogous to figures 1.11-1.14 are shown below, in which the quality of training for higher percentages of cases is clear, and in contrast to the above experiments, where the few 30 hyperparameter optimization rounds completed in three weeks’ computational time, apparently were not enough to fit the training data to a similar quality as for the three lesser percentages.

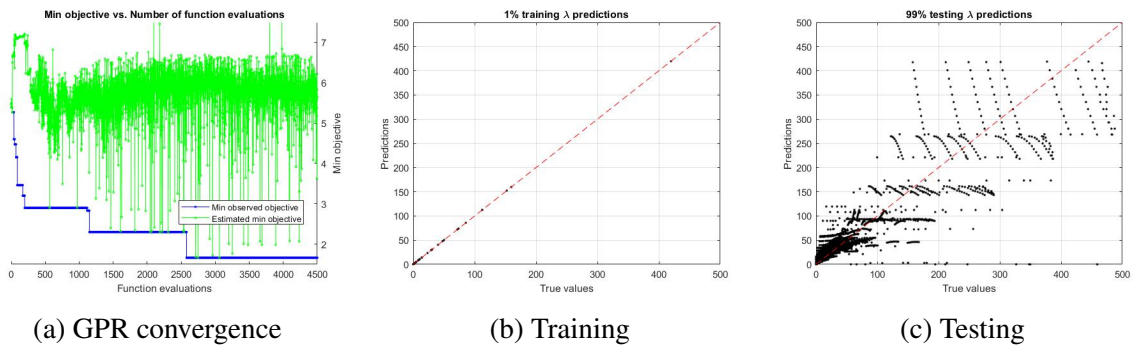


Figure 1.19: Quality of λ predictions, training on 1% of a parameter space satisfying $r \approx \sigma^2$

In addition to these improved results (noting also the much greater agreement with the testing cases, as well as the lower hyperparameter optimization loss values), the predictions of V are significantly better, as shown in plots analogous to figures 1.15-1.18.

It is interesting to observe that, in figures 1.24b-1.28b, the progressive worsening of the results

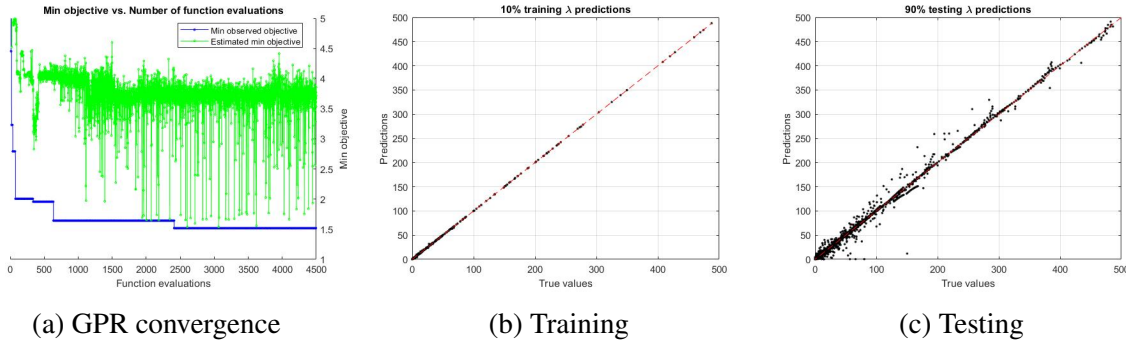


Figure 1.20: Quality of λ predictions, training on 10% of a parameter space satisfying $r \approx \sigma^2$

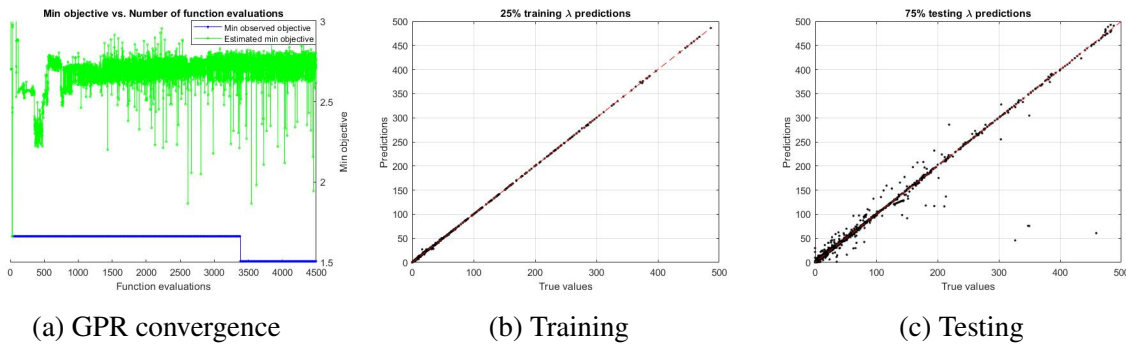
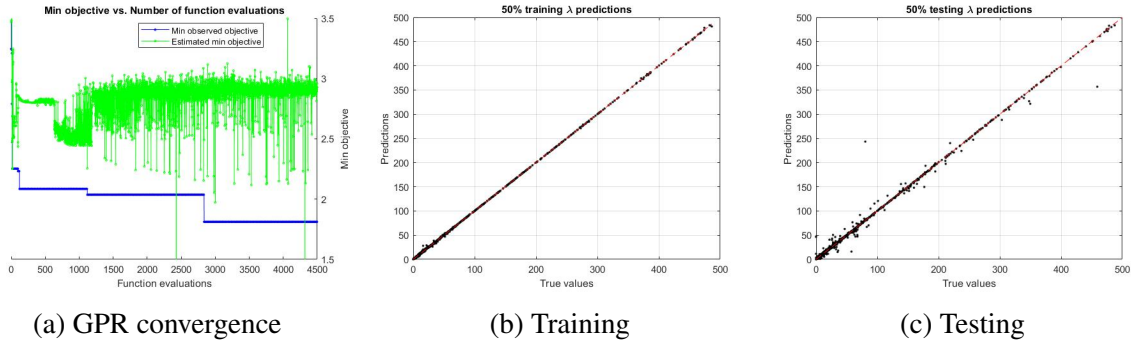
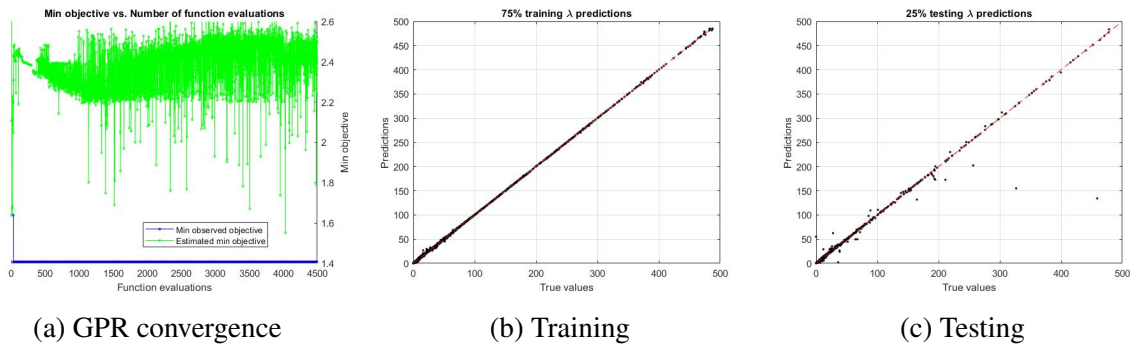


Figure 1.21: Quality of λ predictions, training on 25% of a parameter space satisfying $r \approx \sigma^2$

with decreased training percentage, appears to occur in a structured manner: Namely, there are localized deviations for true values 0, 200, 400, 600, and 800 and greater, which are mostly, but not completely obscured in the case of only 1% training. These values coincide with parameter values of S_0 and K , which may indicate difficulty predicting option prices in cases where $A_T \approx S_0 T$ and $V_T \approx S_0 - K$. This can occur when $r \approx 0 \approx \sigma$, so that $(\forall 0 \leq t \leq T) S_t \approx S_0$. Since such options are not of much practical interest, a failure to predict their prices correctly is not particularly damning. Shown in blue in figures 1.24b-1.28b, are cases in which $r \neq \sigma^2$: Though these do appear to occupy some of the deviations in these plots, particularly as over- rather than underestimates of the true option prices, the pattern may not conclusively demonstrate the increased accuracy of the formula in theorem 1.2.8 for $r \approx \sigma^2$, but this isn't too surprising, recalling that the current experiments were determined by restricting to cases already satisfying this requirement.

As noted above, the following results concerning the estimates of λ ($\tilde{\lambda}$) and V (\tilde{V}), also are much improved over their analogues in table 1.7.

While the rates of negative $\tilde{\lambda}$ values and failed \tilde{V} evaluations are similarly low as in the prior experiments, vanishing in half of the instances in the latter case, the corresponding MAE values, whilst behaving sensibly as before with respect to the training percentage, are much lower, reflecting the graphical improvements already observed. Consideration of similar mean average percentage error (MAPE) values is uninformative, due to the percentage errors for near-zero

Figure 1.22: Quality of λ predictions, training on 50% of a parameter space satisfying $r \approx \sigma^2$ Figure 1.23: Quality of λ predictions, training on 75% of a parameter space satisfying $r \approx \sigma^2$

values being many orders of magnitude higher than the median. This problem is taken into account in defining the tabulated $q_{p\%}$ values:

$$q_{p\%} \equiv \min_{0 \leq q \leq \max V} \{q : \text{MAPE}_{V \geq q} \leq p\%\}$$

In words, $q_{p\%}$ is the smallest nonnegative number (when it exists), such that the MAPE restricted to values of V bounded below by $q_{p\%}$, is no more than $p\%$. The introduction to section 14.3 of Wilmott, Dewynne, and Howison [315] notes that the error of assuming a deterministic interest rate is typically 2%, and so corresponding $q_{2\%}$ values are computed and tabulated above: Rather than judging the V estimates as compared to ‘accurate’ MC values, or benchmarks obtained via other accepted methods within the BS framework, it may be sensible to allow a 2% deviation from such values, as this is the minimal error potentially incurred by use of the BS framework, which in particular assumes a fixed, constant interest rate. And as it is more crucial to accurately price expensive options, the $q_{2\%}$ values indicate the cheapest option prices which may be sufficiently accurately predicted, on average, via the above methodology.

Note that the $q_{p\%}$ values address which *true option values are sufficiently accurately predicted*, rather than which *predicted option values are sufficiently accurate*: Some of the true values near zero in figures 1.24b-1.28b have large predictions which are obviously not within 2% of the true value, and which were indeed the reason for defining $q_{p\%}$, as the predictions in the current and preceding experiments quite clearly tend to exceed rather than undershoot the true values. This is in line with the observation from figures 1.5-1.7, that the functions λ apparently

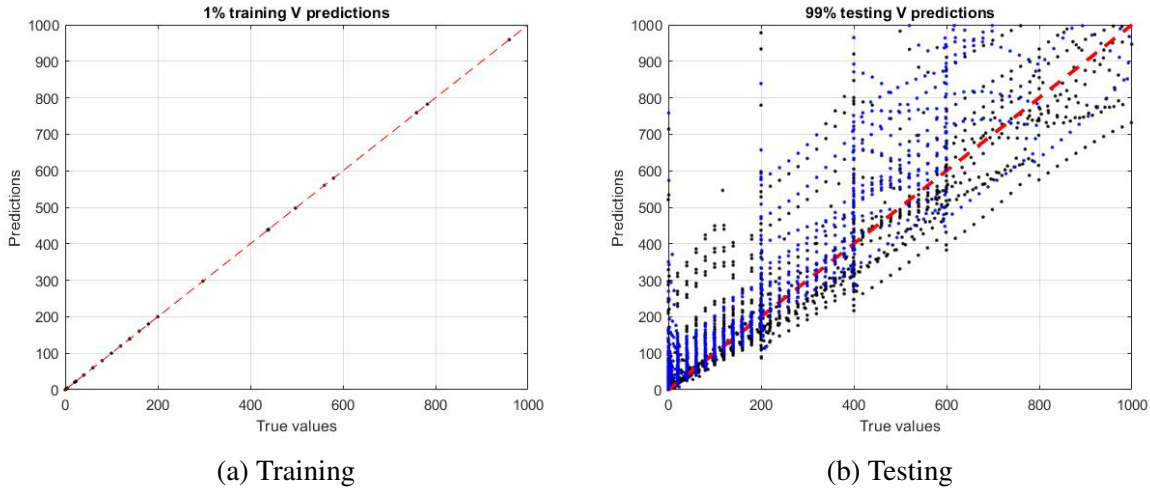


Figure 1.24: Quality of V predictions, training on 1% of a parameter space satisfying $r \approx \sigma^2$

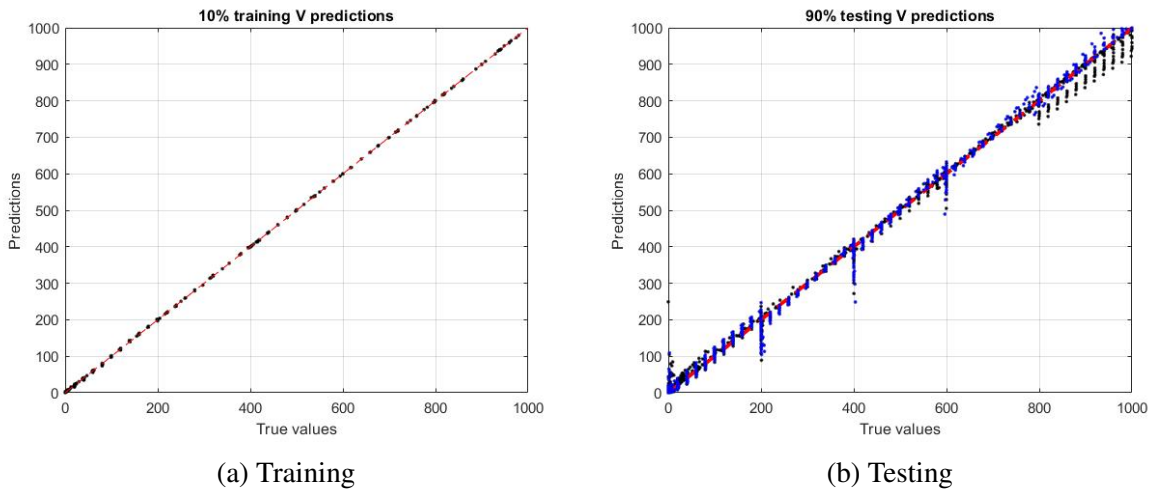


Figure 1.25: Quality of V predictions, training on 10% of a parameter space satisfying $r \approx \sigma^2$

occur near, but not along, curves where at least one derivative of V , i.e. one first-order option sensitivity/Greek vanishes. Similarly defining ‘ $\tilde{q}_{p\%}$ ’ values in terms of a lower bound on the *predictions* of V , yields much higher bounds in some cases, which also do not monotonically decrease in testing with the training percentage, as do the $q_{2\%}$ values tabulated above. To emphasize, the $q_{p\%}$ values indicate which *true* option values are sufficiently accurately priced, on average, rather than which sufficiently large *predicted* option values should be regarded as sufficiently accurate, also on average. This detracts somewhat from the practical utility of the $q_{p\%}$ values, but another set of experiments, bounding the included true option prices further away from zero, rather than requiring them to be strictly positive, may help: In addition to avoiding training a model to correctly identify worthless options, which led to their exclusion in the current set of experiments, training would also not be wasted on sufficiently cheap ones. And it would help bound the grotesquely large (percentage) errors for such options, in which case the ‘ $\tilde{q}_{p\%}$ ’ values might be as informative, and therefore more practically useful, than the $q_{p\%}$

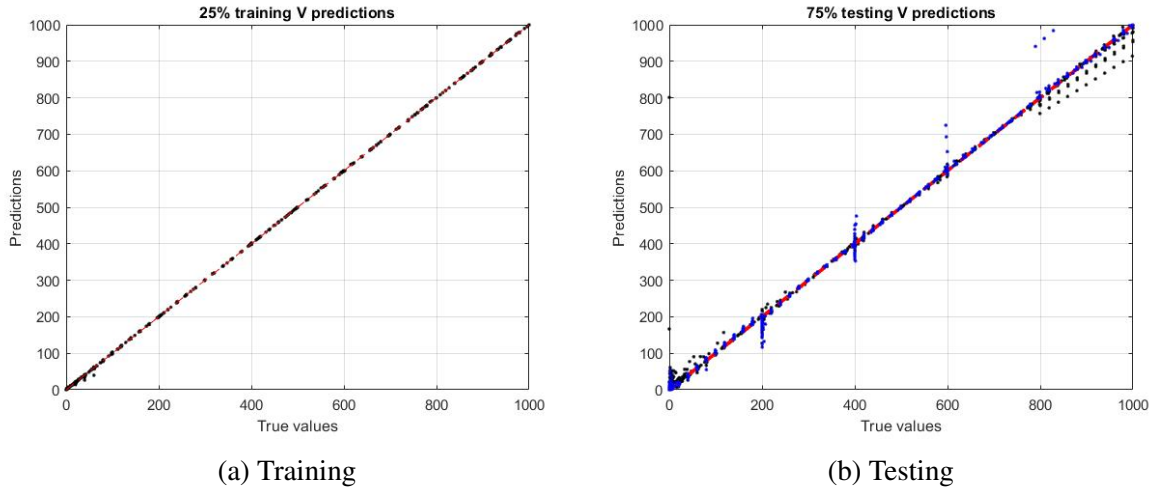


Figure 1.26: Quality of V predictions, training on 25% of a parameter space satisfying $r \approx \sigma^2$

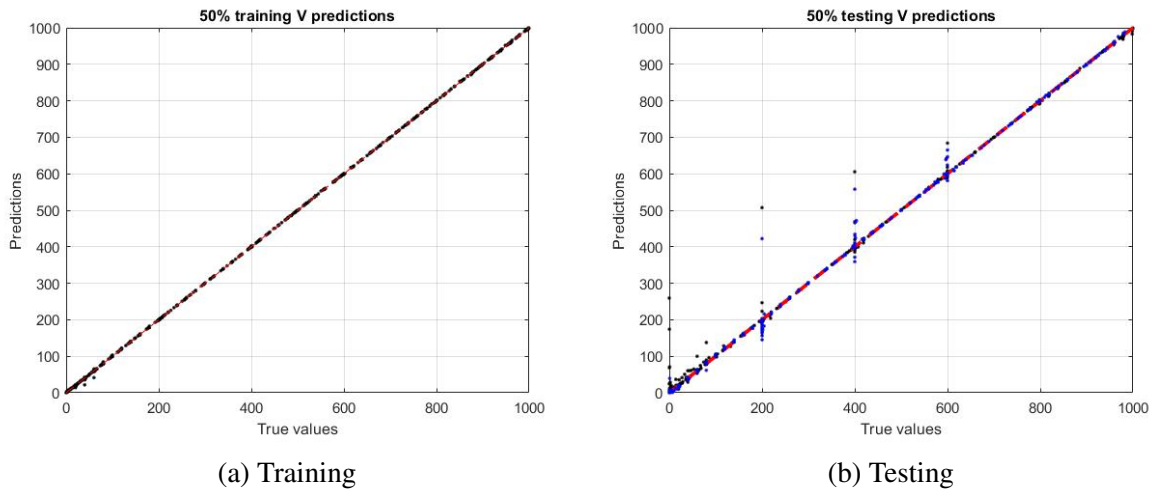


Figure 1.27: Quality of V predictions, training on 50% of a parameter space satisfying $r \approx \sigma^2$

values. But how to do this bounding, e.g., uniformly across training and testing, and/or across training percentages, and/or with larger or lesser bounds than the $q_{p\%}$ values, is not clear, and so is left as one topic for a deeper exploration in a dedicated paper. But it is conjectured that such an exercise, while valuable in its generation of meaningful ' $\tilde{q}_{p\%}$ ' values for practical use, is not likely to otherwise significantly alter the results of the preceding experiments, in any case only slightly modifying as it does the training/testing parameter space.

In prediction from a fitted GPR, the result is a sum of two terms (an additive noise term from the fitted GPR being dropped): One is a linear combination of constants, training predictors (here, σ , r , T , K , and S_0), and squares of the latter. The highest order of training predictors included (zeroth, first, or second) is determined by the basis function (respectively, constant, linear, or quadratic), which in turn is determined during the hyperparameter optimization. The weights are determined in fitting the GPR, and in any case this first additive predictive term is constant

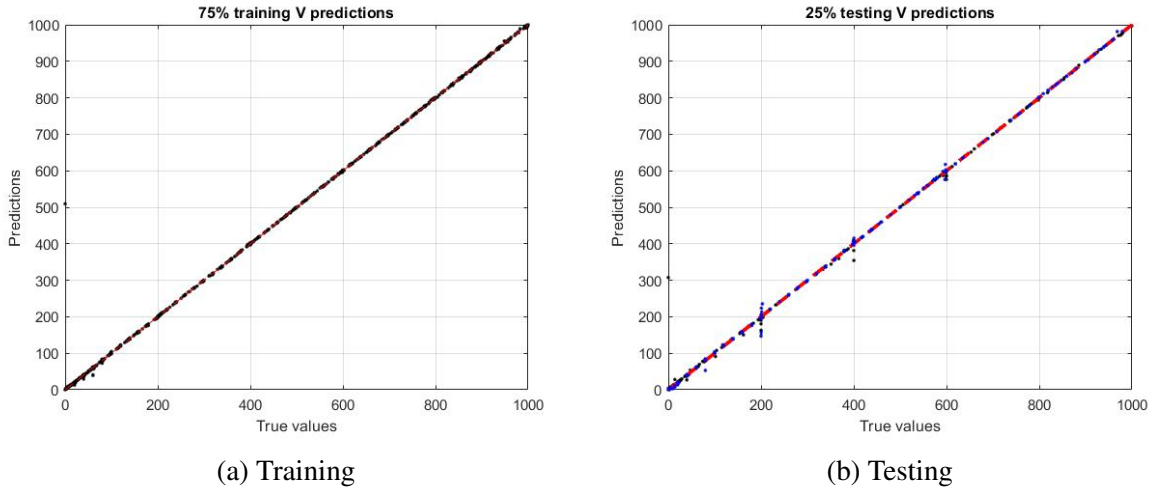


Figure 1.28: Quality of V predictions, training on 75% of a parameter space satisfying $r \approx \sigma^2$

with respect to the predictors for which a prediction is to be computed. The second term is a linear combination of kernel/covariance functions, the weights of this combination again being determined during fitting. The kernel function used, and its internal parameters, are determined as part of hyperparameter optimization. It may also implement automatic relevance determination (ARD); see, e.g., Neal [230]. In either case, all of the considered kernel functions are smooth (see, e.g., Williams and Rasmussen [314]): (Squared) exponential, Matern $^{3/2}$ and $^{5/2}$, and rational quadratic. As such, the resulting predictions $\tilde{\lambda}(\sigma, r, T, K, S_0)$ are also smooth, and may sensibly be substituted into PDE (1.7) via formula (1.43) in theorem 1.2.8. Doing so analytically is impractical, as the explicit expression for $\tilde{\lambda}$ itself is cumbersome to write, including as it does the sum of thousands of terms (of like form) involving many more parameters, the latter typically long decimal numbers. And this is simply to write $\tilde{\lambda}$ explicitly: Deriving its PDE from the two noted formulas (discussed in the preceding section), and substituting its already formidable expression therein, appears unlikely to simplify appreciably. But the point, and another major reason for choosing GPRs over neural networks, e.g., is that *numerical* derivatives may be sensibly computed for $\tilde{\lambda}$, and so substituted into its PDE to assess the extent to which the latter is satisfied. This likely would not work for an hyperparameter optimized neural network, as many of the (e.g., so-called ‘hard’) activation functions are non-smooth, a problem severely compounded by the highly ‘composite’ nature of neural networks, i.e., their prediction via often long series of composed linear combinations of activation functions, further composed over numerous network layers. This is, of course, in addition to the much greater challenge of hyperparameter optimizing the significantly more flexible/parameterized neural networks, noted above.

Direct GPR predictions of the function $V(S_0, 0, 0; \sigma, r, T, K, S_0)$

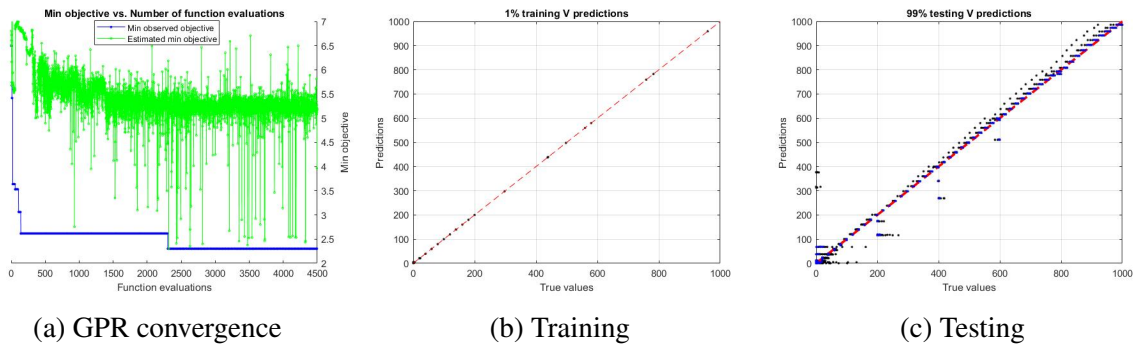
One further concern, is how the preceding results compare to direct GPR predictions of V , or equivalently, what is the impact upon such of intermediate application of formula (1.43) to GPR predictions of λ , supposing the two GPR applications to be of equal quality. To this end, in the case $r \approx \sigma^2$ already studied, analogous experiments were run, but predicting V instead

Train %	$\tilde{\lambda}$ train MAE	$\tilde{\lambda}$ test MAE	\tilde{V} train MAE	\tilde{V} test MAE	Train $q_{2\%}$	Test $q_{2\%}$
1	0.0037	12.576	0.031	51.78	0.01	999.24
10	0.2519	1.673	0.921	6.92	0.92	201.58
25	0.2346	1.766	0.765	4.07	4.78	40.04
50	0.2272	1.248	0.750	2.96	4.78	33.07
75	0.2631	1.635	1.058	2.09	20.01	20.03

Train %	$\tilde{\lambda} < 0$ train (%)	$\tilde{\lambda} < 0$ test (%)	\tilde{V} train error (%)	\tilde{V} test error (%)	\tilde{V} train (%) $> 1.02V_{\max}$	\tilde{V} test (%) $> 1.02V_{\max}$
1	2.5000	0.7042	0	0.2641	0	10.717
10	0.4283	1.3602	0	0.0729	0	0.413
25	0.6843	0.9078	0.0855	0	0	0.469
50	0.7867	1.4373	0.0437	0	0	0.218
75	0.8398	1.0610	0	0	0	0.177

Table 1.8: Performance measures of V and λ predictions in a parameter space with $r \approx \sigma^2$

of λ .³⁷ Plots analogous to figures 1.19-1.23 are shown below:

Figure 1.29: Quality of direct V predictions, training on 1% of a parameter space with $r \approx \sigma^2$

It appears the results these figures depict and those of the preceding subsection are of comparable quality, and the following values, analogous to their applicable counterparts in table 1.8, provide further detail:

The \tilde{V} testing MAE values are uniformly better, but the training values are neither uniformly better nor worse, though they are quite small in both cases. Similarly, both cases had no instances of predictions being discarded as exceeding by at least 2% the maximum ‘true’ price in the data set, but the testing values were much improved in the case of direct prediction of V ; in fact, all but one such value were zero. Perhaps the most notable improvement in this case, is in regard to the $q_{2\%}$ training, and particularly testing values: In the former case, these

³⁷Incidentally, to the best of the author’s knowledge, this also constitutes the first use of GPRs as a surrogate model for (direct) prediction of (arithmetic) Asian option prices.

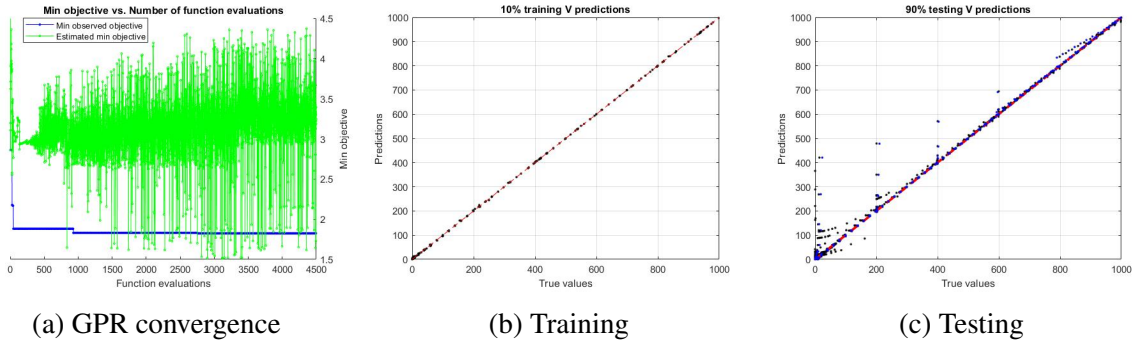


Figure 1.30: Quality of direct V predictions, training on 10% of a parameter space with $r \approx \sigma^2$

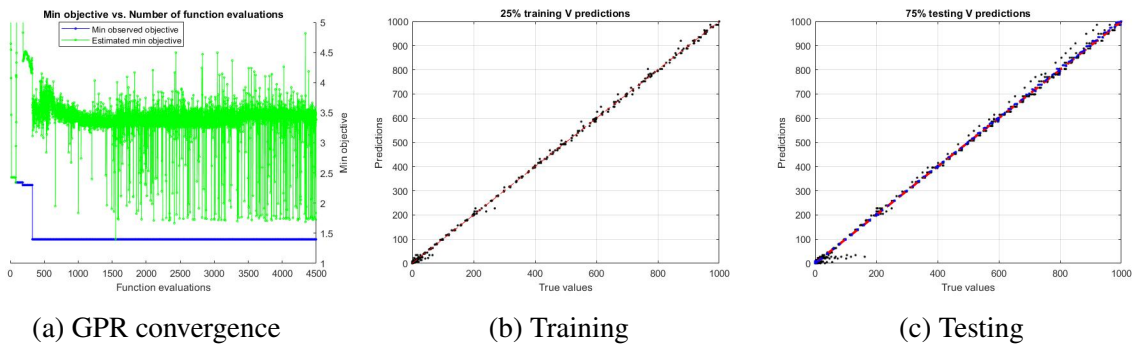
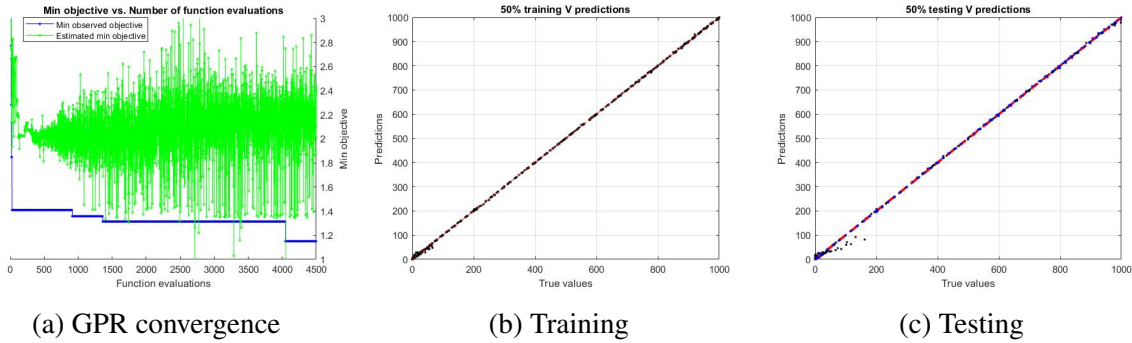
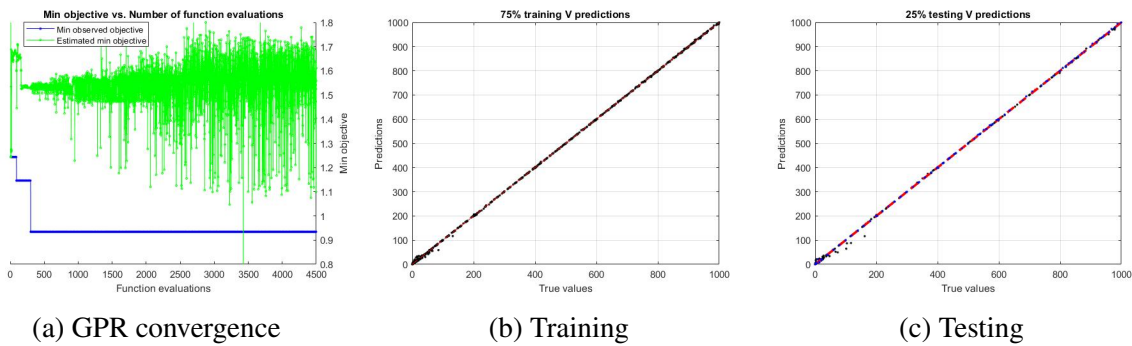


Figure 1.31: Quality of direct V predictions, training on 25% of a parameter space with $r \approx \sigma^2$

values were worse (larger) for the lowest three training percentages, but dropped to 2.18 for the greatest two, as compared to 20.01 in the case of 75% training data, when predicting V indirectly through direct predictions of λ . The arguably more important testing $q_{2\%}$ values, however, were uniformly (and substantially) improved when directly predicting V , falling to 3.29 for the greatest two training percentages, as compared to 20.03 in the case of 75% training data, when predicting V indirectly through direct predictions of λ . Thus, direct GPR prediction of V yields more accurate predictions for lower option prices, in particular, within 2% of the ‘true’ values, in the range $[3.29, 20.03]$. Direct prediction of V is therefore likely practically preferable, particularly with regard to cheaper options. However, this same approach yields significantly more negative predictions of V , in all but one case between one and five percent of the time, compared to similarly many cases of errors being no more than 1.5% when predicting V through λ , pooling errors due to negative predictions of λ and failed predictions of V for positive predictions of λ . But for modest data set sizes, the additional few percentages of easily-detected ‘wasted’ MC simulations for ‘true’ V values, may not be considerable. And direct prediction of V is not bound to respect the constraint, $r \approx \sigma^2$, of the preceding subsection, so in general, this approach is again likely practically preferable. A similar study with a more general parameter space, similar to that of section 1.2.3, would be needed to verify this conjecture for direct prediction of V , and may be a suitable topic for another paper.

Figure 1.32: Quality of direct V predictions, training on 50% of a parameter space with $r \approx \sigma^2$ Figure 1.33: Quality of direct V predictions, training on 75% of a parameter space with $r \approx \sigma^2$

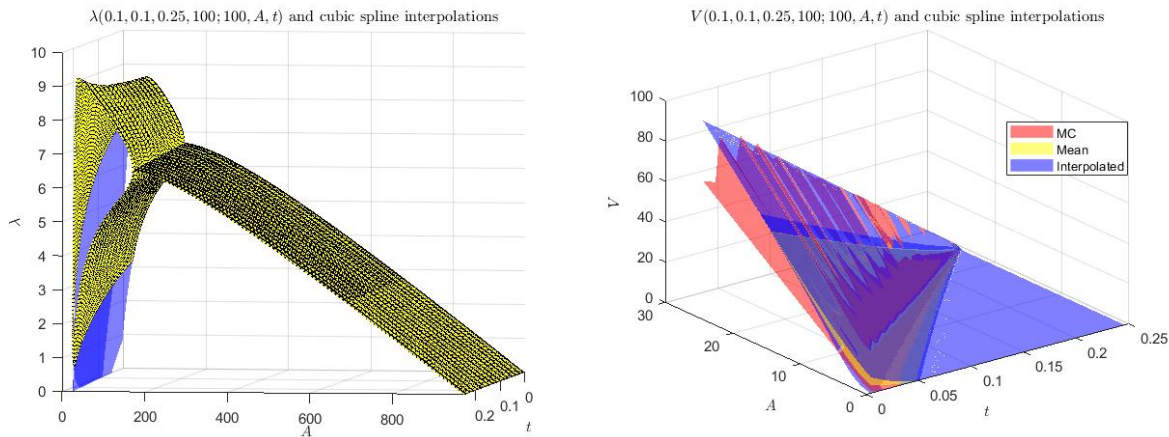
1.2.4 Results for $t > 0$

Formula (1.45) yields an alternate procedure for estimating the function $\lambda(S, A, t; \sigma, r, T, K)$, in particular, when $t > 0$, i.e., *after* the contract is written: It permits to generate arbitrarily fine curves/surfaces from this function, in the case $A \geq KT$, which may then be extrapolated and used to predict prices for values $A < KT$. Naturally, the extrapolation worsens the further out-of-the-money an option is, and so this approach may be unreliable for values of A near zero, e.g., when $t = 0$ and necessarily $A = 0$. This approach is thus complementary to that of the preceding sections, but has the advantage of being much smaller and more rapid: Rather than build a GPR for a large variable/parameter space, as done there, only values $A \geq KT$ need to be considered for given σ, r, K, T, S , and t , in order to solve for $\lambda(S, A, t; \sigma, r, T, K)$ such that $V(S, A, t; \sigma, r, T, K; \lambda)$ equals the corresponding values stipulated by formula (1.45). Various curve-fitting techniques may be applied to extrapolate the function $\lambda(S, A, t; \sigma, r, T, K)$, but a small investigation indicated that cubic splines are best: Among all other applicable default options in `Matlab R2019b`, and Taylor series taken of sufficiently large order to converge for nonnegative values of A , cubic splines performed best. Other `Matlab` built-ins performed poorly, and the Taylor series often required up to 100 terms for convergence on all $A \geq 0$, resulting in effectively linear fits, only a small ‘penalty’ off of the first-order fit. But it is certainly possible that even better results could obtain from a more thorough treatment of this curve-fitting problem, which could be of interest in a separate paper. The results presented here are similar vs. those of section 1.2.3, i.e. for similarly ‘representative’ values of the parameters

Train %	\tilde{V} train MAE	\tilde{V} test MAE	$\tilde{V} < 0$ train (%)	$\tilde{V} < 0$ test (%)	\tilde{V} train (%) $> 1.02V_{\max}$	\tilde{V} test (%) $> 1.02V_{\max}$	Train $q_{2\%}$	Test $q_{2\%}$
1	0.511	9.29	5	3.5871	0	0	2.56	400.08
10	0.688	3.21	2.9979	3.4005	0	0.049	4.74	23.53
25	1.832	2.43	0.3422	1.3470	0	0	8.75	13.28
50	0.443	0.85	2.2290	3.0052	0	0	2.18	3.29
75	0.419	0.73	2.9540	2.5641	0	0	2.18	3.29

Table 1.9: Performance measures of direct V predictions in a parameter space with $r \approx \sigma^2$

and variables, as a study across the space of such values is no longer needed. Respectively displayed in figures 1.34a and 1.34b, are predictions of $\lambda(100, A, t; 0.1, 0.1, 0.25, 100)$ and corresponding prices from theorem 1.2.8.



(a) Extrapolations of $\lambda(100, A, t; 0.1, 0.1, 0.25, 100)$ (b) Corresponding and Monte Carlo prices

Figure 1.34: Prices matching cubic spline-extrapolations of $\lambda(100, A, t; 0.1, 0.1, 0.25, 100)$

Two sheets are observed in the $\lambda(100, A, t; 0.1, 0.1, 0.25, 100)$ surface, corresponding to the pairs of surface intersection curves noted in figures 1.5 through 1.7. These two sheets apparently also bifurcate from a single sheet for values $A \gg KT$. As such, various curves were separately fitted, for each value of t considered, only to values corresponding to one or the other sheet, neglecting the common sheet from which these bifurcate. The corresponding prices are respectively seen to constitute upper and lower predictions. These match one another for near- and deeply-out-of-the-money options, indicated by regions of light blue shading in figure 1.34b. In the former case, these values match those stipulated by formula (1.45) when $A = KT = 25$, and make apparent the inadequacy of the MC estimates in this region: Applying the built-in implementation of Longstaff and Schwartz [210] in Matlab, with a default 1000 sample paths and using antithetic variates, a ‘jaggedness’ is prominent for $A = 25$, despite using a time grid with, at coarsest, a time step of size 0.0025. In fact, this jaggedness remains apparent wherever the MC prices are positive, potentially indicating a serious problem, at the

very least, in the built-in `MatLab` implementation of Longstaff and Schwartz [210].³⁸ Barring issues with this specific implementation, this plot alone glaringly illustrates shortcomings of MC estimates of arithmetic Asian option prices: Even with 1000 sample paths and antithetic variates, such estimates' deviation is apparently up to the level of $\pm\$20$, or roughly $\pm 25\%$, judging from the values near $A = 25$ and $t = 0$. And the MC estimates for at least near-the-money options apparently *underestimate* those (commonly) predicted from both extrapolated sheets of $\lambda(100, A, t; 0.1, 0.1, 0.25, 100)$ in figure 1.34b: Given that the latter match the exact values of formula (1.45) when $A = 25$, where the MC jaggedness is particularly exacerbated, the extrapolated price predictions resulting from theorem 1.2.8 may be sensibly regarded as correct, in contrast to the MC estimates. Furthermore, their *overestimation* of such estimates may partly exonerate the observed overestimation in figures 1.15 through 1.18; to some extent, this may indicate the ability of the GPR to *correct* bias in the MC training data. The implicit treatment of such data as inexact/noisy, and the use of an inexact GPR for training sets of more than (a default) 10000 data points, is also further justified by this one indication of both the bias and variance of MC estimates of arithmetic Asian option prices.

Further out-of-the-money, the price predictions corresponding to extrapolated values of $\lambda(100, A, t; 0.1, 0.1, 0.25, 100)$ begin to diverge and bound the MC estimates; their arithmetic mean is an improvement on both. But as indicated, this is expected from the nature of extrapolation, and although further study may (perhaps significantly) improve such extrapolation, it appears appropriate to regard the proposed method as *complementary* to MC estimation, more (less) useful for nearer- (deeper-out-of-)the-money options. And in all cases, it is certainly faster.

In some ways, the proposed method of this subsection may be more striking than the general GPR constructed earlier, as it is clearly superior to MC price estimation for sufficiently near-the-money options, and is much easier to implement, smaller, and faster. But as it performs worst when $A = 0$, in particular when $t = 0$, it is naturally complementary to the GPR approach, in a similar manner as it is to MC estimation. Finally, note that the results discussed for the method of this subsection appear robust to various unilateral parameter and variable changes: Doubling or halving the maturity, taking $r = \sigma^2 = 0.01$, and considering individually $K = 10 = S$, the resulting plots are qualitatively identical to figures 1.34a and 1.34b. Up to scaling the axes, results analogous to figure 1.34a are identical at a glance, and 'essential' features of the $\lambda(100, A, t; 0.1, 0.1, 0.25, 100)$ surface are preserved: Though the differential geometry, e.g., curvature, of the surface may vary, as well as the curve determining its bifurcation, the bifurcation into two sheets itself and the surface 'shape' perceived at a glance are the same.

1.3 Results

Recall that there are four primary European Asian options, one for each choice between put/call and fixed-/average-strike, with payoffs tabulated as follows:

³⁸To confirm proper usage, it was verified that prices with $A \geq 25$ match those of formula (1.45); apparently this formula is also used in the built-in `MatLab` implementation of Longstaff and Schwartz [210].

Option	Payoff
Fixed-strike put	$\max\{0, K - A\}$
Fixed-strike call	$\max\{0, A - K\}$
Average-strike put	$\max\{0, A - S_T\}$
Average-strike call	$\max\{0, S_T - A\}$

The preceding only considered fixed-strike European Asian call options, but pricing formulas for the other three cases follow directly from that in theorem 1.2.6 and the put-call parity and fixed-average symmetry relations in section 4.3.4 of Kwok [185]. Furthermore, appropriately differentiating such parity and symmetry relations, hedging parameter formulas for the three cases not considered above follow directly from those in corollary 1.2.7. In short, despite considering only one of the four tabulated cases in the foregoing, the pricing and hedging results of the latter immediately yield corresponding results for the three remaining cases. The same holds for the S_T -independent result of theorem 1.2.8.

Extending the results obtained to non-constant, deterministic rates of interest, $r(t)$, and volatility, $\sigma(t)$, is in principle straightforward, but likely to be very computationally burdensome. See, e.g., chapter 8 of Wilmott, Dewynne, and Howison [315] for the general procedure, and also for the inclusion of dividends, to which similar comments can be made: Accommodating a constant, continuously-compounded dividend rate, q , or non-constant but deterministic rate, $q(t)$, can in principle be approached as in the preceding, but is likely to further complicate the computations involved. In particular, these comments hold with regard to the approach taken via Lie point symmetries of the pricing PDE (1.7), since the ‘time’ symmetry (1.27) was not used to generate the invariant solution, $u(x, t)$, given in lemma 1.2.3. In fact, additional symmetries could arise for appropriate functional forms $r(t)$, $\sigma(t)$, and/or $q(t)$. Note that, implicit in the present discussion is the assumption that such functions reflect *continuous* dividend payments and/or averages: In the discrete case, which uses Dirac delta functionals, $\delta(t)$, the classical BS equation is solved in a ‘sequential’ manner which requires numerical implementation and is otherwise not much more complicated, mathematically speaking, than a BS European option. Chapters 8-12 of Wilmott, Dewynne, and Howison [315], e.g., give an overview of this methodology and its application to a variety of exotic options.

Regarding future work, there are a number of possible directions not already thoroughly discussed: One already discussed prior to problem 1, is the possible analytical solution of the PDE for $\lambda(S, A, t; \sigma, r, T, K, S_0)$, substituting formula (1.43) into pricing PDE (1.7). As noted there, this appears to be quite involved, an assertion supported by the many numerical experiments conducted to shed light on the behavior of λ . Another question of interest is the S_T -independent hedging parameters/Greeks implied by those of corollary 1.2.7: The pricing formula (1.43) of theorem 1.2.8 cannot sensibly be differentiated to obtain these parameters, at least not without distinguishing between S within/out the PDF for S_T , in precisely the same sense that formula (1.43) does not satisfy pricing PDE (1.7), at least with λ held constant with respect to all of S , A , and t . But the expectation with respect to S_T of the parameters of corollary 1.2.7 can in principle be analytically taken, to generate analogous hedging parameters to the pricing formula (1.43) of theorem 1.2.8. However, whilst some simplifications would arise through the latter formula, and the overall process would differ little if at all from that in obtaining it, the com-

putations are likely to be much lengthier and more error-prone, even using Maple or another CAS, due both to the length of the parameters of corollary 1.2.7 and, in particular, the presence of powers of $\ln y$ in the corresponding multiplicative integral terms. Having obtained such formulas, numerical experiments could be conducted in much the same way as for the pricing formula (1.43), as before with the ‘true’ values being computed from the Matlab built-in implementation of MC Asian option hedging à la Longstaff and Schwartz [210]. It is expected that such results would be qualitatively similar to those already obtained. But one interesting possibility would be to ascertain whether the various ‘ $\tilde{\lambda}$ ’ values, across pricing and hedging experiments, correspond as expected from analytical substitution of $\lambda(S, A, t; \sigma, r, T, K, S_0)$ into formula (1.43) and then pricing PDE (1.7). That is, whether the ‘ $\tilde{\lambda}$ ’ values for hedging experiments, correspond in some appropriate way to derivatives of those for the pricing experiments already conducted. But whether and how to do this is not yet clear, and judging by the complexity of the hedging parameters of corollary 1.2.7, compared to the pricing formula (1.40), it may be reasonable to expect similarly complicated results in relation to pricing formula (1.43). And noting the work devoted to numerical experiments for this formula alone, those for the four hedging parameters of corollary 1.2.7 may well be expected to require at least four, if not many more, times the effort and space. For these reasons, such experiments may best be left to a companion paper.

Similarly, note that other hedging parameters, e.g., the ‘Vega,’ V_σ , rho, V_r , ‘Vera,’ $V_{r\sigma}$, epsilon/psi, V_q , supposing inclusion of a constant continuously-compounded dividend rate, q , as discussed above, and higher-order and cross-partials of V with respect to S , A , t , σ , r , and/or q , may be obtained just as were the hedging parameters of corollary 1.2.7. But unlike the latter, these additional parameters are not needed to verify the satisfaction of pricing PDE (1.7) by formula (1.40), which was indeed the ultimate purpose of corollary 1.2.7. The latter nonetheless provided arguably the most practically essential parameters, Δ and Γ , along with the important Θ and novel, say ‘Asian Greek,’ V_A . As discussed above, these represent a starting point for hedging formulas analogous to the pricing formula (1.43), numerical experiments similar to those already performed for the latter, and of course, the same comments apply to the additional hedging parameters excluded from corollary 1.2.7.

Finally, a comment is in order regarding the fact that the numerical experiments were performed only for $t = 0$, implying that $A = 0$ and $S = S_0$: The case for arbitrary $t \in [0, T]$ can be considered as an option struck at time t with maturity $T - t$ and the same parameters σ , r , and K , with initial price S_t . This is accommodated by the numerical experiments, since a broad range of maturities and initial prices were considered, but $A_t \equiv \int_0^t S_t dt$ is nonzero unless $S_t = 0$ for some $t \approx 0$. It is conjectured that numerical experiments with a similarly broad range of A values would likely not substantively change the results obtained, and the choice to set $t = 0 = A$ was motivated by the priority to price an option when it’s written, but this is another possible avenue for future work: Adding additional variability to A could only provide further insight into the nature of the pricing formula (1.43). But to emphasize, taking $A = 0$ does not remove its influence on the results obtained: Whilst it may vanish from the pricing formula (1.43), the structure of the latter was obtained based on the dynamics of A and how it enters the payoff (1.8). In fact, the noted form, $A_t \equiv \int_0^t S_t dt$, was chosen over alternatives at an early

stage, partly for the simplification (both algebraically and numerically) that necessarily $A = 0$ if $t = 0$, when it is most crucial, in practice, to price an option. Lastly, note also that an increase in A increases both the likelihood of exercise and the resulting payoff, and so should increase V , which is also suggested by the presence of A/T in the first of the two summands multiplying the lognormal MGF in pricing formula (1.43): Having seen that more expensive (specifically, than about \$20 in the case of 75% training) options are more likely to be priced sufficiently accurately, cases with positive A are therefore in some sense less urgent than those studied, with $A = 0$. And the effect of increasing A may be similar to that of S_0 , seen to be apparently linear in figures 1.5d and 1.7d; or of decreasing K , also apparently linear in the planar portion above the horizontal $V = 0$ in figure 1.6, and the apparent similarity between corresponding plots in figures 1.5 and 1.7, save for a positive constant subtracted from the former, and neglecting the plots concerning S_0 (figures 1.5d and 1.7d).

1.4 Conclusions

The pricing and hedging of (arithmetic) Asian options continues to pose a difficult problem. This has prompted the development of many different solution approaches, which vary in terms of their theoretical utility and practicality. The present paper contributes to both aspects by explicitly establishing multiple families of exact solutions, parameterized by tuples of complex numbers, to relaxed power mean Asian option pricing problems. The solutions approximately satisfy the full problem, and in one case converge to exact solutions under certain parametric restrictions. A number of hedging parameters/Greeks are derived in verifying this family, which may well be useful in their own right. The family consists of (optimal) invariant solutions, constructed first for the pricing PDE of an Asian option with generalized *power* mean, via this equation's (optimal) Lie point symmetry group. Prior to particularizing to the case of an arithmetic average, and subsequently in this case, the results substantially correct, generalize, extend, and make more rigorous those of Taylor and Glasgow [296], Caister, Govinder and O'Hara [55] and [56], and Antoniou [15]. They also offer useful simplifications for implementing numerical analysis of the pricing and hedging problems. Finally, a lengthy series of numerical experiments explores the behavior of the family of invariant solutions, estimating mappings from the problem variables and parameters to particular family members, such that accurate prices generated via MC simulations are reliably predicted, and for near-the-money options, variant MC simulations are vastly outperformed. In this way, it is possible to consistently price sufficiently expensive options, which are naturally of priority in practice.

Chapter 2

Intraday market predictability: A machine learning approach

Abstract

This chapter analyses and demonstrates the predictability of intraday market returns. Conducting apparently the largest existing study of five-minute market returns using state-of-the-art machine learning models trained on the cross-section of lagged market index constituent returns to forecast five-minute market returns, it is shown that regularized linear models such as lasso and elastic nets along with nonlinear tree-based models such as random forests yield significant predictability. Ensemble models that combine individual model predictions perform the best across time and their return predictability translates into economically significant profits with Sharpe ratios after transaction costs of 0.98. These results provide strong evidence that intraday market returns are predictable during short time horizons, beyond what can be explained by transaction costs. It is further shown that the lagged constituent returns hold significant predictive information that is not contained in lagged market returns or in lagged price trend and liquidity characteristics. Consistent with the hypothesis that predictability is driven by slow-moving trader capital, predictability decreased after decimalization, and market returns are more predictable during the middle of the day, on days with high volatility or high illiquidity, and in years of financial crisis.

2.1 Introduction

The predictability of the aggregate market is a central topic in financial economics. While long-horizon (i.e., monthly or quarterly) market predictability has been extensively studied, intraday (i.e., within a trading day) market predictability has received relatively less attention. Traders require time to incorporate new information about cash flows and discount rates, and over short time horizons equity prices can differ from their adjusted fundamental values, particularly when market frictions are high. This process may introduce short-horizon predictability in equity returns and raises several interesting questions. First of all, if markets are predictable at short horizons what is the magnitude of this predictability? Secondly, is intraday return predictability economically profitable, and if so, does this profitability survive transaction costs?

Finally, if markets are predictable it is interesting to know which characteristics, e.g., lagged liquidity or price trends, are in fact important for predicting intraday returns.

Motivated by these questions, intraday market predictability is studied using a cross-section of lagged returns of the market and its constituent stocks as predictor variables.¹ This paper is apparently the first to conduct such a study, and it is speculated that the lack of previous studies on this topic may be due to statistical challenges associated with the high-dimensional inputs and the computational difficulties of estimating models using large panels of high-frequency data. The first issue is surmounted by using a variety of cutting-edge machine learning models necessary to accommodate the long list of predictors and rich functional forms. Considered are candidate methods from Gu, Kelly and Xiu [152] and Hastie, Tibshirani and Friedman [157], including linear models with regularization and dimension reduction using lasso (LAS), elastic net (EN), and principal component regression (PCR), and nonlinear tree-based models like random forests (RF) and gradient-boosted regression trees (GBRT) along with artificial neural networks (ANN). Also considered are the ensemble mean and median of these models. The baseline model uses all the five-minute returns in an expanding estimation window. Training machine learning models on such a large data panel is computationally challenging and this second issue is overcome using the Apache Sparkling Water and H2O.ai computing framework, which allows us to efficiently estimate machine learning models on large datasets. For example, in October, 2016, the estimation window covers 285 months and contains roughly 450,000 five-minute returns for each S&P constituent, requiring approximately 48 hours to estimate all models.

The null hypothesis throughout this chapter is that markets are not predictable, since any predictability should be removed by active traders. The alternative hypothesis is that the information in lagged returns is *not* instantly reflected in market prices, and as a result, lagged returns *are* predictive of short-term market returns. These hypotheses are tested by examining the statistical predictability and the economic significance thereof for each of the machine learning models considered. If models can forecast returns and consistently profit in so doing, then markets are likely to be predictable on short-horizons. The second alternative hypothesis is that there is predictability beyond that explained by transaction costs. This is tested by evaluating the economic significance of model predictions after accounting for realistic transaction costs. If model predictions remain profitable after such costs, then there is likely incorrect pricing beyond that explained by transaction costs.

To examine if there is statistical predictability of intraday market returns over five-minute time intervals, the estimated models are trained on lagged returns with an expanding estimation window from 1993 to 2016. For non-ensemble models, linear as well as nonlinear, out-of-sample R^2 (R_{OOS}^2) values range up to 2.00% for LAS, followed by 1.95% for EN, and 1.71% for the nonlinear RF model. The ensemble mean (median) model yields R_{OOS}^2 of 2% (2.01%), illustrating the strength of combined forecasts. However, most of this predictability is concentrated in the pre-decimalization period from 1993 to 2000. In the early post-decimalization period

¹While individual stocks are likely more predictable than the market, market predictability is studied first to avoid issues related to data mining and concerns about lack of liquidity.

from 2001 to 2004, the LAS, EN, and RF models still have positive R_{OOS}^2 values of 0.91%, 0.81%, and 0.85%, respectively. The ensemble mean and median models have respective post-decimalization R_{OOS}^2 values of 1.04% and 1.01%. However, during the late decimalization period from 2005 to 2016, model predictability is found to significantly decrease due to decreased transaction costs, but remains positive for the LAS, PCR, and ensemble models.

Next, to examine the economic significance of model predictions, a market-timing strategy that buys (sells) the market on positive (negative) predictions is considered. The results demonstrate that a small intraday R_{OOS}^2 value can yield large economic profits, especially given the numerous trading opportunities available at a five-minute interval. Using the baseline model from 1993 to 2016, all models are found to have positive returns. The LAS, EN, RF, and GBRT models have the highest non-ensemble statistical predictability and so too high annualized returns (Sharpe ratios²) of 191%, 188%, 198%, and 192% (2.71, 2.68, 2.90, and 2.84). The ensemble mean and median have annualized returns (Sharpe ratios) of 205% and 204% (2.90 and 2.82). Again, most returns are concentrated in the pre-decimalization period. However, in the late post-decimalization period from 2005 to 2016, all models still earn economically significant profits, with ANN earning the lowest returns (Sharpe ratios) of 16% (0.83). The consistently positive returns and high Sharpe ratios provide further evidence that intraday markets are predictable.

Lastly, economic significance is analyzed after transaction costs. To do this the market-timing strategy is modified to only trade when the signal is strong, i.e., when the model prediction exceeds the transaction cost. Using the baseline model from 1993 to 2016, all models, with the exception of ANN, are found to have positive returns even after accounting for transaction costs. The PCR, RF, and ensemble mean and median models have Sharpe ratios of 0.68, 0.77, 0.67, and 0.98, respectively, vastly exceeding the Sharpe ratio of 0.48 for the benchmark buy-and-hold SPDR S&P 500 (SPY) portfolio. Again, most returns are concentrated in the pre-decimalization period. In the late post-decimalization period from 2005 to 2016, the PCR and RF models have annual returns (Sharpe ratios) of 9% and 8% (0.88 and 0.81) after transaction costs. The ensemble mean and median have respective annual returns (Sharpe ratios) of 4% and 5% (0.35 and 0.68) after transaction costs. As a comparison, from 2005 to 2016 the SPY returned 7% with a Sharpe ratio of 0.46. Thus, in this recent sample PCR, RF, and the ensemble median models continue to earn significantly higher Sharpe ratios than the market does even after transaction costs, providing strong evidence that markets are predictable even after accounting for transaction costs.

It is hypothesized that the demonstrated significant predictability of intraday market returns through time is driven by slow-moving trader capital, i.e., by infrequent portfolio rebalancing (Bogousslavsky [40] and Duffie [101]), at a high frequency. If some traders rebalance their portfolios infrequently, they may be slow to incorporate shocks to individual stock returns into the aggregate market, particularly when traders face severe volatility or illiquidity. To analyse this further, it is examined whether the results differ within the trading day, in periods of high versus low volatility or illiquidity, and during periods of financial crisis. First, since traders are

²Of annualized returns to their volatility values.

most active at the beginning and end of each day, predictability is expected to be low during those times. The results show that predictability is indeed stronger when traders are less active and exhibits an inverse-U shape. Second, during periods of high volatility and high illiquidity, traders encounter significant market frictions. Consistent with the hypothesis, predictability and its economic significance are found to increase when market volatility and illiquidity are high. Finally, predictability is also expected to be relatively high during crisis periods due to attending, significant market frictions. The results demonstrate that predictability is indeed stronger during the subprime mortgage and EU debt crises.

The baseline predictor variables used here are the cross-section of lagged intraday returns for the market constituents. A natural comparison to these cross-sectional models are autoregressive models for the market return itself, since the predictability results could simply be capturing intraday momentum. For example, Heston, Korajczyk and Sadka [160] find significant auto-correlation of half-hour returns at daily intervals and Gao et al. [133] find that the first half-hour return of the SPY predicts the last. Thus, as an additional analysis, results of the cross-sectional model are validated by contrast against results obtained with AR(1), AR(p) where the lag-order p is chosen to minimize the validation error, and AR(500) models estimated using ordinary least-squares (OLS), LAS, EN, and PCR. The baseline cross-sectional models are confirmed to significantly outperform the autoregressive models, indicating that the cross-section of lagged constituent returns has significant predictive information that is not contained in lagged market returns.

As a final additional analysis intraday market predictability is evaluated using additional lagged stock characteristics as predictors. This analysis includes market beta, momentum, illiquidity, extreme returns, trading volume, volatility, skewness, and kurtosis, all estimated over the previous day.³ Also considered is the lagged bid-ask spread. In most cases adding additional variables is found to *decrease* model predictability, indicating that these characteristics do not help lagged returns predict the market portfolio returns. These findings have important implications for the possible economic mechanisms that drive such predictability. For example, could the predictability be caused by intraday momentum, as argued by Heston, Korajczyk and Sadka [160] and Gao et al. [133]? This explanation seems unlikely given that price trend variables fail to improve model predictability. A big picture implication of the findings is that a careful exploration of the economic mechanisms driving predictability is warranted.

The chapter is related to at least three existing strands of literature. First, the results are naturally related to the recent literature examining intraday return predictability using lagged returns and trading volume. Chordia, Roll and Subrahmanyam [78] and [79] study the predictability of short-run stock returns, finding that intraday returns cannot be predicted by past prices, but that order imbalances do forecast short-horizon returns. Heston, Korajczyk and Sadka [160] find significant autocorrelation of returns at daily intervals, for up to 20 days. Gao et al. [133] demonstrate that the first half-hour return of the SPY market exchange-traded fund (ETF) predicts the last half-hour return. Bogousslavsky [40] theoretically establishes that seasonality in intraday returns may be caused by traders' infrequent rebalancing. Chincó,

³Gu, Kelly and Xiu [152] demonstrate that price trend and liquidity have the strongest predictive ability.

Clark-Joseph and Ye [74] use a LAS model on the cross-section of NYSE lagged returns to show that one-minute returns are predictable. These studies use *linear* models to forecast returns, whereas the present models use information from the entire *cross-section* of lagged returns as well as other characteristics, and permit more flexible functional forms accommodating variable interactions and other *nonlinear* effects. Furthermore, Ke, Kelly and Xiu [172] and Renault [258] show that text data forecasts intraday returns.

The chapter also relates to the rapidly expanding literature applying machine learning techniques in financial economics.⁴ The chapter is most closely related to Gu, Kelly and Xiu [152], which applies an extensive array of machine learning techniques to the problem of predicting equity risk premiums (see also Fischer and Krauss [120], Long, Lu and Cui [209] and Marković et al. [215]).⁵ Bianchi, Büchner and Tamoni [35] consider the prediction of *bond*, rather than stock, risk premiums, and other financial applications of machine learning include their use for derivatives pricing (Ye and Zhang [319]), hedge fund selection and return prediction (Chen, Wu and Tindall [68]), credit risk management (Barboza, Kimura and Altman [26]), portfolio management and optimization (Yun et al. [321] and Day and Lin [95]), cryptocurrency (Dutta, Kumar and Basu [106] and Alessandretti et al. [10]), stochastic discount factors (Korsaye, Quaini and Trojani [183]), and factor models (Bryzgalova, Pelger and Zhu [50], Chen, Pelger and Zhu [71], Feng, Polson and Xu [119], Gu, Kelly and Xiu [153], and Kelly, Pruitt and Su [175]).

Finally, the present study is similar in spirit to the literature on aggregate market return predictability, which focuses on long horizons over months, quarters and years.⁶ Fama [115] established the general view that return predictability at long horizons is driven by time-varying expected returns, consistent with market efficiency. The research differs fundamentally, since the established predictability and its economic value is driven by market frictions at high frequencies.

The chapter is organized as follows: Section 2.2 introduces the considered machine learning methods and explains the evaluation of intraday predictability. Section 2.3 presents results for statistical predictability, economic significance before and after transaction costs, and a robustness check using different training window sizes. Section 2.4 examines if the results differ across the time of day, across levels of volatility and illiquidity, or during crisis periods, whether autoregressive models can also forecast returns, and whether predictability can be improved by including additional input variables. Section 2.5 concludes. The appendices contain further details on the data used, variable cleaning, the stock characteristics used as inputs, and on the implementation of the machine learning methods, in general, and hyperparameter

⁴See Weigand [310] for a recent concise summary of asset pricing via machine learning and Heston and Sinha [161] and Hajek and Barushka [154] for reviews of the history of financial applications of machine learning.

⁵Other papers similar to that of Gu, Kelly and Xiu [152], which are also concerned with the use of machine learning for predicting returns, include those of Chong, Han and Park [76], Feng, He and Polson [118], Kelly, Pruitt and Su [176], Sutherland, Jung and Lee [291], and Xue et al. [317].

⁶See, among others, Ang and Bekaert [13], Campbell and Thompson [57], Chen, Da and Zhao [70], Kelly and Pruitt [174], Neely et al. [233], and Welch and Goyal [311]. See Koijen and Van Nieuwerburgh [181] and Rapach and Zhou [256] for recent surveys.

optimization, in particular.

2.2 Methodology

Given T high-frequency price observations, indexed by $1 \leq t \leq T$, and denoting by p_t the natural logarithm of the t^{th} observed price, the corresponding logarithmic return is given by

$$r_t \equiv p_t - p_{t-1} \quad (2.1)$$

Of particular interest in this chapter is the market return, denoted by r_t^M , and whether or not it can be predicted. To assess this, one of the largest empirical studies of market prediction in the high-frequency literature is conducted. The primary dataset is the trade and quotes (TAQ) database, containing intraday transaction data for all stocks on the New York stock exchange (NYSE), American stock exchange (AMEX), NASDAQ, and other American regional exchanges, from February, 1993, to October, 2016. The goal is to predict five-minute changes in the aggregate market, which is proxied by the SPDR S&P 500 (SPY) ETF. The baseline predictor variables are lagged returns of the SPY and S&P 500 constituents. The latter are obtained from the center for research in security prices (CRSP) and updated monthly.⁷

Given a vector of lagged characteristics of the SPY and S&P 500 constituents, \mathbf{X}_t^ℓ , indexed by integers, $0 \leq \ell \leq 500$, the objective of the present chapter is to use state of the art machine learning methods to approximate the empirical model given by $f : \mathbb{R}^{n \times K} \rightarrow \mathbb{R}$:

$$r_{t+1}^M = f(\mathbf{X}_t^\ell) \quad (2.2)$$

(Here, $n \in \mathbb{N}$ denotes the number of observations.) In the baseline model, the covariates are lagged returns, such that $\mathbf{X}_t^\ell = r_t^\ell$, resulting in $K = 501$ covariates. Additionally considered are models that include other firm-level characteristics, such that $\mathbf{X}_t^\ell = [r_t^\ell \ z_t^\ell]$, where $[\cdot]$ denotes matrix concatenation, and z_t^ℓ (z_t^ℓ with $\ell \geq 1$) is the market beta (of firm $\ell \geq 1$), illiquidity, kurtosis, maximum, minimum, momentum, skewness, volatility, trading volume, or bid-ask spread. Models so expanded have $K = 1002$ covariates (501 each characteristics and lagged returns).⁸

Introduced subsequently are several methods for estimating the function f in the empirical model (2.2). Considered are individual (non)linear models and so-called ensemble ones that weigh together multiple such individuals. Finally, it is explained how, given an estimator of f , predictability can be assessed both statistically and economically.

2.2.1 Linear models

Consider the optimization problem

$$\inf_{\beta \in \mathbb{R}^K} m[\mathbf{y} - \mathbf{X}\beta] \quad (2.3)$$

⁷Appendix B.1 provides further details on the databases used.

⁸The only exception to this is the expanded model with z_t^ℓ equal to market beta which only has $K = 1001$ covariates, since the SPY market beta $z_t^0 = 1$. Appendix B.2 details these characteristics' calculation.

Here, in the machine learning literature terminology, the market returns are *targets*, denoted \mathbf{y} , an $n \times 1$ column vector, and the predictors, e.g., lagged returns or other characteristics, are denoted \mathbf{X} , an $n \times K$ matrix in case of K predictors. In equation (2.3), $m[\cdot]$ denotes a metric or loss for the fit of the model and β denotes the relevant parameters in the model. For example, in the case of OLS regression the metric is taken to be the Euclidean/ ℓ^2 norm, i.e., $m \equiv \|\cdot\|_2$, and the solution to optimization problem (2.3) is given by the classical OLS estimator,

$$\hat{\beta} \equiv (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Given many predictors, the simple linear regression model easily becomes inefficient, leading to in-sample over-fitting which is detrimental to the objective of out-of-sample prediction. Presented are two approaches to deal with this; regularization and principle component regression.

Regularization

One primary objective of *regularization* techniques (Friedman, Hastie and Tibshirani [125]) is to avoid over-fitting in statistical models. This is often accomplished by adding a penalty term to the optimization problem (2.3):

$$\inf_{\beta \in \mathbb{R}^K} \{m[\mathbf{y} - \mathbf{X}\beta] + \lambda n[\beta]\} \quad (2.4)$$

Here, the functional $n[\cdot]$, often a norm, penalizes non-zero parameters and the *regularization parameter* λ regulates the impact of the penalty as a multiplicative scale. A classical example is ridge regression, in which optimization problem (2.4) is modified to

$$\inf_{\beta \in \mathbb{R}^K} \left\{ \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_2^2 \right\}$$

The smoothness of ridge regression (Marquardt [216]) resulting from using the ℓ^2 norm is computationally advantageous, but may result in many ‘near-but-non-zero’ coefficients and so may not reduce the dimensionality of the optimization problem in a sufficient manner. Considered instead are lasso regression (Tibshirani [300]) and elastic nets (Zou and Hastie [329]).

In the case of lasso regression (LAS) the optimization problem (2.4) is modified to

$$\inf_{\beta \in \mathbb{R}^K} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \right\}$$

Hence LAS employs the computationally difficult (i.e., non-smooth) ℓ^1 norm, but has the resulting advantage that many coefficients are driven to zero exactly, leaving out only those of sufficient predictive importance. The resulting β , is non-zero only for those predictors which most significantly determine the target and may therefore be of much lower dimension than the original problem which is indeed one of the primary objectives of the use of regularization.

In the case of an elastic net regression (EN) the optimization problem (2.4) is modified to

$$\inf_{\beta \in \mathbb{R}^K} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \alpha \lambda_1 \|\beta\|_1 + \frac{1 - \alpha}{2} \lambda_2 \|\beta\|_2^2 \right\}$$

Hence EN convexly combines ridge and lasso penalties to balance these two competing properties; smoothness and perfect elimination of unimportant predictors. Here, $\alpha \in [0, 1]$ is the coefficient of the convex combination of ℓ^1 and ℓ^2 norms of the regression coefficients, β , and in general, each may have its own regularization parameter, respectively, λ_1 and λ_2 . In the particular cases of $\alpha \in \{0, 1\}$, elastic nets respectively reduce to ridge and lasso regressions.⁹

For expository convenience, explicit regularization problems have been presented primarily with respect to OLS regression, excepting the general additive problem (2.4). But all considered forms of regularization may be extended further, naturally to generalized linear models, but also to all other considered machine learning techniques. Throughout, the common theme in lasso (ridge) regression problems is to add penalty terms involving regularization-parameter scaled ℓ^1 (ℓ^2) norms of particular models' (hyper-)parameters, and in the case of elastic nets, to similarly add convex combinations of such penalty terms. In other words, the discussion captures the spirit of these regularization techniques, as applied to all considered machine learning methods, which are most easily explicated for OLS regression.

Principal component regression

Principal component regression (PCR) is a dimension-reduction technique used to summarize variation within a dataset using a small number of linear combinations thereof (Jolliffe [169]). Given a dataset \mathbf{X} , consisting of n observations of K predictors, PCR solves

$$\sup_{\mathbf{w} \in \mathbb{R}^K} \frac{\mathbf{w}^T \Sigma \mathbf{w}}{\mathbf{w}^T \mathbf{w}}$$

Here, $\mathbf{w} \in \mathbb{R}^K$ are the predictor weight vectors and Σ is the covariance of the predictors. The motivation for PCR is clear from this formulation: Since an eigenvector of Σ , \mathbf{w} , solves the eigenvalue problem, $\Sigma \mathbf{w} = \lambda \mathbf{w}$, for a corresponding eigenvalue of Σ , λ , the best *eigenvector* solution of this problem is obviously that for which the corresponding eigenvalue, λ , is largest. This follows since the Raleigh quotient to be optimized, $\mathbf{w}^T \Sigma \mathbf{w} / \mathbf{w}^T \mathbf{w}$, which measures normalized variation/variance of the dataset along the weight vector, \mathbf{w} , simplifies in this case to λ . So in applying PCR, it is necessary only to compute an eigenvalue decomposition of Σ , sort its eigenvalues, and take as many corresponding eigenvectors as are desired principal components, to yield the principal directions. Normalizing each by its corresponding square-root eigenvalue, the desired principal components are obtained.

In the context of optimization problem (2.3), PCR is first applied to the predictors, \mathbf{X} . Supposing $\kappa \ll K$ components are desired, the projected predictors, $\mathbf{Z} \equiv (\mathbf{X} - \mu) \mathbf{W}$, are used in a linear regression yielding the classical OLS estimator given by

$$\hat{\beta} \equiv (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y}$$

Here, μ is the mean of the predictors and $\mathbf{W} \equiv [\mathbf{w}_{(1)} \quad \mathbf{w}_{(2)} \quad \cdots \quad \mathbf{w}_{(\kappa)}]$ is the matrix of principal components to be used. The resulting (demeaned) projected predictors, \mathbf{Z} , are thus only of

⁹Appendix C explains how the hyperparameters α , λ_1 , and λ_2 , or in the case of LAS only λ , are selected.

dimension $\kappa \ll K$, yielding a potentially significant reduction in the number of predictors and resulting regression model complexity, while preserving, to the extent possible, the richness of variation in the original data.¹⁰

2.2.2 Nonlinear models

More generally, the fundamental problem to be addressed is the solution of

$$\inf_{\mathbf{f} \in \mathcal{F}} m[\mathbf{y} - \mathbf{f}(\mathbf{X})] \quad (2.5)$$

Here, \mathcal{F} denotes the set of functions from which candidate predictors, $\mathbf{f} : \mathbb{R}^{n \times K} \rightarrow \mathbb{R}^n$, are drawn and m denotes a metric or loss for the fit of the model. Several linear models have been considered and now are two classes of nonlinear ones, those each based on decision trees and artificial neural networks.

Tree-based models

Random forests and gradient-boosted regression trees constitute two classes of ensemble methods, which collect multiple simple estimators (in both cases, decision trees) in order to produce more robust estimates.

Decision trees are nonparametric, hierarchical sequences of decisions, which optimally construct, based on the training and validation data, sequences of decisions to classify or regress arbitrary input predictors. Individual decision trees train in logarithmic time with the number of training points, but over-fitting is common, as more decisions (greater ‘depth’) are needed to better model training data, which may not generalize well. Individual trees often behave chaotically, too, in that the optimal structure may change drastically in response to the addition or removal of a handful of training data. Ensembles of trees mitigate the resulting large variance of individual decision trees, and avoid over-fitting by restriction to simple individuals across the ensemble, but inherit from such individuals some degree of the advantages of interpretation and efficient training. The fundamentals and many refinements of decision tree training are provided by Breiman et al. [47].

Random forest (RF) models (Breiman [46]) independently and pseudorandomly generate decision trees, which are separately trained and whose predictions are then averaged to yield the ensemble prediction. In RF prediction robustness improvements are achieved via variance reductions implicit in the law of large numbers. Given predictors \mathbf{X} , denote an individual decision tree’s estimator, say that of the i^{th} tree generated in an ensemble, by $\mathbf{g}_i(\mathbf{X})$. Supposing there are N decision trees in the ensemble, the prediction of RF is simply

$$\mathbf{f}_{\text{RF}}(\mathbf{X}) \equiv \frac{1}{N} \sum_{i=1}^N \mathbf{g}_i(\mathbf{X})$$

¹⁰Selecting the value of κ can be a complex problem. Appendix C explains how this hyperparameter is set.

That is, RF amounts to considering the (arithmetic) average of the predictions of individual trees in the forest.¹¹

Each individual decision tree in the ensemble is trained on data drawn pseudorandomly with replacement from the training data. I.e., each individual tree is trained on a bootstrapped sample of the training data, and aggregated in making predictions via the average of those from each individual, and so random forests constitute one instance of bootstrap aggregated (*bagged*) predictors (Breiman [45]). This *bagging* yields many individuals with uncorrelated prediction errors and less variance on average, whereas individuals tend to have high variance and overfit. The construction may however introduce estimator bias, which is the problem addressed by gradient-boosted regression trees. The bagging implicit in random forests also addresses the NP-completeness of the problem of training a *globally* optimal individual decision tree, which necessitates the use of heuristics, including typical greedy implementations, in training such individuals and bagging mitigates much of the bias that such heuristics introduce during individuals' training.

By the law of large numbers upon which random forests are justified, results improve with the number of trees in the forest, but non-linearly and plateau at some point, beyond which computing effort is unnecessarily expended. The more features considered when splitting any tree node, the greater the variance but the lesser the bias, and vice versa. For similar statistical reasons, these same three hyperparameters are of greatest interest in the construction of gradient-boosted regression trees.

As their name suggests, gradient-boosted regression trees (GBRT) implement *boosting*, a second case of ensemble methods. As opposed to averaging methods, the simple estimators are sequentially (and so not independently) generated in a manner which progressively eliminates bias from the ensemble (a standard reference is Schapire and Freund [266]). Boosting refers to the notion of developing a strong learner (a predictor with metric, m , approaching zero on arbitrary data) from a weak one (a predictor which performs marginally better than random guessing). Schapire [265] first affirmatively answered this *hypothesis boosting* question, and the adaptive resampling and combining, or *arcing*, algorithm of Freund and Schapire [123] is regarded as the canonical method for achieving such boosting in machine learning. *Gradient* boosting refers to the generalization of this and other boosting algorithms to the case of arbitrary differentiable loss functions, first achieved explicitly by Friedman [126] and [127].

In the case of GBRT, the ensemble prediction is a weighted combination of individual predictions, with weights γ_i determined as part of training leading to the following representation:

$$\mathbf{f}_{\text{GBRT}}(\mathbf{X}) \equiv \sum_{i=1}^N \gamma_i \mathbf{g}_i(\mathbf{X})$$

¹¹The crucial hyperparameters to optimize for RF are the number of trees in the forest, the maximum number of predictors or *features* considered at each node of each tree, and the maximum number of nodes between the leaves and root of any tree, known as the *depth*. Appendix C explains how these are selected.

Specifying $\mathbf{f}_0 \equiv 0$ and denoting $\mathbf{f}_{\text{GBRT}} \equiv \mathbf{f}_N$, the following holds for $1 \leq i \leq N$

$$\mathbf{f}_i(\mathbf{X}) = \mathbf{f}_{i-1}(\mathbf{X}) + \gamma_i \mathbf{g}_i(\mathbf{X})$$

At each stage, i , of training, the decision tree predictor, \mathbf{g}_i , is greedily chosen to solve

$$\inf_{\mathbf{g} \in \mathcal{G}} \mathcal{L}[\mathbf{y}, \mathbf{f}_{i-1}(\mathbf{X}) + \gamma_i \mathbf{g}(\mathbf{X})] \quad (2.6)$$

Here, \mathcal{L} is some loss function, typically expressed as a sum of losses, say L , between corresponding targets, $y_j \in \mathbf{y} \equiv \{y_j\}$, and predictors, $\mathbf{x}_j \in \mathbf{X} \equiv \{\mathbf{x}_j\}$, as

$$\mathcal{L}[\mathbf{y}, \mathbf{f}_{i-1}(\mathbf{X}) + \gamma_i \mathbf{g}_i(\mathbf{X})] \equiv \sum_{j=1}^n L[y_j, \mathbf{f}_{i-1}(\mathbf{x}_j) + \gamma_i \mathbf{g}_i(\mathbf{x}_j)] \quad (2.7)$$

Note that the solution to optimization problem (2.6) is conditional on the current ensemble, \mathbf{f}_{i-1} , as opposed to the independent generation of individual decision tree predictors, \mathbf{g}_i , in RF.

In the case where the loss, L , is differentiable, gradient boosting solves optimization problem (2.6) via gradient descent as

$$\mathbf{f}_i(\mathbf{X}) = \mathbf{f}_{i-1}(\mathbf{X}) - \gamma_i \sum_{j=1}^n \partial_{\mathbf{f}} L[y_j, \mathbf{f}_{i-1}(\mathbf{x}_j)]$$

Here, the weights, γ_i , are chosen to optimize a loss similar to that of formula (2.7) given by

$$\inf_{\gamma > 0} \sum_{j=1}^n L(y_j, \mathbf{f}_{i-1}(\mathbf{x}_j) - \gamma \partial_{\mathbf{f}} L[y_j, \mathbf{f}_{i-1}(\mathbf{x}_j)])$$

Thus, the weights can be seen as step sizes/learning rates in this gradient descent procedure.¹²

The seminal paper by Friedman [127] combines gradient boosting with bagging: Each consecutive decision tree is trained on a random sample of the training data, as in the case of random forests, but as for gradient-boosted regression trees, the sampling is done without replacement, so that consecutive trees are trained on exponentially decreasing shares of the training data.

Artificial neural networks

As their name suggests, artificial neural networks (ANNs) are motivated by the neural networks found in animal brains, and theoretical neuroscientific models thereof (see Bishop [37] and [36]). The most general-purpose and well-known neural network architecture is feedforward in which an *input layer* consisting of the inputs, $\mathbf{y}^0 \equiv \mathbf{x}_i$, is followed by a sequence of *hidden*

¹²In addition to the hyperparameters discussed for RF, there are in the case of GBRT the loss function, L , used for training and the additional uniform multiplier/scale, ν , which may be factored out of the step sizes, γ_i , as an explicit learning rate, additionally permitting ‘shrinkage’ regularization (Fawcett [117]). A smaller learning rate typically leads to better testing performance, but requires more training steps/additional decision trees in the ensemble, to maintain a given training error. Appendix C provides further details.

layers, each consisting of a number of neurons. Initially, the inputs are weighted by a set of learned parameters and added to another, the so-called *bias*, to yield an input for each neuron in the succeeding layer. Denoting by k_1 the number of neurons in this layer and $k_0 \equiv K$ the number of inputs, there are thus $k_1(k_0 + 1)$ parameters which yield this layer's inputs as

$$(\forall 1 \leq j \leq k_1) x_j^1 \equiv \mathbf{y}^0 \mathbf{w}_j^1 + b_j^1$$

Here, \mathbf{w}_j^1 is the learned column vector of weights for the row vector of inputs, \mathbf{y}^0 , and b_j^1 the corresponding learned additive bias. To introduce non-linearity into the output of a neuron $y_j^1 \equiv \phi_j^1(x_j^1)$ where ϕ_j^1 is the so-called activation function of neuron j .

More concisely the mapping may be written as

$$\begin{aligned} \mathbf{x}^1 &\equiv \mathbf{y}^0 \mathbf{W}^1 + \mathbf{b}^1 \\ \mathbf{y}^1 &= \mathbf{\Phi}^1(\mathbf{x}^1) \end{aligned} \tag{2.8}$$

Here, \mathbf{W}^1 is the matrix with columns \mathbf{w}_j^1 , \mathbf{b}^1 is the row vector with entries b_j^1 , and $\mathbf{\Phi}^1$ is the mapping from the row vector \mathbf{x}^1 , with entries x_j^1 , to the row vector \mathbf{y}^1 , with entries, y_j^1 . In succeeding hidden layers, if applicable, the process is iterated as follows:

$$\begin{aligned} (\forall 2 \leq \nu \leq N) \mathbf{x}^\nu &\equiv \mathbf{y}^{\nu-1} \mathbf{W}^\nu + \mathbf{b}^\nu \\ \mathbf{y}^\nu &= \mathbf{\Phi}^\nu(\mathbf{x}^\nu) \end{aligned} \tag{2.9}$$

Here, N is the total number of hidden layers. Finally, the *output layer* yields the predictions,

$$\begin{aligned} \mathbf{x}^{N+1} &\equiv \mathbf{y}^N \mathbf{W}^{N+1} + \mathbf{b}^{N+1} \\ \mathbf{y}^{N+1} &= \mathbf{\Phi}^{N+1}(\mathbf{x}^{N+1}) \end{aligned} \tag{2.10}$$

In the present application, the output layer is linear and the number of neurons, k_{N+1} , is naturally set to one, resulting in

$$\mathbf{f}_{\text{ANN}}(\mathbf{X}) \equiv \mathbf{y}^N \mathbf{W}^{N+1} + \mathbf{b}^{N+1} \tag{2.11}$$

Hence, the neurons are simply linearly aggregated into the forecast.

Collecting all weights and biases across layers, the feedforward network has a total parameter count of $\sum_{\nu=0}^N k_{\nu+1}(k_\nu + 1)$ as there are $\sum_{\nu=0}^N k_{\nu+1}$ biases and $\sum_{\nu=0}^N k_{\nu+1}k_\nu$ weights and such networks may indeed be highly parametric. Classically, these are optimized via stochastic gradient descent, but several adaptive/'momentum'-based generalizations have been proposed.¹³

¹³In ANN, the optimization algorithm employed for training, the loss function it uses, the ℓ^1 and ℓ^2 regularization parameters, the number of epochs and batches, and the level of *dropout* (Srivastava et al. [290]), the percentage of training data discarded in each epoch to avoid over-fitting/regularize the optimization problem, all constitute *hyperparameters* which impact the quality of predictions from a model with trained *parameters*. However, the actual network architecture, characterized by the number of hidden layers, the number of neurons in each layer, and the activation function associated with each neuron, often have an even greater impact. Appendix C provides further details on these hyperparameters.

2.2.3 Ensemble methods

Ensemble methods seek to combine multiple predictors' results, for both variance reduction and 'crowd wisdom' purposes. For example, various measures of central tendency or location parameters, including the median and alternate (weighted) means, such as the harmonic, geometric and arithmetic mean, may act as ensemble aggregation methods and represent an 'average' prediction based on all the predictors' outputs.¹⁴ For regression problems, the two most prominent machine learning ensemble aggregation methods are boosting and bagging, as respectively outlined for GBRT and RF models. Abstractly, given any of these ensemble aggregation methods, say, $AM \in \{\text{Bag, boost, (weighted) arithmetic/geometric/harmonic mean, median}\}$, etc., and a set of predictors' outputs, e.g., $\{\mathbf{f}_{\text{LAS}}(\mathbf{X}), \mathbf{f}_{\text{EN}}(\mathbf{X}), \mathbf{f}_{\text{PCR}}(\mathbf{X}), \mathbf{f}_{\text{RF}}(\mathbf{X}), \mathbf{f}_{\text{GBRT}}(\mathbf{X}), \mathbf{f}_{\text{ANN}}(\mathbf{X})\}$, the ensemble prediction is

$$\mathbf{f}_{\text{AM}}(\mathbf{X}) \equiv \text{AM}(\{\mathbf{f}_{\text{LAS}}(\mathbf{X}), \mathbf{f}_{\text{EN}}(\mathbf{X}), \mathbf{f}_{\text{PCR}}(\mathbf{X}), \mathbf{f}_{\text{RF}}(\mathbf{X}), \mathbf{f}_{\text{GBRT}}(\mathbf{X}), \mathbf{f}_{\text{ANN}}(\mathbf{X})\})$$

The sequel considers the (arithmetic) average and the median as aggregation methods and as such does not introduce any additional hyperparameters.

It should be noted that regularization may also be applied to the ensemble aggregation methods just as it may be in the particular cases of GBRT and RF models.¹⁵ The machine learning ensemble aggregation methods are more flexible than straightforward computation of some measure of central tendency or location parameter of the various models' predictions, as the former permit regularization and hyperparameter tuning. Which ensemble aggregation method to use is itself a hyperparameter optimization problem, albeit a small one which may be largely mitigated by simply computing many ensembles, e.g., boosting, bagging, and a variety of 'popular' measures of central tendency or location parameters. The justification for any particular aggregation method may come from either the statistical (variance-reduction and bias) properties of bagging or boosting, or laws of large numbers, central limit theorems, or other asymptotic results applicable to the computed measures of central tendency or location parameters. Also relevant are considerations of how the ensemble aggregation is to be interpreted, e.g., as a 'representative vote/consensus/average' from a 'panel of experts' or (bootstrap) aggregation of *independent, weak* learners; or a *dependent* sequence of weak learners progressively correcting their predecessors' biases; respectively for measures of central tendency or location parameters, bagging, or boosting.

2.2.4 Evaluation criteria

The predictability of market returns is assessed by using two types of criteria: purely statistical criteria based on a relevant metric for out of sample model fit and economic criteria based on the obtained returns from trading on a given model's predictions.

¹⁴Timmermann [301] provides an overview of forecast combinations and Genre et al. [142] show that forecast combinations using a simple average often outperform methods that rely on estimated combination weights.

¹⁵In fact, Koren [182] aggregated model predictions using GBRT, in the winning solution of the Netflix Prize.

Statistical significance

To evaluate the predictive performance of the high-frequency market return forecasts, the out-of-sample R^2 metric proposed by Gu, Kelly and Xiu [152] is calculated. Given the market return, r_{t+1}^M , and a corresponding model prediction given the history up to time t , \hat{r}_{t+1}^M , the out-of-sample R^2 is calculated as

$$R_{\text{OOS}}^2 \equiv 1 - \frac{\sum_{t \in \text{Test}} (r_{t+1}^M - \hat{r}_{t+1}^M)^2}{\sum_{t \in \text{Test}} (r_{t+1}^M)^2} \quad (2.12)$$

Note that the R_{OOS}^2 metric is only calculated over the test samples, indexed by times t in the set Test. The denominator of R_{OOS}^2 is the sum of squared market returns without demeaning. As discussed by Gu, Kelly and Xiu [152], the historical mean underperforms a zero forecast. The historical mean return is noisy, resulting in artificially high estimates of R^2 . Hence, zero rather than the historical mean is taken as a benchmark.¹⁶

Following Gao et al. [133] and Chinco, Clark-Joseph and Ye [74], also considered are the estimated coefficients from running a simple predictive regression given by

$$\mathbb{E}r_{t+1}^M = \alpha + \beta \hat{r}_{t+1}^M \quad (2.13)$$

Here, r_{t+1}^M is the market return and \hat{r}_{t+1}^M the model prediction for five-minute time interval $t + 1$.

Economic significance

In addition to evaluating predictive performance, the ability of each model to time the market is assessed. A trading strategy is implemented that takes a long (short) position in the market if the model predicts a positive (negative) return. The profitability can be expressed as $\pi \equiv \sum_{t \in \text{Test}} \pi_t$ where the individual daily profits, π_t , are defined as

$$\pi_t(\hat{r}_{t+1}^M) \equiv \begin{cases} -r_{t+1}^M & \text{if } \hat{r}_{t+1}^M < 0, \\ r_{t+1}^M & \text{if } \hat{r}_{t+1}^M > 0, \\ 0 & \text{otherwise} \end{cases} \quad (2.14)$$

Reported are annualized excess arithmetic returns and Sharpe ratios of this trading strategy across the test sample. The Sharpe ratio is calculated as the monthly excess return divided by the corresponding standard deviation and scaled by $\sqrt{12}$.

The economic significance of the trading strategy could be completely driven by small return fluctuations such as the bid-ask bounce discussed by Roll [263], which is not useful to traders. To assess the models' ability to predict larger returns, a trading strategy is considered that accounts for transaction costs. Given national best bid (ask) price, Bid_t (Ask_t), and

¹⁶Benchmarking against the historical mean, the R^2 increases by approximately 0.01% across methods.

midquote $M_t = \frac{\text{Bid}_t + \text{Ask}_t}{2}$, transaction costs are estimated by the relative national-best bid-offer (NBBO) spread, given by

$$\text{Spread}_t = \frac{\text{Ask}_t - \text{Bid}_t}{M_t} \quad (2.15)$$

Following Chincó, Clark-Joseph and Ye [74], the economic significance of returns after transaction costs are evaluated by implementing a trading strategy that only trades when model predictions exceed the transaction costs. The returns of this trading strategy are

$$\phi_t(\hat{r}_{t+1}^M) \equiv \begin{cases} -r_{t+1}^M - \text{Spread}_t & \text{if } |\hat{r}_{t+1}^M| > \text{Spread}_t \text{ and } \hat{r}_{t+1}^M < 0, \\ r_{t+1}^M - \text{Spread}_t & \text{if } |\hat{r}_{t+1}^M| > \text{Spread}_t \text{ and } \hat{r}_{t+1}^M > 0, \\ 0 & \text{otherwise} \end{cases} \quad (2.16)$$

Again, reported are annualized excess returns and Sharpe ratios for this trading strategy. For models to be profitable, their predicted returns must be properly directed and exceed transaction costs, that is $\hat{r}_{t+1}^M > 0$ and $|\hat{r}_{t+1}^M| > \text{Spread}_t$ for a long position to be implemented. This simple trading strategy presents a significant hurdle for model validation and provides a benchmark for returns available to traders. A simple strategy is used to avoid data mining concerns, but more sophisticated strategies can yield higher returns.

2.3 Empirical results

The sequel presents results for the eight models considered: lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), artificial neural networks (ANN), and the ensemble (arithmetic) average (MEAN) and median (MED). Results for OLS are omitted from the analysis, it being found that the simple linear model is highly inaccurate, resulting in a negative out-of-sample R^2 in most periods. All models are estimated using Sparkling Water from H2O.ai, which combines the machine learning algorithms of H2O with the big data capabilities of Apache Spark, and permits efficient estimation of machine learning models on much larger datasets than the existing literature. Models are trained using all intraday and overnight observations, however results are reported from 9:35 to 3:55 only excluding the opening and closing returns to avoid concerns regarding the accuracy of these auction-based prices and overnight/weekend effects. For each model, hyperparameters are tuned in the validation set using random search on mean squared error.

The predictive performance of each model is evaluated by out-of-sample R^2 , R_{OOS}^2 , and the predictive regression coefficients and t -statistics. Economic significance is evaluated by the excess returns and Sharpe ratios of the market timing trading strategy with and without transaction costs. In each case, model performance is examined over the entire testing sample from April, 1993, to October, 2016. The results are also presented for significant sub-periods, including the 1/8 tick size sample from 1993 to 1996, the 1/16 tick size sample from 1997 to 2000, the early post-decimalization sample from 2001 to 2004, and the late post-decimalization sample from 2005 to 2016. Considered finally are the results' robustness to using models trained on 1, 4, 7, 10, 16, 22, 34, and 58 months of returns instead of the baseline expanding window.

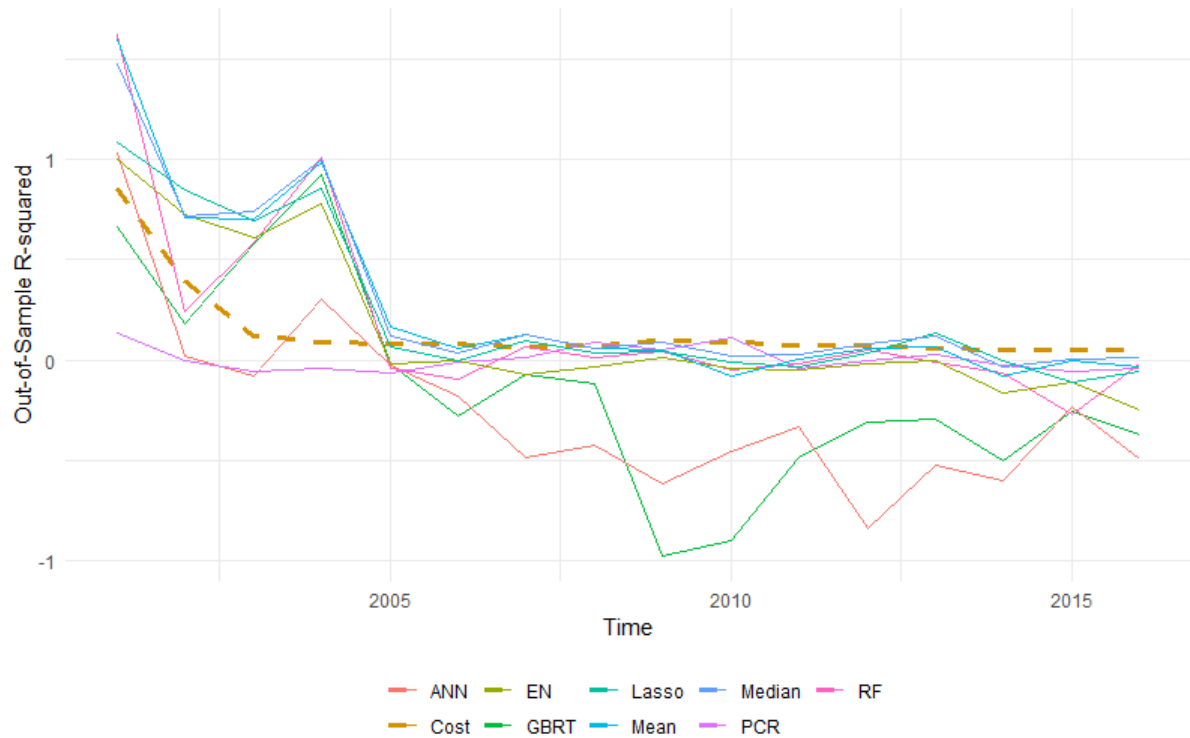
2.3.1 Market predictability

Table 2.1 reports coefficients from the predictive regression (2.13) along with Newey-West t -statistics with 79 lags, R^2_{OOS} percentages, and the ratio of R^2_{OOS} to mean transaction cost [$R^2_{\text{OOS}}/(\text{Bid-Ask Spread scaled by } 1,000)$] for the eight machine learning techniques considered. Panel A presents results for the entire testing sample from April 1993 to October 2016. Across models, slope coefficients are roughly 1 with most t -statistics exceeding 40. The magnitude of the intercept and slope coefficients are not exactly zero and one respectively, since the relationship between realized and predicted returns is non-linear (e.g., due to the prevalence of zero return observations in five-minute returns). Table 2.1 shows that all models predict the market better than a naive forecast of zero, since each of them have a positive R^2_{OOS} . The LAS and EN models have R^2_{OOS} of 2% and 1.95%, respectively. These regularized linear models have the highest R^2_{OOS} among non-ensemble models. Among non-linear models, RF performs best with an R^2_{OOS} of 1.71%. Surprisingly, ANN performs poorly despite doing well in Gu, Kelly and Xiu [152]. This difference is likely because these authors forecast using an ensemble of neural networks, whereas the present analysis is limited by computational time to a single neural network prediction each period, which typically has higher variance forecasts. Across all models, the ensemble mean and median perform best with respective R^2_{OOS} values of 2% and 2.01%. Note that most models have an R^2_{OOS} exceeding 1.6%, which is the R^2_{OOS} documented by Gao et al. [133] for 30-minute returns and roughly twice the monthly R^2_{OOS} reported by Gu, Kelly and Xiu [152]. Having a similar level of R^2_{OOS} at a higher frequency (like five-minutes in this chapter) than at a lower frequency is interesting since more trades can be carried out based on the predictability. The findings differ from those of Chordia, Roll and Subrahmanyam [78], who use linear models and find that lagged returns cannot forecast five-minute equity returns. The results reveal that by including lagged returns of market constituents using high-dimensional models, market returns are predictable using only lagged returns.

Next, the predictability is decomposed by time period based on observed structural breaks caused by changes to tick size and transaction costs. Panel B presents results for the early sample from 1993 to 1996, coinciding with a 1/8 tick size. This sample has the strongest predictability, with R^2_{OOS} exceeding 6% across all models and slope t -statistics exceeding 30. This finding makes sense intuitively, since predictability should be high when a large tick size prevents traders from bringing prices to their fundamental values. Also, trading volumes were relatively low and transaction costs were high during this time period, which further increased market frictions. Panel C presents results for the period with 1/16 tick size from 1997 to 2000. This period has the second highest predictability with R^2_{OOS} exceeding 3% for all models except PCR. The $R^2_{\text{OOS}}/\text{Cost}$ ratio decreased by roughly 90% relative to the previous period, demonstrating that most of the decreases in predictability were due to factors other than the decreases in transaction costs, and possibly due to technological improvements for traders (i.e., faster computer driven trading).¹⁷

Beginning in 2001, exchanges adopted decimal (\$0.01) trading ticks, which is the tick size in the remainder of the sample. For this reason, 2001 to 2016 is referred to as the *post-*

¹⁷The simple intuition for this is the following: if the decrease in R^2_{OOS} could be fully explained by a decrease in transaction costs, then the $R^2_{\text{OOS}}/\text{Cost}$ ratio should remain the same across periods.

Figure 2.1: R^2_{OOS} and trading costs

Plot of the post-decimalization R^2_{OOS} for the SPY using the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). Also plotted is the mean transaction cost scaled by 1,000.

decimalization period. Panel D reports results for the early post-decimalization period from 2001 to 2004. Coefficients are statistically significant at the 1% level across all models, and while the R^2_{OOS} are lower than pre-decimalization, they remain relatively high. Furthermore, the R^2_{OOS}/Cost ratio decreased by roughly 50% relative to the previous period, indicating that much of the decrease in predictability can be explained by decreases in transaction costs. The highest non-ensemble R^2_{OOS} are for the LAS, EN, and RF models at 0.91%, 0.81%, and 0.85%, respectively. The ensemble mean and median reach an R^2_{OOS} of 1.04% and 1.01%, respectively. These R^2_{OOS} percentages remain in line with the market predictability results of Gao et al. [133] and Gu, Kelly and Xiu [152] for the 30-minute and monthly time horizons, respectively. It should be noted that from 2001 to 2004 there was a significant decrease in transaction costs with no change to tick size. Figure 2.1 plots the mean transaction cost (Bid-Ask Spread scaled by 1,000) against the R^2_{OOS} for each model during the post-decimalization period. Consistent with the present hypothesis, the decrease in transaction costs in the early post-decimalization period significantly decreased the R^2_{OOS} across all models.

Finally, panel E reports results for the late post-decimalization period from 2005 to 2016. The LAS and PCR models have positive R^2_{OOS} of 0.02% and 0.04%, respectively, during this period.

Interestingly, the PCR model has a positive R_{OOS}^2 during this period despite having the lowest R_{OOS}^2 in all previous periods. This is likely because the subprime and European debt crises occurred (late) post-decimalization, and as a result of asset correlation increasing to near one in these crisis periods, returns had a strong factor structure.¹⁸ The EN, RF, GBRT, and ANN models have negative R_{OOS}^2 of -0.04%, -0.01%, -0.39%, and -0.45%, respectively, indicating that they perform worse than a naive constant prediction of zero. The poor performance of the non-linear models suggests they may be over-fitting a simpler predictive relationship during this recent period. The ensemble mean and median, however, have the highest R_{OOS}^2 of 0.04% and 0.06% respectively, benefiting from the strong performance of PCR during crisis periods. Thus, models which generally perform poorly may sometimes improve ensemble models.

At a first glance, it may appear that after 2005 all of the predictability is gone. However, it is argued that this is not the case. In particular, figure 2.1 shows that the R_{OOS}^2 for the two ensemble methods stays consistently positive, although it is small in magnitude, and the ensemble median R_{OOS}^2 is positive in every year except for 2015. Relative to the previous period, the R_{OOS}^2 decreased by roughly 95%, but the $R_{\text{OOS}}^2/\text{Cost}$ ratio only decreased by 33-66%, indicating that some of the decrease in predictability can be explained by decreases in transaction costs. Given the substantial decrease in transaction costs for five-minute returns, a high R_{OOS}^2 is unreasonable, and a small but positive R_{OOS}^2 is expected. As argued by Campbell and Thompson [57] and Rapach and Zhou [256], even a small R^2 can generate economically large portfolio returns. This is especially true given the number of trading opportunities at a five-minute interval.

In summary, the first set of results demonstrate that the market is remarkably predictable at the five-minute frequency. This predictability is highest prior to the decimalization of exchanges in 2001. Post-decimalization, markets became faster in integrating lagged intraday information. However, the R_{OOS}^2 remains positive for the LAS, PCR, and ensemble models, demonstrating that some predictability persists even after the decimalization of exchanges and indicating that decreases in transaction costs also decreased other market frictions.

2.3.2 Economic significance

As noted, even a small predictability at five-minute intervals can result in large and economically significant returns. This section therefore evaluates the economic significance of the trained models' predictions by implementing the simple trading strategy specified in equation (2.14). Since the models are optimized to minimize forecasting error, the economic significance of forecasts provides an indirect evaluation of model performance. Table 2.2 presents the annualized excess returns, Newey-West t -statistics with one lag and Sharpe ratios, denoted SR in the table, of the market timing strategy. Columns (1) - (8) present results for the eight machine learning models and columns (9) and (10) report the intraday SPY returns as well as the benchmark buy-and-hold SPY returns, respectively, for comparison.

Panel A of table 2.2 reports the results for the entire sample from 1993 to 2016. Rankings across methods are mostly consistent with their R_{OOS}^2 percentages, and all models have positive

¹⁸In section 2.4.3 it is confirmed that the PCR model indeed performs well during both of these crisis periods.

returns and Sharpe ratios. Among the non-ensemble models, LAS, EN, RF, and GBRT have the highest R^2_{OOS} and also have high returns (Sharpe ratios) of, respectively, 191%, 188%, 198%, and 192% (2.71, 2.68, 2.90, and 2.84), indicating that dimension reduction is important for predicting returns. In particular, the tree-based RF and GBRT have the highest non-ensemble returns (Sharpe ratios), showing that modeling non-linearities and interaction effects is important for return prediction. The ANN model has lower returns (Sharpe ratios) of 172% (2.66) due to its high variance predictions. PCR has positive but relatively low excess returns of 75% matching its low R^2_{OOS} over the entire sample. Consistent with their R^2_{OOS} percentages, the ensemble mean and median have the highest returns (Sharpe ratios) of, respectively, 205% and 204% (2.90 and 2.82), illustrating that combining forecasts yields higher economic predictability. It is noteworthy that every model significantly out-performs the benchmark buy-and-hold SPY strategy, which yields excess returns (Sharpe ratios) of 7% (0.48).

Next, panel B of table 2.2 reports results for the 1/8 tick period from 1993 to 1996. The excess returns (Sharpe ratios) of all models exceed 300% (5). The large economic returns are consistent with the large R^2_{OOS} observed for this period. Panel C reports results for the 1/16 tick period from 1997 to 2000. While table 2.1 showed that the R^2_{OOS} decreased relative to the previous period, surprisingly returns and Sharpe ratios increased during this period for nearly all models. This volatile period contains both the Asian crisis and the dot-com bust, and in section 2.4.3 it is shown that the economic significance of model predictions increase during financial crises. PCR has a highly negative R^2_{OOS} during this period resulting in a low predictive return, suggesting that there is not a strong factor structure during this period.

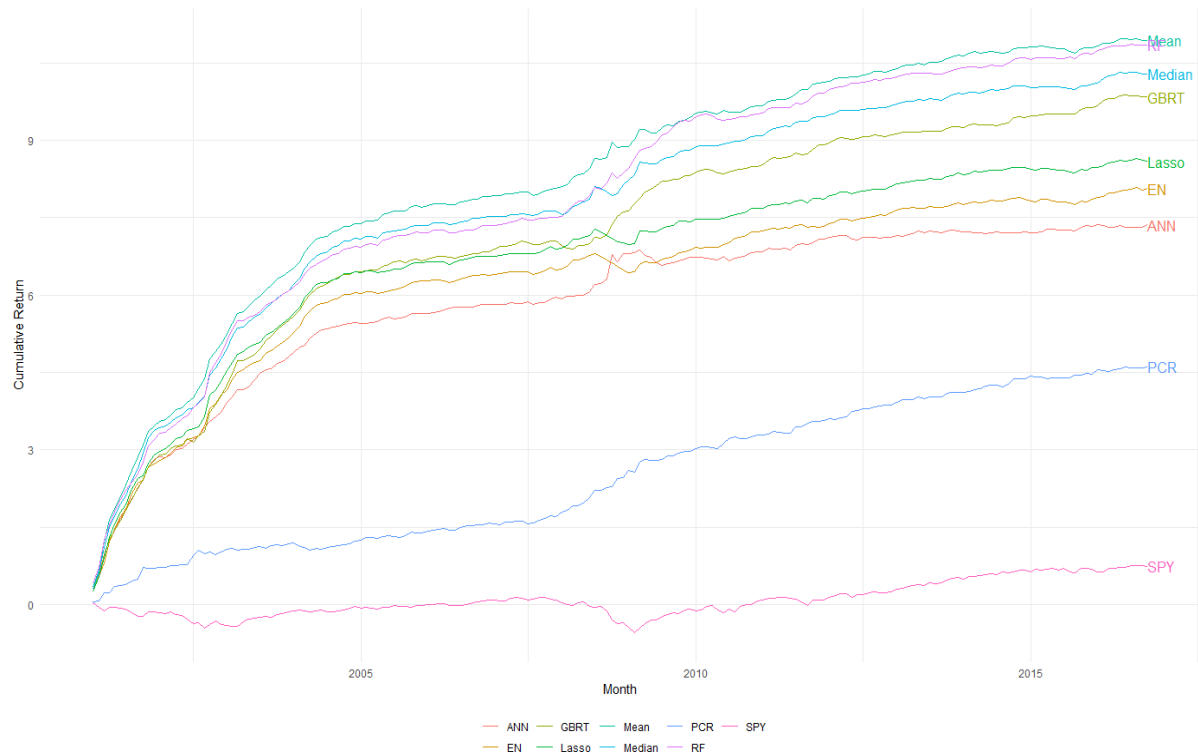
Considered finally are the results in the post-decimalization period. Panel D reports results for the early post-decimalization period from 2001 to 2004 and shows that, consistent with the finding that the R^2_{OOS} decreased post-decimalization, excess returns and Sharpe ratios, although significant and larger than the benchmark buy-and-hold SPY returns, are lower than in previous periods. Panel E reports results for the late post-decimalization period from 2005 to 2016. Table 2.1 showed that this period had a substantial decrease in R^2_{OOS} , and consequently returns and Sharpe ratios are significantly lower than in previous periods. However, returns and Sharpe ratios remain high relative to the benchmark buy-and-hold SPY. In particular, all models have returns exceeding 16% and most models have Sharpe ratios exceeding one. In comparison, the buy-and-hold SPY returns (Sharpe ratios) were 7% (0.46).

Figure 2.2 plots the cumulative excess returns for each model and of the SPY from 2001 to 2016. Every model has higher cumulative returns than the market portfolio with the RF and ensemble models having significantly higher cumulative returns. In summary, the consistency of the results across models supports the hypothesis that the intraday market is predictable and demonstrates that the predictability is economically significant. Next it is shown that profitability remains even with (large) transaction costs.

2.3.3 Economic significance after transaction costs

The established economic significance does not account for transaction costs, which are substantial when trading at a five-minute frequency. If models are only accurate for small return

Figure 2.2: Cumulative returns by model



Plot of the post-decimalization cumulative excess returns of the market-timing strategy for the SPY using the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED).

fluctuations but cannot forecast large returns, then they are not useful to traders. This section shows that the predictability of the market portfolios is economically significant even after accounting for transaction costs. It is additionally documented that after 2005, R_{OOS}^2 , excess returns, and transaction costs significantly decreased at the same time, so it is not obvious if model predictions remain profitable after accounting for transaction costs in this recent sample.

Table 2.3 presents the annualized excess returns and Sharpe ratios of the market timing strategy with transaction costs along with the average percentage of executed trades. Panel A of table 2.3 reports the results for the entire sample from 1993 to 2016. Columns (1) - (8) present results for the eight machine learning models and columns (9) and (10) report the intraday SPY and the benchmark buy-and-hold SPY returns respectively. As expected, all returns are significantly lower due to transaction costs and infrequent trading. However, even after accounting for transaction costs, every model (with the exception of ANN) has positive returns. Among non-ensemble models, PCR and RF have the highest returns (Sharpe ratios), respectively yielding 5% and 6% (0.68 and 0.77). The two ensemble models have high returns (Sharpe ratios), both yielding 6% (0.67 and 0.98). As a benchmark, a buy-and-hold SPY strategy has 7% return and a Sharpe ratio of 0.48. Thus, even after transaction costs, the PCR, RF, and ensemble mod-

els outperform holding the market.¹⁹ These findings demonstrate that such models can predict large returns that exceed the transaction costs very well. The results are similar in magnitude to the strategies in Gao et al. [133] and Chinco, Clark-Joseph and Ye [74] that have after-transaction cost annualized returns of 4.46% (Sharpe ratio of 0.98) and 4.92%, respectively. Consistent with the previous sections, ANN predictions have a high variance and hence fail to forecast large returns. Similarly, the linear LAS and EN models had high R^2_{OOS} , but the statistical predictability did not translate well into economic profitability, earning returns (Sharpe ratios) after transaction costs of 3% and 1% (0.42 and 0.20), respectively. It is shown in section 2.4.3 that LAS and EN had highly negative returns during the subprime crisis, since these models generally perform worse when predictors are highly correlated (Wang et al. [308]).²⁰

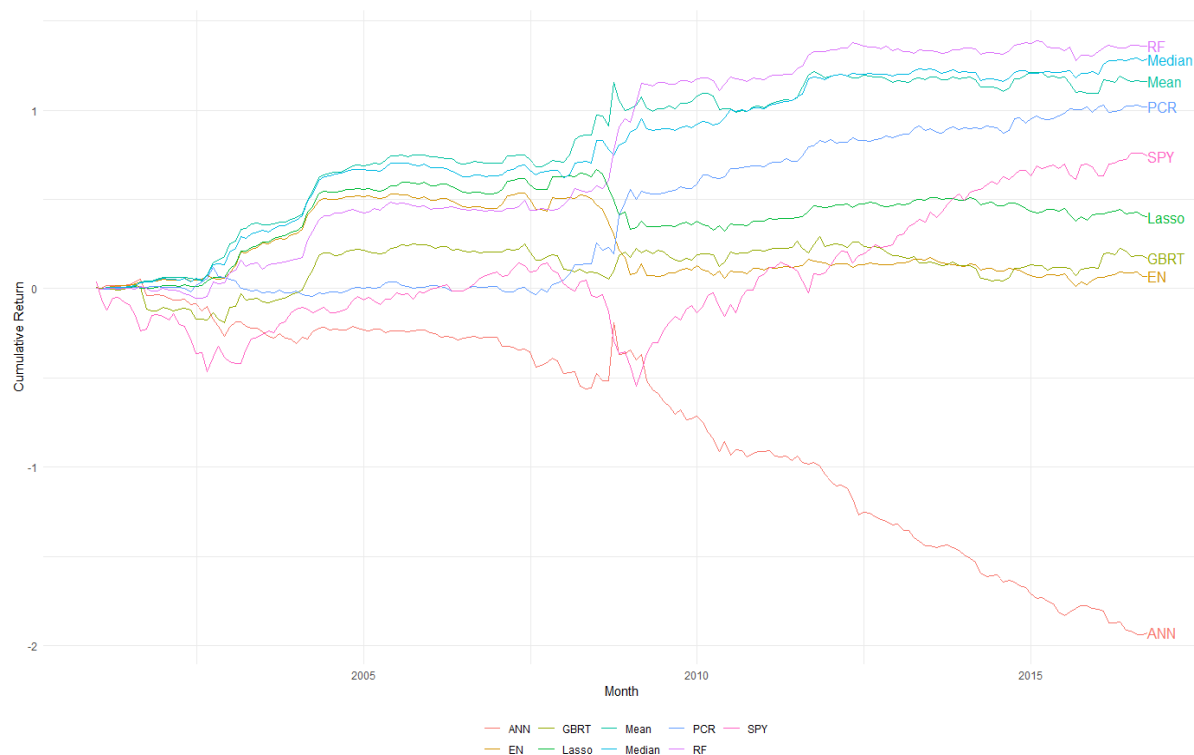
Next, panel B of table 2.3 reports results for the 1/8 tick period from 1993 to 1996. The models' returns are low since they only trade roughly 1% of the time. However, the large and positive Sharpe ratios of at least 0.8 indicate model predictions are economically significant even after paying transaction costs. The LAS, EN, GBRT, and ensemble models have Sharpe ratios exceeding two and so the benchmark buy-and-hold SPY Sharpe ratio of 1.6. Panel C of table 2.3 reports results for the 1/16 tick period from 1997 to 2000, where models trade less than 0.25% of the time due to large transaction costs. Whereas table 2.2 showed that pre-transaction cost returns are higher than in the previous period, table 2.3 shows that after accounting for transaction costs returns (Sharpe ratios) are lower than in the previous period. LAS, EN, PCR, RF, ANN, and the ensemble models, though, continue to have higher Sharpe ratios than the buy-and-hold benchmark Sharpe ratio of 0.71.

Finally, panels D and E report results for the post-decimalization period from 2001 to 2016, where models trade much more frequently due to lower transaction costs. Panel D reports results for the early post-decimalization period from 2001 to 2004 and shows that excess returns (Sharpe ratios) increase for the LAS, EN, RF, GBRT, and ensemble models relative to the period before 2001 due to the significantly lower transaction costs. All models, except ANN, earn much higher returns (Sharpe ratios) than the benchmark -0.01% (-0.07) for the buy-and-hold SPY. Panel E of table 2.3 reports results for the late post-decimalization period from 2005 to 2016. Previous sections documented that this period had substantially lower R^2_{OOS} returns, and Sharpe ratios relative to previous periods. However, due to the decrease in transaction costs, the after-transaction cost returns and Sharpe ratios are still high. During this period, the

¹⁹When regressing the trading strategy returns after transaction costs onto the benchmark buy-and-hold SPY it is found that the alphas are positive and statistically significant at the 10% level for all models except ANN. The betas, on the other hand, are mostly small, and if they are large and significant, they tend to be negative indicating that the trading strategy if anything hedges systematic risk. Thus, there is strong evidence that the trading strategy out-performs the benchmark portfolio and little evidence that the models are simply buying systematic risk.

²⁰Unreported results confirm that when correlations among constituent returns increase LAS and EN perform worse while PCR performs better. Intuitively, LAS and EN may encounter difficulties when stocks are highly correlated. For example, if several highly correlated stocks are relevant for prediction, then LAS may only select one from the group and shrinks the rest to zero (Zou and Hastie [329]). EN mitigates this issue by using the ridge penalty. However, the ridge penalty forces the estimated coefficients of highly correlated predictors to be close together, which is problematic since the coefficients on the predictor stocks likely have different magnitudes or different signs (Wang et al. [308]). Conversely, PCR has stronger predictability when correlations increase, since the model creates new predictors that summarize the variation of the constituent stocks.

Figure 2.3: Cumulative returns after transaction costs by model



Plot of the post-decimalization cumulative excess returns of the market-timing strategy with transaction costs for the SPY using the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED).

benchmark buy-and-hold SPY earned 7% returns with a 0.46 Sharpe ratio. The PCR and RF models out-perform the benchmark with returns (Sharpe ratios) of 9% and 8% (0.88 and 0.81), respectively, which was shown in section 2.4.3 to be partially due to their strong performance during the subprime crisis. The ensemble median also beat the buy-and-hold SPY with returns (Sharpe ratios) of 5% (0.68).²¹

Figure 2.3 plots the cumulative excess returns after transaction costs for each model and of the SPY from 2001 to 2016. After accounting for transaction costs, the PCR, RF and ensemble models have higher cumulative returns relative to the market, despite only trading infrequently. Even accounting for the drop in R_{OOS}^2 after 2005, the after-transaction cost returns remain consistently large. That is, even after accounting for transaction costs, the considered models earn

²¹These results are robust to assuming fixed transaction costs of, say, 0.01% or 0.1% instead of the bid-ask spread (the median post-decimalization spread was 0.008%). In particular, with 0.01% fixed transaction costs economic profitability generally increases across models and though the models seldom trade with a 0.1% fixed transaction cost Sharpe ratios remain positive for the LAS, EN, and ensemble models. This verifies that the results are not driven by models that only trade when the spread is low.

economically significant returns with low variance. Demonstrated are economic gains available to traders using such model forecasts supporting the hypothesis that markets are predictable at the five-minute frequency. However, the market is notably less predictable post-decimalization, particularly after 2005, as expected, evidenced by the lower returns and Sharpe ratios.

2.3.4 Robustness to training sample size

The baseline model uses an expanding window for training and one month each for validation and for out-of-sample testing. If the documented predictive relationship is stable, then forecasting accuracy should be increasing in training sample size, since more observations yield lower variance forecasts. However, financial time series are notorious for containing structural breaks, time-varying volatility, and other nonstationarities (Timmermann [302]). A shorter training sample may therefore outperform a longer one if the empirical model in equation (2.2) changes due to this nonstationarity, in which case using a longer training sample may yield biased forecasts. Tested are the importance of various training periods by evaluating R_{OOS}^2 and after-transaction cost Sharpe ratios using 58-, 34-, 22-, 16-, 10-, 7-, 4-, and 1-month samples for training and compare the results with the baseline training duration.

Panel A of table 2.4 reports the post-decimalization R_{OOS}^2 of the eight machine learning models using different training window lengths. This sample includes several possible structural breaks, including the subprime and EU debt crises. Across nearly all models, R_{OOS}^2 is increasing in training size, indicating that the predictive relationship is mostly stable. However, the R_{OOS}^2 do not increase monotonically, which demonstrates that non-stationarities do have some effect on forecasts. This is particularly apparent comparing the 58-month estimation window to expanding, suggesting that results could be improved by starting the expanding window later in the sample. Furthermore, the R_{OOS}^2 are positive for the ensemble models across nearly all training periods and positive for most individual models. Panel B of table 2.4 reports the after-transaction cost Sharpe ratios using different training window lengths for the post-decimalization period. Consistent with the results for the R_{OOS}^2 , model Sharpe ratios are mostly increasing in training size across models.

The results in table 2.4 first of all show that the predictability findings for intraday market returns are largely robust to using different training window specifications. Secondly, they importantly show that predictability increases with the training window size. Chincó, Clark-Joseph and Ye [74] theorize that market predictability could be driven by very short-term sparse signals. The results indicate instead that predictability may be consistently exploiting inefficiencies across time and is not necessarily driven by infrequent signals.

2.4 Additional analysis

The hypothesis for intraday predictability is based on slow-moving trader capital. This section examines whether the results differ within the trading day, in periods of high versus low volatility or illiquidity, and during periods of financial crisis. Finally, the baseline model performance

is compared to that of autoregressive models for the market return and it is considered whether forecasting accuracy may be improved by including additional lagged variables.

2.4.1 Time-of-day patterns

This section tests the intraday implications of slow traders on intraday predictability. Since traders are most active at the beginning and end of each day, predictability is expected to be low during those times. Also, since intraday trading volume exhibits a ‘U’ shape, predictability is expected to exhibit an inverse-U shape.

Panel A of table 2.5 reports the R_{OOS}^2 of each half-hour interval in the post-decimalization period. For every model, the R_{OOS}^2 is low in the first half hour and last hour of the day, when traders are most active. On the other hand, these models have high R_{OOS}^2 between 10:00 - 15:00, demonstrating that nearly all of the predictability occurs during the middle of the day when traders are less active. Panel B of table 2.5 reports after-transaction cost Sharpe ratios of each half-hour interval in the post-decimalization period. For most models, the Sharpe ratios are highest between 10:00 - 15:00, demonstrating that the increased R_{OOS}^2 translates into economically meaningful returns after transaction costs. The Sharpe ratio for the intraday SPY in column (9) does not display the same pattern, showing that the models are not simply buying the SPY portfolio.

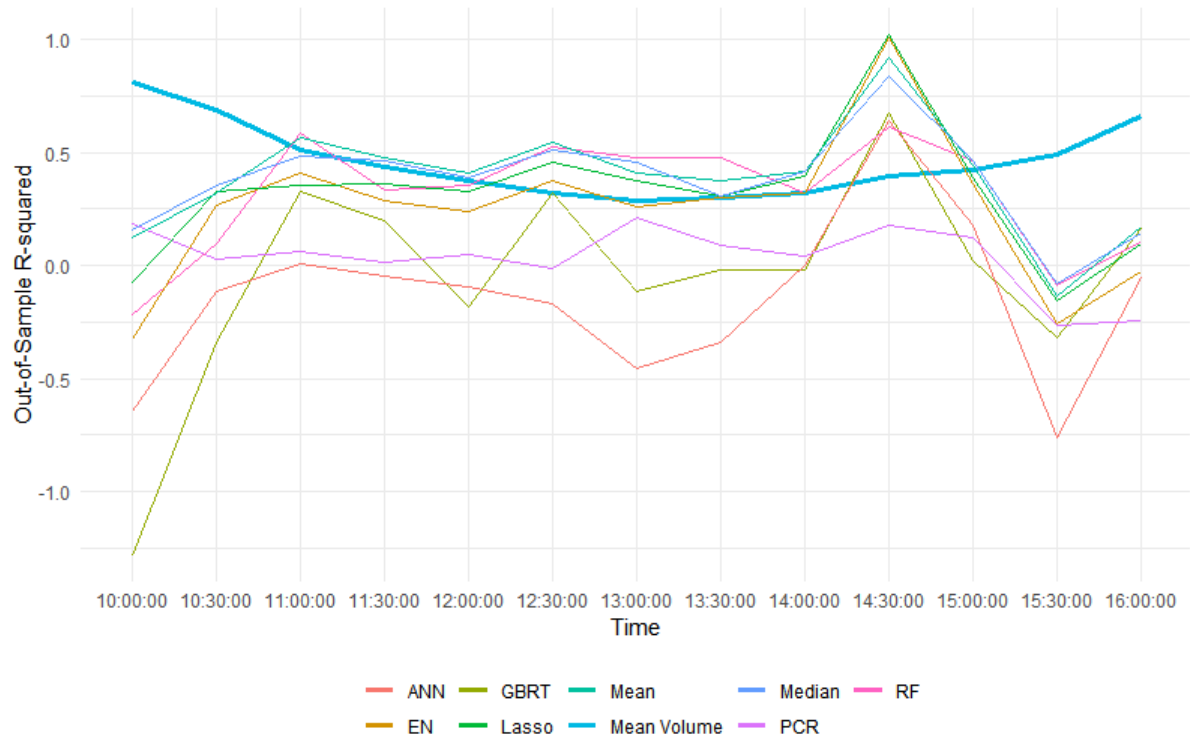
Figure 2.4 plots the R_{OOS}^2 and mean trading volume of each half-hour interval in the post-decimalization period. Across all models (except ANN), there is a remarkably similar inverse-U pattern, suggesting that these models are approximating the same function. Every model’s R_{OOS}^2 jumps up between 14:00 - 14:30, suggesting that this is the least active time for traders. Every model’s R_{OOS}^2 jumps down between 15:00 - 15:30, suggesting that traders are active at this time, which is consistent with the finding of Lou, Polk and Skouras [212], that institutional investors tend to initiate trades near the close. The discovery that predictability is stronger when traders are less active is consistent with the hypothesis that the predictability is driven by slow traders.

Gao et al. [133] find that returns in the first half hour forecast returns in the last half hour. The results in table 2.5 show that the predictability of the models are orthogonal to their findings and complement their paper meaningfully. Economically, there are likely several sources of risk and trading behaviors driving intraday patterns in returns.

2.4.2 Volatility and illiquidity effects

During periods of high volatility or illiquidity (i.e., periods of low liquidity), traders encounter higher market frictions. According to the slow trader hypothesis, predictability should be higher when volatility or illiquidity is higher. This hypothesis is tested by first sorting days into three equal groups (tertiles) based on their daily realized volatility and Amihud [12] measure of illiquidity, respectively. The predictability is then studied individually for each tertile.

Panel A of table 2.6 reports the R_{OOS}^2 of each volatility group during the post-decimalization

Figure 2.4: R^2_{OOS} and trading volume

Plot by time of day of the post-decimalization R^2_{OOS} for the SPY using the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). Also plotted is the mean SPY trading volume.

period. Consistent with the hypothesis, R^2_{OOS} is strictly increasing in volatility across every model (except PCR). Likewise, panel B of table 2.6 shows that after-transaction cost Sharpe ratios are increasing in volatility for most models. Note that column (9) shows that the intraday SPY Sharpe ratio decreases in volatility, so model Sharpe ratios relative to the benchmark are strongly increasing across the volatility tertiles.²²

Panel A of table 2.7 reports the R^2_{OOS} of each illiquidity group during the post-decimalization period. Interestingly, R^2_{OOS} is small for low- and mid-, but very large for high-illiquidity days. This suggests that most of the predictability occurs during days when the market is illiquid and when it's more expensive for traders to rebalance their portfolios. Panel B of table 2.7 reports after-transaction cost Sharpe ratios of each illiquidity group during the post-decimalization period. For most models, Sharpe ratios are increasing in illiquidity.

²²To examine this further, a volatility timing strategy was considered that trades only when the previous day's volatility is in a given tertile group using only past information to create the groups, i.e., with no look-ahead bias. The results show that Sharpe ratios are increasing in volatility for all models except LAS and EN, demonstrating that the ex-post volatility analysis could be converted into a tradeable volatility timing strategy.

In summary, predictability and economic significance increase when market volatility and illiquidity are high. These findings are similar to Gao et al. [133] who show that 30-minute predictability is higher on high volatility and illiquidity days. The results are consistent with their findings and with the hypothesis that market predictability is driven by slow-moving capital from traders that face market frictions. These findings also support the argument in Chordia, Roll and Subrahmanyam [78] and [79] that prices can be separated from fundamentals and are predictable in short horizons due to insufficient liquidity.

2.4.3 Impact of financial crisis

The recent 2005 to 2016 period contains the subprime mortgage and European debt crises. These crisis periods are associated with significant market frictions. According to the hypothesis, predictability should be higher during these crisis periods. This is tested in this section by studying several interesting sub-periods of the late post-decimalization period illustrating crisis and non-crisis periods.

Panel A of table 2.8 reports the R_{OOS}^2 of crisis and non-crisis periods. Across most models, the R_{OOS}^2 during the subprime and EU crises are higher than during the 2010 - 2011 and 2014 - 2016 periods. However, the R_{OOS}^2 was high during the earliest 2005-2007 period, possibly due to technological factors (i.e., less sophisticated trading). PCR had the highest R_{OOS}^2 during the subprime crisis and a positive R_{OOS}^2 during the EU debt crisis. This finding confirms the assertion in section 2.3.1 that PCR has stronger predictability when equities are more correlated and follow a strong factor structure, as is the case during a financial crisis. From 2014 - 2016, model predictability is negative across models. This may be because the period was relatively stable, but could also be caused by reduced trader frictions.²³ These results demonstrate that the predictability is higher during a financial crisis. In the slow moving theory of capital, traders may not invest in arbitrage opportunities if they have insufficient capital or can invest in assets with higher expected returns (Duffie [101]). These results suggest that during financial crises, the market portfolio may be particularly predictable due to slow moving capital, findings similar to Gao et al. [133], who demonstrate increased 30-minute predictability during the subprime crisis.

The economic significance of crisis and non-crisis periods is somewhat mixed. Panel B of table 2.8 reports the pre-transaction cost Sharpe ratios of crisis and non-crisis periods. The pre-transaction cost Sharpe ratios are mostly consistent with the R_{OOS}^2 results, with the exception of the 2010 - 2011 period which earned higher Sharpe ratios than expected given the low R_{OOS}^2 . Panel C of table 2.8 reports the after-transaction cost Sharpe ratios. Surprisingly, the results differ from the findings in panels A and B for the 2005 - 2007 period and the EU debt crisis period. It appears that high transaction costs during these periods removed most of the predictive profits. Similarly, the 2010 - 2011 period had the highest after-transaction cost Sharpe ratios, potentially due to reduced transaction costs after the subprime crisis. Interestingly, the LAS and EN models had extremely low returns during the financial crisis, likely because these

²³It will be interesting to study if predictability increases during the ongoing Coronavirus crisis once high-frequency data becomes available for this period.

models often perform poorly with highly correlated predictors (cf. section 2.3.3).

2.4.4 Comparison to autoregressive models

A natural comparison to the cross-sectional models with lagged intraday returns for the market constituents are autoregressive models for the market return itself, since the predictability results could simply be capturing intraday momentum. For example, Heston, Korajczyk and Sadka [160] find significant auto-correlation of half-hour returns at daily intervals and Gao et al. [133] find that the first half-hour return of the SPY predicts the last. This section evaluates the predictability of linear models estimated on up to 500 lagged SPY returns using the same specifications as the baseline machine learning models. Considered are a simple AR(1) that uses the SPY return with one lag (i.e., the baseline model without constituent returns) and an AR(p) model where the number of lagged returns (up to 500) is chosen to minimize the validation mean squared error. Also considered are the linear LAS, EN, and PCR models to perform dimension reduction on the 500 lagged SPY returns.

Panel A of table 2.9 reports the R^2_{OOS} of the linear autoregressive models together with the linear baseline models for comparison. During the overall period from 1993 to 2016, LAS and EN generally have the highest R^2_{OOS} among AR models of 0.6% and 0.58%, suggesting that certain market return lags contain predictive information. However, these regularized models only slightly improve on the simplest AR(1) model's R^2_{OOS} of 0.55%, indicating that the most recent lag is the most important for prediction. This is reinforced by the AR(p), which uses a median of 29 lags and performs worse than the AR(1) model with an R^2_{OOS} of 0.53%. The AR(500) model without dimension reduction, i.e. estimated with OLS, performs poorly due to the high-dimensional inputs as expected. In comparison, the baseline LAS and EN models significantly outperform every AR model with an R^2_{OOS} of 2% and 1.95% respectively. Additionally, the baseline PCR's R^2_{OOS} of 0.31% is higher than the AR(500) PCR model's R^2_{OOS} of -0.07%. These results demonstrate that there is significant predictive information embedded in the lagged returns of the S&P 500 constituents.

Panel B of table 2.9 reports the after-transaction cost Sharpe ratios of the linear autoregressive models together with the linear baseline models for comparison. During the overall period from 1993 to 2016, every AR model has a negative Sharpe ratio except for the AR(1), which has a Sharpe ratio of zero. In contrast, the baseline linear models all have positive Sharpe ratios, with the PCR's Sharpe ratio of 0.68 even outperforming the buy-and-hold S&P 500's Sharpe ratio of 0.48. These results demonstrate that using the lagged returns of the S&P 500 constituents is necessary for improving the economic significance of the predictions. Panel B also shows that this holds true for all considered sub-periods.

2.4.5 Effect of additional variables

Finally, it is considered whether including additional lagged characteristics of the S&P 500 constituents may improve forecasting accuracy. Due to computational constraints (both in terms of memory requirements and training time), only considered are results using the four-month training sample during the post-decimalization period. Characteristics that proxy short-

term changes in liquidity and trading trends are considered. The characteristics include firm-level market beta, momentum, maximum, minimum, volatility, illiquidity, trading volume, kurtosis, and skewness calculated over preceding days as well as the lagged observed bid-ask spread.

Table 2.10 reports R_{OOS}^2 when including different characteristics during the post-decimalization period. The first row reports R_{OOS}^2 for the baseline model using lagged returns only. Panel A shows that the LAS model is improved by including illiquidity, momentum, or skewness, indicating that there may be some index constituents with useful characteristics for predicting the market return. PCR is also slightly improved by adding momentum and volume. However, for most models including any single characteristic reduces the R_{OOS}^2 . One possibility for the poor performance of the price trend and liquidity characteristics is that they are estimated over the preceding day and potentially noisy. However when the lagged bid-ask spread is included, which is not estimated, the R_{OOS}^2 also decreases across most models. These results may suggest that including additional characteristics simply adds noise to the model without providing additional predictive information. While adding these characteristics generally does not improve market forecasts beyond the baseline model, it remains possible that such characteristics may help predict individual stock returns.

Panel B of table 2.10 reports the after-transaction cost Sharpe ratios when including different characteristics during the post-decimalization period. This panel shows, similarly to panel A, that including some characteristics may slightly improve the Sharpe ratios of certain models. For example, LAS, EN, PCR, RF, ANN, and the ensemble median can be slightly improved by including some characteristics. In particular, the median ensemble after including the bid-ask spread achieves a Sharpe ratio of 0.44. But these results are not robust and could be by chance.

One concern is that the increased data dimension from additional characteristics may increase the estimation error. However, the predictability of only the characteristics was analyzed, removing lagged returns from the model, and again found no evidence of predictability. In summary, little evidence is found that including other characteristics can improve predictions. Among the tested characteristics, only lagged returns consistently predict market returns.

2.5 Conclusion

This chapter apparently conducts the largest existing study of intraday market return predictability using state-of-the-art machine learning models trained on the cross-section of lagged market index constituent returns and other characteristics to forecast five-minute market returns over the longest possible time period. The chapter demonstrates that there is significant statistical predictability of intraday market returns and establishes that this return predictability translates into economically significant profits even after accounting for transaction costs. Furthermore, the lagged constituent returns are shown to hold significant predictive information that is not contained in lagged market returns or in lagged price trend and liquidity characteristics.

Specifically, the chapter shows that regularized linear models such as lasso and elastic nets and nonlinear tree-based models such as random forests yield the largest positive out-of-sample R^2 percentages. Linear models such as principal component regression had high out-of-sample R^2 percentages during the subprime and EU debt crises, providing returns that hedge these crisis states. Ensemble models that combine individual model predictions perform the best across time and the return predictability from these models translates into economically significant profits with Sharpe ratios after transaction costs of 0.98. This Sharpe ratio is much higher than what is obtained from holding the index intraday and significantly exceeds the Sharpe ratio of the benchmark buy-and-hold strategy.

Across time, it is shown that the statistical predictability has suffered somewhat as transaction costs were reduced post-decimalization. It is argued that this strongly suggests that predictability could be a result of slow-moving trader capital. Consistent with the hypothesis of slow traders, market returns are also shown to be more predictable during the middle of the day when trading activity is lower, on days with high volatility or high illiquidity where prices can be driven further away from their fundamental values, and during years of financial crisis which are plagued by market frictions. Nevertheless, the best ensemble models retain some predictability and trading based on these models' signals remains profitable throughout the sample, even after adjusting for transaction costs.

The results provide strong evidence that market returns are predictable over short horizons. The empirical findings suggest that further investigation into the economic mechanisms driving such short-horizon predictability is warranted. The late-informed investor explanation in Gao et al. [133] is supported by the evidence. Another explanation of Chinco and Fos [75] theorizes that computational complexity of traders' rebalancing introduces noise. It is believed this can explain some of the predictability results, wherein models appear able to capture the systematic behaviour of traders' rebalancing. In particular, this could explain the persistent profitability of the studied models in recent years. However, verifying these economic mechanisms requires further investigation which is left for future research.

Table 2.1: Market predictability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED
Panel A: Overall						Period (1993 - 2016)		
Intercept	-0.64	-0.59	-0.31	-0.45	-0.36	-0.32	-0.64	-0.67
<i>t</i> -stat	-4.02	-3.72	-1.98	-2.84	-2.28	-1.98	-4.04	-4.24
Slope	1.38	1.29	0.65	1.22	0.90	0.83	1.53	1.54
<i>t</i> -stat	47.26	46.06	24.14	54.11	43.23	42.49	58.28	58.97
R^2_{OOS}	2.00	1.95	0.31	1.71	1.64	1.40	2.00	2.01
$R^2_{\text{OOS}}/\text{Cost}$	0.12	0.11	0.02	0.10	0.09	0.08	0.12	0.12
Panel B: 1/ 8 Tick Size						Period (1993 - 1996)		
Intercept	-0.22	-0.17	-0.16	-0.14	-0.32	0.16	-0.21	-0.26
<i>t</i> -stat	-1.16	-0.90	-0.95	-0.84	-1.68	0.79	-1.29	-1.59
Slope	1.55	1.45	0.92	1.70	1.18	1.17	1.59	1.65
<i>t</i> -stat	36.53	35.80	44.92	41.48	38.60	41.24	42.96	41.07
R^2_{OOS}	9.27	9.32	6.52	6.78	8.95	7.41	9.16	9.14
$R^2_{\text{OOS}}/\text{Cost}$	1.08	1.09	0.76	0.79	1.05	0.87	1.07	1.07
Panel C: 1/16 Tick Size						Period (1997 - 2000)		
Intercept	-1.71	-1.67	-0.61	-1.28	-1.22	-1.04	-1.67	-1.73
<i>t</i> -stat	-3.91	-3.82	-1.40	-2.88	-2.71	-2.23	-3.80	-3.93
Slope	1.42	1.38	0.18	1.27	1.05	0.92	1.62	1.56
<i>t</i> -stat	35.52	35.31	2.77	46.73	41.86	38.57	43.56	45.09
R^2_{OOS}	3.72	3.69	-0.76	3.48	3.45	3.21	3.60	3.66
$R^2_{\text{OOS}}/\text{Cost}$	0.07	0.07	-0.01	0.06	0.06	0.06	0.07	0.07
Panel D: Early Post-Decimalization						Period (2001 - 2004)		
Intercept	-0.65	-0.57	-0.42	-0.25	-0.30	-0.14	-0.55	-0.57
<i>t</i> -stat	-1.40	-1.24	-0.91	-0.55	-0.68	-0.32	-1.19	-1.24
Slope	1.14	0.99	0.59	0.87	0.63	0.61	1.36	1.38
<i>t</i> -stat	10.52	9.67	4.67	16.09	13.46	10.87	15.12	14.65
R^2_{OOS}	0.91	0.81	0.03	0.85	0.48	0.38	1.04	1.01
$R^2_{\text{OOS}}/\text{Cost}$	0.03	0.03	0.00	0.03	0.02	0.01	0.03	0.03
Panel E: Late Post-Decimalization						Period (2005 - 2016)		
Intercept	-0.10	-0.03	-0.12	-0.07	0.08	0.07	-0.14	-0.20
<i>t</i> -stat	-0.43	-0.11	-0.52	-0.30	0.35	0.29	-0.60	-0.84
Slope	0.58	0.38	0.64	0.48	0.13	0.10	0.68	0.82
<i>t</i> -stat	3.70	3.17	4.80	4.85	1.53	1.25	4.02	4.59
R^2_{OOS}	0.02	-0.04	0.04	-0.01	-0.39	-0.45	0.04	0.06
$R^2_{\text{OOS}}/\text{Cost}$	0.01	-0.01	0.01	0.00	-0.12	-0.14	0.01	0.02

Reported are coefficients from the predictive regression (2.13) along with Newey-West *t*-statistics with 79 lags, R^2_{OOS} percentages, and R^2_{OOS} scaled by transaction cost for the SPY using the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), and ensemble mean (MEAN) and median (MED). The predictors used are lagged returns for the market and all the S&P 500 constituents. Results are reported for the full, 1/16 tick size, 1/8 tick size, early post-decimalization, and late post-decimalization samples.

Table 2.2: Excess Returns and Sharpe ratios

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	Intraday SPY	Hold SPY
Panel A: Overall									Period (1993 - 2016)	
Return	1.91	1.88	0.75	1.98	1.92	1.72	2.05	2.04	-0.03	0.07
<i>t</i> -stat	4.81	4.92	4.62	4.50	4.26	5.11	4.60	4.91	-1.11	2.19
SR	2.71	2.68	1.97	2.90	2.84	2.66	2.90	2.82	-0.24	0.48
Panel B: 1/ 8 Tick Size									Period (1993 - 1996)	
Return	3.74	3.73	3.15	3.47	3.47	3.39	3.83	3.88	-0.01	0.11
<i>t</i> -stat	5.73	6.00	10.73	6.43	6.35	8.13	6.77	6.72	-0.37	2.62
SR	5.63	5.60	8.49	6.81	5.73	6.90	6.63	6.19	-0.17	1.19
Panel C: 1/16 Tick Size									Period (1997 - 2000)	
Return	5.61	5.57	0.33	5.72	5.61	5.11	5.79	5.83	-0.14	0.12
<i>t</i> -stat	25.49	26.00	1.35	16.05	18.06	18.92	17.92	20.74	-1.78	1.62
SR	12.35	12.20	1.01	11.30	12.73	10.35	11.73	11.28	-0.89	0.71
Panel D: Early Post-Decimalization									Period (2001 - 2004)	
Return	1.61	1.51	0.31	1.74	1.61	1.37	1.85	1.77	-0.04	-0.01
<i>t</i> -stat	5.12	5.74	3.32	5.37	6.31	4.06	4.46	4.60	-0.63	-0.12
SR	4.55	4.62	1.56	4.29	4.84	4.42	4.67	4.64	-0.35	-0.07
Panel E: Late Post-Decimalization									Period (2005 - 2016)	
Return	0.18	0.17	0.29	0.33	0.29	0.16	0.30	0.27	0.01	0.07
<i>t</i> -stat	4.48	3.89	7.75	6.02	4.84	3.27	6.22	4.96	0.32	1.37
SR	1.36	1.41	2.10	2.06	2.03	0.83	1.80	1.73	0.09	0.46

Reported are the annualized excess returns, *t*-statistics, and Sharpe ratios for the SPY market-timing strategy. Models include the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). The independent variables are lagged returns of S&P 500 constituents. Results are reported for the full, 1/16 tick size, 1/8 tick size, early post-decimalization, and late post-decimalization samples.

Table 2.3: Excess Returns and Sharpe ratios with transaction costs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	Intraday SPY	Hold SPY
Panel A: Overall									Period (1993 - 2016)	
Return	0.03	0.01	0.05	0.06	0.02	-0.07	0.06	0.06	-0.03	0.07
<i>t</i> -stat	1.71	0.77	3.29	2.70	1.10	-3.87	3.38	4.27	-1.11	2.19
SR	0.42	0.20	0.68	0.77	0.24	-0.66	0.67	0.98	-0.24	0.48
% Trades	11.63	12.73	11.45	12.39	14.86	19.26	11.21	11.00		
Panel B: 1/ 8 Tick Size									Period (1993 - 1996)	
Return	0.04	0.05	0.04	0.01	0.07	0.03	0.05	0.04	-0.01	0.11
<i>t</i> -stat	4.79	4.86	2.68	1.22	2.82	2.77	4.98	4.28	-0.37	2.62
SR	2.50	2.60	1.48	0.80	2.27	1.20	2.70	2.34	-0.17	1.19
% Trades	0.47	0.60	1.53	0.27	1.25	0.75	0.52	0.48		
Panel C: 1/16 Tick Size									Period (1997 - 2000)	
Return	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	-0.14	0.12
<i>t</i> -stat	1.83	1.67	1.48	1.88	0.46	1.38	2.15	2.42	-1.78	1.62
SR	1.01	0.95	0.75	0.97	0.21	1.14	1.04	0.98	-0.89	0.71
% Trades	0.11	0.12	0.07	0.13	0.20	0.25	0.08	0.10		
Panel D: Early Post-Decimalization									Period (2001 - 2004)	
Return	0.14	0.13	0.00	0.11	0.06	-0.06	0.17	0.17	-0.04	-0.01
<i>t</i> -stat	3.17	3.05	0.05	2.09	0.85	-1.29	3.17	3.70	-0.63	-0.12
SR	2.00	1.97	0.02	1.20	0.49	-0.61	2.20	2.19	-0.35	-0.07
% Trades	10.57	11.31	7.23	12.76	15.01	17.54	10.05	10.05		
Panel E: Late Post-Decimalization									Period (2005 - 2016)	
Return	-0.01	-0.04	0.09	0.08	0.00	-0.14	0.04	0.05	0.01	0.07
<i>t</i> -stat	-0.58	-1.41	3.07	1.95	-0.18	-4.44	1.41	2.32	0.32	1.37
SR	-0.18	-0.49	0.88	0.81	-0.05	-0.98	0.35	0.68	0.09	0.46
% Trades	19.45	21.35	19.88	20.29	24.12	32.18	18.78	18.37		

Reported are the annualized excess returns, *t*-statistics, and Sharpe ratios for the SPY market-timing strategy with transaction costs. Models include the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). The independent variables are lagged returns of S&P 500 constituents. Results are reported for the full, 1/16 tick size, 1/8 tick size, early post-decimalization, and late post-decimalization samples.

Table 2.4: Market predictability (percentage R^2_{OOS}) and profitability (Sharpe ratio) by training duration post-decimalization

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Post-Decimalization Out-of-Sample R^2								
Training months	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED
Expanding	0.30	0.22	0.04	0.26	-0.12	-0.19	0.35	0.35
58 month	0.33	0.30	-0.06	0.25	-0.13	-0.15	0.38	0.39
34 month	0.23	0.26	0.09	0.23	-0.29	-0.20	0.36	0.38
22 month	0.29	0.27	0.07	0.22	-0.42	-0.30	0.33	0.34
16 month	0.19	0.18	0.08	-0.01	-0.39	-0.36	0.28	0.30
10 month	0.07	0.05	0.07	-0.06	-0.36	-0.29	0.22	0.24
7 month	0.05	0.05	0.00	-0.03	-0.36	-0.42	0.20	0.21
4 month	0.05	0.05	-0.08	-0.22	-0.63	-0.65	0.12	0.14
1 month	-0.09	-0.09	-0.33	-0.68	-1.35	-0.40	-0.08	-0.04
Panel B: Post-Decimalization Sharpe Ratio after Transaction Costs								
	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED
Expanding	0.34	0.06	0.72	0.90	0.12	-0.90	0.69	1.04
58 month	0.26	0.21	-0.16	0.85	0.23	-0.73	0.80	0.81
34 month	-0.07	-0.21	0.19	0.69	-0.30	-0.72	0.45	0.56
22 month	-0.13	-0.26	0.43	0.19	-0.30	-1.27	0.50	0.37
16 month	-0.10	-0.28	0.41	0.06	-0.24	-1.04	0.41	0.57
10 month	-0.11	-0.14	0.27	-0.43	-0.64	-0.54	0.50	0.55
7 month	-0.04	-0.06	0.28	-0.03	-0.22	-0.63	0.59	0.61
4 month	-0.41	-0.40	0.02	-0.30	-0.66	-0.89	0.34	0.15
1 month	-0.16	-0.15	-0.40	-1.50	-2.13	-1.29	-0.29	0.09

Reported are the out-of-sample predictive R^2 percentages and annualized Sharpe ratios for the SPY market-timing strategy with transaction costs. Models include the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). Compared are R^2 values across 1-, 4-, 7-, 10-, 16-, 22-, 34-, and 58-month training windows. The independent variables are lagged returns of S&P 500 constituents. Reported are results for the *post-decimalization* subsamples.

Table 2.5: Market predictability (percentage R^2_{OOS}) and profitability (Sharpe ratio) by time post-decimalization

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Post-Decimalization Out-of-Sample R^2									
Time	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	
9:35 - 10:00	-0.06	-0.21	0.21	-0.20	-1.26	-0.72	0.12	0.19	
10:00 - 10:30	0.34	0.26	0.05	0.09	-0.27	-0.17	0.33	0.36	
10:30 - 11:00	0.38	0.40	0.10	0.55	0.29	-0.02	0.56	0.49	
11:00 - 11:30	0.41	0.32	0.03	0.35	0.20	-0.07	0.49	0.47	
11:30 - 12:00	0.32	0.24	-0.02	0.29	-0.22	-0.11	0.38	0.38	
12:00 - 12:30	0.47	0.41	-0.01	0.48	0.32	-0.21	0.54	0.51	
12:30 - 13:00	0.39	0.26	0.24	0.48	-0.20	-0.46	0.40	0.45	
13:00 - 13:30	0.31	0.33	0.10	0.43	-0.05	-0.28	0.38	0.31	
13:30 - 14:00	0.37	0.26	0.01	0.34	-0.04	-0.07	0.38	0.39	
14:00 - 14:30	0.99	0.96	0.19	0.62	0.65	0.62	0.90	0.84	
14:30 - 15:00	0.42	0.35	0.17	0.47	0.01	0.12	0.44	0.47	
15:00 - 15:30	-0.16	-0.27	-0.26	-0.08	-0.32	-0.77	-0.13	-0.09	
15:30 - 15:55	0.10	-0.01	-0.26	0.13	0.11	-0.11	0.17	0.14	
Panel B: Post-Decimalization Sharpe Ratio after Transaction Costs									
Time	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	Intraday SPY
9:35 - 10:00	-0.20	-0.35	0.23	0.26	-0.64	-0.63	0.55	0.21	-0.25
10:00 - 10:30	0.57	0.47	0.60	0.08	0.19	-0.19	0.50	0.69	-0.02
10:30 - 11:00	0.18	0.23	0.35	0.07	0.17	0.03	0.62	0.31	-0.21
11:00 - 11:30	-0.11	-0.30	-0.08	0.00	0.18	-0.17	0.08	-0.08	-0.29
11:30 - 12:00	0.13	0.15	0.21	0.52	0.08	-0.23	0.40	0.32	0.12
12:00 - 12:30	0.00	-0.09	0.44	0.38	0.20	-0.50	0.28	0.24	-0.04
12:30 - 13:00	0.22	0.32	0.40	0.39	-0.52	-0.89	0.18	0.62	0.47
13:00 - 13:30	0.13	0.11	0.55	0.74	0.23	-0.29	0.18	0.57	0.16
13:30 - 14:00	0.17	0.00	0.09	0.39	-0.26	-0.59	0.34	0.34	-0.31
14:00 - 14:30	0.01	0.19	0.51	0.44	0.64	-0.17	0.27	0.34	-0.23
14:30 - 15:00	0.31	0.17	0.49	0.52	0.24	-0.30	0.35	0.59	0.43
15:00 - 15:30	-0.13	-0.30	-0.30	-0.34	-0.51	-0.99	-0.59	-0.39	0.27
15:30 - 15:55	-0.11	-0.28	-0.44	0.43	0.40	0.27	-0.05	-0.11	-0.09

Reported are the out-of-sample predictive R^2 percentages and annualized Sharpe ratios for the SPY market-timing strategy with transaction costs. Models include the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). Compared are results across 30-minute windows throughout the trading day. The independent variables are lagged returns of S&P 500 constituents. Reported are results for the *post-decimalization* subsamples.

Table 2.6: Market predictability (percentage R^2_{OOS}) and profitability (Sharpe ratio) by volatility in the post-decimalization period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Post-Decimalization Out-of-Sample R^2									
Volatility	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	
Low	0.09	-0.02	-0.03	0.02	-0.41	-0.76	0.12	0.17	
Mid	0.19	0.11	-0.09	0.11	-0.11	-0.37	0.24	0.25	
High	0.34	0.26	0.07	0.31	-0.09	-0.11	0.40	0.39	
Panel B: Post-Decimalization Sharpe Ratio after Transaction Costs									
Volatility	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	Intraday SPY
Low	0.47	0.09	0.20	0.23	-0.81	-1.79	-0.03	0.44	2.88
Mid	0.40	0.39	0.43	0.28	0.26	-1.08	0.81	0.89	-0.20
High	0.14	-0.15	0.73	1.04	0.23	-0.33	0.60	0.87	-0.93

Reported are the out-of-sample predictive R^2 percentages and annualized Sharpe ratios for the SPY market-timing strategy with transaction costs. Models include the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). Compared are results across three groups sorted on realized volatility. The independent variables are lagged returns of S&P 500 constituents. Reported are results for the *post-decimalization* subsamples.

Table 2.7: Market predictability (percentage R^2_{OOS}) and profitability (Sharpe ratio) by illiquidity in the post-decimalization period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Post-Decimalization Out-of-Sample R^2									
Illiquidity	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	
Low	0.00	-0.08	0.06	-0.06	-0.41	-0.46	0.04	0.04	
Mid	0.01	-0.04	0.02	0.02	-0.43	-0.34	0.05	0.07	
High	0.76	0.66	0.04	0.69	0.37	0.11	0.85	0.83	
Panel B: Post-Decimalization Sharpe Ratio after Transaction Costs									
Illiquidity	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	Intraday SPY
Low	-0.25	-0.43	0.42	0.11	-0.34	-1.08	0.24	0.26	0.33
Mid	-0.02	-0.21	0.93	0.83	0.11	-0.47	0.23	0.59	-0.22
High	1.03	0.85	0.18	1.03	0.47	-0.56	1.50	1.22	-0.07

Reported are the out-of-sample predictive R^2 percentages and annualized Sharpe ratios for the market-timing strategy with transaction costs. Models include the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). Compared are results across three groups sorted on Amihud [12] illiquidity. The independent variables are lagged returns of S&P 500 constituents. Reported are results for the *post-decimalization* subsamples.

Table 2.8: Market predictability (percentage R_{OOS}^2) and profitability (Sharpe ratio) by crisis period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Post-Decimalization Out-of-Sample R^2										
	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED		
2005-2007	0.07	-0.05	0.00	0.02	-0.14	-0.39	0.11	0.10		
Subprime	0.04	-0.02	0.08	0.02	-0.37	-0.48	0.06	0.07		
2010-2011	-0.02	-0.05	0.03	-0.03	-0.67	-0.39	-0.03	0.02		
EU crisis	0.08	-0.01	0.01	0.03	-0.30	-0.70	0.06	0.10		
2014-2016	-0.07	-0.17	-0.04	-0.14	-0.36	-0.41	-0.03	0.00		
Panel B: Post-Decimalization Sharpe Ratio before Transaction Costs										
	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	Intraday SPY	Hold SPY
2005-2007	1.44	1.13	1.66	1.86	1.93	2.26	2.32	1.85	-0.93	0.56
Subprime	1.06	0.96	3.11	3.78	2.67	0.94	2.22	1.91	0.07	-0.40
2010-2011	2.18	2.07	2.22	1.87	2.32	1.14	2.42	3.22	0.20	0.54
EU crisis	3.07	2.88	2.82	3.08	2.30	0.84	3.52	3.08	1.34	2.29
2014-2016	0.79	0.94	1.70	1.68	1.88	0.51	0.91	1.35	0.61	0.68
Panel C: Post-Decimalization Sharpe Ratio after Transaction Costs										
	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED	Intraday SPY	Hold SPY
2005-2007	0.21	-0.04	0.22	-0.17	-0.48	-1.11	-0.36	-0.35	-0.93	0.56
Subprime	-1.09	-1.55	1.38	2.00	-0.12	-0.52	0.73	0.87	0.07	-0.40
2010-2011	0.77	0.30	1.97	1.07	0.38	-1.35	0.69	1.82	0.20	0.54
EU crisis	0.63	-0.22	0.68	0.14	-1.05	-2.68	-0.09	0.57	1.34	2.29
2014-2016	-0.58	-0.38	0.64	0.12	0.17	-2.18	-0.06	0.40	0.61	0.68

Reported are the out-of-sample predictive R^2 percentages and annualized Sharpe ratios for the SPY market-timing strategy without and with transaction costs. Models include the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). The independent variables are lagged returns of S&P 500 constituents. Reported are results for the late post-decimalization subsample which includes the subprime crisis, during 2008-2009, and the European debt crisis, during 2012-2013.

Table 2.9: Market predictability (percentage R^2_{OOS}) and profitability (Sharpe ratio) for autoregressive models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Out-of-Sample R^2									
	Autoregressive Models						Baseline Models		
	AR(1)	AR(p)	OLS	LAS	EN	PCR	LAS	EN	PCR
1993-2016	0.55	0.53	-0.49	0.58	0.60	-0.07	2.00	1.95	0.31
1993-1996	0.22	0.37	-4.51	0.12	0.40	-0.04	9.27	9.32	6.52
1997-2000	1.87	1.95	1.07	1.78	1.79	-0.01	3.72	3.69	-0.76
2001-2004	0.17	0.08	-0.81	0.30	0.27	-0.13	0.91	0.81	0.03
2005-2016	-0.10	-0.20	-0.67	-0.02	-0.03	-0.08	0.02	-0.04	0.04
Panel B: Sharpe Ratio after Transaction Costs									
	Autoregressive Models						Baseline Models		
	AR(1)	AR(p)	OLS	LAS	EN	PCR	LAS	EN	PCR
1993-2016	0.00	-0.63	-1.25	-0.27	-0.38	-0.03	0.42	0.20	0.68
1993-1996	-0.49	-0.71	-0.56	-0.68	-0.68	0.14	2.50	2.60	1.48
1997-2000	0.72	0.73	0.41	0.49	0.65	0.33	1.01	0.95	0.75
2001-2004	0.10	-0.39	-2.08	0.63	0.27	0.38	2.00	1.97	0.02
2005-2016	-0.04	-0.91	-1.70	-0.50	-0.62	-0.10	-0.18	-0.49	0.88

Reported are the out-of-sample predictive R^2 percentages and annualized Sharpe ratios for the SPY market-timing strategy with transaction costs. Models include the AR(1), AR(p), lasso (LAS), elastic net (EN), and principal component regression (PCR). The independent variables are 500 lagged returns of S&P 500. Results are reported for the full, 1/16 tick size, 1/8 tick size, early post-decimalization, and late post-decimalization samples.

Table 2.10: Market predictability (percentage R^2_{OOS}) and profitability (Sharpe ratio) with additional characteristics in the post-decimalization period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Post-Decimalization Out-of-Sample R^2								
	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED
Base	0.05	0.05	-0.08	-0.22	-0.63	-0.65	0.12	0.14
BETA	-0.98	-1.41	-2.44	-1.92	-1.54	-4.10	-0.39	-0.03
ILLIQ	0.09	-0.44	-8.73	-2.70	-7.74	-3.22	-0.62	0.05
KURT	-0.64	-2.27	-1.1E+18	-2.12	-1.62	-4.25E+23	-1.18E+22	-0.01
MAX	0.03	-0.50	-0.10	-0.56	-1.14	-1.27	-0.02	0.09
MIN	-0.37	-0.36	-0.10	-5.53	-0.68	-1.38	-0.25	0.01
MOM	0.07	-1.24	-0.07	-0.42	-2.59	-1.05	-0.08	0.08
SKEW	0.06	-1.13	-4.8E+12	-0.72	-1.59	-1.8E+15	-5.5E+13	-0.01
VOL	-0.02	-0.35	-0.16	-1.69	-2.26	-1.09	-0.07	0.07
VOLUME	-1.95	-2.24	-0.06	-3.05	-7.49	-1.35	-0.47	-0.24
SPREAD	-1.51	-1.20	-3.67	-0.12	-0.42	-1.40	-0.08	0.13
Panel B: Post-Decimalization Sharpe Ratio after Transaction Costs								
	LAS	EN	PCR	RF	GBRT	ANN	MEAN	MED
Base	-0.41	-0.40	0.02	-0.30	-0.66	-0.89	0.34	0.15
BETA	-0.27	-0.30	0.14	-0.80	-2.18	-1.38	-0.17	-0.08
ILLIQ	-0.29	-0.36	-0.48	-1.12	-2.77	-1.29	-0.95	-0.18
KURT	-0.47	-0.46	0.22	-0.90	-1.74	-1.19	-0.07	0.11
MAX	-0.21	-0.31	-0.27	-0.71	-1.76	-1.00	-0.53	0.02
MIN	-0.18	-0.18	-0.06	-0.47	-1.28	-0.99	-0.20	0.06
MOM	-0.29	-0.32	0.12	-0.80	-1.72	-0.92	-0.18	0.07
SKEW	-0.26	-0.37	0.27	-0.62	-1.58	-1.00	-0.10	0.32
VOL	-0.53	-0.61	-0.21	-0.94	-2.33	-1.14	-0.17	-0.12
VOLUME	-0.24	-0.27	0.28	-1.02	-2.75	-1.11	-0.68	0.08
SPREAD	-0.21	-0.25	-0.03	0.16	-0.56	-1.00	0.17	0.44

Reported are the out-of-sample predictive R^2 percentages and annualized Sharpe ratios for the SPY using the lasso (LAS), elastic net (EN), principal component regression (PCR), random forest (RF), gradient-boosted regression trees (GBRT), neural network (ANN), mean ensemble (MEAN), and median ensemble (MED). In addition to lagged returns (LAG), the independent variables include market beta (BETA), illiquidity (ILLIQ), kurtosis (KURT), maximum (MAX), minimum (MIN), momentum (MOM), skewness (SKEW), volatility (VOL), or volume (VOLUME) of S&P 500 constituents calculated over the previous day as well as the percent bid-ask spread (SPREAD). See appendix B.2 for details on how these characteristics are calculated.

Chapter 3

Asian Δ -hedging and S&P 500 ensemble predictive trading profitability

Abstract

A profitability comparison between Asian option Δ -hedging, using the results of chapter 1, and machine learning (mean) ensemble predictive trading the S&P 500, using the results of chapter 2, is presented. Interpreting the upper- and lower-bounds on arithmetic Asian option prices, given in section 1.2.4, as hypothetically respectively representing ask and bid values, it is found that hedging profitability depends largely on the ability of the option holder to secure prices closer to the bid value, and that, settling midway between the bid and ask, significant profits are consistently accumulated during the years 2004-2016. This is managed with infrequent rebalancing of a self-financing portfolio, according to a backward-difference estimate of the Δ -hedging parameter. In contrast, ensemble predictive trading the S&P 500 yields comparatively very small returns, despite trading much more frequently. Given the robustness of the latter profitability to transaction costs, demonstrated in chapter 2, the same robustness follows as a corollary of the superior hedging returns achieved.

3.1 Introduction

As one means of ‘bridging the gap’ between the lengthier chapters 1 and 2, a profitability comparison is presented between Asian option Δ -hedging and machine learning mean ensemble predictive trading the S&P 500, using the results of both chapters. The hedging strategy achieves far superior returns (on average) with infrequent periods (days) hedging a single arithmetic Asian option via a portfolio self-financed by short-selling the underlying (S&P 500), and rebalancing every five minutes according to a backward-difference estimate of the Δ -hedging parameter. Ensemble predictive trading the S&P 500 much more frequently (every five minutes of each day) and in greater quantity (one S&P 500 ‘share’), further enforcing the superiority of the arithmetic Asian option Δ -hedging strategy.

This brief chapter proceeds as follows: The arithmetic Asian option Δ -hedging strategy, and the methodology for its computation, are presented in section 3.2, using results from chapter

1. Results for the final (and otherwise arbitrarily chosen) trading-day of available S&P 500 returns data, December 30, 2016, are presented in section 3.3, followed by results over the period including calendar years 2004-2016. Conclusions follow in section 3.4.

3.2 Methods

Suppose a trader decides at the beginning of each day, say, $d \in \{1, 2, \dots, D\}$, whether to hedge a single fixed-strike arithmetic Asian call option on an S&P 500 underlying, self-financing a hedging portfolio by short-selling enough underlying units to cover the option purchase. (The option is taken to have maturity at the end of the day, in part to avoid complications arising from discontinuities in the underlying price between days.) For the remainder of the day, the trader is committed to rebalancing the portfolio every five minutes to contain the same single option, and Δ units of the underlying, depositing or borrowing/withdrawing from a risk-free bond account as required. At the beginning of day d , the trader observes the underlying price, $S_{d,0}$, and considers whether to hedge an Asian option with price, $V_{d,0}$, taken according to whether $V_{d,0} \leq \bar{V}$. I.e., the trader is unwilling to short-sell more than \bar{V} worth of underlying to self-finance the portfolio, choosing instead to do nothing for the remainder of the day, and preserve any existing profits or losses. In case the trader hedges on day d , the portfolio is balanced at each five-minute period, say, $m \in \{1, 2, \dots, M\}$, such that the following number of underlying units are held:

$$\Delta \equiv V_S \approx \frac{V_{d,m} - V_{d,m-1}}{S_{d,m} - S_{d,m-1}} \equiv \Delta_{d,m} \quad (3.1)$$

The trader observes the underlying price history, including $\{S_{d,m}, S_{d,m-1}\}$, and option prices, $\{V_{d,m}, V_{d,m-1}\}$, computed (by an exogenous ‘nature’ or ‘oracle’) using formula (1.43) as in section 1.2.4. This again yields both upper- and lower-bound option prices, which are interpreted respectively as ask and bid values. Both prices yield their own Δ -estimates via the backward-difference formula (3.1). This method for estimating the option Δ -hedging parameter is preferred to a similar procedure as in section 1.2.4, applied to the corresponding parameter of corollary 1.2.7, for two reasons: First, this procedure proved *very* numerically sensitive, to the extent that initial λ_0 -value guesses were most often accepted outright, indicating a (nearly) flat price V along underlying S curves. Second, and not unrelated, the hedging parameters of corollary 1.2.7 depend explicitly on the terminal underlying price, S_T , and so must be evaluated numerically via MC simulations, which drastically increase running time and add uncertainty to such hedging parameter estimates. [Nor is it attractive to attempt analytical derivation of a corresponding Δ -hedging parameter as in theorem 1.2.8: Though this certainly appears feasible, it would likely require much more effort and prove much more error-prone than the derivation of formula (1.43), which was already quite arduous using even Maple to simplify where possible.]

Each day, d , the risk-free interest rate, r_d , and volatility of the underlying S&P 500, σ_d , are moment-matched (for a lognormal distribution) to the index data that day. These data are described in appendix B, but only the five-minutely sampled index/‘market’ prices including years 2004-2016 are used; i.e., the individual S&P 500 constituents are not explicitly considered, though they may implicitly account heavily for the corresponding mean predictive trading

strategy taken from chapter 2 and presently compared with a simpler, (advantageously) far less liquid, and much more profitable (on average) arithmetic Asian option Δ -hedging strategy. The superiority of the latter is further enforced by the fact that it doesn't require recourse to the individual constituents of the S&P 500, nor does it require three months of such historical data to render trading decisions; only the present and immediately past option and underlying prices are needed to compute the Δ -hedging parameter (3.1). [The three-month predictions were used, as they give the mean ensemble predictive trading strategy the least advantage over arithmetic Asian option Δ -hedging, compared to the longer prediction windows considered (with improved results) in chapter 2.] Of course, each option is computed using the corresponding underlying price, as well as the *daily* index price history to compute (via the trapezoidal rule) the arithmetic average, $A_{d,m}$. But this is done exogenous to the trader, by an 'oracle' able to divine the daily risk-free rate, r_d , and underlying volatility, σ_d .

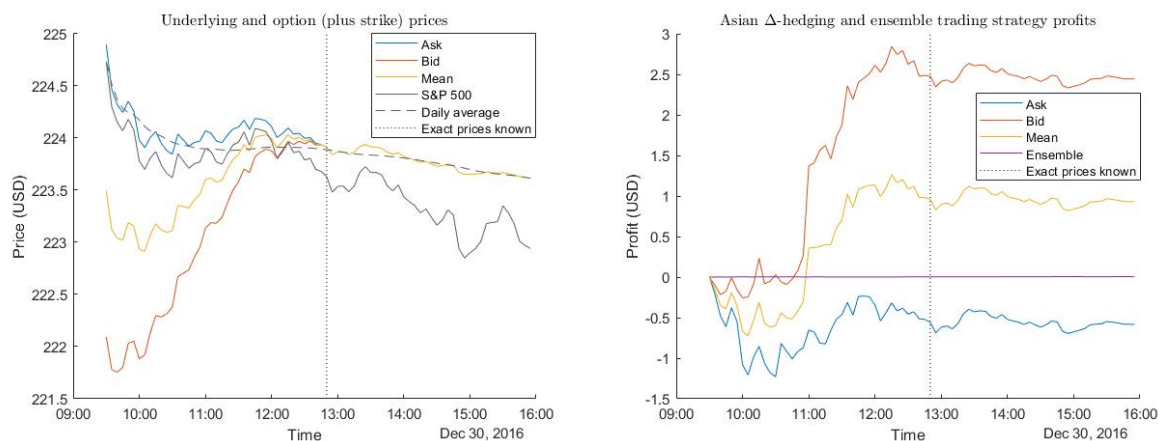
3.3 Results

Shown in figures 3.1-3.3 are a number of results for the otherwise-arbitrarily-chosen, last trading day of available five-minutely S&P 500 price data, December 30, 2016. On that day, and all days in the years including 2004-2016 studied subsequently, the strike price is set to half of the daily starting price, i.e., $K_d = S_{d,0}/2$ on day d . This is done in order to avoid similar numerical difficulties which arise for such strike values when attempting to estimate the arithmetic Asian Δ -hedging parameter as done for the price itself in section 1.2.4, described following equation (3.1). Namely, initial λ_0 guesses tend to be accepted immediately, resulting in price discontinuities. This is due to the extreme sensitivity of prices V to λ demonstrated in figures 1.5-1.7, and both numerical challenges may be related to the small (annualized) scales, $\sigma_d^2(T-t)$, involved in options with daily maturities, and are apparently exacerbated for larger strike values which remove call options further from the money, where the estimation quality demonstrated in figures 1.34 deteriorates and the option price flattens to zero. (The desirable behavior is demonstrated in figure 1.2.3 with $\lambda_{0,\text{ask}} = 10 = 10\lambda_{0,\text{bid}}$; the difficulty occurs when such curves 'bend and flatten' past some time, no longer updating beyond the initial guesses, taken as the preceding values to somewhat mitigate the challenge.) Naturally, the tendency for this problem to arise in estimating V increases as the time to maturity, $T-t$, falls, and lower strike values ensure eventual exercise earlier in the day, allowing the exact option price (1.45) to be used, which also immediately yields exact hedging parameters, in particular,

$$\Delta \equiv V_S = \frac{1 - e^{-r(T-t)}}{rT} \quad (3.2)$$

Figure 3.1a presents the S&P 500 price and its daily average on December 30, 2016, along with bid, ask, and their mean option (plus strike) prices; the time when eventual exercise is guaranteed is indicated vertically. The option prices converge to the exact value, known past this time, in a manner similar to figure 1.34b: The fact that the upper bound/ask price tracks more closely the underlying, than the bid and even their mean, accounts for the profits achieved, rather than losses, from Δ -hedging portfolios including a single arithmetic Asian call option purchased at the lesser prices, displayed in figure 3.1b. These are clearly realized during the period when the bid (and mean) price is increasing most rapidly. Supposing the trader is able

to settle at the beginning of the day to purchase one option at the mean bid-ask price, one key takeaway is that the resulting Δ -hedging strategy dramatically dominates that applying the mean ensemble predictions of chapter 2: The latter, though profitable on the day considered, are three orders of magnitude lower, despite trading on the same order of units of underlying [one every five minutes, as opposed to the (differences of) the Δ -hedging ratios shown in figure 3.2b]. Interestingly, the most profitable ‘bid’ Δ -hedging strategy trades quite erratically in the underlying prior to certain exercise, again most notably (and naturally) when the bid itself is increasing most rapidly. The much greater apparent fit of the ‘ask’ Δ -hedging ratios with the values once exercise is certain, perhaps indicate that the upper (ask) option price is tighter than the lower (bid) bound, more so than the closer tracking of the underlying observed and noted in figure 3.1a. Essentially, the bidder is able to profit by capitalizing at the beginning of the day with an artificially low bid, compared to a more reasonable ask, and so once settling midway with a willing seller. Note also that the interpretation, valid for European options in the BS framework, of the Δ -hedging parameter as the probability of exercise, is clearly *not* valid for arithmetic Asian options: The true values (3.2) decay exponentially to zero as the time to maturity, $T - t$, vanishes. Figure 3.3b shows the range of figure 3.2b within the unit interval.



(a) Option prices converge to S&P 500 average (b) Profitability comparison with ensemble trading

Figure 3.1: Profitability comparison for December 30, 2016

To further compare the arithmetic Asian Δ -hedging and ensemble trading strategies, consider the ‘charm’ of Haug [158] in the case that the exact option price (1.45) is used:

$$\Delta_t \equiv V_{S_t} = -\frac{e^{-r(T-t)}}{T} < 0 \quad (3.3)$$

This may otherwise be approximated by a backward finite difference, as for formula (3.1):

$$\Delta_t \approx \frac{\Delta_{d,m} - \Delta_{d,m-1}}{t_{d,m} - t_{d,m-1}} \quad (3.4)$$

Here, of course, $t_{d,m(-1)}$ denotes the time, measured in years from an arbitrary but common starting date, at the beginning of five-minute time interval $m(-1)$ on day d . Restricting attention to the *sign* of the charm (3.4), figure 3.4a compares the arithmetic Asian bid and ask

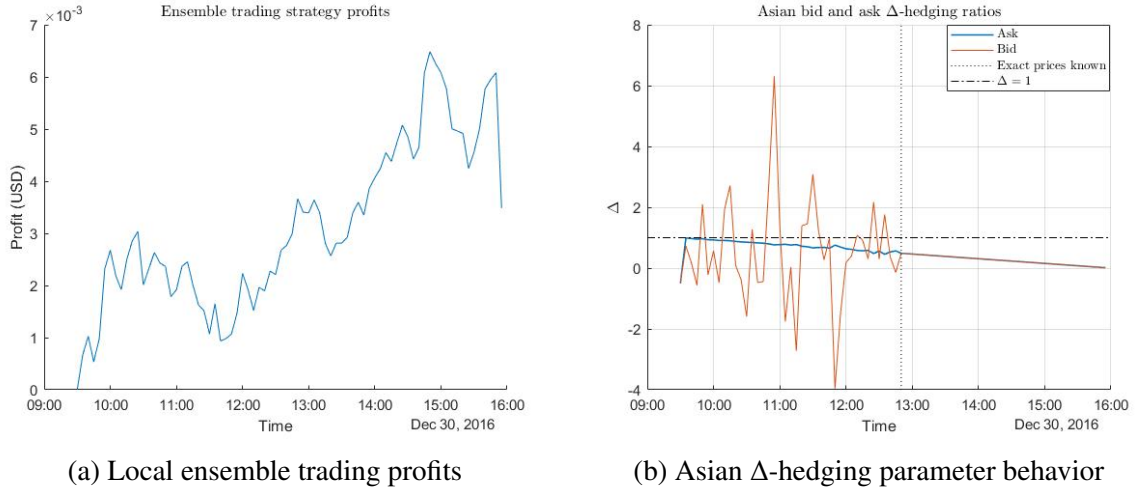


Figure 3.2: Further profitability comparison for December 30, 2016

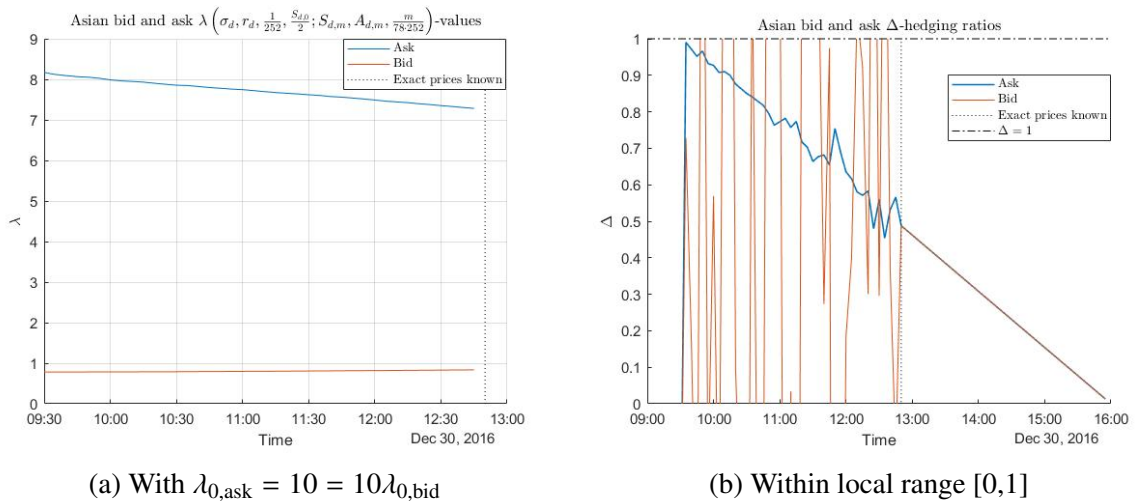


Figure 3.3: Yet further profitability comparison for December 30, 2016

Δ -hedging with the ensemble trading strategy on December 30, 2016, in terms only of whether the underlying is purchased or sold during each five-minute interval following the initial self-financing by short-selling the underlying and prior to the certainty of eventual exercise, when the exact formula (3.3) is applied and all three strategies sell off underlying for the remainder of the day. Indeed, the mean ensemble trading strategy sells one unit of underlying every five minutes throughout the day, resulting in the modest profits depicted in figure 3.2a. In the period shown in figures 3.4, the arithmetic Asian bid (ask) Δ -hedging strategy sells underlying 48.72% (66.67%) of the time, versus 100% for the mean ensemble trading strategy, and the bid and ask strategies simultaneously buy and sell underlying only 28.21% of the time. But as discussed and displayed in figures 3.2b and 3.4b, it is clear that the arithmetic Asian bid (ask) Δ -hedging strategy sells underlying often in multiple (small fractional) units, as opposed to the uniform unit sold every five minutes by the mean ensemble strategy, resulting in the comparatively (moderately) large profits (losses) given in figure 3.1b.

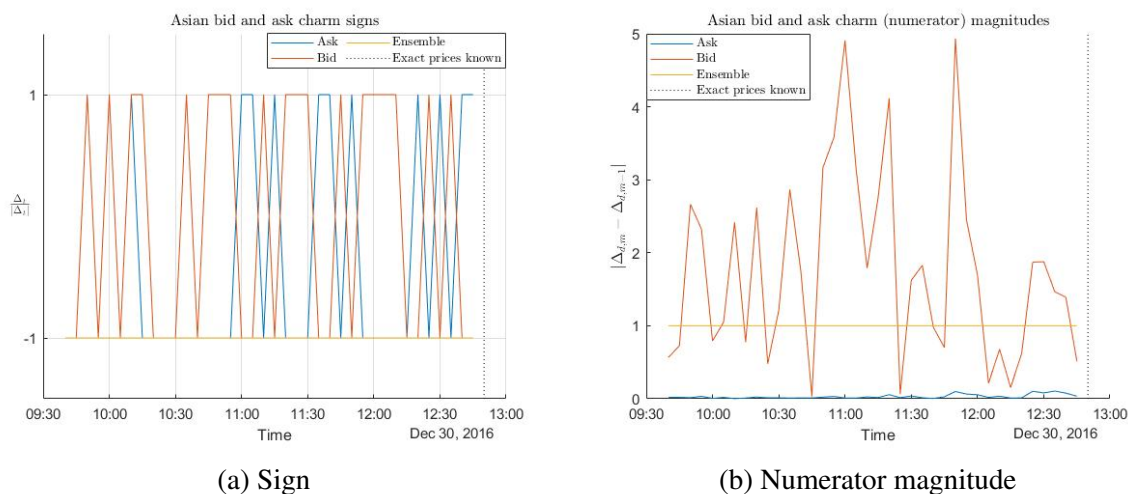


Figure 3.4: Charm prior to certain exercise on December 30, 2016

Considering the inclusive period 2004-2016 (for convenience as the data is differently formatted during the period 1993-2003), figure 3.5a demonstrates sustained profits (losses) of the bid and mean (ask) arithmetic Asian Δ -hedging strategy, as in figure 3.1b for December 30, 2016, *assuming the trader is unwilling to balance the Δ -hedging portfolio at a cost of more than \$200 each day*. That is, at the beginning of each day, the trader computes the cost of balancing the Δ -hedging portfolio, and only considers proceeding so long as it does not exceed \$200. This is one way of gauging the risk preferences of the trader. To the nearest (ten thousand) dollar(s), the trader never (always) balances during the period 2004-2016 when observing a maximum allowable balancing cost up to (exceeding) \$36 (\$190000). Shown in figure 3.6a are ask (bid) profit paths in blue (red), from light to dark shades determined by a log-linear space between these two prices: Perhaps most apparent is that (intuitively) many of the profit steps occur at common times, and there may be greater path volatility in earlier years. To better illustrate the effect of maximum allowable balancing cost, cumulative profits for the entire period 2004-2016 are shown in figures 3.6c-3.6e, respectively bounding above the plotted allowable balancing costs at \$190000, \$3000, and \$70: These bounds are chosen to illustrate both the behavior over the entire log-linear range of (1000) allowable balancing costs, as well as earlier periods of high volatility in greater detail. A number of conclusions follow: The ask (bid) profit paths are mostly negative (positive), as expected, such that the mean profit is positive between allowable borrowing costs of roughly \$46.5 and \$650, peaking over a ‘sweet spot’ between \$52 and \$64.5, in which there is comparatively little volatility in both bid and ask, and especially the mean profits. Thus, a trader settling each period midway between the bid and ask value, performs well when moderately risk-averse, spending only up to about \$650 on daily balancing, and particularly so when more acutely risk-averse, restricting daily balancing to be under roughly \$65, but not so risk-averse as to avoid balancing altogether (i.e., below \$37) or achieve negative profits (below \$46.5). Traders able to settle at the bid price, however, are best served to be risk-loving, not restricting daily balancing costs to be any value below roughly \$1100; traders settling at the ask price are similarly most poorly served.

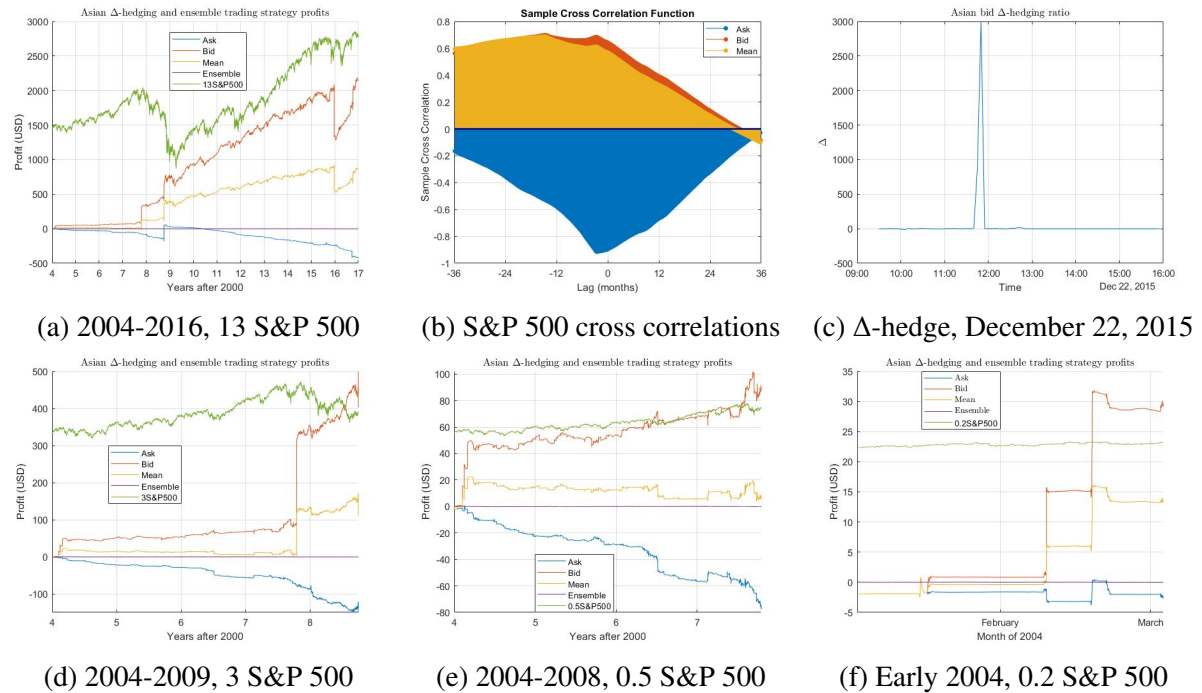


Figure 3.5: Profitability comparison with ensemble trading and multiples of the S&P 500

The remaining five figures 3.5 further examine aspects of the primary one, figure 3.5a: Noting that (13 times) the S&P 500 apparently tracks quite well the bid curve in the latter figure, versions on the reduced time scales from January 2004 to the jumps roughly occurring near the ends of 2009, 2008, and February 2004, are respectively shown in figures 3.5d-3.5f. The visual similarity is retained in each figure, and validated in figure 3.5b, which indicates the sample cross-correlation between (multiples of) the S&P 500 and the ask, bid, and mean arithmetic Asian Δ -hedging strategies, over the entire period 2004-2016, for lags between plus and minus three years: All strategies exhibit statistically significant sample cross correlations with the S&P 500 at all such lags; the bid and mean strategies' cross correlation functions are very similar, as the profitability of the latter clearly tends near half that of the former, with the ask profitability being nearly an order of magnitude less (in absolute value) than that of the bid. The bid and mean cross correlations also persist at higher values than the ask for large negative lags, but peak at smaller values (near 0.7), whereas the bid is minimized near -0.9! Finally, figure 3.5c provides insight into the loss, exceeding 600 USD, of the bid (and about half that on the mean) strategy on December 22, 2015, compared with little variation in the S&P 500, ask profitability, and bid and ask arithmetic Asian option prices on the same day: The bid (and mean) profitability depends otherwise on only the Δ -hedging parameter value (and the bond account, which is also so directly determined). The corresponding plot in figure 3.5c clarifies the problem: The S&P 500 values varied little between 11:40, 11:45, 11:50, and 11:55 on December 22, 2015, particularly as compared to the arithmetic Asian bid prices estimated at

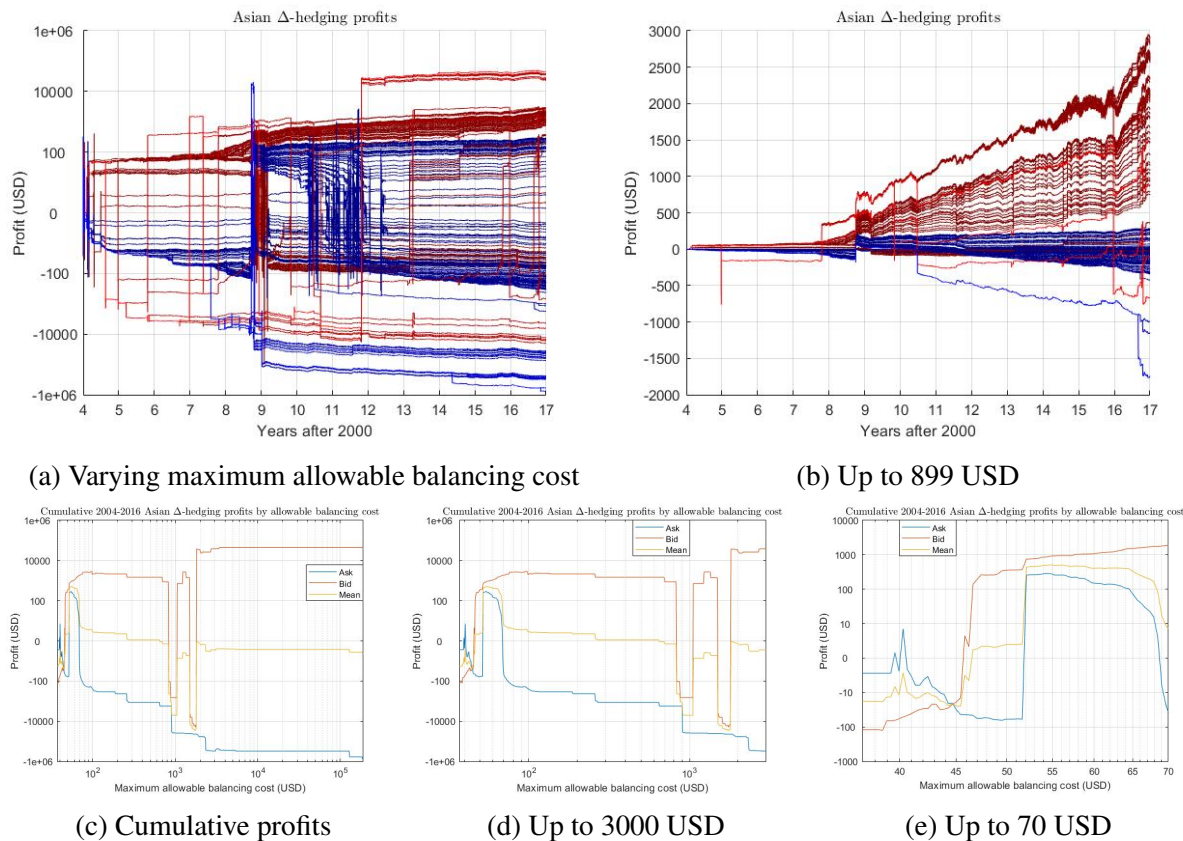


Figure 3.6: Profitability by maximum allowable balancing cost

the same times, resulting in backward-difference Δ -hedging parameter estimates (3.1) orders of magnitude larger than others issued that and most other days. I.e., the trader blindly accepts the anomalous orders to buy and sell hundreds or thousands of ‘shares’ in the S&P 500 between 11:40 and 11:55 on December 22, 2015, despite the amounts ‘typically’ issued being orders of magnitude smaller. In this case, this is done at a small fraction of a percentage loss, which for the thousands of shares traded and hundreds of dollars per share, yields the loss exceeding 600 USD (300 USD) of the bid (mean) strategy, negating about 25% of the total profitability accrued to the strategy over the preceding four years, or about one third of the strategy lifetime. As such, the foregoing analysis in part acts as a cautionary tale for both applying (in particular, first-order) finite differences to estimate hedging parameters, and ‘mindless’ automated trading. But with the other jumps in figure 3.5a apparently being due to similar phenomena, and noting that the same phenomenon yields very large gains on rare days, it is conceivable that the strategy may be less profitable long-term, if some ‘safeguard’ were introduced to ignore outlier trade orders; particularly, since the bid and mean strategies tend not to trade on many days, and as such significantly accrue greater wealth due to compounding at the risk-free interest rate on such days. But such ‘slow and steady’ strategies which filter outlying trade orders, compared to the ‘high risk, high reward’ ones studied in figures 3.5, may be appropriate in some circumstances and chosen by certain risk(-averse) preferences: They are not presently modeled, both since there are myriad ways to enact such filtering, and the strategies currently

studied are ‘pure’ Δ -hedging, in that the only decision is whether to self-finance the strategy at the beginning of each day; but this is one possible direction of further study. Additionally, plausible profitability curves for the ‘safe and steady’ strategies can readily be imagined from those of figures 3.5, simply by visually ‘removing’ the jumps and scaling toward the abscissa, to account for reduced profitability or loss due to the compounding of smaller such values.

As far as the arithmetic Asian Δ -hedging strategy is concerned, it is clear that profitability depends heavily on both the ability of the trader to secure options on a daily basis nearer to the bid than the ask price, and to be willing to accordingly balance at ‘moderately large’ costs, e.g., between roughly \$46.5 and \$650 during the entire period 2004–2016. And figure 3.5a makes clear that, when achieved, such profits dwarf those of the mean ensemble trading strategy, despite trading in comparable quantities of the underlying and much less frequently to boot. It is seen in chapter 2 that the latter profits are robust to trading costs, so that those of the arithmetic Asian Δ -hedging strategy certainly are also. To conclude the results and present an alternate perspective of those given in chapter 2, a graphical analysis is conducted of the mean ensemble trading strategy (without transaction costs), shown in figure 3.7a (and overlaid with its profitability in figure 3.9c and net periods holding the underlying in figure 3.10a) during the period 2004–16, and several realized monthly statistics of the underlying (S&P 500) price path considered as predictors in chapter 2, overlaid by the trading strategy in figures 3.7b–3.9b: Illiquidity, kurtosis, extrema, momentum, skewness, volatility, and volume. Monthly statistics are used, as that is the prediction window by which the trading strategy is determined in chapter 2, and perhaps to some extent as a result, and apparent in figure 3.7a, trades typically change direction only once every two months over the period 2004–2016; see also the corresponding histogram in figure 3.10b. Momentum refers to the sum of lagged (over the preceding month) underlying returns, as a notion of whether the underlying has recently altogether been ‘up’ or ‘down.’ Illiquidity is computed according to Amihud [12], and is intuitively seen to decrease in time in figure 3.7b. An alternate viewpoint of the mean ensemble trading strategy is explored in figures 3.10b and 3.10c: The waiting times between changes in trading direction are shown as a histogram in the former, both when buying or selling and restricted to the former, from which the latter are immediately inferred. A Poisson arrival process for the changes in trading direction has exponential waiting times, fitted to the three histograms’ data in figure 3.10c, along with nonparametric kernel estimates and similarly peaked gamma distributions. Visually, the fitted curves do not appear immediately inconsistent with one another or the corresponding histograms, possibly intuiting similarly simple dynamics for trading strategy decision changes, determined via a mean ensemble of several machine learning predictions in chapter 2. In particular, the memoryless property of the idealized exponential waiting time distribution may offer a clue as to why hyperparameter optimized recursive (in particular, long short-term memory) neural networks did not fit the high-frequency S&P 500 returns data addressed in chapter 2 better than the other models considered, and which results are as such not included therein.

Finally, to better assess the correlation between the mean ensemble trading strategy and realized monthly underlying statistics overlaid in figures 3.7b–3.9b, corresponding sample cross-correlation functions are plotted in figures 3.11. Those for maximum, minimum, volatility, volume, and especially momentum, all have typical ‘peaked bell curve’ forms, and the oth-

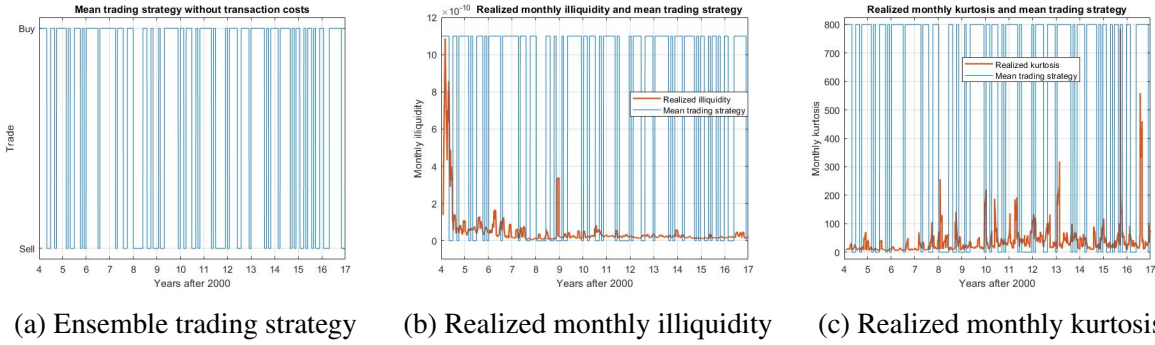


Figure 3.7: Mean strategy and realized monthly illiquidity and kurtosis

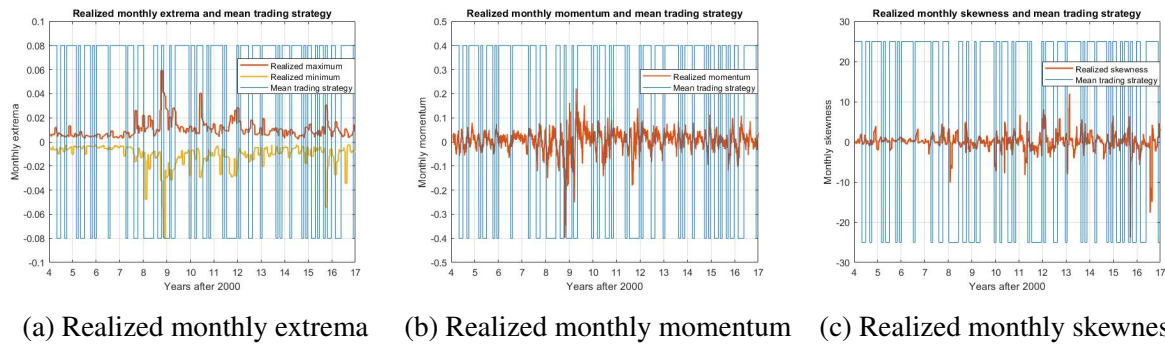


Figure 3.8: Realized monthly extrema, momentum and skewness

ers are apparently somewhat periodic; that for illiquidity notably is much greater for negative lags, perhaps at least partly because illiquidity itself decreases in time. Beyond these rough observations, it is notable that, even with forward and backward lags of up to three years, the trading strategy is apparently overwhelmingly more often than not significantly correlated with each of the realized monthly statistics considered (with 95% confidence intervals about the null hypothesis of no correlation being shown in blue in each plot). The scale of all correlations is also remarkable; at least 15% in either direction in all cases, and in some markedly more so. Again, particularly the case of momentum in figure 3.11e, with a peak correlation over 0.5 at approximately one month’s backward lag, consistent with predictability of the trading strategy. Similarly for volatility in figure 3.11g with correlation about -0.4, maximum in figure 3.11c at about 0.35, volume in figure 3.11h at about -0.3, and minimum in figure 3.11d at about -0.25. In other words, observing one month in advance each of the realized underlying statistics over the preceding month, the trading strategy is more likely to buy than sell, as in decreasing order of importance, momentum is positive, volatility is low, maximum is positive, volume is small, and minimum is negative. In short, persistent positive returns over the preceding month, with a large spread and yet small volatility and trading volume, are highly conducive to buying the underlying, and ideally profitability. This may initially appear contrary to the fact that these same realized monthly statistics were found *not* to be useful predictors of profitability in chapter 2. Notwithstanding the fact that many other predictors are involved there (i.e., the lagged individual constituents of the S&P 500), the present results are in fact not surprising: They consider the correlation of the various realized monthly statistics with the mean ensemble

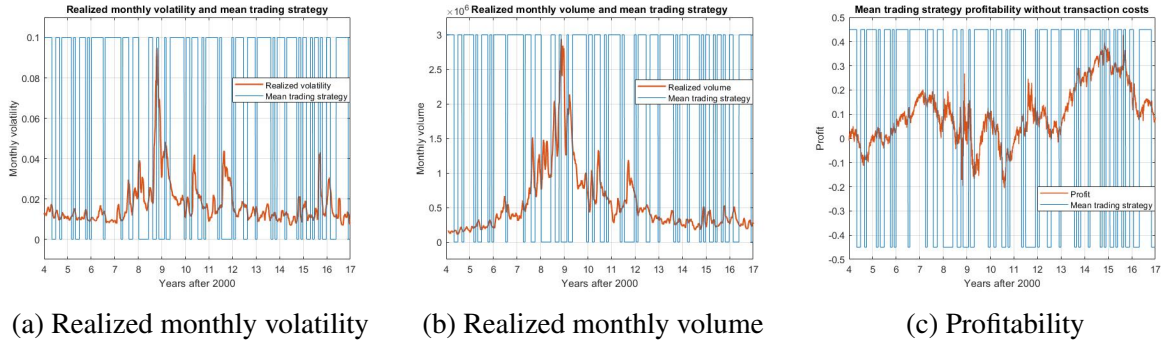


Figure 3.9: Mean strategy profitability and realized monthly volatility and volume

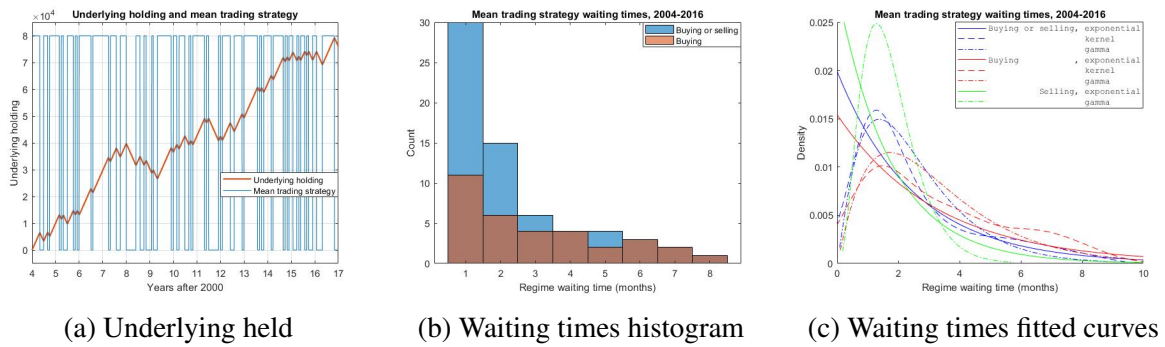


Figure 3.10: Mean strategy net holding and waiting times

ble *trading strategy*, the correlation of which with its own profitability is found in chapter 2 to be (significantly) a fraction of a percentage, so that correlation between the trading strategy and the realized monthly statistics is unlikely to reach to profitability of the former. Also, as noted, figures 3.11 demonstrate significant correlation with an overwhelming majority of lags, forward and backward, up to three years: Clearly, much of the variation of the trading strategy cannot be explained by the top five correlated statistics lagged one month behind. Nonetheless, figures 3.11 plausibly provide a natural interpretation of the mean ensemble trading strategy derived in chapter 2, and the small but significant correlation between that and its profitability, provides some evidence that ‘intuitive’ trading strategies backed by such technical analyses may, albeit barely, ‘beat the market.’

3.4 Conclusions

A profitability comparison between Asian option Δ -hedging, using the results of chapter 1, and machine learning (mean) ensemble predictive trading the S&P 500, using the results of chapter 2, is presented. Interpreting the upper- and lower-bounds on arithmetic Asian option prices, given in section 1.2.4, as hypothetically respectively representing ask and bid values, it is found that hedging profitability depends heavily on both the ability of the trader to secure

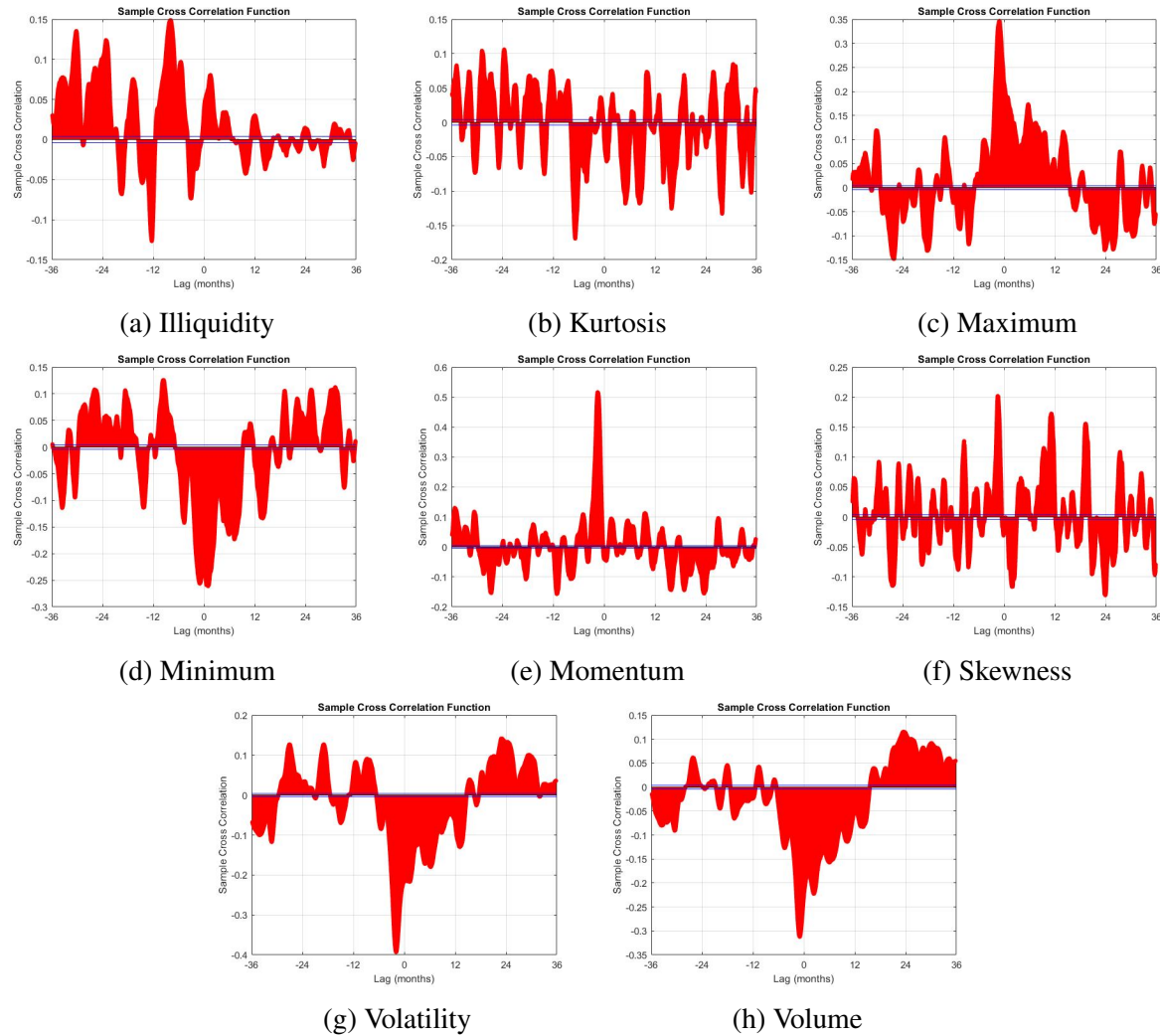


Figure 3.11: Mean trading strategy and realized monthly statistical sample cross correlations

options on a daily basis nearer to the bid than the ask price, and to be willing to accordingly balance at ‘moderately large’ costs, e.g., between roughly \$46.5 and \$650 during the entire period 2004-2016. And when achieved, such profits dwarf those of the mean ensemble trading strategy, despite trading in comparable quantities of the underlying and much less frequently to boot. The latter profits being robust to trading costs, so too are those of the arithmetic Asian Δ -hedging strategy. An array of primarily graphical analyses of sample correlations between the mean ensemble trading strategy derived in chapter 2 and a variety of realized monthly statistics of the underlying S&P 500, provide a natural interpretation of the strategy, and the small but significant correlation between that and its profitability, evidence that ‘intuitive’ trading strategies backed by such technical analyses may ‘beat the market.’

Conclusion

The preceding three essays in financial asset pricing concern *illiquid* markets; one on the partial differential equation pricing and hedging of *power mean* Asian options, via their (optimal) Lie point symmetry groups, leading to practical pricing formulas. Another presents high-frequency machine learning predictions of S&P 500 returns, statistically significantly demonstrating short-horizon market predictability, and economically significantly profitable (beyond transaction costs) trading strategies. The third compares the profitability of the latter mean ensemble strategies and Asian option Δ -hedging. Interpreting bounds on arithmetic Asian option prices as ask and bid values, hedging profitability depends largely on securing prices closer to the bid value, and settling midway between the bid and ask, significant profits are consistently accumulated during the years 2004-2016. In contrast, ensemble predictive trading the S&P 500 yields comparatively very small returns, despite trading much more frequently.

The pricing and hedging of (arithmetic) Asian options poses a difficult problem, which has prompted the development of several solution approaches, differing in theoretical informativeness and practicality. Multiple families of exact solutions to relaxed power mean Asian option pricing boundary-value problems are explicitly established, which approximately satisfy the full pricing problem, and in one case, converge to exact solutions under certain parametric restrictions. Corresponding hedging parameters/Greeks are derived. This family consists of (optimal) invariant solutions, constructed for the pricing partial differential equations of a class of continuous/generalized power mean Asian options, via their (optimal) Lie point symmetry groups. Numerical experiments, in particular, an exhaustive series of Gaussian process regressions, explore the behavior of this family, and lead to accurate, reliable, consistent pricing models suitable for use in practice.

The second chapter analyses and demonstrates the predictability of intraday market returns. Conducting apparently the largest existing study of five-minute market returns using state-of-the-art machine learning models trained on the cross-section of lagged market index constituent returns to forecast five-minute market returns, it is shown that regularized linear models such as lasso and elastic nets along with nonlinear tree-based models such as random forests yield significant predictability. Ensemble models that combine individual model predictions perform the best across time and their return predictability translates into economically significant profits with Sharpe ratios after transaction costs of 0.98. These results provide strong evidence that intraday market returns are predictable during short time horizons, beyond what can be explained by transaction costs. It is further shown that the lagged constituent returns hold significant predictive information that is not contained in lagged market returns or in lagged

price trend and liquidity characteristics. Consistent with the hypothesis that predictability is driven by slow-moving trader capital, predictability decreased after decimalization, and market returns are more predictable during the middle of the day, on days with high volatility or high illiquidity, and in years of financial crisis.

The third chapter presents a profitability comparison between Asian option Δ -hedging, using the results of chapter 1, and machine learning (mean) ensemble predictive trading the S&P 500, using the results of chapter 2. Interpreting the upper- and lower-bounds on arithmetic Asian option prices, given in section 1.2.4, as hypothetically respectively representing ask and bid values, it is found that hedging profitability depends largely on the ability of the option holder to secure prices closer to the bid value, and that, settling midway between the bid and ask, significant profits are consistently accumulated during the years 2004-2016. This is managed with infrequent rebalancing of a self-financing portfolio, according to a backward-difference estimate of the Δ -hedging parameter. In contrast, ensemble predictive trading the S&P 500 yields comparatively very small returns, despite trading much more frequently. Given the robustness of the latter profitability to transaction costs, demonstrated in chapter 2, the same robustness follows as a corollary of the superior hedging returns achieved.

Collectively, in addition to developing extensive, original results for a newly-defined class of power mean Asian options, and in so doing apparently also conducting the first Gaussian process regression pricing of arithmetic Asian options, apparently the largest existing study of five-minute market returns was conducted. The two contributions are tied together via a detailed profitability comparison of derived hedging/trading strategies of each, further emphasizing that both are particularly informative for illiquid markets. These results provide theoretical insights with practical implications, aiming for use and added credibility (and profitability!) in financial practice. For derivatives generally and Asian options, in particular, proper pricing encourages their trade, promoting investment, economic growth, financial institutions' reputations, and social welfare; as do accurate machine learning predictions of intraday market returns, which yield profitability by statistical arbitrage and theoretical insights into the market predictability associated with trader (in)activity.

Bibliography

- [1] Allen Abrahamson et al. Efficient path-dependent valuation using lattices: Fixed and floating strike asian options. Technical report, Citeseer, 2003.
- [2] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1948.
- [3] George Adomian. *Solving frontier problems of physics: the decomposition method*, volume 60. Springer Science & Business Media, 2013.
- [4] Donald Aingworth, Rajeev Motwani, and Jeffrey D Oldham. Accurate approximations for asian options. In *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, pages 891–900. Society for Industrial and Applied Mathematics, 2000.
- [5] Marco Airolidi. A perturbative moment approach to option pricing. *arXiv preprint cond-mat/0401503*, 2004.
- [6] Marco Airolidi*. A moment expansion approach to option pricing. *Quantitative Finance*, 5(1):89–104, 2005.
- [7] Jiro Akahori et al. Some formulae for a new type of path-dependent option. *The Annals of Applied Probability*, 5(2):383–388, 1995.
- [8] Karhan Akcoglu, Ming-Yang Kao, and Shuba V Raghavan. Fast pricing of european asian options with provable accuracy: single-stock and basket options. In *European Symposium on Algorithms*, pages 404–415. Springer, 2001.
- [9] Fares Al-Azemi and Ovidiu Calin. Asian options with harmonic average. *Applied Mathematics & Information Sciences*, 9(6):2803, 2015.
- [10] Laura Alessandretti, Abeer ElBahrawy, Luca Maria Aiello, and Andrea Baronchelli. Anticipating cryptocurrency prices using machine learning. *Complexity*, 2018, 2018.
- [11] Diego Amaya, Peter Christoffersen, Kris Jacobs, and Aurelio Vasquez. Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118(1):135–167, 2015.
- [12] Yakov Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1):31–56, 2002.
- [13] Andrew Ang and Geert Bekaert. Stock return predictability: Is it there? *The Review of Financial Studies*, 20(3):651–707, 2007.
- [14] Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299, 2006.
- [15] Solomon M Antoniou. The arithmetic average case of asian options: Reducing the pde to the black-scholes equation through lie symmetries.
- [16] Solomon M Antoniou. Bond pricing models with closed-form solutions.
- [17] Solomon M Antoniou. A general solution to the stochastic equity volatility equation.
- [18] Solomon M Antoniou. Portfolio optimization and stochastic control. closed-form solutions to the hamilton-jacobi-bellman equation through lie symmetries.
- [19] Solomon M Antoniou. Pricing asian options: the geometric average case. *SKEMSYS report*.
- [20] Hrayr Arahamian and Bacef Maddah. Pricing asian options via compound gamma and orthogonal polynomials. *Applied Mathematics and Computation*, 264:21–43, 2015.
- [21] Taha Aziz, Aeeman Fatima, and Chaudry Masood Khalique. Integrability analysis of the partial differential equation describing the classical bond-pricing model of mathematical finance. *Open Physics*, 16(1):766–779, 2018.
- [22] Belal E Baaquie, Claudio Coriano, and Marakani Srikant. Hamiltonian and potentials in derivative pricing models: exact results and lattice simulations. *Physica A: Statistical Mechanics and its Applications*, 334(3-4):531–557, 2004.
- [23] Ahmet Bakaloglu, Taha Aziz, and FM Mahomed. Invariant criteria for the zero-coupon bond pricing vasicek and cox-ingersoll-ross models. *New Trends in Mathematical Sciences*, 5(2):29, 2017.
- [24] Turan G Bali, Nusret Cakici, and Robert F Whitelaw. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99(2):427–446, 2011.
- [25] Laura Ballotta and Andreas E Kyprianou. A note on the α -quantile option. *Applied Mathematical Finance*, 8(3):137–144, 2001.
- [26] Flavio Barboza, Herbert Kimura, and Edward Altman. Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83:405–417, 2017.
- [27] Ole E Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008.
- [28] Jérôme Barraquand and Thierry Pudet. Pricing of american path-dependent contingent claims. *Mathematical Finance*, 6(1):17–51, 1996.
- [29] Pauline Barrieu, A Rouault, and M Yor. A study of the hartman-watson distribution motivated by numerical problems related to the pricing of asian options. *Journal of Applied Probability*, 41(4):1049–1058, 2004.
- [30] Harry Bateman. Tables of integral transforms. 1954.
- [31] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1):281–305, 2012.
- [32] Jean Bertoin, Loïc Chaumont, and Marc Yor. Two chain-transformations and their applications to quantiles. *Journal of Applied Probability*, 34(4):882–897, 1997.
- [33] Omishwary Bhatoo, Arshad Ahmad Iqbal Peer, Eitan Tadmor, Désiré Tangman, and Aslam Aly El Faidal Saib. Efficient conservative second-order central-upwind schemes for option-pricing problems. *Journal of Computational Finance*, 22(5), 2019.
- [34] Omishwary Bhatoo, Arshad Ahmad Iqbal Peer, Eitan Tadmor, Désiré Yannick Tangman, and Aslam Aly El Faidal Saib. Conservative third-order central-upwind schemes for option pricing problems. *Viennan Journal of Mathematics*, 47(4):813–833, 2019.
- [35] Daniele Bianchi, Matthias Büchner, and Andrea Tamoni. Bond risk premia with machine learning. *forthcoming in The Review of Financial Studies*, 2020.
- [36] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [37] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.
- [38] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.
- [39] George W Bluman and Sukeyuki Kumei. *Symmetries and differential equations*, volume 81. Springer Science & Business Media, 2013.
- [40] Vincent Bogousslavsky. Infrequent rebalancing, return autocorrelation, and seasonality. *The Journal of Finance*, 71(6):2967–3006, 2016.
- [41] G Bormetti, Guido Montagna, N Moreni, and Oreste Nicosini. Pricing exotic options in a path integral approach. *Quantitative Finance*, 6(1):55–66, 2006.
- [42] Phelim Boyle and Alexander Potapchik. Prices and sensitivities of asian options: A survey. *Insurance: Mathematics and Economics*, 42(1):189–211, 2008.
- [43] Phelim P Boyle. Options: A monte carlo approach. *Journal of financial economics*, 4(3):323–338, 1977.
- [44] Y Bozhkov and S Dimas. Enhanced group analysis of a semi linear generalization of a general bond-pricing equation. *Communications in Nonlinear Science and Numerical Simulation*, 54:210–220, 2018.
- [45] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- [46] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [47] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. Classification and regression trees. belmont, ca: Wadsworth. *International Group*, 432:151–166, 1984.
- [48] Mark Broadie and Paul Glasserman. Estimating security price derivatives using simulation. *Management science*, 42(2):269–285, 1996.
- [49] Iurii Brychkov. *Integral transforms of generalized functions*.
- [50] Svetlana Bryzgalova, Markus Pelger, and Jason Zhu. Forest through the trees: Building cross-sections of stock returns. *Available at SSRN 3493458*, 2019.
- [51] Herbert Buchholz. *The confluent hypergeometric function: with special emphasis on its applications*, volume 15. Springer Science & Business Media, 2013.
- [52] Peter S Bullen. *Handbook of means and their inequalities*, volume 560. Springer Science & Business Media, 2013.
- [53] Ning Cai. Pricing and hedging of quantile options in a flexible jump diffusion model. *Journal of applied probability*, 48(3):637–656, 2011.
- [54] Ning Cai, Nan Chen, and Xiangwei Wan. Occupation times of jump-diffusion processes with double exponential jumps and the pricing of options. *Mathematics of Operations Research*, 35(2):412–437, 2010.
- [55] NC Caister, KS Govinder, and JG O’Hara. Optimal system of lie group invariant solutions for the asian option pde. *Mathematical methods in the applied sciences*, 34(11):1353–1365, 2011.
- [56] Nicolette C Caister, JOHN G O’HARA, and Keshlan S Govinder. Solving the asian option pde using lie symmetry methods. *International Journal of Theoretical and Applied Finance*, 13(08):1265–1277, 2010.
- [57] John Y Campbell and Samuel B Thompson. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531, 2008.

- [58] Peter Carr and Michael Schröder. On the valuation of arithmetic-average asian options: the geman-yor laplace transform revisited. *arXiv preprint math/0102080*, 2001.
- [59] Peter Carr and Michael Schröder. Bessel processes, the integral of geometric brownian motion, and asian options. *Theory of Probability & Its Applications*, 48(3):400–425, 2004.
- [60] Horatio Scott Carslaw and John Conrad Jaeger. Conduction of heat in solids. *Oxford: Clarendon Press, 1959, 2nd ed.*, 1959.
- [61] Pierre Cartier and Cecilie DeWitt-Morette. A rigorous mathematical foundation of functional integration. In *Functional Integration*, pages 1–50. Springer, 1997.
- [62] Zhongdi Cen, Anbo Le, and Aimin Xu. Finite difference scheme with a moving mesh for pricing asian options. *Applied Mathematics and Computation*, 219(16):8667–8675, 2013.
- [63] George Chacko and Sanjiv Das. Pricing interest rate derivatives: a general approach. *The review of financial studies*, 15(1):195–241, 2002.
- [64] George Chacko and Sanjiv Ranjan Das. Average interest. Technical report, National Bureau of Economic Research, 1997.
- [65] Prasad Chalasani, Somesh Jha, and Isaac Saias. Approximate option pricing. *Algorithmica*, 25(1):2–21, 1999.
- [66] Prasad Chalasani, Somesh Jha, and Ashok Varikooty. Accurate approximations for european-style asian options. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE, 1997.
- [67] K Charalambous and PGL Leach. Financial derivatives and lie symmetries. *Transactions of the Royal Society of South Africa*, 70(1):1–7, 2015.
- [68] Jiaqi Chen, Wenbo Wu, and Michael L Tindall. Hedge fund return prediction and fund selection: A machine-learning approach. *Occasional Paper*, 16:04, 2016.
- [69] Kuan-Wen Chen and Yuh-Dauh Lyuu. Accurate pricing formulas for asian options. *Applied Mathematics and Computation*, 188(2):1711–1724, 2007.
- [70] Long Chen, Zhi Da, and Xinlei Zhao. What drives stock price movements? *The Review of Financial Studies*, 26(4):841–876, 2013.
- [71] Luyang Chen, Markus Pelger, and Jason Zhu. Deep learning in asset pricing. Available at SSRN 3350138, 2019.
- [72] Tatiana P Chernogorova and Lubin G Vulkov. Two splitting methods for a fixed strike asian option. In *International Conference on Numerical Analysis and Its Applications*, pages 214–221. Springer, 2012.
- [73] Terry HF Cheuk and Ton CF Vorst. Average interest rate caps. *Computational Economics*, 14(3):183–196, 1999.
- [74] Alex Chincio, Adam D Clark-Joseph, and Mao Ye. Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1):449–492, 2019.
- [75] Alex Chincio and Vyacheslav Fos. The sound of many funds rebalancing. 2019.
- [76] Eunsuk Chong, Chulwoo Han, and Frank C Park. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83:187–205, 2017.
- [77] Kam Yoon Chong and John G O'Hara. Lie symmetry analysis of a fractional black-scholes equation. In *AIP Conference Proceedings*, volume 2153, page 020007. AIP Publishing LLC, 2019.
- [78] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics*, 76(2):271–292, 2005.
- [79] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Liquidity and market efficiency. *Journal of Financial Economics*, 87(2):249–268, 2008.
- [80] Tarun Chordia, Avanidhar Subrahmanyam, and V Ravi Anshuman. Trading activity and expected stock returns. *Journal of financial Economics*, 59(1):3–32, 2001.
- [81] Gennaro Cibelli, Sergio Polidoro, and Francesco Rossi. Sharp estimates for geman–yor processes and applications to arithmetic average asian options. *Journal de Mathématiques Pures et Appliquées*, 129:87–130, 2019.
- [82] R Cimpoiu and R Constantinescu. New symmetries and particular solutions for 2d black-scholes model. 2013.
- [83] Gonzalo Cortazar and Eduardo S Schwartz. The valuation of commodity contingent claims. *Journal of Derivatives*, 1(4):27–39, 1994.
- [84] Massimo Costabile, Ivar Massabó, and Emilio Russo. An adjusted binomial model for pricing asian options. *Review of Quantitative Finance and Accounting*, 27(3):285–296, 2006.
- [85] John C Cox, Stephen A Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of financial Economics*, 7(3):229–263, 1979.
- [86] J David Cummins and Hélyette Geman. An asian option approach to the valuation of insurance futures contracts. *Review of Futures Markets*, 13:517–517, 1994.
- [87] Tian-Shyr Dai, Guan-Shieng Huang, and Yuh-Dauh Lyuu. Extremely accurate and efficient tree algorithms for asian options with range bounds. In *2002 NTU International Conference on Finance, National Taiwan University, Taiwan*, pages 5–2, 2002.
- [88] Tian-Shyr Dai and Yuh-Dauh Lyuu. Efficient, exact algorithms for asian options with multiresolution lattices. *Review of Derivatives Research*, 5(2):181–203, 2002.
- [89] Tian-Shyr Dai and Yuh-Dauh Lyuu. An exact subexponential-time lattice algorithm for asian options. *Acta informatica*, 44(1):23–39, 2007.
- [90] Tian-Shyr Dai, Jr-Yan Wang, and Hui-Shan Wei. An ingenious, piecewise linear interpolation algorithm for pricing arithmetic average options. In *International Conference on Algorithmic Applications in Management*, pages 262–272. Springer, 2007.
- [91] Tian-Shyr Dai, Jr-Yan Wang, and Hui-Shan Wei. Adaptive placement method on pricing arithmetic average options. *Review of Derivatives Research*, 11(1-2):83, 2008.
- [92] Angelos Dassios. The distribution of the quantile of a brownian motion with drift and the pricing of related path-dependent options. *The Annals of Applied Probability*, pages 389–398, 1995.
- [93] Angelos Dassios. Sample quantiles of additive renewal reward processes. *Journal of applied probability*, 33(4):1018–1032, 1996.
- [94] AH Davison and S Mamba. Symmetry methods for option pricing. *Communications in Nonlinear Science and Numerical Simulation*, 47:421–425, 2017.
- [95] Min-Yuh Day and Jian-Ting Lin. Artificial intelligence for etf market prediction and portfolio optimization. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1026–1033, 2019.
- [96] Cécile DeWitt-Morette and Antoine Polacci. *Functional Integration: Basics and Applications*, volume 361. Springer Science & Business Media, 2013.
- [97] S Dimas, K Andriopoulos, D Tsoubelis, and PGL Leach. Complete specification of some partial differential equations that arise in financial mathematics. *Journal of Nonlinear Mathematical Physics*, 16(sup01):73–92, 2009.
- [98] Kemal Dinçer Dingerç and Wolfgang Hörmann. Control variates and conditional monte carlo for basket and asian options. *Insurance: Mathematics and Economics*, 52(3):421–434, 2013.
- [99] Kemal Dinçer Dingerç and Wolfgang Hörmann. Improved monte carlo and quasi-monte carlo methods for the price and the greeks of asian options. In *Proceedings of the Winter Simulation Conference 2014*, pages 441–452. IEEE, 2014.
- [100] Satya D Dubey. Compound gamma, beta and f distributions. *Metrika*, 16(1):27–31, 1970.
- [101] Darrell Duffie. Presidential address: Asset price dynamics with slow-moving capital. *The Journal of Finance*, 65(4):1237–1267, 2010.
- [102] Daniel Dufresne. Laguerre series for asian and other options. *Mathematical Finance*, 10(4):407–428, 2000.
- [103] Daniel Dufresne. Asian and basket asymptotics. 2002.
- [104] Daniel Dufresne. The log-normal approximation in financial and other computations. *Advances in Applied Probability*, 36(3):747–773, 2004.
- [105] Daniel Dufresne and Hanbo Li. Pricing asian options: Convergence of gram-charlier series. 2014.
- [106] Anirudha Dutta, Saket Kumar, and Meheli Basu. A gated recurrent unit approach to bitcoin price prediction. *Journal of Risk and Financial Management*, 13(2):23, 2020.
- [107] Stephen A Easton. Valuing bonds with embedded average price options. *Australian Journal of Management*, 21(1):29–40, 1996.
- [108] SO Edeki, GO Akinlabi, and O González-Gaxiola. Adomian decomposition method for analytical solution of a continuous arithmetic asian option pricing model. *Telkomika*, 17(2):866–872, 2019.
- [109] Sunday O Edeki, Tanki Motsepa, Chaudry Masood Khalique, and Grace O Akinlabi. The greek parameters of a continuous arithmetic asian option pricing model via laplace adomian decomposition method. *Open Physics*, 16(1):780–785, 2018.
- [110] Sunday O EDEKI, Olabisi O UGBEBOR, and Paul O OGUNDILE. Analytical solutions of a continuous arithmetic asian model for option pricing using projected differential transform method. *Engineering Letters*, 27(2), 2019.
- [111] P Embrechts, LCG Rogers, M Yor, et al. A proof of dassios' representation of the α -quantile of brownian motion with drift. *The Annals of Applied Probability*, 5(3):757–767, 1995.
- [112] Javier Esparza, Jose L Lopez, and Javier Sesma. Zeros of the whittaker function associated to coulomb waves. *IMA journal of applied mathematics*, 63(1):71–87, 1999.
- [113] Christian-Oliver Ewald, Olaf Menkens, and Sai Hung Marten Ting. Asian and australian options: A common perspective. *Journal of Economic Dynamics and Control*, 37(5):1001–1018, 2013.
- [114] Alexander Eydeland and Hélyette Geman. Fundamentals of electricity derivatives. *Energy Modelling and the Management of Uncertainty*, pages 35–43, 1999.
- [115] Eugene F Fama. Efficient capital markets: Ii. *The journal of finance*, 46(5):1575–1617, 1991.
- [116] Viviana Fanelli, Lucia Maddalena, and Silvana Musti. Asian options pricing in the day-ahead electricity market. *Sustainable cities and society*, 27:196–202, 2016.
- [117] Tom Fawcett. Using rule sets to maximize roc performance. In *Proceedings 2001 IEEE international conference on data mining*, pages 131–138. IEEE, 2001.
- [118] Guan hao Feng, Jingyu He, and Nicholas G Polson. Deep learning for predicting asset returns. *arXiv preprint arXiv:1804.09314*, 2018.
- [119] Guan hao Feng, Nick Polson, and Jianeng Xu. Deep learning in characteristics-sorted factor models. Available at SSRN 3243683, 2019.
- [120] Thomas Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669, 2018.
- [121] PA Forsyth, KR Vetzal, and R Zvan. Convergence of lattice and pde methods for pricing asian options. In *Proceedings of the Ninth Annual Derivative Securities Conference, Boston*. Citeseer, 1999.
- [122] Georgios Fofas and Mats G Larson. Valuing asian options using the finite element method and duality techniques. *Journal of computational and applied mathematics*, 222(1):144–158, 2008.
- [123] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.
- [124] Avner Friedman. *Generalized functions and partial differential equations*. Courier Corporation, 2005.
- [125] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.
- [126] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [127] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.
- [128] Michael C Fu, Dilip B Madan, and Tong Wang. Pricing continuous asian options: a comparison of monte carlo and laplace transform inversion methods. *Journal of Computational*

- Finance*, 2(2):49–74, 1999.
- [129] Hideharu Funahashi and Masaaki Kijima. A unified approach for the pricing of options relating to averages. *Review of Derivatives Research*, 20(3):203–229, 2017.
- [130] Gianluca Fusai. Pricing asian options via fourier and laplace transforms. *Journal of computational finance*, 7(3):87–106, 2004.
- [131] Gianluca Fusai, Marina Marena, and Andrea Roncoroni. Analytical pricing of discretely monitored asian-style options: Theory and application to commodity markets. *Journal of Banking & Finance*, 32(10):2033–2045, 2008.
- [132] Gianluca Fusai and Aldo Tagliani. An accurate valuation of asian options using moments. *International Journal of Theoretical and Applied Finance*, 5(02):147–169, 2002.
- [133] Lei Gao, Yufeng Han, Sophia Zhengzi Li, and Guofu Zhou. Market intraday momentum. *Journal of Financial Economics*, 129(2):394–414, 2018.
- [134] Marcellino Gaudenzi, Antonino Zanette, and Maria Antonietta Lepellere. The singular points binomial method for pricing american path-dependent options. *The Journal of Computational Finance*, 14(1):29, 2010.
- [135] Rafail K Gazizov and Nail H Ibragimov. Lie symmetry analysis of differential equations in finance. *Nonlinear Dynamics*, 17(4):387–407, 1998.
- [136] Hélyette Geman. The bermuda triangle: weather, electricity, and insurance derivatives. *The Journal of Alternative Investments*, 3(1):61–69, 2000.
- [137] Hélyette Geman. From bachelier and lundberg to insurance and weather derivatives. *Mathematical Physics Studies*, Kluwer Academic Publishers, pages 81–95, 2000.
- [138] Hélyette Geman. Functionals of brownian motion in path-dependent option valuation. In *European Congress of Mathematics*, pages 379–390. Springer, 2001.
- [139] Hélyette Geman. Time changes, laplace transforms and path-dependent options. *Computational Economics*, 17(1):81–92, 2001.
- [140] Hélyette Geman. Towards a european market of electricity: spot and derivatives trading. In *Presented at Joint IEA/NEA workshop, 25–26 March, 2003, Paris, France.*, 2002.
- [141] Hélyette Geman and Marc Yor. Bessel processes, asian options, and perpetuities. *Mathematical finance*, 3(4):349–375, 1993.
- [142] Véronique Genre, Geoff Kenny, Aidan Meyler, and Allan Timmermann. Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, 29(1):108–121, 2013.
- [143] Thomas Gerstner and Marco Noll. Randomized multilevel quasi-monte carlo path simulation. In *Recent Developments in Computational Finance: Foundations, Algorithms and Applications*, pages 349–369. World Scientific, 2013.
- [144] Michael B Giles. Multilevel monte carlo path simulation. *Operations research*, 56(3):607–617, 2008.
- [145] Michael B Giles, Kristian Debrabant, and Andreas Rössler. Analysis of multilevel monte carlo path simulation using the milstein discretisation. *Discrete & Continuous Dynamical Systems-B*, 24(8):3881, 2019.
- [146] Mike Giles. Improved multilevel monte carlo convergence using the milstein scheme. In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer, 2008.
- [147] Paul Glasserman, Philip Heidelberger, and Perwez Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical finance*, 9(2):117–152, 1999.
- [148] J Goard. New solutions to the bond-pricing equation via lie’s classical method. *Mathematical and Computer Modelling*, 32(3-4):299–313, 2000.
- [149] M Barry Goldman, Howard B Sosin, and Mary Ann Gatto. Path dependent options: “buy at the low, sell at the high”. *The Journal of Finance*, 34(5):1111–1127, 1979.
- [150] Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. *arXiv preprint arXiv:1302.4389*, 2013.
- [151] Yuri N Grigoriev, Vladimir F Kovalev, Sergey V Meleshko, and Nail H Ibragimov. *Symmetries of integro-differential equations: with applications in mechanics and plasma physics*. Springer, 2010.
- [152] Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. Technical report, National Bureau of Economic Research, 2018.
- [153] Shihao Gu, Bryan Kelly, and Dacheng Xiu. Autoencoder asset pricing models. *Journal of Econometrics*, 2020.
- [154] Petr Hajek and Aliaksandr Barushka. Integrating sentiment analysis and topic detection in financial news for stock movement prediction. In *Proceedings of the 2nd International Conference on Business and Information Management*, pages 158–162, 2018.
- [155] Chuan-Hsiang Han and Yongzeng Lai. Generalized control variate methods for pricing asian options. *Journal of Computational Finance*, 14(2):87, 2010.
- [156] MS Hashemi. On black-scholes equation, method of heir-equations, nonlinear self-adjointness and conservation laws. 2016.
- [157] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.
- [158] Espen Gaarder Røug. *The complete guide to option pricing formulas*, volume 2. McGraw-Hill New York, 2007.
- [159] Isabel Hernández, Consuelo Mateos, Juan Núñez, and Ángel F Tenorio. Lie theory: applications to problems in mathematical finance and economics. *Applied Mathematics and Computation*, 208(2):446–452, 2009.
- [160] Steven L Heston, Robert A Korajczyk, and Ronnie Sadka. Intraday patterns in the cross-section of stock returns. *The Journal of Finance*, 65(4):1369–1407, 2010.
- [161] Steven L Heston and Nitish Ranjan Sinha. News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, 73(3):67–83, 2017.
- [162] Akos Horvath and Peter Medvegyev. Pricing asian options: A comparison of numerical and simulation approaches, twenty years later. 2016.
- [163] William Wei-Yuan Hsu and Yuh-Dauh Lyuu. A convergent quadratic-time lattice algorithm for pricing european-style asian options. *Applied mathematics and computation*, 189(2):1099–1123, 2007.
- [164] John C Hull and Alan D White. Efficient procedures for valuing european and american path-dependent options. *The Journal of Derivatives*, 1(1):21–31, 1993.
- [165] Hideki Iwaki, Masaaki Kijima, and Toshihiro Yoshida. Approximate valuation of average options. *Annals of Operations Research*, 45(1):131–145, 1993.
- [166] Edwin T Jaynes. Where do we stand on maximum entropy? *The maximum entropy formalism*, 15, 1979.
- [167] Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91, 1993.
- [168] Lishang Jiang and Min Dai. Convergence of binomial tree methods for european/american path-dependent options. *SIAM Journal on Numerical Analysis*, 42(3):1094–1109, 2004.
- [169] I.T. Jolliffe and Springer-Verlag. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002.
- [170] Nengjia Ju. Pricing asian and basket options via taylor expansion. *Journal of Computational Finance*, 5(3):79–103, 2002.
- [171] Bosiu C Kaibe and John G O’Hara. Symmetry analysis of an interest rate derivatives pde model in financial mathematics. *Symmetry*, 11(8):1056, 2019.
- [172] Zheng Tracy Ke, Bryan T Kelly, and Dacheng Xiu. Predicting returns with text data. Technical report, National Bureau of Economic Research, 2019.
- [173] Ahmed Kebaier et al. Statistical romberg extrapolation: a new variance reduction method and applications to option pricing. *The Annals of Applied Probability*, 15(4):2681–2705, 2005.
- [174] Bryan Kelly and Seth Pruitt. The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2):294–316, 2015.
- [175] Bryan Kelly, Seth Pruitt, and Yinan Su. Instrumented principal component analysis. *Available at SSRN 2983919*, 2019.
- [176] Bryan T Kelly, Seth Pruitt, and Yinan Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524, 2019.
- [177] Angelien GZ Kemna and ACF Vorst. A pricing method for options based on average asset values. *Journal of Banking & Finance*, 14(1):113–129, 1990.
- [178] Chaudry Masood Khaliq and Tanki Motsepa. Lie symmetries, group-invariant solutions and conservation laws of the vasicek pricing equation of mathematical finance. *Physica A: Statistical Mechanics and its Applications*, 505:871–879, 2018.
- [179] Timothy Klassen. Simple, fast and flexible pricing of asian options. *Final version in: Journal of Computational Finance*, 4(3):89–124, 2001.
- [180] Steen Koekebakker, Roar Adland, and Sigbjørn Sødal. Pricing freight rate options. *Transportation Research Part E: Logistics and Transportation Review*, 43(5):535–548, 2007.
- [181] Ralph SJ Koijen and Stijn Van Nieuwerburgh. Predictability of returns and cash flows. *Annu. Rev. Financ. Econ.*, 3(1):467–491, 2011.
- [182] Yehuda Koren. The bellco solution to the netflix grand prize. *Netflix prize documentation*, 81(2009):1–10, 2009.
- [183] Sofonias A Korsaye, Alberto Quaini, and Fabio Trojani. Smart sdfs. *Available at SSRN 3475451*, 2019.
- [184] Alpesh Kumar, Lok Pati Tripathi, and Mohan K Kadalbajoo. A numerical study of asian option with radial basis functions based finite differences method. *Engineering Analysis with Boundary Elements*, 50:1–7, 2015.
- [185] Yue-Kuen Kwok. *Mathematical models of financial derivatives*. Springer, 2008.
- [186] Jean-Bernard Lasserre, Tomas Prieto-Rumeau, and Mihail Zervos. Pricing a class of exotic options via moments and sdp relaxations. *Mathematical Finance*, 16(3):469–494, 2006.
- [187] PGL Leach and K Andriopoulos. Newtonian economics. In *Proceedings of 10th International Conference in Modern group analysis*, volume 134, page 142, 2005.
- [188] PGL Leach and K Andriopoulos. Algebraic solutions of some partial differential equations arising in financial mathematics and related areas. *Bulletin of the Greek Mathematical Society*, 54:15–27, 2007.
- [189] PGL Leach, RM Morris, and A Paliathanasis. The algebraic properties of the space-and time-dependent one-factor model of commodities. *Quaestiones Mathematicae*, 40(1):91–106, 2017.
- [190] PGL Leach, JG O’Hara, and W Sinkala. Symmetry-based solution of a model for a combination of a risky investment and a riskless investment. *Journal of mathematical analysis and applications*, 334(1):368–381, 2007.
- [191] Edmond Levy. Pricing european average rate currency options. *Journal of International Money and Finance*, 11(5):474–491, 1992.
- [192] Weiping Li and Su Chen. Pricing and hedging of arithmetic asian options via the edgeworth series expansion approach. *The Journal of Finance and Data Science*, 2(1):1–25, 2016.
- [193] Shijun Liao. *Homotopy analysis method in nonlinear differential equations*. Springer, 2012.
- [194] Sophus Lie and G Scheffers. Lectures on differential equations with known infinitesimal transformations. *BG Teubner, Leipzig*, 1891.
- [195] Tiong Wee Lim. Pricing and hedging asian options: a recursive integration approach.
- [196] Tiong Wee Lim. Performance of recursive integration for pricing european-style asian options. *preprint*, 2002.
- [197] Vadim Linetsky. The path integral approach to financial modeling and options pricing. *Computational Economics*, 11(1-2):129–163, 1997.
- [198] Vadim Linetsky. Exact pricing of asian options: an application of spectral theory. *Working Paper*, 2002.
- [199] Vadim Linetsky. Exotic spectra. *Risk*, pages 85–89, 2002.
- [200] Vadim Linetsky. Spectral expansions for asian (average price) options. *Operations Research*, 52(6):856–867, 2004.
- [201] John Lintner. Security prices, risk, and maximal gains from diversification. *The journal of finance*, 20(4):587–615, 1965.
- [202] Yifang Liu and Deng-Shan Wang. Symmetry analysis of the option pricing model with dividend yield from financial markets. *Applied mathematics letters*, 24(4):481–486, 2011.
- [203] CF Lo, PH Yuen, and CH Hui. Option risk measurement with time-dependent parameters. *International Journal of Theoretical and Applied Finance*, 3(03):581–589, 2000.

- [204] Chi-Fai Lo. Lie-algebraic approach for pricing zero-coupon bonds in single-factor interest rate models. *Journal of Applied Mathematics*, 2013, 2013.
- [205] Chi-Fai Lo and CH Hui. Pricing multi-asset financial derivatives with time-dependent parameters—lie algebraic approach. *International Journal of Mathematics and Mathematical Sciences*, 32(7):401–410, 2002.
- [206] Chi Fai Lo and Cho-Hoi Hui. Lie-algebraic approach for pricing moving barrier options with time-dependent parameters. *Journal of Mathematical Analysis and Applications*, 323(2):1455–1464, 2006.
- [207] Chi Fai Lo, PH Yuen, and Cho-Hoi Hui. Constant elasticity of variance option pricing model with time-dependent parameters. *International Journal of Theoretical and Applied Finance*, 3(04):661–674, 2000.
- [208] Peter Løchte Jørgensen. Lognormal approximation of complex path-dependent pension scheme payoffs. *The European Journal of Finance*, 13(7):595–619, 2007.
- [209] Wen Long, Zhichen Lu, and Lingxiao Cui. Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, 164:163–173, 2019.
- [210] Francis A Longstaff and Eduardo S Schwartz. A two-factor interest rate model and contingent claims valuation. *The Journal of Fixed Income*, 2(3):16–23, 1992.
- [211] Roger Lord. Partially exact and bounded approximations for arithmetic asian options. Available at SSRN 678041, 2005.
- [212] Dong Lou, Christopher Polk, and Spyros Skouras. A tug of war: Overnight versus intraday expected returns. *Journal of Financial Economics*, 2019.
- [213] Ghassan RI Majed. A survey of financial derivatives utilised within the petroleum industry. *OPEC review*, 20(1):87–115, 1996.
- [214] Michael D Marozzi. An adaptive extrapolation discontinuous galerkin method for the valuation of asian options. *Journal of computational and applied mathematics*, 235(12):3632–3645, 2011.
- [215] Ivana Marković, Miloš Stojanović, Jelena Stanković, and Milena Stanković. Stock market trend prediction using ahp and weighted kernel ls-svm. *Soft Computing*, 21(18):5387–5398, 2017.
- [216] Donald W Marquardt. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612, 1970.
- [217] TSHDISO Masebe and JACOB Manale. New symmetries of black-scholes equation. *Proc. AMCM*, 13:221–231, 2013.
- [218] Andrew Matacz. Path dependent option pricing: the path integral partial averaging method. In *AFA 2001 New Orleans Meetings*, 1999.
- [219] P McCullagh and JA Nelder. Generalized linear models., 2nd edn.(chapman and hall: London). *Standard book on generalized linear models*, 1989.
- [220] Farshid Mehrdoust. A new hybrid monte carlo simulation for asian options pricing. *Journal of Statistical Computation and Simulation*, 85(3):507–516, 2015.
- [221] Sergey V Meleshko. *Methods for constructing exact solutions of partial differential equations: mathematical and analytical techniques with applications to engineering*. Springer Science & Business Media, 2006.
- [222] Moshe Arye Milevsky and Steven E Posner. Asian options, the sum of lognormals, and the reciprocal gamma distribution. *Journal of financial and quantitative analysis*, 33(3):409–422, 1998.
- [223] Ryoza Miura. A note on look-back options based on order statistics. *Hitsubashi Journal of Commerce and Management*, pages 15–28, 1992.
- [224] Kyoung-Sook Moon, Yunju Jeong, and Hongjoong Kim. An efficient binomial method for pricing asian options. *Economic Computation & Economic Cybernetics Studies & Research*, 50(2), 2016.
- [225] Manuel Moreno and Javier F Navas. Australian options. *Australian Journal of Management*, 33(1):69–93, 2008.
- [226] Tanki Motsepa, Chaudry Masood Khaliq, and Moltati Molati. Group classification of a general bond-option pricing equation of mathematical finance. In *Abstract and Applied Analysis*, volume 2014. Hindawi, 2014.
- [227] V Naicker, K Andriopoulos, and PGL Leach. Symmetry reductions of a hamilton-jacobi-bellman equation arising in financial mathematics. *Journal of Nonlinear Mathematical Physics*, 12(2):268–283, 2005.
- [228] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.
- [229] C Nasim. On k-transform. *International Journal of Mathematics and Mathematical Sciences*, 4, 1981.
- [230] Radford M Neal. *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media, 2012.
- [231] Edwin H Neave. A frequency distribution method for valuing average options. *ASTIN Bulletin: The Journal of the IAA*, 27(2):173–205, 1997.
- [232] Edwin H Neave and George L Ye. Pricing asian options using path bundling. Available at SSRN 284888, 2000.
- [233] Christopher J Neely, David E Rapach, Jun Tu, and Guofu Zhou. Forecasting the equity risk premium: the role of technical indicators. *Management science*, 60(7):1772–1791, 2014.
- [234] Jørgen Aase Nielsen, Klaus Sandmann, et al. *Asian exchange rate options under stochastic interest rates: Pricing as a sum of delayed payment options*. Rheinische Friedrich-Wilhelms-Universität Bonn, 1998.
- [235] Aleksandr Aleksandrovich Novikov and NE Kordzakhia. Lower and upper bounds for prices of asian-type options. *Proceedings of the Steklov Institute of Mathematics*, 287(1):225–231, 2014.
- [236] Alexander Novikov and Nino Kordzakhia. On lower and upper bounds for asian-type options: a unified approach. *arXiv preprint arXiv:1309.2383*, 2013.
- [237] Chuma Raphael Nwozo and Sunday Emmanuel Fadugba. Performance measure of laplace transforms for pricing path dependent options. *International Journal of Pure and Applied Mathematics*, 94(2):175–197, 2014.
- [238] Ken'ichiro Ohta, Kunihiko Sadakane, Akiyoshi Shioura, and Takeshi Tokuyama. A fast, accurate, and simple method for pricing european-asian and saving-asian options. *Algorithmica*, 42(2):141–158, 2005.
- [239] Peter J Olver. *Applications of Lie groups to differential equations*, volume 107. Springer Science & Business Media, 2000.
- [240] L Vo Ovsyannikov. *Lectures on the theory of group properties of differential equations*. World Scientific Publishing Co Inc, 2013.
- [241] John G O'Hara, Christodoulos Sophocleous, and Peter GL Leach. Symmetry analysis of a model for the exercise of a barrier option. *Communications in Nonlinear Science and Numerical Simulation*, 18(9):2367–2373, 2013.
- [242] Andronikos Paliathanasis, K Krishnakumar, KM Tamizhmani, and Peter GL Leach. Lie symmetry analysis of the black-scholes-merton model for european options with stochastic volatility. *Mathematics*, 4(2):28, 2016.
- [243] Andronikos Paliathanasis, Richard M Morris, and Peter GL Leach. Lie symmetries of $(1+2)$ nonautonomous evolution equations in financial mathematics. *Mathematics*, 4(2):34, 2016.
- [244] VIJAY K PARMAR. Pricing asian average strike options-a lie algebraic approach. 2014.
- [245] Andrea Pascucci. *PDE and martingale methods in option pricing*. Springer Science & Business Media, 2011.
- [246] Kuldeep Singh Patel and Mani Mehra. A numerical study of asian option with high-order compact finite difference scheme. *Journal of Applied Mathematics & Computing*, 57(1-2):467–491, 2018.
- [247] Kuldeep Singh Patel and Mani Mehra. High-order compact finite difference scheme for pricing asian option with moving boundary condition. *Differential Equations and Dynamical Systems*, 27(1-3):39–56, 2019.
- [248] Andreas Pechtl. Some applications of occupation times of brownian motion with drift in mathematical finance. *Advances in Decision Sciences*, 3(1):63–73, 1999.
- [249] Jacques Pézier and Johanna Scheller. Average price portfolio insurance as optimal implementation of life-cycle investment strategies. 2012.
- [250] Carolyn E Phelan, Daniele Marazzina, and Guido Germano. Pricing methods for α -quantile and perpetual early exercise options based on spitzer identities. *Quantitative Finance*, pages 1–20, 2020.
- [251] Martial Phélippé-Guinvarc'H and Jean Cordier. An option of the average european futures prices for an efficient hog producer risk management. 2010.
- [252] Refet Polat. Lie symmetries of the black-scholes equation. *Selçuk Journal of Applied Mathematics*, 11(2):85–91, 2010.
- [253] Tanvi Rai and Siddhartha P Chakrabarty. Pricing asian call option with average strike using a non-uniform grid. *Journal of Interdisciplinary Mathematics*, 16(2-3):191–201, 2013.
- [254] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Numerical gaussian processes for time-dependent and nonlinear partial differential equations. *SIAM Journal on Scientific Computing*, 40(1):A172–A198, 2018.
- [255] Germán I Ramírez-Espinoza and Matthias Ehrhardt. Conservative and finite volume methods for the convection-dominated pricing problem. *Advances in Applied Mathematics and Mechanics*, 5(6):759–790, 2013.
- [256] David Rapach and Guofu Zhou. Forecasting stock returns. In *Handbook of economic forecasting*, volume 2, pages 328–383. Elsevier, 2013.
- [257] Patrick Rebenstros, Brajesh Gupta, and Thomas R Bromley. Quantum computational finance: Monte carlo pricing of financial derivatives. *Physical Review A*, 98(2):022321, 2018.
- [258] Thomas Renault. Intraday online investor sentiment and return patterns in the us stock market. *Journal of Banking & Finance*, 84:25–40, 2017.
- [259] Peter Ritchken and L Sankarasubramanian. Averaging and deferred payment yield agreements. *The Journal of Futures Markets (1986-1998)*, 13(1):23, 1993.
- [260] Peter Ritchken, L Sankarasubramanian, and Anand M Vijh. The valuation of path dependent contracts on the average. *Management Science*, 39(10):1202–1213, 1993.
- [261] Tarcísio M Rocha Filho and Annibal Figueiredo. [sade] a maple package for the symmetry analysis of differential equations. *Computer Physics Communications*, 182(2):467–476, 2011.
- [262] L Chris G Rogers and Zo Shi. The value of an asian option. *Journal of Applied Probability*, 32(4):1077–1088, 1995.
- [263] Richard Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of finance*, 39(4):1127–1139, 1984.
- [264] PG Rooney. On the η and h η transformations. *Canad. J. Math*, 32:1021–1044, 1980.
- [265] Robert E Schapire. The strength of weak learnability. *Machine learning*, 5(2):197–227, 1990.
- [266] Robert E Schapire and Yoav Freund. Boosting: Foundations and algorithms. *Kybernetes*, 2013.
- [267] Michael Schober, David K Duvenaud, and Philipp Hennig. Probabilistic ode solvers with runge-kutta means. In *Advances in neural information processing systems*, pages 739–747, 2014.
- [268] M Schröder. On the valuation of arithmetic-average asian options: integral representations, oktober 1997, revised november 1999.
- [269] Michael Schröder. On the valuation of arithmetic-average asian options: Laguerre series and theta integrals. *arXiv preprint math/0012072*, 2000.
- [270] Michael Schröder. Analytical ramifications of derivatives valuation: Asian options and special functions. *arXiv preprint math/0202298*, 2002.

- [271] Michael Schröder. Laguerre series in contingent claim valuation, with applications to asian options. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 15(3):491–531, 2005.
- [272] Michael Schröder. On ladder height densities and laguerre series in the study of stochastic functionals. i. basic methods and results. *Advances in applied probability*, 38(4):969–994, 2006.
- [273] Michael Schröder. On ladder height densities and laguerre series in the study of stochastic functionals. ii. exponential functionals of brownian motion and asian option values. *Advances in applied probability*, 38(4):995–1027, 2006.
- [274] Michael Schröder. On constructive complex analysis in finance: explicit formulas for asian options. *Quarterly of applied mathematics*, 66(4):633–658, 2008.
- [275] Michael Schröder. On arithmetic-average asian power options: closed forms and explicit methods for valuation. *Quarterly Journal of Mechanics and Applied Mathematics*, 66(1):1–27, 2013.
- [276] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.
- [277] Lue Shen, Chu Song Leheng Chen, and Jing Qian. Pricing asian options on com-modities with garch model. 2019.
- [278] Akiyoshi Shioura and Takeshi Tokuyama. Efficiently pricing european? asian options? ultimate implementation and analysis of the amo algorithm. *Information processing letters*, 100(6):213–219, 2006.
- [279] Steven E Shreve. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer Science & Business Media, 2004.
- [280] Gheorghe Silberberg. Discrete symmetries of the black-scholes equation. In *Proceedings of 10th International Conference in Modern Group Analysis*, volume 190, page 197, 2005.
- [281] JP Singh and S Prabakaran. Group properties of the black scholes equation & its solutions. *Electronic Journal of Theoretical Physics*, 5(18), 2008.
- [282] W Sinkala, PGL Leach, and JG O'hara. Invariance properties of a general bond-pricing equation. *Journal of Differential Equations*, 244(11):2820–2835, 2008.
- [283] W Sinkala, PGL Leach, and JG O'hara. Zero-coupon bond prices in the vasiček and cir models: Their computation as group-invariant solutions. *Mathematical methods in the Applied Sciences*, 31(6):665–678, 2008.
- [284] Winter Sinkala. On the generation of arbitrage-free stock price models using lie symmetry analysis. *Computers & Mathematics with Applications*, 72(5):1386–1393, 2016.
- [285] Winter Sinkala, PGL Leach, and John G O'Hara. An optimal system and group-invariant solutions of the cox-ingersoll-ross pricing equation. *Applied mathematics and computation*, 201(1-2):95–107, 2008.
- [286] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.
- [287] Christodoulos Sophocleous and Peter GL Leach. Algebraic aspects of evolution partial differential equations arising in financial mathematics. *Applied Mathematics & Information Sciences*, 4(3):289–305, 2010.
- [288] Christodoulos Sophocleous, Peter GL Leach, and Konstantinos Andriopoulos. Algebraic properties of evolution partial differential equations modelling prices of commodities. *Mathematical methods in the applied sciences*, 31(6):679–694, 2008.
- [289] Christodoulos Sophocleous, John G O'Hara, and Peter GL Leach. Symmetry analysis of a model of stochastic volatility with time-dependent parameters. *Journal of computational and applied mathematics*, 235(14):4158–4164, 2011.
- [290] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [291] Ian Sutherland, Yesuk Jung, and Gunhee Lee. Statistical arbitrage on the koshi 200: An exploratory analysis of classification and prediction machine learning algorithms for day trading. *Journal of Economics and International Business Management*, 6(1):10–19, 2018.
- [292] Radoslaw Szymtkowski and Sebastian Bielski. An orthogonality relation for the whitaker functions of the second kind of imaginary order. *Integral transforms and special functions*, 21(10):739–744, 2010.
- [293] KM Tamizhmani, K Krishnakumar, and PGL Leach. Algebraic resolution of equations of the black-scholes type with arbitrary time-dependent parameters. *Applied Mathematics and Computation*, 247:115–124, 2014.
- [294] Désiré Yannick Tangman, Ashvin Gopaul, and Muddun Bhuruth. Exponential time differencing with runge-kutta time stepping for convectively dominated financial problems. *University of Mauritius Research Journal*, 15(1):373–386, 2009.
- [295] DY Tangman, AAI Peer, N Rambeerich, and M Bhuruth. Fast simplified approaches to asian option pricing. *The Journal of Computational Finance*, 14(4):3, 2011.
- [296] Stephen Taylor and Scott Glasgow. A novel reduction of the simple asian option and lie-group invariant solutions. *International Journal of Theoretical and Applied Finance*, 12(08):1197–1212, 2009.
- [297] Chinthananda Tellambura and Damith Senaratne. Accurate computation of the mgf of the lognormal distribution and its application to sum of lognormals. *IEEE Transactions on Communications*, 58(5), 2010.
- [298] Nico M Temme. *Special functions: An introduction to the classical functions of mathematical physics*. John Wiley & Sons, 2011.
- [299] GWP Thompson. *Fast narrow bounds on the value of Asian options*. Citeseer.
- [300] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
- [301] Allan Timmermann. Forecast combinations. *Handbook of Economic Forecasting*, 1:135–196, 2006.
- [302] Allan Timmermann. Elusive return predictability. *International Journal of Forecasting*, 24(1):1–18, 2008.
- [303] Stuart M Turnbull and Lee Macdonald Wakeman. A quick algorithm for pricing european average options. *Journal of financial and quantitative analysis*, 26(3):377–389, 1991.
- [304] Felisa J Vázquez-Abad and Daniel Dufresne. Accelerated simulation for pricing asian options. In *1998 Winter Simulation Conference. Proceedings (Cat. No. 98CH36274)*, volume 2, pages 1493–1500. IEEE, 1998.
- [305] Jan Vecer. Asian options on the harmonic average. *Quantitative Finance*, 14(8):1315–1322, 2014.
- [306] Jian Wang, Jungyup Ban, Seongjin Lee, and Changwoo Yoo. Comparative study of numerical algorithms for the arithmetic asian option. *Journal of the Korean Society for Industrial and Applied Mathematics*, 22(1):75–89, 2018.
- [307] Kehluh Wang and Ming-Feng Hsu. Numerical valuation of asian options with higher moments in the underlying distribution. In *Handbook of Quantitative Finance and Risk Management*, pages 587–603. Springer, 2010.
- [308] Sijian Wang, Bin Nan, Saharon Rosset, and Ji Zhu. Random lasso. *The Annals of Applied Statistics*, 5(1):468, 2011.
- [309] James Wei and Edward Norman. Lie algebraic solution of linear differential equations. *Journal of Mathematical Physics*, 4(4):575–581, 1963.
- [310] Alois Weigand. Machine learning in empirical asset pricing. *Financial Markets and Portfolio Management*, 33(1):93–104, 2019.
- [311] Ivo Welch and Amit Goyal. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508, 2008.
- [312] Rafal Weron. Market price of risk implied by asian-style electricity options and futures. *Energy Economics*, 30(3):1098–1115, 2008.
- [313] Sander Willems. Asian option pricing with orthogonal polynomials. *Quantitative Finance*, 19(4):605–618, 2019.
- [314] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.
- [315] P Wilmott, J Dewynne, and S Howison. Option pricing: Mathematical models and computation. 1993.
- [316] Yang Xuan-Liu, Zhang Shun-Li, and Qu Chang-Zheng. Symmetry breaking for black-scholes equations. *Communications in Theoretical Physics*, 47(6):995, 2007.
- [317] Jingming Xue, SiHang Zhou, Qiang Liu, Xinwang Liu, and Jianping Yin. Financial time series prediction using l2, l1f-elm. *Neurocomputing*, 277:176–186, 2018.
- [318] Semen B Yakubovich. *Index transforms*. World Scientific, 1996.
- [319] Tingting Ye and Liangliang Zhang. Derivatives pricing via machine learning. Available at SSRN 3352688, 2019.
- [320] Marc Yor. The distribution of brownian quantiles. *Journal of applied probability*, 32(2):405–416, 1995.
- [321] Hyungbin Yun, Minhyeok Lee, Yeong Seon Kang, and Junhee Seok. Portfolio management via two-stage deep learning with a joint cost. *Expert Systems with Applications*, 143:113041, 2020.
- [322] Matthew D Zeiler. Adadelta: An adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.
- [323] Armen H Zemanian. *Distribution theory and transform analysis: an introduction to generalized functions, with applications*. Courier Corporation, 1987.
- [324] Boxiang Zhang, Yang Yu, and Weiguang Wang. Numerical algorithm for delta of asian option. *The Scientific World Journal*, 2015, 2015.
- [325] Chaojun Zhang and Xiaoqun Wang. Quasi-monte carlo-based conditional pathwise method for option greeks. *Quantitative Finance*, 20(1):49–67, 2020.
- [326] Peng Zhang. Path integral and asian options. *arXiv preprint arXiv:1008.4841*, 2010.
- [327] Peter G Zhang. *Exotic options: a guide to second generation options*. World Scientific, 1998.
- [328] Suhua Zhang and Yongzeng Lai. Efficient multiple control variate method with applications to exotic option pricing. *Communications in Statistics-Theory and Methods*, pages 1–18, 2019.
- [329] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.
- [330] R Zvan, PA Forsyth, and KR Vetzal. A general finite element approach for pde option pricing models. *methods*, 19:3, 1998.
- [331] Robert Zvan, Peter A Forsyth, and Kenneth Roy Vetzal. *Robust numerical methods for PDE models of Asian options*. PhD thesis, Citeseer, 1996.
- [332] Robert Zvan, Peter A Forsyth, and KR Vetzal. A finite volume approach for contingent claims valuation. *IMA Journal of Numerical Analysis*, 21(3):703–731, 2001.

Appendix A

Exponential operator formulas

This appendix establishes all results claimed to follow ‘trivially’ from Taylor’s theorem in Proposition 14 of Parmar [244]. In fact, it is seen that the proofs by Taylor’s theorem are rather nontrivial (those involving the Laplacian, ∇^2 , are not attempted in this manner), but a formal operator approach yields them readily, including a vast generalization of the first three.

By Taylor’s theorem, given (real analytic) $f \in C^\omega(\mathbb{R}, \mathbb{R})^1$ and $x, t \in \mathbb{R}$,

1. If $|t| < R$, the radius of convergence of the Taylor series expansion of $f(x) \in C^\omega(\mathbb{R}, \mathbb{R})$,

$$f(x+t) = \left[\sum_{n=0}^{\infty} \frac{t^n}{n!} \partial_x^n \right] f(x) = e^{t\partial_x} f(x) \quad (\text{A.1})$$

This remains valid, replacing t by functions $f(t) : \mathbb{R} \rightarrow \mathbb{R}$, e.g., $c_0(t) \equiv -\int_0^t [\sigma^2(s)/2 + r(s)] ds$, as in (the proof of) Proposition 2 of Parmar [244]; *as long as* $|f(t)| < R$. Given that the radii of convergence of Taylor series expansions of solutions of pricing PDE (1.7), e.g., as simplified to the reducing pricing PDEs (1.30) or (1.51), are not known in advance, if they exist at all, *this condition cannot be taken for granted*. Having done so may partly explain why the otherwise apparently correct proof of Proposition 2 of Parmar [244] yields a solution which careful scrutiny revealed to be invalid, i.e., not to satisfy pricing PDE (1.7).

2. Note that ($\forall 0 \leq \ell \in \mathbb{Z}$) $(x\partial_x)^0 x^\ell \equiv 1x^\ell \equiv x^\ell$, and if $(x\partial_x)^k x^\ell = \ell^k x^\ell$ for some $0 \leq k, \ell \in \mathbb{Z}$, then $(x\partial_x)^{k+1} x^\ell = x\partial_x \ell^k x^\ell = \ell^{k+1} x^\ell$. Denoting by $f(x) \equiv \sum_{k=0}^{\infty} a_k x^k$ the convergent power series expansion at $x \in \mathbb{R}$ of $f \in C^\omega(\mathbb{R}, \mathbb{R})$,

$$e^{tx\partial_x} f(x) = \sum_{n,k=0}^{\infty} \frac{t^n (x\partial_x)^n}{n!} a_k x^k = \sum_{n,k=0}^{\infty} \frac{(tk)^n}{n!} a_k x^k = \sum_{k=0}^{\infty} a_k (e^t x)^k = f(e^t x) \quad (\text{A.2})$$

Only this result is used by Lo, Yuen and Hui [207] and [203] and Lo and Hui [206]; the entirety of the exponential function implies that the validity of formula (A.2) requires

¹The proofs require $f \in C^\omega(\mathbb{R}, \mathbb{R})$ to have convergent Taylor series expansions for all $x \in \mathbb{R}$, i.e., to be real analytic, so that simply *smooth* $f \in C^\infty(\mathbb{R}, \mathbb{R})$ will not do. [Indeed, result 1 is equivalent to $f \in C^\omega(\mathbb{R}, \mathbb{R})$.] E.g., the differentiability class of *bump* functions, $C^\infty(\mathbb{R}, \mathbb{R}) \supset C_0^\infty(\mathbb{R}, \mathbb{R}) \not\subset C^\omega(\mathbb{R}, \mathbb{R})$.

only that $f \in C^\omega(\mathbb{R}, \mathbb{R})$, in contrast to the preceding and subsequent results, in which concerns regarding the radius of convergence of the Taylor series expansion of f arise.

3. Note that $(\forall 0 \leq \ell \in \mathbb{Z}) (x^2 \partial_x)^0 x^\ell \equiv 1 x^\ell \equiv (\ell)_0 x^{\ell+0}$,² and if $(x^2 \partial_x)^k x^\ell = (\ell)_k x^{\ell+k}$ for some $0 \leq k, \ell \in \mathbb{Z}$, then $(x^2 \partial_x)^{k+1} x^\ell = x^2 \partial_x (\ell)_k x^{\ell+k} = (\ell)_{k+1} x^{\ell+k+1}$. Then if $|tx| < 1$,

$$e^{tx^2 \partial_x} f(x) = \sum_{n,k=0}^{\infty} \frac{t^n (x^2 \partial_x)^n}{n!} a_k x^k = \sum_{n,k=0}^{\infty} \frac{(k)_n (tx)^n}{n!} a_k x^k = \sum_{k=0}^{\infty} a_k \left(\frac{x}{1-tx} \right)^k = f \left(\frac{x}{1-tx} \right)$$

(This follows since the hypergeometric functions $(\forall b \in \mathbb{C}) F(k, b; b; tx) = (1-tx)^{-k}$; see section 5.1. of Temme [298].)

More generally, denote $g_\ell(x, t) \equiv e^{tx^\ell \partial_x} f(x)$ and note that $\partial_t g_\ell(x, t) = x^\ell \partial_x g_\ell(x, t)$ and $g_\ell(x, 0) = f(x)$, yielding a characteristic, $t + c_\ell(x)$, with

$$c_\ell(x) \equiv \begin{cases} \ln x & 1 = \ell \\ \frac{x^{1-\ell}}{1-\ell} & 0 \leq 1 \neq \ell \in \mathbb{N} \end{cases}$$

Thus, given the general solutions, $g_\ell(x, t) \equiv G_\ell[t + c_\ell(x)]$, it follows that $G_\ell[c_\ell(x)] = g_\ell(x, 0) = f(x)$, and thence $G_\ell(x) \equiv f[c_\ell^{-1}(x)]$, so that $g_\ell(x, t) = f(c_\ell^{-1}[t + c_\ell(x)])$. This clearly reduces to results 1-3 in the respective cases of $\ell \in \{0, 1, 2\}$, and in general,

$$(\forall 0 \leq 1 \neq \ell \in \mathbb{Z}) \quad e^{tx^\ell \partial_x} f(x) = f \left(\left[(1-\ell)t + x^{1-\ell} \right]^{\frac{1}{1-\ell}} \right) \quad (\text{A.3})$$

This formal operator approach is thus vastly more expedient than the initial proofs via induction and power series. It also readily yields the modified argument of $f \in C^\omega(\mathbb{R}, \mathbb{R})$, which is not apparent upon examining terms of the sequences $\{(\forall 1 \leq k-1, \ell \in \mathbb{N}) (x^\ell \partial_x)^k\}$:

$$\begin{aligned} a_{\ell,2} &\equiv (x^\ell \partial_x)^2 = \ell & x^{2\ell-1} \partial_x + & x^{2\ell} \partial_x^2 \\ a_{\ell,3} &\equiv (x^\ell \partial_x)^3 = \ell(2\ell-1) & x^{3\ell-2} \partial_x + 3\ell & x^{3\ell-1} \partial_x^2 + x^{3\ell} \partial_x^3 \\ a_{\ell,4} &\equiv (x^\ell \partial_x)^4 = \ell(2\ell-1)(3\ell-2) & x^{4\ell-3} \partial_x + \ell(11\ell-4) & x^{4\ell-2} \partial_x^2 + 6\ell x^{4\ell-1} \partial_x^3 + x^{4\ell} \partial_x^4 \end{aligned}$$

And though the coefficients of $a_{\ell,k+1}$ are linearly determined by those of $a_{\ell,k}$, the addition of a new (nonzero) coefficient with each sequence element yields an unwieldy and opaque recurrence relation which does not well lend itself to the general formula (A.3), nor in nearly so straightforward a fashion as the foregoing operator approach, which is also used for the remaining result.

4. Given $1 \leq n \in \mathbb{N}$, $x \in \mathbb{R}^n$, and $f \in C^\infty(\mathbb{R}^n, \mathbb{R})$, denote $g(x, t) \equiv e^{t\nabla^2} f(x)$ and note that $\partial_t g(x, t) = \nabla^2 g(x, t)$, i.e., the (n -dimensional case of) the heat equation (1.35) with unit thermal conductivity. Furthermore, the initial condition, $g(x, 0) = f(x)$, yields the

²Here, the Pochhammer symbol, $(\ell)_k \equiv \prod_{n=0}^{k-1} \ell + n$, is also known as an ascending factorial.

(n -dimensional case of) the Cauchy problem addressed in section 1.2.1, with general solution given by (the n -dimensional case of) Poisson's formula (1.38):

$$g(x, t) = (4\pi t)^{-\frac{n}{2}} \int_{\mathbb{R}^n} g(\chi, 0) e^{-\frac{(x-\chi)^T(x-\chi)}{4t}} d\chi = (4\pi t)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f(\chi) e^{-\frac{(x-\chi)^T(x-\chi)}{4t}} d\chi \quad (\text{A.4})$$

This remains valid, replacing t by functions $f(t) : \mathbb{R} \rightarrow \mathbb{R}$ satisfying $f(0) = 0$, e.g., $c_+(t) \equiv \int_0^t \sigma^2(s) ds$, as in (the proof of) Proposition 2 of Parmar [244]. But as in the discussion following Taylor series (A.1), the validity of this, along with analogous application of the formulas (A.3) in case of $\ell \in \{0, 1, 2\}$, as in (the proof of) Proposition 12 of Parmar [244], may be subject to unknown, stringent conditions on the variables and parameters of solutions of pricing equation (1.7), if such conditions exist at all. This point is emphasized by the great ease of deriving formulas (A.3) and (A.4) via formal operator theoretic arguments, as opposed to the more rigorous [and subject to the stronger condition $f \in C^\omega(\mathbb{R}^n, \mathbb{R}) \subseteq C^\infty(\mathbb{R}^n, \mathbb{R})$] approach via Taylor's formula (A.1), used to derive results 1-3, and seen to be much more intricate and less intuitive. Thus, these additional oversights may also partly explain [in the case of formula (A.4)] the failure of the solution 'derived' in Proposition 2 of Parmar [244] to satisfy pricing PDE (1.7).

Appendix B

Data

Further details are provided on the trade and quote (TAQ) data cleaning procedures and calculation of the additional characteristics used in the empirical analysis is described.

B.1 Trade and quotes cleaning

The TAQ data require substantial cleaning due to contamination from market microstructure noise. Used are the Monthly TAQ Second database up to 2003 and the Daily TAQ Millisecond/Microsecond database from 2004-2016, with noisy observations filtered following procedures similar to Barndorff-Nielsen et al. [27]. Returns are primarily calculated using prices from the trades database. Bid and ask observations from the quotes database are also used for cleaning trades and calculating transaction costs. For both trades and quotes, entries with time stamps outside of the trading day (9:30 AM to 4:00 PM) are removed. Entries with a zero or negative bid, ask, or price are also removed. For each stock, only entries from the exchange with the highest volume in each month are kept.

For only the quotes database, the national-best bid-ask (NBBO) is constructed following procedures from Keintz (<https://support.sas.com/resources/papers/proceedings16/11201-2016.pdf>) using quotes from all exchanges at a second interval. All entries with a negative spread (ask minus bid) and quotes larger than 50 times the median spread on each day, are removed.

For only the trades data, entries with corrected trades ($CORR \neq 0$) are deleted. Entries with an abnormal sale condition are removed, only keeping entries with COND equal to E, F, @, *, @E, @F, *E and *F. Multiple trades with the same second timestamp are replaced with an entry with the median price. Finally, quotes are used to discipline the trade prices: Entries are removed if their price is above the ask plus spread or below the bid minus spread.

The cleaned second-by-second price, bid and ask data is then aggregated into five-minute intervals and merged with the daily center for research in security prices (CRSP) data file by the CUSIP key. Merging the TAQ and CRSP data is challenging, because their tickers often differ. Additionally, tickers change in time due to mergers, acquisitions, and other corporate events. Instead, the TAQ and CRSP databases are merged using CUSIPs. The CUSIP identifier of each

stock is obtained by merging trades with the TAQ master files. Finally, each stock is indexed by CRSP PERMINOs, which are unique and do not change in time.

B.2 Additional firm characteristics

In addition to returns with one lag, other lagged characteristics of the SPY and S&P 500 constituents are considered. These characteristics proxy short-term changes in liquidity and price trends. High-frequency analogs are created of characteristics from the asset pricing literature that have been shown to predict returns, including firm-level market beta, momentum, illiquidity, maximum, minimum, and trading volume. Also considered are higher order moments, including volatility, skewness, and kurtosis. These characteristics are all calculated over preceding days. The particular variables used are the following:

Market beta: The market beta is motivated by the capital asset pricing model of Sharpe [276] and Lintner [201]. At time $1 \leq t \leq T$, given (SPY market) return of stock $1 \leq \ell \leq 500$, r_t^ℓ (r_t^m), it is based on a single factor model given by

$$r_t^\ell = \alpha^\ell + \beta^\ell(r_t^m) + \epsilon_t^\ell$$

This is computed for each five-minute interval using a rolling window of preceding days, yielding one overnight and 78 intraday observations. BETA of stock ℓ during each five-minute interval, t , is the least-squares estimate, $\hat{\beta}^\ell$.

Momentum: Motivated by Jegadeesh and Titman [167], the MOM of stock ℓ during each five-minute interval, t , is the cumulative return for the preceding day, yielding one overnight and 78 intraday observations; that is, $\sum_{i=1}^T r_i^\ell$ with $T = 79$.

Illiquidity: Motivated by Amihud [12], the ILLIQ of stock ℓ , during each five-minute interval, t , is the ratio of the absolute stock return, $|r_t^\ell|$, to the dollar trading volume, averaged in a day excluding the overnight volume, i.e., over the $T = 78$ previous trading intervals in the day. Given the corresponding five-minute trading volume (price times number of shares traded) in dollars at time t , VOLD_t^ℓ , this yields

$$\text{ILLIQ}^\ell \equiv \frac{1}{T} \sum_{t=1}^T \frac{|r_t^\ell|}{\text{VOLD}_t^\ell}$$

Maximum (minimum) return: Motivated by Bali, Cakici and Whitelaw [24], the MAX (MIN) of stock ℓ during each five-minute interval, t , is defined as the maximum (minimum) five-minute return within the preceding day (i.e., the previous 79 observations), that is, $\max_{1 \leq t \leq T} r_t^\ell$ ($\min_{1 \leq t \leq T} r_t^\ell$).

Trading volume: Motivated by Chordia, Subrahmanyam and Anshuman [80], the VOLUME of stock ℓ during each five-minute interval, t , is defined as the average number of shares traded

within the preceding day excluding the overnight volume (i.e., the previous 78 observations):

$$\text{VOLUME}^\ell \equiv \frac{1}{T} \sum_{t=1}^T \text{SharesTraded}_t^\ell$$

Volatility: Motivated by Ang et al. [14], the VOL of stock ℓ during each five-minute interval, t , is the standard deviation of five-minute returns within the preceding day (i.e., the previous 79 observations):

$$\text{VOL}^\ell \equiv \sqrt{\frac{1}{T} \sum_{t=1}^T (r_t^\ell - \bar{r}^\ell)^2} \equiv \text{sd}(r_t^\ell)$$

Here, $\bar{r}^\ell \equiv \frac{1}{T} \sum_{t=1}^T r_t^\ell$ denotes the corresponding arithmetic mean.

Skewness (kurtosis): Motivated by Amaya et al. [11], the SKEW (KURT) of stock ℓ during each five-minute interval, t , is the skewness (kurtosis) of five-minute returns within the preceding day (i.e., the previous 79 observations):

$$\text{SKEW}^\ell \equiv \frac{1}{T} \sum_{t=1}^T \frac{(r_t^\ell - \bar{r}^\ell)^3}{\text{sd}(r_t^\ell)^3} \quad \left(\text{KURT}^\ell \equiv \frac{1}{T} \sum_{t=1}^T \frac{(r_t^\ell - \bar{r}^\ell)^4}{\text{sd}(r_t^\ell)^4} \right)$$

Appendix C

Hyperparameters

Discussed in greater detail are the most relevant hyperparameters tuned for each of the considered models. First, applicable to all models, inherent in regularization is the particular binary problem of *whether* to regularize or not. Generally, the hyperparameter optimization problem is that of tuning the regularization parameters, λ in the case of lasso (LAS) regularization, with the binary problem of whether this parameter is zero or not. In elastic net (EN) regularization, the convex combination coefficient $\alpha \in [0, 1]$ is another hyperparameter to be optimized, with the endpoints $\{0, 1\}$ respectively corresponding to the discrete choices of only ridge or only lasso regularization, respectively. In principal component regression (PCR) the number of principal components, $1 \leq \kappa \leq K$, can be considered a hyperparameter to be optimized.¹

There are two crucial hyperparameters for random forest (RF) models: (1) the number of decision trees (or base learners) in the forest, and (2) the maximum depth of each of these trees, i.e., the maximum path length from any tree root to any of its leaves. More trees in the forest yield greater prediction variance reduction at increased computational cost. Restricting maximum tree depth limits the complexity, both saving computational time and avoiding over-fitting. Similar goals may be achieved less directly, e.g., by limiting the number of training samples per leaf, which generally decreases with increased tree size. The same hyperparameters are important for gradient-boosted regression tree (GBRT) models, in addition to the explicit *learning rate*, ν , which may be factored out from the step sizes or weights γ_i . A smaller learning rate typically leads to better testing performance, but requires more training steps/additional decision trees in the ensemble, to maintain a given training error. Though these are the principal hyperparameters to consider for RF and GBRT models, several other, generally less important ones do exist. For example, bounding above the number of predictors determining node splits in base learners is relatively important for RF models in particular. Greater bounds permit more complexity, resulting in bias reduction at the expense of increased prediction variance and computational cost. Note that base learner complexity may be directly limited by bounding above the number of node splits or the total number of tree nodes, i.e., the tree size, itself.

For artificial neural networks (ANN), in particular, the loss function used for training is important. Standard choices include (the implemented) mean squared error (MSE), as well as

¹Other heuristics for determining κ are to accept the smallest value which achieves some level of total explained variance or to increase κ so long as the increase in variance explained exceeds a given threshold.

mean absolute error, logarithmic hyperbolic cosine, and a variety of others adapted to more specific scenarios.² Model overfitting is guarded against by adding ℓ^1 and ℓ^2 regularization to the loss function. Other important factors involved in training include the number of iterations, or epochs, and the level of *dropout*, the percentage of training data discarded in each epoch to avoid over-fitting and regularize the optimization problem (Srivastava et al. [290]).

While the optimization algorithm employed for training, the loss function it uses, the number of epochs, and the level of dropout, all constitute important *hyperparameters* impacting the quality of predictions from ANN models with trained *parameters*, the actual network architecture, i.e. the number of hidden layers, the number of neurons in each hidden layer, and the particular activation function associated with each neuron, may often have an even greater impact on the results. The activation functions used are the rectified linear unit (ReLU), $\phi(x) = \max\{0, x\}$ (Nair and Hinton [228]); the maxout, $\phi(x_1, x_2) = \max\{x_1, x_2\}$ (Goodfellow et al. [150]); and the hyperbolic tangent, $\phi(x) = \tanh x$. A hyperparameter grid is used, fixing three hidden layers. Models are estimated using stochastic gradient descent (SGD) optimization with the ADADELTA (Zeiler [322]) adaptive learning rate method for faster estimation. Dropout and early stopping are used as regularization to prevent overfitting.

Table C.1 summarizes the set of hyperparameters that are tuned for each of the models considered. To tune these at a particular time, the corresponding data is partitioned into training, testing, and validation sets and the optimization problem (2.5) is repeatedly solved using training predictors and targets, validating at each iteration to determine whether to terminate training and minimize over-fitting. All training decisions aim to minimize the loss metric, $m[\cdot]$, modifying internal parameters, such as β for linear models, according to some optimization algorithm, e.g., classical SGD in the case of ANN models. Validation MSE is considered when tuning the hyperparameters. For LAS and EN, λ is selected using coordinate descent (Friedman, Hastie and Tibshirani [125]). For PCR grid search is used. For all other algorithms, random search is used with 50 models, which is more efficient than grid search (Bergstra and Bengio [31]). Contrary to Gu, Kelly and Xiu [152], ensembles of neural networks are not used due to computational limitations.

²Of particular interest is the Huber loss function, a combination of the ℓ^1 and ℓ^2 norms that permits control of the sensitivity to outliers, used in Gu, Kelly and Xiu [152]. Preliminary results show that this metric may outperform the MSE metric used for both the ANN and the GBRT models.

Table C.1: Model hyperparameters

LAS	λ
EN	λ , $\alpha = \{0, 0.25, 0.5, 0.75, 1\}$
PCR	$\kappa = 1-99$ by 2
RF	#trees = 100-700 by 100, tree depth = 1-40 by 5, #minimum rows = {1, 25, 50}, columns randomly selected = floor(#number of features * {.05, .15, .25, .333, .4})
GBRT	#trees = 100-700 by 100, tree depth = 2-40 by 3, learning rate = 0.01 - 0.1 by 0.03, sample rate = 0.5-1 by 0.2, column sample rate = 0.1 - 1 by 0.2
ANN	Activation Functions = Relu, Maxout, Tanh; with and without dropout, Hidden nodes = {1000,500,10}, {100,50,10}, {16,14,12},{20,15,10},{25,17,10},{15,10,5}, epochs = {50,100}, dropout ratio = {0, 0.1, 0.2}, max $w^2 = \{10, 100, 1000, 3.4028235e+38\}$, $\ell^1 = \{0, 0.00001, 0.0001\}$, $\ell^2 = \{0, 0.00001, 0.0001\}$, rho (rate time decay) = {0.9, 0.95, 0.99, 0.999}, epsilon (rate time smoothing) = {1e-10, 1e-8, 1e-6, 1e-4}

Sets of tuned hyperparameters for each model considered.

Curriculum Vitae

Name:	Dillon Huddleston	
Post-Secondary Education and Degrees:	University of Calgary Calgary, AB 2008 - 2013 B.Sc.	
	University of Western Ontario London, ON 2014 - 2015 M.A.	
	University of Western Ontario London, ON 2015 - 2021 Ph.D.	
Honours and Awards:	Mitacs Research Training Award	2020
	Ontario Graduate Scholarship (Recruitment in 2015)	2015-2016, 2018
	Gervan Fearon PhD Award in Economics	2016
	Department of Mathematics and Statistics Silver Medallion	2014
	Archibald Wayne Dingman Memorial Scholarship	2013
	Louise McKinney Scholarship	2009, 2013
Related Work Experience:	Quantitative cryptoasset model development at Lightning Capital	2021
	Ongoing technical consultation through PeoplePerHour	2020-2021
	Research Assistant for Greg Pavlov and Maria Goltsman	2020-2021
	Research Assistant for Ozden Sungur	2018
	Referee for <i>Insurance and Risk Management Review</i>	2018
	Tutor at OneClass, Chegg Inc., TutorOcean	2015-2021
	ValNav Product Developer at Energy Navigator Inc.	2014
	Engineering Intern in Gas Turbine Development at Alstom Power	2011-2012
Publications:	Intraday Market Predictability: A Machine Learning Approach (with F. Liu & L. Stentoft) Journal of Financial Econometrics, Invited by the Chief Editors.	2021