Essays on Rational Portfolio Theory

Ph.D. Thesis

Simon Ellersgaard Nielsen
MPhysPhil MASt MSc

2016

The Department of Mathematical Sciences
The University of Copenhagen

This dissertation has been submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at the Ph.D. School of The Faculty of Science, University of Copenhagen.

Author:	Simon Ellersgaard Nielsen, The Department of Mathematical Sciences University of Copenhagen Universitetsparken 5, 2100 Copenhagen Ø, Denmark e-mail: ellersgaard@math.ku.dk e-mail: s_ellersgaard@yahoo.com
Supervisor:	Rolf Poulsen, University of Copenhagen

Submission date:	January 4th, 2016.
Assessment Committee:	Mogens Steffensen, University of Copenhagen Nicole Branger, University of Muenster Claus Munk, Copenhagen Business School
ISBN:	978-87-7078-939-4

If you're anxious for to shine in the high aesthetic line,
as a man of culture rare,
You must get up all the germs of the transcendental terms,
and plant them everywhere.
You must lie upon the daisies and discourse in novel phrases
of your complicated state of mind,
The meaning doesn't matter if it's only idle chatter
of a transcendental kind.
And everyone will say, as you walk your mystic way,
"If this young man expresses himself in terms too deep for me,
Why, what a very singularly deep young man
this deep young man must be!"

W. S. Gilbert - *The Aesthete*

Optimisation consists in finding the mathematically optimal policy that an economic agent could pursue. For instance, what is the "optimal" quantity you should allocate to stocks? It involves complicated mathematics and thus raises the barrier to entry by non-mathematically trained scholars. I would not be the first to say that this optimisation set back the social science by reducing it from the intellectual and reflective discipline it was becoming to an attempt at an "exact science". By "exact science", I mean a second-rate engineering problem for those who want to pretend that they are in a physics department - so-called physics envy. In other words, an intellectual fraud.

N.N. Taleb - *The Black Swan*

Foreword

1 From Model Risk to Optimal Asset Allocation

1.1 The Problem of Measure $\mathbb{P}$

Inherent to the stochastic modelling of financial markets is the (often tacit) specification of a *measure space* $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω captures the set of future outcomes, $\mathscr{F}$ is a σ-algebra codifying the events we would like to consider, and $\mathbb{P} : \mathscr{F} \mapsto [0,1]$ is a probability measure which assigns weights thereto in accordance with certain axioms in the manner of Kolmogorov. Besides the glaringly obvious problem of defining the very notion of probability (leave that to the philosophers), there is the somewhat more tangible issue relating to the epistemic inaccessibility of $\mathbb{P}$. Specifically, while it is often assumed in the literature that everybody agrees on which events occur with which probabilities, it is quite clear (to the point of being a statistical banality) that this does not pertain to the real world. *Au contraire*, there is considerable so-called Knightian uncertainty surrounding which model (and therefore: which probability measure) should be employed in our financial modelling: knowledge of $\mathbb{P}$ is manifestly *a posteriori*, being a property of the world which we infer through the study of past realisations. Unfortunately, this data is but a shadow on the wall in Plato's cave, yielding only a rudimentary understanding of which governing dynamics is truly at play. For any given time series, an infinite number of congruent models prevail, each with their own historically induced probability measure $\{\mathbb{H}_1, \mathbb{H}_2, \ldots\}$. This kind of radical underdetermination is problematic since financial modelling manifestly is an abstract idealisation of hyper-complex facts of this world. For one reason or another, all social science models are fundamentally wrong[1], suggesting that the employ of such golden rules as Occam's razor scarcely will be meaningful in selecting a superior $\mathbb{H}$ (the intuition being that the right model, if anything, should be more complicated than what's posited).

The point here is forcefully communicated through an example which has its roots in Markowitz' modern portfolio theory. Consider the simple interest-free 1-period model

[1] Though some *may* be useful.

in which the rate of return of n risky assets is multivariate normally distributed, $\mathbf{R} \sim \mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{R}]$ is the expected return, and $\boldsymbol{\Sigma} = \mathbb{C}\text{ov}[\mathbf{R}]$ is the associated covariance matrix which we assume non-singular. Moreover, let $\boldsymbol{\pi}$ be a vector of portfolio weights placed on the n risky assets, such that the expected return is $\mu_\pi = \boldsymbol{\pi}^\intercal \boldsymbol{\mu}$ and the variance is $\sigma_\pi^2 = \boldsymbol{\pi}^\intercal \boldsymbol{\Sigma} \boldsymbol{\pi}$. Now, suppose we are interested in those portfolios which *for fixed returns yield the minimal possible variance*: solving the Lagrange multiplier problem [5] we find that this class of portfolios traces out a parabola (a *mean-variance frontier*) in (σ_π^2, μ_π)-space:

$$\sigma_\pi^2(\mu_\mu) = d^{-1}(a - 2b\mu_\pi + c\mu_\pi^2),$$

where $a \equiv \boldsymbol{\mu}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, $b \equiv \mathbf{1}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, $c \equiv \mathbf{1}^\intercal \boldsymbol{\Sigma}^{-1} \mathbf{1}$, and $d \equiv ac - b^2$. Here, the singularly most important portfolio attainable (the so-called *market portfolio*, *M*) is that which admits the highest possible return per unit volatility. Solving the optimisation problem, the associated *tangency portfolio* is given by

$$\boldsymbol{\pi}^* = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}},$$

from which we can compute $M = (\sigma_\pi^{*2}, \mu_\pi^*)$. Now the question we are dying to ask is "to which extent are these results robust to model risk" - here interpreted as uncertainty surrounding the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$? Specifically, if we are to perform sample estimates of the mean and the covariance of a model which truthfully has the distribution $\mathbf{R} \sim \mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ per unit time, how much do the estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ respectively perturb the picture above? Following Ellersgaard, Lando, and Poulsen [4] the answers turn out to be, in turn, "a lot" and "not much at all" - owing to the near telescoping nature of the drift-estimator. In particular, frontiers which utilise $\hat{\boldsymbol{\mu}}$ will tend to be wildly scattered with respect to the true frontier, while those frontiers which make use of $\hat{\boldsymbol{\Sigma}}$ only vary at a modest level. This highlights a profound issue at the heart of quantitative asset management: since drifts manifestly are the kind of quantities we care about, we must brace ourselves for considerable model risk in our modern portfolio theoretical recommendations. Remarkably, this simple observation rarely is disseminated in business schools!

1.2 The Problem of Measure $\mathbb{Q}$

Unsurprisingly, the uncertainty surrounding measure $\mathbb{P}$ extends to the risk-neutral valuation of derivative securities. Recall that a fundamental result of arbitrage-free markets is the existence of a measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, such that discounted asset prices are $\mathbb{Q}$-martingales. Specifically, the claim is that if Φ_T is the pay-off of a security at time T, then there exists a (not necessarily unique) $\mathbb{Q} \sim \mathbb{P}$ such that the time $t \leq T$ price of the security is

$$V_t = \mathbb{E}^{\mathbb{Q}}[B_t^T \Phi_T | \mathscr{F}_t], \tag{1}$$

where B_t^T is a (possibly stochastic) discount factor representing the time t price of a zero-coupon bond with terminal value equal to unity. Conceptually, $\mathbb{Q}$ (while formally satisfying the axioms of probability) has nothing to do with objective chances of events per se, but rather should be construed as the financial market's conception of fair option pricing. Specifically, we see that $\forall F \in \mathscr{F}$, $\mathbb{Q}(F)$ is equivalent to the fair value of an security with terminal pay-off $\Phi_T = 1/B_t^T$ if F occurs (zero otherwise).

Insofar as the market is *complete*, roughly understood as the case where the number of risky traded securities is at or above the number of random sources [1], $\mathbb{Q}$ is uniquely defined in terms of $\mathbb{P}$. Knightian uncertainty in $\mathbb{P}$ thus trickles through to $\mathbb{Q}$. Nevertheless, even if we were to imagine ourselves in the fortuitous (and I daresay: implausible) circumstances that $\mathbb{P}$ is known with certainty, uncertainty in $\mathbb{Q}$ can still prevail if the market is *incomplete*. Sadly, market incompleteness is an almost inexorable fact of life qua the abundant empirical support for random discontinuities in stock prices (jumps), and non-tradeable state variables such as stochastic volatility (for a thorough review of these issues, including a coherent measure for model risk, we refer the reader to Cont [5]).

Some insight into quantifying model risk may be provided through a meta-theorem we call *the fundamental theorem of derivative trading*. In elementary terms, we here concern ourselves with continuous time models marked by uncertainty in the volatility component

$$\mathbb{Q}_i: \qquad dS_t = rS_t dt + \sigma_{i,t} S_t dW_t^{\mathbb{Q}_i},$$

where r is the risk-free rate, $\sigma_i : [0,T] \mapsto (0,\infty)$, and W is a Brownian motion. Suppose the real volatility is the unknown quantity $\sigma_{t,r}$. Insofar as we Δ-hedge a derivative V written on S based on a $\mathbb{Q}_i$-assumption, how much do we stand to gain/lose over time on a mark-to-market basis? Indeed, what kind of P&L-evolution can we expect for various choices of the hedge volatility? Initial steps towards answering such questions are taken by El Karoui, Jeanblanc-Picque, and Shreve [3], who establish that the hedge-error, on an incremental basis, is of the form

$$d\text{P\&L} = \text{sgn}(\text{derivative position})\tfrac{1}{2}(\sigma_{t,r}^2 - \sigma_{t,i}^2)S_t^2 \partial_{ss}^2 V_t dt.$$

In the first paper of this dissertation, Ellersgaard, Jönsson, and Poulsen provide an abstract generalisation of their work, and consider the same questions from an empirical perspective.

1.3 The Problem of the Rational Investor

Finally, a few words on rationality and optimal asset allocation. Quantitative economics has long been under fire for operating with what many perceive to be a misguided obsession with rationality (Taleb, with his characteristic rumbustious demeanour, is one such fiery opponent - see the quotation above). This criticism, however, seems to be largely anchored in the curious belief that economics should be a purely *descriptive* discipline, with no room for such idealistic pursuits as *prescriptivism*: the art of telling people what they *ought* to do. What underlies this callous dismissal is at best nebulous: surely, there is considerable interest in establishing how one can secure the optimal realisation of one's wealth (say) even under idealised circumstances - yet, on a more benign reading, there is perhaps *some* merit to the critique. Specifically, one might argue that prescriptivists have not with sufficient emphasis underlined the limitations of their work, indeed have not integrated their prescriptions more closely with what people actually demand. For example, Markowitz' 1-period investment model described above suffers from some rather glaring limitations, including the disregard of: (i) consumption, (ii) dynamic portfolio updating based on new information, (iii) the interplay between investing for the sake of future consumption, (iv) non-normality of returns (see Munk [5]). Of course, since the heyday of Merton, these issues have been mitigated considerably, through the principles of continuous time stochastic control theory. Specifically, the type of problems we are nowadays interested are analogous to determining the optimal rate of consumption process $\{c_s^*\}_{s\in[0,T]}$, alongside the optimal portfolio weights $\{\boldsymbol{\pi}_s^*\}_{s\in[0,T]}$ such that

$$\{c_s^*, \boldsymbol{\pi}_s^*\}_{s\in[0,T]} = \underset{\{c_s,\boldsymbol{\pi}_s\}_{s\in[0,T]}\in\mathscr{A}}{\operatorname{argmax}} \; \mathbb{E}\left[\alpha \int_0^T e^{-\delta s} u(c_s)ds + (1-\alpha)e^{-\delta T}u(\mathscr{W}_T)\right], \quad (2)$$

where $\{\mathscr{W}_t\}_{t\in[0,T]}$ is the controlled stochastic wealth process of the investor, $\alpha \in [0,1]$ is a relative importance weight between continuous consumption and bequest, $\delta \in \mathbb{R}_+$ is a subjective discounting factor, and $u : \mathbb{R}^+ \mapsto \mathbb{R}$ is a von-Neumann Morgenstern utility function. Nonetheless, the operative word remains *mitigated*: despite the success story nested in the solutions to problems a la (2), there is still considerable room for improvement. For example, the notion that anyone cares about their *continuous* rate of consumption is blatantly false (rather we care about periodic withdrawals). Indeed, utility in itself is an area which demands further attention to square it with prevalent investor habits (Kahneman's prospect theory might be worthy of attention). It will be interesting to see what future research has to say on these matters.

In this dissertation we present four control studies into rational portfolio theory (in the classical prescriptivist's sense). Some of the issues we investigate include (A) analysing the method by which the non-linear Hamilton-Jacobi-Bellman equations should be solved numerically in connection with Merton type optimisation problems. A thorough review of the explicit and implicit methods is provided in one and more spatial dimensions. (B) Exposing the optimal investment ratios for a utility maximising investor who trades in bonds and stocks in a stochastic volatility environment. Various models are considered, including their effect in empirical trading experiments. (C) Extending the above analysis to include the derivatives markets. How much do the portfolio weights change? What is the effect of hedging stochastic volatility per se versus merely including a second asset? Is the bond-stock-derivative strategy truly superior when applied to real market data? (D) Hedging derivatives in a limit order book when one has the option of placing both limit and market orders. Assuming a certain tolerance towards deviating from a targeted hedge strategy, when should a rational investor place which type of order?

Overall, this thesis has been enjoyable to write. We hope the reader will enjoy reading it as well.

References

1. Björk, *Arbitrage Theory in Continuous Time*, Oxford University Press, 3rd edition.
2. Cont, *Model Uncertainty and its Impact on the Pricing of Derivative Instruments*, Mathematical Finance, Vol. 16, No. 3 (July 2006), pp. 519-547.
3. El Karoui, N: and Jeanblanc-Picque, M. and Shreve, S. *Robustness of the Black and Scholes Formula*, Mathematical Finance, 1998, Vol. 8, pp. 93-126.
4. Ellersgaard, Lando, and Poulsen, *Notes for Finance 1 (and More)*, 2015. Unpublished lecture notes.
5. Munk, *Dynamic Asset Allocation*. Unpublished lecture notes. December 18, 2013 edition http://mit.econ.au.dk/vip_htm/cmunk/noter/dynassal.pdf.

Acknowledgements

Several people have played a pivotal role in the creation of this dissertation and should rightfully be credited. First and foremost I wish to thank my primary supervisor, Rolf Poulsen, who despite my somewhat unconventional academic background took me under his academic wing. Throughout he has helped shape my thought process, whilst simultaneously giving me enough academic freedom to walk my own ways. Analogous accolades are extended to my "supervisor abroad", Mark H.A. Davis, who was kind enough to let me spend a year at my alma mater: Imperial College London. The masterly insights of these gentlemen into the field of financial mathematics are greatly appreciated.

Furthermore, I have on several occasions collaborated with a fellow doctoral candidate: Swede extraordinaire: Martin Jönsson. Martin's keen attention to technicals detail as well as his admirable flair for numerical programming and handling of large data sets have greatly enhanced the quality of my work. I thank him profusely and wish him all the best with his personal project.

Finally, during the course of my Ph.D. studies multiple people have engaged in valuable discussions about finance and the world by and large. In particular, I wish to thank Jannick Schreiner, Andre Ribeiro, Romano Trabalzini, Ivo Mihaylov, Pierre Blacque, Harry Zheng, Goran Peskir, Carl Shneider, Christian Jørgensen, Michael Olsen, and of course Nithya Sridharan.

For the printed edition several layout changes and typo corrections have been implemented. I wish to thank my thesis committee, Mogens Steffensen, Claus Munk, and Nicole Branger for their thorough and helpful preliminary assessment report.

London, February 2016 *Simon Ellersgaard Nielsen*

Abstract Summary

This dissertation is comprised of five research papers written during the period January 2013 - December 2015. Their abstracts are:

- **The Fundamental Theorem of Derivative Trading**. When estimated volatilities are not in perfect agreement with reality, delta hedged option portfolios will incur a non-zero profit-and-loss over time. There is, however, a surprisingly simple formula for the resulting hedge error, which has been known since the late 90s. We call this The Fundamental Theorem of Derivative Trading. This paper is a survey with twists of that result. We prove a more general version of it and discuss various extensions (including jumps) and applications (including deriving the Dupire-Gyöngy-Derman-Kani formula). We also consider its practical consequences both in simulation experiments and on empirical data thus demonstrating the benefits of hedging with implied volatility.

- **Numerical Stochastic Control Theory with Applications in Finance**. Analytic solutions to HJB equation in mathematical finance are relatively hard to come by, which stresses the need for numerical procedures. In this paper we provide a self-contained exposition of the finite-horizon Markov chain approximation method as championed by Kushner and Dupuis. Furthermore, we provide full details as to how well the algorithm fares when we deploy it in the context of Merton type optimisation problems. Assorted issues relating to implementation and numerical accuracy are thoroughly reviewed, including multidimensionality and the positive probability requirement, the question of boundary conditions, and the choice of parametric values.

- **Stochastic Volatility for Utility Maximisers Part I**. From an empirical perspective, the stochasticity of volatility is manifest, yet there have been relatively few attempts to reconcile this fact with Merton's theory of optimal portfolio selection for wealth maximising agents. In this paper we present a systematic analysis of the optimal asset allocation in a derivative-free market for the Heston model, the 3/2 model, and a Fong Vasicek type model. Under the assumption that the market price of risk is proportional to volatility, we can derive closed form expressions for the optimal portfolio using the formalism of Hamilton-Jacobi-Bellman. We also perform an empirical investigation,

which strongly suggests that there in reality are no tangible welfare gains associated with hedging stochastic volatility in a bond-stock economy.

- **Stochastic Volatility for Utility Maximisers Part II**. Using martingale methods we derive bequest optimising portfolio weights for a rational investor who trades in a bond-stock-derivative economy characterised by a generic stochastic volatility model. For illustrative purposes we then proceed to analyse the specific case of the Heston economy, which admits explicit expressions for plain vanilla Europeans options. By calibrating the model to market data, we find that the demand for derivatives is primarily driven by the myopic hedge component. Furthermore, upon deploying our optimal strategy on real market prices, we find only a very modest improvement in portfolio wealth over the corresponding strategy which only trades in bonds and stocks.

- **Optimal Hedge Tracking Portfolios in a Limit Order Book**. In this paper we develop a control theoretic solution to the manner in which a portfolio manager optimally should track a targeted Δ, given that he wishes to hedge a short position in European call options the underlying of which is traded in a limit order book. Specifically, we are interested in the interplay between posting limit and market orders respectively: when should the portfolio manager do what (and at what price)? To this end, we set up an Hamilton-Jacobi-Bellman quasi variational inequality which we can solve numerically. Our scheme is shown to be monotone, stable, and consistent and thence, modulo a comparison principle, convergent in the viscosity sense. Finally, we provide a concrete numerical study, comparing our algorithm with more naïve approaches to delta-hedging.

Further to these papers we provide an extensive appendix which summarises standard results from martingale pricing, PDE methods (analytic and numerical), and stochastic control theory. The work presented in this section is mostly (though not exclusively) derivative, and can be read as a friendly reminder with respect to the body of theory underpinning the research articles.

Dansk Resumé

Denne afhandling består af fem forskningsartikler, som er skrevet i perioden januar 2013 - december 2015. Deres resuméer er som følger:

- **Den Fundamentale Teorem om Derivathandel**. Når estimerede volatiliteter ikke er i perfekt overensstemmelse med virkeligheden, vil delta-hedgede optioner med tiden afføde en profit eller et tab ulig nul. Ikke desto mindre findes der en overraskende simpel formel for den resulterende hedge-fejl, hvilken har været kendt siden de sene 90ere. Vi kalder denne formel Den Fundamentale Teorem om Derivathandel. Denne artikel er en oversigt med modifikationer af dette resultat. Vi beviser en mere generel version deraf og diskuterer forskellige viderebygninger (inklusiv spring) samt applikationer (inklusiv en udledning af Dupire-Gyöngy-Derman-Kani formlen). Vi anskuer også dens praktiske konsekvenser både i simulering samt i forbindelse med empirisk data, og demonstrerer således gavnligheden af at hedge med den implicitte volatilitet.

- **Numerisk Stokastisk Kontrolteori med Applikationer i Finansiering**. Analytiske løs-ninger til HJB ligninger i matematisk finansiering er en relativ sjælden foreteelse, hvilket understreger nødvendigheden af numeriske procedurer. I denne artikel fremlægger vi en eksposition af Markov-kæde approksimationsmetoden i endelig tid, først etableret af Kushner og Dupuis. Endvidere anskues detaljerne af hvor godt algoritmen egentlig klarer sig, når vi anvender den i optimeringsproblemer a la Merton. Forskellige problemstillinger relaterende til implementering og numerisk præcision gennemgås nøje, inklusiv multi-dimensionalitet og kravet om positive sandsynligheder, spørgsmålet om randbetingelser, og valget af parametriske værdier.

- **Stokastisk Volatilitet for Nyttemaksimerende Investorer Del I**. Fra et empirisk perspektiv er stokastisk volatilitet veletableret. Dog er der beklageligvis gjort relativt få forsøg på at forene dette faktum med Mertons teori for optimale porteføjevalg for velfærdsmaksimerende investorer. I denne artikel præsenterer vi en systematisk analyse af optimal aktivfordeling i et derivatfrit marked for Heston-modellen, 3/2-modellen, samt en model a la Fong Vasicek. Under den antagelse at markedsprisen for risiko er proportional med volatilitet kan vi udlede lukkede formler for den den optimale porteføłgevægt qua Hamilton-Jacobi-Bellman formalismen. Vi foretager også en empirisk analyse, som stærkt peger i retningen af, at der i praksis ikke er noget at hente på at hedge stokastisk volatilitet i en obligations-aktie økonomi.

- **Stokastisk Volatilitet for Nyttemaksimerende Investorer Del II**. Ved at gøre brug af martingale metoder udvikler vi arv-optimerende porteføljevægte for rationelle investorer, som handler i en obligations-aktie-derivat økonomi karakteriseret ved en generisk stokastisk volatilitetsmodel. For at illustrere vores resultater fortsætter vi da med at analysere Heston modellen som et specifikt eksempel - thi denne tillader eksplicitte løsninger til vanilla Europæiske optioner. Ved at kalibrere vores model til data fra markedet opdager vi, at kravet om derivater primært er drevet af det myopiske hedge komponent. Endvidere: når vi søsætter vores optimale strategi på faktiske markedspriser, finder vi en enddog meget begrænset forbedring i velfærd - i forhold til den tilsvarende strategi som kun handler i obligationer og aktier.

- **Optimale Hedge-sporende Porteføljer i en Ordrebog**. I denne artikel udvikler vi en kontrolteoretisk løsning til metoden hvorigennem en porteføljebestyrer optimalt skal spore et givent Δ, givet at han ønsker at hedge en kort position i Europæiske optioner hvis underliggende aktiv handles in en ordrebog (limit order book). Mere specifikt er vi interesserede i dualiteten mellem det at poste limit-ordrer samt det at poste market-ordrer: hvornår skal porteføljebestyreren gøre hvad (og til hvilken pris?). For at besvare dette spørgsmål fremsætter vi en Hamilton-Jacobi-Bellman kvasi-variationsulighed, som kan løses numerisk. Vores numeriske procedure vises at være monoton, stabil og konsistent og derfor, modulo et sammenligningsprincip, konvergent. Slutteligt præsenterer vi et konkret numerisk studie, hvori vi sammenligner vores algoritme med mere naive tilgange til delta-hedging.

Foruden disse papier fremturer vi med et ekstensivt appendix, som opsummerer standardresultater fra martingale prisfastsættelse, PDE metoder (analytiske og numeriske), samt stokastisk kontrolteori. Arbejdet som præsenteres i denne sektion er mestendels (men ikke 100%) derivativt og kan læses som en venlig erindring angående den teori, som ligger til grund for forskningsartiklerne.

Contents

Part I
Model Risk

Chapter 1
The Fundamental Theorem of Derivative Trading
Exposition, Extensions, and Experiments

Simon Ellersgaard, Martin Jönsson, and Rolf Poulsen

Abstract When estimated volatilities are not in perfect agreement with reality, delta hedged option portfolios will incur a non-zero profit-and-loss over time. There is, however, a surprisingly simple formula for the resulting hedge error, which has been known since the late 90s. We call this The Fundamental Theorem of Derivative Trading. This paper is a survey with twists of that result. We prove a more general version of it and discuss various extensions (including jumps) and applications (including deriving the Dupire-Gyöngy-Derman-Kani formula). We also consider its practical consequences both in simulation experiments and on empirical data thus demonstrating the benefits of hedging with implied volatility.

Key words: Delta Hedging, Model Uncertainty, Volatility Arbitrage.

All authors are with The Department of Mathematical Sciences
Universitetsparken 5, 2100 Copenhagen Ø, Denmark
e-mail: ellersgaard@math.ku.dk, e-mail: maj@math.ku.dk, and e-mail: rolf@math.ku.dk.
Poulsen is the corresponding author.

1.1 A Meditation on the Art of Derivative Hedging

Introduction. Of all possible concepts within the field mathematical finance, that of *continuous time derivative hedging* indubitably emerges as the central pillar. First used in the seminal work by Black and Scholes [6],[1] it has become the cornerstone in the determination of no-arbitrage prices for new financial products. Yet a disconnect between this body of abstract mathematical theory and real world practise prevails. Specifically, successful hedging relies crucially on us having near perfect information about the model that drives the underlying asset. Even if we boldly adopt the standard stochastic differential equation paradigm of asset pricing, it remains to make exact specifications for the degree to which the price process reacts to market fluctuations (i.e. to specify the diffusion term, the volatility). Alas, volatility blatantly transcends direct human observation, being, as it were, a Kantian *Ding an sich*[2] of which we only have approximate knowledge.

One such source comes from measuring the standard deviation of past log returns over time (this is tantamount to assuming that the model can at least *locally* be approximated as a geometric brownian motion). Yet this process raises uncomfortable questions pertaining to statistical measurement: under ordinary circumstances, increasing the sample space should narrow the confidence interval around our sample parameter. Only here, there is no a priori way of telling when a model undergoes a drastic structural change.[3] Inevitably, this implies that extending the time series of log returns too far into the past might lead to a less accurate estimator, as we might end up sampling from a governing dynamics that is no longer valid. Of course, we may take *some* measures against this issue, by trying our luck with ever more intricate time series analyses until we stumble upon a model the parameters of which satisfy our arbitrary tolerance for statistical significance. Nevertheless, in practise this procedure invariably boils down to checking some finite basket of models and selecting the best one from the lot. Furthermore, unknown structural breaks continue to pose a problem no matter what.

Alternatively, we might try to extract an implied volatility from the market by fitting our model to observed option prices. Nevertheless the inadequacy of the methodology quickly becomes apparent: first, implied volatility might be ill-defined as it is the case for certain exotic products such as barrier options. Secondly, it is quite clear that the market hysteria which drives the prices of traded options need not capture the market hysteria which drives the corresponding market for the underlying asset. Fair pricing ultimately boils down to understanding the true nature of the underlying product: not to mimic the collective madness of option traders.

---

[1] That Black and Scholes along with Merton were the first is the general consensus, although the paper by Haug and Taleb [14] shows that the view is not universal.

[2] Literally, *thing in itself* or the *noumenon*. Kant held that there is a distinction between the way things appear to observers (phenomena) and the way reality actually is construed (noumena).

[3] This scenario is not at all implausible. Unlike the physical sciences where the fundamental laws are assumed to have no *sufficient reason* to change (in the Leibnizian or Occamian sense), this philosophical principle would hardly withstand scrutiny in a social science context. Asset price processes are fundamentally governed by market agents and their reactions to various events (be they self-induced or exogenous). There is really no reason to assume that these market players will not drastically change their opinions at some point (for one reason or the other).

Whilst volatility at its core remains elusive to us, the situation is perhaps not as dire as one might think. Specifically, we can develop a formal understanding of the profit-&-loss we incur upon hedging a portfolio with an erroneous volatility - at least insofar as we make some moderate assumptions of the dynamical form of the underlying assets. To give a concrete example of this, consider the simple interest rate free framework presented in Andreasen [2] where the price process of a single non-dividend paying asset is assumed to follow the real dynamics

$$dX_t = \mu_{t,r} X_t dt + \sigma_{t,r} X_t dW_t.$$

Let V_t^i be the value of an option that trades in the market at a certain implied volatility σ_i (possibly quite different from the epistemically inaccessible $\sigma_{t,r}$). Now if we were to set up a hedge of a long position on such an option, using σ_i as our hedge volatility, an application of Itō's formula, coupled with the Black-Scholes equation, shows that the infinitesimal value change in the hedge portfolio

$$\Pi_t = V_t^i - \partial_x V_t^i \cdot X_t,$$

is

$$d\Pi_t = \tfrac{1}{2}(\sigma_{t,r}^2 - \sigma_i^2) X_t^2 \partial_{xx}^2 \Pi_t dt, \tag{1.1}$$

which generally is non-zero unless $\sigma_i = \sigma_{t,r}$. For reasons that will become clearer below, the importance of this result is of such magnitude that Andreasen dubs it *The Fundamental Theorem of Derivative Trading*. Indeed, a more abstract variation of it will be the central object of study in this paper.

To the best of our knowledge, quantitative studies into erroneous delta-hedging leading to a result like (1.1) first appeared in a paper on the robustness of the Black-Scholes formula by El Karoui et al. [13]. They viewed the result as a largely negative one: unless volatility is bounded (which it is not in any stochastic volatility model) then there is no simple super-replication strategy. Subsequently, various sources have re-derived the result (with various tweaks) - most prominently the works of Gibson et al. [17], Henrard [20], Mahayani et al. [21], Rasmussen [22], Carr [7], and Ahmad and Wilmott [1]. Today, the gravity of erroneous Δ-hedging is unquestionably more widely appreciated, yet the Fundamental Theorem of Derivative Trading continues to fly largely under the radar in academia and industry.

Overview. The structure of this paper is as follows: in section 1.2 we state and prove a new, generalised version of the Fundamental Theorem of Derivative Trading and discuss its various implications for hedging strategies and applications (some of which might prove surprising). In section 1.3 we expose the implications of adding a jump process to the framework, thus emphasising the relative ease with which the original proof can be adapted. Finally, section 1.4 presents an empirical investigation into what actually happens to our portfolio when we hedge using various volatilities.

1.2 The Fundamental Theorem of Derivative Trading

1.2.1 Derivation

Model Set-up. Consider a financial market comprised of a risk-free money account as well as n risky assets, each of which pays out a continuous dividend yield. We assume all assets to be infinitely divisible as to the amount which may be held, that trading takes place continuously in time and that no trade is subject to financial friction. Formally, we imagine the information flow of this world to be captured by the stochastic basis $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$, where Ω represents all possible states of the economy, $\mathbb{P}$ is the physical probability measure, and $\mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}$ is a filtration which satisfies the *usual conditions*.[4] The price processes of the risky assets, $\boldsymbol{X}_t = (X_{1t}, X_{2t}, ..., X_{nt})^\intercal$, are assumed to follow the real dynamics[5]

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_t}[\boldsymbol{\mu}_r(t, \widetilde{\boldsymbol{X}}_t)dt + \boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)d\boldsymbol{W}_t], \tag{1.2}$$

where $\mathbf{D}_{\boldsymbol{X}}$ is the $n \times n$ diagonal matrix $\mathrm{diag}(X_{1t}, X_{2t}, ..., X_{nt})$, and $\boldsymbol{W}_t = (W_{1t}, W_{2t}, ..., W_{nt})^\intercal$ is an n-dimensional standard Brownian motion adapted to $\mathbb{F}$. Furthermore, $\boldsymbol{\mu}_r : [0,\infty) \times \mathbb{R}^{n+m} \mapsto \mathbb{R}^n$ and $\boldsymbol{\sigma} : [0,\infty) \times \mathbb{R}^{n+m} \mapsto \mathbb{R}^{n\times n}$ are deterministic functions, sufficiently well-behaved for the SDE to have a unique strong solution (in particular, we assume the regularity conditions

$$\int_t^s |\mathbf{D}_{\boldsymbol{X}_u}\boldsymbol{\mu}_r(u, \widetilde{\boldsymbol{X}}_u)|du < \infty, \qquad \int_t^s |\mathbf{D}_{\boldsymbol{X}_u}\boldsymbol{\sigma}_r(u, \widetilde{\boldsymbol{X}}_u)|^2 du < \infty, \tag{1.3}$$

hold a.s. $\forall t \leq s$, where the first norm is to be understood in the Euclidian sense, whilst the latter should be construed in the matrical sense).[6] Finally, we define $\widetilde{\boldsymbol{X}}_t$ as the $n+m$ dimensional vector $(\boldsymbol{X}_t; \boldsymbol{\chi}_t)$ where $\boldsymbol{\chi}_t = (\chi_{1t}, \chi_{2t}, ..., \chi_{mt})^\intercal$ has the interpretation of an m-dimensional state variable, the exact dynamical nature of which is not integral to what follows.[7]

In what follows we consider the scenario of what happens when we hedge an option on $\boldsymbol{X}_t$, ignorant of the existence of the state variable $\boldsymbol{\chi}_t$, as well as the form of $\boldsymbol{\mu}_r(\cdot,\cdot)$ and $\boldsymbol{\sigma}_r(\cdot,\cdot)$. Specifically, we shall imagine that we are misguided to the extent that we would model the dynamics of $\boldsymbol{X}_t$ as a local volatility model with diffusion matrix $\boldsymbol{\sigma}_h(t, \boldsymbol{X}_t)$. Similar assumptions pertain to the market, although here we label the "implied" diffusion

---

[4] Specifically, it satisfies right-continuity, $\cap_{s\geq t}\mathscr{F}_s = \mathscr{F}_t \ \forall t \geq 0$ (if we move incrementally forward in time there will be no jump in information), and completeness, i.e. $\mathscr{F}_0$ contains all $\mathbb{P}$ null sets.

[5] The nomenclature "real dynamics" is ripe with unfortunate connotations of Platonic realism (ontological significance of mathematical objects). Strictly speaking, this is not what we require, but rather the *expressive adequacy* of a model: i.e. it's ability to adequately capture the financial events unfolding.

[6] Specifically, if $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{n\times d}$, the the Euclidian norm is defined as $|\boldsymbol{x}| \equiv (\sum_{i=1}^n x_i^2)^{1/2}$, while the matrical norm is $|\boldsymbol{A}| \equiv (\sum_{i=1}^n \sum_{j=1}^d A_{ij}^2)^{1/2}$.

[7] Nonetheless, a common assumption in the stochastic volatility literature is obviously to let $\boldsymbol{\chi}$ be driven by a stochastic differential equation of the form $d\boldsymbol{\chi}_t = \boldsymbol{m}(\boldsymbol{\chi}_t)dt + \boldsymbol{v}(\boldsymbol{\chi}_t)d\boldsymbol{W}_t + \bar{\boldsymbol{v}}(\boldsymbol{\chi}_t)d\bar{\boldsymbol{W}}_t$, where $\bar{\boldsymbol{W}}_t$ is second standard Brownian motion (independent of the first), and $\boldsymbol{m}$, $\boldsymbol{v}$ and $\bar{\boldsymbol{v}}$ are dimensionally consistent, regularity conforming vectors and matrices.

matrix $\boldsymbol{\sigma}_i(t, \boldsymbol{X}_t)$ to distinguish it from our personal belief. Irrespective of which dynamical specification is being made, we maintain that regularity conditions analogous to (1.3) remain satisfied. Finally, a cautionary remark: throughout these pages we use r and i to emphasise that the volatility is *real* and *implied* respectively, whilst h refers to an arbitrary *hedge* volatility. For a comprehensible reading, it is incumbent that the reader keeps these definitions in mind.

Theorem 1.1. ***The Fundamental Theorem of Derivative Trading.*** *Let $V_t = V(t, \boldsymbol{X}_t) \in \mathscr{C}^{1,2}([0,\infty) \times \mathbb{R}^n)$ be the price process of a European option with terminal pay-off $V_T = g(\boldsymbol{X}_T)$. Assume we at time $t = 0$ acquire such an option for the market-price V_0^i, with the associated (not necessarily uniquely determined) implied volatility $\boldsymbol{\sigma}_i(0, \boldsymbol{X}_0)$. Furthermore, suppose we set out to Δ-hedge our position, but remain under the impression that the correct volatility ought, in fact, to be $\boldsymbol{\sigma}_h(0, \boldsymbol{X}_0)$, leading to the fair price V_0^h. Then the present value of the* ***profit-&-loss*** *we incur from holding such a portfolio over the interval $\mathbb{T} = [0, T]$ is*

$$P\&L_{\mathbb{T}}^h = V_0^h - V_0^i + \tfrac{1}{2}\int_0^T e^{-\int_0^t r_u du}\, \mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t} \boldsymbol{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t) \mathbf{D}_{\boldsymbol{X}_t} \nabla^2_{\boldsymbol{xx}} V_t^h] dt, \tag{1.4}$$

where $r_u = r(u, \boldsymbol{X}_u)$ is the locally risk free rate, $\nabla^2_{\boldsymbol{xx}}$ is the Hessian operator, and

$$\boldsymbol{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t) \equiv \boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)\boldsymbol{\sigma}_r^\intercal(t, \widetilde{\boldsymbol{X}}_t) - \boldsymbol{\sigma}_h(t, \boldsymbol{X}_t)\boldsymbol{\sigma}_h^\intercal(t, \boldsymbol{X}_t). \tag{1.5}$$

is a matrix which takes values in $\mathbb{R}^{n\times n}$.

Proof: Let $\{\Pi_t^h\}_{t\in[0,T]}$ be the value process of the hedge portfolio long one option valued according to the implied market conception, $\{V_t^i\}_{t\in[0,T]}$, and short $\{\Delta_t^h = \nabla_{\boldsymbol{x}} V_t^h\}_{t\in[0,T]}$ units of the underlying with value process $\{\boldsymbol{X}_t\}_{t\in[0,T]}$, where $\nabla_{\boldsymbol{x}}$ is the gradient operator. We suppose the money account B is chosen such that the net value of the position is zero:

$$\Pi_t^h = V_t^i + B_t - \nabla_{\boldsymbol{x}} V_t^h \bullet \boldsymbol{X}_t = 0,$$

where $\bullet$ is the dot product. Now consider the infinitesimal change to the value of this portfolio over the interval $[t, t+dt]$, where $t \in [0, T)$. From the *self-financing condition* we have that

$$d\Pi_t^h = dV_t^i + r_t B_t dt - \nabla_{\boldsymbol{x}} V_t^h \bullet (d\boldsymbol{X}_t + \boldsymbol{q}_t \circ \boldsymbol{X}_t dt),$$

where $\boldsymbol{q}_t = (q_1(t, X_{1t}), q_2(t, X_{2t}), ..., q_n(t, X_{nt}))^\intercal$ codifies the continuous dividend yields and $\circ$ is the Hadamard (entry-wise) product.[8] Jointly, the two previous equations entail that

$$d\Pi_t^h = dV_t^i - \nabla_{\boldsymbol{x}} V_t^h \bullet (d\boldsymbol{X}_t - (r_t \boldsymbol{\iota} - \boldsymbol{q}_t) \circ \boldsymbol{X}_t dt) - r_t V_t^i dt, \tag{1.6}$$

where $\boldsymbol{\iota} = (1, 1, ..., 1)^\intercal \in \mathbb{R}^n$.

[8] Per definition, if A and B are matrices of equal dimensions, then $(A \circ B)_{ij} = A_{ij} B_{ij}$.

Now consider the option valued under $\boldsymbol{\sigma}_h(t,\boldsymbol{X}_t)$; from the multi-dimensional Itō formula (see for instance Björk [2], p. 65.) we have that

$$dV_t^h = \{\partial_t V_t^h + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}_r^\intercal(t,\widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}_r(t,\widetilde{\boldsymbol{X}}_t)]\}dt + \nabla_{\boldsymbol{x}}V_t^h \bullet d\boldsymbol{X}_t, \tag{1.7}$$

where we have used the fact that $\boldsymbol{X}_t$ is governed by (1.2). Meanwhile, V_t^h satisfies the multi-dimensional Black Scholes equation for dividend paying underlyings (see for instance Björk, Theorem 13.1 and Proposition 16.7),

$$r_tV_t^h = \partial_t V_t^h + \nabla_{\boldsymbol{x}}V_t^h \bullet ((r_t\boldsymbol{\iota} - \boldsymbol{q}_t)\circ \boldsymbol{X}_t) + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}_h^\intercal(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}_h(t,\boldsymbol{X}_t)]. \tag{1.8}$$

Combining this expression with the Itō expansion we obtain,

$$\begin{aligned} 0 =\ & -dV_t^h + r_tV_t^h dt + \nabla_{\boldsymbol{x}}V_t^h \bullet (d\boldsymbol{X}_t - (r_t\boldsymbol{\iota} - \boldsymbol{q}_t)\circ \boldsymbol{X}_t dt) \\ & + \tfrac{1}{2}\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\big(\boldsymbol{\sigma}_r(t,\widetilde{\boldsymbol{X}}_t)\boldsymbol{\sigma}_r^\intercal(t,\widetilde{\boldsymbol{X}}_t) - \boldsymbol{\sigma}_h(t,\boldsymbol{X}_t)\boldsymbol{\sigma}_h^\intercal(t,\boldsymbol{X}_t)\big)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt \end{aligned} \tag{1.9}$$

where we have used the fact that the trace is invariant under cyclic permutations of its constituent matrices. Finally, defining $\boldsymbol{\Sigma}_{rh}(t,\widetilde{\boldsymbol{X}}_t)$ as in (1.5), and adding (1.9) to (1.6) we obtain

$$\begin{aligned} d\Pi_t^h =\ & dV_t^i - dV_t^h - r_t(V_t^i - V_t^h)dt + \tfrac{1}{2}\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{rh}(t,\widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt \\ =\ & e^{\int_0^t r_u du}d(e^{-\int_0^t r_u du}(V_t^i - V_t^h)) + \tfrac{1}{2}\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{rh}(t,\widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt. \end{aligned} \tag{1.10}$$

Whilst a perfect hedge would render this infinitesimal value-change in the portfolio *zero*, this is clearly not the case here. In fact, upon discounting (1.10) back to the present ($t = 0$) and integrating up the infinitesimal components, we find that net profit-&-loss incurred over the life-time of the portfolio is

$$\begin{aligned} P\&L_{\mathbb{T}}^h &= \int_0^T d(e^{-\int_0^t r_u du}(V_t^i - V_t^h)) + \int_0^T e^{-\int_0^t r_u du}\tfrac{1}{2}\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{rh}(t,\widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt \\ &= V_0^h - V_0^i + \tfrac{1}{2}\int_0^T e^{-\int_0^t r_u du}\ \mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{rh}(t,\widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt. \end{aligned}$$

where

$$P\&L_{\mathbb{T}}^h \equiv \int_0^T e^{-\int_0^t r_u du} d\Pi_t^h,$$

and the last line makes use of the fact that $V_T^i = V_T^h = g(\boldsymbol{X}_T)$. This is the desired result. □

Remark 1.1. A few observations on this proof are in order: first, the relative simplicity of (1.4) clearly boils down to the assumption that the market is perceived to be driven by a local volatility model. If this assumption is dropped equation (1.8) no longer holds. Secondly, it should be clear that the value of the *P&L* changes sign if we are short on the derivative and long the underlying. Thirdly, the market price of the derivative enters only though the initial price V_0. That is because we look at the profit-&-loss accrued over the entire life-time of the portfolio. The case of marking-to-market requires further analysis and/or assumption. We will elaborate on this in the following subsection.

Remark 1.2. From a generalist's perspective, theorem 1 suffers from a number of glaring limitations: for instance, the governing asset price dynamics only considers Brownian stochasticity, the hedge is assumed to be a workaday Δ-hedge, and the option type is vanilla European in the sense that the terminal pay-off is determined by the instantaneous price of the underlying assets. Fortunately, the Fundamental Theorem can readily be extended in various directions: e.g. it can be shown that if $V_t = V(t, X_t, A_t)$ is an Asian option written on the continuous average A_t of the underlying process X_t, then the Fundamental Theorem remains form invariant. In section 1.3 we consider one particularly topical dynamical modification viz. the incorporation of possible market crashes through jump diffusion.

1.2.2 The Implications for Δ-Hedging.

From a first inspection, the Fundamental Theorem quite clearly demonstrates that reasonably successful hedging is possible *even* under significant model uncertainty. Indeed, as Davis [10] puts it "without some robustness property of this kind, it is hard to imagine that the derivatives industry could exist at all". In this section, we dive further into the implications of what happens to our portfolio, by considering the case where we hedge with (a) the real volatility, and (b) the implied volatility.

Hedging With the Real Volatility. Suppose we happen to be bang-on our estimate of the real volatility matrix in our Δ-hedge, i.e. let $\boldsymbol{\sigma}_h(t, \boldsymbol{X}_t) = \boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)$ a.s. $\forall t \in [0, T]$, then $\boldsymbol{\Sigma}_{rr}(t, \widetilde{\boldsymbol{X}}_t) = \mathbf{0}$ and the present valued profit-&-loss amounts to

$$P\&L^r_{\mathbb{T}} = V^r_0 - V^i_0,$$

which is manifestly deterministic.[9] However, we observe that this relies crucially on us holding the portfolio until expiry of the option. Day-to-day fluctuations of the profit-&-loss still vary stochastically (erratically) as it is vividly demonstrated by combining equation (1.9) (where $h = i$) with equation (1.6) (where $h = r$):

$$\begin{aligned} d\Pi^r_t = &\ \tfrac{1}{2}\,\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{x}\boldsymbol{x}} V^i_t]dt \\ &+ \nabla_{\boldsymbol{x}}\left(V^i_t - V^r_t\right) \bullet \left\{(\boldsymbol{\mu}^r_t - r_t\boldsymbol{\iota} + \boldsymbol{q}_t)\circ \boldsymbol{X}_t dt + \mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)d\boldsymbol{W}_t\right\}, \end{aligned}$$

cf. the explicit dependence of the Brownian increment. As for the *profitability* of the Δ-hedging strategy, this is a complex issue which ultimately must be studied on a case-by-case basis. However, for options with positive *vega*,[10] it suffices to require that the real volatility everywhere exceeds the implied volatility.

[9] Obviously, this can only be the case if there is no underlying state variable.

[10] A clear example of vega being manifestly positive would be European calls and puts, which satisfy the assumptions needed to derive the Black-Scholes formula. Explicitly, $\nu \equiv \frac{\partial V}{\partial \sigma} = S_t e^{-\delta(T-t)}\phi(d_1)\sqrt{T-t} > 0$ where ϕ is the standard normal pdf and d_1 has the usual definition.

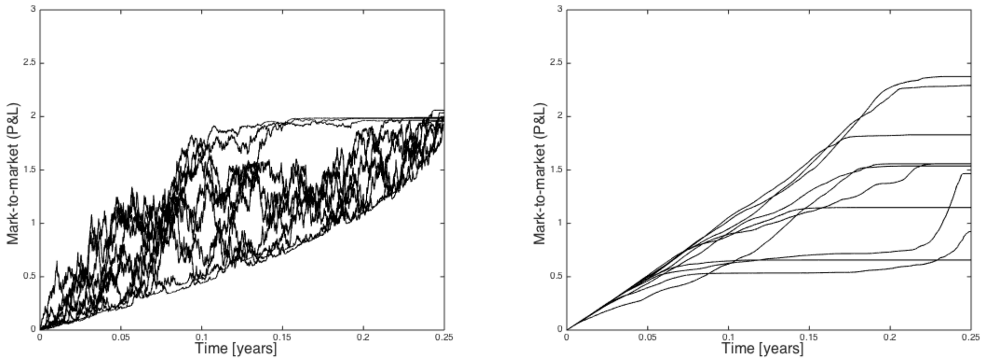

Fig. 1.1 Left: Delta hedging a portfolio assuming that $\sigma_h = \sigma_r$. The parameter specifications are: $r = 0.05$, $\mu = 0.1$, $\sigma_i = 0.2$, $\sigma_r = 0.3$, $S_0 = 100$, $K = 100$, $q = 0$ and $T = 0.25$. The portfolio is rebalanced 5000 times during the lifetime of the option. Observe that while the P&L fluctuates randomly along the path of S_t due to the presence of dW_t, the accumulated P&L at the maturity of the option is the deterministic quantity $\Pi_T = e^{rT}(V_0^r - V_0^i)$. From the Black-Scholes formula it follows that $V_0^r = 6.583$ and $V_0^i = 4.615$ so $\Pi_{T=1} = 1.993$. The fact that our ten paths only approximately hit this terminal value is attributable to the discretisation of the hedging which should be done in continuous time. **Right:** Delta hedging a portfolio assuming that $\sigma_h = \sigma_i$.The parameter specifications are as before. Evidently, the accumulated P&L stays highly path dependent for the *entire* duration of the option. However, the curves per se are smooth, which highlights that $d\Pi_t^i$ does not depend explicitly on the Brownian increment.

Hedging With the Implied Volatility. Suppose instead we hedge the portfolio using the implied volatility matrix $\boldsymbol{\sigma}_i(t, \boldsymbol{X}_t)\ \forall t \in [0, T]$, then the associated present-valued profit-&-loss is of the form

$$P\&L_{\mathbb{T}}^i = \frac{1}{2}\int_0^T e^{-\int_0^t r_u du}\, \mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t} \boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t} \nabla_{\boldsymbol{xx}}^2 V_t^i] dt.$$

As we find ourselves integrating over the stochastic process $\boldsymbol{X}_t$, this profit-&-loss is manifestly stochastic. Notice though that $d\Pi_t^i$ here does *not* depend explicitly on the Brownian increment (the daily profit-and-loss is $\mathscr{O}(dt)$) which gives rise to point that "bad models cause bleeding - not blow-ups". As for the *profitability* of the strategy, again this is a complex issue: however, insofar as $\boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t) \circ \nabla_{\boldsymbol{xx}}^2 V_t^i$ is *positive definite* a.s. for all $t \in [0, T]$, then we're making a profit with probability one. To see this, recall that the trace can be written as[11]

$$\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t} \boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t} \nabla_{\boldsymbol{xx}}^2 V_t^i] = \boldsymbol{X}_t^\intercal (\boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t) \circ \nabla_{\boldsymbol{xx}}^2 V_t^i)\boldsymbol{X}_t,$$

In particular, if $\boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t) \circ \nabla_{\boldsymbol{xx}}^2 V_t^i$ is positive definite at all times, i.e.

$$\forall t \in [0, T]\ \ \forall \boldsymbol{X}_t \in \mathbb{R}^n : \quad \boldsymbol{X}_t^\intercal (\boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t) \circ \nabla_{\boldsymbol{xx}}^2 V_t^i)\boldsymbol{X}_t > 0,$$

[11] This follows from the general identity for matrices $\mathbf{A}$ and $\mathbf{B}$ of corresponding dimensions: $\mathbf{x}^\intercal(\mathbf{A} \circ \mathbf{B})\mathbf{y} = \mathrm{tr}[\mathbf{D_x A D_y B}^\intercal]$ where $\mathbf{x}$ and $\mathbf{y}$ are vectors.

then $P\&L^i_{\mathbb{T}} > 0$. A sufficient condition for this to be the case is that $\boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t)$ and $\nabla^2_{\boldsymbol{xx}} V^i_t$ individually are positive definite $\forall t$, as demonstrated by the *Schur Product Theorem*.

Wilmott's Hedge Experiment. The points imbued in the previous two paragraphs are forcefully demonstrated in the event that there is only one risky asset in existence, the derivative is a European call option and all volatilities are assumed constant. Based on Wilmott and Ahmad, Figure 1.1 clearly illustrates the behaviour of the profit-&-loss paths insofar as we hedge with (a) the real volatility, and (b) the implied volatility. Again, the main insights are as follows: hedging V^i_t with the real volatility causes the $P\&L$ of the portfolio to fluctuate erratically over time, only to land at a deterministic value at maturity. On the other hand, hedging V^i_t with the implied volatility yields smoother (albeit still stochastic) P&L curves. Nonetheless, here there is no way of telling what the P&L actually amounts to at maturity.

Rather perturbingly, both strategies blatantly suggest the relative ease with which we can make *volatility arbitrage*. Specifically, assuming that the historical volatility is a reasonable proxy for the real volatility, $\sigma_{\text{hist}} \approx \sigma_r$, and that $\sigma_{\text{hist}} > \sigma_i$ ($\sigma_{\text{hist}} < \sigma_i$), it would suffice to go long (short) on the hedge portfolio for $\mathbb{P}(P\&L_{\mathbb{T}} \geq 0) = 1$ and $\mathbb{P}(P\&L_{\mathbb{T}} > 0) > 0$.

Reality, of course, is not always as simple as our abstract idealisations, wherefore we dedicate section four to an empirical investigation of Wilmott's hedge experiment.

1.2.3 Applications

Due to the presence of the real volatility, the exact nature of which transcends our epistemic domain, one might reasonably ponder whether the Fundamental Theorem conveys any practical points besides those of the preceding subsection. Using two poignant (even if somewhat eccentric) examples, we will argue that the gravity of the Fundamental Theorem propagates well into risk management and volatility surface calibration. Zero rates and dividends will be assumed throughout.

Example 1.1. Let $V_t(T,K)$ be the price process of a European strike K maturity T call or put option, written on an underlying which obeys Geometric Brownian Motion, $dX_t = \mu_r X_t dt + \sigma_r X_t dW_t$, where μ_r, σ_r are constants. Suppose we Δ-hedge a long position on V_t at the *implied* volatility, $\sigma_h = \sigma_i$, then the Fundamental Theorem implies that

$$P\&L^i_{\mathbb{T}} = \tfrac{1}{2} \int_0^T (\sigma_r^2 - \sigma_i^2) X_t^2 \Gamma^i_t \, dt,$$

where

$$\Gamma^i_t \equiv \frac{\phi(d^i_1)}{X_t \sigma_i \sqrt{T-t}},$$

is the option's gamma, $\phi : \mathbb{R} \mapsto \mathbb{R}_+$ is the standard normal pdf and

$$d^i_1 \equiv \tfrac{1}{\sigma_i \sqrt{T-t}} \left\{ \ln(X_t/K) + \tfrac{1}{2}\sigma_i^2 (T-t) \right\}.$$

Since $\forall t\ \Gamma_t^i > 0$ the strategy is profitable if and only if $\sigma_r^2 > \sigma_i^2$. Furthermore, by maximising the integrand with respect to X_t we find that the $P\&L_{\mathbb{T}}^i$ is maximal when

$$X_t^* = Ke^{\frac{1}{2}\sigma_i^2(T-t)},$$

Specifically, upon evaluating the integral explicitly we find that

$$\max_{X_t} P\&L_{\mathbb{T}}^i = \sqrt{\frac{T}{2\pi}}\frac{K}{\sigma_i}(\sigma_r^2 - \sigma_i^2).$$

Using elementary statistics we can compute a confidence interval for the real volatility based on historical observations. Hence, we can compute a confidence interval for the maximal profit-&-loss we might face upon holding the hedge portfolio till expiry.

Example 1.2. Let $V_t = C_t(T,K)$ be the price process of a European strike K maturity T call option written on an underlying price process X. As in (1.2) we assume the fundamental dynamics to be of the form

$$dX_t = \mu_r(t,\widetilde{X}_t)X_t dt + \sigma_r(t,\widetilde{X}_t)X_t dW_t,$$

where $\widetilde{X}_t$ is defined as the $(1+m)$-dimensional vector $(X_t;\chi_t)$ and χ is a state variable. Also, we suppose

$$\mathbb{E}\left[\int_0^T \sigma_r^2(t,\widetilde{X}_t)X_t^2 dt\right] < \infty,$$

and that there exists an equivalent martingale measure, $\mathbb{Q}$, which renders X_t a martingale (recall the risk free rate is assumed zero):[12]

$$dX_t = \sigma_r(t,\widetilde{X}_t)X_t dW_t^{\mathbb{Q}}.$$

Now consider the admittedly somewhat contrived scenario of a Δ-hedged portfolio, long one unit of the call, for which σ_h and σ_i are both zero.[13] The associates value process is

$$\Pi_t^i = C_t^i(T,K) + B_t - \partial_x C_t^h(T,K)\cdot X_t = (X_t - K)^+ + B_t - \mathbb{1}_{\{X_t > K\}}X_t, \tag{1.11}$$

[12] Obviously, such an existence claim is not altogether innocuous. Indeed, the measure change is here further complicated by the fact that we have not made formal specifications for the dynamical form of the state variable χ_t. However, insofar as we adopt the standard dynamical assumption $d\chi_t = \boldsymbol{m}(\chi_t)dt + \boldsymbol{v}(\chi_t)dW_t + \bar{\boldsymbol{v}}(\chi_t)d\bar{\boldsymbol{W}}_t$, our existence claim is tantamount to positing the existence of a market price of risk vector $\boldsymbol{\theta} \in \mathbb{R}^m$ which renders the process $L(T) = L_X(T)L_\chi(T)$ a true martingale, where

$$L_X(T) \equiv \exp\left\{-\int_0^T \frac{\mu_r(t,\widetilde{X}_t)}{\sigma_r(t,\widetilde{X}_t)}dW_t - \frac{1}{2}\int_0^T \frac{\mu_r^2(t,\widetilde{X}_t)}{\sigma_r^2(t,\widetilde{X}_t)}dt\right\},$$

and

$$L_\chi(T) \equiv \exp\left\{-\int_0^T \boldsymbol{\theta}_t^\intercal d\bar{\boldsymbol{W}}_t - \frac{1}{2}\int_0^T |\boldsymbol{\theta}_t|^2 dt\right\}.$$

[13] To be precise, the contrived part is the assumption that the call trades at zero volatility; less so that we hedge it at zero volatility. The latter corresponds to a so-called *stop-loss strategy*, see Carr [8].

where $\mathbb{1}_{\{X_t>K\}}$ is the indicator function. The important point here is that $(X_t-K)^+$ may be reinterpreted as the terminal pay-off of a strike K maturity t call option (obviously, the specification $\sigma_h=\sigma_i=0$ is paramount here). Substituting (1.11) into the infinitesimal form of the Fundamental Theorem,

$$d\Pi_t^i = \tfrac{1}{2}(\sigma_r^2(t,\widetilde{X}_t)-\sigma_i^2)X_t^2\partial_{xx}^2 C_t^i(T,K)dt,$$

we find that

$$d((X_t-K)^+ + B_t - \mathbb{1}_{\{X_t>K\}}X_t) = \tfrac{1}{2}\sigma_r^2(t,\widetilde{X}_t)X_t^2\delta(X_t-K)dt, \tag{1.12}$$

where we once again have made use of $\sigma_i=0$, alongside the fact that $\partial_x\mathbb{1}_{\{X_t>K\}}$ is the Dirac delta-function $\delta(X_t-K)$. Taking the risk neutral expectation of (1.12), conditional on $\mathscr{F}_0$, the left-hand side reduces to

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[LHS] &= \mathbb{E}^{\mathbb{Q}}[d(X_t-K)^+] + \mathbb{E}^{\mathbb{Q}}[dB_t - \mathbb{1}_{\{X_t>K\}}dX_t] \\
&= d\mathbb{E}^{\mathbb{Q}}[(X_t-K)^+] - \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{X_t>K\}}dX_t] \\
&= dC_0^r(t,K) - \mathbb{E}^{\mathbb{Q}}[\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{X_t>K\}}dX_t|\mathscr{F}_t]] \\
&= dC_0^r(t,K) - \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{X_t>K\}}\mathbb{E}^{\mathbb{Q}}[dX_t|\mathscr{F}_t]] \\
&= dC_0^r(t,K),
\end{aligned} \tag{1.13}$$

where the second line uses $r=0$ (whence $dB_t=0$), whilst the third line uses the law of iterated expectations and the fact that $\mathbb{E}^{\mathbb{Q}}[(X_t-K)^+]$ is the time zero price of a strike K maturity t call option. Finally, the fourth line follows from the $\mathscr{F}_t$-measurability of $\mathbb{1}_{\{X_t>K\}}$, whilst the fifth line exploits the martingale property $\mathbb{E}^{\mathbb{Q}}[dX_t]=0$.

As for the right-hand side, define the joint density

$$f^{\mathbb{Q}}_{\sigma_r^2,X_t}(\sigma^2,x)d\sigma^2dx = \mathbb{Q}(\{\sigma^2\le\sigma_r^2\le\sigma^2+d\sigma^2\}\cap\{x\le X_t\le x+dx\}),$$

then

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[RHS] &= \tfrac{1}{2}\iint_{\mathbb{R}_+^2}\sigma^2x^2\delta(x-K)f^{\mathbb{Q}}_{\sigma_r^2,X_t}(\sigma^2,x)d\sigma^2dxdt \\
&= \tfrac{1}{2}\iint_{\mathbb{R}_+^2}\sigma^2x^2\delta(x-K)f^{\mathbb{Q}}_{\sigma_r^2}(\sigma^2|X_t=x)f^{\mathbb{Q}}_{X_t}(x)d\sigma^2dxdt \\
&= \tfrac{1}{2}\int_{\mathbb{R}_+}x^2\delta(x-K)f^{\mathbb{Q}}_{X_t}(x)\left\{\int_{\mathbb{R}_+}\sigma^2 f^{\mathbb{Q}}_{\sigma_r^2}(\sigma^2|X_t=x)d\sigma^2\right\}dxdt \\
&\equiv \tfrac{1}{2}\int_{\mathbb{R}_+}x^2\delta(x-K)f^{\mathbb{Q}}_{X_t}(x)\mathbb{E}^{\mathbb{Q}}[\sigma_r^2(t,\widetilde{X}_t)|X_t=x]dxdt \\
&= \tfrac{1}{2}K^2f^{\mathbb{Q}}_{X_t}(K)\mathbb{E}^{\mathbb{Q}}[\sigma_r^2(t,\widetilde{X}_t)|X_t=K]dt.
\end{aligned} \tag{1.14}$$

Recalling that $\partial_K\mathbb{E}^{\mathbb{Q}}[(X_t-K)\mathbb{1}_{\{X_t>K\}}] = -\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{X_t>K\}}]$, and $-\partial_K\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{X_t>K\}}] = \mathbb{E}^{\mathbb{Q}}[\delta(X_t-K)]$ we arrive at the Breeden-Litzenberger formula

$$f^{\mathbb{Q}}_{X_t}(K) = \partial^2_{KK} C^r_0(t,K). \tag{1.15}$$

Combining equations (1.13), (1.14), and (1.15) we thus have that

$$\frac{dC^r_0}{dt}(t,K) = \tfrac{1}{2}\partial^2_{KK} C^r_0(t,K) K^2 \mathbb{E}^{\mathbb{Q}}[\sigma^2_r(t,\widetilde{X}_t)|X_t = K],$$

which using the change of notation[14] $t = T$ amounts to the celebrated Dupire-Gyöngy-Derman-Kani formula

$$\mathbb{E}^{\mathbb{Q}}[\sigma^2_r(T,\widetilde{X}_T)|X_T = K] = \frac{\partial_T C^r_0(T,K)}{\frac{1}{2}K^2 \partial^2_{KK} C^r_0(T,K)}, \tag{1.16}$$

- see e.g. Dupire [12] or Derman and Kani [11]. Using some amount of extrapolation,[15] the righthand side is empirically measurable, hence (1.16) provides a way of calibrating the volatility surface to observed call option prices in the market.

Remark 1.3. In Wittgensteinian terms we must "throw away the ladder" to arrive at this final conclusion, [28] prop. 6.54. Hitherto, we have assumed that the real parameters (r) are fundamentally unobservable, whilst the implied parameters (i) are those we are exposed to in the market. Yet, no such distinction exists in the works of Dupire et al., whence the r superscript in (1.16) really ought to be dropped.

Remark 1.4. The above derivation is arguably unconventional and neither rigorous nor the quickest way to demonstrate (1.16). In fact, the entire point of setting $\sigma_i = 0$ is essentially to extract the Itō-(Tanaka) formula applied to $(X_t - K)^+$, from which Derman et al.'s derivation takes its starting point. We keep the derivation here, as it provides a curious glimpse into how two philosophically quite distinct theorems can be interconnected.

1.3 The Gospel of the Jump

Following remark 1.2, it is worthwhile exploring how the Fundamental Theorem can be adapted to new terrain. For instance, it is well known that Brownian motion in itself does not adequately capture the sporadic discontinuities that emerge in stock price processes. Hence, it is opportune to scrutinise the effect of a jump diffusion process, which in turn will give rise to another valuable lesson on the profitability of imperfect hedging.

Already, it is a well-known fact that exact hedges generally do not exist in a jump economy where the true dynamics of the underlying is perfectly disseminated (see e.g. Shreve [23] or Privault [19]). It is thus of some theoretical interest to see how this preexisting hedge error is further complicated under the model error framework of the Fundamental Theorem. We note that this problem has been treated (with various degrees of rigour) in

[14] We do this to emphasise that t is the *maturity* of the option (not its value at time t).

[15] Exactly how to do this extrapolation has turned out to be sufficiently non-trivial to spurn numerous papers and successive quant-of-the-year awards a-decade-and-a-half later, see Andreasen and Huge [3] (pure local volatility), Guyon, J. and Henry-Labordère [18] (decorated stochastic volatility models).

Andreasen [2] and Davis [10] when the hedge volatility is implied. Our main contribution is to generalise to the multi-dimensional framework with arbitrary specifications for the volatility and jump distribution. For an overview of multi-dimensional jump-diffusion theory we refer the reader to the appendix.

Suppose the real dynamics of the underlying price process obeys

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_t}[\boldsymbol{\mu}_r(t,\widetilde{\boldsymbol{X}}_t)dt + \boldsymbol{\sigma}_r(t,\widetilde{\boldsymbol{X}}_t)d\boldsymbol{W}_t] + \mathbf{D}_{\boldsymbol{X}_{t-}}d\boldsymbol{Y}_t \tag{1.17}$$

where $\{\boldsymbol{Y}_t\}_{t\geq 0}$ is an n-dimensional vector of independent compound Poisson processes. Specifically, the j^{th} component is given by

$$[\boldsymbol{Y}_t]_j \equiv Y_t^j = \sum_{k=1}^{N_t^j} Z_k^j,$$

where $\{N_t^j\}_{t\geq 0}$ is an intensity-λ_j Poisson process, and $\{Z_k^j\}_{k\geq 1}$ is a sequence of relative jump-sizes, assumed to be i.i.d. square-integrable random variables with cumulative distribution function (cdf) $\nu_j : \mathbb{R} \mapsto [0,1]$. For shorthand, we shall refer to the vectors $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_n)^\intercal$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, ..., \nu_n)^\intercal$ as the intensity and cdf of $\boldsymbol{Y}_t$.

Oblivious to the true nature of (1.17), we imagine that pricing and hedging should be performed (with obvious notation) under the tuple $\langle \boldsymbol{\phi}, \boldsymbol{\lambda}_h^{\mathbb{Q}}, \boldsymbol{\nu}_h^{\mathbb{Q}}, \boldsymbol{\sigma}_h(t,\boldsymbol{X}_t), \mathbb{Q}\rangle$, where $\mathbb{Q}$ is the risk neutral measure

$$\begin{aligned} d\mathbb{Q}_{\boldsymbol{\phi},\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}} = {} & \exp\left\{ \int_0^T \boldsymbol{\phi}_s \bullet d\boldsymbol{W}_s - \tfrac{1}{2}\int_0^T |\boldsymbol{\phi}_s|^2 ds - \sum_{j=1}^n (\lambda_{h,j}^{\mathbb{Q}} - \lambda_{h,j})T \right\} \\ & \cdot \prod_{j=1}^n \prod_{k=1}^{N_t^j} \frac{\lambda_{h,j}^{\mathbb{Q}} d\nu_{h,j}^{\mathbb{Q}}(Z_k^j)}{\lambda_{h,j} d\nu_{h,j}(Z_k^j)} d\mathbb{P}^h, \end{aligned} \tag{1.18}$$

such that $\{\boldsymbol{\phi}_t\}_{t\geq 0}$ is a bounded adapted n-dimensional process, and $\boldsymbol{\lambda}_h^{\mathbb{Q}}, \boldsymbol{\nu}_h^{\mathbb{Q}}$ respectively represent the jump intensity and jump-size distribution under $\mathbb{Q}$. Specifically, the price of an option with terminal pay-off $g(\boldsymbol{X}_T)$ is determined as

$$V_t^h = \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u du} g(\boldsymbol{X}_T) | \mathscr{F}_t^{\boldsymbol{X}}],$$

with the underlying supposedly driven by

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_t}[r_t \boldsymbol{\iota} dt + \boldsymbol{\sigma}_h(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}] + \mathbf{D}_{\boldsymbol{X}_{t-}}[d\boldsymbol{Y}_t - \boldsymbol{\lambda}_h^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]]$$

with $\boldsymbol{Z}_1 = (Z_1^1, Z_1^2, ..., Z_1^n)^\intercal$, and $\mathbb{Q}$ has been specified such that

$$\boldsymbol{\mu}_h(t,\boldsymbol{X}_t) + \boldsymbol{\lambda}_h^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1] + \boldsymbol{\sigma}_h(t,\boldsymbol{X}_t)\boldsymbol{\phi}_t = r_t \boldsymbol{\iota}, \tag{1.19}$$

is satisfied almost everywhere.[16]

[16] It should be clear the $\mathbb{Q}$ is not uniquely determined. In fact, for (1.19) to admit only one solution, we would require that either (i) $\boldsymbol{\lambda}_h = \boldsymbol{\lambda}_h^{\mathbb{Q}} = 0$ (there are *no* jumps), in which case we recover the standard

Remark 1.5. We emphasise that (1.18) is a risk neutral measure transformation of the *hedge* dynamics with the associated measure $\mathbb{P}^h$. This is to be contrasted with example 2 in subsection 1.2.3 in which $\mathbb{Q}$ is the risk neutral measure of the *real* dynamics.

Theorem 1.2. ***The Fundamental Theorem of Derivative Trading with Jumps.*** *Let $V_t = V(t, \boldsymbol{X}_t) \in \mathscr{C}^{1,2}([0,\infty)\times\mathbb{R}^n)$ be the price process of a European option with terminal pay-off $V_T = g(\boldsymbol{X}_T)$. Assume we at time $t = 0$ acquire such an option for the market-price V_0^i, with the associated (not necessarily uniquely determined) implied volatility $\boldsymbol{\sigma}_i(0, \boldsymbol{X}_0)$. Furthermore, suppose we set out to Δ-hedge our position, but remain under the impression that the correct volatility ought, in fact, to be $\boldsymbol{\sigma}_h(0, \boldsymbol{X}_0)$, leading to the fair price V_0^h. Then the present value of the* ***profit-&-loss*** *we incur from holding such a portfolio over the interval $\mathbb{T} = [0, T]$ is*

$$
\begin{aligned}
P\&L_{\mathbb{T}}^h = V_0^h - V_0^i + \tfrac{1}{2}\int_0^T e^{-\int_0^t r_u du}\, \mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t} \boldsymbol{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t) \mathbf{D}_{\boldsymbol{X}_t} \nabla^2_{\boldsymbol{x}\boldsymbol{x}} V_t^h] dt,& \\
+ \int_0^T \sum_{j=1}^n e^{-\int_0^t r_u du} \Big\{ \Big(\Delta_j V_t^h(t, \boldsymbol{X}_{t-}) - X_{j,t-} Z^j_{N_t^j} \partial_{x_j} V_t^h \Big) dN_t^j & \qquad (1.20) \\
- \lambda^{\mathbb{Q}}_{h,j} \Big(\mathbb{E}^{\mathbb{Q}}[\Delta_j V_t^h(t, \boldsymbol{x})]|_{\boldsymbol{x}=\boldsymbol{X}_{t-}} - X_{j,t-}\mathbb{E}^{\mathbb{Q}}[Z_1^i] \partial_{x_j} V_t^h \Big) dt \Big\},&
\end{aligned}
$$

where

$$
\Delta_j V_t^h(t, \boldsymbol{X}_{t-}) \equiv V^h(t, \boldsymbol{X}_{t-} \circ (\boldsymbol{\iota} + \hat{\mathbf{e}}_j Z^j_{N_t})) - V^h(t, \boldsymbol{X}_{t-}),
$$

represents the change in value of the option when the underlying jumps in the j^{th} component, and $[\hat{\mathbf{e}}_j]_k = \delta_{j,k}$ is a unit vector in $\mathbb{R}^n$.

Sketch Proof: The proof runs in parallel with that of theorem 1. Specifically, the analogue of expression (1.6) is

$$
d\Pi_t^h = dV_t^i - \nabla_{\boldsymbol{x}} V_t^h \bullet (d\boldsymbol{X}_t^{\text{cont.}} - (r_t \boldsymbol{\iota} - \boldsymbol{q}_t) \circ \boldsymbol{X}_t dt) - \nabla_{\boldsymbol{x}} V_t^h \bullet d\boldsymbol{Y}_t - r_t V_t^i dt,
$$

where $d\boldsymbol{X}_t^{\text{cont.}}$ is the continuous part of (1.17) i.e.

$$
d\boldsymbol{X}_t^{\text{cont.}} = \mathbf{D}_{\boldsymbol{X}_t}[\boldsymbol{\mu}_r(t, \widetilde{\boldsymbol{X}}_t) dt + \boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t) d\boldsymbol{W}_t].
$$

Furthermore, in analogy with (1.7) and (1.8) we have the Itō formula

$$
\begin{aligned}
dV_t^h = \{\partial_t V_t^h + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}_r^{\intercal}(t, \widetilde{\boldsymbol{X}}_t) \mathbf{D}_{\boldsymbol{X}_t} \nabla^2_{\boldsymbol{x}\boldsymbol{x}} V_t^h \mathbf{D}_{\boldsymbol{X}_t} \boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)]\} dt + \nabla_{\boldsymbol{x}} V_t^h \bullet d\boldsymbol{X}_t^{\text{cont.}},& \\
+ \sum_{j=1}^n [V^h(t, \boldsymbol{X}_{t-} \circ (\boldsymbol{\iota} + \hat{\mathbf{e}}_j Z^j_{N_t})) - V^h(t, \boldsymbol{X}_{t-})] dN_t^j,&
\end{aligned}
$$

Girsanov theorem with $\boldsymbol{\phi}_t = \boldsymbol{\sigma}_h^{-1}(\boldsymbol{\mu}_h - r\boldsymbol{\iota})$, or (ii) when $\boldsymbol{\sigma}_h = 0$ (there are *only* jumps) in which case $\boldsymbol{\lambda}_h^{\mathbb{Q}} = (\boldsymbol{\mu}_h - r_t \boldsymbol{\iota}) \oslash \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]$ where $\oslash$ is Hadamard division.

and the partial integro-differential equation for pricing purposes

$$\begin{aligned} r_t V_t^h &= \partial_t V_t^h + \nabla_{\boldsymbol{x}} V_t^h \bullet ((r_t \boldsymbol{\iota} - \boldsymbol{q}_t) \circ \boldsymbol{X}_t) + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}_h^\intercal(t,\boldsymbol{x}) \mathbf{D}_{\boldsymbol{X}_t} \nabla^2_{\boldsymbol{x}\boldsymbol{x}} V_t \mathbf{D}_{\boldsymbol{X}_t} \boldsymbol{\sigma}_h(t,\boldsymbol{X})] \\ &\quad + \textstyle\sum_{j=1}^n \lambda_{h,j}^{\mathbb{Q}} \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\hat{\mathbf{e}}_j Z_1^j)) - V(t,\boldsymbol{x}) - x_j Z_1^j \partial_{x_j} V(t,\boldsymbol{x})]_{\boldsymbol{x}=\boldsymbol{X}_{t-}}, \end{aligned}$$

Combining these three expressions as above yields the desired result. □

Remark 1.6. The last two lines in (1.20) (which we denote by $P\&L_J$) represent the present-valued profit-&-loss brought about by our inability to hedge the jump risk completely. Letting $J_{j,h}^{\mathbb{Q}}(dt \times dz^j)$ be a Poisson random measure with intensity

$$\mathbb{E}^{\mathbb{Q}}[J_{j,h}^{\mathbb{Q}}(dt \times dz^j)] = \lambda_{h,j}^{\mathbb{Q}} dt d\nu_{h,j}^{\mathbb{Q}}(z^j),$$

for $j = 1,2,...,n$, we have that the jump contribution to the profit-and-loss may be written as

$$P\&L_J = \int_0^T \int_{\mathbb{R}} \sum_{j=1}^n e^{-\int_0^t r_u du} \{\Delta_j V_t^h(t,\boldsymbol{X}_{t-}) - z^j X_{j,t-} \partial_{x_j} V_t^h\} J_{j,h}^{\mathbb{Q}}(dt \times dz_j). \tag{1.21}$$

This highlights that if V is convex in all of its components (a property it will inherit from the payoff function under mild conditions) then $\forall j : \Delta V > \partial_{x_j} V \Delta X_j$ whence the integrand in $P\&L_J$ is positive. Thus, our hedge portfolio actually benefits from jumps (in either direction) of any of the underlying price process. Conversely, if we had shorted the option, the hedge profit would obviously take a hit in the event of a jump (in Talebian terms, holding a hedge portfolio with a short option position corresponds to "picking pennies in front of a steam roller"[17]). A vivid illustration of this point is provided in figure 1.2 for an option written on a single underlying.

1.4 Insights From Empirics: On Arbitrage and Erraticism

Inspired by Wilmott's theoretical hedge experiment, we now look into the empirical performance of Δ-hedging strategies based on (I) forecasted implied volatilities and (II) forecasted actual (i.e. historical) volatilities. Specifically, we are interested in the properties of the accumulated P&L, insofar as we Δ-hedge, till expiry, a three-month call-option on the S&P500 index, initially purchased at-the-money. We investigate a totality of 36 such portfolios over disjoint intervals between July 2004 and July 2013. This involves market data on both the underlying index and on options. Daily data on the S&P500 index is readily and freely available. For option data, we combine a 2004-2009 data set from a major commercial bank[18] with more recent prices from OptionMetrics obtained via the Wharton Financial Database.

[17] [27] p.19, "'*Most traders were just "picking pennies in front of a steam roller," exposing themselves to the high impact rare event yet sleeping like babies, unaware of it.*"

[18] The bank shall remain nameless, but the data can be downloaded from http://www.math.ku.dk/ rolf/Svend/http://www.math.ku.dk/∼rolf/Svend/

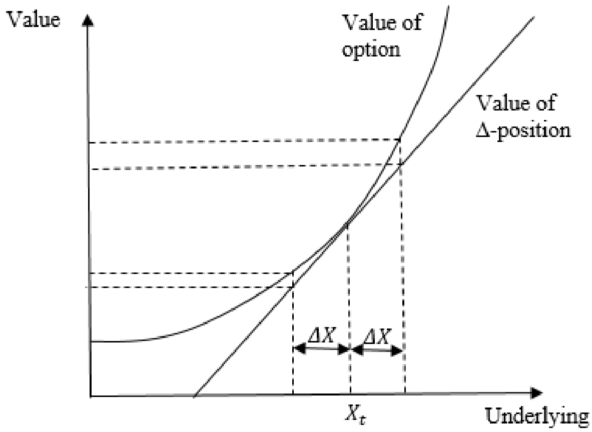

Fig. 1.2 Suppose we Δ-hedge a long position in an option with a convex pricing function. Insofar as a jump in the underlying occurs, $X_t \mapsto X_t \pm \Delta X_t$, it follows that the value of the option will exceed the value of the Δ-position. Hence, our net *P&L* benefits from such an occurrence. Obviously, the converse will be true if we hold a short position in the option.

Whilst ATM call option prices straightforwardly are obtained from the data set, the (forecasted) implied and actual volatilities require a bit of manipulation. In case of the former, we define the daily implied volatility, over the life-time of the portfolio, as the ATM implied volatility of corresponding tenor obtained at the portfolio purchasing date (the resulting volatility process is illustrated by the black curve in Figure 1.3). In case of the latter, we require a suitable volatility model fitted to historical data in order to predict the "actual" volatility process. Specifically, we define the daily actual volatility, over the life-time of the portfolio, as the conditional expectation of a volatility model which has been fitted to market data from the previous portfolio period. In this context, we observe that models with lognormal volatility dynamics generally have more empirical support than, say, Heston's model (see Gatehral and Jaisson [16] and their references). The Exponential General Autoregressive Conditional Heteroskedasticity model (EGARCH(1,1)) has proven particularly felicitous in the context of S&P 500 forecasting (see Awartani and Corradi [4]) - a result we assume applies universally for each of the 36 portfolios investigated. Thus, we hold it to be the case that daily log returns, r_t, can be modelled as $r_t = \mu + \varepsilon_t$, where μ is the mean return, and ε_t has the interpretation of a hetereoskedastic error. In particular, ε_t is construed to be the product between a white noise process, $z_t \sim N(0,1)$, and a daily standard deviation, σ_t, which obeys the relation

$$\log \sigma_t^2 = \alpha_0 + \alpha_1 \log \sigma_{t-1}^2 + \alpha_2 \left[\frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \sqrt{\frac{2}{\pi}} \right] + \alpha_3 \frac{\varepsilon_{t-1}}{\sigma_{t-1}}, \tag{1.22}$$

where $\alpha_0, \alpha_1, \alpha_2$ and α_3 are constants. The resulting volatility process is illustrated by the light grey curve in Figure 1.3.

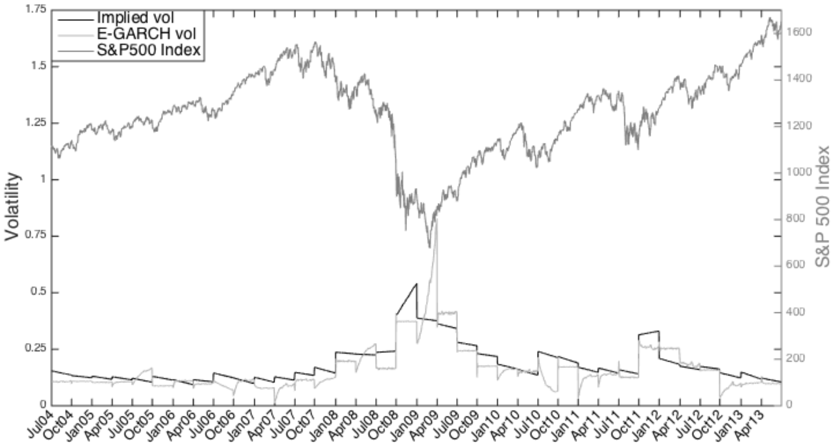

Fig. 1.3 The top **grey curve** is the S&P500 Index plotted from July 2004 to July 2013 [units on right hand axis]. The tic-dates on the time axis have deliberately been chosen to match the purchasing dates $\{t_i\}_{i=1}^{36}$ of the 36 delta-hedged portfolios under investigation (each of which is of three months' duration). The **light grey curve** is the actual (stochastic) volatility estimated from a lognormal volatility model. Specifically, every time segment between purchasing dates $[t_i, t_{i+1})$ reflects a Monte Carlo simulated forecast based upon an EGARCH(1,1) fitted to market data from the previous time segment $[t_{i-1}, t_i)$. Finally, the **black curve** is the three-month ATM implied volatility. Specifically, every time segment between purchasing dates $[t_i, t_{i+1})$ is a static forecast based upon ATM implied volatility data from the purchasing date t_i. Both volatility curves have their units on the left hand axis.

A few remarks on the estimated volatility processes are in order. First, we clearly see that volatility can change dramatically during the life-time of a portfolio. We also see that implied volatility typically is higher than actual volatility. This oft-reported result can be explained theoretically by the stochastic volatility having a market price of risk attached, see for instance Henderson et al. [15]. Finally, there is a clear negative correlation between stock returns and volatility during the financial turmoil which followed the Lehman default in September 2008. All in all, reality unsurprisingly turns out to be a bit more complicated than the set-up in Wilmott's experiment. Still and all, does its main messages carry over? To test this, we perform a hedge experiment with the following design:

- For any given portfolio, we compute the daily implied volatilities $\{\sigma_t^{\text{imp}}\}_{t=1}^{63}$ and the daily actual volatilities $\{\sigma_t^{\text{act}}\}_{t=1}^{63}$ as outlined above. We assume there are 63 trading days over a three months period (labelled by $t = 1, 2, \ldots, 63$) and let S_t, r_t and q_t denote the time t value of the index, interest rate and dividend yield.
- For each of the two hedging strategies $x \in \{\sigma^{\text{imp}}, \sigma^{\text{act}}\}$ we do the following: If $\sigma_1^{\text{act}} < \sigma_1^{\text{imp}}$ we short the call ($\gamma = -1$); otherwise, we go long the call ($\gamma = +1$). Then, we set up the delta neutral portfolio $\Pi_1 = B_1 - \gamma\Delta_1^{\text{BS}}(x_1)S_1 + \gamma C_1^{\text{BS}}(\sigma_1^{\text{imp}})$ s.t. $\Pi_1 = 0$, where $\Delta_1^{\text{BS}}(x_1)$ is the well-known Black-Scholes delta.

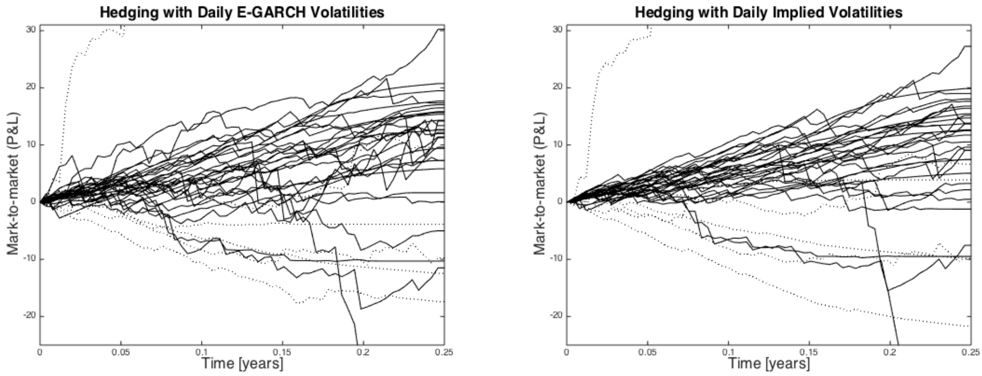

Fig. 1.4 Panels (a) (actual) and (b) (implied) show the path-for-path hedge error behaviour for the 36 non-overlapping three-month hedges. Dotted paths correspond to cases where we initially take a long position in the option.

- For $t = 2, 3, ..., 63$ we do the following: compute the time t value of the portfolio set up the previous day: $\tilde{\Pi}_t = B_{t-1}e^{r_{t-1}\Delta t} - \gamma\Delta_t^{\text{BS}}(x_t)S_t e^{q_{t-1}\Delta t} + \gamma C_t^{\text{BS}}(\sigma_t^{\text{imp}})$. The quantity $dP\&L_t = \tilde{\Pi}_t - \Pi_{t-1}$ defines the profit-&-loss accrued over the interval $[t-1, t]$. Next, we rebalance the portfolio such that it, once again, is delta-neutral, $\Pi_t = B_t - \gamma\Delta_t^{\text{BS}}(x_t)S_t + \gamma C_t^{\text{BS}}(\sigma_t^{\text{imp}})$, where B_t is chosen in accordance with the self-financing condition: $\tilde{\Pi}_t = \Pi_t$.
- Finally, at the option expiry, we compute the terminal P&L, as well as its lifetime quadratic variation, $\sum_{t=1}^{63} |dP\&L_t|^2/63$.

The 36 hedge error (or P&L) paths and the distributions of the quadratic variation of the two methods are shown in Figure 1.4. Table 1.1 reports descriptive statistics and a statistical tests of various hypotheses.

Quantity	**Mean (m)**	**Std. Dev. (sd)**	**Notes (Hypotheses Tests)**
Hedge error, actual volatility	7.7	17.3	The mean hedge error is statistically greater than zero (p-value = 1%) when we hedge with the actual vol. forecast.
Hedge error, implied volatility	7.7	15.6	The mean hedge error is statistically greater than zero (p-value = 1%) when we hedge with the implied vol. forecast. We cannot reject the hypothesis that the standard deviations $sd_{act} = sd_{imp}$ are distinct (p-value = 55%).
Quadratic var., actual volatility	1.2	2.1	
Quadratic var., implied volatility	0.81	2.0	The mean quadratic variation of the hedge error when we hedge with the actual vol. forecast is statistically less than the quadratic variation when we hedge with the implied vol. forecast (p-value = 1.4%).

Table 1.1 Summary statistics and hypothesis tests for different hedge strategies.

First, we note (top panels figure 1.4) that even though implied volatility typically is above actual volatility, this far from creates volatility arbitrage. Hedge errors for the two methods readily become negative. A primary explanation for this is the randomness of volatility. Our Δ-hedged strategy only makes us a profit if realised volatility ends up "on the right side" of initial implied volatility. And that we don't know for sure until after the hedging period is over; we have to base our decisions on forecasts; initial forecasts even, for the fundamental theorem to apply. Notice though that the averages for both hedge errors are significantly positive. This shows that there is a risk premium that can be picked up, most often by selling options and Δ-hedging them. Because the hedge is not perfect, this compensation is anticipated. The question is, is it financially significant? In theory the hedged portfolio has an initial cost of zero, so it is not obvious how to define a rate of return, but the initial option price would seem a reasonable (possibly conservative) benchmark for the collateral that would need to be posted on a hedged short call option position. From column three in Table 1.2 the average option price is \$ 49.2. Comparing this to the means ($\sim$ 7.7; remember this is over a three-month horizon) and standard deviations ($\sim$ 15.5; ditto) of the hedge errors in Table 1.1 shows that the gains are also significant in economic terms. Put differently, a crude calculation

$$\frac{4 \cdot \frac{7.7}{49.2} - 0.02}{\sqrt{4} \cdot \frac{15.5}{49.2}},$$

gives annualised Sharpe-ratios around 1.

If we look just at the terminal hedge errors, then the difference in riskiness (as measured by standard deviation) between hedging with actual and hedging with implied volatility is in no way statistically significant (the p-value for equality of variances is 55%). Also, the correlation between the terminal hedge error from the two approaches is 0.97. However, if we consider the quadratic variations as the measure of riskiness, then the picture changes. The average quadratic variation of the implied hedge error (0.81) is only two-thirds of the average quadratic variation of the actual hedge error (1.2) (a paired t-test for equality yields a p-value of 1.4%).

All in all this shows that volatility arbitrage is difficult, but the following insight from Wilmott's experiment stands: If you are in the business of hedging, then the use of implied volatility should make you sleep better at night.

1.5 Conclusion

In the world of finance, no issue is more pressing than that of hedging our risks, yet remarkably little attention has been paid to the risk brought about by the possibility that our models might be wrong. To remedy this deplorable situation, we have in this paper derived a meta-theorem that quantifies the P&L of a Δ-hedged portfolio with an erroneous volatility specification. *Meta-* to the extent that one of the constituent parameters (the real volatility) is transcendental; yet, also a theorem with some very concrete "real world" corollaries. For instance, it was shown that hedging with the implied volatility gives rise to smooth (i.e. $\mathscr{O}(dt)$) P&L-paths, whilst any other hedge volatility yields erratic (i.e. $\mathscr{O}(dW_t)$) P&L paths. In a somewhat quirkier context, the Dupire-Gyöngy-Derman-Kani formula for volatility surface calibration was shown to be a corollary.

Whilst the theorem proved in section one is more general than the versions typically found in the literature, it does not go *far enough*. Extensive empirical support has been added to the case of discontinuities in the stock price process: thus, in the Gospel of the Jump we extended the Fundamental Theorem to include compound Poisson processes, which came with the revelation that jumps unambiguously hurt you when you try to hedge short put and call option positions.

One of the most conspicuous implications of the Fundamental Theorem is undoubtedly the apparent ease with which arbitrage can be made: e.g. in the constant parameter framework of Wilmott's experiment, a free lunch is guaranteed insofar as we can establish $\max\{\sigma_r, \sigma_i\}$ (in case of the former, we go long on the option - otherwise, we short it). Studying this strategy empirically, we find that the mean P&L indeed is in the positive; nonetheless, qua a significant dispersion the profit readily turns negative: the statistical arbitrage accordingly relies on us being willing to take so some significant hits along the way. Indeed, this is without even factoring in the non-negligible role of transaction costs. On the other hand, there is strong evidence that hedging at the implied volatility does yield smoother P&L paths.

References

1. Ahmad, R. and Wilmott, P., *Which Free Lunch Would You Like Today, Sir?: Delta Hedging, Volatility Arbitrage and Optimal Portfolios*, *Wilmott Magazine*, 2005, November, pp. 64-79.
2. Andreasen, J. *Derivatives - The View from the Trenches*, 2003. Presentation at the University of Copenhagen. http://www.math.ku.dk/ rolf/jandreasen.pdf.
3. Andreasen, J. and Huge, B. *Volatility Interpolation*, *Risk Magazine*, 2011, March, pp. 76-79.
4. Awartani, B. and Corradi, V., *Predicting the Volatility of the S&P-500 Stock Index via GARCH Models: The Role of Asymmetries*, *International Journal of Forecasting*, 2005, Vol. 21, pp. 167-183.
5. Björk, T. *Arbitrage Theory in Continuous Time*, Oxford University Press, 2009. 3rd Edition.
6. Black, F. and Scholes, M., *The Pricing of Options and Corporate Liabilities*, *The Journal of Political Economy*, 1973, Vol. 81, pp. 637-654.
7. Carr, P., *FAQ's in Option Pricing Theory*, 2002. Working paper, available at http://www.math.nyu.edu/research/carrp/papers/pdf/faq2.pdf.
8. Carr, P. and Jarrow, R., *The Stop-Loss Start-Gain Paradox and Option Valuation: A New Decomposition into Intrinsic and Time Value*, *The Review of Financial Studies*, 1990, Vol. 3, pp. 469-492.
9. Cont and Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC Financial Mathematics Series. 2004.
10. Davis, M.H.A. *Black Scholes Formula*, *Encyclopaedia of Quantitative Finance*, 2010, Vol. 1, pp. 201-207.
11. Derman, E. and Kani, I., *Stochastic Implied Trees : Arbitrage Pricing with Stochastic Term and Strike Structure of Volatility*, *International Journal of Theoretical and Applied Finance*, 1998, pp. 7-22.
12. Dupire, B., *Pricing with a Smile*, *Risk Magazine*, 1994, January, pp. 18-20.
13. El Karoui, N: and Jeanblanc-Picque, M. and Shreve, S. *Robustness of the Black and Scholes Formula*, *Mathematical Finance*, 1998, Vol. 8, pp. 93-126.
14. Haug, E. and Taleb. N., *Option Traders Use (Very) Sophisticated Heuristics, Never the Black-Scholes-Merton Formula*, *Journal of Economic Behaviour and Organization*, 2011, Vol. 77, pp. 97-106.
15. Henderson, V. and Hobson, D. and Howison, S. and Kluge, T., *A Comparison of Option Prices Under Different Pricing Measures in a Stochastic Volatility Model with Correlation*, *Review of Derivatives Research*, 2005, Vol. 8, pp. 5-25.
16. Gatheral, J. and Jaisson, T. and Rosenbaum, M., *Volatility is Rough*, 2014. Working paper. Available at http://arxiv.org/abs/1410.3394.
17. Gibson, R. and Lhabitant, F. and Pistre, N. and Talay, D., *Interest Rate Model Risk: What are we Talking About?*, *Journal of Risk*, 1999, Vol. 1, pp. 37-62.
18. Guyon, J. and Henry-Labordère, P., *Being Particular about Calibration*, *Risk Magazine*, 2012, January, pp. 92-97.
19. Hanson, *Applied Stochastic Processes and Control for Jump-Diffusions: Modeling, Analysis, and Computation (Advances in Design and Control)*, 2007.
20. Henrard, M., *Parameter Risk in the Black and Scholes Model*, 2001. Working paper, http://128.118.178.162/eps/ri/papers/0310/0310002.pdf.
21. Mahayni, A. and Schlögl, E. and Schlögl, L., *Robustness of Gaussian Hedges and the Hedging of Fixed Income Derivatives*, 2001. Working paper. http://www.qfrc.uts.edu.au/research/research_papers/rp19.pdf
22. Rasmussen, N., *Hedging with a Misspecified Model*, 2001. Working paper at SSRN. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=260754.
23. Runggaldier, W. *Jump Diffusion Models*, in *Handbook of Heavy Tailed Distributions in Finance*, 2003 Elsevier Science B.V.
24. Privault, N., *Notes on Stochastic Finance*, 2013. Lecture Notes FE6516, http://www.ntu.edu.sg/home/nprivault/indext.html.
25. Runggaldier, *Jump Diffusion Models*, in *Handbook of Heavy Tailed Distributions in Finance*, 2003 Elsevier Science B.V.
26. Shreve, S., *Stochastic Calculus for Finance II*, 2008. Springer.
27. Taleb, N., *The Black Swan - The Impact of the Highly Improbable*, Penguin Books, Revised Edition 2010.
28. Wittgenstein, L. *Tractatus Logico-Philosophicus*, 1922.

Appendix: Multi-dimensional Jumps

In this section we establish Girsanov's Theorem and a pricing PDE for multi-dimensional jump-diffusion models. The equivalent results for 1-dimensional models are ubiquitous - see for instance Cont and Tankov [16], Privault [19] or Runggaldier [43].

The Radon-Nikodym Derivative

A Generalised Girsanov Theorem for Jump-Diffusion Processes. Let $\{\boldsymbol{W}_t\}_{t\in[0,T]}$ be a d_w-dimensional vector of independent Wiener processes on the filtered probability space $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}_{t\in[0,T]})$. One the same space, let $\{\boldsymbol{Y}_t\}_{t\in[0,T]}$ be a d_y-dimensional vector of independent compound Poisson processes, the i^{th} component of which is

$$Y_t^i = \sum_{k=1}^{N_t^i} Z_k^i,$$

where $N_t^i \sim \text{Pois}(\lambda_i t)$, $\lambda_i > 0$ is an intensity parameter, and $\{Z_k^i\}_{k\in\mathbb{N}}$ is a sequence of i.i.d. random variables with jump distribution $d\nu_i(z)$. Finally, let $\{\phi_t\}_{t\in[0,T]}$ be a d_w-dimensional Girsanov kernel (some bounded, adapted process), then the processes

$$\left\{\boldsymbol{W}_t^{\mathbb{Q}} := \boldsymbol{W}_t - \int_0^t \phi_s ds\right\}_{t\in[0,T]}, \quad \text{and} \quad \left\{\tilde{\boldsymbol{Y}}_t^{\mathbb{Q}} := \boldsymbol{Y}_t - \boldsymbol{\lambda}^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]t\right\}_{t\in[0,T]}$$

are martingales under the probability measure $\mathbb{Q}$ defined as

$$d\mathbb{Q}_{\phi,\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}} = \mathscr{E}(\phi \star \boldsymbol{W})(T) \cdot e^{-\sum_{i=1}^n (\lambda_i^{\mathbb{Q}} - \lambda_i)T} \prod_{i=1}^n \prod_{k=1}^{N_t^i} \frac{\lambda_i^{\mathbb{Q}} d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)} d\mathbb{P}$$

where $\mathscr{E}(\phi \star \boldsymbol{W})(T) = e^{\int_0^T \phi_s \bullet d\boldsymbol{W}_s - \frac{1}{2}\int_0^T |\phi_s|^2 ds}$ is the Doleans exponential with respect to $\boldsymbol{W}_t$, and we have defined $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_{d_y})^\intercal$, and $\boldsymbol{\nu} = (\nu_1, \nu_2, ..., \nu_{d_y})^\intercal$.

Proof. The diffusion part is well known from Girsanov's theorem and will not be treated here. Instead we will show that for any bounded measurable function $f : \mathbb{R}^{d_y} \mapsto \mathbb{R}$, the following equivalence obtains

$$\mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[f(\boldsymbol{Y}_T)] = \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_T)\frac{d\mathbb{Q}}{d\mathbb{P}}\right],$$

for the defined measure $\mathbb{Q}$. To this end, define $\boldsymbol{Y}_t^h$ as the vector $\boldsymbol{Y}_t$ with the upper limit of the summation, N_t^i, replaced by some fixed number $h^i \in \mathbb{N}_0$ for all i. Then the RHS can be written as

$$
\begin{aligned}
& e^{-\sum_{i=1}^{d_y}(\lambda_i^{\mathbb{Q}}-\lambda_i)T}\mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t)\prod_{i=1}^{d_y}\prod_{k=1}^{N_t^i}\frac{\lambda_i^{\mathbb{Q}}d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)}\right] \\
&= e^{-\sum_{i=1}^{d_y}(\lambda_i^{\mathbb{Q}}-\lambda_i)T}\sum_{h^1=0}^{\infty}\cdots\sum_{h^{d_y}=0}^{\infty}\mathbb{P}\left(\bigcap_{i=1}^{d_y}\{N_T^i=h^i\}\right)\cdot \\
&\quad \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t)\prod_{i=1}^{d_y}\prod_{k=1}^{N_t^i}\frac{\lambda_i^{\mathbb{Q}}d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)}\middle|\bigcap_{i=1}^{d_y}\{N_T^i=h^i\}\right] \\
&= e^{-\sum_{i=1}^{d_y}(\lambda_i^{\mathbb{Q}}-\lambda_i)T}\sum_{h^1=0}^{\infty}\cdots\sum_{h^{d_y}=0}^{\infty}\prod_{i=1}^{d_y}\mathbb{P}\left(N_T^i=h^i\right)\mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i}\frac{\lambda_i^{\mathbb{Q}}d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)}\right] \\
&= e^{-\sum_{i=1}^{d_y}(\lambda_i^{\mathbb{Q}}-\lambda_i)T}\sum_{h^1=0}^{\infty}\cdots\sum_{h^{d_y}=0}^{\infty}\prod_{i=1}^{d_y}\frac{e^{-\lambda_i T}(\lambda_i T)^{h_i}}{k^i!}\mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i}\frac{\lambda_i^{\mathbb{Q}}d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)}\right] \\
&= \sum_{h^1=0}^{\infty}\cdots\sum_{h^{n}=0}^{\infty}\prod_{i=1}^{d_y}\frac{e^{-\lambda_i^{\mathbb{Q}}T}(\lambda_i^{\mathbb{Q}}T)^{h_i}}{k^i!}\mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i}\frac{d\nu_i^{\mathbb{Q}}(Z_k^i)}{d\nu_i(Z_k^i)}\right] \\
&= \sum_{h^1=0}^{\infty}\cdots\sum_{h^{d_y}=0}^{\infty}\prod_{i=1}^{d_y}\frac{e^{-\lambda_i^{\mathbb{Q}}T}(\lambda_i^{\mathbb{Q}}T)^{h_i}}{k^i!}\int_{\mathbb{R}^{h^1}}\cdots\int_{\mathbb{R}^{h^{d_y}}}f(y_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i}\frac{d\nu_i^{\mathbb{Q}}(z_k^i)}{d\nu_i(z_k^i)}d\nu_i(z_k^i) \\
&= \sum_{h^1=0}^{\infty}\cdots\sum_{h^{d_y}=0}^{\infty}\prod_{i=1}^{d_y}\frac{e^{-\lambda_i^{\mathbb{Q}}T}(\lambda_i^{\mathbb{Q}}T)^{h_i}}{k^i!}\int_{\mathbb{R}^{h^1}}\cdots\int_{\mathbb{R}^{h^{d_y}}}f(y_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i}d\nu_i^{\mathbb{Q}}(z_k^i) \\
&= \sum_{h^1=0}^{\infty}\cdots\sum_{h^{n}=0}^{\infty}\prod_{i=1}^{d_y}\mathbb{Q}\left(N_T^i=h^i\right)\mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}\left[f(\boldsymbol{Y}_t^h)\right] \\
&= \sum_{h^1=0}^{\infty}\cdots\sum_{h^{d_y}=0}^{\infty}\mathbb{Q}\left(\bigcap_{i=1}^{d_y}\{N_T^i=h^i\}\right)\mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}\left[f(\boldsymbol{Y}_t)\middle|\bigcap_{i=1}^{d_y}\{N_T^i=h^i\}\right]
\end{aligned}
$$

which by the law of total probability corresponds to the LHS, $\mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[f(\boldsymbol{Y}_T)]$ as desired. To show independence of increments under $\mathbb{Q}$, let $\xi_s = d\mathbb{Q}/d\mathbb{P}(s)$, and let f and g be two bounded measurable functions. Suppose $s < t \leq T$ then

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[f(\boldsymbol{Y}_s)g(\boldsymbol{Y}_t-\boldsymbol{Y}_s)] &= \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[f(\boldsymbol{Y}_s)g(\boldsymbol{Y}_t-\boldsymbol{Y}_s)\xi_t] \\
&= \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[f(\boldsymbol{Y}_s)\xi_s]\mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[g(\boldsymbol{Y}_t-\boldsymbol{Y}_s)\xi_t/\xi_s] \\
&= \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[f(\boldsymbol{Y}_s)\xi_s]\mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[g(\boldsymbol{Y}_t-\boldsymbol{Y}_s)\xi_t] \\
&= \mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[f(\boldsymbol{Y}_s)]\cdot\mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[g(\boldsymbol{Y}_t-\boldsymbol{Y}_s)].
\end{aligned}
$$

$\square$

Measure Changes in Jump-Diffusion dynamics

To appreciate the gravity of this result, consider the jump-diffusion dynamics of the n-dimensional stock price process

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_{t-}}[\boldsymbol{\mu}(t,\boldsymbol{X}_t)dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t] + \mathbf{D}_{\boldsymbol{X}_t}\theta(t,\boldsymbol{X}_{t-})d\boldsymbol{Y}_t \tag{1.23}$$

where $\mathbf{D}_{\boldsymbol{X}_{t-}} = \mathrm{diag}(X_{1,t-}, X_{2,t-}, ..., X_{n,t-})$, $\boldsymbol{\mu} : [0,\infty) \times \mathbb{R}^n \mapsto \mathbb{R}^n$, $\boldsymbol{\sigma} : [0,\infty) \times \mathbb{R}^n \mapsto \mathbb{R}^{n\times d_w}$ and $\theta : [0,\infty) \times \mathbb{R}^n \mapsto \mathbb{R}^{n\times d_y}$. For the purposes of no-arbitrage pricing we want to determine the measure $\mathbb{Q}$ such that the discounted process

$$\boldsymbol{X}_t^* := e^{-\int_0^t r_u du}\boldsymbol{X}_t,$$

is a martingale. From Ito's lemma, it can readily be deduced that

$$\begin{aligned} d\boldsymbol{X}_t^* = &\mathbf{D}^*_{\boldsymbol{X}_t}[(\boldsymbol{\mu}(t,\boldsymbol{X}_t) - r_t\boldsymbol{\iota} + \boldsymbol{\theta}(t,\boldsymbol{X}_t)\boldsymbol{\lambda}\circ\mathbb{E}_{\boldsymbol{\nu}}[\boldsymbol{Z}_1])dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t] \\ &+ \mathbf{D}^*_{\boldsymbol{X}_{t-}}\theta(t,\boldsymbol{X}_{t-})(d\boldsymbol{Y}_t - \boldsymbol{\lambda}\circ\mathbb{E}_{\boldsymbol{\nu}}[\boldsymbol{Z}_1]dt), \end{aligned}$$

where we have added and subtracted $\boldsymbol{\theta}(t,\boldsymbol{X}_t)\boldsymbol{\lambda}\circ\mathbb{E}_{\boldsymbol{\nu}}[\boldsymbol{Z}_1]dt$, where $\boldsymbol{Z}_1 := (Z_1^1, Z_1^2, ..., Z_1^{d_y})^\intercal$. Using the measure transformation in the theorem above this transforms to

$$\begin{aligned} d\boldsymbol{X}_t^* = &\mathbf{D}^*_{\boldsymbol{X}_t}[(\boldsymbol{\mu}(t,\boldsymbol{X}_t) - r_t\boldsymbol{\iota} + \theta(t,\boldsymbol{X}_t)\boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1] + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)\boldsymbol{\phi}_t)dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}] \\ &+ \mathbf{D}^*_{\boldsymbol{X}_{t-}}\theta(t,\boldsymbol{X}_{t-})d\tilde{\boldsymbol{Y}}_t^{\mathbb{Q}}. \end{aligned}$$

Hence, $\boldsymbol{X}_t^*$ is a $\mathbb{Q}$-martingale iff the tuple $\langle\boldsymbol{\phi}_t, \boldsymbol{\lambda}^{\mathbb{Q}}, \boldsymbol{\nu}^{\mathbb{Q}}, \mathbb{Q}\rangle$ is chosen such that

$$\boldsymbol{\mu}(t,\boldsymbol{X}_t) + \boldsymbol{\theta}(t,\boldsymbol{X}_t)\boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1] + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)\boldsymbol{\phi}_t = r_t\boldsymbol{\iota}, \tag{1.24}$$

almost everywhere. Needless to say, the infinite number of tuples which satisfies the no-arbitrage condition (1.24) ruins our chances of establishing *unique* prices for financial derivatives depending on the underlying jump diffusion dynamics. This is obvious by recalling that the discounted price process of V also should be a martingale, whence:

$$V(t,\boldsymbol{X}_t) = \mathbb{E}_{t,\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}\left[e^{-\int_t^T r_u du}g(\boldsymbol{X}_T)\right], \tag{1.25}$$

where $g(\boldsymbol{X}_T)$ is the terminal pay-off.

Example: The Merton Model.

Let S_t be the price process of a stock which obeys the following dynamics under the $\mathbb{P}$ measure, $S_t = S_0 e^{\mu t + \sigma W_t + Y_t}$, or in differential form

$$dS_t = S_t[(\mu + \tfrac{1}{2}\sigma^2)dt + \sigma dW_t] + S_{t-}(e^{Z_t} - 1)dN_t.$$

Transforming to the measure[19] $\mathbb{Q}_{0,\lambda^{\mathbb{Q}},\nu^{\mathbb{Q}}}$ we have $W_t = W_t^{\mathbb{Q}}$ and

$$dS_t = S_t[(\mu + \tfrac{1}{2}\sigma^2 + \lambda^{\mathbb{Q}}\mathbb{E}_{\nu^{\mathbb{Q}}}[e^{Z_1} - 1])dt + \sigma dW_t^{\mathbb{Q}}] \\ + S_{t-}[(e^{Z_t} - 1)dN_t - \lambda^{\mathbb{Q}}\mathbb{E}_{\nu^{\mathbb{Q}}}[e^{Z_1} - 1]dt],$$

which implies the no-arbitrage condition

$$r = \mu + \tfrac{1}{2}\sigma^2 + \lambda^{\mathbb{Q}}\mathbb{E}_{\nu^{\mathbb{Q}}}[e^{Z_1} - 1],$$

and, of course, $S_t = S_0 \exp\{\mu t + \sigma W_t^{\mathbb{Q}} + Y_t\}$. Assume $\{Z_k\}_{k\in\mathbb{N}}$ is a sequence of independent $N(\delta, \zeta^2)$ distributed random variables under $\mathbb{Q}_{0,\lambda^{\mathbb{Q}},\nu^{\mathbb{Q}}}$. Then by the law of total expectation, the value of a derivative paying out $\Theta(S_T)$ at maturity is

$$V_t = e^{-r\tau - \lambda^{\mathbb{Q}}\tau} \sum_{k=0}^{\infty} \frac{(\lambda^{\mathbb{Q}}\tau)^k}{k!} \mathbb{E}_{t,\lambda^{\mathbb{Q}},\nu^{\mathbb{Q}}}[\Theta(S_t e^{\mu\tau + k\delta + X})],$$

where $\tau := T - t$ and

$$X := \sigma(W_T - W_t) + \sum_{i=1}^{k}(Z_i - \delta) \sim N(0, \sigma^2\tau + k\zeta^2).$$

In particular, if the option is a vanilla call, $\Phi(S_T) = (S_T - K)^+$, we find that

$$\begin{aligned} V_t &= e^{-r\tau - \lambda^{\mathbb{Q}}\tau} \sum_{k=0}^{\infty} \frac{(\lambda^{\mathbb{Q}}\tau)^k}{k!} \mathbb{E}_{t,\lambda^{\mathbb{Q}},\nu^{\mathbb{Q}}}\left[(S_t e^{(r - \frac{1}{2}\sigma^2 - \lambda^{\mathbb{Q}}(e^{\delta + \frac{1}{2}\zeta^2} - 1)\tau + k\delta + X} - K)^+\right] \\ &= e^{-r\tau - \lambda^{\mathbb{Q}}\tau} \sum_{k=0}^{\infty} \frac{(\lambda^{\mathbb{Q}}(T-t))^k}{k!} C^{\mathrm{BS}}\left(S_t e^{\frac{1}{2}k\zeta^2 - \lambda^{\mathbb{Q}}(e^{\delta + \frac{1}{2}\zeta^2} - 1)\tau + k\delta}, K, \sigma^2 + \tfrac{k\zeta^2}{\tau}, r, \tau\right), \end{aligned}$$

where $C^{\mathrm{BS}} := C^{\mathrm{BS}}(S_t, K, \sigma^2, r, \tau)$ is the Black-Scholes price. The first line follows from the no-arbitrage condition alongside

$$\mathbb{E}_{\nu^{\mathbb{Q}}}[e^{Z_1}] = e^{\delta + \frac{1}{2}\zeta^2}.$$

The last line from the relation

$$e^{-r\tau}\mathbb{E}[(xe^{X - \frac{1}{2}v^2 + r\tau} - K)^+] = C^{\mathrm{BS}}(x, K, v^2/\tau, r, \tau),$$

where $v := \sigma^2\tau + k\zeta^2$. Hence the value of a call option under a jump-diffusion dynamics can be written as an infinite superposition of Black Scholes call prices.

[19] Strictly speaking, this is not Merton's choice of measure. He assumes the jump intensity and distribution to be invariant under $\mathbb{P} \mapsto \mathbb{Q}$, while the Wiener process changes.

PIDE Methods

The Partial Integro-Differential Pricing Equation (PIDE) Consider a jump diffusion dynamics of the form (1.23), and let $V_t = V(t, \boldsymbol{X}_t)$ be a derivative the value of which is contingent upon it. Let $\langle \phi_t, \boldsymbol{\lambda}^{\mathbb{Q}}, \boldsymbol{\nu}^{\mathbb{Q}}, \mathbb{Q}\rangle$ be a tuple such that the no-arbitrage condition (1.24) is satisfied. Then

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_t}[r_t\boldsymbol{\iota}dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}] + \mathbf{D}_{\boldsymbol{X}_{t-}}\theta(t,\boldsymbol{X}_{t-})[d\boldsymbol{Y}_t - \boldsymbol{\lambda}^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]dt]],$$

and $V_t = V(t, \boldsymbol{x})$ satisfies the PIDE

$$\begin{aligned} r_t V_t = {} & \partial_t V_t + r_t \boldsymbol{x} \bullet \nabla_{\boldsymbol{x}} V_t + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^{\intercal}(t,\boldsymbol{x})\mathbf{D}_{\boldsymbol{x}}\nabla^2_{\boldsymbol{x}\boldsymbol{x}}V_t\mathbf{D}_{\boldsymbol{x}}\boldsymbol{\sigma}(t,\boldsymbol{x})] \\ & + \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\textstyle\sum_{i=1}^{d_y} \lambda_i^{\mathbb{Q}}\{V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{x})Z_1^i)) - V(t,\boldsymbol{x})\} \\ & - (\mathbf{D}_{\boldsymbol{x}}\theta(t,\boldsymbol{x})\boldsymbol{\lambda}^{\mathbb{Q}} \circ \boldsymbol{Z}_1) \bullet \nabla_{\boldsymbol{x}} V(t,\boldsymbol{x})], \end{aligned}$$

where $\nabla_{\boldsymbol{x}}$ is the gradient operator, $\nabla^2_{\boldsymbol{x}\boldsymbol{x}}$ is the Hessian operator, $\boldsymbol{\iota} := (1,1,...,1)^{\intercal} \in \mathbb{R}^n$, and $\boldsymbol{\theta}_{:,i}(t,\boldsymbol{x})$ denotes the i^{th} column of the matrix $\theta(t,\boldsymbol{x})$. Particularly, when $n = d_y$ and $\theta = \mathbb{I}$ is the identity matrix then

$$\begin{aligned} r_t V_t = {} & \partial_t V_t + r_t \boldsymbol{x} \bullet \nabla_{\boldsymbol{x}} V_t + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^{\intercal}(t,\boldsymbol{x})\mathbf{D}_{\boldsymbol{x}}\nabla^2_{\boldsymbol{x}\boldsymbol{x}}V_t\mathbf{D}_{\boldsymbol{x}}\boldsymbol{\sigma}(t,\boldsymbol{x})] \\ & + \textstyle\sum_{i=1}^{n} \lambda_i^{\mathbb{Q}}\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\hat{\mathbf{e}}_i Z_1^i)) - V(t,\boldsymbol{x}) - x_i Z_1^i \partial_{x_i} V(t,\boldsymbol{x})], \end{aligned}$$

where $\hat{\mathbf{e}}_i$ is a unit vector in the i^{th} direction.

Proof. Suppose a jump occurs in the i^{th} component of the compound Poisson process $\boldsymbol{Y}_t$: $Y_t^i = Y_{t-}^i + Z_t^i$. From the governing dynamics (1.23), this means that the stock price process jumps by $\Delta\boldsymbol{X}_t = \boldsymbol{X}_{t-} \circ (\boldsymbol{\iota} + \theta_{:,i}(t,\boldsymbol{X}_{t-})Z_t^i)$. Defining the continuous SDE

$$d\boldsymbol{X}_t^{\text{cont.}} := \mathbf{D}_{\boldsymbol{X}_t}[r_t\boldsymbol{\iota}dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}] - \mathbf{D}_{\boldsymbol{X}_{t-}}\theta(t,\boldsymbol{X}_{t-})\boldsymbol{\lambda}^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]dt],$$

we find by Ito's lemma,

$$\begin{aligned} dV(t,\boldsymbol{X}_t) &= \partial_t V(t,\boldsymbol{X}_t)dt + \nabla_{\boldsymbol{x}} V(t,\boldsymbol{X}_t) \bullet d\boldsymbol{X}_t^{\text{cont.}} \\ &\quad + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^{\intercal}(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{x}\boldsymbol{x}}V(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)]dt \\ &\quad + \textstyle\sum_{i=1}^{d_y}(V(t,\boldsymbol{X}_{t-}+\Delta\boldsymbol{X}_t) - V(t,\boldsymbol{X}_{t-}))dN_t^i \\ &= \partial_t V(t,\boldsymbol{X}_t)dt + \nabla_{\boldsymbol{x}} V(t,\boldsymbol{X}_t) \bullet (\mathbf{D}_{\boldsymbol{X}_t} r_t \boldsymbol{\iota})dt \\ &\quad + \nabla_{\boldsymbol{x}} V(t,\boldsymbol{X}_t) \bullet (\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}) \\ &\quad - \nabla_{\boldsymbol{x}} V(t,\boldsymbol{X}_t) \bullet (\mathbf{D}_{\boldsymbol{X}_{t-}}\theta(t,\boldsymbol{X}_{t-})\boldsymbol{\lambda}^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1])dt \\ &\quad + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^{\intercal}(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{x}\boldsymbol{x}}V(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)]dt \end{aligned}$$

$$
\begin{aligned}
&+\textstyle\sum_{i=1}^{d_y}(V(t,\boldsymbol{X}_{t-}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{X}_{t-})Z^i_{N^i_t}))-V(t,\boldsymbol{X}_{t-}))dN^i_t\\
=\ &\partial_t V(t,\boldsymbol{X}_t)dt+\nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_t}r_t\boldsymbol{\iota})dt\\
&-\nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_{t-}}\theta(t,\boldsymbol{X}_{t-})\boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1])dt\\
&+\tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^\intercal(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{x}\boldsymbol{x}}V(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)]dt\\
&+\textstyle\sum_{i=1}^{d_y}\lambda^{\mathbb{Q}}_i\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}\left[(V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{x})Z^i_1))-V(t,\boldsymbol{x}))\right]_{\boldsymbol{x}=\boldsymbol{X}_{t-}}dt\\
&+\nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}^{\mathbb{Q}}_t)\\
&+\textstyle\sum_{i=1}^{d_y}\big\{(V(t,\boldsymbol{X}_{t-}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{X}_{t-})Z^i_{N^i_t}))-V(t,\boldsymbol{X}_{t-}))dN^i_t\\
&-\lambda^{\mathbb{Q}}_i\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}\left[(V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{x})Z^i_1))-V(t,\boldsymbol{x}))\right]_{\boldsymbol{x}=\boldsymbol{X}_{t-}}dt\big\}.
\end{aligned}
$$

Under $\mathbb{Q}$, the expectations of the diffusion term and the *compensated* jump terms (the last three lines) vanish. Furthermore, since

$$V^*_t := e^{-\int_0^t r_u du}V(t,\boldsymbol{X}_t),$$

is a $\mathbb{Q}$ martingale; dV^*_t should be driftless. These facts jointly imply that

$$
\begin{aligned}
&-r_tV(t,\boldsymbol{X}_t)+\partial_t V(t,\boldsymbol{X}_t)+r_t\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\iota}\bullet\nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\\
&-\nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_{t-}}\theta(t,\boldsymbol{X}_{t-})\boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1])\\
&+\tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^\intercal(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{x}\boldsymbol{x}}V(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)]\\
&+\textstyle\sum_{i=1}^{d_y}\lambda^{\mathbb{Q}}_i\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}\left[(V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{x})Z^i_1))-V(t,\boldsymbol{x}))\right]_{\boldsymbol{x}=\boldsymbol{X}_{t-}}=0,
\end{aligned}
$$

which essentially is what we wanted to show. □

Appendix B: Data

Contract	ATM strike	Option price	P&L$_T^{\text{actual}}$	P&L$_T^{\text{implied}}$	Q.V.$^{\text{actual}}$	Q.V.$^{\text{implied}}$
07-Jul-2004	1118.3	36.5852	12.2045	15.1591	0.5269	0.2615
05-Oct-2004	1134.5	33.0392	5.8372	5.0520	0.1683	0.1386
05-Jan-2005	1183.7	34.7050	11.4080	13.6705	0.1975	0.1759
06-Apr-2005	1184.1	34.9985	7.3072	9.0917	0.3162	0.1693
06-Jul-2005	1194.9	34.4864	11.9818	10.5282	0.2974	0.0894
04-Oct-2005	1214.5	37.8141	7.2779	7.4261	0.5384	0.1670
06-Jan-2006	1285.4	37.1621	12.6952	12.4934	0.1539	0.1406
07-Apr-2006	1295.5	38.2703	0.0765	0.5022	0.2827	0.2444
07-Jul-2006	1265.5	45.5356	15.3714	13.7452	0.3655	0.1974
05-Oct-2006	1353.2	42.7682	12.6179	12.6400	0.0904	0.0945
08-Jan-2007	1412.8	45.4682	-5.0096	2.1741	2.4476	1.0569
09-Apr-2007	1444.6	47.0689	19.4885	7.4564	0.7699	0.0865
09-Jul-2007	1531.8	55.8378	-11.4976	-7.5524	1.8603	1.1396
05-Oct-2007	1557.6	63.1625	1.6451	-1.2115	1.2330	0.4542
09-Jan-2008	1409.1	74.2874	9.6117	9.6158	1.1975	0.6555
09-Apr-2008	1354.5	66.2276	17.3617	19.0049	0.8019	0.6270
09-Jul-2008	1244.7	62.8179	-56.9636	-47.0345	8.0872	10.4193
07-Oct-2008	996.2	83.8510	55.3847	51.8900	9.7721	6.4129
09-Jan-2009	890.3	69.9489	14.1892	3.2637	3.0947	0.4083
10-Apr-2009	856.6	62.9702	30.2400	27.2551	0.5701	0.4336
09-Jul-2009	882.7	49.8464	-12.4499	-9.8467	0.1245	0.1039
07-Oct-2009	1057.6	49.0640	17.0496	18.0507	0.2944	0.2135
07-Jan-2010	1141.7	42.2410	16.4989	16.4106	0.2595	0.1990
09-Apr-2010	1194.4	36.6784	-10.3121	-9.5031	0.5463	0.5578
09-Jul-2010	1078.0	52.2001	15.6833	17.6455	3.0501	0.3326
07-Oct-2010	1158.1	50.6050	20.7394	19.8607	0.1926	0.2166
06-Jan-2011	1273.8	43.6970	9.4015	11.7384	0.3762	0.2400
07-Apr-2011	1333.5	44.7866	13.3942	13.8116	0.3490	0.3055
08-Jul-2011	1343.8	43.0900	-3.8722	3.8883	0.1692	0.3196
06-Oct-2011	1165.0	73.4417	14.3245	16.8015	0.8601	0.7112
06-Jan-2012	1277.8	53.9770	-17.4158	-21.6739	0.3472	0.1853
09-Apr-2012	1382.2	48.9735	-9.7517	-9.9641	0.4760	0.4018
09-Jul-2012	1352.5	47.5814	15.8417	15.4475	0.3181	0.3184
05-Oct-2012	1460.9	42.9608	11.2648	9.0422	3.0925	0.8156
09-Jan-2013	1461.0	43.7355	17.6935	14.7094	0.2747	0.1360
11-Apr-2013	1593.4	39.2535	9.4000	6.6261	1.0037	0.7546

Table 1.2 The first column lists the purchasing dates of the 36 contracts. Column two shows the ATM strikes at which the contracts are purchased and column three show the prices at which this happens. The fourth column gives the terminal P&L for each contract, when the hedge is performed with an "actual" (EGARCH(1,1)) volatility forecast. Column five likewise, but when the hedge is with the implied volatilities. Finally, columns six and seven give the quadratic variation, defined as $\sum_{i=1}^{N} |dP\&L_i|^2/N$, where $N = 63$ is the number of trading days, for the entire actual and implied paths respectively.

Part II
Optimal Asset Allocation

Chapter 2
Numerical Stochastic Control Theory with Applications in Finance
The Markov Chain Approximation Method in a Finite-Horizon Mertonian Portfolio Optimisation Context

Simon Ellersgaard

Abstract Analytic solutions to HJB equation in mathematical finance are relatively hard to come by, which stresses the need for numerical procedures. In this paper we provide a self-contained exposition of the finite-horizon Markov chain approximation method as championed by Kushner and Dupuis. Furthermore, we provide full details as to how well the algorithm fares when we deploy it in the context of Merton type optimisation problems. Assorted issues relating to implementation and numerical accuracy are thoroughly reviewed, including multidimensionality and the positive probability requirement, the question of boundary conditions, and the choice of parametric values.

Key words: Numerical Stochastic Control, Markov Chain Approximations, Merton's Portfolio Problem, Labour Income

---

The Department of Mathematical Sciences
Universitetsparken 5, 2100 Copenhagen Ø, Denmark
e-mail: ellersgaard@math.ku.dk.

2.1 Introduction

It is well known to anyone who has dabbled in continuous time continuous state stochastic financial control problems that searching for closed form solutions is an onerous and all-too-often fruitless endeavour. Given the far-reaching scope of the field, from classical portfolio optimisation a la Merton [12], [35], to modern day market micro-structure as in Almgren [3], and Stoikov et al. [6], it is quite clear that numerical control procedures are called for. Indeed, a plurality of such methods already prevails, from hands-on discretisations of the Hamilton-Jacobi-Bellman equation (see e.g. Forsyth et al. [24], [27]), to the employ of sophisticated forward-backward SDE techniques (see e.g. Ludwig et al. [17]). Most prominently, perhaps, stands Kushner and Dupuis' monograph on the so-called *Markov chain approximation method* [31], which although predominantly concerned with control-independent diffusions, straight-forwardly can be generalised in this direction, [16]. This paper pertains to this latter method, the central premise of which is to substitute the continuous time continuous state controlled state process by a discrete time discrete state Markov chain. As shown in [31], by choosing the transition probabilities of the Markov chain based on standard finite difference discretisations of the governing PDE, convergence can be established.

The purpose of this paper is two-fold: first, it serves as an exegesis of the Markov chain approximation method for anyone who seeks a swift and comparatively unconvoluted theoretical overview of the area. To this end, even numbered sections aim to equip the reader with the fundamental theoretical tools needed in order to implement numerical control problems in simple environments where uncertainty is driven by a Brownian motion. For the reader well-versed in the Markov chain approximation method, these sections can safely be skimmed if not downright skipped. The second, and arguably more intriguing part of this paper, is the odd numbered sections, which provide a detailed account of just how well the Markov chain approximation method fares when we deploy it in a Merton type portfolio optimisation context. As it will soon become clear, actual implementations require a non-negligible amount of *Fingerspitzengefühl*, the key lessons of which are not communicated by Kushner and Dupuis. For example, inexact boundary conditions turn out to have a corrosive effect on the accuracy of the numerical controls for a rather large region into the finite difference grid. However, we show that if an exact relationship between adjacent grid nodes can be established, this problem altogether dissipates. Another question concerns the sanctity of one of the central requirements of the algorithm: *positive transition probabilities*. Given the prevalence of multi-dimensional problems for which negative probabilities are hard to avoid, it is only reasonable to test how well the Markov chain approximation method fares in this domain. Our numerical studies suggest that negative probability schemes indeed have something to offer, although one must tread carefully.

Admittedly, this is not the first paper to deal the numerical implementation of Merton type problems. Clear examples include Fitzpatrick et al. [10] and Munk [22], [23]. Nevertheless, these papers all deal with the infinite horizon single state process case, and only focus on the implicit implementation procedure. We, on the other hand, focus on finite horizon investment problems both from an explicit and implicit perspective, with generalisations to higher spatial dimensions.

2.2 The Finite Horizon Stochastic Control Problem

We set out by restricting our attention to mono-dimensional controlled diffusion processes over finite temporal horizons. Let $\mathbb{T} = [0,T]$ where $T < \infty$ be the time interval of interest, and let $X_t : \Omega \times \mathbb{T} \mapsto \mathbb{R}$ be the stochastic process (the state variable) we are trying to control. As convention would have it, the latter is assumed to inhabit the stochastic basis $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} = \{\mathscr{F}_t^X\}_{t\in\mathbb{T}}$ is the canonical filtration of X. Defining the mono-dimensional Brownian motion $\{W_t\}_{t\in\mathbb{T}}$ we stipulate the governing dynamics of X as

$$dX_s^\alpha = b(s, X_s^\alpha, \alpha_s)ds + \sigma(s, X_s^\alpha, \alpha_s)dW_s, \tag{2.1}$$

where $X_t^\alpha = x$, and $\alpha_s = \alpha(s, X_s)$ is an $\mathscr{F}_s^X$-adapted Markovian control, which takes values in the control space $\mathbb{A} \subset \mathbb{R}^m$, and b and σ are continuous functions $b : \mathbb{T} \times \mathbb{R} \times \mathbb{A} \mapsto \mathbb{R}$ and $\sigma : \mathbb{T} \times \mathbb{R} \times \mathbb{A} \mapsto \mathbb{R}$ chosen such as to guarantee the existence of a unique strong solution.[1]

The fundamental control problem we are trying solve is that of maximising

$$W(t,x,\alpha) = \mathbb{E}_{t,x}\left[\int_t^T e^{-\int_t^s \beta_u du} f(s, X_s^\alpha, \alpha_s)ds + e^{-\int_t^T \beta_u du} g(X_T^\alpha)\right], \tag{2.2}$$

for all $(t,x) \in \mathbb{T} \times \mathbb{R}$ and $\alpha \in \mathscr{A}(t,x)$. Here, $f : \mathbb{T} \times \mathbb{R} \times \mathbb{A} \mapsto \mathbb{R}$ is a running reward function, whilst $g : \mathbb{R} \mapsto \mathbb{R}$ is a terminal reward function, both of which typically are taken to satisfy quadratic growth conditions. $\beta_u = \beta(u, X_u^\alpha, \alpha_u)$ is a function $\beta : \mathbb{T} \times \mathbb{R} \times \mathbb{A} \mapsto \mathbb{R}$ which present-values the reward functions as appropriate. Finally, we denote by $\mathscr{A}(t,x)$ the subset of controls which are *admissible*, i.e. the $\mathscr{F}_t^X$-adapted $\mathbb{A}$-valued controls which minimally satisfy $\mathbb{E}[\int_t^T |e^{-\int_t^s \beta_u du} f(s, X_s^\alpha, \alpha_s)|ds] < \infty$.

The crux of the matter is that the maximisation problem (2.2) is not exactly trivial. Defining the (optimal) *value function*

$$V(t,x) = \sup_{\alpha \in \mathscr{A}(t,x)} W(t,x,\alpha), \tag{2.3}$$

one may invoke the *dynamic programming principle* (DPP)

$$V(t,x) = \sup_{\alpha \in \mathscr{A}(t,x)} \mathbb{E}_{t,x}\left[\int_t^\tau e^{-\int_t^s \beta_u du} f(s, X_s^\alpha, \alpha_s)ds + e^{-\int_t^\tau \beta_u du} V(\tau, X_\tau^\alpha)\right], \tag{2.4}$$

for any stopping time τ, in order to set up the *Hamilton-Jacobi-Bellman* (HJB) equation

$$0 = \partial_t V + \sup_{a \in \mathbb{A}} \left\{-\beta_t V + b(t,x,a)\partial_x V + \tfrac{1}{2}\sigma^2(t,x,a)\partial_{xx}^2 V + f(t,x,a)\right\}, \tag{2.5}$$

s.t. $V(T,x) = g(x)$. The most striking feature of the HJB equation is arguably the change in the supremum: while the original control problem asked us to optimise over the set of $\mathbb{A}$-valued control processes $\{\alpha_s, s \geq 0\}$, we are now faced with "merely" having to

[1] Here, standard assumptions are that the SDE satisfies (I) the uniform Lipschitz condition $\exists K \in (0,\infty)$ s.t. $\forall t \in \mathbb{T}$, $\forall x,y \in \mathbb{R}$ and $\forall \alpha \in \mathbb{A}$: $|b(t,x,\alpha) - b(t,y,\alpha)| + |\sigma(t,x,\alpha) - \sigma(t,y,\alpha)| \leq K|x-y|$, and (II) $\mathbb{E}[\int_0^T |b(t,0,\alpha)|^2 + |\sigma(t,0,\alpha)|^2 dt] \leq \infty$. Notice though, that these are fairly strict assumptions which rule out square root processes (CIR models).

optimise over the set $\mathbb{A}$. Insofar as a solution, ϕ, can be found to (2.5) we may invoke a *verification procedure* to check if ϕ coincides with the value function (i.e. to check if we have indeed solved the problem). Assuming the technicalities above, it suffices to check if (a) $\phi \in \mathscr{C}^{1,2}([0,T) \times \mathbb{R}) \cap \mathscr{C}^0(\mathbb{T} \times \mathbb{R})$ and if (b) ϕ satisfies a quadratic growth condition. The reader is referred to Fleming & Soner [21], Pham [18], and Ross [41], for details.

2.3 The Merton Problem: An Analytic Reminder

The benchmark result against which we shall be comparing most of our numerical procedures stems right from the foundations of stochastic control theory in continuous time finance. Specifically, we are interested in Merton's quintessential problem of portfolio optimisation, [12], [35], over finite investment horizons $\mathbb{T} = [0,T]$. For the reader's convenience, we here provide a cursory overview of the problem.[2]

Let X be the running wealth of an investor who trades continuously in a risk free asset and a stock, the respective price processes of which obey the usual dynamical equations

$$dB_t = rB_t dt, \quad \text{and} \quad dS_t = \mu S_t dt + \sigma S_t dW_t,$$

where r, μ and σ are constant parameters. If $\theta_t : \Omega \times \mathbb{T} \mapsto \mathbb{R}$ denotes the total [dollar] amount the investor has in stocks, and $c_t : \Omega \times \mathbb{T} \mapsto \mathbb{R}_+$ denotes his rate of consumption, it follows from the *self-financing condition* that

$$dX_s^{\theta,c} = [rX_s^{\theta,c} + \theta_s(\mu - r) - c_s]ds + \theta_t \sigma dW_s, \tag{2.6}$$

where $X_t^{\theta,c} = x$ is the initial endowment (assumed non-negative). Deriving utility from both his life-time consumption rate as well as his terminal bequest, the investor's problem is that of determining an optimal control pair $(\theta_t^*, c_t^*,) \in \mathscr{A}(t,x)$ s.t.

$$W(t,x,\theta,c) = \mathbb{E}_{t,x}\left[\int_t^T e^{-\beta(s-t)} u(c_s) ds + e^{-\beta(T-t)} u(X_T^{\theta,c})\right], \tag{2.7}$$

is maximal, where

$$\mathscr{A}(t,x) := \left\{(\theta,c) : \int_t^T c_s ds + \int_t^T \theta_s^2 ds < \infty \ \& \ X_t^{\theta,c} \geq 0 \text{ a.s. } \forall t \in \mathbb{T}\right\}.$$

The discount factor, β, is assumed constant, and the utility function $u : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is assumed to be isoelastic, i.e. $u(x) = x^{1-\gamma}/(1-\gamma)$, where γ codifies the investor's level of risk aversion. In this paper we will restrict our attention to the root functions $\gamma \in (0,1)$. Setting up the HJB equation

$$\beta V = \partial_t V + \sup_{(\theta,c) \in \mathbb{R} \times \mathbb{R}_+} \left\{[rx + \theta(\mu - r) - c]\partial_x V + \tfrac{1}{2}\theta^2 \sigma^2 \partial_{xx}^2 V + \frac{c^{1-\gamma}}{1-\gamma}\right\}, \tag{2.8}$$

[2] For a more thorough survey we refer to Munk [15].

s.t. $V(T,x) = x^{1-\gamma}/(1-\gamma)$, we find the *first order conditions*:

$$\theta^*(t,x) = -\frac{(\mu - r)}{\sigma^2}\frac{\partial_x V}{\partial^2_{xx} V}, \quad \text{and} \quad c^*(t,x) = (\partial_x V)^{-1/\gamma}. \tag{2.9}$$

Using the ansatz $V(t,x) = g(t)^{\gamma} x^{1-\gamma}(1-\gamma)^{-1}$, where g is some deterministic function of time, we can reduce this problem to a Bernoulli equation, the solution to which is

$$g(t) = A^{-1}(1 + [A-1]e^{-A(T-t)}),$$

where $A = [\beta - r(1-\gamma)]/\gamma - \frac{1}{2}(1-\gamma)[\mu - r]^2/(\gamma^2\sigma^2)$. Crucially, this allows us to calculate explicit expressions for the first order conditions of (2.8) viz.

$$\theta^*(t,x) = \frac{(\mu - r)x}{\gamma\sigma^2}, \quad \text{and} \quad c^*(t,x) = \frac{x}{g(t)}, \tag{2.10}$$

which, from verification, constitute the desired optimal controls.

2.4 Towards a Trinomial/Explicit Markov Chain Approximation

2.4.1 Establishing the Approximation

Consider the discretisation $\mathscr{T}^{\delta} \times \mathscr{R}^h = \{0, \delta, 2\delta, ..., N\delta = T\} \times \{x_{\min}, x_{\min} + h, x_{\min} + 2h, ..., x_{\min} + Ih =: x_{\max}\}$ of the full state space $\mathbb{T} \times \mathbb{R}$ of section 2.2, where, $x_{\min}$ and $x_{\max}$ are artificially imposed lower and upper boundaries, whilst h and δ are fixed spatial and temporal separations in the grid. Let $\{\xi_n^{h,\delta} | n \in \mathbb{N}_0\}$ be a controlled discrete parameter Markov chain on $\mathscr{R}^h$ that approximates (in a sense soon to be spelled out) the controlled process X_t^{α}. The stochastic evolution of said chain is determined by the set of probabilities $\{p^{h,\delta}(x,y|a,t) : x,y \in \mathscr{R}^h\}$ where $p^{h,\delta}(x,y|a,t)$ denotes the conditional probability that the Markov chain jumps from state x to state y given that the control $a \in \mathbb{A}$ is applied at time t. Needless to say, if these probabilities are to be construed as Markov chain transition probabilities, then they must satisfy the basic requirements of positivity and summing to unity. Let $a_n^{h,\delta} = a^{h,\delta}(n\delta, \xi_n^{h,\delta})$ denote the random variable, which is the actual control action for the chain, applied at the discrete time $n\delta$. We say the control policy $a^{h,\delta} = \{a_n^{h,\delta} | n \in \mathbb{N}_0\}$ for the chain is *admissible* provided that (i) $a_n^{h,\delta} \in \mathbb{A}$ and (ii) the chain has the Markov property under that policy:

$$\mathbb{P}[\xi_{n+1}^h = y | \xi_i^{h,\delta}, a_i^{h,\delta}, i \le n] = p^h(\xi_n^{h,\delta}, y | a_n^{h,\delta}, n\delta).$$

Let $\mathscr{A}^{h,\delta}(n\delta, x)$ denote the set of admissible controls given that $\xi_n^{h,\delta} = x$ at time $n\delta$. We may then state the approximation to (2.2) as

$$W^{h,\delta}(n\delta,x,a^{h,\delta}) = \mathbb{E}\Bigg[\sum_{i=n}^{N-1} e^{-\sum_{j=n}^{i-1}\beta(j\delta,\xi_j^{h,\delta},a_j^{h,\delta})\delta} f(i\delta,\xi_i^{h,\delta},a_i^{h,\delta})\delta + e^{-\sum_{j=n}^{N-1}\beta(j\delta,\xi_j^{h,\delta},a_j^{h,\delta})\delta} g(\xi_N^{h,\delta})\Bigg|\xi_n^{h,\delta}=x\Bigg],$$

whilst the discretisation of the value function (4.1) is defined as

$$V^{h,\delta}(n\delta,x) = \sup_{a^{h,\delta}\in\mathscr{A}^{h,\delta}(n\delta,x)} W^{h,\delta}(n\delta,x,a^{h,\delta}). \tag{2.11}$$

Crucially, in order to secure convergence of $V^{h,\delta}(n\delta,x)$ to $V(n\delta,x)$ as $h\to 0$, it is incumbent that the Markov chain approximation is chosen in accordance with *local consistency conditions*. Specifically, defining $\Delta\xi_n^{h,\delta} := \xi_{n+1}^{h,\delta} - \xi_n^{h,\delta}$ we require that $\sup_{n,\omega}|\Delta\xi_n^{h,\delta}|\to 0$ as $h\to 0$ as well as

$$\mu_n^{h,\delta}(x,a) := \mathbb{E}[\Delta\xi_n^{h,\delta}|\xi_n^{h,\delta}=x,a_n^{h,\delta}=a] = b(n\delta,x,a)\delta + o(\delta), \tag{2.12a}$$

$$\Sigma_n^{h,\delta}(x,a) := \mathbb{E}[(\Delta\xi_n^{h,\delta}-\mu_n^{h,\delta}(x,a))^2|\xi_n^{h,\delta}=x,a_n^{h,\delta}=a] = \sigma^2(n\delta,x,a)\delta + o(\delta), \tag{2.12b}$$

$\forall x\in\mathscr{R}^h$ and $\forall a\in\mathbb{A}$. Here $o(y)$ is defined as a function which is small relative to y i.e: $\lim_{y\to 0} o(y)/y = 0$.

2.4.2 Extracting the Solution

Insofar as we have a locally consistent Markov chain approximation to X_t^α, we may solve for the discretised value function (2.11) through repeated application of the explicit DPP:[3]

$$V^{h,\delta}(n\delta,x) = \sup_{a\in\mathbb{A}}\left[f(n\delta,x,a)\delta + e^{-\beta(n\delta,x,a)\delta}\sum_{y\in\mathscr{R}^h} p^{h,\delta}(x,y|a,n\delta)V^{h,\delta}((n+1)\delta,y)\right] \tag{2.13}$$

starting from the terminal condition $V^{h,\delta}(N\delta,x) = g(x)$ and working our way incrementally backwards in time. As in a garden variety (linear) explicit procedure, the grid component $V^{h,\delta}(n\delta,x)$ is given entirely in terms of known quantities $\{V^{h,\delta}((n+1)\delta,y) : y\in\mathscr{R}^h\}$, albeit with the added caveat that we must maximise the expression over all $a\in\mathbb{A}$ at every step in the process. Evidently, this aspect should (insofar as possible) be handled through the employ of the associated *first order conditions* i.e. the a^* which renders the partial

[3] This equation is a one-step discrete approximation to (2.4) i.e.

$$V^{h,\delta}(n\delta,x) = \sup_{a\in\mathbb{A}}\left[f(n\delta,x,a)\delta + e^{-\beta(n\delta,x,a)\delta}\mathbb{E}\left[V^{h,\delta}((n+1)\delta,y)\Big|\xi_n^{h,\delta}=x,a_n^{h,\delta}=a\right]\right]$$

derivative of the RHS of (2.13) with respect to a equal to zero. A considerably more time consuming procedure involves a search over a bounded mesh of possible controls at every point in the grid $\mathscr{T}^{\delta} \times \mathscr{R}^{h}$.

The final piece left of the puzzle is that of how we go about constructing a locally consistent Markov chain in the first place. A luminous "beacon in night" is in this context a fairly flexible programme involving finite difference approximations of the governing differential equation. Specifically, consider the PDE formally satisfied by (2.2). From Feynmac-Kac's theorem we have

$$\beta_t W = \partial_t W + b(t,x,a)\partial_x W + \tfrac{1}{2}\sigma^2(t,x,a)\partial_{xx}^2 W + f(t,x,a), \tag{2.14}$$

or in explicit discretised terms

$$\begin{aligned}\beta_{n\ \delta} W^{h,\delta}((n+1)\delta,x,a) =& \delta^{-1}[W^{h,\delta}((n+1)\delta,x,a) - W^{h,\delta}(n\delta,x,a)] \\ &+ b(n\delta,x,a)\mathscr{D}_x W^{h,\delta}((n+1)\delta,x,a) \\ &+ \tfrac{1}{2}\sigma^2(n\delta,x,a)\mathscr{D}_{xx}^2 W^{h,\delta}((n+1)\delta,x,a) + f(n\delta,x,a).\end{aligned} \tag{2.15}$$

The key step consists of defining the difference operators $\mathscr{D}_x$ and $\mathscr{D}_{xx}^2$ such that equation (2.15) yields transition probabilities which are uniformly non-negative. For this reason we forgo the otherwise cherished usage of central differencing and propose the following *up-wind* scheme:

- If $b(t,x,a) \geq 0$ let $\mathscr{D}_x W(t,x,a) = \mathscr{D}_x^+ W(t,x,a) := h^{-1}[W(t,x+h,a) - W(t,x,a)]$.
- If $b(t,x,a) < 0$ let $\mathscr{D}_x W(t,x,a) = \mathscr{D}_x^- W(t,x,a) := h^{-1}[W(t,x,a) - W(t,x-h,a)]$.
- Let $\mathscr{D}_{xx}^2 W(t,x,a) = h^{-2}[W(t,x+h,a) - 2W(t,x,a) + W(t,x-h,a)]$.

Defining $[y]^+ = \max\{y,0\}$ and $[y]^- = \max\{-y,0\}$, items 1 & 2 may be written as

- $b(t,x,a)\mathscr{D}_x W(t,x,a) = h^{-1}[W(t,x+h,a) - W(t,x,a)][b(t,x,a)]^+ - h^{-1}[W(t,x,a) - W(t,x-h,a)][b(t,x,a)]^-$.

Substituting these approximations into (2.15) and rearranging we obtain

$$\begin{aligned}&W^{h,\delta}(n\delta,x,a) \\ &\quad= \left(\delta h^{-1}[b(n\delta,x,a)]^+ + \tfrac{1}{2}\delta h^{-2}\sigma^2(n\delta,x,a)\right) W^{h,\delta}((n+1)\delta,x+h,a) \\ &\quad+ \left(1 - \delta\beta(n\delta,x,a) - \delta h^{-1}|b(n\delta,x,a)| - \delta h^{-2}\sigma^2(n\delta,x,a)\right) W^{h,\delta}((n+1)\delta,x,a) \\ &\quad+ \left(\delta h^{-1}[b(n\delta,x,a)]^- + \tfrac{1}{2}\delta h^{-2}\sigma^2(n\delta,x,a)\right) W^{h,\delta}((n+1)\delta,x-h,a) \\ &\quad+ f(n\delta,x,a)\delta,\end{aligned}$$

where we have made use of the fact that $|y| = [y]^+ + [y]^-$. More succinctly, we can write this (with obvious notation) as

$$W^{h,\delta}(n\delta,x,a) = \sum_{y \in \mathscr{R}^h(x)} p^{h,\delta}(x,y|a,n\delta) W^{h,\delta}((n+1)\delta,y,a) + f(n\delta,x,a)\delta,$$

where $\mathscr{R}^h(x) := \{x+h, x, x-h\}$, whence we have clear candidates for the transition probabilities. However, $p^{h,\delta}(x,x+h|a,n\delta), p^{h,\delta}(x,x|a,n\delta)$ and $p^{h,\delta}(x,x-h|a,n\delta)$ do not sum to unity. This prompts us to invoke a *renormalisation*: specifically, defining $\mathscr{N}_a^{h,\delta} := 1/(1-\delta\beta(n\delta,x,a))$ we propose the following

$$p^{h,\delta}(x,x+h|a,n\delta) = \mathscr{N}_a^{h,\delta}\left(\delta h^{-1}[b(n\delta,x,a)]^+ + \tfrac{1}{2}\delta h^{-2}\sigma^2(n\delta,x,a)\right), \tag{2.16a}$$

$$\begin{aligned} p^{h,\delta}(x,x|a,n\delta) = \mathscr{N}_a^{h,\delta}\Big(&1-\delta\beta(n\delta,x,a) \\ &-\delta h^{-1}|b(n\delta,x,a)| - \delta h^{-2}\sigma^2(n\delta,x,a)\Big), \end{aligned} \tag{2.16b}$$

$$p^{h,\delta}(x,x-h|a,n\delta) = \mathscr{N}_a^{h,\delta}\left(\delta h^{-1}[b(n\delta,x,a)]^- + \tfrac{1}{2}\delta h^{-2}\sigma^2(n\delta,x,a)\right), \tag{2.16c}$$

$$p^{h,\delta}(x,y|a,n\delta) = 0, \qquad \forall y \notin \mathscr{R}^h(x), \tag{2.16d}$$

which clearly satisfy $\sum_{y\in\mathscr{R}^h(x)} p(x,y|n\delta,a) = 1$. Furthermore, insofar as $1 > \delta\beta(n\delta,x,a) + \delta h^{-1}|b(n\delta,x,a)| + \delta h^{-2}\sigma^2(n\delta,x,a)$ $\forall x \forall n \forall a$, or, equivalently,

$$\delta < \left[\beta(n\delta,x,a) + h^{-1}|b(n\delta,x,a)| + h^{-2}\sigma^2(n\delta,x,a)\right]^{-1}, \tag{2.17}$$

we can also guarantee the $p^{h,\delta}(x,x|a,n\delta)$ stays uniformly non-negative (the requirement of non-negativity is obviously satisfied by $p^{h,\delta}(x,x+h|a,n\delta)$ and $p^{h,\delta}(x,x-h|a,n\delta)$). Finally, it is easy to show that the probabilities satisfy the local consistency conditions (2.12). Thus, for (2.12a) we find that

$$\begin{aligned} \mu_n^{h,\delta}(x,a) &= h\cdot p^{h,\delta}(x,x+h|a,n\delta) + 0\cdot p^{h,\delta}(x,x|a,n\delta) - h\cdot p^{h,\delta}(x,x-h|a,n\delta) \\ &= h\mathscr{N}_a^{h,\delta}\left(\delta h^{-1}[b(n\delta,x,a)]^+ - \delta h^{-1}[b(n\delta,x,a)]^-\right) \\ &= (1-\delta\beta(n\delta,x,a))^{-1} b(n\delta,x,a)\delta \\ &= b(n\delta,x,a)\delta + o(\delta), \end{aligned}$$

as desired. Similarly, for (2.12b)

$$\begin{aligned} \Sigma_n^{h,\delta}(x,a) &= (h - b(n\delta,x,a)\delta)^2\cdot p^{h,\delta}(x,x+h|a,n\delta) + (b(n\delta,x,a)\delta)^2\cdot p^{h,\delta}(x,x|a,n\delta) \\ &\quad + (-h-b(n\delta,x,a)\delta)^2\cdot p^{h,\delta}(x,x-h|a,n\delta) + o(\delta) \\ &= \mathscr{N}_a^{h,\delta}\Big(h^2[\delta h^{-1}[b(n\delta,x,a)]^+ + \tfrac{1}{2}\delta h^{-2}\sigma^2(n\delta,x,a)] \\ &\quad + h^2[\delta h^{-1}[b(n\delta,x,a)]^- + \tfrac{1}{2}\delta h^{-2}\sigma^2(n\delta,x,a)]\Big) + o(\delta) \\ &= (1-\delta\beta(n\delta,x,a))^{-1}\Big(\delta\sigma^2(n\delta,x,a)\Big) + o(\delta) \\ &= \sigma^2(n\delta,x,a)\delta + o(\delta). \end{aligned}$$

Hence, the proposed transition probabilities are locally consistent.

Remark 2.1. In practice, the drift term $b(n\delta,x,a)$ is often found to be a linear sum of several different components. This allows us to deploy a method known as *splitting the operator*, which can reduce the computational complexity of (2.16) somewhat. Suppose $b(n\delta,x,a) = \sum_{i=1}^{k} b_i(n\delta,x,a)$ and that we apply the upwind criterion to each component of b individually rather than b as a whole. Specifically, writing $b(n\delta,x,a)\mathscr{D}_x W$ as $\sum_{i=1}^{k} b_i(n\delta,x,a)\mathscr{D}_{x,i}W$, suppose we choose a forward or backward differencing of $\mathscr{D}_{x,i}W$ based on the sign of $b_i(n\delta,x,a)$ $\forall i$, thus replacing $[b(n\delta,x,a)]^+$ in (2.16a) with $\sum_{i=1}^{k}[b_i(n\delta,x,a)]^+$, and $[b(n\delta,x,a)]^-$ in (2.16c) with $\sum_{i=1}^{k}[b_i(n\delta,x,a)]^-$. The advantage of this is apparent if the signs of the individual components are a priori given!

Remark 2.2. The skeptical reader might wonder whether our commitment to positive probabilities is mandatory. After all, there seems to be a growing community of iconoclasts who argue for the cogency of negative probabilities in finance, cf. Haug [13], Meissner et al. [18] and Zvan et el. [29]. Unfortunately, this heterodoxy is a dangerous game for our present purposes: specifically, upon proving convergence of the Markov chain approximation from viscosity principles, Kusher and Dupuis assume the *monotonicity property* (assumption A2.1. p. 449 [31]), which manifestly requires all probabilities to be non-negative.[4] Thus, we shall continue to abide by the positivity criterion (here, equation (2.17)), even though this inexorably will force us to adopt extremely small time steps in a Mertonian context.

2.5 The Trinomial/Explicit Method and Merton's Problem

The DPP (2.13) and its tripartite probability structure (2.16), naturally allows for two different algorithmic interpretations with clear analogies in numerical option pricing. On the one hand, we can view it as recipe for a *trinomial tree*; on the other, as a full-fledged *finite difference grid.* Whilst this prima facie might appear like a minor detail, the difference of which boils down to algorithmic run time (the number of grid points evaluated), we shall argue that the rabbit hole goes deeper. Specifically, if we opt for the finite difference interpretation, then inevitably we will have to make specifications for the boundary points (a somewhat nebulous endeavour). This issue is side-stepped with a trinomial model, albeit at the cost of interlocking the number time steps with respect to the number of space steps, thereby complicating the positivity criterion (2.17) further. We provide a full exposition of these issues below: first, however, it is worthwhile phrasing the DPP and the associated transition probabilities for the Merton problem.

In line with non-negativity assumption on financial wealth we restrict x to the state-space $\mathscr{R}_1^h = \{0,h,2h,...,Ih = x_{\max}\}$. Let $\theta(n\delta,x)$ and $c(n\delta,x)$ be the discrete controls of the problem, both of which are supposed to be bounded by the interval $[0,Kx]$, for some constant K.[5] Then $\forall x \in \{h,2h,...,(I-1)h\}$ the DPP may be stated as

[4] For an illuminating account on convergence of numerical HJB schemes to viscosity solutions the reader is referred to Forsyth et al. [24] particularly lemma 5.3 and theorem 5.1.

[5] Bounding different controls by different constants is obviously quite feasible: the reader should make a personal judgement call as to what makes sense in a given context.

$$V^{h,\delta}(n\delta,x) = \sup_{(\theta,c)\in\mathbb{R}\times\mathbb{R}_+}\left[\frac{c(n\delta,x)^{1-\gamma}}{1-\gamma}\delta + e^{-\beta\delta}\sum_{y\in\mathscr{R}^h(x)} p^{h,\delta}(x,y|\theta,c,n\delta)V^{h,\delta}((n+1)\delta,y)\right], \tag{2.18}$$

where $p^{h,\delta}(x,y|\theta,c,n\delta)$ is shorthand notation for $p^{h,\delta}(x,y|\theta(n\delta,x),c(n\delta,x),n\delta)$ and $\mathscr{R}^h(x) = \{x+h,x,x-h\}$. Specifically, we define the transition probabilities

$$\begin{aligned}
p^{h,\delta}(x,x+h|\theta,c,n\delta) &= \frac{1}{1-\delta\beta}\left(\delta h^{-1}\left(rx+\theta(n\delta,x)(\mu-r)\right)+\tfrac{1}{2}\delta h^{-2}\theta^2(n\delta,x)\sigma^2\right),\\
p^{h,\delta}(x,x|\theta,c,n\delta) &= \frac{1}{1-\delta\beta}\Big(1-\delta\beta-\delta h^{-1}\left(rx+\theta(n\delta,x)(\mu-r)+c(n\delta,x)\right)\\
&\qquad -\delta h^{-2}\theta^2(n\delta,x)\sigma^2\Big),\\
p^{h,\delta}(x,x-h|\theta,c,n\delta) &= \frac{1}{1-\delta\beta}\left(\delta h^{-1}c(n\delta,x)+\tfrac{1}{2}\delta h^{-2}\theta^2(n\delta,x)\sigma^2\right),\\
p^{h,\delta}(x,y|\theta,c,n\delta) &= 0, \qquad \forall y\notin\mathscr{R}^h(x),
\end{aligned}$$

where we have made use of the splitting of the operator technique.[6] Finally, two straightforward differentiations of (2.18) yield the first order conditions (FOCs)

$$\begin{aligned}
\theta^*(n\delta,x) &= -\frac{(\mu-r)}{\sigma^2}\frac{\mathscr{D}_x^+V^{h,\delta}((n+1)\delta,x)}{\mathscr{D}_{xx}^2V^{h,\delta}((n+1)\delta,x)},\\
c^*(n\delta,x) &= \left(\frac{e^{-\beta\delta}}{1-\beta\delta}\mathscr{D}_x^-V^{h,\delta}((n+1)\delta,x)\right)^{-1/\gamma}
\end{aligned}$$

where the differencing operators $\mathscr{D}_x^+, \mathscr{D}_x^-$ and $\mathscr{D}_{xx}^2$ are as defined above, and we enforce the restriction $\theta^*(n\delta,x), c^*(n\delta,x) \in [0,KIh]$ $\forall n\forall x$. Given these analytic expressions for the optimal controls, the problem we are trying to solve is as simple as computing

$$V^{h,\delta}(n\delta,x) = \left[\frac{c^*(n\delta,x)^\gamma}{1-\gamma}\delta + e^{-\beta\delta}\sum_{y\in\mathscr{R}^h(x)} p^{h,\delta}(x,y|\theta^*,c^*,n\delta)V^{h,\delta}((n+1)\delta,y)\right],$$

incrementally backwards in time: $n = N-1, N-2, ..., 1, 0$.

[6] Specifically, since $x \geq 0$ $[rx]^+ = rx$ ($\Rightarrow [rx]^- = 0$); since $\mu > r$ the position in the risky asset must be positive whence $[\theta(\mu-r)]^+ = \theta(\mu-r)$ ($\Rightarrow [\theta(\mu-r)]^- = 0$). Lastly, since consumption is non-negative $[-c]^- = c$ ($\Rightarrow [-c]^+ = 0$).

				-64.645					-74.676
				-17.553					-26.199
				-10.483					-16.438
				-7.4424					-11.819
			⋰	-5.7583				⋰	-9.1145
		-5.0536		-4.6915			-7.7463		-7.3362
	-4.3111	-4.2530		-3.9562		-6.5044	-6.4357		-6.0774
-3.7686	-3.7199	-3.6708	…	-3.4192	-5.5826	-5.5220	-5.4606	…	-5.1392
	-3.2710	-3.2285		-3.0100		-4.7625	-4.7065		-4.4129
		-2.8811		-2.6880			-4.1060		-3.8340
				-2.4280					-3.3618
			⋱	-2.2138				⋱	-2.9691
				-2.0342					-2.6375
				-1.8815					-2.3538
				-1.7501					-2.1082
$t=0$	$t=\delta$	$t=2\delta$	…	$t=7\delta$	$t=0$	$t=\delta$	$t=2\delta$	…	$t=7\delta$

Table 2.1 The percentage errors (defined as $100 \cdot (a_{\text{approx.}} - a_{\text{true}})/a_{\text{true}}$) for the optimal stock investment (**left**) and the optimal consumption (**right**) at various times in the trinomial tree. The row number corresponds to the corresponding integer multiple of h in the state space. In particular, the $t=0$ node occurs at $x = 8h$.

γ	β	r	μ	σ	$x_{\max}$	K
0.5	0.02	0.05	0.1	0.3	100	1.5

Table 2.2 The parametric specifications for the risk aversion γ, the subjective discounting β, the risk free rate r, the stock drift and volatility μ and σ, the upper bound on the state space $x_{\max}$, and the constant curbing the controls from above K.

2.5.1 The Trinomial Method

Let I be an even number ($\in 2\mathbb{N}$). The trinomial method is exactly what the name suggests: a recombining tree diagram where every node has exactly three child nodes (or, equivalently, a grid which cuts off both end nodes of its state space whenever we move backwards in time). Thus, at expiry, we set out by evaluating $V^{h,\delta}$ for all $x \in \{0, h, ..., (I-1)h, Ih\}$ using the terminal condition. Then, one time step prior, we compute $V^{h,\delta}$ for all $x \in \{h, ..., (I-1)h\}$ using the DPP and continue thusly until $\frac{1}{2}I$ time steps into the past we compute $V^{h,\delta}$ for the singleton $x \in \{\frac{1}{2}Ih\}$. The advantages of this procedure over a full finite difference grid are clear: first, it saves us the trouble of having to specify boundary conditions for $V^{h,\delta}(\cdot, 0)$ and $V^{h,\delta}(\cdot, Ih)$. Secondly, assuming we are indeed only interested in the centre grid controls, there is a clear reduction in the number of nodes we need to evaluate ($(\frac{1}{2}I+1)^2$ for the tree versus $(\frac{1}{2}I+1)(I+1)$ for an analogously sized grid). However, the fact that we interlock the number of spatial separations in the grid $\#(h) = I$ with the number of temporal separations $\#(\delta) = N = \frac{1}{2}I$ does not bode well for the positivity criterion

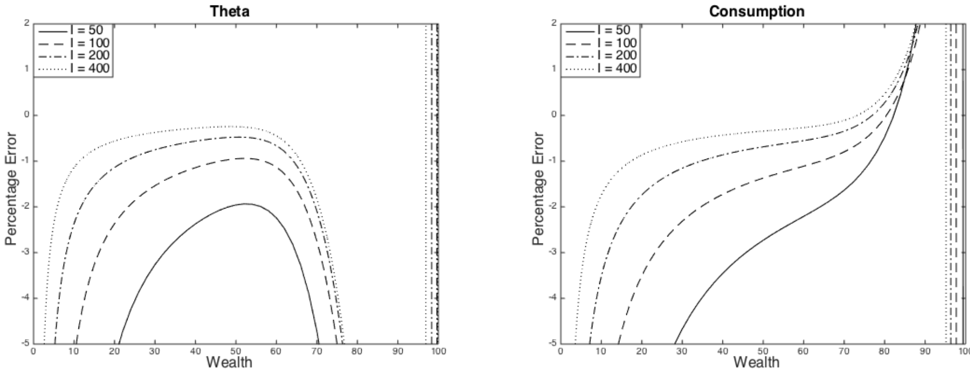

Fig. 2.1 The percentage errors of the numerically computed optimal stock investment (**left**) and the optimal consumption (**right**) computed for various values of I. We assume $T = 1$ and $N = \lceil \beta + (r + (\mu - r)K + K)I + \sigma^2 K^2 I^2 \rceil$.

$$\forall x \forall \theta \forall c : \delta < \left[\beta + h^{-1}\left(rx + \theta(n\delta, x)(\mu - r) + c(n\delta, x)\right) + h^{-2}\theta^2(n\delta, x, a)\sigma^2\right]^{-1}, \tag{2.19}$$

cf. equation (2.17). Specifically, upon setting $\delta = T/(\frac{1}{2}I)$, $h = x_{\max}/I$, $x = x_{\max}$ and $\theta = c = Kx_{\max}$ we get the *worst case scenario* inequality, which (if satisfied) surely will guarantee the positivity of all transition probabilities (and thence the convergence of the algorithm):

$$T < \tfrac{1}{2}I\left[\beta + (r + (\mu - r)K + K)I + \sigma^2 K^2 I^2\right]^{-1}. \tag{2.20}$$

This is a serious constraint. Suppose momentarily $I \in \mathbb{R}_+$: upon viewing the RHS as a function $f : \mathbb{R}_+ \mapsto \mathbb{R}$ of I, we find that f assumes a maximum at

$$I^* = \sqrt{\frac{\beta}{\sigma^2 K^2}}.$$

For realistic parametric values, $f(I^*)$ might prove considerably less than the desired level of T. Indeed, the true solution space (I should be a positive even number) will curb the allowed T values even further. Thus, the trinomial model might not be able to solve the Merton problem for the set of parameters we desire. Or at least not without some modification: e.g. by introducing varying time steps in line with the fact that the state space decreases as we move backwards in time.

To get a feel for the gravity of this, consider a set-up with the parametric choices given in table 3.2. We find $I^* = 0.943$ and $f(I^*) = 0.300$: i.e. (2.20) makes us consider $T \leq 0.300$. Suppose we set $T = 0.1$: since $f(16) = 0.103$ an acceptable grid specification is $I = 16$ ($dt = 0.0125$, $h = 6.25$). Table 3.1 is a (partial) report of the percentage errors of the numerical controls incurred from the trinomial method. Unsurprisingly, the large grid spacing ($h = 6.25$) gives rise to non-negligible errors when evaluating the differencing operators $\mathscr{D}_x^{\pm}V$, $\mathscr{D}_{xx}^2 V$ *in lieu* of the derivatives $\partial_x V$ and $\partial_{xx}^2 V$. Indeed, these errors are most pronounced around $x = 0$ where the square root utility function ($\gamma = 0.5$) experiences the

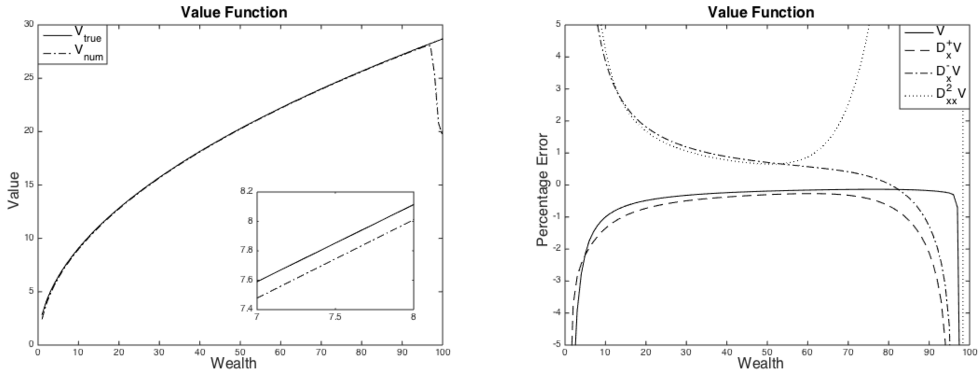

Fig. 2.2 The analytic and numerical value functions (with miniature zoom) (**left**) and the percentage errors of the numerically computed value function including various derivatives (**right**). We assume $I = 100$, $T = 1$ and $N = \lceil \beta + (r + (\mu - r)K + K)I + \sigma^2 K^2 I^2 \rceil$.

sharpest rise. For the mother of all nodes (the trunk of the tree, $(t,x) = (0,50)$) we find the optimal numerical controls $(\theta^*, c^*) = (53.46, 42.76)$ vs. their analytic counterparts $(\theta^*, c^*) = (55.56, 45.29)$ and thence percentage errors of $(-3.77, -5.58)\%$.

2.5.2 The Explicit Method

The easiest way to circumvent the problems of the previous subsection is obviously to solve (2.18) as a full ("rectangular") finite difference grid, thereby decoupling $\#(\delta)$ from $\#(h)$. Obviously, we will still need to satisfy inequality (2.19) - however, this is now as easy as evaluating the RHS and specifying the δ accordingly. The main obstacle is undoubtedly the requirement that we must now make specifications for the boundary conditions alongside the grid. Compared to numerical problems in option pricing, this is considerably more obscure. E.g. whilst it is plausible that an option deep in or out of the money has vanishing gamma, an analogous argument does not carry over to a Mertonian value function problem.

A standing point of this paper is that the nebulosity of the boundary conditions is detrimental to the accuracy of our numerical controls for a *seizable* chunk of the state space. To this end, consider the *Dirichlet* conditions $\forall n \in \{0, 1, ..., N-1\}$:

$$V^{h,\delta}(n\delta, 0) = 0, \quad \text{and} \quad V^{h,\delta}(n\delta, Ih) = V^{h,\delta}(T, Ih).$$

The philosophy here is simple: a bankrupt investor ($x = 0$) can neither consume, nor build up a bequest. Thus, assuming $\gamma \in (0,1)$ his value function is nil. At the other extreme ($x = Ih$), for relatively short investment horizons we do not expect drastic changes in the value function: hence, we may approximate the upper boundary based on the terminal condition. To test the performance of the explicit method we plot the percentage errors of the numerically computed optimal controls in figure 2.1 for various values of I (and an N just big enough to satisfy the worst case scenario inequality). These plots vividly demonstrate the grave shortcomings of our numerical method for large parts of the upper and lower state space. For low wealth levels this is hardly surprising: here the gradient of the utility function is at its steepest, and we wouldn't expect our relatively coarse grained difference operators to capture this adequately. The fact that the Dirichlet boundary happens to be analytically exact at $x = 0$ does to some degree seem to compensate for this as we increase I and N. On the other hand, the accuracy of θ^* near the upper boundary scarcely profits from increasing the fine-graining of the grid: seemingly, the mis-specification (however modest) of the upper Dirichlet boundary effectively kills our chances of reasonable convergence. Figure 2.2 offers a deeper investigation: the numerical value function is highly accurate for most wealth levels except at the upper boundary. The percentage error of the numerically computed second derivative is catastrophic in this region, which in turn propagates to the optimal stock investment (recall θ^* supervenes upon the ratio $\mathscr{D}_x^+ V / \mathscr{D}_{xx}^2 V$).

Remark 2.3. We previously stressed the importance of utilising the (numerical) first order conditions upon coding the stochastic control problem. Nonetheless, there may be situations in which this is not feasible; i.e. where one has to resort to numerical optimisation of the DPP in stead. To test the feasibility of this method for the problem at hand we discretised the space of possible controls $\mathbb{A} = [0,150] \times [0,150]$ into a mesh $\mathbb{A}_m$ of equidistant separation 0.01. At $t = T - \delta$ we performed a maximisation of the DPP over the full control mesh for every node in the optimal value grid. However, based on the principle that controls do not tend to vary drastically across incrementally close periods of time, all other maximisations where done over a subspace of $\mathbb{A}_m$ tailor-made to the optimal control pair of the subsequent time step. Specifically, if $(\theta^*(n\delta,x), c^*(n\delta,x)) = (\bar{\theta},\bar{c})$ was found to be optimal for $V^{h,\delta}(n\delta,x)$, then the optimal solution for $V^{h,\delta}((n-1)\delta,x)$ was assumed to lie no further than ± 10 (in any direction) of $(\bar{\theta},\bar{c})$. The numerical results of this method are illustrated in figure 2.3. While the algorithmic run time *with* the FOCs was found to be only 2.45 seconds on a 2.5 GHz Intel Core i5 processor, the corresponding run time *without* the FOCs was a staggering 1590.36 seconds.

2.6 Towards an Implicit Method

2.6.1 A New Type of Markov Chain

In subsection 2.4.1 we introduced a controlled discrete parameter Markov chain $\{\xi_n^{h,\delta} | n \in \mathbb{N}_0\}$ on the state space $\mathscr{R}^h = \{x_{\min}, x_{\min} + h, x_{\min} + 2h, \ldots, x_{\min} + Ih =: x_{\max}\}$ in approximation of the controlled process X_t^α. Clearly, space and time play fundamentally different roles in this picture: while $\xi_n^{h,\delta}$ moves dynamically through space in accordance with locally consistent transition probabilities, time is a passive index ordered in terms of multiples of δ. The fundamental difference between the *explicit* and *implicit* method is precisely that the latter promotes time to a full state variable. Specifically, we now consider a controlled discrete parameter Markov chain $\{\zeta_n^{h,\delta} | n \in \mathbb{N}_0\}$ on the full time-space

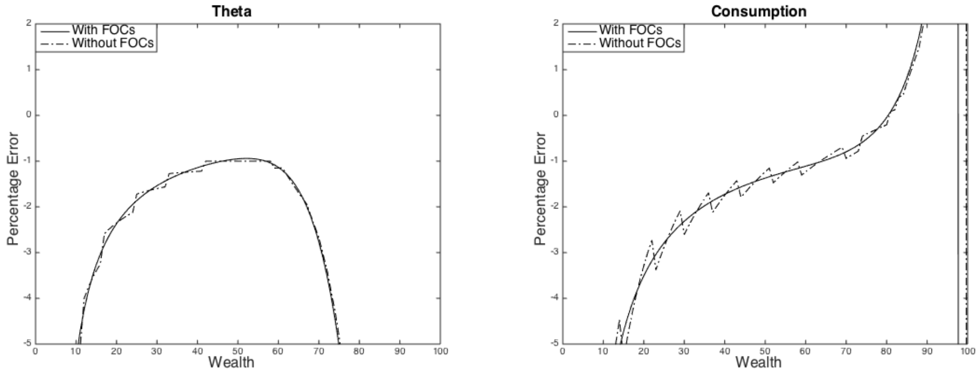

Fig. 2.3 The percentage errors of the numerically computed optimal stock investment (**left**) and the optimal consumption (**right**) with and without the use of the first order conditions. We assume $I = 100$, $T = 1$ and $N = \lceil \beta + (r + (\mu - r)K + K)I + \sigma^2 K^2 I^2 \rceil$.

grid $\mathscr{T}^\delta \times \mathscr{R}^h = \{0, \delta, 2\delta, \dots, N\delta = T\} \times \{x_{\min}, x_{\min} + h, x_{\min} + 2h, \dots, x_{\min} + Ih =: x_{\max}\}$, such that transition probabilities no longer carry the chain solely through space, but also through time.

Denote by $p^{h,\delta}(s,x;t,y|a)$ the conditional probability that the Markov chain jumps from state $(s,x) \in \mathscr{T}^\delta \times \mathscr{R}^h$ to state $(t,y) \in \mathscr{T}^\delta \times \mathscr{R}^h$ given that the control $a \in \mathbb{A}$ is applied. Furthermore, let

$$\Delta t_n^{h,\delta} := \Delta t^{h,\delta}(n\delta, \zeta_n^{h,\delta}, a_n^{h,\delta})$$

denote a positive *interpolation interval*, such that $t_n^{h,\delta} = \sum_{j=0}^{n-1} \Delta t_j^{h,\delta}$. Then the discrete approximations to (2.2) and (4.1) may be stated as

$$\begin{aligned} W^{h,\delta}(n\delta, x, a^{h,\delta}) = \mathbb{E}\Bigg[&\sum_{i=n}^{N-1} e^{-\sum_{j=n}^{i-1} \beta(j\delta, \zeta_j^{h,\delta}, a_j^{h,\delta}) \Delta t_j^{h,\delta}} f(i\delta, \zeta_i^{h,\delta}, a_i^{h,\delta}) \Delta t_i^{h,\delta} \\ &+ e^{-\sum_{j=n}^{N-1} \beta(j\delta, \zeta_j^{h,\delta}, a_j^{h,\delta}) \Delta t_j^{h,\delta}} g(\zeta_N^{h,\delta}) \Bigg| \zeta_n^{h,\delta} = (n\delta, x) \Bigg], \end{aligned}$$

and

$$V^{h,\delta}(n\delta, x) = \sup_{a^{h,\delta} \in \mathscr{A}^{h,\delta}(n\delta, x)} W^{h,\delta}(n\delta, x, a^{h,\delta}).$$

Again, *local consistency* of some sort is the basic requirement for the convergence of the Markov chain. Let $\zeta_{n,0}^{h,\delta}$ and $\zeta_{n,1}^{h,\delta}$ denote the temporal and spatial parts of $\zeta_n^{h,\delta}$ respectively. We then require that $\sup_{n,\omega} |\Delta \zeta_{n,1}^{h,\delta}| \to 0$ as $h \to 0$ as well as

$$\mu_n^{h,\delta}(x,a) := \mathbb{E}[\Delta \zeta_{n,1}^{h,\delta} | \zeta_{n,1}^{h,\delta} = x, a_n^{h,\delta} = a] = b(n\delta, x, a) \Delta t_n^{h,\delta} + o(\Delta t_n^{h,\delta}), \tag{2.21a}$$

$$\mathbb{E}[(\Delta\zeta_{n,1}^{h,\delta} - \mu_n^{h,\delta}(x,a))^2|\zeta_{n,1}^{h,\delta} = x, a_n^{h,\delta} = a] = \sigma^2(n\delta,x,a)\Delta t_n^{h,\delta} + o(\Delta t_n^{h,\delta}), \quad (2.21b)$$

$\forall x \in \mathscr{R}^h$ and $\forall a \in \mathbb{A}$, i.e. the spatial component of the chain must be consistent.

2.6.2 Extracting the Solution

As the designator clearly insinuates, the basic point of the *implicit method* is to set up a DPP, in which any given grid node is coupled to multiple grid nodes at the preceding time step (contrast this to the *explicit method* in which any given grid node is coupled to multiple grid nodes at the subsequent time step). Specifically, the implicit DPP we wish to solve is of the form

$$\begin{aligned} V^{h,\delta}(n\delta,x) = \sup_{a\in\mathbb{A}} \Bigg[& f(n\delta,x,a)\Delta t_n^{h,\delta} + e^{-\beta(n\delta,x,a)\Delta t_n^{h,\delta}} \sum_{y\in\mathscr{R}^h} p^{h,\delta}(n\delta,x;n\delta,y|a)V^{h,\delta}(n\delta,y) \\ & + e^{-\beta(n\delta,x,a)\Delta t_n^{h,\delta}} p(n\delta,x;(n+1)\delta,x|a)V^{h,\delta}((n+1)\delta,x) \Bigg], \end{aligned} \quad (2.22)$$

with the usual terminal condition $V^{h,\delta}(N\delta,x) = g(x)$. Notice that by virtue of implicitness the only non-zero cross temporal transition probability is the one that takes the spatial state into itself. Generalisations to this are quite feasible, although it would take us into the domain of so-called θ-*schemes*.

Before we derive locally consistent expressions for the transition probabilities and the interpolation interval, let us briefly consider how one should go about solving an expression like (2.22). Writing the implicit DPP on the general matrix form

$$\sup_{a\in\mathbb{A}} \left[\mathbf{M}_n^{h,\delta}(a)\mathbf{V}_n^{h,\delta} + \mathbf{q}_n^{h,\delta}(a) \right] = \mathbf{V}_{n+1}^{h,\delta} \quad (2.23)$$

where $\mathbf{V}_n^{h,\delta} := (V^{h,\delta}(n\delta,x_{\min}), V^{h,\delta}(n\delta,x_{\min}+h), ..., V^{h,\delta}(n\delta,x_{\min}+Ih))^\intercal$ and $\mathbf{q}_n^{h,\delta}(a)$ are vectors in $\mathbb{R}^{I+1}$, and $\mathbf{M}_n^{h,\delta}(a)$ is a matrix of transition probabilities in $\mathbb{R}^{(I+1)\times(I+1)}$, the challenges we face become quite apparent. Specifically, qua the supremum operator, (2.23) transcends the garden variety linear system of equations, which allows for immediate computation of $\mathbf{V}_n^{h,\delta}$ in terms of $\mathbf{V}_{n+1}^{h,\delta}$. Rather, the control dependence of $\mathbf{M}_n^{h,\delta}(a)$ and $\mathbf{q}_n^{h,\delta}(a)$ will generally render the LHS highly non-linear in $\mathbf{V}_n^{h,\delta}$ and thus expression (2.23) that much more difficult to solve.

2.6.2.1 Approximations in Policy Space

One of the easiest procedures to overcome the non-linearity issue of (2.23) is indubitably the employ of *policy space iterations*, based on a sequential computation of increasingly more accurate values for $\mathbf{V}_n^{h,\delta}$ and a. Specifically, let $a_0 \in \mathbb{A}$ be some initial admissible feedback policy (a reasonable starting point is to use the optimal control from the subsequent time step, $(n+1)\delta$). Then a first approximation $\mathbf{V}_{n,0}^{h,\delta}$ to $\mathbf{V}_n^{h,\delta}$ may be computed by solving the system

$$\mathbf{M}_n^{h,\delta}(a_0)\mathbf{V}_{n,0}^{h,\delta} + \mathbf{q}_n^{h,\delta}(a_0) = \mathbf{V}_{n+1}^{h,\delta}.$$

Indeed, knowledge of $\mathbf{V}_{n,0}^{h,\delta}$ allows us to compute an updated estimate, a_1, of a based on:[7]

$$a_1 = \underset{a\in\mathbb{A}}{\operatorname{argmax}} \left[\mathbf{M}_n^{h,\delta}(a)\mathbf{V}_{n,0}^{h,\delta} + \mathbf{q}_n^{h,\delta}(a)\right],$$

Clearly, this procedure can now be repeated all over again. Indeed for a general $k = 0,1,2,\ldots$ we may perform the iterative steps

$$\mathbf{M}_n^{h,\delta}(a_k)\mathbf{V}_{n,k}^{h,\delta} + \mathbf{q}_n^{h,\delta}(a_k) = \mathbf{V}_{n+1}^{h,\delta}, \tag{2.24a}$$

$$a_{k+1} = \underset{a\in\mathbb{A}}{\operatorname{argmax}} \left[\mathbf{M}_n^{h,\delta}(a)\mathbf{V}_{n,k}^{h,\delta} + \mathbf{q}_n^{h,\delta}(a)\right]. \tag{2.24b}$$

Under mild conditions it can be shown that $\mathbf{V}_{n,k}^{h,\delta} \to \mathbf{V}_n^{h,\delta}$ as $k \to \infty$.[8] In practical terms, a reasonable place to stop the algorithm is when the value function stops changing notably in successive iterations: e.g. when

$$|\mathbf{V}_{n,k+1}^{h,\delta} - \mathbf{V}_{n,k}^{h,\delta}|_\infty < \varepsilon,$$

where ε is some small positive number.

2.6.2.2 On ps and Δts

Finally, let's expose the procedure by which we obtain the relevant transition probabilities and interpolation interval. Analogously to subsection 2.4.2 we consider an implicit upwind discretisation of (2.14)

$$\begin{aligned}\beta_{n\,\delta} W^{h,\delta}(n\delta,x,a) = {} & \delta^{-1}[W^{h,\delta}((n+1)\delta,x,a) - W^{h,\delta}(n\delta,x,a)] \\ & + [b(n\delta,x,a)]^+ \mathscr{D}_x^+ W^{h,\delta}(n\delta,x,a) - [b(n\delta,x,a)]^- \mathscr{D}_x^- W^{h,\delta}(n\delta,x,a) \\ & + \tfrac{1}{2}\sigma^2(n\delta,x,a)\mathscr{D}_{xx}^2 W^{h,\delta}(n\delta,x,a) + f(n\delta,x,a),\end{aligned} \tag{2.25}$$

where the differencing operators $\mathscr{D}_x^+, \mathscr{D}_x^-$ and $\mathscr{D}_{xx}^2$ are as defined above. This expression can be rearranged as

[7] Needless to say, this step is optimally handled through the FOCs.

[8] See Kusher & Dupuis [31] theorem 6.2.1.

$$\begin{aligned} W^{h,\delta}(n\delta,x,a) = & \frac{\left(h^{-1}[b(n\delta,x,a)]^+ + \frac{1}{2}h^{-2}\sigma^2(n\delta,x,a)\right)}{Q^{h,\delta}(n\delta,x,a)} W^{h,\delta}(n\delta,x+h,a) \\ & + \frac{\left(h^{-1}[b(n\delta,x,a)]^- + \frac{1}{2}h^{-2}\sigma^2(n\delta,x,a)\right)}{Q^{h,\delta}(n\delta,x,a)} W^{h,\delta}(n\delta,x-h,a) \\ & + \frac{\delta^{-1}}{Q^{h,\delta}(n\delta,x,a)} W^{h,\delta}((n+1)\delta,x,a) + f(n\delta,x,a)\frac{1}{Q^{h,\delta}(n\delta,x,a)}, \end{aligned} \tag{2.26}$$

where we have defined

$$Q^{h,\delta}(n\delta,x,a) := \beta(n\delta,x,a) + \delta^{-1} + h^{-1}|b(n\delta,x,a)| + h^{-2}\sigma^2(n\delta,x,a). \tag{2.27}$$

(2.26) is of the form

$$\begin{aligned} W^{h,\delta}(n\delta,x,a) = & \sum_{y\in\mathscr{R}_0^h(x)} p^{h,\delta}(n\delta,x;n\delta,y|a) W^{h,\delta}((n+1)\delta,y,a) \\ & + p^{h,\delta}(n\delta,x;(n+1)\delta,x|a) W^{h,\delta}(n\delta,x,a) + f(n\delta,x,a)\Delta t_n^{h,\delta}, \end{aligned}$$

where $\mathscr{R}_0^h(x) := \{x+h, x-h\}$, which provides us with clear candidates for the transition probabilities and interpolation interval. However, again the associated $p^{h,\delta}(n\delta,x;n\delta,x+h|a)$, $p^{h,\delta}(n\delta,x;n\delta,x-h|a)$ and $p^{h,\delta}(n\delta,x;(n+1)\delta,x|a)$ fail to sum to unity unless $\beta = 0$. To compensate for this fact, we introduce a non-zero probability that the Markov chain stays the same

$$\begin{aligned} p(n\delta,x;n\delta,x|a) &= 1 - \sum_{y\in\mathscr{R}_0^h(x)} p^{h,\delta}(n\delta,x;n\delta,y|a) - p^{h,\delta}(n\delta,x;(n+1)\delta,x|a) \\ &= \frac{\beta(n\delta,x,a)}{Q^{h,\delta}(n\delta,x,a)}. \end{aligned}$$

All in all, we are therefore have

$$p^{h,\delta}(n\delta,x;n\delta,x+h|a) = \frac{h^{-1}[b(n\delta,x,a)]^+ + \frac{1}{2}h^{-2}\sigma^2(n\delta,x,a)}{Q^{h,\delta}(n\delta,x,a)}, \tag{2.28a}$$

$$p^{h,\delta}(n\delta,x;n\delta,x-h|a) = \frac{h^{-1}[b(n\delta,x,a)]^- + \frac{1}{2}h^{-2}\sigma^2(n\delta,x,a)}{Q^{h,\delta}(n\delta,x,a)}, \tag{2.28b}$$

$$p^{h,\delta}(n\delta,x;(n+1)\delta,x|a) = \frac{\delta^{-1}}{Q^{h,\delta}(n\delta,x,a)}, \tag{2.28c}$$

$$\begin{aligned} p^{h,\delta}(n\delta,x;n\delta,x|a) = & 1 - \sum_{y\in\mathscr{R}_0^h(x)} p^{h,\delta}(n\delta,x;n\delta,y|a) \\ & - p^{h,\delta}(n\delta,x;(n+1)\delta,x|a), \end{aligned} \tag{2.28d}$$

and

$$p^{h,\delta}(n\delta,x;n\delta,y|a,n\delta) = 0, \qquad \forall y \notin \mathscr{R}^h(x), \tag{2.28e}$$

along with the interpolation interval

$$\Delta t_n^{h,\delta} = \frac{1}{Q^{h,\delta}(n\delta,x,a)}. \tag{2.29}$$

It is readily seen that these probabilities are non-negative and comply with local consistency (2.21). Nevertheless, it remains computationally somewhat troubling (albeit mathematically correct) that the DPP (2.22) involves an optimisation over all $a \in \mathbb{A}$ of an expression which is heavily control dependent in its numerator and its denominator. To mitigate this tediousness, one may opt for redefining the denominator as the control independent quantity

$$\bar{Q}^{h,\delta}(n\delta,x) := \sup_{a\in\mathbb{A}} Q^{h,\delta}(n\delta,x,a), \tag{2.30}$$

assuming, of course, that $\bar{Q}^{h,\delta}(n\delta,x)$ is finite. If this fails to be the case, one can consider imposing artificial upper bounds on the controls (indeed, this will be the case for the Merton problem treated below - see also Fitzpatrick & Fleming [10]). The resulting transition probabilities are form-invariant expressions with respect to (2.28), albeit with the obvious proviso that (2.28d) will *not* equal $\beta(n\delta,x,a)/\bar{Q}^{h,\delta}(n\delta,x,a)$.

2.7 The Implicit Method and Merton's Problem

2.7.1 Set-up

Recall the definition of $Q^{h,\delta}$, (2.27), which in a Mertonian context this boils down to

$$Q^{h,\delta}(n\delta,x,\theta,c) = \beta + \delta^{-1} + h^{-1}\left(rx + \theta(n\delta,x)(\mu - r) + c(n\delta,x)\right) + h^{-2}\theta^2(n\delta,x,a)\sigma^2.$$

Clearly, the corresponding $\bar{Q}^{h,\delta}$, (2.30), is only finite insofar as we bound the controls from above (again, we will assume that $\forall n \forall x : \theta(n\delta,x), c(n\delta,x) \in [0,Kx]$). With this constraint

$$\bar{Q}^{h,\delta}(n\delta,x) = \beta + \delta^{-1} + h^{-1}\left(rx + Kx(\mu - r) + Kx\right) + h^{-2}K^2x^2\sigma^2.$$

Moreover, to save ourselves the trouble of writing multiple inverted hs in the transition probabilities, define

$$\tilde{Q}^{h,\delta}(n\delta,x) := h^2\bar{Q}^{h,\delta}(n\delta,x),$$

then $\forall x \in \{h, 2h, \ldots, (I-h)h\}$

$$
\begin{aligned}
p^{h,\delta}(n\delta,x;n\delta,x+h|\theta,c) &= \frac{h\big(rx+\theta(n\delta,x)(\mu-r)\big)+\frac{1}{2}\theta(n\delta,x)^2\sigma^2}{\tilde{Q}^{h,\delta}(n\delta,x)},\\
p^{h,\delta}(n\delta,x;n\delta,x-h|\theta,c) &= \frac{hc(n\delta,x)+\frac{1}{2}\theta(n\delta,x)^2\sigma^2}{\tilde{Q}^{h,\delta}(n\delta,x)},\\
p^{h,\delta}(n\delta,x;(n+1)\delta,x|\theta,c) &= \frac{h^2\delta^{-1}}{\tilde{Q}^{h,\delta}(n\delta,x)},\\
p^{h,\delta}(n\delta,x;n\delta,x|\theta,c) &= 1-\sum_{y\in\mathscr{R}_0^h(x)} p^{h,\delta}(n\delta,x;n\delta,y|\theta,c)-p^{h,\delta}(n\delta,x;(n+1)\delta,x|a),\\
p^{h,\delta}(n\delta,x;n\delta,y|\theta,c) &= 0, \qquad \forall y\notin\mathscr{R}^h(x),
\end{aligned}
\tag{2.31}
$$

where $\Delta t_n^{h,\delta}=h^2/\tilde{Q}^{h,\delta}(n\delta,x)$. The main obstacle is again specifying plausible boundary conditions for the lower and upper boundaries. At $x=0$ we maintain that bankruptcy corresponds to a zero-consumption zero-investment strategy at all points in time, i.e. $V^{h,\delta}(0)=0$ with

$$
\begin{aligned}
p(n\delta,0;n\delta,0|\theta,c) &= 1,\\
p(n\delta,0;n\delta,y|\theta,c) &= 0, \qquad \forall y\neq 0.
\end{aligned}
$$

For $x=Ih$ the situation remains less transparent. Following [10] and [15] we make the assumption that there's a vanishing probability of leaving the grid, i.e.

$$
\begin{aligned}
p^{h,\delta}(n\delta,Ih;n\delta,Ih-h|\theta,c) &= \frac{hc(n\delta,Ih)+\frac{1}{2}\theta(n\delta,Ih)^2\sigma^2}{\tilde{Q}^{h,\delta}(n\delta,Ih)},\\
p^{h,\delta}(n\delta,Ih;(n+1)\delta,Ih|\theta,c) &= \frac{h^2\delta^{-1}}{\tilde{Q}^{h,\delta}(n\delta,x)},\\
p^{h,\delta}(n\delta,Ih;n\delta,Ih|\theta,c) &= 1-p(n\delta,Ih;n\delta,Ih-h|\theta,c)-p(n\delta,Ih;(n+1)\delta,Ih|\theta,c),\\
p^{h,\delta}(n\delta,Ih;n\delta,y|\theta,c) &= 0, \qquad \forall y\notin\{Ih-h,Ih\}.
\end{aligned}
\tag{2.32}
$$

Thus, $\forall x\in\{h,2h,...,(I-1)h\}$ we have the implicit DPP

$$
\begin{aligned}
V^{h,\delta}(n\delta,x) = \sup_{(\theta,c)\in\mathbb{R}\times\mathbb{R}_+}\Bigg[&\frac{c(n\delta,x)^{1-\gamma}}{1-\gamma}\frac{h^2}{\tilde{Q}^{h,\delta}(n\delta,x)}+e^{-\frac{\beta h^2}{\tilde{Q}^{h,\delta}(n\delta,x)}}\sum_{y\in\mathscr{R}^h(x)} p^{h,\delta}(n\delta,x;n\delta,y\\
&|\theta,c)V^{h,\delta}(n\delta,y)+e^{-\frac{\beta h^2}{\tilde{Q}^{h,\delta}(n\delta,x)}}p(n\delta,x;(n+1)\delta,x|\theta,c)V^{h,\delta}((n+1)\delta,x)\Bigg],
\end{aligned}
$$

with terminal condition $V^{h,\delta}(N\delta,x)=x^{1-\gamma}(1-\gamma)^{-1}$. An analogous[9] expression holds for $V^{h,\delta}(n\delta,Ih)$ whilst $V^{h,\delta}(n\delta,0)=0$. We solve the DPP using iterations in policy space: specifically, at a given iterative step $k\in\mathbb{N}_0$ we solve the tridiagonal linear system

[9] I.e. with probabilities as in (2.32).

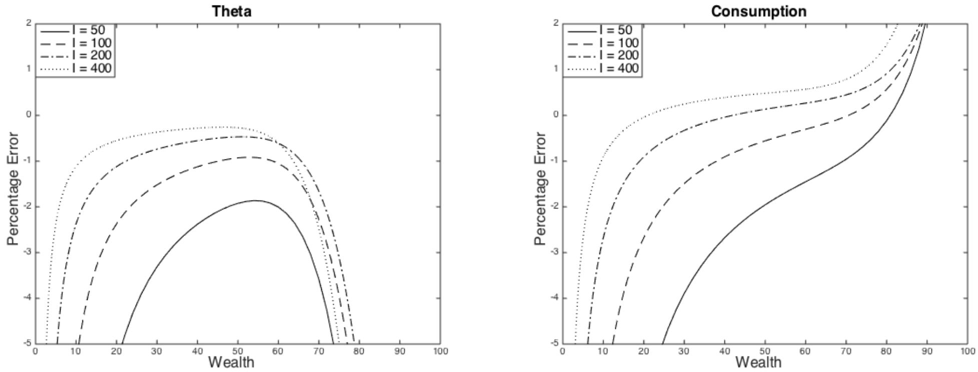

Fig. 2.4 The percentage errors of the numerically computed optimal stock investment (**left**) and the optimal consumption (**right**) computed for various values of I. We assume $T = 1$ and $N = 10$. On average 3 policy iterations per time step are performed.

$$\mathbf{M}_n^{h,\delta}(\theta_k, c_k)\mathbf{V}_{n,k}^{h,\delta} + \mathbf{q}_n^{h,\delta}(c_k) = -h^2\delta^{-1}\mathbf{V}_{n+1}^{h,\delta}, \tag{2.33}$$

where $\mathbf{V}_{n,k}^{h,\delta} \in \mathbb{R}^{I+1}$ is a vector the $(i+1)^{\text{th}}$ component of which is $V_k^{h,\delta}(n\delta, ih)$, and $\mathbf{q}_n^{h,\delta}(c_k) \in \mathbb{R}^{I+1}$ is a vector the first component of which is 0, and more generally, the $(i+1)^{\text{th}} > 1$ component of which is $h^2 \exp\{\beta h^2/\tilde{Q}^{h,\delta}(n\delta, ih)\} c_k(n\delta, ih)^{1-\gamma}/(1-\gamma)$. Furthermore, $\mathbf{M}_n^{h,\delta}(\theta_k, c_k)$ is the $(I+1)\times(I+1)$ tridiagonal matrix

$$\mathbf{M}_n^{h,\delta}(\theta_k, c_k) = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ M_{1,0} & M_{1,1} & M_{1,2} & 0 & \cdots & 0 \\ 0 & M_{2,1} & M_{2,2} & M_{2,3} & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & M_{I,I-1} & M_{I,I} \end{pmatrix},$$

with components

$$\begin{aligned} M_{i,i-1} &= hc_k(n\delta, ih) + \tfrac{1}{2}\theta_k(n\delta, ih)^2\sigma^2, & i \in \mathbb{N}^I \\ M_{i,i+1} &= h\big(rih + \theta_k(n\delta, ih)(\mu - r)\big) + \tfrac{1}{2}\theta_k(n\delta, ih)^2\sigma^2, & i \in \mathbb{N}^{I-1} \\ M_{i,i} &= \tilde{Q}^{h,\delta}(n\delta, ih)(1 - e^{\beta h^2/\tilde{Q}^{h,\delta}(n\delta, ih)}) - M_{i,i-1} - M_{i,i+1} - h^2\delta^{-1}, & i \in \mathbb{N}^{I-1} \\ M_{I,I} &= \tilde{Q}^{h,\delta}(n\delta, Ih)(1 - e^{\beta h^2/\tilde{Q}^{h,\delta}(n\delta, Ih)}) - M_{I,I-1} - h^2\delta^{-1}, & \end{aligned}$$

where $\mathbb{N}^I = \{1, 2, ..., I\}$ and $\mathbb{N}^{I-1} = \{1, 2, ..., I-1\}$. The associated policy update (cf. equation (2.24b)) is performed through the FOCs:

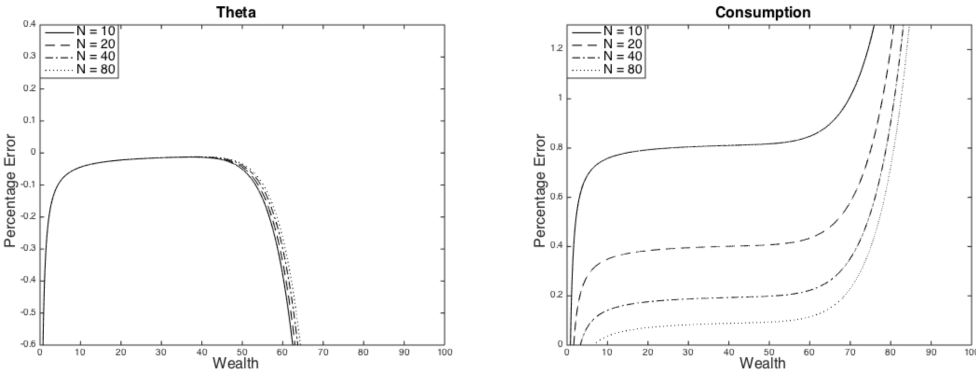

Fig. 2.5 The percentage errors of the numerically computed optimal stock investment (**left**) and the optimal consumption (**right**) computed for various values of N. We assume $T = 1$ and $I = 10^4$.

$$\theta_{k+1}(n\delta, ih) = -\frac{(\mu - r)}{\sigma^2} \frac{\mathscr{D}_x^+ V_k^{h,\delta}((n+1)\delta, ih)}{\mathscr{D}_{xx}^2 V_k^{h,\delta}((n+1)\delta, ih)}, \qquad i \in \mathbb{N}^{I-1}$$

$$c_{k+1}(n\delta, ih) = \left(e^{-\frac{\beta h^2}{\tilde{Q}^{h,\delta}(n\delta, ih)}} \mathscr{D}_x^- V_k^{h,\delta}((n+1)\delta, ih) \right)^{-1/\gamma}, \quad i \in \mathbb{N}^I$$

with all other control values zero: $\theta_{k+1}(n\delta, 0) = \theta_{k+1}(n\delta, Ih) = c_{k+1}(n\delta, 0) = 0$. Again we enforce $\theta(n\delta, ih), c(n\delta, ih) \in [0, KIh]\ \forall n, \forall i$.

Remark 2.4. At a given time step $n\delta < T$, it is opportune to set the initial controls $\theta_0(n\delta, :)$ and $c_0(n\delta, :)$ equal to $\theta^*((n+1)\delta, :)$ and $c^*((n+1)\delta, :)$ respectively.

Remark 2.5. Equation (2.33) can aptly be solved with Thomas' algorithm. If we insist on inverting the matrix, a Gaussian elimination procedure would cost us $\mathscr{O}((I+1)^3)$ binary operations per inversion. On the other hand, if we simply aim to solve the problem (and we do), an $\mathscr{O}(I+1)$ tridiagonal matrix algorithm will do just fine.

2.7.2 Results

We continue to work under the parameter specifications in table 3.2 and set the policy convergence parameter ε to 0.0001. Figure 2.4 plots the percentage errors of the numerically computed optimal controls, for various levels of I, with $T = 1$ and N fixed at 10 ($dt = 0.1$). Despite the comparatively large time steps, we find that the picture is almost identical to the corresponding explicit case, cf. figure 2.1, with the following provisos: (a) There is a small loss in accuracy in the optimal investment strategy between $I = 200$ and $I = 400$. (b) The optimal consumption seems to overshoot the 0% level. Further grid refinements (up to $I = 5 \cdot 10^4$) indicate that the upper boundary does not deteriorate much further. Indeed, if we simultaneously refine the grid along the temporal axis (recall, this is what was done in

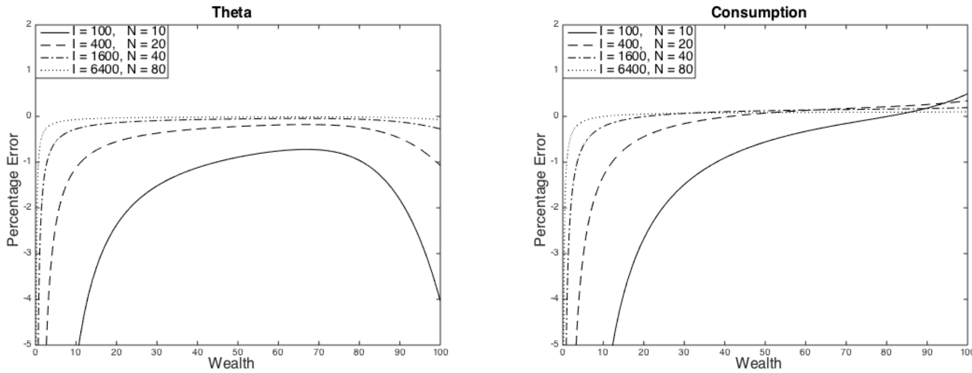

Fig. 2.6 The percentage errors of the numerically computed optimal stock investment (**left**) and the optimal consumption (**right**) computed for various values of I and N, assuming a relational upper boundary, (2.34).

the explicit case in order to satisfy the worst case scenario inequality), the accuracy of the controls will improve. This is vividly illustrated in figure 2.5 for $I = 10^4$: notice in particular the improvements in the numerically computed optimal consumption, which quickly reverts back to the desired zeroth level of percentage error.

Ultimately, figures 2.4 and 2.5 are also a testimony to the fact that whilst increasing I quickly dampens numerical imprecision at low wealth levels, there is relatively little to be gained at the other end of the wealth spectrum. Only by augmenting N as well do we experience some rather modest improvements in the percentage error for the upper boundary. Thus, we are once more left with the impression that our inexact upper boundary effectively kills our chances of numerical accuracy in that region - at least for reasonable levels of computational time. Obviously, providing an analytically exact boundary defeats the very purpose of a numerical routine in the first place, so for most practical purposes it seems we have to *bite the bullet*. A somewhat milder (but certainly not innocuous) strategy would be the deployment of an accurate ansatz for the value function at the upper boundary. Specifically, from the linearity of the wealth dynamics (2.6) it is at least reasonable to conjecture that if (θ^*, c^*) is an optimal control pair at (t,x) then $(k\theta^*, kc^*)$ will be optimal for (t,kx). Thus, from (2.7) we find that $V(t,kx) = k^{1-\gamma}V(t,x)$, which for $k = x^{-1}$ leads to the conjecture $V(t,x) = g(t)^\gamma x^{1-\gamma}/(1-\gamma)$, where $g(t) := (1-\gamma)V(t,1)$ is some function of time only. Enforcing this equation at the upper boundary, it is readily shown that

$$V^{h,\delta}(n\delta,(I+1)h) = \left(1+\frac{1}{I}\right)^{1-\gamma} V^{h,\delta}(n\delta, Ih). \tag{2.34}$$

We call this the *relational boundary*. Thus, at $x = Ih$, rather than using the inward probabilities (2.32) of Fitzpatrick & Fleming, we retain the expressions (2.31): a transition to $V^{h,\delta}(n\delta,(I+1)h)$ is simply handled through (2.34). Similarly, a FOC $(\theta(n\delta, Ih))$ which depends on $V^{h,\delta}(n\delta,(I+1)h)$ can be handled through (2.34). The results speak for themselves: in figure 2.6 we see that we effectively have eradicated numerical imprecision for

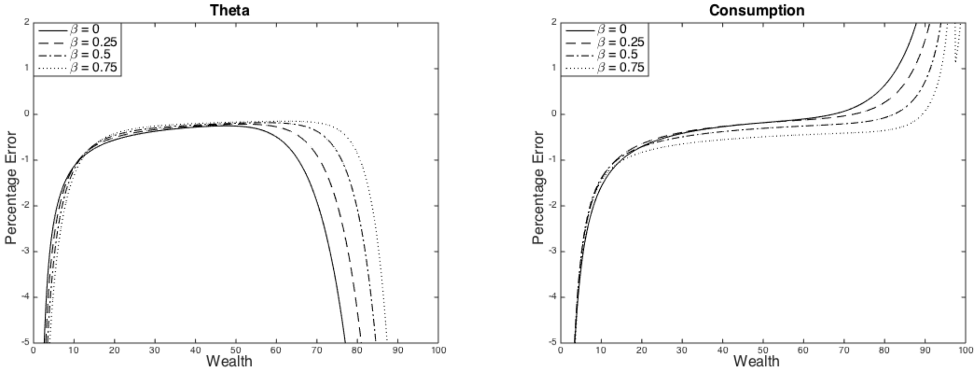

Fig. 2.7 The percentage errors of the numerically computed optimal stock investment (**left**) and the optimal consumption (**right**) computed for various values of β. We assume $T = 1$, $N = 50$ and $I = 400$.

high wealth levels for suitably fine grids. The price we had to pay was the correct assumption that the optimal value function is separable in space and time.

Finally, returning to the original boundary of Fitzpatrick & Fleming, let us say a few words about parametric choices and numerical reliability. In Munk [15] it is demonstrated that the subjective discount factor, β, is heavily correlated with the accuracy of the numerical procedure *in the infinite horizon case*, to the point that $\beta \approx 0.1$ yields deplorably inaccurate numerical controls, while $\beta \approx 0.8$ yields admirably accurate numerical controls. Hence, Munk argues that for economically plausible (i.e. low) values of β that grid must be so designed such that one can ignore "a rather wide neighbourhood of the imposed upper boundary". Having worked with $\beta = 0.02$ throughout this paper, the magnitude of the discounting factor is clearly less of an issue in the finite horizon case. However, as figure 2.7 clearly illustrates, this does not belie the fact that higher values of β generally lead to better numerical accuracy.

A similar conclusion extends to the level of risk aversion, γ, cf. figure 2.8. Low risk aversion ($\gamma \approx 0.35$) leads to numerical instability. On the other hand, a high risk aversion can significantly reduce the inaccuracy at the upper boundary. This is hardly surprising: at $x = Ih$, as γ increases, the *curvature*[10] of utility function (and thence also the optimal value function) decreases, which ultimately entails higher accuracy for the difference operators - particularly $\mathscr{D}^2_{xx}$.

2.8 A Tale From Higher Dimensions

Hitherto our concern has solely been with stochastic control problems with a singular spatial dimension. Prima facie, a generalisation to higher dimensions might seem like a conceptually trivial extension of what has already been covered (modulo the exponen-

[10] Per definitionem, if $f : \mathbb{R} \mapsto \mathbb{R}$ is a $\mathscr{C}^2$ function, then the curvature is given by $\kappa = |f''|/(1+f'^2)^{3/2}$.

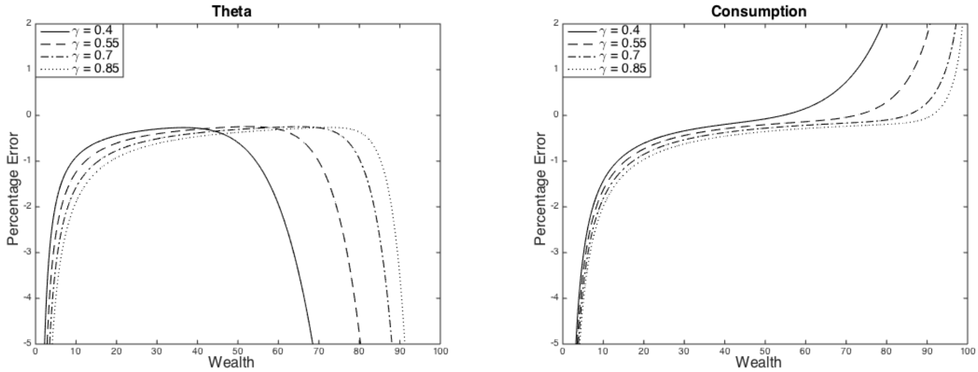

Fig. 2.8 The percentage errors of the numerically computed optimal stock investment (**left**) and the optimal consumption (**right**) computed for various values of γ. We assume $T = 1$, $N = 50$ and and $I = 400$.

tial increase in computational complexity - Bellman's so-called *curse of dimensionality*). Nonetheless, as we shall see, the verisimilitude of this claim rests heavily upon a rather severe constraint on the diffusion matrix, which invariably will bring us into trouble for Merton type problems. For our present purposes, we shall restrict our attention to an *implicit implementation*. The DPP we wish to solve is therefore still of the generic form (2.22), where y now runs over a multi-dimensional space grid.

We reinterpret equation (2.1) as an m-dimensional process, where $b : \mathbb{T} \times \mathbb{R}^m \times \mathbb{A} \mapsto \mathbb{R}^m$, $\sigma : \mathbb{T} \times \mathbb{R}^m \times \mathbb{A} \mapsto \mathbb{R}^{m \times r}$ and W is an $r-$dimensional standard Brownian motion. The PDE satisfied by (2.2) is the multi-dimensional extension of (2.14) viz.

$$\beta_t W = \partial_t W + \sum_{i=1}^{m} b_i(t,x,a)\partial_{x_i} W + \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}(t,x,a)\partial^2_{x_i x_j} W + f(t,x,a), \tag{2.35}$$

where we have defined $x := (x_1, x_2, ..., x_m)^\intercal$ and $S(t,x,a) := \sigma(t,x,a)\sigma^\intercal(t,x,a)$. Again, to procure transition probabilities for the approximating Markov chain $\{\zeta_n^{h,\delta} | n \in \mathbb{N}_0\}$ on the multi-dimensional grid $\mathscr{T}^\delta \times \mathscr{R}^{h_1} \times ... \times \mathscr{R}^{h_m} = \{0, \delta, ..., N\delta\} \times \{x_{1,\min}, x_{1,\min} + h_1, ..., x_{1,\min} + h_1 I_1 =: x_{1,\max}\} \times ... \times \{x_{m,\min}, x_{m,\min} + h_m, ..., x_{m,\min} + h_m I_m =: x_{m,\max}\}$ we discretise (2.35) in the *upwind* sense in analogy with (2.25). Let $\hat{\mathbf{e}}_i = (0, 0, ..., 1_i, 0, ..., 0)^\intercal \in \mathbb{R}^m$ be a unit vector in the direction of the i^{th} spatial dimension, then the relevant difference operators may be stated as

- $\mathscr{D}^+_{x_i} W(t,x,a) := h_i^{-1}[W(t, x + \hat{\mathbf{e}}_i h_i, a) - W(t,x,a)]$ (used if $b_i(t,x,a) \geq 0$).
- $\mathscr{D}^-_{x_i} W(t,x,a) := h_i^{-1}[W(t,x,a) - W(t, x - \hat{\mathbf{e}}_i h_i, a)]$ (used if $b_i(t,x,a) < 0$).
- $\mathscr{D}^2_{x_i x_i} W(t,x,a) = h_i^{-2}[W(t, x + \hat{\mathbf{e}}_i h_i, a) - 2W(t,x,a) + W(t, x - \hat{\mathbf{e}}_i h_i, a)]$.

Furthermore, we introduce the upwind cross derivatives:

- $\mathscr{D}^{+2}_{x_i x_j} W(t,x,a) := (2h_i h_j)^{-1}[2W(t,x,a) + W(t,x+\hat{\mathbf{e}}_i h_i + \hat{\mathbf{e}}_j h_j, a) + W(t,x-\hat{\mathbf{e}}_i h_i - \hat{\mathbf{e}}_j h_j, a) - W(t,x+\hat{\mathbf{e}}_i h_i, a) - W(t,x-\hat{\mathbf{e}}_i h_i, a) - W(t,x+\hat{\mathbf{e}}_j h_j, a) - W(t,x-\hat{\mathbf{e}}_j h_j, a)]$ (used if $i \neq j$ and $S_{ij}(t,x,a) \geq 0$).
- $\mathscr{D}^{-2}_{x_i x_j} W(t,x,a) := -(2h_i h_j)^{-1}[2W(t,x,a) + W(t,x+\hat{\mathbf{e}}_i h_i - \hat{\mathbf{e}}_j h_j, a) + W(t,x-\hat{\mathbf{e}}_i h_i + \hat{\mathbf{e}}_j h_j, a) - W(t,x+\hat{\mathbf{e}}_i h_i, a) - W(t,x-\hat{\mathbf{e}}_i h_i, a) - W(t,x+\hat{\mathbf{e}}_j h_j, a) - W(t,x-\hat{\mathbf{e}}_j h_j, a)]$ (used if $i \neq j$ and $S_{ij}(t,x,a) < 0$).

Using these approximations in the discrete approximation of (2.35) we find (after the usual "unity" rescaling [36]) that the transition probabilities are of the form

$$\begin{aligned} p^{h,\delta}(n\delta,x;n\delta,x\pm\hat{\mathbf{e}}_i h_i|a) = \frac{1}{Q^{h,\delta}(n\delta,x,a)}\bigg(&\frac{[b_i(n\delta,x,a)]^{\pm}}{h_i} \\ &-\sum_{j\neq i}\frac{|S_{ij}(n\delta,x,a)|}{2h_i h_j} + \tfrac{1}{2}\frac{S_{ii}(n\delta,x,a)}{h_i^2}\bigg), \end{aligned} \tag{2.36a}$$

$$p^{h,\delta}(n\delta,x;n\delta,x\pm(\hat{\mathbf{e}}_i h_i + \hat{\mathbf{e}}_j h_j)|a) = \frac{S^+_{ij}(n\delta,x,a)}{2h_i h_j Q^{h,\delta}(n\delta,x,a)} \tag{2.36b}$$

$$p^{h,\delta}(n\delta,x;n\delta,x\pm(\hat{\mathbf{e}}_i h_i - \hat{\mathbf{e}}_j h_j)|a) = \frac{S^-_{ij}(n\delta,x,a)}{2h_i h_j Q^{h,\delta}(n\delta,x,a)} \tag{2.36c}$$

$$p^{h,\delta}(n\delta,x;(n+1)\delta,x|a) = \frac{\delta^{-1}}{Q^{h,\delta}(n\delta,x,a)}, \tag{2.36d}$$

$$\begin{aligned} p^{h,\delta}(n\delta,x;n\delta,x|a) = 1 &- \sum_{y\in\mathscr{R}^h_0(x)} p^{h,\delta}(n\delta,x;n\delta,y|a) \\ &- p^{h,\delta}(n\delta,x;(n+1)\delta,x|a), \end{aligned} \tag{2.36e}$$

where, $p^{h,\delta}(n\delta,x;n\delta,y|a) = 0 \;\forall y \notin \mathscr{R}^h(x)$, and we have defined

$$\begin{aligned} Q^{h,\delta}(n\delta,x,a) := \beta(n\delta,x,a) + \delta^{-1} &+ \sum_{i=1}^m h_i^{-1}|b_i(n\delta,x,a)| \\ + \sum_{i=1}^m h_i^{-2} S_{ii}(n\delta,x,a) &- \sum_{i=1}^m\sum_{j\neq i}(2h_i h_j)^{-1}|S_{ij}(n\delta,x,a)|. \end{aligned} \tag{2.37}$$

Now, $\mathscr{R}^h(x) := \{\forall i \forall j \text{ s.t. } i \neq j | x, x\pm\hat{\mathbf{e}}_i h_i, x\pm\hat{\mathbf{e}}_i h_i \pm \hat{\mathbf{e}}_j h_j\}$, while $\mathscr{R}^h_0(x) := \mathscr{R}^h(x)\backslash\{x\}$. The interpolation interval remains

$$\Delta t_n^{h,\delta} = \frac{1}{Q^{h,\delta}(n\delta,x,a)}.$$

Remark 2.6. Again, it is computationally advantageous to redefine $Q^{h,\delta}(n\delta,x,a)$ without control dependence in analogy with (2.30).

Provided that the formulae in (2.36) can be interpreted as probabilities, they satisfy the basic convergence requirement of being locally consistent in a sense analogous to (2.21).[11] Nonetheless, qua the negative coefficient on $|S_{ij}(n\delta,x,a)|$ there is potentially an issue with respect to keeping the probabilities uniformly non-negative. In fact, a closer look at (2.36a) and (2.37) reveals that a *sufficient condition* for non-negativity is the requirement that $\forall i \in \mathbb{N}$, $\forall x \in \mathbb{R}^m$ and $\forall a \in \mathbb{A}$

$$\frac{1}{h_i} S_{ii}(n\delta,x,a) \geq \sum_{j\neq i} \frac{1}{h_j} |S_{ij}(n\delta,x,a)|. \tag{2.38}$$

The problem with this constraint is that it fails to obtain for a rather large class of problems of interest in mathematical economics. While there in some cases are easy fixes to it, the general case is less encouraging. Consider e.g. the case where S is a matrix independent of (x,a), which fails to satisfy (2.38), then the problem is as simple as rotating the coordinate system until the inequality *is* satisfied. But what about more traditional "asset allocation type" problems such as

$$\begin{aligned} dX_t &= b_1(t,X_t,Y_t,a)dt + \sigma_1(t,X_t,Y_t,a)dW_{1t}, \\ dY_t &= b_2(t,Y_t)dt + \sigma_2(t,Y_t)\{\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}\}, \end{aligned}$$

where ρ is a correlation coefficient between X and Y, and $\sigma_1, \sigma_2 \geq 0$? Assuming a uniformly equidistant grid spacing, condition (2.38) reduces to the diagonal dominance conditions

$$\sigma_1 \geq |\rho|\sigma_2, \quad \text{and} \quad \sigma_2 \geq |\rho|\sigma_1,$$

which generally won't be simultaneously satisfied (for $\rho = \pm 1$ they are always mutually inconsistent unless $\sigma_1 = \sigma_2$ a.s., while they for the diagonal case $\rho = 0$ clearly always are consistent). To combat this, one could try an adaptive discretisation approach *a la* Wang et al. [27], where a suitable differencing scheme is custom picked for each individual grid node, with the aim of securing non-negative probabilities. However, such a search over uniformly positive coefficients will inevitably constitute a non-trivial exercise in coding for dimensionality $m \geq 1$, thus bringing the *practicality* of the approach into doubt. In [16] Kushner likewise describes a method which through the employ of non-local transitions aims to reduce the relative numerical noise of the off-diagonal elements in S - a method he stoically characterises as requiring some flexibility on the part of the programmer.

2.9 A Multi-dimensional Take on the Labour Income Problem

In this section our aim is to test the performance of the multi-dimensional Markov chain approximation method on a bivariate Merton type optimisation problem with a known analytic solution, viz. the case where the investor consumes and invests as before, whilst receiving a stochastic labour income. As it will quickly become apparent, this problems

[11] σ^2 should be read as $\sigma\sigma^\intercal = S$.

fails to satisfy the crucial inequality (2.38), wherefore convergence to the right solution *cannot* be guaranteed. Nonetheless, there is surely some academic interest in scrutinising the accuracy of the multi-dimensional algorithm beyond the assumptions to which it is confined. Accurate results could be an indicator, that reasonably similar optimisation problems can be solved numerically although they fail to satisfy the criterion of positive probabilities.

2.9.1 The Labour Income Problem

We imagine that the Mertonian investor also receives an exogenous stochastic endowment at the rate Y_t, henceforth described as *labour income*. The modified Merton problem may thence be stated as

$$V(t,x,y) = \sup_{(\theta,c)\in\mathscr{A}(t,x,y)} \mathbb{E}_{t,x,y}\left[\int_t^T e^{-\beta(s-t)}u(c_s)ds + e^{-\beta(T-t)}u(X_T^{\theta,c})\right],$$

$$\text{s.t. } dX_s^{\theta,c} = [rX_s^{\theta,c} + \theta_s(\mu-r) + Y_s - c_s]ds + \theta_s\sigma dW_s,$$

where $(X_t^{\theta,c}, Y_t) = (x,y)$, and u (once again) is identified as an isoelastic utility function $u(x) = x^{1-\gamma}/(1-\gamma)$. For simplicity, we take Y to be governed by

$$dY_s = pY_s dt + qY_s dW_s,$$

where p,q are constant parameters, and the random source on Y_t has been chosen to move in lockstep with the stock price process, thus rendering income risk perfectly hedgeable by the traded financial securities. Clearly, this assumption does not find its grounding in empirics, but rather our scheming intentions of deriving a simple closed-form benchmark against which our numerical algorithms can be compared. The key insight is that labour income effectively translates to receiving a "dividend" of magnitude $Y_s ds$ over the time increment $[s, s+ds]$. Hence the present value of the investor's future income (his so-called *human wealth*) can be computed as

$$H(t,y) := \mathbb{E}_{t,y}^{\mathbb{Q}}\left[\int_t^T e^{-r(s-t)}Y_s ds\right],$$

where $\mathbb{Q}$ is the *risk neutral measure* defined through the process $\xi_t := \mathbb{E}_t[d\mathbb{Q}/d\mathbb{P}] = \exp\{-\frac{1}{2}\lambda^2 t - \lambda W_t\}$, $\lambda := (\mu-r)/\sigma$ being the *market price of risk* of the stock. Using the Abstract Bayes' Theorem[12], we may re-express human wealth under the $\mathbb{P}$-measure:

$$H(t,y) = \xi_t^{-1}\mathbb{E}_{t,y}\left[\int_t^T e^{-r(s-t)}\xi_s Y_s ds\right].$$

In particular, since the Y process admits an explicit solution of the form $Y_s = y\exp\{(p - \frac{1}{2}q^2)(s-t) + q(W_s - W_t)\}$ under $\mathbb{P}$, we find, after a few manipulations, that

[12] See Björk [10], proposition B.41.

$$H(t,y) = \begin{cases} \frac{y}{r-p+q\lambda}\left(1 - e^{-(r-p+q\lambda)(T-t)}\right), & \text{if } r-p+q\lambda \neq 0, \\ y(T-t), & \text{if } r-p+q\lambda = 0. \end{cases} \tag{2.39}$$

Effectively, the total wealth of the investor at time t is therefore of the magnitude $x + H(t,y)$, and it makes sense to hypothesise that Merton's optimal consumption and investment strategies are scaled accordingly. Specifically, for the governing HJB equation

$$\beta V = \partial_t V + \sup_{(\theta,c)\in\mathbb{R}\times\mathbb{R}_+} \Big\{ [rx + \theta(\mu - r) + y - c]\partial_x V + \tfrac{1}{2}\theta^2\sigma^2\partial^2_{xx}V + yp\partial_y V + \tfrac{1}{2}y^2q^2\partial^2_y V + y\theta q\sigma\partial^2_{xy}V + \frac{c^{1-\gamma}}{1-\gamma} \Big\},$$

with terminal condition $v(w,y,T) = w^{1-\gamma}/(1-\gamma)$ and associated FOCs

$$\theta^*(t,x,y) = -\frac{(\mu - r)}{\sigma^2}\frac{\partial_x V}{\partial^2_{xx}V} - \frac{yq}{\sigma}\frac{\partial^2_{xy}V}{\partial^2_{xx}V}, \quad \text{and} \quad c^*(t,x,y) = (\partial_x V)^{-1/\gamma}, \tag{2.40}$$

we make the ansatz

$$V(t,x,y) = g(t)^\gamma (w + H(y,t))^{1-\gamma}(1-\gamma)^{-1}, \tag{2.41}$$

in direct analogy with the Merton case, albeit scaled for the auxiliary wealth brought about by labour income. Indeed, upon solving the problem we find that g is defined as in section 2.3, while

$$\theta^*(t,x,y) = \frac{(\mu - r)x}{\gamma\sigma^2} + \frac{H(t,y)}{\sigma}\left(\frac{\mu - r}{\gamma\sigma} - q\right), \quad \text{and} \quad c^*(t,x,y) = \frac{x + H(t,y)}{g(t)}. \tag{2.42}$$

As a sanity check, note that we recover Merton's original control functions upon setting the human wealth equal to zero. For a more detailed exegesis of the problem we refer the reader to Munk [15]. For a study of labour income in incomplete markets see Duffie et al. [7], [8], and Munk [23].

2.9.2 The Implicit Implementation

From the linearity of the human wealth in the y variable (2.39) and equation (2.41), the value function is clearly homogenous of degree $1-\gamma$ in (x,y). In the spirit of Davis et al. [6] the governing HJB equation can therefore be reduced to a single spatial variable, $z = x/y$, which in turn warrants a drastic simplification of the numerical procedure (not to mention positive transition probabilities!). However, as suggested above, there is some interest in putting the multi-dimensional Markov framework to the test despite the issue of negative probabilities: at the very least the implementation procedure is sufficiently non-trivial to deserve some amount of elucidation.

Thus, let us discretise the labour income problem on the three dimensional lattice $\mathscr{T}^{\delta} \times \mathscr{R}^{h_x} \times \mathscr{R}^{h_y} = \{0, \delta, ..., N\delta =: T\} \times \{x_{\min}, x_{\min} + h_x, ..., x_{\min} + Ih_x =: x_{\max}\} \times \{y_{\min}, y_{\min} + h_y, ..., y_{\min} + Jh_y =: y_{\max}\}$. Based on (2.37) and the following (artificial) upper boundary on the controls $\forall n \forall x \forall y : \theta(n\delta, x, y), c(n\delta, x, y) \leq K(x+y)$ we define the control free quantity

$$
\begin{aligned}
\bar{Q}^{h,\delta}(n\delta, x, y) = \delta\beta + 1 + \delta h_x^{-1}\Big(rx + K(x+y)(\mu - r) + y + K(x+y)\Big) + \delta h_y^{-1} py \\
+ \delta h_x^{-2} K^2 (x+y)^2 \sigma^2 + \delta h_y^{-2} y^2 q^2 - \delta h_x^{-1} h_y^{-1} y K(x+y) q\sigma,
\end{aligned}
$$

where we have factored out a δ^{-1}. The associated non-zero transition "probabilities" (2.36) take the form[13]

$$
\begin{aligned}
p^{h,\delta}(n\delta, x, y; n\delta, x+h_x, y|\theta, c) &= \frac{\delta}{\bar{Q}^{h,\delta}}\Big(\frac{rx + \theta(\mu - r) + y}{h_x} - \tfrac{1}{2}\frac{y\theta\sigma q}{h_x h_y} + \tfrac{1}{2}\frac{\theta^2\sigma^2}{h_x^2}\Big), \\
p^{h,\delta}(n\delta, x, y; n\delta, x-h_x, y|\theta, c) &= \frac{\delta}{\bar{Q}^{h,\delta}}\Big(\frac{c}{h_x} - \tfrac{1}{2}\frac{y\theta\sigma q}{h_x h_y} + \tfrac{1}{2}\frac{\theta^2\sigma^2}{h_x^2}\Big), \\
p^{h,\delta}(n\delta, x, y; n\delta, x, y+h_y|\theta, c) &= \frac{\delta}{\bar{Q}^{h,\delta}}\Big(\frac{yp}{h_y} - \tfrac{1}{2}\frac{y\theta\sigma q}{h_x h_y} + \tfrac{1}{2}\frac{y^2 q^2}{h_y^2}\Big), \\
p^{h,\delta}(n\delta, x, y; n\delta, x, y-h_y|\theta, c) &= \frac{\delta}{\bar{Q}^{h,\delta}}\Big(-\tfrac{1}{2}\frac{y\theta\sigma q}{h_x h_y} + \tfrac{1}{2}\frac{y^2 q^2}{h_y^2}\Big), \\
p^{h,\delta}(n\delta, x, y; n\delta, x+h_x, y+h_y|\theta, c) &= \frac{\delta y\theta q\sigma,}{2\bar{Q}^{h,\delta} h_x h_y} \\
p^{h,\delta}(n\delta, x, y; n\delta, x-h_x, y-h_y|\theta, c) &= \frac{\delta y\theta q\sigma}{2\bar{Q}^{h,\delta} h_x h_y}, \\
p^{h,\delta}(n\delta, x, y; (n+1)\delta, x, y|\theta, c) &= \frac{1}{\bar{Q}^{h,\delta}}, \\
p^{h,\delta}(n\delta, x, y; n\delta, x, y|\theta, c) &= 1 - (\text{sum of the probabilities above}),
\end{aligned}
$$

where we for notational simplicity have suppressed the arguments on $\bar{Q}$, θ and c. The interpolation interval is $\Delta t_n^{h,\delta} = \delta / \bar{Q}^{h,\delta}(n\delta, x, y)$. Notice that condition (2.38) simultaneously requires $yqh_x \leq \theta\sigma h_y$ and $\theta\sigma h_y \leq yqh_x$, which obviously fails to be the case. Negative probabilities are therefore to be expected.

Remark 2.7. As for the boundary conditions (the four surfaces of the grid characterised by $x = x_{\min}, x = x_{\max}, y = y_{\min}$ and $y = y_{\max}$), we adopt a relational boundary approach in analogy with (2.34). From (2.41) we find that

$$
V^{h,\delta}(n\delta, x_{\max} + h_x, y_{\min} + h_y j) = G_{+,j}^{n} V^{h,\delta}(n\delta, x_{\max}, y_{\min} + h_y j).
$$

where we have introduced the notation

[13] Again, we have made use of splitting of the operator.

$$G^n_{+,j} := \left(1 + \frac{h_x}{x_{\max} + H(n\delta, y_{\min} + jh_y)}\right)^{1-\gamma}.$$

(Think of $G^n_{+,j}$ as an operator which raises the index $i = I$ on the operand by *plus* one, whilst keeping the j index as is). On the other side of the grid, we can introduce $G^n_{-,j}$ such that $G^n_{-,j}V(n\delta, x_{\min}, y_{\min} + jh_y) = V(n\delta, x_{\min} - h_x, y_{\min} + jh_y)$ (the operator $G^n_{-,j}$ now lowers the index $i = 0$ of the operand by *minus* one). Analogous coefficients are found for the y-variable: $G^n_{i,+}$ and $G^n_{i,-}$, as for the edges of the grid G^n_{++}, G^n_{+-}, G^n_{-+}, G^n_{--} (all through the employ of (2.41)).

Again, the trick to solving the DPP for the labour income problem

$$\begin{aligned}
V^{h,\delta}(n\delta, x) = \sup_{(\theta,c)\in\mathbb{R}\times\mathbb{R}_+} \Bigg[& \frac{c(n\delta,x,y)^{1-\gamma}}{1-\gamma} \frac{\delta}{\bar{Q}^{h,\delta}(n\delta,x,y)} \\
& + e^{-\frac{\beta\delta}{\bar{Q}^{h,\delta}(n\delta,x,y)}} \sum_{(z,z')\in\mathscr{R}^h(x,y)} p^{h,\delta}(n\delta,x,y;n\delta,z,z'|\theta,c)V^{h,\delta}(n\delta,z,z') \\
& + e^{-\frac{\beta\delta}{\bar{Q}^{h,\delta}(n\delta,x,y)}} p(n\delta,x,y;(n+1)\delta,x,y|\theta,c)V^{h,\delta}((n+1)\delta,x,y) \Bigg],
\end{aligned}$$

is to rewrite the system on the form

$$\mathbf{M}^{h,\delta}_n(\theta_k, c_k)\mathbf{V}^{h,\delta}_{n,k} + \mathbf{q}^{h,\delta}_n(c_k) = \mathbf{V}^{h,\delta}_{n+1},$$

which can be solved using iterations in policy space. In this connection it is computationally advantageous to think carefully about the ordering the set of indices $\{(i,j)|0 \le i \le I, 0 \le j \le J\}$. At the very least one should choose an ordering which renders $\mathbf{M}^{h,\delta}_n$ *block tridiagonal* and optimally also one which minimises the associated *bandwidth* (essentially: the size of the individual blocks). A particularly natural choice is

$$(i,j) \in \{(0,0),(0,1),...,(0,J);(1,0),(1,1),...,(1,J);...;(I,0),(I,1),...,(I,J)\}, \quad (2.43)$$

which yields the block tridiagonal matrix system

$$\begin{pmatrix}
\bar{\mathbf{M}}^n_0 \boldsymbol{\Gamma}^n_0 + \hat{\mathbf{M}}^n_0 & \tilde{\mathbf{M}}^n_0 & 0 & 0 & \cdots & 0 \\
\bar{\mathbf{M}}^n_1 & \hat{\mathbf{M}}^n_1 & \tilde{\mathbf{M}}^n_1 & 0 & \cdots & 0 \\
0 & \bar{\mathbf{M}}^n_2 & \hat{\mathbf{M}}^n_2 & \tilde{\mathbf{M}}^n_2 & \cdots & 0 \\
\vdots & & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \bar{\mathbf{M}}^n_I & \hat{\mathbf{M}}^n_I + \tilde{\mathbf{M}}^n_I \boldsymbol{\Gamma}^n_I
\end{pmatrix}
\begin{pmatrix} \mathbf{V}^n_0 \\ \mathbf{V}^n_1 \\ \mathbf{V}^n_2 \\ \vdots \\ \mathbf{V}^n_I \end{pmatrix}
+ \begin{pmatrix} \mathbf{q}^n_0 \\ \mathbf{q}^n_1 \\ \mathbf{q}^n_2 \\ \vdots \\ \mathbf{q}^n_I \end{pmatrix}
= \begin{pmatrix} \mathbf{V}^{n+1}_0 \\ \mathbf{V}^{n+1}_1 \\ \mathbf{V}^{n+1}_2 \\ \vdots \\ \mathbf{V}^{n+1}_I \end{pmatrix}, \quad (2.44)$$

where $\mathbf{V}_i^n$ and $\mathbf{q}_i^n$ are $(J+1)$-dimensional vectors the $(j+1)^{\text{th}}$ components of which respectively amount to $V_k^{h,\delta}(n,i,j)$ and

$$\delta \exp\{\beta\delta/\bar{Q}^{h,\delta}(n,i,j)\}\frac{c_k(n,i,j)^{1-\gamma}}{1-\gamma},$$

for $i = 0,1,2,...,I$ and $j = 0,1,2,...,J$. For ease of notation, we have here defined

$$(n,i,j) := (n\delta, x_{\min} + ih_x, y_{\min} + jh_y).$$

As for the block components, we define the $(J+1)\times(J+1)$ blocks $\boldsymbol{\Gamma}_0^n := \operatorname{diag}(G_{-,0}^n, G_{-,1}^n, ..., G_{-,J}^n)$, $\boldsymbol{\Gamma}_I^n := \operatorname{diag}(G_{+,0}^n, G_{+,1}^n, ..., G_{+,J}^n)$, alongside the $\mathbf{M}$s which generally are tridiagonal matrices

$$ {}^x\mathbf{M}_i^n := \begin{pmatrix} {}^xa_{i,0}^n\,{}^x\gamma_{i,0}^n + {}^xb_{j,0}^n & {}^xc_{i,0}^n & 0 & 0 & \cdots & 0 \\ {}^xa_{i,1}^n & {}^xb_{i,1}^n & {}^xc_{i,1}^n & 0 & \cdots & 0 \\ 0 & {}^xa_{i,2}^n & {}^xb_{i,2}^n & {}^xc_{i,2}^n & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 0 & {}^xa_{i,J}^n & {}^xb_{i,J}^n + {}^x\gamma_{i,J}^n\,{}^xc_{i,J} \end{pmatrix}, $$

where the superscript x represents a "bar","hat" or "tilde" over the following object, and ${}^xa$, ${}^xb$, and ${}^xc$ are coefficients to be spelled out momentarily. Again, for relational boundary purposes, we have introduced the quantities $\gamma_{i,0}^n := G_{i-1,-}^n$, $\hat{\gamma}_{i,0}^n := G_{i,-}^n$, $\tilde{\gamma}_{i,0}^n := G_{i+1,-}^n$, and $\gamma_{i,J}^n := G_{i-1,+}^n$, $\hat{\gamma}_{i,J}^n := G_{i,+}^n$, $\tilde{\gamma}_{i,J}^n := G_{i+1,+}^n$ for all i - except the four corner cases where $\gamma_{0,0}^n := G_{-,-}^n/G_{-,0}^n$, $\tilde{\gamma}_{I,0}^n := G_{+,-}^n/G_{+,0}^n$, $\gamma_{0,J}^n := G_{-,+}^n/G_{-,J}^n$, and $\tilde{\gamma}_{I,J}^n := G_{+,+}^n/G_{+,J}^n$. Finally, a few short manipulations of the DPP reveal that

$$\begin{aligned}
\bar{a}_{i,j}^n &= -\frac{\delta y\theta_k(n,i,j)q\sigma}{2h_xh_y}, \\
\bar{b}_{i,j}^n &= \delta\Big(-\frac{c_k(n,i,j)}{h_x} + \tfrac{1}{2}\frac{y\theta_k(n,i,j)\sigma q}{h_xh_y} - \tfrac{1}{2}\frac{\theta_k(n,i,j)^2\sigma^2}{h_x^2}\Big), \\
\bar{c}_{i,j}^n &= 0, \\
\hat{a}_{i,j}^n &= \delta\Big(\tfrac{1}{2}\frac{y\theta_k(n,i,j)\sigma q}{h_xh_y} - \tfrac{1}{2}\frac{y^2q^2}{h_y^2}\Big), \\
\hat{b}_{i,j}^n &= \bar{Q}^{h,\delta}(e^{\beta\delta/\bar{Q}^{h,\delta}(n,i,j)} - 1) - \bar{a}_{i,j}^n - \bar{b}_{i,j}^n - \hat{a}_{i,j}^n - \hat{c}_{i,j}^n - \tilde{b}_{i,j}^n - \tilde{b}_{i,j}^n + 1, \\
\hat{c}_{i,j}^n &= \delta\Big(-\frac{yp}{h_y} + \tfrac{1}{2}\frac{y\theta_k(n,i,j)\sigma q}{h_xh_y} - \tfrac{1}{2}\frac{y^2q^2}{h_y^2}\Big) \\
\tilde{a}_{i,j}^n &= 0, \\
\tilde{b}_{i,j}^n &= \delta\Big(-\frac{rx + \theta_k(n,i,j)(\mu - r) + y}{h_x} + \tfrac{1}{2}\frac{y\theta_k(n,i,j)\sigma q}{h_xh_y} - \tfrac{1}{2}\frac{\theta_k(n,i,j)^2\sigma^2}{h_x^2}\Big),
\end{aligned}$$

$$\tilde{c}^{n}_{i,j} = -\frac{\delta y \theta_k(n,i,j) q \sigma}{2 h_x h_y},$$

$\forall i \forall j$. From the DPP we also find the discretised FOCs

$$\theta_{k+1}(n,i,j) = -\frac{(\mu - r)}{\sigma^2} \frac{\mathscr{D}^{+}_{x} V^{h,\delta}_{k}(n,i,j)}{\mathscr{D}^{2}_{xx} V^{h,\delta}_{k}(n,i,j)} - \frac{(y_{\min} + i h_y) q}{\sigma} \frac{\mathscr{D}^{+2}_{xy} V^{h,\delta}_{k}(n,i,j)}{\mathscr{D}^{2}_{xx} V^{h,\delta}_{k}(n,i,j)},$$
$$c_{k+1}(n,i,j) = \left(e^{-\frac{\beta\delta}{\bar{Q}^{h,\delta}(n,i,j)}} \mathscr{D}^{-}_{x} V^{h,\delta}_{k}(n,i,j) \right)^{-1/\gamma},$$

through which we update our controls (compare these expressions with (2.40)). Again, the boundaries $i = 0, i = I, j = 0$ and $j = J$ are handled through the relational operators and the all controls are suitably constrained.

Remark 2.8. The central idea behind writing the DPP on block tridiagonal form, (2.44), is that we can solve the system through a generalised block form of Thomas' algorithm. Were we to invert the Brobdingnagian $(I+1)(J+1) \times (I+1)(J+1)$ matrix using a Gaussian procedure we would incur $\mathscr{O}((I+1)^3(J+1)^3)$ binary operations. With the generalised Thomas algorithm this number we can solve the system with $\mathscr{O}((I+1)(J+1)^3)$ operations. Incidentally, this power asymmetry highlights the importance of choosing an ordering which keeps the bandwidth minimal: essentially, the ordering (2.43) is only to be preferred if $J \leq I$. If $J > I$, it would be better to have a system with blocks of dimensionality $(I+1) \times (I+1)$ as per the ordering $(i,j) \in \{(0,0),(1,0),...,(I,0);(1,1),(2,1),...,(I,1);...;(1,J),(2,J),...,(I,J)\}$. Hence order $\mathscr{O}((J+1)(I+1)^3)$ in complexity. Further details are provided in the appendix.

2.9.3 Results

We assume the same parameters as in table 3.2. Furthermore, let the drift and diffusion of the income process be $p = 0.04$ and $q = 0.1$ respectively, and let the state space be restricted to $[x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}] = [20,100] \times [10,50]$. Figures 2.9 and 2.10 plot the percentage errors of the numerically computed optimal controls for various values of I, J, and N. Considering that the algorithm comes with no guarantee of convergence, the results are (for the current parametric choices) pleasantly accurate. Part of this success story must be ascribed to the use of relational boundaries. We experimented with various Dirichlet alternatives and found that even mild perturbations could wreck considerable havoc. It is interesting to note that increasing N not necessarily improves the accuracy of the algorithm. Furthermore, while the illustrated surfaces seemingly become increasingly more accurate as we increase the spatial fine-graining, numerical experiments also indicate that there is an upper threshold for I, J beyond which the numerical controls can become exceedingly inaccurate. Jointly, these facts clearly point in the direction that the algorithm does *not* converge, which hardly is surprising given that our negative probabilities violate

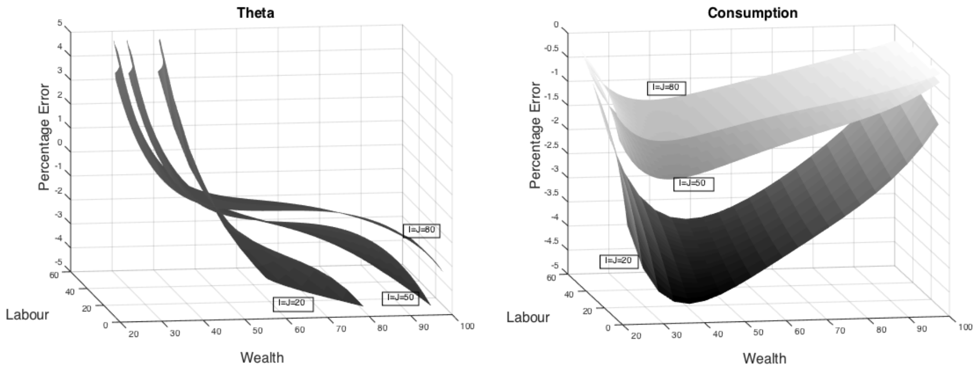

Fig. 2.9 The percentage errors of the numerically computed optimal stock investment (**left**) and the optimal consumption (**right**) for various values of I and J (i.e. various values of spatial grid refinements). We assume $T = 1$ and $N = 50$.

the monotonicity property. Nonetheless, this does not mean that the procedure is altogether useless: very reasonable results can be obtained for a wide range of parameters (the percentage errors come in at less than 2% for the dominant part of the grid). Thus, if everything else fails, a multi-dimensional Markov chain approximation with negative probabilities can still serve as a beacon in the night for the working economist. Ultimately, one must experiment with different grid specifications to test the robustness of the numerical results: indeed, ask oneself the dangerously nebulous question: are my results *reasonable*?

2.10 Conclusion

Our path through the numerical landscape of financial control theory has been long and winded, and it is well worth summarising some of our key findings. The main advantage of the *explicit* Markov chain approximation is the fact that it is straight-forward to implement. Alas, simplicity comes at the cost of satisfying the probabilistic positivity requirement (2.17), which for Merton type problem entails exceedingly small time steps. For a grid interpretation of the explicit method this is a nuisance, but does not cause difficulties in principle. As for the trinomial interpretation, a "worst case scenario" take on the inequality may altogether rule out the existence of a converging solution. To sidestep this issue, one may opt for the *implicit* Markov chain approximation instead. Whilst this prima facie calls for a solution to a highly non-linear system of equations, one may invoke the *iterations in policy space algorithm* (2.24) in order to render the system (iteratively) linear and thence susceptible to a tridiagonal matrix algorithm. Irrespective of the procedure being implemented a further key insight pertains to the immense computational benefit of using discretised FOCs upon updating the controls. Searches over discrete control spaces are better avoided (or should at the very least be restricted to locally plausible regions based upon subsequent values of optimality).

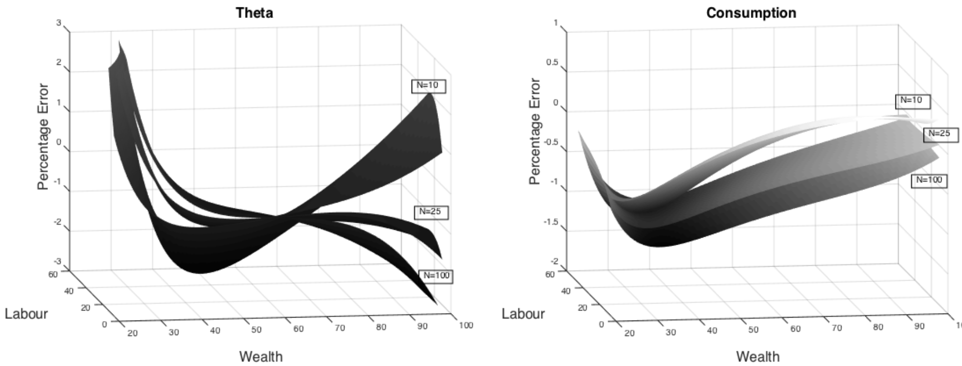

Fig. 2.10 The percentage errors of the numerically computed optimal stock investment (**left**) and the optimal consumption (**right**) for various values of N (i.e. various values of temporal grid refinements). We assume $T = 1$, $I = 100$, and $J = 100$.

In a Mertonian context, both procedures were found to be wildly inaccurate near the upper and lower boundaries. Considering the general opaqueness of what constitutes adequate boundaries this is hardly surprising: however, it is at least somewhat unsettling that the numerical accuracy near the upper boundary seemingly benefits so very little from increasing the grid refinement (and invariably, the computational run-time). Further numerical studies indicated that one may take some measures against this by choosing appropriate parameters (essentially, high values for β and γ), although this admittedly may fly in the face of empirical data. A better (but admittedly also bolder) move is that of introducing a relational upper boundary (2.34) based on the correct ansatz of time-space separability of the value function. This solved the problem satisfactorily even for relatively coarse grained grid specifications.

Finally, while a multi-dimensional extensions to the Markov chain approximation are theoretically possible, keeping the transition probabilities non-negative for Merton type problems proves to be considerably more complex. Given the relevance of this class of problems, and the absence of easy fixes, we put the algorithm to the test by studying the (non-spatially reduced) labour income problem, in which negative transitions are manifest. The numerically computed optimal controls turned out to be surprisingly accurate, despite the absence of algorithmic monotonicity. However, the performance for various parameters and grid specifications is also sufficiently erratic, in order for the method to be deemed potentially dangerous.

References

1. Almgren, *Optimal Trading with Stochastic Liquidity and Volatility*, *SIAM J. FINANCIAL MATH.*, Vol. 3, pp. 163-181.
2. Avellaneda and Stoikov, *High-frequency Trading in a Limit Order Book*, *Quantitative Finance*, Vol. 8, No. 3, April 2008, 217-224.
3. Björk, *Arbitrage Theory in Continuous Time*, Oxford University Press, 3rd edition.
4. Burden and Faires, *Numerical Analysis*. Brooks/Cole, 9th edition.
5. Davis and Norman, *Portfolio Selection with Transaction Costs*. *Mathematics of Operations Research*, 1990.15, 676-713.
6. Duffie, *Dynamic Asser Pricing Theory*. Princeton University Press, 3rd edition.
7. Duffie and Zariphopoulou, *Optimal Investment and Undiversifiable Income Risk*, *Mathematical Finance*, Vol. 3, No. 2 (April 1993), 135-148.
8. Duffie, Fleming, Soner, and Zariphopoulou, *Hedging in Incomplete Markets with HARA Utility*, *Journal of Economic Dynamics and Control*, 21 (1997) 753-782.
9. Ellersgaard, *On Thomas' Algorithm: A Study of Efficient Solutions to Finite Difference Problems in Mathematical Finance*. Unpublished manuscript. Available at https://sellersgaard.files.wordpress.com/2013/09/thomasalgorithm.pdf
10. Fitzpatrick and Fleming, *Numerical Methods for an Optimal Investment-Consumption Model*, *Mathematics of Operations Research*. Vol. 16, No. 4, November 1991.
11. Fleming and Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd edition. *Springer*. ISBN-10: 0-387-26045-5, ISBN-13: 978-0387-260457.
12. Forsyth and Labahn, *Numerical methods for controlled Hamilton-Jacobi-Bellman PDEs in finance*, *Journal of Computational Finance*. 11:2 (2007/2008: Winter) 1-44.
13. Haug, *Why so Negative to Negative Probabilities?*, *Wilmott Magazine*, (2004) Sep/Oct, pp. 34-38. http://www.espenhaug.com/NegativeProbabilitiesHaug.pdf
14. Kushner and Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd edition. Springer. ISBN 0-387-95139-3.
15. Kushner, *Numerical Methods for Stochastic Control Problems in Continuous Time*, *SIAM J. Control and Optimization*. Vol. 28, No. 5, pp. 999-1048, September 1990.
16. Kushner, *Consistency Issues for Numerical Methods for Variance Control, with Applications to Optimization in Finance*, *IEEE Transactions On Automtic Control*, Vol. 44, No. 12, December 1999.
17. Ludwig, Sirignano, Huang, and Papanicolaou, *A Forward-Backward Algorithm for Stochastic Control Problems*, http://web.stanford.edu/ jasirign/pdf/FBSDE-Algorithm_v3.pdf.
18. Meissner and Burgin, *Negative Probabilities in Financial Modeling*. Working paper (February 28, 2011). http://ssrn.com/abstract=1773077 or http://dx.doi.org/10.2139/ssrn.1773077.
19. Merton, *Continuous-Time Finance*. *Blackwell*. ISBN 0-631-15847-2, ISBN 0-631-18508-9.
20. Merton, *Lifetime Portfolio Selection under Uncertainty: the Continuous-Time Case*. *The Review of Economics and Statistics* 51 (3) pp. 247-257 (1969).
21. Munk, *Dynamic Asset Allocation*. Unpublished lecture notes. 2013. http://mit.econ.au.dk/vip_htm/cmunk/noter/dynassal.pdf.
22. Munk, *The Markov Chain Approximation Approach for Numerical Solution of Stochastic Control Problems: Experiences from Merton's Problem*. *Applied Mathematics and Computation*, 2003,136(1): pp. 47-77.
23. Munk, *Optimal Consumption/Investment Policies with Undiversifiable Income Risk and Liquidity Constraints* . *Journal of Economic Dynamics and Control,* , 2000, 24(9): pp. 1315-1343.
24. Munk, *Numerical Methods for Continuous-Time, Continuous-State Stochastic Control Problems*, Working paper. https://dl.dropboxusercontent.com/u/65974535/MyPapers/MCHAIN.pdf.
25. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications*. *Stochastic Modelling and Applied Probability 61*, Springer. e-ISBN: 978-3-540-89500-8.
26. Ross, *Stochastic Control in Continuous Time*. Unpublished lecture notes. http://www.swarthmore.edu/NatSci/kross1/Stat220notes.pdf.
27. Wang and Forsyth, *Maximal Use of Central Differencing for Hamilton-Jacobi-Bellman PDEs in Finance*, *SIAM J. Numer. Anal.*, 46(3), 1580?1601.

28. Zariphopoulou, *A Solution Approach to Valuation with Unhedgeable Risks*, *Finance and Stochastics*, 5, pp. 61–82, 2001.
29. Zvan, Forsyth, and Vetzal, *Negative coefficients in two-factor option pricing models*, *Journal of Computational Finance*, 01 Oct 2003.

Appendix A: The Generalised Thomas Algorithm

First, recall the fundamentals of Thomas' algorithm: suppose we have a matrix system of the form

$$\mathbf{Ax} = \mathbf{y},$$

where $\mathbf{x},\mathbf{y} \in \mathbb{R}^I$ and $\mathbf{A} \in \mathbb{R}^{I\times I}$ is a tridiagonal matrix. Insofar as we wish to solve for $\mathbf{x}$ we could perform an explicit matrix inversion of $\mathbf{A}$, thus incurring an order of $\mathscr{O}(I^3)$ binary operations in the process. However, it is much more sensible to exploit the tridiagonality of $\mathbf{A}$ by first decomposing the matrix into $\mathbf{LU}$ form, where $\mathbf{L},\mathbf{U} \in \mathbb{R}^{I\times I}$ are lower and upper triangular matrices (the only non-zero elements of which are in the leading diagonal and the 'next-to" leading diagonal). Solving for $\mathbf{x}$ in $\mathbf{LUx} = \mathbf{y}$, is now a two-step process: first, we solve for $\mathbf{x}' \in \mathbb{R}^J$ in the system

$$\mathbf{Lx}' = \mathbf{y},$$

simply by working our way row-by-row downwards through the system. Secondly, we solve for the desired object $\mathbf{x}$ in the system

$$\mathbf{Ux} = \mathbf{x}',$$

by working our way row-by-row upwards through the system. The total number of binary system incurred in the process is of $\mathscr{O}(I)$ - a considerable simplification over the Gaussian algorithm.

The Extension

Unsurprisingly perhaps, the exact same principles can be extended to matrix systems of a block tridiagonal nature (the only caveat being the fact that matrix multiplication is manifestly non-commutative whence one must take slightly more care in the derivation). Specifically consider the case where $\mathbf{Ax} = \mathbf{y}$ is a matrix system such that $\mathbf{x},\mathbf{y} \in \mathbb{R}^{I\cdot J}$ and $\mathbf{A} \in \mathbb{R}^{I\cdot J\times I\cdot J}$ is a block tridiagonal matrix, with $J \times J$-dimensional block components ${}^x\mathbf{A}_i$ where $x \in \{[\text{blank}], \wedge, \sim\}$:

$$\mathbf{A} := \begin{pmatrix} \hat{\mathbf{A}}_1^n & \tilde{\mathbf{A}}_1^n & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A}_2^n & \hat{\mathbf{A}}_2^n & \tilde{\mathbf{A}}_2^n & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_3^n & \hat{\mathbf{A}}_3^n & \tilde{\mathbf{A}}_3^n & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{A}_{J-1}^n & \hat{\mathbf{A}}_{J-1}^n & \tilde{\mathbf{A}}_{J-1}^n \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_J^n & \hat{\mathbf{A}}_J^n \end{pmatrix}.$$

We now perform a block **LU** decomposition, to which end we introduce the lower and upper triangular block matrices:

$$\mathbf{L} := \begin{pmatrix} \mathbb{I} & 0 & 0 & 0 & \dots & 0 \\ \mathbf{L}_2 & \mathbb{I} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{L}_3 & \mathbb{I} & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & 0 & \mathbf{L}_{I-1} & \mathbb{I} & 0 \\ 0 & \dots & 0 & 0 & \mathbf{L}_I & \mathbb{I} \end{pmatrix}, \quad \mathbf{U} := \begin{pmatrix} \mathbf{H}_1 & \mathbf{U}_1 & 0 & 0 & \dots & 0 \\ 0 & \mathbf{H}_2 & \mathbf{U}_2 & 0 & \dots & 0 \\ 0 & 0 & \mathbf{H}_3 & \mathbf{U}_3 & & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & & 0 & 0 & \mathbf{H}_{I-1} & \mathbf{U}_{I-1} \\ 0 & \dots & 0 & 0 & 0 & \mathbf{H}_I^n \end{pmatrix},$$

where $\mathbf{L}_i, \mathbf{H}_i$ and $\mathbf{U}_i$ are in $\mathbb{R}^{J \times J}$ and $\mathbb{I}$ is the identity matrix of the same space. Thomas' generalised algorithm then boils down to the pseudo-code exhibited in the algorithm below in direct analogy with the simple tridiagonal matrix case.

Remark 2.9. Suppose the matrix blocks in turn are tridiagonal matrices, as it will be the case for PDE problems with two spatial variables and purely local transitions. Can one imbed a Thomas algorithm within Thomas' generalised algorithm to solve for the $\mathbf{L}_i$s and $\mathbf{y}_i$s (thus sidestepping an explicit inversion of the $\mathbf{H}_i$s)? Emphatically: *no*. This follows from the simple fact that the $\mathbf{H}_i$s generally won't be tridiagonal beyond the case $i = 1$.

Complexity

We may now compute the reduction in computational complexity as follows: from the I full matrix inversions of the $\mathbf{H}_i$s we incur from Gaussian elimination: $I \cdot (\frac{1}{3}J^3 + J^2 + \frac{1}{3}J)$ multiplications & divisions and $I \cdot (\frac{1}{3}J^3 + \frac{1}{2}J^2 - \frac{5}{6}J)$ additions & subtractions. The remaining multiplications & divisions in the algorithm amount to $5IJ - 4J$, whilst the remaining additions & subtractions amount to $3IJ - 3J$. Thus, under the assumption that $J \leq I$, the overall complexity of the algorithm scales as $\mathscr{O}(IJ^3)$ - a substantial improvement over Gauss' $\mathscr{O}(I^3J^3)$.[14] Insofar as $I \leq J$ we could obtain the "mirror" result $\mathscr{O}(JI^3)$ by choosing the ordering such that the block matrices have dimensionality $I \times I$ instead. For further details about tridiagonal matrix methods in mathematical finance we refer the reader to Ellersgaard [15] or Appendix B in this dissertation.

[14] In actual implementations of Thomas' generalised algorithm one might find that it is appropriate to manipulate the indices, which will engender further binary operations. However, the totality of these prove to be considerably smaller than the leading order of magnitude.

Thomas' Generalised Algorithm

$$
\begin{array}{llcl}
\text{Set} & \mathbf{H}_1 & := & \hat{\mathbf{A}}_1; \\
 & \mathbf{U}_1 & := & \tilde{\mathbf{A}}_1; \\
\text{For } i = 2, \ldots, I-1 \text{ set} & \mathbf{L}_i & := & \mathbf{A}_i [\mathbf{H}_{i-1}]^{-1}; \\
 & \mathbf{H}_i & := & \hat{\mathbf{A}}_i - \mathbf{L}_i \mathbf{U}_{i-1}; \\
 & \mathbf{U}_i & := & \tilde{\mathbf{A}}_i; \\
\text{Set} & \mathbf{L}_I & := & \mathbf{A}_I [\mathbf{H}_{I-1}]^{-1}; \\
 & \mathbf{H}_I & := & \hat{\mathbf{A}}_I - \mathbf{L}_I \mathbf{U}_{I-1}; \\
\text{Set} & \mathbf{x}'_1 & := & \mathbf{y}_1; \\
\text{For } i = 2, \ldots, I \text{ set} & \mathbf{x}'_i & := & \mathbf{y}_i - \mathbf{L}_i \mathbf{x}'_{i-1}; \\
\text{Set} & \mathbf{x}_I & := & [\mathbf{H}_I]^{-1} \mathbf{x}'_I; \\
\text{For } i = I-1, \ldots, 1 \text{ set} & \mathbf{x}_i & := & [\mathbf{H}_i]^{-1} (\mathbf{x}'_i - \mathbf{U}_i \mathbf{x}_{i+1});
\end{array}
$$

Chapter 3
Stochastic Volatility for Utility Maximisers Part I
The Bond-Stock Economy

Simon Ellersgaard and Martin Jönsson

Abstract From an empirical perspective, the stochasticity of volatility is manifest, yet there have been relatively few attempts to reconcile this fact with Merton's theory of optimal portfolio selection for wealth maximising agents. In this paper we present a systematic analysis of the optimal asset allocation in a derivative-free market for the Heston model, the 3/2 model, and a Fong Vasicek type model. Under the assumption that the market price of risk is proportional to volatility, we can derive closed form expressions for the optimal portfolio using the formalism of Hamilton-Jacobi-Bellman. We also perform an empirical investigation, which strongly suggests that there in reality are no tangible welfare gains associated with hedging stochastic volatility in a bond-stock economy.

Key words: Merton's Portfolio Problem, Stochastic Volatility, HJB Equation

---

Both authors are with The Department of Mathematical Sciences
Universitetsparken 5, 2100 Copenhagen Ø, Denmark
e-mail: ellersgaard@math.ku.dk and e-mail: maj@math.ku.dk.
Ellersgaard is corresponding author.

3.1 Introduction

Since Merton's seminal paper on lifetime portfolio selection under uncertainty [12] the use of stochastic control theory in the study of dynamic optimal portfolio selection has become a cornerstone in the field of mathematical economics. Countless extensions and modifications have been proposed to Merton's original argument, including, but not limited to, stochastic interest rates and labour wages, inflation risk and home ownership (for a highly readable exegesis of these topics, the reader is referred to Munk [15] and the references therein). By and large, these extensions centre around solving ever more involuted variations of the non-linear Hamilton-Jacobi-Bellman equation. Indeed, this paper is no exception in this regard.[1] Curiously though, there have been relatively few attempts at unifying Merton's classical portfolio problem with stochastic volatility. Given the overwhelming empirical support which underpins the latter (indeed, the tremendous theoretical interest it has received from the quantitative finance community) this is surprising. In fact, it is only during the past decade that a small number of papers have emerged offering closed-form expressions for optimal portfolio choices in stochastic volatility economies. Most prominently, perhaps, is the work by Liu [19][2] and Liu and Pan [20] (a) in which an investor seeks to optimise her bequest in a Hestonian market [17] with access to a risk free money account and a stock, (b) in which essentially the same scenario is extended to include derivatives. For a formal derivation of (a) see Kraft [18] who offers a full-fledged verification argument.[3]. Another paper worthy of mention is that of Branger and Hansis [5], in which optimal "buy-and-hold" strategies are studied for an isoelastic (CRRA) investor who can also trade in a stock option. Here the market is again Hestonian, albeit with the interesting caveat that they allow for correlated jumps in the driving processes of the stock price and the variance. Fundamentally, Branger and Hansis seek to uncover the utility gains obtained through the incorporation of derivatives and the losses incurred due to the omission of risk factors and erroneous estimates of risk premia. Finally, we refer the reader to a paper by Chacko and Viceira [10] in which an optimal consumption process under stochastic volatility is desired for an investor with Epstein-Zin utility. Working with what effectively translates to a $3/2$-model in a stock-bond economy, Chacko and Viceira manage to derive closed-form approximations to the optimal consumption level, followed by a perspective from empirics. Conceptually, our paper bears its strongest ties to this study. Nonetheless, the differences between our approach and that of Chacko and Viceira also remain pronounced: e.g. our optimisation problem pertains to terminal wealth maximisation - not continuous rate of consumption. Moreover, our stochastic volatility models are different, as is our empirical performance test.

---

[1] Nevertheless, the HJB approach no longer monopolises the market cf. the comparatively recent developments in martingale / convex-duality theory which in some sense provide a much more æsthetic and general approach to the field (see e.g. Pham [18]).

[2] Liu takes this as an instantiation of a more general body of theory involving state variables of "quadratic form".

[3] Non-Lipschitzian models (of which the CIR process is an example) do not fall within that general spectrum of SDEs for which verification is considered well established (see for instance Zariphopoulou [35]).

The structure of this paper is as follows: In section 3.2 we lay down the foundations for our study, by discussing the assumptions enforced upon the market and the investor. In section 3.3 we formalise the optimal investment problem using principles of stochastic control theory and perform a standard dimensional reduction of the governing differential equation based on the linearity of the wealth dynamics. Sections 3.4 and 3.5 are dedicated to the solution of two concrete problems viz. the case where the variance process follows a Heston model, and the case where it follows a 3/2 model. Upon making concrete specifications for the market price of risk, the former portfolio weight is shown to be independent of the level of the variance, whilst the same thing cannot be said for the latter. Section 3.6 is a generalisation of the material from section 3.3, by allowing for multiple risky assets and multiple state variables. Section 3.7 exemplifies this extension by considering a half-forgotten model proposed by Fong and Vasicek in the early 90s [31], [32]. Fundamentally, the concern here is the inclusion of fixed income products the values of which are dependent upon a stochastic yield-volatility model. Finally, section 3.8 constitutes an empirical investigation into the extent to which rational investors should be concerned about hedging stochastic volatility as opposed to just following a garden variety Merton strategy. Upon calibrating our models to market data, we observe that hedging stochastic volatility within the confines of our market assumptions, in practice does not lead to tangible welfare gains. Section 3.9 concludes.

3.2 Model Set-up

3.2.1 The Economy

Consider a financial market satisfying the usual *Black-Scholes-Merton assumptions* i.e. a market characterised by continuous trading and the absence of arbitrage, a market where all assets are infinitely divisible as to the amount which may be held and where no trade is subject to transaction costs or taxation (*friction*). For simplicity, we initially consider the rudimentary financial landscape where there are just two tradeable assets in existence, viz. a risk risk-free money account (a *bond*), represented by the price process $\{B_t\}_{t\geq 0}$, and a singular risky asset (a *stock*), represented by the price process $\{S_t\}_{t\geq 0}$. In concrete dynamical terms, we shall assume that the money account grows according to the deterministic equation

$$dB_t = rB_t dt,$$

where r is a constant risk free rate (this assumption is relaxed later). On the other hand, the risky asset is assumed to be governed by the generic stochastic volatility model

$$\begin{aligned} dS_t &= S_t\{(r+\sqrt{V_t}\lambda_1(V_t))dt+\sqrt{V_t}dW_{1t}\},\\ dV_t &= \alpha(t,V_t)dt+\beta(t,V_t)(\rho dW_{1t}+\sqrt{1-\rho^2}dW_{2t}), \end{aligned} \tag{3.1}$$

where $\{V_t\}_{t\geq 0}$ is the instantaneous variance process, and the random components W_1 and W_2 are independent Wiener processes. Here, we have introduced λ_1 as the *market price of risk* associated with W_1 (which, equivalently, may be thought of as the *Sharpe ratio* of the stock) and the parameter $\rho \in [-1,1]$ to codify an instantaneous correlation between the stock price and the variance

$$\rho = \mathbb{C}\mathrm{orr}[dS_t, dV_t].$$

The coefficients $\alpha : [0,\infty) \times \mathbb{R}^+ \mapsto \mathbb{R}$ (the *drift* of the variance) and $\beta : [0,\infty) \times \mathbb{R}^+ \mapsto \mathbb{R}$ (the *volatility* of variance) are continuous deterministic functions, which satisfy certain technical conditions in order to guarantee the existence of a unique strong solution. As usual, all stochastic processes are assumed to inhabit a filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$, where Ω represents all possible states of the economy, and $\mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}$ is a filtration which satisfies the *usual conditions*.

Remark 3.1. Observe that λ_1 is assumed dependent upon the instantaneous variance. Specifically, it turns out to be useful to posit that

$$\lambda_1(V_t) = \lambda\sqrt{V_t},$$

where $\lambda \in \mathbb{R}^+$. Nonetheless, until this becomes sufficiently obvious, we will work with the general function λ_1.

Remark 3.2. The fact that the market is assumed void of derivatives is a shortcoming worth highlighting. Specifically, our findings in what follows can at best be said to be applicable to unsophisticated investors who take positions in a bond-stock mixture.

3.2.2 The Investor's Problem

We consider an investor who trades in the financial market over a known temporal horizon $\mathbb{T} = [0,T]$ where $T \in (0,\infty)$. If $\mathscr{W}_t^\pi$ represents her total wealth at time t and π_t represents the fraction of wealth she places on the risky asset (with the remaining wealth being deposited in the risk free money account), a quick application of the *self-financing condition* shows that her total wealth evolves according to the stochastic differential equation

$$d\mathscr{W}_t^\pi = \mathscr{W}_t^\pi\{(r + \pi_t\sqrt{V_t}\lambda_1(V_t))dt + \pi_t\sqrt{V_t}dW_{1t}\}. \tag{3.2}$$

Fundamentally, we assume that the the investor is interested in determining the functional form of the risky portfolio weight which will maximise the expected utility of her terminal wealth ("bequest" at time T). Specifically, the goal is to compute an optimal admissible feedback control law $\pi_t^* = \pi^*(t, \mathscr{W}_t, V_t)$ for some function $\pi^* : \mathbb{T} \times \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}$ such that

$$\{\pi_s^*\}_{s\in[t,T]} = \underset{\{\pi_t\}\in\mathscr{A}(t,w,v)}{\operatorname{argmax}} \mathbb{E}_{t,w,v}\left[e^{-\delta(T-t)}u(\mathscr{W}_T^\pi)\right],$$

where $u : \mathbb{R}^+ \mapsto \mathbb{R}$ is a von Neumann-Morgenstern utility function and δ is the investor's subjective discount factor. Throughout this paper, we will assume that the space of admissible controls is:

$$\mathscr{A}(t,w,v) = \left\{ \pi : \int_t^T \pi_u^2 du < \infty \text{ a.s. for all } T \geq t, \text{ s.t. } (\mathscr{W}_t, V_t) = (w,v) \right\}.$$

In particular, we do not enforce any restrictions on short-selling or leveraging. As for the utility function, u, we adopt the convention of isoelasticity, meaning that

$$u(x) = \begin{cases} (x^{1-\gamma} - 1)/(1-\gamma), & \text{for } \gamma \neq 1, \\ \ln(x), & \text{for } \gamma = 1, \end{cases}$$

where γ is a positive[4] parameter which codifies the investor's level of risk aversion. Obviously, optimal choices are unaffected by translations of the utility function along the ordinate axis, which means that the -1 in the numerator of u ($\gamma \neq 1$) can and will be dropped in an optimisation context. The main point of retaining the -1 in the definition stems from the fact that $(x^{1-\gamma} - 1)/(1-\gamma)$ formally converges to $\ln(x)$ as $\gamma \to 1$ as it may be verified by applying l'Hôspital's rule. Accordingly, we will also be able to consider the optimal financial decisions of log investors throughout this paper simply by letting γ go to unity.

3.3 Towards Rigour

3.3.1 The HJB Formalism

Defining the optimal value function $J : \mathbb{T} \times \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}$

$$J(t,w,v) \equiv \sup_{\{\pi_t\} \in \mathscr{A}(t,w,v)} \mathbb{E}_{t,w,v} \left[e^{-\delta(T-t)} \frac{(\mathscr{W}_T^{\pi})^{1-\gamma}}{1-\gamma} \right], \tag{3.3}$$

we may use standard control theoretic arguments to show that J necessarily satisfies the *Hamilton-Jacobi-Bellman* (HJB) equation

$$\begin{aligned} \delta J = \partial_t J + \sup_{\pi_t \in \mathbb{R}} \Big\{ & w(r + \pi_t \sqrt{v} \lambda_1(v)) \partial_w J + \alpha(t,v) \partial_v J \\ & + \tfrac{1}{2} w^2 \pi_t^2 v \partial_{ww}^2 J + \tfrac{1}{2} \beta^2(t,v) \partial_{vv}^2 J + \rho w \pi_t \sqrt{v} \beta(t,v) \partial_{wv}^2 J \Big\}, \end{aligned}$$

with the terminal condition $J(T,w,v) = w^{1-\gamma}/(1-\gamma)$. Differentiating the bracketed expression with respect to π_t gives is the *first order condition*

[4] The positivity of γ renders the utility function concave and thus the agent *risk averse*, as opposed to risk neutral ($\gamma = 0$) or risk loving ($\gamma < 0$).

$$\pi_t^* = -\frac{\lambda_1(v)\partial_w J}{w\sqrt{v}\partial^2_{ww}J} - \frac{\rho\beta(t,v)\partial^2_{wv}J}{w\sqrt{v}\partial^2_{ww}J}. \tag{3.4}$$

Hence, the HJB equation may be reformulated as

$$\begin{aligned} \delta J &= \partial_t J + rw\partial_w J + \alpha(t,v)\partial_v J + \tfrac{1}{2}\beta^2(t,v)\partial_{vv}J \\ &\quad - \tfrac{1}{2}\lambda_1^2(v)\frac{(\partial_w J)^2}{\partial^2_{ww}J} - \tfrac{1}{2}\rho^2\beta^2(t,v)\frac{(\partial^2_{wv}J)^2}{\partial^2_{ww}J} - \rho\beta(t,v)\lambda_1(v)\frac{\partial_w J\partial^2_{wv}J}{\partial^2_{ww}J}. \end{aligned} \tag{3.5}$$

3.3.2 A Dimensional Reduction

From the linearity of the wealth dynamics (3.2) it follows that the optimal strategy π_t^* must be independent of the level of wealth (specifically, multiplying $\mathscr{W}_t$ by an arbitrary constant k renders the dynamics form-invariant: thus, there should be no need for tampering with the control). Using this insight, the optimal value function may therefore be written as

$$\begin{aligned} J(t,kw,v) &= \mathbb{E}_{t,w,v}[e^{-\delta(T-t)}u(k\mathscr{W}_T^*)] \\ &= k^{1-\gamma}\mathbb{E}_{t,w,v}[e^{-\delta(T-t)}u(\mathscr{W}_T^*)] \\ &= k^{1-\gamma}J(t,w,v). \end{aligned}$$

Setting $k = 1/w$ we arrive at the result that the optimal value function is separable in wealth and time-volatility:

$$J(t,w,v) = g(t,v)^\gamma \frac{w^{1-\gamma}}{1-\gamma}, \tag{3.6}$$

where $g(t,v)^\gamma \equiv (1-\gamma)J(t,1,v)$. Inserting this expression into the HJB equation (3.5) we obtain, after considerable simplification[5],

$$\begin{aligned} 0 &= \partial_t g - \left(\tfrac{\delta}{\gamma} - \hat{\gamma}r - \tfrac{1}{2}\gamma^{-1}\hat{\gamma}\lambda_1^2(v)\right)g + \left(\alpha(t,v) + \hat{\gamma}\rho\beta(t,v)\lambda_1(v)\right)\partial_v g \\ &\quad + \tfrac{1}{2}\beta^2(t,v)\partial^2_{vv}g - \tfrac{1}{2}(1-\gamma)(1-\rho^2)\beta^2(t,v)\frac{(\partial_v g)^2}{g}, \end{aligned} \tag{3.7}$$

where $g(T,v) = 1$ (cf. the terminal condition above) and we have introduced the parameter $\hat{\gamma} \equiv (1-\gamma)/\gamma$. Moreover, if we combine the ansatz above with (3.4), the first order condition simplifies to

$$\pi_t^* = \frac{\lambda_1(v)}{\gamma\sqrt{v}} + \frac{\rho\beta(t,v)\partial_v g}{\sqrt{v}g}. \tag{3.8}$$

Specifically, we see that the rational investor ought to amend her Mertonian stock-wealth (as codified by $\lambda_1(v)/(\gamma\sqrt{v})$) by the correction $\rho\beta(t,v)\partial_v g/(\sqrt{v}g)$ to hedge against fluc-

[5] Here we use that $\partial_t J = \hat{\gamma}^{-1}g^{\gamma-1}\partial_t g w^{1-\gamma}$, $\partial_w J = g^\gamma w^{-\gamma}$, $\partial^2_{ww}J = -\gamma g^\gamma w^{-\gamma-1}$, $\partial_v J = \hat{\gamma}^{-1}g^{\gamma-1}\partial_v g w^{1-\gamma}$, $\partial^2_{vv}J = -\gamma g^{\gamma-2}(\partial_x g)^2 w^{1-\gamma} + \hat{\gamma}^{-1}g^{\gamma-1}\partial^2_{vv}g w^{1-\gamma}$ and $\partial^2_{wv}J = \gamma g^{\gamma-1}\partial_v g w^{-\gamma}$.

tuations in the underlying state variable, v. Insofar as the latter is zero, the investor is said to be *myopic*. Already we see that this occurs under at least two different circumstances, both of which relish in intuitive appeal: viz. the event when volatility is deterministic ($\beta(t,v) = 0$) and the event when volatility is completely uncorrelated with the price of the risky asset ($\rho = 0$) (whence the risky asset *per se* provides no hedge against fluctuations in the underlying state variable).

To proceed any further we now have to make concrete assumptions about the nature of the stochastic volatility model, i.e. the functional form of α and β. To "set the scene", as it were, we initially expose Liu's [19] investigation of optimal portfolio choices when volatility obeys a Heston (CIR) model, only to generalise immediately to the case where the Heston parameters are made time-dependent. Subsequently, we consider having volatility follow a 3/2-model, which is equivalent to stipulating that volatility^{-1} is CIR. All proofs are relegated to the appendix.

3.4 Experiences From the Heston Model

One of the most popular volatility models in the derivatives industry, cherished for the fact that it admits quasi-analytic expressions for European call and put options, is the Heston model, [17]. Dynamically, it corresponds to the Cox-Ingersoll-Ross (CIR) equation,

$$dV_t = \kappa(\theta - V_t)dt + \xi\sqrt{V_t}(\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}), \tag{3.9}$$

where κ, θ and ξ are positive parameters which respectively signify the speed of mean reversion, the long run mean and the so-called volatility of variance. To secure the the variance process V_t is non-explosive, these parameters are typically taken to satisfy the Feller condition: $2\kappa\theta \geq \xi^2$.

Theorem 3.1. ***The Heston Optimal Portfolio.*** *Under the assumptions that (I) the volatility model is Hestonian and (II) the market price of risk of the stock is proportional to volatility, $\lambda_1(v) = \lambda\sqrt{v}$, it follows that the optimal portfolio weight to place on the stock is*

$$\pi_t^* = \frac{\lambda}{\gamma} + \rho\xi\hat{\gamma}B(T-t), \tag{3.10}$$

where we have introduced the function

$$B(\tau) = \frac{\lambda^2}{\gamma}\frac{(e^{\hat{\eta}\tau} - 1)}{(\hat{\kappa}+\hat{\eta})(e^{\hat{\eta}\tau} - 1) + 2\hat{\eta}}, \tag{3.11}$$

and the parameters $\hat{\kappa} \equiv \kappa - \hat{\gamma}\rho\xi\lambda$ and $\hat{\eta} \equiv \sqrt{\hat{\kappa}^2 - \hat{\gamma}\xi^2(\rho^2 + \gamma[1-\rho^2])\lambda^2/\gamma}$.

Proof. The key insight is that by positing $\lambda_1(v) = \lambda\sqrt{v}$ we obtain a differential equation which admits an exponential affine solution. See appendix A for details. □

A few remarks on the optimal portfolio weight are in order: first, π_t^* does *not* depend on the instantaneous volatility, v. Whilst this *prima facie* is a surprising result (to the extent that one perhaps would expect an inverse correlation between stock holding and volatility), note that we effectively have imposed an offsetting condition by engineering the risk premium to be proportional to volatility. Secondly, as $\gamma \to 1$, $\pi^* \to \lambda/\gamma$ (the Merton portfolio) which is in accordance with the theorem that log investors are myopic in their optimal investment strategy (see e.g. Munk [15]). Thirdly, a few words about the behaviour of the B-function: assuming $\hat{\kappa} \in \mathbb{R}^+$ then $B'(\tau) > 0$ and $B''(\tau) < 0$ whence B is a monotonically increasing and concave function of the time to maturity.[6] It is also bounded from above: $B(\tau) \to \lambda^2/(\gamma(\hat{\kappa}+\hat{\eta}))$ as $\tau \to \infty$, whence

$$\lim_{\tau\to\infty} \pi_t^* = \frac{\lambda}{\gamma} + \frac{\rho\xi\hat{\gamma}\lambda^2}{\gamma(\hat{\kappa}+\hat{\eta})}.$$

On the other hand, as $\tau \to 0$, $B(\tau) \to 0$ whence $\pi^* \to \lambda/\gamma$ which fits the intuition that short horizon investors increasingly disregard volatility fluctuations (they become increasingly myopic in their investment strategies as the TTM decreases). Finally, observe that the investors position vis-à-vis the Merton portfolio is intimately linked to the signs of ρ and $\hat{\gamma}$. If both parameters are positive or negative, then the Heston portfolio takes a more aggressive position in the stock market. Conversely, if one and only one of them is negative, the investor will be more prone towards the risk free asset. Empirically, one often finds that $\rho < 0$ for equity, while $\gamma > 1$ ($\hat{\gamma} < 0$), which corresponds to the first scenario.

Now, over longer temporal horizons it is scarcely plausible to assume that the financial landscape will remain sufficiently static to guarantee the constancy of the Heston parameters $\{\kappa,\theta,\xi,\rho\}$. Whilst it is easy to envision all sorts of stochastic dependencies, we are at the same time faced with the burden of making our model analytically tractable. To this end, consider the simplest of extensions, in which $\{\kappa,\theta,\xi,\rho\}$ are made to be deterministic functions of time. Specifically, let us consider the volatility model

$$dV_t = \kappa(t)(\theta(t) - V_t)dt + \xi(t)\sqrt{V_t}(\rho(t)dW_{1t} + \sqrt{1-\rho(t)^2}dW_{2t}), \tag{3.12}$$

where $\kappa : [0,\infty) \mapsto (0,\infty)$, $\theta : [0,\infty) \mapsto (0,\infty)$, $\xi : [0,\infty) \mapsto (0,\infty)$ and $\rho : [0,\infty) \mapsto [-1,1]$ are defined such that the SDE for the variance process has a unique strong solution and, typically, also such that the time-dependent Feller condition is satisfied, i.e. $2\kappa(t)\theta(t) > \xi^2(t)$ for all $t \in [0,\infty)$. Finally, we will assume that $\kappa(t)$, $\xi(t)$ and $\rho(t)$ are (or may be reasonably approximated as) *piecewise constant functions*, continuous from the left, whilst no such restriction is placed on $\theta(t)$. Specifically, let $[t_0,t_n] = [t,T]$ be some finite temporal horizon over which the investor trades, then we assume the existence of a finite number of discontinuity points $t_1 < t_2 < ... < t_{n-1} \in (t,T)$ such that $\bar{\kappa}_1, \bar{\xi}_1, \bar{\rho}_1$ are constant over the half-closed interval $(t_{n-1},t_n]$, $\bar{\kappa}_2, \bar{\xi}_2, \bar{\rho}_2$ are constant over the half-closed interval

[6] We calibrated the Heston model 44 times over he course of eleven years and consistently found $\hat{\kappa} \in \mathbb{R}^+$.

$(t_{n-2}, t_{n-1}]$, ... and $\bar{\kappa}_n, \bar{\xi}_n, \bar{\rho}_n$ are constant over the closed interval $[t_0, t_1]$. The union of these n disjoint subintervals is clearly $[t, T]$.

Theorem 3.2. ***The Time-dependent Heston Optimal Portfolio.*** *Under the assumptions that (I) the volatility model is time-dependent Hestonian with κ, ξ and ρ being piecewise constant functions, continuous from the left, and (II) the market-price-of risk of the stock is proportional to volatility, $\lambda_1(v) = \lambda\sqrt{v}$ where $\lambda \in \mathbb{R}^+$, it follows that the optimal portfolio weight to place on the stock is*

$$\pi_t^* = \frac{\lambda}{\gamma} + \bar{\rho}_n \bar{\xi}_n \hat{\gamma} B_n(T-t),$$

where $\bar{\rho}_n$, $\bar{\xi}_n$ are the values of ρ and ξ at time t, and $\{B_k\}_{k=1}^n$ are complicated functions that are computed sequentially for $k = 1, \ldots, n$ according to equation (3.46) in the appendix. For each k, the function B_k depends on $\bar{\kappa}_i$, $\bar{\rho}_i$, $\bar{\xi}_i$ for $i = 1, \ldots, k$, the piecewise constant values of κ, ρ, ξ.

Proof. Again, from $\lambda_1(v) = \lambda\sqrt{v}$ we obtain a differential equation which admits an exponential affine solution. This allows us to solve the system sequentially backwards in time. See appendix A for details. □

3.5 Experiences From the 3/2 Model

Following empirical studies into S&P100 implied volatilities by Jones [13] and Bakshi, Ju, and Yang [2], a stylised fact of the variance process diffusion exponent is its proximity to 3/2 rather than the 1/2 inherent to the Heston model. This prompts us to look into the non-affine volatility model simply known as the 3/2 model:

$$dV_t = \kappa V_t(\theta - V_t)dt + \xi V_t^{3/2}(\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}), \tag{3.13}$$

where θ and ξ retain their interpretations from before, but the speed of mean reversion now has been made dependent upon the instantaneous variance: κV_t. Unlike the Heston model, there is here no"non-explosivity" restriction on the choice of parameters,[7] although we maintain that κ, θ and ξ are all positive. To solve the optimal portfolio problem under the $3/2$-model we make two key assumptions: first, inspired from the previous subsection we will take $\lambda_1(v) = \lambda\sqrt{v}$. Secondly, we boldly assume *market completeness* by fixing a perfect negative correlation between S_t and V_t:

[7] To see this, use Itō's lemma to show that $\{V_t^{-1}\}$ is a CIR (Heston) model with speed of mean reversion $\kappa\theta$, long run mean $(\kappa + \xi^2)/(2\kappa)$ and volatility of variance $-\xi$. Substituting these parameters into the Feller condition we obtain, after a few lines of manipulation, $\kappa \geq -\xi^2/2$ which always is satisfied.

$$\rho = -1.$$

Whilst this inexorably is an approximation, it is perhaps not an altogether unreasonable one - at least some of the time. Drimus [13], for instance, finds that $\rho = -0.99$ in a calibration of the $3/2$-model to the S&P 500 index.[8] Our own calibrations are less encouraging in this respect (see the empirical section).

Theorem 3.3. ***The 3/2 Optimal Portfolio.*** *Under the assumptions that (I) the volatility model is* $3/2$ *and (II) the market price of risk of the stock is proportional to volatility,* $\lambda_1(v) = \lambda\sqrt{v}$*, and (III) the stock and variance processes are perfectly negatively correlated, it follows that the optimal portfolio weight to place on the stock is*

$$\pi_t^* = \frac{\lambda}{\gamma} + \xi a\left(1 + \frac{z}{\hat{\zeta}}\frac{M(a+1;\hat{\zeta}+1;z)}{M(a;\hat{\zeta};z)}\right), \tag{3.14}$$

where $M(a;\hat{\zeta};z) = {}_1F_1(a;\hat{\zeta};z)$ *is the confluent hypergeometric function (Kummer's function of the first kind)*

$$M(a;\hat{\zeta};z) \equiv \sum_{n=0}^{\infty}\frac{a^{(n)}z^n}{\hat{\zeta}^{(n)}n!} = 1 + \frac{a}{\hat{\zeta}}z + \frac{a(a+1)}{\hat{\zeta}(\hat{\zeta}+1)}\frac{z^2}{2} + \frac{a(a+1)(a+2)}{\hat{\zeta}(\hat{\zeta}+1)(\hat{\zeta}+2)}\frac{z^3}{6} + \cdots$$

and we have defined the variable $z = z(v, T-t)$*:*

$$z \equiv \frac{2\kappa\theta}{\xi^2 v(1 - e^{\kappa\theta(T-t)})},$$

and the parameters $a = -\hat{\omega} + \sqrt{\hat{\omega}^2 - \frac{\hat{\gamma}\lambda^2}{\gamma\xi^2}}$, $\hat{\zeta} \equiv 1 + 2(a+\hat{\omega})$ *where* $\hat{\omega} \equiv \frac{1}{2} + \frac{\hat{\gamma}\xi\lambda + \kappa}{\xi^2}$.

Proof. Beyond the specification $\lambda_1(v) = \lambda\sqrt{v}$, the key element in the proof is here the ansatz that the governing differential equation only depends on (t,v) through the intervening variable $y(t,v) \equiv \int_t^T e^{\int_t^u \kappa\theta ds}du \cdot v$. This allows us, through a series of coordinate transformations, to obtain a confluent hypergeometric differential equation. See appendix A for details. A cursory introduction to confluent hypergeometric functions and their properties is provided in appendix C. $\square$

We make the following observations: firstly, unlike the Merton-Heston problem, the optimal control here retains a dependence upon z (and thence on the instantaneous variance v).

---

[8] To highlight the limited comparative difference between setting $\rho = -1$ and $\rho = -0.99$ we ran 10^6 Monte Carlo simulations of a 3/2-driven stock path over a period of three months using the parameters $(S_0, V_0, \mu, \kappa, \theta, \xi) = (100, 0.2450, 0.15, 22.84,\ 0.4669, 8.560)$. The mean *terminal* price difference between the $\rho = -1$ and $\rho = -0.99$ paths was found to be 3.7807 with a standard deviation of 3.1861, whilst the mean price difference *over the lifetime of the stock* was evaluated to 2.2788 with a standard deviation of 1.5477. The difference between the two scenarios is, in other words, relatively modest over short temporal horizons - even if not altogether negligible.

Secondly, it is a well-known fact that the ratio $M(a+1;\hat{\zeta}+1;z)/M(a;\hat{\zeta};z)$ can be written as a Gaussian continued fraction. Finally, as for the limiting behaviour of the control, we observe that as $\tau \equiv (T-t) \to \infty$ $y \to \infty$ whence $z \uparrow 0$ and $M \to 1$. Thus, for long temporal horizons the optimal portfolio strategy converges to

$$\lim_{\tau\to\infty} \pi_t^* = \frac{\lambda}{\gamma} + \xi a.$$

On the other hand, as $\tau \to 0$, $y \to 0$ and $z \downarrow -\infty$ so $M(a+1;\hat{\zeta}+1;z)/M(a;\hat{\zeta};z) \to -\hat{\zeta}/z$, where we have used that for large (negative) zs

$$M(a;\hat{\zeta};z) \sim (-z)^a \frac{\Gamma(\hat{\zeta})}{\Gamma(\hat{\zeta}-a)},$$

where Γ is the gamma function, in conjunction with the elementary relation $\Gamma(\hat{\zeta}+1) = \hat{\zeta}\Gamma(\hat{\zeta})$. Once again, the upshot is that short term investment horizons are myopic: $\pi_t^* \to \lambda/\gamma$ as $\tau \to 0$.

3.6 The Multi-Asset Multi-Factor Extension

3.6.1 Towards Realism

One notable limitation of the financial landscape discussed above is inevitably the assumption that there is just one tradeable risky asset in existence. In a similar vein, working with a single-factor model (the volatility) seems comically reductionistic: at the very least, our model should be able to accommodate multiple volatilities as well as a stochastic short rate (to encompass the bond market). Thus, for the sake of financial plurality, we introduce an abstract generalisation of the dynamics (3.1). Specifically, letting $\boldsymbol{S}_t = (S_{1t}, S_{2t}, ..., S_{Nt})$ represent an N-dimensional vector codifying the price processes of N risky assets, and letting $\boldsymbol{V}_t = (V_{1t}, V_{2t}, ..., V_{Mt})$ represent an M-dimensional state variable, we suppose

$$\begin{aligned} d\boldsymbol{S}_t &= \mathbf{D}_{\boldsymbol{S}}\{(r(\boldsymbol{V}_t)\boldsymbol{\iota} + \boldsymbol{\sigma}(t,\boldsymbol{V}_t)\boldsymbol{\lambda}_1(\boldsymbol{V}_t))dt + \boldsymbol{\sigma}(t,\boldsymbol{V}_t)d\boldsymbol{W}_{1t}\}, \\ d\boldsymbol{V}_t &= \boldsymbol{\alpha}(t,\boldsymbol{V}_t)dt + \boldsymbol{\beta}_1(t,\boldsymbol{V}_t)d\boldsymbol{W}_{1t} + \boldsymbol{\beta}_2(t,\boldsymbol{V}_t)d\boldsymbol{W}_{2t}. \end{aligned} \tag{3.15}$$

Here, $\boldsymbol{W}_1 \in \mathbb{R}^N$ and $\boldsymbol{W}_2 \in \mathbb{R}^M$ are independent Wiener processes (internally, as well as with respect to each other), where $\boldsymbol{\lambda}_1(\boldsymbol{V}_t)$ is the market price of risk associated with the former. Furthermore, $\boldsymbol{\sigma} : [0,\infty) \times \mathbb{R}^M \mapsto \mathbb{R}^{N\times N}$, $\boldsymbol{\alpha} : [0,\infty) \times \mathbb{R}^M \mapsto \mathbb{R}^M$, $\boldsymbol{\beta}_1 : [0,\infty) \times \mathbb{R}^M \mapsto \mathbb{R}^{M\times N}$ and $\boldsymbol{\beta}_2 : [0,\infty) \times \mathbb{R}^M \mapsto \mathbb{R}^{M\times M}$ are progressively measurable functions, which, again, satisfy certain technical conditions. Finally, the short rate r has been made dependent upon the state variable $\boldsymbol{V}_t$ and we have introduced the notation $\mathbf{D}_{\boldsymbol{S}} : \mathbb{R}^N \mapsto \mathbb{R}^{N\times N}$ for the diagonalisation of $\boldsymbol{S}$, i.e. $\mathbf{D}_{\boldsymbol{S}} \equiv \mathrm{diag}(S_1, S_2, ..., S_N)$, and $\boldsymbol{\iota}$ as the the N-dimensional vector of ones $(1,1,...,1)^\intercal$.

Regarding the investor, our assumptions from before remain intact, i.e. we are still contemplating a situation in which the goal is to maximise expected utility of discounted bequest $\mathbb{E}\left[e^{-\delta T}u(\mathscr{W}_T)\right]$. However, as she is now balancing a portfolio comprised on N risky assets and one risk free asset, her self-financing portfolio now evolves as

$$d\mathscr{W}_t^{\boldsymbol{\pi}} = \mathscr{W}_t^{\boldsymbol{\pi}}\{(r(\boldsymbol{V}_t) + \boldsymbol{\pi}_t^{\intercal}\boldsymbol{\sigma}\boldsymbol{\lambda}_1(\boldsymbol{V}_t))dt + \boldsymbol{\pi}_t^{\intercal}\boldsymbol{\sigma}d\boldsymbol{W}_{1t}\}.$$

where $\boldsymbol{\pi}_t = (\pi_{1t}, \pi_{2t}, ..., \pi_{Nt})$ is an N-dimensional vector of controls corresponding to the weights she places on each of the risky assets.[9]

3.6.2 The HJB Equation

Defining the optimal value function $J : \mathbb{T} \times \mathbb{R}^+ \times \mathbb{R}^{N+} \mapsto \mathbb{R}$ as in (5.4) with v and π replaced by the appropriate vector quantities $\boldsymbol{v}$ and $\boldsymbol{\pi}$, it can be shown that the governing multi-dimensional HJB equation is of the form[10]

$$\begin{aligned}\delta J = \partial_t J + \sup_{\boldsymbol{\pi}_t \in \mathbb{R}^N} \Big\{ & w(r(\boldsymbol{v}) + \boldsymbol{\pi}_t^{\intercal}\boldsymbol{\sigma}(t,\boldsymbol{v})\boldsymbol{\lambda}_1(\boldsymbol{v}))\partial_w J + \boldsymbol{\alpha}(t,\boldsymbol{v})^{\intercal}\nabla_{\boldsymbol{v}} J \\ & + \tfrac{1}{2}w^2\boldsymbol{\pi}_t^{\intercal}\boldsymbol{\sigma}(t,\boldsymbol{v})\boldsymbol{\sigma}^{\intercal}(t,\boldsymbol{v})\boldsymbol{\pi}_t\partial^2_{ww}J + \tfrac{1}{2}\mathrm{tr}[\nabla^2_{\boldsymbol{v}\boldsymbol{v}}J\boldsymbol{\Sigma}] + w\boldsymbol{\pi}_t^{\intercal}\boldsymbol{\sigma}(t,\boldsymbol{v})\boldsymbol{\beta}^{\intercal}(t,\boldsymbol{v})\nabla_{\boldsymbol{v}}\partial_w J\Big\},\end{aligned} \tag{3.16}$$

where $J(T,w,\boldsymbol{v}) = w^{1-\gamma}/(1-\gamma)$ is the terminal condition, and we have introduced the following notation: (i)

$$\boldsymbol{\Sigma} \equiv \boldsymbol{\beta}_1(t,\boldsymbol{v})\boldsymbol{\beta}_1^{\intercal}(t,\boldsymbol{v}) + \boldsymbol{\beta}_2(t,\boldsymbol{v})\boldsymbol{\beta}_2^{\intercal}(t,\boldsymbol{v}),$$

(ii) tr as the trace operator, (iii) $\nabla_{\boldsymbol{v}} \equiv (\partial_{v_1}, \partial_{v_2}, ..., \partial_{v_N})^{\intercal}$ as the gradient operator and (iv) $\nabla^2_{\boldsymbol{v}\boldsymbol{v}} \equiv \nabla_{\boldsymbol{v}} \otimes \nabla_{\boldsymbol{v}}$ as the Hessian operator. Differentiating partially with respect to $\boldsymbol{\pi}_t$ and equating to zero, we find that the associated FOC is

$$\boldsymbol{\pi}_t^* = -(\boldsymbol{\sigma}^{\intercal}(t,\boldsymbol{v}))^{-1}\boldsymbol{\lambda}(\boldsymbol{v})\frac{\partial_w J}{w\partial^2_{ww}J} - (\boldsymbol{\sigma}^{\intercal}(t,\boldsymbol{v}))^{-1}\boldsymbol{\beta}_1^{\intercal}(t,\boldsymbol{v})\frac{\nabla_{\boldsymbol{v}}\partial_w J}{w\partial^2_{ww}J}. \tag{3.17}$$

If we once again conjecture a solution of the form

$$J(t,w,\boldsymbol{v}) = \frac{g(t,\boldsymbol{v})^{\gamma}w^{1-\gamma}}{1-\gamma},$$

tedious calculations show that (3.16) and (3.17) jointly entail that g must satisfy

$$\begin{aligned}0 = \partial_t g - \left(\tfrac{\delta}{\gamma} - \hat{\gamma}r(\boldsymbol{v}) - \tfrac{1}{2}\gamma^{-1}\hat{\gamma}||\boldsymbol{\lambda}_1(\boldsymbol{v})||^2\right)g + \Big(\boldsymbol{\alpha}(t,\boldsymbol{v}) + \hat{\gamma}\boldsymbol{\beta}_1(t,\boldsymbol{v})\boldsymbol{\lambda}_1(\boldsymbol{v})\Big)^{\intercal}\nabla_{\boldsymbol{v}} g \\ + \tfrac{1}{2}\mathrm{tr}[\nabla^2_{\boldsymbol{v}\boldsymbol{v}}g\boldsymbol{\Sigma}] - \tfrac{1}{2}(1-\gamma)\frac{1}{g}\nabla_{\boldsymbol{v}}^{\intercal}g\boldsymbol{\beta}_2(t,\boldsymbol{v})\boldsymbol{\beta}_2^{\intercal}(t,\boldsymbol{v})\nabla_{\boldsymbol{v}} g,\end{aligned} \tag{3.18}$$

[9] The residual weight $1 - \sum_{i=1}^N \pi_{it}$ is placed on the risk free asset.

[10] For elementary introductions to multi-dimensional HJB equations see [7] and [15].

subject to the boundary condition $g(T,\boldsymbol{v})=1$, where we reiterate that $\hat{\gamma}\equiv(1-\gamma)/\gamma$. We may also rewrite the FOC as

$$\pi_t^* = \frac{1}{\gamma}(\boldsymbol{\sigma}^\intercal(t,\boldsymbol{v}))^{-1}\boldsymbol{\lambda}_1(\boldsymbol{v}) + \frac{1}{g}(\boldsymbol{\sigma}^\intercal(t,\boldsymbol{v}))^{-1}\boldsymbol{\beta}_1^\intercal(t,\boldsymbol{v})\nabla_{\boldsymbol{v}} g, \tag{3.19}$$

which, of course, is nothing but Merton's $(M+2)$-fund separation result, as it may be verified by writing $\boldsymbol{\beta}_1^\intercal(t,\boldsymbol{v})\nabla_{\boldsymbol{v}} g$ as $\sum_{j=1}^M \boldsymbol{\beta}_{1:j}^\intercal(t,\boldsymbol{v})\partial_{v_j} g$, where $\boldsymbol{\beta}_{1:j}^\intercal$ is the entire j^{th} column of $\boldsymbol{\beta}_1^\intercal$.

3.7 Experiences From a Fong-Vasicek Type Model

Multi-factor models rarely wallow in analytic elegance, and alas our exemplification is no exception. What follows is an attempt to set up a more realistic security market than what we have hitherto been exposed to, whilst keeping the factor number low. Specifically, one of the assumptions made in the single-factor model was the constancy of the interest rate. Given that the variance of changes in the yield of treasury bonds is manifestly fluctuating over time (see e.g. [32]), this is clearly undesirable. Thus, we argue that a realistic financial market model, at the very least should be able to encompass a bond market where the short rate of interests is modelled dynamically over time. Keeping with the spirit of this paper, consider modelling r_t through the stochastic volatility model

$$dr_t = \kappa_r(\theta_r - r_t)dt + \sqrt{V_t}dW_{1t}, \tag{3.20a}$$

$$dV_t = \kappa_v(\theta_v - V_t)dt + \xi\sqrt{V_t}\{\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}\}, \tag{3.20b}$$

where W_{1t} and W_{2t} are independent Wiener processes. Following Fong & Vasicek [31], [32], the implications on bond pricing of such a model are well known. Specifically,

Theorem 3.4. *If the short rate of interest, r_t, is driven by the system of SDEs (3.20), and the market prices of risk satisfy*

$$\lambda_1(v) = \lambda_1\sqrt{v} \quad \textit{and} \quad \lambda_2(v) = \frac{[\lambda_2 - \rho\lambda_1]\sqrt{v}}{\sqrt{1-\rho^2}}, \tag{3.21}$$

where λ_1, λ_2 are positive constants, then the time t price of a zero-coupon bond maturing at time $T \geq t$ is given by

$$P_t^T = \exp\left\{-\bar{A}(T-t) - \bar{B}_1(T-t)r - \bar{B}_2(T-t)v\right\}, \tag{3.22}$$

where $(r_t, V_t) = (r, v)$, and $\bar{A}$, $\bar{B}_1$ and $\bar{B}_2$ are functions of the time to maturity, which satisfy the system of differential equations:

$$\bar{A}'(\tau) = \kappa_r\theta_r\bar{B}_1(\tau) + \kappa_v\theta_v\bar{B}_2(\tau), \tag{3.23a}$$
$$\bar{B}_1'(\tau) = 1 - \kappa_r\bar{B}_1(\tau), \tag{3.23b}$$
$$\bar{B}_2'(\tau) = -\left\{\lambda_1\bar{B}_1(\tau) + \tfrac{1}{2}\bar{B}_1^2(\tau)\right\} - \{\kappa_v + \lambda_2\xi + \xi\rho\bar{B}_1(\tau)\}\bar{B}_2(\tau) - \tfrac{1}{2}\xi^2\bar{B}_2^2(\tau), \tag{3.23c}$$

subject to the boundary conditions $\bar{A}(0) = \bar{B}_1(0) = \bar{B}_2(0) = 0$. *Furthermore, the dynamics of the price of a zero-coupon bond is given by*

$$dP_t^T = P_t^T\left\{(r + \varphi_P(T-t))dt + \sigma_{P1}(T-t)dW_{1t} + \sigma_{P2}(T-t)dW_{2t}\right\},$$

where

$$\varphi_P(\tau) = \lambda_1(v)\sigma_{P1}(\tau) + \lambda_2(v)\sigma_{P2}(\tau), \tag{3.24a}$$
$$\sigma_{P1}(\tau) = -\sqrt{v}\bar{B}_1(\tau) - \sqrt{v}\rho\xi\bar{B}_2(\tau), \tag{3.24b}$$
$$\sigma_{P2}(\tau) = -\xi\sqrt{v}\sqrt{1-\rho^2}\bar{B}_2(\tau). \tag{3.24c}$$

We note that analytic expressions exist for the functions $\bar{A}(\tau)$ $\bar{B}_1(\tau)$ *and* $\bar{B}_2(\tau)$*: see Selby & Strickland [29] or Fong & Vasicek [32]) for details.*

Proof: The result follows from a standard no-arbitrage argument. Modulo some sign conventions and non-significant differences in definitions[11] the proof is fully exposed in [32] and will not be reproduced here. □

The Model. Suppose we have a market with $N > 2$ tradeable assets out of which two are zero-coupon bonds of different maturities $\{P_t^{T_1}, P_t^{T_2}\}$, and the remaining $N-2$ assets are stocks $\{S_t^1, S_t^2, ..., S_t^{N-2}\}$. The bonds are assumed to be governed by the Fong-Vasicek model presented above, while he price processes of the stocks follow the quasi-Hestonian dynamics

$$dS_t^j = S_t^j\left\{(r_t + \sqrt{V_t}\psi_j(V_t))dt + \sqrt{V_t}\sum_{i=1}^{j+2}k_{ji}dW_{it}\right\}, \tag{3.25}$$

[11] A brief remark on the market prices of risk is in order. Specifically, the market prices of risk, $q(v)$ and $p(v)$, referred to in both [29] and [32] are related to $\lambda_1(v)$ and $\lambda_2(v)$ through the linear transformation $q(v) = \lambda_1(v)$ and $p(v) = \lambda_1(v)\rho + \lambda_2(v)\sqrt{1-\rho^2}$. The difference boils down to a matter of definition: whilst we define the market price of risk qua the dynamics *written on standard form*, the standardisation criterion is dropped in [29] and [32]. Specifically, we define the market price of risk as $\boldsymbol{\sigma}\mathbf{L}\boldsymbol{\lambda}^i \equiv (\boldsymbol{\mu} - r\mathbf{1})$, where $\boldsymbol{\mu} - r\mathbf{1}$ is the excess return, and $\mathbf{L}$ is the lower triangular matrix in the Cholesky decomposition of the covariance matrix of the Wiener vector. Fong & Vasicek disregard the standardisation (meaning that they define $\boldsymbol{\sigma}\boldsymbol{\lambda}^d \equiv (\boldsymbol{\mu} - r\mathbf{1})$). In general, if $\boldsymbol{\lambda}^i$ is the risk vector associated with the former definition, and $\boldsymbol{\lambda}^d$ is the risk vector associated with the latter, we may transform between the two using $\mathbf{L}\boldsymbol{\lambda}^i = \boldsymbol{\lambda}^d$ where $\mathbf{L}$ is the lower triangular matrix in the Cholesky decomposition of the covariance matrix of the Wiener vector.

where $1 \le j \le N-2$ and V_t is driven by the SDE given in (3.20b) [we stress that there is no empirical support for the assumption that bonds and stocks should be governed by the same variance process - we invoke it for pure mathematical convenience]. All additional Wiener increments defined thusly are assumed independent of each other. Also, the Sharpe ratio $\psi_j(V_t)$ is given by $\sum_{i=1}^{j+2} k_{ji}\lambda_i(V_t)$, where $\lambda_i(V_t)$ is the market price of risk associated with W_i. Interpretation-wise, we may construe (3.25) as the *standardised version* of the situation where every stock price effectively has three random sources: two coming from the bond market (W_1, W_2) and one inherent to the stock itself W_j^* This, in turn, has non-zero correlation with all other shocks inherent to the other stocks, thus captivating the notion of a pervasive systematic risk. Cholesky decomposing the vector $(W_1^*, W_2^*, ..., W_{N-2}^*)$ the dynamics will take on the form of (3.25).

Written on the form (3.15) we are accordingly dealing with a model where $\boldsymbol{S}_t = (P_t^{T_1}, P_t^{T_2}, S_t^1, ..., S_t^{N-2})^\intercal$, $\boldsymbol{\lambda}_1(V_t) = (\lambda_1(V_t), \lambda_2(V_t), ..., \lambda_N(V_t))^\intercal$, and $d\boldsymbol{W}_{1t} = (dW_{1t}, dW_{2t}, ..., dW_{Nt})^\intercal$ are vectors in $\mathbb{R}^N$, whilst

$$\sigma(t, V_t) = \begin{pmatrix} \sigma_{P1}(T_1 - t) & \sigma_{P2}(T_1 - t) & 0 & 0 & \cdots & 0 \\ \sigma_{P1}(T_2 - t) & \sigma_{P2}(T_2 - t) & 0 & 0 & \cdots & 0 \\ k_{11}\sqrt{V_t} & k_{12}\sqrt{V_t} & k_{13}\sqrt{V_t} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ k_{N-3,1}\sqrt{V_t} & k_{N-3,2}\sqrt{V_t} & \cdots & & k_{N-3,N-1}\sqrt{V_t} & 0 \\ k_{N-2,1}\sqrt{V_t} & k_{N-2,2}\sqrt{V_t} & \cdots & & k_{N-2,N-1}\sqrt{V_t} & k_{N-2,N}\sqrt{V_t} \end{pmatrix}, \tag{3.26}$$

is a matrix in $\mathbb{R}^{N\times N}$. Furthermore, $\boldsymbol{V}_t = (r_t, V_t)^\intercal$, $\boldsymbol{\alpha}(t, \boldsymbol{V}_t) = (\kappa_r(\theta_r - r_t), \kappa_v(\theta_v - V_t))^\intercal \in \mathbb{R}^2$,

$$\boldsymbol{\beta}_1(t, V_t) = \begin{pmatrix} \sqrt{V_t} & 0 & 0 \cdots 0 \\ \xi\rho\sqrt{V_t} & \xi\sqrt{1-\rho^2}\sqrt{V_t} & 0 \cdots 0 \end{pmatrix}, \tag{3.27}$$

is in $\mathbb{R}^{2\times N}$, and $\boldsymbol{\beta}_2(t, \boldsymbol{V}_t) = d\boldsymbol{W}_{2t} = 0$.

To obtain a complete specification, we will also need to make concrete assumptions about the market price of risk vector $\boldsymbol{\lambda}_1(V_t)$. Here, we will stick with the Fong-Vasicek assumption (3.21) with respect to $\lambda_1(V_t), \lambda_2(V_t)$ and set $\lambda_j(V_t) = \lambda_j\sqrt{V_t}$ for all $j > 2$, where λ_j is a positive constant.

The Spatially Reduced HJB Equation. Suppose we define the constant

$$\lambda_\rho^2 \equiv \frac{\lambda_1^2}{1-\rho^2} - \frac{2\rho\lambda_1\lambda_2}{1-\rho^2} + \frac{\lambda_2^2}{1-\rho^2} + \lambda_3^2 + ... + \lambda_N^2,$$

then (3.18) takes on the form

$$\begin{aligned} 0 = \partial_t g - \left(\tfrac{\delta}{\gamma} - \hat{\gamma}r - \tfrac{1}{2}\gamma^{-1}\hat{\gamma}\lambda_\rho^2 v\right) g + (\kappa_r(\theta_r - r) + \hat{\gamma}\lambda_1 v)\partial_r g \\ + (\kappa_v(\theta_v - v) + \hat{\gamma}\xi\lambda_2 v)\partial_v g + \tfrac{1}{2}v\partial_{rr}^2 g + \xi\rho v\partial_{rv}^2 g + \tfrac{1}{2}\xi^2 v\partial_{vv}^2 g, \end{aligned} \tag{3.28}$$

subject to the terminal condition $g(T,r,v)=1$, where $g(t,r,v):\mathbb{T}\times\mathbb{R}\times\mathbb{R}^+\mapsto\mathbb{R}$. Given that the coefficients in this PDE are linear functions of r and v, we guess the exponential-affine solution

$$g(t,r,v)=\exp\left\{-\tfrac{\delta}{\gamma}(T-t)+\hat{\gamma}A(T-t)+\hat{\gamma}B_1(T-t)r+\hat{\gamma}B_2(T-t)v\right\}. \tag{3.29}$$

Combining (3.28) and (3.29) and using the fact that the resulting expression must hold for all values of r and v we obtain the coupled differential equations

$$\begin{aligned}
A'(\tau) &= \kappa_r\theta_r B_1(\tau)+\kappa_v\theta_v B_2(\tau), && (3.30a)\\
B_1'(\tau) &= 1-\kappa_r B_1(\tau), && (3.30b)\\
B_2'(\tau) &= \left\{\tfrac{1}{2}\gamma^{-1}\lambda_\rho^2+\hat{\gamma}\lambda_1 B_1(\tau)+\tfrac{1}{2}\hat{\gamma}B_1^2(\tau)\right\}\\
&\quad -\{\kappa_v-\hat{\gamma}\lambda_2\xi-\hat{\gamma}\xi\rho B_1(\tau)\}B_2(\tau)+\tfrac{1}{2}\hat{\gamma}\xi^2 B_2^2(\tau), && (3.30c)
\end{aligned}$$

subject to the boundary conditions $A(0)=B_1(0)=B_2(0)=0$. Notice that this system is identical to (3.23) provided we let $\gamma\to\infty$ ($\hat{\gamma}\to-1$).

Theorem 3.5. *The functions $A(\tau)$, $B_1(\tau)$ and $B_2(\tau)$ are given explicitly by*

$$\begin{aligned}
A(\tau) &= \theta_r(\tau-B_1(\tau))-\frac{2\kappa_v\theta_v}{\xi^2\hat{\gamma}}\ln\left(\frac{L(\tau)}{L(0)}\right), && (3.31a)\\
B_1(\tau) &= \frac{1-e^{-\kappa_r\tau}}{\kappa_r}, && (3.31b)\\
B_2(\tau) &= -\frac{\kappa_r\hat{p}}{\xi^2\hat{\gamma}}e^{-\kappa_r\tau}+\frac{2\kappa_r}{\xi^2\hat{\gamma}}G(\tau), && (3.31c)
\end{aligned}$$

where we have defined the following functions

$$\begin{aligned}
L(\tau) &= \sum_{j=1}^{2}K_j e^{-\kappa_r\hat{\beta}_j\tau-\frac{1}{2}\hat{p}e^{-\kappa_r\tau}}M(\hat{a}_j,\hat{\zeta}_j,\hat{q}e^{-\kappa_r\tau}),\\
G(\tau) &= \sum_{j=1}^{2}\frac{K_j e^{-\kappa_r\hat{\beta}_j\tau}}{L(\tau)e^{\frac{1}{2}\hat{p}e^{-\kappa_r\tau}}}\Big\{\hat{\beta}_j M(\hat{a}_j,\hat{\zeta}_j,\hat{q}e^{-\kappa_r\tau})\\
&\quad +\hat{q}e^{-\kappa_r\tau}\frac{\hat{a}_j}{\hat{\zeta}_j}M(\hat{a}_j+1,\hat{\zeta}_j+1,\hat{q}e^{-\kappa_r\tau})\Big\},
\end{aligned}$$

where $M(a,\zeta,z)$ is Kummer's function and K_1,K_2 are constants given by the relation $K_2=\Xi K_1$ where

$$\Xi=\frac{[\beta_1-\frac{\hat{p}}{2}]M(\hat{a}_1,\hat{\zeta}_1,\hat{q})+\hat{q}\frac{\hat{a}_1}{\hat{\zeta}_1}M(\hat{a}_1+1,\hat{\zeta}_1+1,\hat{q})}{[\frac{\hat{p}}{2}-\beta_2]M(\hat{a}_2,\hat{\zeta}_2,\hat{q})-\hat{q}\frac{\hat{a}_2}{\hat{\zeta}_2}M(\hat{a}_2+1,\hat{\zeta}_2+1,\hat{q})}.$$

Furthermore, we have introduced the following parameters:

$$(a)\quad \hat{p} \equiv \frac{\xi}{\kappa_r^2}[i\sqrt{\hat{\gamma}^2(1-\rho^2)} - \hat{\gamma}\rho], \qquad (b)\quad \hat{q} \equiv \hat{p} + \frac{\rho\xi\hat{\gamma}}{\kappa_r^2},$$

$$(c)\quad \hat{a}_j \equiv \frac{\hat{\zeta}_j}{2} + \frac{i\hat{\gamma}\rho(1-\hat{\vartheta})}{2\sqrt{\hat{\gamma}^2(1-\rho^2)}}\left\{1 - \frac{\hat{\gamma}\xi(1+\lambda_1\kappa_r)}{\rho(1-\hat{\vartheta})\kappa_r^2}\right\}, \qquad (d)\quad \hat{\zeta}_j \equiv 2\hat{\beta}_j + 1 - \hat{\vartheta},$$

$$(e)\quad \hat{\beta}_1 \equiv \hat{\beta}, \qquad (f)\quad \hat{\beta}_2 \equiv \hat{\vartheta} - \hat{\beta},$$

$$(g)\quad \hat{\beta} \equiv \frac{\hat{\vartheta}}{2} - \frac{1}{2}\sqrt{\hat{\vartheta}^2 - \frac{2\hat{\gamma}\xi^2}{\kappa_r^2}\left[\frac{\lambda_\rho^2}{2\gamma} + \frac{\hat{\gamma}\lambda_1}{\kappa_r} + \frac{\hat{\gamma}}{2\kappa_r^2}\right]}, \qquad (h)\quad \hat{\vartheta} \equiv \frac{\kappa_v}{\kappa_r} - \frac{\hat{\gamma}\xi\lambda_1}{\kappa_r} - \frac{\hat{\gamma}\xi\rho}{\kappa_r^2},$$

where $i = \sqrt{-1}$ is the complex unit.

Proof. See appendix A. □

Optimal Controls. We are now in a position to compute the optimal portfolio weights for our Fong-Vasicek model.

Theorem 3.6. *For the general N-asset case, the optimal portfolio weights $\boldsymbol{\pi}_t^* = (\pi_{t,B_1}^*, \pi_{t,B_2}^*, \pi_{t,S_1}^*, \ldots, \pi_{t,S_{N-2}}^*)^\intercal$ are given by*

$$\boldsymbol{\pi}_t^* = \frac{1}{\gamma}(\tilde{\boldsymbol{\sigma}}^\intercal(t))^{-1}\tilde{\boldsymbol{\lambda}}_1 + \frac{\hat{\gamma}}{d(t)}\begin{pmatrix} \bar{B}_1(T_2-t)B_2(T-t) - \bar{B}_2(T_2-t)B_1(T-t) \\ \bar{B}_2(T_1-t)B_1(T-t) - \bar{B}_1(T_1-t)B_2(T-t) \\ \mathbf{0} \end{pmatrix}, \tag{3.32}$$

where $B_1(\tau)$ and $B_2(\tau)$ are the functions defined in (3.31b) and (3.31c), and $\bar{B}_1(\tau)$ and $\bar{B}_2(\tau)$ are the zero-coupon bond price functions which solve (3.23b) and (3.23c).[12] *Furthermore we have defined the following quantities: (i) $\tilde{\boldsymbol{\sigma}}(t) \equiv \boldsymbol{\sigma}(t,v)/\sqrt{v} \in \mathbb{R}^{N\times N}$, (ii) $\tilde{\boldsymbol{\lambda}}_1 = \boldsymbol{\lambda}_1(v)/\sqrt{v} \in \mathbb{R}^N$, (iii) $\mathbf{0} = (0,0,\ldots,0)^\intercal \in \mathbb{R}^{N-2}$ and (iv)*

$$d(t) \equiv \bar{B}_1(T_1-t)\bar{B}_2(T_2-t) - \bar{B}_2(T_1-t)\bar{B}_1(T_2-t).$$

In the particular case where $\underline{N=3}$ (two bonds and one stock are being traded), the optimal bond weights are explicitly calculated as

$$\pi^*_{t,B_1} = \frac{1}{\gamma\xi\sqrt{1-\rho^2}d(t)}\Bigg\{\bar{B}_1(T_2)\left[\frac{(\lambda_2-\rho\lambda_1)}{\sqrt{1-\rho^2}} - \frac{k_2\lambda_3}{k_3}\right]$$
$$+\xi\bar{B}_2(T_2)\left[\sqrt{1-\rho^2}\left(\frac{k_1\lambda_3}{k_3}-\lambda_1\right)+\rho\left(\frac{(\lambda_2-\rho\lambda_1)}{\sqrt{1-\rho^2}}-\frac{k_2\lambda_3}{k_3}\right)\right]\Bigg\}$$
$$+\frac{\hat{\gamma}}{d(t)}\left\{\bar{B}_1(T_2-t)B_2(T-t)-\bar{B}_2(T_2-t)B_1(T-t)\right\},$$

$$\pi^*_{t,B_2} = \frac{1}{\gamma\xi\sqrt{1-\rho^2}d(t)}\Bigg\{\bar{B}_1(T_1)\left[\frac{k_2\lambda_3}{k_3} - \frac{(\lambda_2-\rho\lambda_1)}{\sqrt{1-\rho^2}}\right]$$
$$+\xi\bar{B}_2(T_1)\left[\sqrt{1-\rho^2}\left(\lambda_1-\frac{k_1\lambda_3}{k_3}\right)+\rho\left(\frac{k_2\lambda_3}{k_3}-\frac{(\lambda_2-\rho\lambda_1)}{\sqrt{1-\rho^2}}\right)\right]\Bigg\}$$
$$+\frac{\hat{\gamma}}{d(t)}\left\{\bar{B}_2(T_1-t)B_1(T-t)-\bar{B}_1(T_1-t)B_2(T-t)\right\}.$$

Moreover, the optimal stock weight is given by

$$\pi^*_{t,S_1} = \frac{\lambda_3}{\gamma k_3}.$$

Proof: We derive (3.32) based upon (3.19), where $\boldsymbol{\sigma}$ is defined in (3.26), $\boldsymbol{\beta}_1$ is defined in (3.27), and $g(t,r,v)$ is defined in (3.29). The key insights are the following: since all components of $(\boldsymbol{\sigma}^\intercal)^{-1}$ are proportional to $1/\sqrt{v}$, and all components of $\boldsymbol{\lambda}_1(v)$ are proportional to $\sqrt{v}$, the "tangency portfolio"-component $\frac{1}{\gamma}(\boldsymbol{\sigma}^\intercal)^{-1}\boldsymbol{\lambda}_1(v)$ is independent of v. Furthermore, since $\boldsymbol{\beta}_1^\intercal(\partial_r g, \partial_v g)$ is a vector in $\mathbb{R}^N$ which is non-zero only in the first two components, we only need to calculate the first two columns of the inverse matrix $(\boldsymbol{\sigma}^\intercal)^{-1}$ (of which only components $(1,1),(1,2),(2,1)$, and $(2,2)$ will be non-zero) in order to specify the "hedge portfolio"-component $g^{-1}(\boldsymbol{\sigma}^\intercal)^{-1}\,\boldsymbol{\beta}_1^\intercal(\partial_r g,\partial_v g)$. To this end, we use the following matrix result: let $\mathbf{A}_{(n)}$ be an $n\times n$ matrix, the *bottom row* of which is non-zero only in the rightmost entry, $\mathbf{A}_{n,n}$, then its inverse is given by[13]

$$\begin{pmatrix} \mathbf{A}_{(n-1)} & \mathbf{A}_{(1:n-1),n} \\ \mathbf{0}^\intercal_{(n-1)} & \mathbf{A}_{n,n}\end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1}_{(n-1)} & -\mathbf{A}^{-1}_{(n-1)}\mathbf{A}_{(1:n-1),n}\mathbf{A}^{-1}_{n,n} \\ \mathbf{0}^\intercal_{(n-1)} & \mathbf{A}^{-1}_{n,n}\end{pmatrix},$$

[13] This is a straightforward corollary of the *Banachiewicz identity* for blockwise inversion of matrices. Suppose $\mathbf{A},\mathbf{B},\mathbf{C}$ and $\mathbf{D}$ are matrix blocks such that $\mathbf{A}$ and $\mathbf{D}$ are square matrices, then

$$\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}^{-1} = \begin{pmatrix}\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\\ -(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\end{pmatrix},$$

insofar as $\mathbf{A}$ and $(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})$ are invertible.

where $\mathbf{A}_{(n-1)} \in \mathbb{R}^{(n-1)\times(n-1)}$, $\{\mathbf{0}_{(n-1)}, \mathbf{A}_{(1:n-1),n}\} \in \mathbb{R}^{n-1}$ and $\mathbf{A}_{n,n} \in \mathbb{R}$. Clearly, $\boldsymbol{\sigma}^\mathsf{T}_{(N)} := \boldsymbol{\sigma}^\mathsf{T}$ is such a matrix. Indeed, so is the sequence of matrices $\boldsymbol{\sigma}^\mathsf{T}_{(N-1)}$, $\boldsymbol{\sigma}^\mathsf{T}_{(N-2)}, \ldots, \boldsymbol{\sigma}^\mathsf{T}_{(3)}$ so we can apply the rule repeatedly to show that

$$(\boldsymbol{\sigma}^\mathsf{T})^{-1}_{1:2,1:2} = \begin{pmatrix} \sigma_{P1}(T_1 - t) & \sigma_{P1}(T_2 - t) \\ \sigma_{P2}(T_1 - t) & \sigma_{P2}(T_2 - t) \end{pmatrix}^{-1},$$

where $\sigma_{P1}(\tau)$, $\sigma_{P2}(\tau)$ are defined in (3.24b) and (3.24c). □

Unsurprisingly we find that only the two bonds are used to hedge against interest rate risk (as codified by W_1, W_2 in the Fong-Vasicek model) whilst stocks are used to hedge each other. More intriguing is the fact that the optimal controls are altogether independent of the state-variables (r, v), i.e. they are purely deterministic. This property is similar to a two-bond one-stock model considered by Brennan and Xia [7], where the interest rate follows a Vasicek model with a stochastic long-run mean. Provided we let $\gamma \to \infty$, we can also recover the property of their model that the hedge portfolio component is $(-\hat{\gamma}, 0, 0, \ldots, 0)$ if $T = T_1$ and $(0, -\hat{\gamma}, 0, \ldots, 0)$ if $T = T_2$.[14]

3.8 The Empirical Perspective

Thus far we have managed to develop some fairly sophisticated correction formulae to the original Merton problem given the presence of stochastic volatility. Nonetheless, two important empirical questions remain unanswered: first, the magnitude of these corrections with respect to market data is not intuitively obvious. Secondly, the empirical performance of the corrected investment strategies versus the Merton strategy must be scrutinised: whilst stochastic volatility has become common dogma in the derivatives industry, it is not clear that our prescriptive ("rational") trading approach will benefit from its incorporation. To develop an understanding for these queries, we propose a study of an investor who trades in a representative market index (the S&P 500 ETF - see figure 3.1), using respectively

1. The Merton portfolio weight, $\pi^{\mathrm{Me}} = \lambda/\gamma$,
2. The Heston portfolio weight, π^{He}_t, cf. equation (3.10),

[14] We also note that the completeness of our model makes it suitable (albeit not straightforward) for studying consumption, i.e. for investigating control problems where the investor tries to maximise $\mathbb{E}_t[\int_t^T u(c_s)ds]$, where u is a utility function (here, assumed isoelastic) and c_s denotes the rate of consumption at time s. Specifically, using the *envelope theorem* $u'(c_t) = \partial_w J(t, w, r, v)$, where J is the optimal value function, as well as the standard separation ansatz $J(t, w, r, v) = g(t, r, v)^\gamma w^{1-\gamma}/(1-\gamma)$, we find that $c_t^* = w/g(t, r, v)$. For the Fong-Vasicek model presented here, the appropriate g-function is now

$$g(t, r, v) = \int_t^T \exp\{-\tfrac{\delta}{\gamma}(s-t) + \hat{\gamma}A(s-t) + \hat{\gamma}B_1(s-t)r + \hat{\gamma}B_2(s-t)v\}ds,$$

where $A(\tau)$, $B_1(\tau)$ and $B_2(\tau)$ are as defined in (3.31a), (3.31b) and (3.31c). Given the complicated hypergeometric nature of these functions, the evaluation of the integral is a task better left for numerical study.

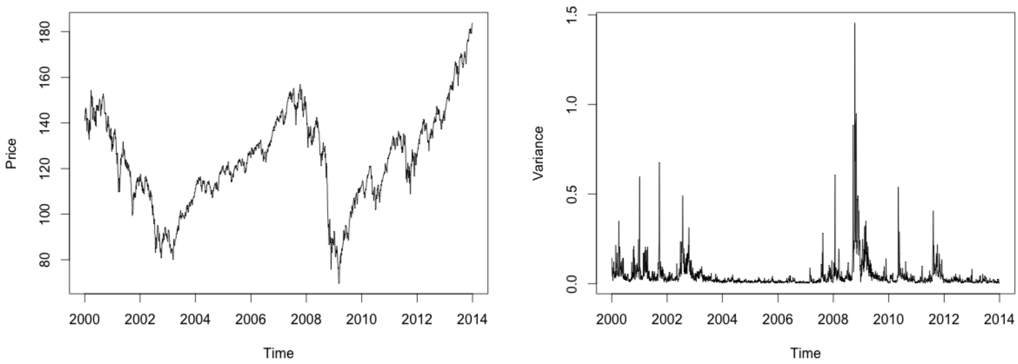

Fig. 3.1 The price process of the S&P500 ETF (left) and the measured variance [per year] (right). Notice the erratic nature of the latter, which corroborates our claim that volatility seems to behave like a stochastic process. The data is from the period 2000-01-03 to 2013-12-31 and contains 1,492 entries. For further details the reader is referred to Appendix D.

3. The $3/2$ portfolio weight, $\pi_t^{3/2}$, cf. equation (3.14).

Specifically, we consider an experimental set-up in which an investor, initially endowed with wealth $w_0 = 1000$ \$, trades in a self-financing manner over a three months horizon, rebalancing her portfolio approximately every second day. At the end of the trading period, we compute the resulting profit-&-loss (P&L) incurred by each of the three investment strategies. Throughout we fix the assumptions $r = 0.02$, $\gamma = 2.5$ and $\lambda = 0.5$. To get *some* semblance of a statistical basis for our analysis we repeat the entire experiment for a totality of 44 non-overlapping three months trading periods such that trading period 1 runs from 2003-01-01 to 2003-03-31, trading period 2 runs from 2003-04-01 to 2003-06-30 etc. Insofar as strategies (2) and (3) are truly superior to strategy (1), this should be reflected by their average return per unit standard deviation ratio.

Remark 3.3. Keeping the risk free rate constant is inexorably an approximation. However, it is not an altogether unreasonable one: all of our conclusions were found to be robust upon substituting a fixed r with the 3-month rate of the U.S. treasury bill.

3.8.1 Calibration Details

For portfolios (2) and (3) we estimate the parameters of the associated volatility model by calibrating the model to market data running over a period of three years, immediately prior to the first day of trading. For example, if a portfolio runs from 2003-01-01 to 2003-03-31, then the associated estimation period runs from 2000-01-01 to 2002-12-32. For both models, we obtain estimates of $\psi = (\kappa, \theta, \xi)$ by numerical optimisation of the log-likelihood function for respective model. In particular, if $v = (v_0, \ldots, v_n)$ is the variance from the estimation period, the log-likelihood function is given by

$$l(\psi;v) = \sum_{i=0}^{n-1} \log\phi(v_{i+1}|v_i;\psi),$$

where $\phi(v|y;\psi)$ denotes the conditional distribution of the variance process.

In Heston's model, the variance follows a CIR process which has a non-central χ^2 conditional distribution. A consistent estimator for this process can be obtained by employing the conditional moments in a Gaussian likelihood and maximising, see Sorensen [23]. Hence, we maximise

$$l^g(\psi;v) = \sum_{i=0}^{n-1} \log\phi(v_{i+1};m_{i+1|i},\sigma^2_{i+1|i}),$$

where ϕ denotes the Gaussian density, while the conditional moments are given by

$$\begin{aligned} m_{i+1|i} &= v_i e^{-\kappa h_i} + \theta(1-e^{-\kappa h_i}), \\ \sigma^2_{i+1|i} &= v_i \frac{\xi^2}{\kappa}(e^{-\kappa h_i} - e^{-2\kappa_i h}) + \theta\frac{\xi^2}{2\kappa}(1-e^{-\kappa h_i})^2. \end{aligned}$$

where the time-spacing between two observations v_{i+1} and v_i is given by h_i.

For the 3/2 model, an application of Ito's lemma shows that $1/V_t$ follows a CIR-process as well with reversion speed $\tilde{\kappa} = \kappa\theta$, long-term mean $\tilde{\theta} = (\kappa+\xi^2)/(\kappa\theta)$ and volatility $\tilde{\xi} = -\xi$. Thus, we employ the same estimator (maximising the Gaussian likelihood with conditional moments) on the inverse-transformed variance data $\{1/v_i, i = 0,\ldots,n\}$ to obtain an estimator for the 3/2 model. See also appendix D.

3.8.2 The Mertonian Benchmark Strategy

We illustrate the P&L paths of the Merton strategy against which we shall be comparing our stochastic volatility strategies in figure 3.2. As suggested, the Merton strategy is characterised by a constant holding in the risky asset of $\pi^{\mathrm{Me}} = \lambda/\gamma = 0.5/2.5 = 0.2$, with the remaining wealth going into the risk free asset. Data for the investment periods and terminal P&L values is reported in table 3.1 in appendix E. Of the 44 portfolios, 28 have a terminal profit which exceeds the initial portfolio value. Unsurprisingly, the biggest losses are incurred in the wake of the Lehman default in 2008. The average relative return of the strategy is 7.6100 \$ with a standard deviation of 15.0471 \$, corresponding to a return per unit std. dev. of 0.5057. The 5% value at risk (VaR) amounts to a loss of 14.5569 \$. The relative returns of the Merton strategy are illustrated in figure 3.3, in which we also exhibit the corresponding numbers for a risk-neutral investor ($\gamma = 1$) who allocates funds equally between risky and risk free assets. As expected, the latter comes with a greater mean return (11.70 \$), but also a considerably higher dispersion (37.89 \$), i.e. a sub-par return per unit std. dev. of only 0.3088.

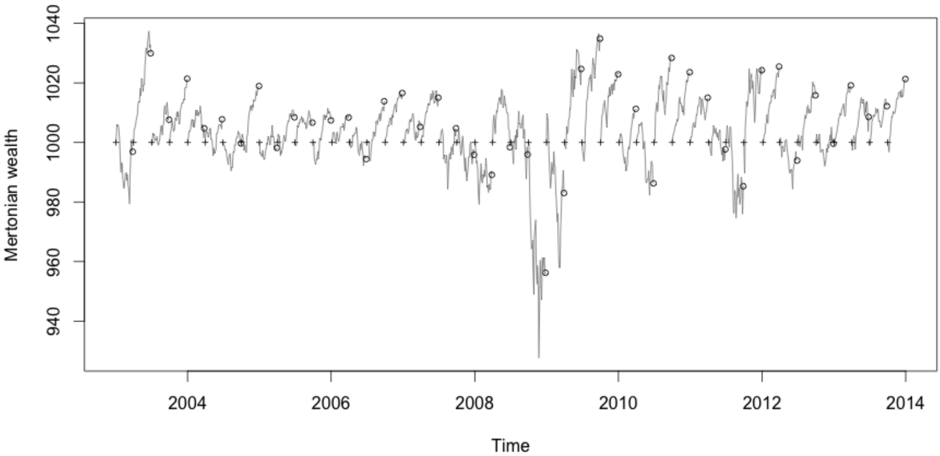

Fig. 3.2 The running wealth from an agent with risk aversion $\gamma = 2.5$ who invests a proportion $\pi = 20\%$ in the S&P500 index and a proportion $1 - \pi = 80\%$ at the risk-free rate 2% during 44 consecutive periods of 3 months each. The initial wealth of each period ($w_0 = 1000$) is marked with a cross while the terminal wealth is marked with a circle.

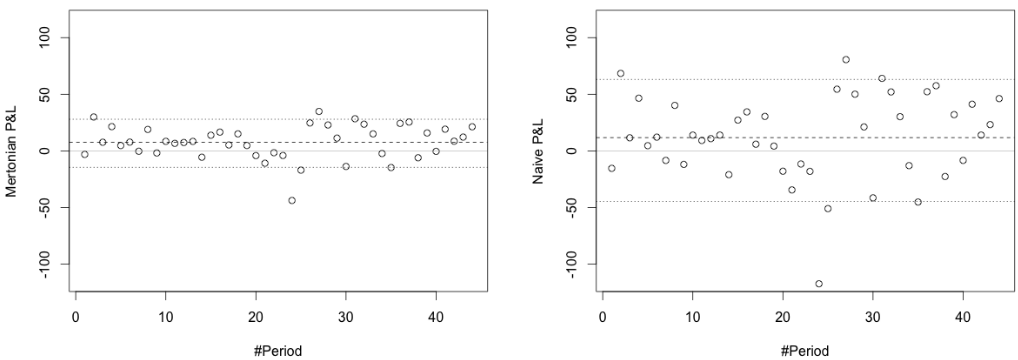

Fig. 3.3 The profit-and-loss of each period from Mertonian optimal investment (left) and the naïve (50-50) investment strategy (right). The mean, 5%- and 95% quantiles are plotted with dashed lines.

3.8.3 The Hestonian Portfolio Strategy

We now repeat the experiment for the Heston portfolio weight, π_t^{He}. The calibrated parameter values κ, θ and ξ are reported in table 3.2 in appendix E for each of the 44 investment periods. The magnitudes of the associated portfolio weights are plotted in figure 3.4. Note

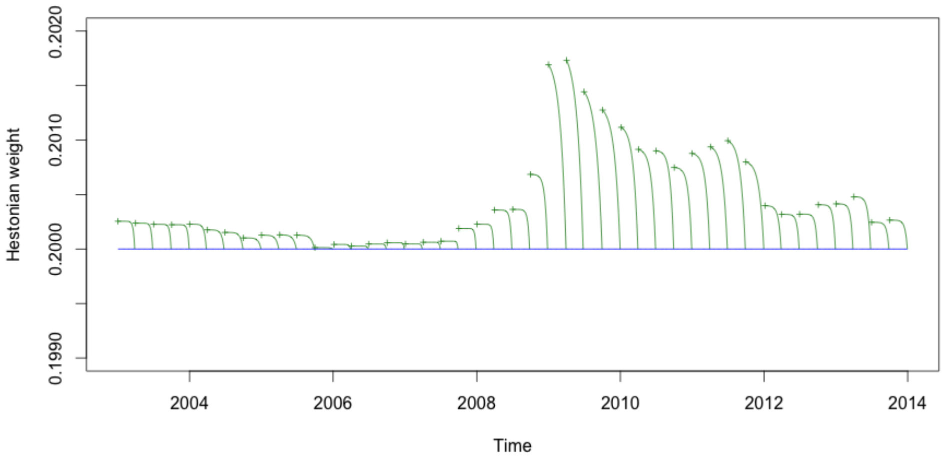

Fig. 3.4 The Hestonian trading weight π_t^{He} used for each of the 44 periods is plotted with a green line while the constant Mertonian weight is plotted in blue. The start of each period is marked with a cross.

that every portfolio weight decreases monotonically to the Mertonian benchmark as the time to maturity goes to zero as suggested by our analysis above. Furthermore, observe that the Heston correction constitutes an extremely modest perturbation from the Merton weight: specifically, the Heston correction consistently suggests that the investor should increase her holding in the risky asset with less than $1/5^{th}$ of a percentage point vis-à-vis the Merton ratio. This answers our question about the magnitude of the volatility correction. As for the performance, upon plotting the running wealth we obtain a graph which is virtually indistinguishable from figure 3.2 (again, data for the investment periods and terminal P&L values is reported in table 3.1 in appendix E). Only by plotting the relative difference between the two graphs (see figure 3.5) do we get an idea about the significance of hedging stochastic volatility: in the relatively homeostatic pre-financial crisis market the effect is virtually non-existent. As the market crashes, the Heston strategy takes some modest hits, which may be symptomatic of using estimated parameters from a quiet period. Finally, in the subsequent period of relative turbulence the effect is a melange of dubious near-insignificant gains and losses. Overall, the average relative return comes out at 7.6161 \$, with a standard deviation of 15.1000 \$, hence a return per std. dev. of 0.5044. The 5% VaR amounts to a loss of 14.6502 \$. Evidently, these numbers provide *no* support whatsoever to the hypothesis that the Heston strategy outperforms the Mertonian benchmark.

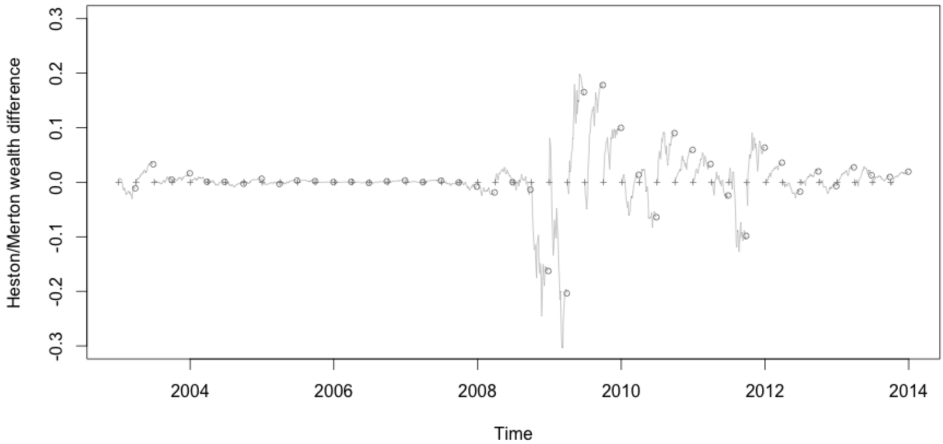

Fig. 3.5 Difference between the running wealth from the Hestonian and Mertonian investment strategies. The units on the ordinate axis are measured in [$] with an initial investment of 1000 $.

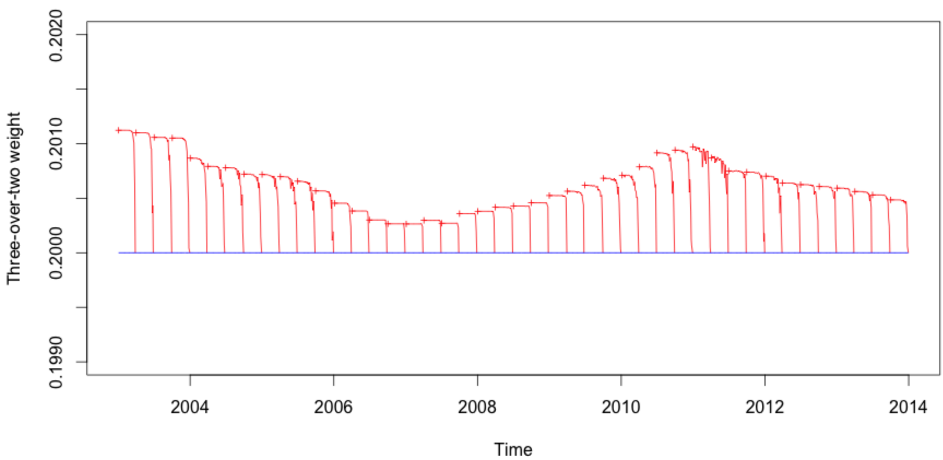

Fig. 3.6 The 3-over-2 weight $\pi_t^{3/2}(V_t)$ in red and the Mertonian weight π^{Me} in blue. We emphasise that the weights here vary both with respect to time to maturity and the running variance.

3.8.4 The 3/2 Portfolio Strategy

Finally, we turn our attention to the performance of the $3/2$ portfolio weight, $\pi_t^{3/2}$, disregarding the unfortunate fact that the $\rho = -1$ assumption used in deriving (3.14) does not

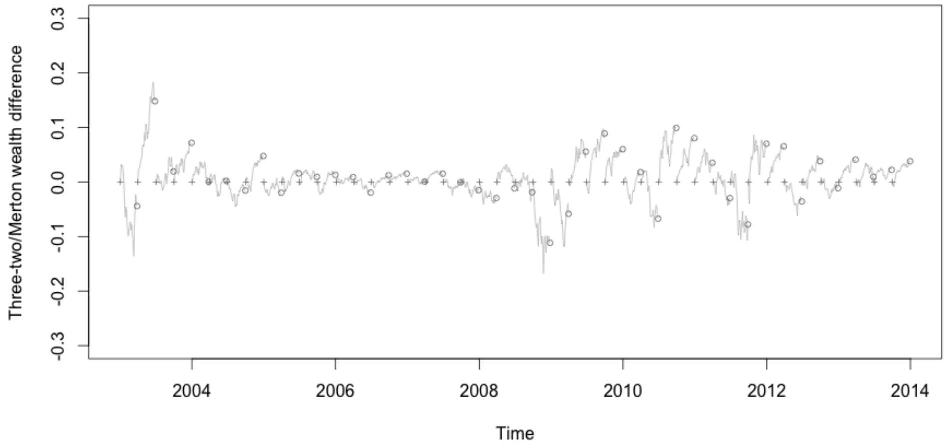

Fig. 3.7 Difference between the running wealth from the 3-over-2 and Mertonian investment strategies. The units on the ordinate axis are measured in [$] with an initial investment of 1000 $.

reflect the actual state of the market (see table 3.2, which also features the calibrated parameters κ, θ and ξ for the 3/2 model). The magnitude of the associated portfolio weights are plotted in figure 3.6. The key difference from the Hestonian case is that the portfolio weight now is a function of the instantaneous variance, wherefore the decay to the Mertonian benchmark as $\tau \to 0$ no longer is monotone. The day-to-day fluctuations brought about by changes in the variance process are nonetheless not big enough to be immediately visible on the figure. Again, the magnitude of the volatility corrections to the Mertonan benchmark remain minuscule, coming in at less than $1/5^{th}$ of a percentage point. We report the terminal P&L of the strategy in table 3.1 and provide a visual overview of the running P&L with respect to the Merton strategy in figure 3.7. Given the minuteness of the portfolio corrections, the associated profits and losses are correspondingly diminutive. In concrete terms we find the average relative return to be 7.6222 $ with a standard deviation of 15.0938 $, corresponding to a return per standard deviation of 0.5050. The 5% VaR amounts to a loss of 14.6331 $. Again, it is clear that the 3/2 volatility correction has no merit whatsoever with respect to the Merton strategy.

3.9 Conclusion

In this paper we have derived closed form expressions for the optimal hedge ratio for a wealth maximiser exposed to various types of stochastic volatility. To this end we had to posit that the Sharpe ratio of the stock is proportional to volatility. Whilst the hedge ratio for the Heston model was found to be independent of the volatility process, the same conclusion does not apply to the 3/2 model or the Fong Vasicek type model - both of which depend on v through confluent hypergeometric functions. Plenty of mathematical research questions are still left unanswered by our study: for example, it is non-obvious if closed form expressions can be derived for more plausible volatility models (such as the lognormal ones - see [16]), and if consumption can be incorporated into the picture.

In order to assess the gravity of our prescriptive standards for rational investments, we put the Heston model and the 3/2 model to the test for an investor who trades in a representative market index. Interestingly, from an empirical perspective there is nothing whatsoever to recommend the notion that stochastic volatility should be hedged in a wealth optimisation context in a bond-stock economy. For both the Heston model and the 3/2 model, the corrections to the Merton ratio are less than $1/5^{th}$ of a percentage point and do nothing to increase the return per unit standard deviation ratio. Indeed, in 17 out of the 44 trading periods, the Merton strategy is marginally better. Either way, the welfare gains and losses induced by the volatility sensitive strategy amounts to less than $1/50\%$ of the initial investment over a three months period. Of course, our analysis here can hardly be said to be void of pitfalls: e.g. it would generally be desirable to (i) work with more realistic market assumptions (in particular, a market with derivatives), (ii) work under a more empirically accurate volatility model. Some steps towards mitigating these rudiments are found in the subsequent chapter.

References

1. Bakshi, Ju, and Yang, *Estimation of continuous time models with an application to equity volatility*, *Journal of Financial Economics*, 2004, 82(1), 227-249.
2. Björk, *Arbitrage Theory in Continuous Time*, Oxford University Press, 3rd edition.
3. Branger and Hansis, *Asset Allocation: How Much Does Model Choice Matter?*, *Journal of Banking & Finance*, July 2012, Volume 36, Issue 7, pp. 1865-1882.
4. Brennan and Xia, *Stochastic Interest Rates and the Bond-Stock Mix*, *European Finance Review* (2000) 4 (2): 197-210.
5. Burden and Faires, *Numerical Analysis*. Brooks/Cole, 9th edition.
6. Carr and Sun, *A New Approach for Option Pricing under Stochastic Volatility*. In *Review of Derivatives Research*, May 2007, Volume 10, Issue 2, pp 87-150.
7. Chacko and Viceira, *Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets*. Rev. Financ. Stud. (Winter 2005) 18 (4): 1369-1402.
8. Dimitroff, Lorenz and Szimayer, *A Parsimonious Multi-Asset Heston Model: Calibration and Derivative Pricing*. 2009, *Fraunhofer ITWM*.
9. Drimus, *Options on realized variance by transform methods: A non-affine stochastic volatility model*. 2011, *Quantitative Finance*.
10. Duffie, *Dynamic Asser Pricing Theory*. Princeton University Press, 3rd edition.
11. Gatheral, Jaisson, Rosenbaum, *Volatility is Rough*, 2014. http://arxiv.org/pdf/1410.3394v1.pdf.
12. Heston, *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*, *The Review of Financial Studies*, 1993, Volume 6, number 2, pp. 327-343.
13. Jones, *The dynamics of stochastic volatility: evidence from underlying and options markets*, *Journal of Econometrics*, 2003, 116, 118-224.
14. Kraft, *Optimal Portfolios and Heston's Stochastic Volatility Model: An Explicit Solution for Power Utility*, Quantitative Finance, Volume 5, Issue 3, 2005, pp. 303-313,
15. Liu, *Portfolio Selection in Stochastic Environments*, in *The Review of Financial Studies*, Vol. 20, No. 1 (Jan., 2007), pp. 1-39.
16. Liu and Pan, *Dynamic Derivative Strategies*, in *Journal of Financial Economics*, Volume 69, Issue 3, September 2003, pp. 401-430.
17. MacDonald, *Properties of the Confluent Hypergeometric Function*. Technical Report No. 84. 1948. Research Laboratory of Electronics, MIT. http://dspace.mit.edu/bitstream/handle/1721.1/4966/RLE-TR-084-14234239.pdf?sequence=1.
18. Merton, *Lifetime Portfolio Selection under Uncertainty: the Continuous-Time Case*. *The Review of Economics and Statistics* 51 (3) pp. 247?257 (1969).
19. Munk, *Dynamic Asset Allocation*.
Unpublished lecture notes. http://mit.econ.au.dk/vip_htm/cmunk/noter/dynassal.pdf.
20. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications*. *Stochastic Modelling and Applied Probability 61*, Springer.
21. Ross, *Stochastic Control in Continuous Time*.
Unpublished lecture notes. http://www.swarthmore.edu/NatSci/kross1/Stat220notes.pdf.
22. Selby and Strickland, *Computing the Fong and Vasicek Pure Discount Bond Price Formula*. *The Journal of Fixed Income*, September 1995, Vol. 5, No. 2: pp. 78-84.
23. Sorensen, *On the Asymptotetics of Estimating Functions*. *Brazilian Journal of Probability and Statistics*, 1999, 13, pp. 111-136.
24. Vasicek and Fong, *Fixed-income Volatility Management*, *The Journal of Portfolio Management*, Volume 17, No. 4, Summer 1991, pages 41-46.
25. Vasicek and Fong, *Interest Rate Volatility as a Stochastic Factor*, 1991, Gifford Fing Associates working paper.
26. Wong and Heyde, *On Changes of Measure in Stochastic Volatility Models*, *Journal of Applied Mathematics and Stochastic Analysis*, Volume 2006, Article ID 18130, pages 1-13.
27. Zariphopoulou, *A Solution Approach to Valuation with Unhedgeable Risks*, *Finance and Stochastics*, 5, pp. 61–82, 2001.

Appendix A: Proofs

Heston's Model

Proof: Substituting (3.9) into (3.7) we observe that all coefficients are either constant or depend on v or $\sqrt{v}$. Insofar as we aim for a solution of the exponential affine form, the latter is clearly undesirable. However, this problem is readily mitigated if we make the bold move of positing that either $\lambda_1(v) = \lambda\sqrt{v}$ (where $\lambda \in \mathbb{R}^+$ is a constant of proportionality) or $\lambda_1(v) = 0$. On empirical grounds we opt for the former, as there is little to suggest that we live in a world of risk-neutrality. Thus, the PDE to be solved is

$$\begin{aligned} 0 &= \partial_t g - \left(\tfrac{\delta}{\gamma} - \hat{\gamma} r - \tfrac{1}{2}\gamma^{-1}\hat{\gamma}\lambda^2 v\right) g + (\kappa(\theta - v) + \hat{\gamma}\rho\xi\lambda v)\,\partial_v g \\ &\quad + \tfrac{1}{2}\xi^2 v \partial^2_{vv} g - \tfrac{1}{2}(1-\gamma)(1-\rho^2)\xi^2 v \frac{(\partial_v g)^2}{g}, \end{aligned} \tag{3.33}$$

with the usual boundary condition, $g(T,v) = 1$. Clearly, this beckons us to attempt a solution of the form

$$g(t,v) = \exp\left\{-(\tfrac{\delta}{\gamma} - \hat{\gamma} r)(T-t) + \hat{\gamma} A(T-t) + \hat{\gamma} B(T-t) v\right\}, \tag{3.34}$$

where $A, B : \mathbb{R} \mapsto \mathbb{R}$ are deterministic functions of the time to maturity $\tau = T - t$, which satisfy the conditions $A(0) = B(0) = 0$. Specifically, upon combining (3.33) and (3.34) we arrive at the PDE[15]

$$0 = \frac{\lambda^2 v}{2\gamma} - A'(\tau) - B'(\tau)v + (\kappa(\theta - v) + \hat{\gamma}\rho\xi\lambda v)B(\tau) + \tfrac{1}{2}\hat{\gamma}\xi^2 v\left(\rho^2 + \gamma[1-\rho^2]\right)B^2(\tau),$$

where $'$ denotes the temporal derivative $d/d\tau$. It is readily seen that this equation is of the form $0 = P(B',B)v + Q(A',B)$, where

$$P(B',B) \equiv \frac{\lambda^2}{2\gamma} - B'(\tau) - (\kappa - \hat{\gamma}\rho\xi\lambda)B(\tau) + \tfrac{1}{2}\hat{\gamma}\xi^2\left(\rho^2 + \gamma[1-\rho^2]\right)B^2(\tau), \tag{3.35a}$$

$$Q(A',B) \equiv -A'(\tau) + \kappa\theta B(\tau). \tag{3.35b}$$

Since $0 = P(B',B)v + Q(A',B)$ should be satisfied for all values of the instantaneous variance v, we are forced to set $P(B',B) = 0$ and $Q(A',B) = 0$. This gives us two coupled (but ultimately solvable) ODEs. In particular, $P(B',B) = 0$ is a Riccati equation with constant coefficients, the solution to which is derived in Lemma 3.1, formula (3.64). Under the assumption that $\hat{\eta} \in \mathbb{R}^+$ (which certainly is satisfied for the empirically plausible parameter choice of $\gamma > 1$) the answer is given by

[15] Here we use that $\partial_v g = \hat{\gamma} B(\tau) g$, $\partial^2_{vv} g = \hat{\gamma}^2 B^2(\tau) g$ and $\partial_t g = -\partial_\tau g = (\gamma^{-1}\delta - \hat{\gamma} r - \hat{\gamma} A'(\tau) - \hat{\gamma} B'(\tau) v) g$.

$$B(\tau) = \frac{\lambda^2}{\gamma} \frac{(e^{\hat{\eta}\tau} - 1)}{(\hat{\kappa} + \hat{\eta})(e^{\hat{\eta}\tau} - 1) + 2\hat{\eta}}, \tag{3.36}$$

From this, we can now use the equation $Q(A', B) = 0 \Leftrightarrow A(\tau) = \kappa\theta \int_0^\tau B(s)ds$ to compute $A(\tau)$ in closed form, although this turns out to be redundant in terms of computing the optimal control π_t^*. We state the result here only for completeness and refer the reader to Lemma 3.2, formula (3.66), for details:

$$A(\tau) = \frac{\kappa\theta}{\hat{\gamma}\xi^2(\rho^2 + \gamma[1 - \rho^2])} \left\{ (\hat{\kappa} + \hat{\eta})\tau + 2\ln \left| \frac{2\hat{\eta}}{(\hat{\kappa} + \hat{\eta})(e^{\hat{\eta}\tau} - 1) + 2\hat{\eta}} \right| \right\}. \tag{3.37}$$

Finally, we are now in a position to calculate the optimal portfolio weight for an investor living in a Hestonian economy: by substituting $\lambda_1 = \lambda\sqrt{v}$ and the exponential affine solution (3.34) into (3.8) we obtain

$$\pi_t^* = \frac{\lambda}{\gamma} + \rho\xi\hat{\gamma}B(T - t). \tag{3.38}$$

which completes the proof. □

The Time-dependent Heston Model

Proof: Imposing the condition $\lambda_1(v) = \lambda\sqrt{v}$ yields a spatially reduced HJB equation with coefficients that are linear in v

$$\begin{aligned} 0 = \partial_t g - \left(\tfrac{\delta}{\gamma} - \hat{\gamma}r - \tfrac{1}{2}\gamma^{-1}\hat{\gamma}\lambda^2 v\right) g + (\kappa(t)(\theta(t) - v) + \hat{\gamma}\rho(t)\xi(t)\lambda v)\partial_v g \\ + \frac{1}{2}\xi^2(t)v\partial_{vv}^2 g - \frac{1}{2}(1 - \gamma)(1 - \rho^2(t))\xi^2(t)v\frac{(\partial_v g)^2}{g}, \end{aligned} \tag{3.39}$$

with terminal condition $g(T, v) = 1$. Clearly, the trick is to solve n PDEs sequentially, starting with the interval $(t_{n-1}, t_n]$ (for which we have the boundary condition $g(t_n, v) = 1$) and working our way backwards in time to the interval $[t, t_1]$. To this end, it will be convenient to work with inverse time $\tau = T - s$ and thence the inverse discontinuity-points $\tau_k = T - t_{n-k}$ for $k = 0, \dots, n - 1$.

Starting with the first subinterval from the end, $s \in (t_{n-1}, T]$, our problem is that of solving a PDE of inverse time over $\tau \in [0, \tau_1)$ with the initial condition $g(0, v) = 1$. Indeed, by setting $\kappa(t) = \bar{\kappa}_1, \xi(t) = \bar{\xi}_1$ and $\rho(t) = \bar{\rho}_1$ we see that the problem for all practical purposes is indistinguishable from the one we solved above. Thus, we guess the solution

$$g(\tau, v) = \exp\left\{ -\left(\tfrac{\delta}{\gamma} - \hat{\gamma}r\right)\tau + \hat{\gamma}A_1(\tau) + \hat{\gamma}B_1(\tau)v \right\}, \quad \tau \in [0, \tau_1), \tag{3.40}$$

where $A_1, B_1 : [0, \tau_1) \mapsto \mathbb{R}$ are functions that satisfy the conditions $A_1(0) = B_1(0) = 0$. Substituting the exponential form of inverse time in equation (3.40) into equation (3.39) yields the PDE

$$\frac{\lambda^2}{2\gamma}v - A_1'(\tau) - B_1'(\tau)v + (\bar{\kappa}_1(\theta(\tau)-v) + \hat{\gamma}\lambda\bar{\rho}_1\bar{\xi}_1 v)B_1(\tau) + \frac{1}{2}\hat{\gamma}\bar{\xi}_1^2 v(\bar{\rho}_1^2 + \gamma(1-\bar{\rho}_1^2))B_1^2(\tau) = 0. \tag{3.41}$$

Since equation (3.41) must hold for all values of v, we end up with the coupled system of ODEs

$$\frac{\lambda^2}{2\gamma} - B_1'(\tau) - (\bar{\kappa}_1 - \hat{\gamma}\lambda\bar{\rho}_1\bar{\xi}_1)B_1(\tau) + \frac{1}{2}\hat{\gamma}\bar{\xi}_1^2(\bar{\rho}_1^2 + \gamma(1-\bar{\rho}_1^2))B_1^2(\tau) = 0 \tag{3.42a}$$

$$-A_1'(\tau) + \bar{\kappa}_1\theta(\tau)B_1(\tau) = 0. \tag{3.42b}$$

In particular, we recognise equation (3.42a) for B_1 as a Riccati equation with constant parameters and initial condition $B_1(0) = 0$. Under the assumption that $\hat{\eta}_1 \geq 0$, we once again obtain

$$B_1(\tau) = \frac{\lambda^2}{\gamma}\frac{e^{\hat{\eta}_1\tau} - 1}{(\hat{\kappa}_1 + \hat{\eta}_1)(e^{\hat{\eta}_1\tau} - 1) + 2\hat{\eta}_1}, \quad \tau \in [0, \tau_1) \tag{3.43}$$

where $\hat{\kappa}_1 = \bar{\kappa}_1 - \hat{\gamma}\lambda\bar{\rho}_1\bar{\xi}_1$ and $\hat{\eta}_1 = \sqrt{\hat{\kappa}_1^2 - \hat{\gamma}\bar{\xi}_1^2(\bar{\rho}_1^2 + \gamma(1-\bar{\rho}_1^2))\gamma^2/\gamma}$. From this, we obtain the value at the terminal point $B_1(\tau_1)$ which will act as an initial condition for B_2 when we proceed to the next time interval and we denote it $B_2^0 = B_1(\tau_1)$. However, before we do this, observe that equation (3.42b) and the initial condition $A_1(0) = 0$ give $A_1(\tau) = \int_0^\tau \bar{\kappa}_1\theta(u)B_1(u)du$ for $\tau \in [0, \tau_1)$ which might be analytically computable depending on the form of θ. In particular, if θ is piecewise constant over the same subintervals as κ, ξ, ρ, we have a closed form solution while we may resort to numerical integration if θ is a non-constant function.

Next, we proceed to the second subinterval $(t_{n-2}, t_{n-1}]$ which translates to solving the PDE of $g(\tau, v)$ for $\tau \in [\tau_1, \tau_2)$. Assuming a solution of the same exponential form as in equation (3.40) for $g(\tau, v)$ with B_2, A_2 and notice that $B_2(\tau_1) = B_1(\tau_1) = B_2^0$ must hold at the initial point, and thus $A_2(\tau_1) = A_1(\tau_1) = A_2^0$ which constitutes the initial value $g(\tau_1, v)$ for the PDE. Hence, we have a system of ODEs, with $k = 2$ (written out for a general $k = 2, \ldots, n$)

$$\frac{\lambda^2}{2\gamma} - B_k'(\tau) - (\bar{\kappa}_k - \hat{\gamma}\lambda\bar{\rho}_k\bar{\xi}_k)B_k(\tau) + \frac{1}{2}\hat{\gamma}\bar{\xi}_k^2(\bar{\rho}_k^2 + \gamma(1-\bar{\rho}_k^2))B_k^2(\tau) = 0 \tag{3.44}$$

$$-A_k'(\tau) + \bar{\kappa}_k\theta(\tau)B_k(\tau) = 0 \tag{3.45}$$

for $\tau \in [\tau_{k-1}, \tau_k)$ with non-zero initial conditions $B_k(\tau_{k-1}) = B_{k-1}(\tau_{k-1}) = B_k^0$ and $A_k(\tau_{k-1}) = A_{k-1}(\tau_{k-1}) = A_k^0$. We recognise equation (3.44) as a Riccati equation with constant parameters and an initial condition that is non-zero. Under the assumption that $\hat{\eta}_k \geq 0$, we obtain the solution from lemma 3.3, equation (3.70):

$$B_k(\tau) = B_k^0 + \frac{2(a_k(B_k^0)^2 + b_kB_k^0 + \frac{\lambda^2}{2\gamma})(e^{\eta_k(\tau-\tau_{k-1})} - 1)}{(\eta_k - b_k - 2a_kB_k^0)(e^{\eta_k(\tau-\tau_{k-1})} - 1) + 2\eta_k}, \quad \tau \in [\tau_{k-1}, \tau_k) \tag{3.46}$$

where $\tau_k = T - t_{n-k}$ with $\tau_0 = 0$, $B_k^0 = B_{k-1}(\tau_{k-1})$ for $k = 2,..,n$ with $B_1^0 = 0$ and we have defined the parameters

$$a_k = \frac{1}{2}\hat{\gamma}\bar{\xi}_k^2(\bar{\rho}_k^2 + \gamma(1-\bar{\rho}_k^2)), \quad b_k = -(\bar{\kappa}_k - \hat{\gamma}\lambda\bar{\rho}_k\bar{\xi}_k), \quad \eta_k = \sqrt{b_k^2 - 4a_k c}$$

for $k = 1,\ldots,n$. Notice that we may include $k = 1$ since by setting $B_1^0 = 0$, equation (3.46) reduces to equation (3.43) for $k = 1$.

Calculating the $\{B_k\}$ sequentially in this manner for $k = 1,2,...,n$, we find that

$$g(s,v) = g(T-s,v) = \exp\left\{-\left(\tfrac{\delta}{\gamma} - \hat{\gamma}r\right)(T-s) + \hat{\gamma}A_n(T-s) + \hat{\gamma}B_n(T-s)v\right\},$$

for $s \in [t,t_1] \Leftrightarrow s \in [\tau_{n-1}, T-t]$. This relies on $\{A_k\}_{k=1}^n$ where equation (3.45) gives $A_k(\tau) = A_k^0 + \int_{\tau_{k-1}}^{\tau} \bar{\kappa}_k\theta(u)B_k(u)du$ for $\tau \in [\tau_{k-1},\tau_k)$ where $A_k^0 = A_{k-1}(\tau_{k-1})$ except for $A_1^0 = 0$. Even if we can compute this in closed form (which is the case if θ is piecewise constant), this turns out to be redundant in terms of computing the optimal control $\pi^*(t,w,v)$. Specifically, it is readily demonstrated that the FOC (3.8) reduces to the expression

$$\pi_t^* = \frac{\lambda}{\gamma} + \bar{\rho}_n\bar{\xi}\hat{\gamma}B_n(T-t),$$

which completes the proof. □

Remark 3.4. We observe that the optimal control once again is independent of the instantaneous variance in analogy with (4.1). Furthermore, under the assumption that θ is piecewise constant with value $\bar{\theta}_k$ for $\tau \in [\tau_{k-1},\tau_k)$, the solution to equation (3.45) with initial condition A_k^0 is given by

$$A_k(\tau) = A_k^0 - \frac{\bar{\kappa}_k\bar{\theta}_k}{2a_k}\left((b_k+\eta_k)(\tau-\tau_{k-1}) + 2\ln\left(\frac{1-\beta_k e^{-\eta_k(\tau-\tau_{k-1})}}{1-\beta_k}\right)\right),$$

where

$$\beta_k = \frac{b_k + \eta_k + 2a_kB_k^0}{b_k - \eta_k + 2a_kB_k^0}.$$

The 3/2 Model

Proof: Combining (3.7) and (3.13) and setting $\rho = -1$ we find that the governing PDE is

$$0 = \partial_t g - \left(\tfrac{\delta}{\gamma} - \hat{\gamma}r - \tfrac{1}{2}\gamma^{-1}\hat{\gamma}\lambda^2 v\right)g + (\kappa v(\theta - v) - \hat{\gamma}\xi\lambda v^2)\partial_v g + \tfrac{1}{2}\xi^2v^3\partial_{vv}^2 g \qquad (3.47)$$

with the terminal condition $g(t,v) = 1$. First consider the substitution

$$g(t,v) = \exp\left\{-(\tfrac{\delta}{\gamma} - \hat{\gamma}r)(T-t)\right\}\cdot C(t,v) \qquad (3.48)$$

which yields

$$0=\partial_t C+\tfrac{1}{2}\gamma^{-1}\hat{\gamma}\lambda^2 vC+\left(\kappa\theta v-(\hat{\gamma}\xi\lambda+\kappa)v^2\right)\partial_v C+\tfrac{1}{2}\xi^2 v^3\partial_{vv}^2 C \tag{3.49}$$

where $C(T,v)=1$. From Carr & Sun [9] it is known that PDEs of this form can be transformed into confluent hypergeometric differential equations. Specifically, insofar that we make the key conjecture that C depends on (t,v) only through the intervening variable $y(t,v)$:

$$y(t,v)\equiv\int_t^T e^{\int_t^s \kappa\theta ds}du\cdot v=\left(\frac{e^{\kappa\theta(T-t)}-1}{\kappa\theta}\right)\cdot v, \tag{3.50}$$

(3.49) reduces to the second order ODE[16]

$$0=\tfrac{1}{2}\xi^2 y^2 C''(y)-[(\hat{\gamma}\xi\lambda+\kappa)y+1]C'(y)+\tfrac{1}{2}\gamma^{-1}\hat{\gamma}\lambda^2 C(y), \tag{3.51}$$

where $'\equiv d/dy$. The boundary condition $C(T,v)=1$ may now be stated as $C(y=0)=1$. Performing one final change of variables

$$C(y)=z^a D(z),\quad \text{s.t.}\quad z\equiv\frac{b}{y},$$

where a and b are constants to be fixed at our convenience, tedious calculations show that

$$\begin{aligned}0=\tfrac{1}{2}\xi^2 z^{a+2}D''(z)&+\left\{[\xi^2(a+1)+(\hat{\gamma}\xi\lambda+\kappa)]z^{a+1}+\frac{1}{b}z^{a+2}\right\}D'(z)\\&+\left\{[\tfrac{1}{2}\xi^2 a(a+1)+(\hat{\gamma}\xi\lambda+\kappa)a+\tfrac{1}{2}\gamma^{-1}\hat{\gamma}\lambda^2]z^a+\frac{a}{b}z^{a+1}\right\}D(z),\end{aligned} \tag{3.52}$$

where we have used the results

$$C'(y)=-\tfrac{a}{b}z^{a+1}D(z)-\tfrac{1}{b}z^{a+2}D'(z), \tag{3.53a}$$

$$C''(y)=\tfrac{a(a+1)}{y^2}z^a D(z)+2\tfrac{(a+1)}{y^2}z^{a+1}D'(z)+\tfrac{1}{y^2}z^{a+2}D''(z). \tag{3.53b}$$

While this hardly looks like a simplification, we note that our freedom in a and b allow us to set the last square bracket to zero, i.e.

$$0=\tfrac{1}{2}\xi^2 a(a+1)+(\hat{\gamma}\xi\lambda+\kappa)a+\tfrac{1}{2}\gamma^{-1}\hat{\gamma}\lambda^2.$$

Solving this quadratic equation for a we find that

$$a_\pm=-\hat{\omega}\pm\sqrt{\hat{\omega}^2-\frac{\hat{\gamma}\lambda^2}{\gamma\xi^2}},\quad\text{where}\quad \hat{\omega}\equiv\frac{1}{2}+\frac{\hat{\gamma}\xi\lambda+\kappa}{\xi^2}, \tag{3.54}$$

where we observe that the discriminant is positive insofar $\gamma>1$. This is the same assumption we made upon solving the Merton-Heston problem. Hence, for $\gamma>1$, a_+ is a real

[16] Here we use that $\partial_v C(y)=C'(y)\frac{y}{v}$, $\partial_{vv}C(y)=C''(y)\left(\frac{y}{v}\right)^2$ and $\partial_t C(y)=-C'(y)[v+\kappa\theta y]$.

positive number, which is the parametric choice we shall be opting for (NB: for ease of notation, the subscript $_+$ will henceforth be supressed). This allows us to rewrite (3.52) as

$$0 = \tfrac{1}{2}\xi^2 z^{a+2} D''(z) + \left\{ [\xi^2(a+1) + (\hat{\gamma}\xi\lambda + \kappa)] z^{a+1} + \frac{z^{a+2}}{b} \right\} D'(z) + \frac{a}{b} z^{a+1} D(z)$$

or equivalently

$$0 = zD''(z) + \left\{ \hat{\zeta} + \frac{2}{b\xi^2} z \right\} D'(z) + \frac{2a}{b\xi^2} D(z),$$

where

$$\hat{\zeta} \equiv 1 + 2(a + \hat{\omega}). \tag{3.55}$$

Finally, by fixing $b = -2/\xi^2$ we arrive at the confluent hypergeometric differential equation

$$0 = zD''(z) + \{\hat{\zeta} - z\} D'(z) - aD(z). \tag{3.56}$$

This is a comforting result in the sense that we can refer to a substantial body of literature to extract the solution (and the various properties thereof). For the reader's convenience, a cursory primer on the mathematics of confluent hypergeometric functions is provided in appendix C. The basic result presented here is that (3.56) admits the general solution

$$D(z) = K_1 M(a; \hat{\zeta}; z) + K_2 U(a; \hat{\zeta}; z),$$

where K_1, K_2 are arbitrary constants, $M(a; \hat{\zeta}; z)$ is Kummer's function of the first kind, and $U(a; \hat{\zeta}; z)$ is the Tricomi function[17]

$$U(a; \hat{\zeta}; z) = \frac{\Gamma(1-\hat{\zeta})}{\Gamma(a-\hat{\zeta}+1)} M(a; \hat{\zeta}; z) + \frac{\Gamma(\hat{\zeta}-1)}{\Gamma(a)} z^{1-\hat{\zeta}} M(a-\hat{\zeta}+1; 2-\hat{\zeta}; z).$$

We guess (correctly) that the boundary conditions $C(0) = 1$ and $C'(0) = 0$ can be met by setting $K_2 = 0$. To determine K_1, recall that as $y \downarrow 0$, $C(y) \downarrow 1$. At the same time, since $z = -2/(\xi^2 y)$, it must be the case that as $y \downarrow 0$, $z \downarrow -\infty$. Now it is a known fact that for large negative zs the confluent hypergeometric function scales as $M(a; \hat{\zeta}; z) \sim \Gamma(\hat{\zeta})(-z)^{-a}/\Gamma(\hat{\zeta} - a)$ where Γ is the gamma function, cf. theorem 3.9. Thus, since $C(y) = z^a D(z) = K_1 z^a M(a; \hat{\zeta}; z)$ which goes to $K_1 \Gamma(\hat{\zeta})/(\Gamma(\hat{\zeta} - a)(-1)^a)$, as $z \downarrow -\infty$, it must be the case that $K_1 = \Gamma(\hat{\zeta} - a)(-1)^a/\Gamma(\hat{\zeta})$ i.e. the solution to (3.56) is

$$D(z) = \frac{\Gamma(\hat{\zeta} - a)(-1)^a}{\Gamma(\hat{\zeta})} M(a; \hat{\zeta}; z). \tag{3.57}$$

We are now in a position to compute the optimal portfolio weight. From (3.8)

---

[17] Notice that M is undefined for $\hat{\zeta} \in \mathbb{Z}^- \cup \{0\}$, whilst U is undefined for $\hat{\zeta} \in \mathbb{Z}$. Neither of these conditions pose a threat to the problem at hand.

$$\pi_t^* = \frac{\lambda}{\gamma} - \xi v \frac{\partial_v g}{g}, \tag{3.58}$$

where we have used that $\rho = -1$, $\lambda_1(v) = \lambda\sqrt{v}$ and $\beta(t,v) = \xi v^{3/2}$. The ratio $\partial_v g/g$ is readily evaluated using the chain rule and Theorem 3.10 in appendix 3.9. The resulting optimal control turns out to be

$$\pi_t^* = \frac{\lambda}{\gamma} + \xi a \left(1 + \frac{z}{\hat{\zeta}} \frac{M(a+1;\hat{\zeta}+1;z)}{M(a;\hat{\zeta};z)}\right), \tag{3.59}$$

where we reiterate that $z = -2/(\xi^2 y)$ where $y = y(v,t)$ is the variable defined in (3.50), a is is the positive solution defined in (3.54) and $\hat{\zeta}$ is the parameter defined in (3.55). $\square$

The Fong-Vasicek Type Model

Proof: The proof of (3.31b) is trivial. To show (3.31c) we combine (3.30c) with (3.31b) to obtain

$$\begin{aligned} B_2'(\tau) = & \left\{ \left[\frac{\lambda_\rho^2}{2\gamma} + \frac{\hat{\gamma}\lambda_1}{\kappa_r} + \frac{\hat{\gamma}}{2\kappa_r^2}\right] - \left[\frac{\lambda_1\hat{\gamma}}{\kappa_r} + \frac{\hat{\gamma}}{\kappa_r^2}\right] e^{-\kappa_r\tau} + \frac{\hat{\gamma}}{2\kappa_r^2} e^{-2\kappa_r\tau} \right\} \\ & - \left\{ \kappa_v - \hat{\gamma}\lambda_2\xi - \frac{\hat{\gamma}\xi\rho}{\kappa_r} + \frac{\hat{\gamma}\xi\rho}{\kappa_r} e^{-\kappa_r\tau} \right\} B_2(\tau) + \tfrac{1}{2}\hat{\gamma}\xi^2 B_2^2(\tau). \end{aligned}$$

Next, we perform a change of dependent variable

$$L(\tau) = \exp\left\{-\tfrac{1}{2}\hat{\gamma}\xi^2 \int_0^\tau B_2(s)ds\right\}, \tag{3.60}$$

such that

$$B_2(\tau) = -\frac{2}{\hat{\gamma}\xi^2}\frac{L'(\tau)}{L(\tau)}, \quad \text{and} \quad B_2'(\tau) = -\frac{2}{\hat{\gamma}\xi^2}\left[\frac{L''(\tau)}{L(\tau)} - \frac{(L'(\tau))^2}{L^2(\tau)}\right], \tag{3.61}$$

to obtain

$$\begin{aligned} & L''(\tau) + \left\{ \kappa_v - \hat{\gamma}\lambda_2\xi - \frac{\hat{\gamma}\xi\rho}{\kappa_r} + \frac{\hat{\gamma}\xi\rho}{\kappa_r} e^{-\kappa_r\tau} \right\} L'(\tau) \\ & + \tfrac{1}{2}\hat{\gamma}\xi^2 \left\{ \left[\frac{\lambda_\rho^2}{2\gamma} + \frac{\hat{\gamma}\lambda_1}{\kappa_r} + \frac{\hat{\gamma}}{2\kappa_r^2}\right] - \left[\frac{\lambda_1\hat{\gamma}}{\kappa_r} + \frac{\hat{\gamma}}{\kappa_r^2}\right] e^{-\kappa_r\tau} + \frac{\hat{\gamma}}{2\kappa_r^2} e^{-2\kappa_r\tau} \right\} L(\tau) = 0, \end{aligned}$$

subject to the boundary condition $L'(\tau = 0) = 0$ (as it will shortly become obvious, one boundary condition suffices to determine the solution of this second order equation). Making the independent variable change

$$x = e^{-\kappa_r \tau},$$

the PDE becomes[18]

$$\kappa_r^2 x^2 L''(x) + \left\{\kappa_r^2 x - \kappa_r \kappa_v x + \hat{\gamma}\lambda_2 \xi \kappa_r x + \hat{\gamma}\xi\rho x - \hat{\gamma}\xi\rho x^2\right\} L'(x)$$
$$+ \tfrac{1}{2}\hat{\gamma}\xi^2 \left\{ \left[\frac{\lambda_\rho^2}{2\gamma} + \frac{\hat{\gamma}\lambda_1}{\kappa_r} + \frac{\hat{\gamma}}{2\kappa_r^2}\right] - \left[\frac{\lambda_1\hat{\gamma}}{\kappa_r} + \frac{\hat{\gamma}}{\kappa_r^2}\right] x + \frac{\hat{\gamma}}{2\kappa_r^2} x^2 \right\} L(x) = 0,$$

with $L'(x=1) = 0$. Finally, we perform the substitution

$$L(x) = x^{\beta} Q(x),$$

where β is a constant to be fixed at our convenience. After a few manipulations this yields[19]

$$xQ''(x) + \left\{2\beta + 1 - \hat{\vartheta} - \frac{\hat{\gamma}\xi\rho}{\kappa_r^2} x\right\} Q'(x) + \left\{ \left(\beta^2 - \beta\hat{\vartheta} + \frac{\hat{\gamma}\xi^2}{2\kappa_r^2}\left[\frac{\lambda_\rho^2}{2\gamma} + \frac{\hat{\gamma}\lambda_1}{\kappa_r} + \frac{\hat{\gamma}}{2\kappa_r^2}\right]\right)\frac{1}{x}\right.$$
$$\left. - \frac{\hat{\gamma}\beta\xi\rho}{\kappa_r^2} - \frac{\hat{\gamma}^2\xi^2}{2\kappa_r^4}[1 + \lambda_1\kappa_r] + \frac{\hat{\gamma}^2\xi^2}{4\kappa_r^4} x \right\} Q(x) = 0,$$

where we have introduced the parameter

$$\hat{\vartheta} \equiv \frac{\kappa_v}{\kappa_r} - \frac{\hat{\gamma}\xi\lambda_1}{\kappa_r} - \frac{\xi\rho\hat{\gamma}}{\kappa_r^2}.$$

Now, it is clearly desirable to choose β such that the coefficient of $\frac{1}{x}$ vanishes. This is a matter of solving a quadratic equation

$$\beta = \frac{\hat{\vartheta}}{2} \pm \frac{1}{2}\sqrt{\hat{\vartheta}^2 - \frac{2\hat{\gamma}\xi^2}{\kappa_r^2}\left[\frac{\lambda_\rho^2}{2\gamma} + \frac{\hat{\gamma}\lambda_1}{\kappa_r} + \frac{\hat{\gamma}}{2\kappa_r^2}\right]}.$$

We pick the left solution and label it $\hat{\beta}$. Thus, we are left with a simple second order equation, the coefficients of which depend linearly on the independent variable x:

$$xQ''(x) + \left\{2\hat{\beta} + 1 - \hat{\vartheta} - \frac{\hat{\gamma}\xi\rho}{\kappa_r^2} x\right\} Q'(x) + \left\{ - \frac{\hat{\gamma}\hat{\beta}\xi\rho}{\kappa_r^2} - \frac{\hat{\gamma}^2\xi^2}{2\kappa_r^4}[1 + \lambda_1\kappa_r] + \frac{\hat{\gamma}^2\xi^2}{4\kappa_r^4} x \right\} Q(x) = 0.$$

It is well-known that the solution of such equations can be expressed in terms of confluent hypergeometric functions (see Theorem 3.11 and Corollary 3.1 in the appendix). Specifically, we find that

[18] Here we use that $x^2 = e^{-2\kappa_r\tau}$, $\frac{d}{d\tau} = -\kappa_r x \frac{d}{dx}$ and $\frac{d^2}{d\tau^2} = \kappa_r^2 x^2 \frac{d^2}{dx^2} + \kappa_r^2 x \frac{d}{dx}$.

[19] The derivatives are $L' = \beta x^{\beta-1} Q + x^{\beta} Q'$ and $L'' = \beta(\beta-1)x^{\beta-2}Q + 2\beta x^{\beta-1}Q' + x^{\beta}Q''$.

$$Q(x) = e^{-\frac{1}{2}\hat{p}x}\{K_1 M(\hat{a},\hat{\zeta},\hat{q}x) + K_2 x^{1-\hat{\zeta}} M(\hat{a}-\hat{\zeta}+1, 2-\hat{\zeta},\hat{q}x)\},$$

where K_1 and K_2 are constants to be fitted to the boundary condition and we have defined the parameters

$$\hat{p} \equiv \frac{\xi}{\kappa_r^2}[i\sqrt{\hat{\gamma}^2(1-\rho^2)} - \hat{\gamma}\rho], \qquad \hat{q} \equiv \hat{p} + \frac{\rho\xi\hat{\gamma}}{\kappa_r^2},$$

$$\hat{a} \equiv \frac{\hat{\zeta}_j}{2} + \frac{i\hat{\gamma}\rho(1-\hat{\vartheta})}{2\sqrt{\hat{\gamma}^2(1-\rho^2)}}\left\{1 - \frac{\hat{\gamma}\xi(1+\lambda_1\kappa_r)}{\rho(1-\hat{\vartheta})\kappa_r^2}\right\}, \quad \hat{\zeta} \equiv 2\hat{\beta} + 1 - \hat{\vartheta}.$$

Recalling that $L(x) = x^{\beta}Q(x)$ and $x = e^{-\kappa_r\tau}$ this means that

$$L(\tau) = K_1 e^{-\kappa_r\hat{\beta}\tau - \frac{1}{2}\hat{p}e^{-\kappa_r\tau}} M(\hat{a},\hat{\zeta},\hat{q}e^{-\kappa_r\tau}) + K_2 e^{-\kappa_r(\hat{\beta}+1-\hat{\zeta})\tau - \frac{1}{2}\hat{p}e^{-\kappa_r\tau}} M(\hat{a}-\hat{\zeta}+1, 2-\hat{\zeta},\hat{q}e^{-\kappa_r\tau}),$$

or, more succinctly,

$$L(\tau) = \sum_{j=1}^{2} K_j e^{-\kappa_r\hat{\beta}_j\tau - \frac{1}{2}\hat{p}e^{-\kappa_r\tau}} M(\hat{a}_j,\hat{\zeta}_j,\hat{q}e^{-\kappa_r\tau}), \tag{3.62}$$

where we have defined $\hat{\beta}_1 = \hat{\beta}$ and $\hat{\beta}_2 = \hat{\vartheta} - \hat{\beta}$, as well as

$$\hat{a}_j \equiv \frac{\hat{\zeta}_j}{2} + \frac{i}{2\sqrt{1-\rho^2}}\left\{\rho(1-\hat{\vartheta}) - \frac{\hat{\gamma}(1+\lambda_1\kappa_r)}{\kappa_r^2}\right\}, \quad \hat{\zeta}_j \equiv 2\hat{\beta}_j + 1 - \hat{\vartheta}.$$

Finally, $B_2(\tau)$ is expressed in terms of $L(\tau)$ in (3.61). From Theorem 3.10 in appendix 3.9 we have that

$$\frac{d}{d\tau}M(\hat{a}_j;\hat{\zeta}_j;\hat{q}e^{-\kappa_r\tau}) = -\hat{q}\kappa_r e^{-\kappa_r\tau}\frac{\hat{a}_j}{\hat{\zeta}}M(\hat{a}_j+1;\hat{\zeta}_j+1;\hat{q}e^{-\kappa_r\tau}).$$

whence $B_2(\tau)$ is of the form

$$B_2(\tau) = -\frac{\kappa_r\hat{p}}{\xi^2\hat{\gamma}}e^{-\kappa_r\tau} + \frac{2\kappa_r}{\xi^2\hat{\gamma}} \cdot \frac{\sum_{j=1}^{2} K_j e^{-\kappa_r\hat{\beta}_j\tau}\left\{\hat{\beta}_j M(\hat{a}_j,\hat{\zeta}_j,\hat{q}e^{-\kappa_r\tau}) + \hat{q}e^{-\kappa_r\tau}\frac{\hat{a}_j}{\hat{\zeta}_j}M(\hat{a}_j+1,\hat{\zeta}_j+1,\hat{q}e^{-\kappa_r\tau})\right\}}{\sum_{j=1}^{2} K_j e^{-\hat{\beta}_j\kappa_r\tau} M(a_j,\hat{\zeta}_j,\hat{q}e^{-\kappa_r\tau})},$$

which is the desired result (3.31c). Clearly, the value of $B_2(\tau)$ is uniquely determined insofar as we know the ratio $\Xi = K_2/K_1$. But this just requires a single boundary condition ($B_2(0) = 0$) - specifically, it can be shown that

$$\Xi \equiv \frac{[\frac{\hat{p}}{2} - \beta_1] M(\hat{a}_1, \hat{\zeta}_1, \hat{q}) - \hat{q}\frac{\hat{a}_1}{\hat{\zeta}_1} M(\hat{a}_1 + 1, \hat{\zeta}_1 + 1, \hat{q})}{[\beta_2 - \frac{\hat{p}}{2}] M(\hat{a}_2, \hat{\zeta}_2, \hat{q}) + \hat{q}\frac{\hat{a}_2}{\hat{\zeta}_2} M(\hat{a}_2 + 1, \hat{\zeta}_2 + 1, \hat{q})}.$$

It remains to demonstrate the validity of (3.31a). From (3.30a) and the boundary condition $A(0) = 0$ we have that

$$A(\tau) = \kappa_r \theta_r \int_0^\tau B_1(s) ds + \kappa_v \theta_v \int_0^\tau B_2(s) ds.$$

Inserting equations (3.31b) and (3.60) into this expression we get

$$A(\tau) = \theta_r \int_0^\tau \left(1 - e^{-\kappa_r s}\right) ds - \frac{2\kappa_v \theta_v}{\hat{\gamma}\xi^2} \ln(L(\tau)).$$

Performing the integral, and using the fact that $L(\tau) = L(\tau)/L(0)$ (from (3.60), L must obviously satisfy $L(0) = 1$), we get the desired result. □

Appendix B: Differential Equations

Lemma 3.1. *Consider the Riccati equation*

$$\frac{dy}{dx}(x) = ay^2(x) + by(x) + c, \tag{3.63}$$

where a,b and c are constant coefficients. Assuming that $y(0) = 0$ the solution is of the form

$$y(x) = \begin{cases} \dfrac{2c(e^{\delta x}-1)}{(\delta - b)(e^{\delta x}-1)+2\delta}, & \text{if } b^2 > 4ac, \\ \dfrac{2cx}{2-bx}, & \text{if } b^2 = 4ac, \\ \dfrac{2c}{\varepsilon \cot\left(\frac{\varepsilon x}{2}\right) - b}, & \text{if } b^2 < 4ac, \end{cases} \tag{3.64}$$

where $\delta \equiv \sqrt{b^2 - 4ac}$ and $\varepsilon \equiv -i\delta$, where i is the complex unit.

Proof: We will consider the three cases individually:

1. Suppose $b^2 > 4ac$: from quadratic factorisation $ay^2 + by + c$ may be expressed as $a(y - r_1)(y - r_2)$, where $r_1 \equiv (-b+\delta)/(2a)$ and $r_2 \equiv (-b-\delta)/(2a)$ are real valued roots and δ is defined above. Hence, (3.63) may be written as

$$\frac{1}{a}\int_0^y \frac{dy}{(y-r_1)(y-r_2)} = x. \tag{3.65}$$

The integral may readily be solved using partial fractions; specifically:

$$\frac{1}{a(r_1-r_2)}\int_0^y \left(\frac{1}{y-r_1} - \frac{1}{y-r_2}\right) dy = \frac{1}{a(r_1-r_2)}\left(\ln\left|1-\frac{y}{r_1}\right| - \ln\left|1-\frac{y}{r_2}\right|\right)$$

Thus,

$$\frac{1}{a(r_1-r_2)} \ln\left|\frac{1-\frac{y}{r_1}}{1-\frac{y}{r_2}}\right| = x,$$

which can be rearranged to the equation

$$y(x) = \frac{r_1 r_2 (e^{a(r_1-r_2)x} - 1)}{r_1 e^{a(r_1-r_2)x} - r_2}.$$

Now it may readily be checked that $r_1 r_2 = \frac{c}{a}$ and $a(r_1 - r_2) = \delta$. Furthermore, $r_1 e^{a(r_1-r_2)x} - r_2 = \frac{1}{2a}[(-b+\delta)e^{\delta x} - (-b-\delta)] = \frac{1}{2a}[(-b+\delta)(e^{\delta x}-1)+2\delta]$. Thus, $y = 2c(e^{\delta x}-1)/((\delta-b)(e^{\delta x}-1)+2\delta)$ as desired.

2. When $b^2 = 4ac$ then $\delta = 0$ and we have a repeated root, $r = -b/(2a)$. The integral on the LHS of (3.65) therefore becomes

$$\frac{1}{a}\int_0^y \frac{dy}{(y-r)^2} = \frac{1}{a}\left(\frac{1}{r-y} - \frac{1}{r}\right) = \frac{y}{ar(r-y)}.$$

Equating this to x (the RHS of (3.65)) we obtain after some manipulation

$$y = \frac{ar^2 x}{1+arx}.$$

Inserting $r = -b/(2a)$ and using the fact that $b^2 = 4ac$ we obtain the desired result[20].

3. When $b^2 < 4ac$ then δ is imaginary and r_1 and r_2 are complex. The formula derived in 1. still applies - however, to highlight that y is real, it is convenient to reformulate the expression in terms of real parameters. Till this end define the parameter $\varepsilon \equiv -i\delta \in \mathbb{R}$ then

$$\begin{aligned} y(x) &= \frac{2c(e^{i\varepsilon x}-1)}{(i\varepsilon - b)(e^{i\varepsilon x}-1)+2i\varepsilon} = \frac{2c(e^{i\varepsilon x}-1)}{i\varepsilon(e^{i\varepsilon x}+1) - b(e^{i\varepsilon x}-1)}, \\ &= \frac{2ce^{i\frac{\varepsilon x}{2}}(e^{i\frac{\varepsilon x}{2}} - e^{-i\frac{\varepsilon x}{2}})}{i\varepsilon e^{i\frac{\varepsilon x}{2}}(e^{i\frac{\varepsilon x}{2}} + e^{-i\frac{\varepsilon x}{2}}) - be^{i\frac{\varepsilon x}{2}}(e^{i\frac{\varepsilon x}{2}} - e^{-i\frac{\varepsilon x}{2}})}, \\ &= \frac{4ic\sin\left(\frac{\varepsilon x}{2}\right)}{2i\varepsilon\cos\left(\frac{\varepsilon x}{2}\right) - 2bi\sin\left(\frac{\varepsilon x}{2}\right)} = \frac{2c}{\varepsilon\frac{\cos\left(\frac{\varepsilon x}{2}\right)}{\sin\left(\frac{\varepsilon x}{2}\right)} - b}, \end{aligned}$$

which is the desired equation. □

Lemma 3.2. *Let a, b and c be constants such that $c \neq b$, $a \neq 0$ and $b \neq 0$ then*

$$\int_0^x \frac{e^{ax}-1}{b(e^{ax}-1)+c}dx = \frac{x}{b-c} + \frac{c}{ab(b-c)}\ln\left|\frac{c}{b(e^{ax}-1)+c}\right|. \qquad (3.66)$$

Proof: Split the integral into two components $I_1 + I_2$:

$$\int_0^x \frac{e^{ax}-1}{b(e^{ax}-1)+c}dx = \int_0^x \frac{e^{ax}}{b(e^{ax}-1)+c}dx - \int_0^x \frac{dx}{b(e^{ax}-1)+c}$$

[20] A less cumbersome derivation involves expanding $e^{\delta x}$ as $1+\delta x+\mathscr{O}(\delta^2)$ in $y = 2c(e^{\delta x}-1)/((\delta - b)(e^{\delta x}-1)+2\delta)$ and letting $\delta \to 0$.

Upon handling the first integral, I_1, on the RHS, use the substitution $u = e^{ax}$ ($\frac{du}{dx} = ae^{ax}$) whence

$$\frac{1}{a}\int_1^u \frac{du}{b(u-1)+c} = -\frac{1}{ab}\ln\left|\frac{c}{b(e^{ax}-1)+c}\right|. \tag{3.67}$$

A similar substitution is used for the second integral, I_2, only now we will also need to invoke the method of partial fractions

$$\begin{aligned} I_2 = -\frac{1}{a}\int_1^u \frac{du}{u[b(u-1)+c]} &= \frac{1}{a(b-c)}\int_1^u \left(\frac{1}{u} - \frac{b}{b(u-1)+c}du\right) \\ = \frac{1}{b-c}x + \frac{1}{a(b-c)}\ln\left|\frac{c}{b(e^{ax}-1)+c}\right| \end{aligned} \tag{3.68}$$

Since $-\frac{1}{ab} + \frac{1}{a(b-c)} = \frac{c}{ab(b-c)}$ we clearly get the desired result (3.66) upon combining I_1 and I_2. $\square$

Lemma 3.3. *Consider the Riccati equation with a non-zero initial condition*

$$y'(x) = ay^2(x) + by(x) + c, \;\; y(0) = y_0 \tag{3.69}$$

where a,b and c are coefficients such that $b^2 > 4ac$. Then the solution is given by

$$y(x) = y_0 - \frac{(b+\delta+2ay_0)(1-e^{-\delta x})}{2a(1-\beta e^{-\delta x})} \tag{3.70}$$

where $\delta = \sqrt{b^2-4ac}$ and $\beta = \frac{b+\delta+2ay_0}{b-\delta+2ay_0}$.

Proof: Solve the quadratic equation $az_0^2 + bz_0 + c = 0$ to obtain $z_0 = \frac{-b\pm\delta}{2a}$. Select the solution with minus, let $z(x) = z_0 - y(x)$ and substitute into equation (3.69) to get

$$\begin{aligned} -z'(x) &= a(z_0 - z(x))^2 + b(z_0 - z(x)) + c \\ &= (az_0^2 + bz_0 + c) + az^2(x) - (2az_0 + b)z(x) \\ \Leftrightarrow z'(x) &= -az^2(x) - \delta z(x), \;\; z(0) = z_0 - y_0. \end{aligned} \tag{3.71}$$

Consider a solution $\tilde{z}(x) = \frac{\delta}{a}\frac{\beta e^{-\delta x}}{(1-\beta e^{-\delta x})}$, where $\beta = \frac{b+\delta+2ay_0}{b-\delta+2ay_0}$ and it may be readily checked that $\tilde{z}(x)$ satisfies the equation $\tilde{z}'(x) = -a\tilde{z}^2(x) - \delta\tilde{z}(x)$. Notice that $\tilde{z}(0) = \frac{\delta\beta}{a(1-\beta)} = \frac{-b-\delta}{2a} - y_0 = z_0 - y_0 = z(0)$ and we have found a solution to equation (3.71) which gives $y(0) = z_0 - z(0) = y_0$. Hence, equation (3.69) has the solution $y(x) = z_0 - z(x)$ that can be rearranged as

$$y(x) = y_0 - \frac{(b+\delta+2ay_0)(1-e^{-\delta x})}{2a(1-\beta e^{-\delta x})} \tag{3.72}$$

and notice that $y_0 = 0$ gives the reduced solution

$$y(x) = -\frac{(b+\delta)(1-e^{-\delta x})}{2a(1-\frac{b+\delta}{b-\delta}e^{-\delta x})} \tag{3.73}$$

of equation (3.69) with an initial condition equal to zero. Equation (3.72) may equally well be written as

$$y(x) = y_0 + \frac{2(ay_0^2 + by_0 + c)(e^{\delta x} - 1)}{(\delta - b - 2ay_0)(e^{\delta x} - 1) + 2\delta}.$$

$\square$

Appendix C: A Primer for the Mathematics of Confluent Hypergeometric Functions

In this section we briefly review key properties of *confluent hypergeometric functions*. For a more thorough introduction the reader is referred to MacDonald [21] or standard online resources such as https://en.wikipedia.org/wiki/Confluent_hypergeometric_function and http://dlmf.nist.gov/13.

Definition 3.1. The *confluent hypergeometric differential equation* is defined as

$$0 = z\frac{d^2y}{dz^2}(z) + (\zeta - z)\frac{dy}{dz}(z) - ay(z) \tag{3.74}$$

where ζ, a and z are unrestricted. It has a regular singularity at the origin and an irregular singularity at infinity.

Theorem 3.7. *There is an analytic solution to (3.74), known as the confluent hypergeometric function (or Kummer's function of the first kind), which is given by the series expansion*

$$M(a;\zeta;z) = \frac{\Gamma(\zeta)}{\Gamma(a)}\sum_{n=0}^{\infty}\frac{\Gamma(a+n)}{\Gamma(\zeta+n)}\frac{z^n}{n!} \tag{3.75}$$

where $\Gamma : \mathbb{R} \mapsto \mathbb{R}$ is the Gamma function: $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$. Alternatively, $M(a;\zeta;z)$ may be expressed it in terms of rising factorials:

$$M(a;\zeta;z) = \sum_{n=0}^{\infty}\frac{a^{(n)}z^n}{\zeta^{(n)}n!}, \tag{3.76}$$

where $x^{(0)} \equiv 1$ and

$$x^{(n)} \equiv x(x+1)(x+2)\cdots(x+n-1).$$

Notice that $M(a;\zeta;z)$ is not defined when $\zeta \in \mathbb{Z}^- \cup \{0\}$.

Theorem 3.8. *As (3.74) is second order, there exists another, independent solution. If ζ is not integer, it is given by the Tricomi function*

$$U(a;\zeta;z) = \frac{\Gamma(1-\zeta)}{\Gamma(a-\zeta+1)}M(a,\zeta,z) + \frac{\Gamma(\zeta-1)}{\Gamma(a)}z^{1-\zeta}M(a-\zeta+1,2-\zeta,z). \tag{3.77}$$

For the case ζ is integer, see [21].

Theorem 3.9. *Asymptotically, for large values $|z|$:*

$$M(a;\zeta;z) \sim \Gamma(\zeta)\left(\frac{e^z z^{a-\zeta}}{\Gamma(a)} + \frac{(-z)^{-a}}{\Gamma(\zeta-a)}\right) \tag{3.78}$$

for $\arg z \in (-\frac{3}{2}\pi, \frac{1}{2}\pi]$. *Use the first term only insofar $(\zeta - a) \in \mathbb{Z}^-$ or when $\Re[z] > 0$. Use the second term only insofar $a \in \mathbb{Z}^-$ or $\Re[z] < 0$.*

Theorem 3.10. *The derivative of $M(a;\zeta,z)$ is given by*

$$\frac{d}{dz}M(a;\zeta,z)=\frac{a}{\zeta}M(a+1;\zeta+1;z). \tag{3.79}$$

Theorem 3.11. *A generic second order ODE with coefficients linear in the independent variable,*

$$(A_0+A_1z)\frac{d^2y}{dz^2}(z)+(B_0+B_1z)\frac{dy}{dz}(z)+(C_0+C_1z)y(z)=0, \tag{3.80}$$

can be transformed into the confluent hypergeometric differential equation (3.74).

Proof: To see this, set $\hat{z}=\frac{A_0+A_1z}{\sqrt{B_1^2-4C_1}}$, where we assume $B_1^2\neq 4C_1$. After a few manipulations we obtain

$$\hat{z}\frac{d^2y}{d\hat{z}^2}(\hat{z})+\left(B_0+\frac{B_1}{\sqrt{B_1^2-4C_1}}\hat{z}\right)\frac{dy}{d\hat{z}}(\hat{z})+\left(\frac{C_0}{\sqrt{B_1^2-4C_1}}+\frac{C_1}{B_1^2-4C_1}\hat{z}\right)y(\hat{z})=0.$$

Next we set $y(\hat{z})=\exp\left\{-\left[1+\frac{B_1}{\sqrt{B_1^2-4C_1}}\right]\frac{\hat{z}}{2}\right\}w(\hat{z})$. After some tedious calculations we arrive at

$$\hat{z}\frac{d^2w}{d\hat{z}^2}(\hat{z})+(B_0-\hat{z})\frac{dw}{d\hat{z}}(\hat{z})-\left(\left[1+\frac{B_1}{\sqrt{B_1^2-4C_1}}\right]\frac{B_0}{2}-\frac{C_0}{\sqrt{B_1^2-4C_1}}\right)w(\hat{z})=0,$$

which indeed is the confluent hypergeometric differential equation. □

Corollary 3.1. *Assume $A_0=0$ and $A_1=1$, then the general solution to (3.80) may be expressed as*

$$y(z)=e^{-\frac{1}{2}pz}\left\{K_1M(a,\zeta,qz)+K_2z^{1-\zeta}M(a-\zeta+1,2-\zeta,qz)\right\},$$

where K_1,K_2 are arbitrary constants, M is Kummer's function, and we have defined the parameters

$$p\equiv\sqrt{B_1^2-4C_1}+B_1,\quad \textit{and}\quad q\equiv p-B_1,$$

and

$$\zeta\equiv B_0,\quad \textit{and}\quad a\equiv\left[1+\frac{B_1}{\sqrt{B_1^2-4C_1}}\right]\frac{B_0}{2}-\frac{C_0}{\sqrt{B_1^2-4C_1}}.$$

Proof: This follows immediately by combining Theorem 3.8 with Theorem 3.11.

□

Appendix D: Empirical Logbook

For the sake of replicability, further details regarding data acquisition alongside the calibration process are provided below.

- The raw data consists of intraday tick-by-tick trade data of the SPDR S&P 500 EFT (SPY) obtained from Wharton Research Data Services:

  https://wrds-web.wharton.upenn.edu/wrds.

  The simple answer to the question "why not use S&P 500 index data (GSPC) directly" is availability - there is no data provider that offers high frequency S&P 500 data for free (indeed, it comes to quite a high price e.g. from Thomas Reuters, we did investigate this). SPY, on the other hand, is available to us from Wharton, at no cost at all.
- Prior to the analysis, the raw data is pre-processed. The data processing consists of a few steps of data cleaning and aggregation. In particular, the irregular tick-by-tick time scale is aggregated to an equidistant 5 minute time grid. Hence, 5-minute high-frequency data as a result for the S&P 500 ETF.
- The variance process is measured on a daily (every second day) basis from the 5-minute data with the realised volatility measure. 192 price observations are used for each variance measurement which yields 1,492 variance observations. This motivates the requirement of high-frequency data - we may as well measure the variance from daily price data (with e.g. an EGARCH model or EWMA model) but this will arguably yield variance measurements of poorer quality.
- With the (daily) variance observations in hand (note: we use the nomenclature *observations* although the data is not directly observed; the latent variance process is measured), we proceed to estimate parameters. The parameters of the variance processes of Heston?s model and the $3/2$ model are estimated with maximum likelihood estimation: numerical optimisation is performed to optimise the likelihood function of each model. Since there is no convenient form of the likelihood function of the square root process (well, there is: the condition distribution is a non-central chi-squared distribution, but an exact likelihood from this distribution turns out to provide a poor estimator) we use an approximate (Gaussian) likelihood based on the moments of the square root process which yields a consistent estimator (this result is due to Sorensen).
- The dynamics of the variance process in the $3/2$ model, say with inherent parameters (κ, θ, ξ), is equivalent to the reciprocal square root process (apply Ito's formula to $1/Y$ where $Y \sim$ square root process) with parameters $(\tilde{\kappa}, \tilde{\theta}, \tilde{\xi})$ obtained by a simple transformation $(\kappa, \theta, \xi) \mapsto (\tilde{\kappa}, \tilde{\theta}, \tilde{\xi})$. This result is used to obtain a likelihood for this model - in effect, reciprocal variance observations are plugged into the approximated likelihood. This gives two options for the parameter estimation: optimise the likelihood w.r.t. reciprocal parameters $(\tilde{\kappa}, \tilde{\theta}, \tilde{\xi})$ directly (and take the inverse transformation to obtain the inherent parameters) or re-parametrise the likelihood to take the inherent parameters (κ, θ, ξ) and optimise w.r.t. these. The latter is the method we use.

Appendix E: Tables

Investment Period	S start	S end	V start	V end	Me end	He end	3/2 end
02-12-31 to 03-03-31	90.47	87.07	0.051	0.042	996.8997	996.8887	996.8557
03-03-31 to 03-06-30	86.24	97.85	0.095	0.022	1,029.943	1,029.976	1,030.091
03-06-30 to 03-09-30	99.12	100.86	0.046	0.020	1,007.643	1,007.647	1,007.662
03-09-30 to 03-12-31	102.02	111.12	0.0551	0.0041	1,021.428	1,021.445	1,021.5
03-12-31 to 04-03-31	111.18	111.61	0.015	0.010	1,004.727	1,004.728	1,004.727
04-03-31 to 04-06-30	112.63	114.77	0.014	0.010	1,007.758	1,007.759	1,007.76
04-06-30 to 04-09-30	113.52	111	0.0053	0.0119	999.6928	999.69	999.6771
04-09-30 to 04-12-31	112.64	121.2	0.0086	0.0046	1,018.942	1,018.948	1,018.99
04-12-31 to 05-03-31	121.34	117.83	0.0089	0.0184	998.1711	998.168	998.1515
05-03-31 to 05-06-30	117.1	119.76	0.0092	0.0048	1,008.439	1,008.442	1,008.455
05-06-30 to 05-09-30	119.91	121.49	0.014	0.016	1,006.657	1,006.659	1,006.666
05-09-30 to 05-12-31	122.76	124.72	0.0066	0.0074	1,007.415	1,007.415	1,007.428
05-12-31 to 06-03-31	127.07	130	0.0206	0.0067	1,008.397	1,008.398	1,008.406
06-03-31 to 06-06-30	130.24	124.16	0.010	0.013	994.4169	994.4156	994.3975
06-06-30 to 06-09-30	127.22	133.49	0.0091	0.0079	1,013.833	1,013.835	1,013.845
06-09-30 to 06-12-31	133.47	142.09	0.0038	0.0042	1,016.587	1,016.59	1,016.602
06-12-31 to 07-03-31	140.71	141.64	0.0075	0.0270	1,005.232	1,005.232	1,005.233
07-03-31 to 07-06-30	143.01	151.05	0.0068	0.0184	1,015.035	1,015.037	1,015.05
07-06-30 to 07-09-30	152.35	152.62	0.0033	0.0111	1,004.762	1,004.761	1,004.761
07-09-30 to 07-12-31	154.23	147.8	0.0064	0.0117	995.8456	995.8376	995.8307
07-12-31 to 08-03-31	144.4	133.66	0.023	0.046	989.1243	989.1056	989.0949
08-03-31 to 08-06-30	131.89	128.04	0.031	0.030	998.4123	998.4124	998.4007
08-06-30 to 08-09-30	126.06	120.57	0.072	0.181	995.9456	995.932	995.9269
08-09-30 to 08-12-31	113.8	86.56	0.241	0.009	956.227	956.0642	956.1155
08-12-31 to 09-03-31	88.33	78.59	0.055	0.098	983.0127	982.809	982.9544
09-03-31 to 09-06-30	83.05	91.625	0.207	0.029	1,024.678	1,024.843	1,024.734
09-06-30 to 09-09-30	91.57	106.17	0.040	0.042	1,034.855	1,035.033	1,034.944
09-09-30 to 09-12-31	102.53	112.37	0.0268	0.0077	1,022.879	1,022.979	1,022.939
09-12-31 to 10-03-31	113.22	117.43	0.0097	0.0182	1,011.294	1,011.307	1,011.312
10-03-31 to 10-06-30	117.79	107.49	0.0044	0.0418	986.2664	986.2024	986.1993
10-06-30 to 10-09-30	101.88	114.57	0.086	0.022	1,028.377	1,028.467	1,028.485
10-09-30 to 10-12-31	114.5	126.03	0.0220	0.0031	1,023.579	1,023.639	1,023.66
10-12-31 to 11-03-31	125.76	132.73	0.008	0.014	1015.03	1,015.063	1,015.065
11-03-31 to 11-06-30	133.3	129.12	0.0045	0.0096	997.6833	997.6591	997.6538
11-06-30 to 11-09-30	131.77	118.81	0.016	0.150	985.2978	985.1994	985.22
11-09-30 to 11-12-31	114.87	125.91	0.095	0.027	1,024.274	1,024.337	1,024.344
11-12-31 to 12-03-31	127.36	141.71	0.021	0.010	1,025.473	1,025.509	1,025.539
12-03-31 to 12-06-30	140.28	133.2	0.011	0.027	993.9371	993.9198	993.9015
12-06-30 to 12-09-30	135.78	143.74	0.033	0.012	1,015.84	1,015.86	1,015.879
12-09-30 to 12-12-31	144.25	141.04	0.020	0.019	999.5934	999.5864	999.5821
12-12-31 to 13-03-31	144.96	156.365	0.0808	0.0031	1,019.178	1,019.205	1,019.219
13-03-31 to 13-06-30	156.66	160.06	0.011	0.032	1,008.565	1,008.578	1,008.575
13-06-30 to 13-09-30	162.22	168.95	0.016	0.014	1,012.198	1,012.208	1,012.22
13-09-30 to 13-12-31	168.88	183.84	0.0154	0.0013	1,021.297	1,021.317	1,021.335

Table 3.1 This table exhibits the trading period [year-month-date], initial index value, terminal index value, initial variance, terminal variance, terminal portfolio value using the Merton weight, terminal portfolio value using the Heston weight, and terminal portfolio value using the 3/2 weight. All portfolios have an initial value of 1000 \$.

Investment Period	He κ	He θ	He ξ	ρ	3/2 κ	3/2 θ	3/2 ξ
02-12-31 to 03-03-31	55.00	0.07	3.23	-0.14	49.21	1.53	52.14
03-03-31 to 03-06-30	54.624	0.068	3.138	-0.137	50.22	1.52	53.22
03-06-30 to 03-09-30	54.158	0.064	3.043	-0.134	51.22	1.42	55.47
03-09-30 to 03-12-31	51.911	0.063	3.025	-0.127	50.45	1.35	55.81
03-12-31 to 04-03-31	38.256	0.052	2.391	-0.121	57.023	0.755	68.075
04-03-31 to 04-06-30	33.059	0.051	2.238	-0.086	61.532	0.679	74.717
04-06-30 to 04-09-30	30.412	0.048	2.062	-0.075	58.686	0.615	76.006
04-09-30 to 04-12-31	29.978	0.042	1.867	-0.055	59.612	0.587	82.302
04-12-31 to 05-03-31	27.156	0.039	1.796	-0.064	53.501	0.568	82.908
05-03-31 to 05-06-30	24.486	0.038	1.773	-0.060	52.830	0.536	85.286
05-06-30 to 05-09-30	24.733	0.034	1.619	-0.065	53.209	0.517	90.848
05-09-30 to 05-12-31	40.392	0.024	1.292	-0.013	63.850	0.768	105.176
05-12-31 to 06-03-31	56.301	0.019	1.114	-0.071	88.32	0.99	131.31
06-03-31 to 06-06-30	73.311	0.015	1.005	-0.066	112.14	1.23	155.70
06-06-30 to 06-09-30	84.324	0.014	0.981	-0.132	173.36	1.37	198.14
06-09-30 to 06-12-31	90.523	0.013	0.947	-0.184	208.32	1.37	224.50
06-12-31 to 07-03-31	92.842	0.012	0.908	-0.159	201.08	1.51	225.60
07-03-31 to 07-06-30	89.061	0.012	1.075	-0.171	152.83	1.36	200.96
07-06-30 to 07-09-30	94.688	0.012	1.071	-0.211	177.38	1.47	220.82
07-09-30 to 07-12-31	64.082	0.015	1.533	-0.262	103.84	1.08	166.48
07-12-31 to 08-03-31	58.765	0.017	1.744	-0.255	96.040	0.909	157.152
08-03-31 to 08-06-30	43.320	0.022	2.334	-0.221	85.26	0.59	143.20
08-06-30 to 08-09-30	42.340	0.023	2.357	-0.217	85.299	0.586	139.242
08-09-30 to 08-12-31	28.985	0.032	2.566	-0.257	80.671	0.449	130.321
08-12-31 to 09-03-31	12.682	0.067	3.197	-0.231	67.893	0.204	113.631
09-03-31 to 09-06-30	11.981	0.082	3.317	-0.220	70.29	0.19	105.56
09-06-30 to 09-09-30	12.090	0.084	3.354	-0.180	63.814	0.206	95.986
09-09-30 to 09-12-31	13.127	0.086	3.348	-0.172	62.474	0.253	86.802
09-12-31 to 10-03-31	16.517	0.088	3.399	-0.183	62.383	0.285	83.116
10-03-31 to 10-06-30	18.730	0.088	3.375	-0.170	58.799	0.355	74.777
10-06-30 to 10-09-30	22.710	0.094	3.947	-0.172	56.427	0.477	64.317
10-09-30 to 10-12-31	23.202	0.093	3.841	-0.150	53.987	0.471	62.541
10-12-31 to 11-03-31	19.413	0.091	3.797	-0.150	39.730	0.433	60.479
11-03-31 to 11-06-30	16.427	0.087	3.508	-0.148	46.653	0.366	66.990
11-06-30 to 11-09-30	15.924	0.085	3.516	-0.152	59.435	0.285	78.835
11-09-30 to 11-12-31	16.700	0.084	3.496	-0.129	62.905	0.269	80.101
11-12-31 to 12-03-31	28.859	0.054	2.963	-0.129	67.139	0.449	84.405
12-03-31 to 12-06-30	36.130	0.041	2.822	-0.135	70.180	0.564	92.809
12-06-30 to 12-09-30	37.650	0.039	2.795	-0.143	70.897	0.623	95.136
12-09-30 to 12-12-31	35.207	0.037	2.798	-0.170	68.193	0.602	97.968
12-12-31 to 13-03-31	34.653	0.036	2.737	-0.174	70.914	0.602	100.466
13-03-31 to 13-06-30	29.300	0.036	2.710	-0.172	70.788	0.555	106.072
13-06-30 to 13-09-30	34.29	0.03	1.74	-0.16	76.645	0.665	112.305
13-09-30 to 13-12-31	32.995	0.028	1.669	-0.174	83.718	0.622	122.909

Table 3.2 This table exhibits the trading period [year-month-date], the three Heston parameters κ, θ, ξ, the correlation ρ, and the three 3/2 parameters κ, θ, ξ. All parameter estimates are made from data stretching over a three years period, prior to the first day of the associated investment period.

Chapter 4
Stochastic Volatility for Utility Maximisers Part II
The Case of Derivatives

Simon Ellersgaard and Martin Jönsson

Abstract Using martingale methods we derive bequest optimising portfolio weights for a rational investor who trades in a bond-stock-derivative economy characterised by a generic stochastic volatility model. For illustrative purposes we then proceed to analyse the specific case of the Heston economy, which admits explicit expressions for plain vanilla Europeans options. By calibrating the model to market data, we find that the demand for derivatives is primarily driven by the myopic hedge component. Furthermore, upon deploying our optimal strategy on real market prices, we find only a very modest improvement in portfolio wealth over the corresponding strategy which only trades in bonds and stocks.

Key words: Merton's Portfolio Problem, Stochastic Volatility, HJB Equation

Both authors are with The Department of Mathematical Sciences
Universitetsparken 5, 2100 Copenhagen Ø, Denmark
e-mail: ellersgaard@math.ku.dk and e-mail: maj@math.ku.dk.
Ellersgaard is corresponding author.

4.1 Introduction

Regardless of whether we are dealing with the running variance associated with a financial time series, or the implied volatility surface extracted from traded option prices, one thing is abundantly clear: contrary to the assumption of the Black-Scholes-Merton model, volatility is far from constant. In fact, universally accepted stylised facts of the economy include the highly erratic nature of the variance process through time (see figure 4.2), or the skew/smile effect characteristic of the implied volatility surface - as reported in Cont and Tankov [5].[1] Naturally, a plethora of possible resolutions to these effects have been proposed on the modelling front, most prominently *local volatility models* in which σ is a deterministic function of the random stock price, and *diffusion-based stochastic volatility models* in which σ is modelled directly as a stochastic differential equation. Both approaches must be considered significant steps towards designing calibratable models to observed market phenomena although neither can be said to be void of imperfections [5]. However, the latter is arguably the more sophisticated of the two, being as it were more readily susceptible to theoretical augmentation. Derivatives pricing likewise becomes a matter of some interest: whilst local volatility models will have us believe that options are perfectly hedgeable using bonds and the underlying stocks (thereby making them formally redundant), this is not so for valuation under stochastic volatility models. Here, incompleteness [10] forces us to make further exogenous assumptions about the behaviour of the market in order to pin down our risk neutral measure, $\mathbb{Q}$. Specifically, to value one option, enough similar traded options must already exist on the market, in order for us to say anything concrete (in somewhat more abstract terms: a supply-and-demand induced market price of risk must prevail).

Surprisingly, while derivative pricing and calibration in connexion with stochastic volatility constitute major research areas in the quant-finance community, relatively few papers deal with the impact of stochastic volatility on portfolio optimisation. In fact, to the best of our knowledge, the first authors to deal explicitly with the issue are Jun Liu and Jun Pan, a little more than a decade ago. The more pedestrian of their analyses is found in [19], in which bequest optimisation in a Heston-driven[2] bond-stock economy is used to illustrate a grander theoretical point about solutions to HJB equations. Briefly, under the assumption that the market price of risk is proportional to volatility, $\lambda_1 = \bar{\lambda}_1 \sigma$, Liu shows that the optimal portfolio weight to be placed on the stock by a rational CRRA investor is

$$\pi_{S,t}^{Liu} = \frac{\bar{\lambda}_1}{\gamma} - \rho \sigma_v \frac{\gamma - 1}{\gamma} L(T-t), \tag{4.1}$$

where $\bar{\lambda}_1/\gamma$ is the Merton ratio [12], and the second term is a stoch-vol *hedge correction* [15] in which L is the deterministic function

[1] In the words of Cont and Tankov (ch. 1): “For equity and foreign exchange options, implied volatilities $\sigma_t(T,K)$ display a strong dependence with respect to the strike price: this dependence may be decreasing (“skew”) or U-shaped (“smile”) and has greatly increased since the 1987 crash”.

[2] See [17] and section 4.4 for an exposition of this model.

$$L(\tau) = \frac{\bar{\lambda}_1^2}{\gamma} \frac{(e^{\eta\tau} - 1)}{(\varepsilon + \eta)(e^{\eta\tau} - 1) + 2\eta},$$

and we have defined the parameters $\varepsilon \equiv \kappa + \frac{\gamma-1}{\gamma}\rho\sigma_v\lambda$ and

$$\eta \equiv \sqrt{\varepsilon^2 + \frac{\gamma-1}{\gamma}\sigma_v^2\,(\rho^2 + \gamma[1-\rho^2])\,\bar{\lambda}_1^2}.$$

Throughout this paper, we refer to this result as **Liu's strategy**. Little is said by Liu on the empirical implications, yet it is well-known that the correction hedge is negligible. For example, our own investigation [15] reveals that the hedge correction is multiple orders of magnitude smaller than the Merton weight for realistic parameter specifications, thus leading to non-measurable improvements in the investor's welfare.

A far richer theoretical account of the role of stochastic volatility in portfolio maximisation is provided in Liu and Pan [20], who extend the above framework to include jumps in the underlying price process, and complete the market by including tradeable derivative securities (specifically, a straddle, chosen for its sensitivity to volatility risk). By solving the relevant HJB equation optimal portfolio weights are provided (in terms of certain partial derivatives); moreover, through the employ of market calibrated parameters Liu and Pan estimate that the primary demand for derivatives is nested in the myopic component of the portfolio weight (rather than the volatility hedge correction). Based on the same parameters they also establish significant improvements in *certainty equivalent wealth* through the act of including derivatives in a utility maximised portfolio.

4.1.1 Overview

This paper might be read as a quasi-exposition of the work above (in the sense that we obtain similar formulae albeit using different techniques) with elements of novelty. Specifically, through martingale considerations we establish the optimal investment strategy for a fairly generic stochastic volatility model. These formulas are instantiations of more general state-variable expressions found in e.g. Munk [15], but we expose them as (i) there is some pedagogical value in seeing how they can be derived in a martingale framework, (ii) they readily can be adapted to more exotic volatility models. Upon specialising to the Heston model, we then proceed to find explicit expressions in the event the derivative is a plain vanilla call or put option. Vis-a-vis the straddle strategy mentioned above this is a minor variation, yet the value of our analysis lies in its great attention to detail, both from a conceptual and computational point of view. In either case, the change to the optimal investment plan is considerable with respect to Liu's strategy for calibrated market parameters. Furthermore, in corroboration of the findings by Liu and Pan we find through Monte Carlo simulation that the hedge component specific to stochastic volatility is irrelevant. Finally, in a novel portfolio rebalancing experiment using real market prices, we show that access to plain vanillas for utility optimisers create only a very modest (dubious) improvement in the Sharpe ratio, in discordance with the quasi-empirical musings of Liu and Pan.

4.2 Problem Set-up

4.2.1 Market Assumptions

Following the path betrodden by Black, Scholes, and Merton we start out by considering a financial landscape which is frictionless, arbitrage free, and allows for continuous trading. Three assets which jointly complete the market are assumed to prevail, viz. a risk-free money account (a *bond*), one fundamental risky security (a *stock*), and one derivative security with a European exercise feature at time $t = T'$. As it is commonplace, we define the dynamical equations of these securities by first introducing the stochastic basis $(\Omega, \mathscr{F}, \mathbb{F} = \{\mathscr{F}_t\}_{t\in[0,T']}, \mathbb{P})$, where Ω represents all possible states of the economy, $\mathbb{P}$ is the real-world probability measure, and $\mathscr{F}_t$ is the augmented natural filtration of two independent Wiener processes W_1 and W_2: i.e.

$$\mathscr{F}_t = \sigma(\mathscr{F}_t^{\boldsymbol{W}} \cup \mathscr{N}),$$

where $\mathscr{F}_t^{\boldsymbol{W}} = \sigma(\{W_{1s}, W_{2s}\}_{s\in[0,t]})$ and $\mathscr{N}$ represents the null sets of $\mathbb{P}$. With this in mind, we specify the price process dynamics of the money account $\{B_t\}_{t\in[0,T]}$ as the deterministic equation,

$$dB_t = rB_t dt, \tag{4.2}$$

where $B_0 = b_0 \in \mathbb{R}^+$ and r is the constant rate of interest. As for the fundamental risky security $\{S_t\}_{t\in[0,T]}$ we posit an SDE model of a rather generic stochastic volatility form, viz.

$$\begin{aligned} dS_t &= \mu_S(t, V_t) S_t dt + \sqrt{V_t} S_t dW_{1t}, \\ dV_t &= \alpha(t, V_t) dt + \beta(t, V_t)(\rho dW_{1t} + \sqrt{1-\rho^2} dW_{2t}), \end{aligned} \tag{4.3}$$

where $(S_0, V_0) = (s_0, v_0) \in \mathbb{R}^{2+}$, and $\{V_t\}_{t\in[0,T']}$ is the variance process which is we assume strictly positive. As for the dynamical constituents: μ_S, α, and β are taken to be real valued deterministic functions $[0, T'] \times \mathbb{R}^+ \mapsto \mathbb{R}$, whilst

$$\rho = \mathbb{C}\mathrm{orr}[dS_t, dV_t] \in (-1, 1),$$

is a Pearson correlation coefficient between the stock variance processes. Finally, as for the European derivative, we envision a one-time pay-off $\Phi(S_{T'})$ based on the magnitude of the contemporaneous stock value at time $t = T'$. Letting $\{D_t = D(t, S_t, V_t)\}_{t\in[0,T']}$ represent the general price process, it follows from Itô's lemma that

$$dD_t = \mu_D(t, S_t, V_t) D_t dt + \sigma_{1D}(t, S_t, V_t) D_t dW_{1t} + \sigma_{2D}(t, S_t, V_t) D_t dW_{2t}, \tag{4.4}$$

where $D_{T'} = \Phi_{T'}$, and $\mu_D, \sigma_{1D}, \sigma_{2D} : [0,T'] \times \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}$ are the functions

$$\begin{aligned} \mu_D(t,s,v) &\equiv D^{-1}\Big[\partial_t D + \mu_S(t,v)s\partial_s D + \alpha(t,v)\partial_v D + \tfrac{1}{2}vs^2\partial^2_{ss}D \\ &\qquad + \tfrac{1}{2}\beta^2(t,v)\partial^2_{vv}D + \rho\beta(t,v)\sqrt{v}s\partial^2_{sv}D\Big], \\ \sigma_{1D}(t,s,v) &\equiv D^{-1}\Big[\rho\beta(t,v)\partial_v D + \sqrt{v}s\partial_s D\Big], \\ \sigma_{2D}(t,s,v) &\equiv D^{-1}\Big[\sqrt{1-\rho^2}\beta(t,v)\partial_v D\Big], \end{aligned} \tag{4.5}$$

assuming, of course, that $D \in \mathscr{C}^{1,2,2}$.

Crucial to our derivations in the subsequent sections, we now enforce the following minimal structure upon the *aggregate* risk preferences of agents trading in our tripartite economy:

Assumption 1 The market prices of risk λ_1 and λ_2 associated with the aleatoric components W_1 and W_2 are functions of v **only**. In concrete terms this means that

$$\lambda_1(v) = \frac{\mu_S(t,v) - r}{\sqrt{v}}, \tag{4.6}$$

and

$$\lambda_2(v) = \frac{\mu_D(t,s,v) - r}{\sigma_{2D}(t,s,v)} - \frac{\sigma_{1D}(t,s,v)}{\sigma_{2D}(t,s,v)}\lambda_1(v). \tag{4.7}$$

We call this the **weak Heston assumption** for reasons which will become clearer below.[3]

Now, from (4.6) we may define the risk-neutral measure $\mathbb{Q}$ through the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\bigg|_{T'} \equiv \xi(T') \equiv \exp\bigg\{-\tfrac{1}{2}\int_0^{T'}\sum_{i=1}^{2}\lambda_i^2(V_t)dt - \int_0^{T'}\sum_{i=1}^{2}\lambda_i(V_t)dW_{it}\bigg\}. \tag{4.8}$$

Assuming the Novikov condition, $\mathbb{E}[\exp\{\frac{1}{2}\int_0^{T'}(\lambda_1^2(V_t) + \lambda_1^2(V_t))dt\}] < \infty$, ξ is a true martingale, $\mathbb{E}[\xi(T')] = 1$, whence $\mathbb{Q}$ is an equivalent local martingale measure (ELMM). All discounted asset prices under $\mathbb{Q}$ are therefore local martingales, which can be verified by combining Girsanov's transformation

$$dW_{it} = -\lambda_i(V_t)dt + dW_{it}^{\mathbb{Q}},$$

for $i = 1,2$ with the price dynamics (4.3) and (4.4). Finally, upon combining the market price of risk (4.7) with the Itô expressions (4.5) we readily find that the partial differential equation governing the price of the derivative is of the form

[3] We emphasise that this is not a vacuous statement: specifically, the weak Heston assumption is not a *gauge freedom*, as it invariably does say *something* about supply and demand in the market. On the other hand, it is not an approximation either: clearly, we have the mathematical freedom to suppose whatever we want here.

$$0 = \partial_t D + rs\partial_s D + \{\alpha(t,v) - \beta(t,v)[\rho\lambda_1(v) + \sqrt{1-\rho^2}\lambda_2(v)]\}\partial_v D \\ + \tfrac{1}{2}vs^2\partial^2_{ss}D + \tfrac{1}{2}\beta^2(t,v)\partial^2_{vv}D + \rho\beta(t,v)\sqrt{v}s\partial^2_{sv}D - rD, \tag{4.9}$$

subject to the terminal condition $D(T', s, v) = \Phi(s)$.

Remark 4.1. We assume the Novikov condition in order to establish the existence of the ELMM. Rather bizarrely, existence is something all too often glossed over in the stochastic volatility literature. Indeed, there are somewhat spectacular examples of stochastic volatility models where the no-arbitrage condition generally breaks down cf. e.g. the Stein & Stein model.[4] Furthermore, it is reasonable to show formally that discounted asset prices are *true martingales* as opposed to strictly local ones. Although failure of the true martingale property does not entail arbitrage, it can lead to peculiarities ("bubble problems") such as the breakdown of put-call parity. For examples and general theory pertaining to these fascinating issues we refer the reader to Wong & Hyde [34].

4.2.2 Investor Assumptions

We consider the case of an investor who trades in the three asset classes in a self-financing manner over the temporal horizon $[0,T] \subseteq [0,T']$, with the intention of maximising the expected discounted utility of her terminal wealth, $\mathscr{W}_T$. Specifically, we are interested in determining the optimal portfolio weights $\pi^*_{S,t}$ and $\pi^*_{D,t}$ which the investor should place on the stock and the derivative[5] such that

$$\{\pi^*_{S,t}, \pi^*_{D,t}\}_{t\in[0,T]} = \underset{\{\pi_{S,t},\pi_{D,t}\}\in\mathscr{L}^2[0,T]}{\operatorname{argmax}} \mathbb{E}[e^{-\delta T}u(\mathscr{W}_T)], \tag{4.10}$$

where $\delta \in \mathbb{R}_+$ is a subjective discounting factor, and $u : \mathbb{R}^+ \mapsto \mathbb{R}$ is a utility function which we assume isoelastic, i.e.

$$u(x) = \frac{x^{1-\gamma}}{1-\gamma},$$

where $\gamma \in \mathbb{R}_+ \backslash \{1\}$ codifies the investor's level of risk aversion. No restrictions on short-selling and leveraging are enforced upon the portfolio weights. However, to rule out arbitrage through doubling-strategies we assume that the weights belong to space of square-integrable processes, which we have denoted by $\mathscr{L}^2$.

Note that from the self-financing condition it follows that the optimal wealth process $\{\mathscr{W}^*_t\}_{t\in[0,T]}$ obeys the dynamics

---

[4] The problem with their is that 0 lies in the interior of V_t which (from λ_1) gives rise to the condition $\mu_S = r$.

[5] We assume that the remaining fraction of the wealth $1 - \pi_{S,t} - \pi_{D,t}$ is allocated to the risk free money account.

$$
\begin{aligned}
d\mathscr{W}_t^* = {} & [r + \pi_{S,t}^* \sqrt{V_t}\lambda_1(V_t) + \pi_{D,t}^*(\sigma_{1D}\lambda_1(V_t) + \sigma_{2D}\lambda_2(V_t))]\mathscr{W}_t^* dt \\
& + [\pi_{S,t}^* \sqrt{V_t} + \pi_{D,t}^* \sigma_{1D}]\mathscr{W}_t^* dW_{1t} + \pi_{D,t}^* \sigma_{2D}\mathscr{W}_t^* dW_{2t},
\end{aligned}
\tag{4.11}
$$

which can be used to set up a Hamilton-Jacobi-Bellman equation for the value function associated with (4.10). This, traditional approach is nonetheless not the route by which we shall be proceeding: rather, we opt for a martingale theoretic approach, which is well-equipped to handle bequest-optimisation problems in complete financial markets.

4.3 The Martingale Solution

4.3.1 The Optimal Wealth Process

It is well-known that the dynamic programming problem highlighted above may be reformulated as a static optimisation problem by solving for the optimal wealth process, whence the optimal portfolio weights can be deduced [10] [3]. Specifically, we are scrutinising the optimisation problem

$$
\mathscr{W}_T^* = \underset{\mathscr{W}_T \in \mathscr{K}_T}{\operatorname{argmax}} \, \mathbb{E}[e^{-\delta T} u(\mathscr{W}_T)],
\tag{4.12}
$$

over the class of adapted self-financing portfolios, $\mathscr{K}_T$, which is to say subject to the static *budget constraint*,

$$
w_0 = \mathbb{E}^{\mathbb{Q}}[e^{-rT}\mathscr{W}_T],
$$

where $\mathscr{W}_0 = w_0$.[6] In *Lagrangian* terms, we are accordingly dealing with

$$
\mathfrak{L} = \mathbb{E}[e^{-\delta T} u(\mathscr{W}_T) - \eta \xi_T e^{-rT} \mathscr{W}_T],
\tag{4.13}
$$

where η is the Lagrange multiplier, and ξ_T is the Radon-Nikodym derivative defined in (4.8), here introduced to write the entire expectation under the $\mathbb{P}$-measure. By differentiating partially with respect to $\mathscr{W}_T$ and equating to zero, we may extract the optimal terminal wealth

$$
\mathscr{W}_T^* = (u')^{-1}(\eta e^{(\delta - r)T}\xi_T),
$$

where $(u')^{-1}(\cdot) = (\cdot)^{-1/\gamma}$ is the inverse marginal utility, i.e.

$$
\mathscr{W}_T^* = \eta^{-1/\gamma} e^{-qT} \xi_T^{-1/\gamma},
$$

where $q \equiv (\delta - r)/\gamma$. To determine the multiplier we combine this expression with the $\mathbb{P}$-budget constraint

$$
w_0 = \mathbb{E}[e^{-rT}\xi_T \mathscr{W}_T^*],
$$

to get

[6] For a formal proof of this result, see Munk [15].

$$\eta^{-1/\gamma} = \frac{w_0}{\mathbb{E}[e^{-(r+q)T}\xi_T^{1-1/\gamma}]}.$$

Thus,

$$\mathscr{W}_T^* = \frac{w_0 e^{rT}\xi_T^{-1/\gamma}}{\mathbb{E}[\xi_T^{1-1/\gamma}]}. \tag{4.14}$$

Now, consider the denominator $\mathbb{E}[\xi_T^{1-1/\gamma}]$. From (4.8) it is not hard to see that this *almost* looks like a $\mathbb{P}$-expectation of a Radon-Nikodym ξ^0 defined as:

$$\left.\frac{d\mathbb{Q}_0}{d\mathbb{P}}\right|_T \equiv \xi^0(T) \equiv \exp\left\{-\tfrac{1}{2}(1-1/\gamma)^2\int_0^T\sum_{i=1}^2\lambda_i^2(V_t)dt-(1-1/\gamma)\int_0^T\sum_{i=1}^2\lambda_i(V_t)dW_{it}\right\}. \tag{4.15}$$

In fact, one may readily check that $\xi_T^{1-1/\gamma}$ and ξ_T^0 are related through

$$\xi_T^{1-1/\gamma} = \xi_T^0\exp\left\{\frac{1-\gamma}{2\gamma^2}\int_0^T\sum_{i=1}^2\lambda_i^2(V_t)dt\right\}, \tag{4.16}$$

whence

$$\mathbb{E}[\xi_T^{1-1/\gamma}] = \mathbb{E}^{\mathbb{Q}_0}\left[\exp\left\{\frac{1-\gamma}{2\gamma^2}\int_0^T\sum_{i=1}^2\lambda_i^2(V_t)dt\right\}\right]. \tag{4.17}$$

The explicit dependence on the Wiener increments has thus been suppressed through a second change of measure. This expectation is sufficiently important to what follows that we are prompted to introduce the function $H:[0,T]\times\mathbb{R}^+\mapsto\mathbb{R}$:

$$H_t = H(t,v) = \mathbb{E}_{t,v}^{\mathbb{Q}_0}\left[\exp\left\{\frac{1-\gamma}{2\gamma^2}\int_t^T\sum_{i=1}^2\lambda_i^2(V_s)ds\right\}\right]. \tag{4.18}$$

To see how this comes in handy, let us determine the optimal wealth process $\mathscr{W}_t^*$ for all times $t\in[0,T]$. From the budget constraint

$$\begin{aligned}
\mathscr{W}_t^* &= \mathbb{E}_{t,v}^{\mathbb{Q}}[e^{-r(T-t)}\mathscr{W}_T^*]\\
&= e^{-r(T-t)}\tfrac{1}{\xi_t}\mathbb{E}_{t,v}[\xi_T\mathscr{W}_T^*]\\
&= e^{-r(T-t)}\tfrac{1}{\xi_t}\tfrac{w_0e^{rT}}{H_0}\mathbb{E}_{t,v}[\xi_T^{1-1/\gamma}]\\
&= \tfrac{e^{rt}w_0}{\xi_tH_0}\mathbb{E}_{t,v}[\xi_T^0\exp\{\tfrac{1-\gamma}{2\gamma^2}\textstyle\int_0^T(\lambda_1^2(V_s)+\lambda_2^2(V_s))ds\}]\\
&= \tfrac{e^{rt}w_0}{\xi_tH_0}\xi_t^0\mathbb{E}_{t,v}^{\mathbb{Q}_0}[\exp\{\tfrac{1-\gamma}{2\gamma^2}\textstyle\int_0^T(\lambda_1^2(V_s)+\lambda_2^2(V_s))ds\}]\\
&= \tfrac{e^{rt}w_0}{\xi_tH_0}\xi_t^0\exp\{\tfrac{1-\gamma}{2\gamma^2}\textstyle\int_0^t(\lambda_1^2(V_s)+\lambda_2^2(V_s))ds\}\mathbb{E}_{t,v}^{\mathbb{Q}_0}[\exp\{\tfrac{1-\gamma}{2\gamma^2}\textstyle\int_t^T(\lambda_1^2(V_s)+\lambda_2^2(V_s))ds\}]
\end{aligned}$$

$$= \tfrac{e^{rt} w_0}{\xi_t H_0} \xi_t^{1-1/\gamma} \mathbb{E}^{\mathbb{Q}_0}_{t,v}[\exp\{\tfrac{1-\gamma}{2\gamma^2} \int_t^T (\lambda_1^2(V_s) + \lambda_2^2(V_s))ds\}],$$

where the second line uses the abstract Bayes' formula (Björk [10], proposition B.41), the third line uses the optimal wealth expression (4.14), the fourth line uses the identity (4.16), the fifth line uses (4.17), the sixth line splits the integral at the point of measurability $\mathscr{F}_t$, and the final line uses (4.16) again. Hence, from the definition of the H function (4.18) we find that the optimal wealth process can be written as

$$\mathscr{W}_t^* = e^{rt} w_0 \frac{H_t}{H_0} \xi_t^{-1/\gamma}. \tag{4.19}$$

4.3.2 Notes on the H-function

Since the H-function (4.18) is assumed to be a function of t and V_t it follows from Itô's lemma and the dynamics (4.3) that

$$dH_t = \mu_H(t,V_t)H_t dt + \sigma_{1H}(t,V_t)H_t dW_{1t} + \sigma_{2H}(t,V_t)H_t dW_{2t}, \tag{4.20}$$

where

$$\begin{aligned}
\mu_H(t,v) &\equiv H^{-1}\left[\partial_t H + \alpha(t,v)\partial_v H + \tfrac{1}{2}\beta^2(t,v)\partial^2_{vv}H\right],\\
\sigma_{1H}(t,v) &\equiv H^{-1}\rho\beta(t,v)\partial_v H,\\
\sigma_{2H}(t,v) &\equiv H^{-1}\sqrt{1-\rho^2}\beta(t,v)\partial_v H.
\end{aligned}$$

Now, from Girsanov's theorem it follows that the $\mathbb{Q}_0$-Brownian increments are related to the $\mathbb{P}$-Brownian increments through

$$dW_{it} = -(1-1/\gamma)\lambda_i(v)dt + dW_{it}^{\mathbb{Q}_0},$$

for $i = 1,2$. Substituting these into the dynamics for the variance V_t (4.3), we find that the drift changes as $\alpha(t,v) \to \alpha^{\mathbb{Q}_0}(t,v)$ where

$$\alpha^{\mathbb{Q}_0}(t,v) \equiv \alpha(t,v) - (1-1/\gamma)\beta(t,v)\left[\rho\lambda_1(v) + \sqrt{1-\rho^2}\lambda_2(v)\right].$$

Thus, from Feynman-Kac we may deduce that H solves the linear PDE

$$0 = \partial_t H + \alpha^{\mathbb{Q}_0}(t,v)\partial_v H + \tfrac{1}{2}\beta^2(t,v)\partial^2_{vv}H + \frac{1-\gamma}{2\gamma^2}\sum_{i=1}^{2}\lambda_i^2(V_t)H, \tag{4.21}$$

subject to the terminal condition $H(T,v) = 1$. The point here is well-worth appreciating: rather than enduring the non-linearity inherent to the HJB formalism, we have transformed the optimisation problem into something as comparatively pedestrian as having to solve

(4.21). Whether we aim for an analytic or a numerical solution, it is clear which approach imbues the greatest allure.

4.3.3 The Optimal Portfolio Weights

Finally, we are in a position to determine the optimal portfolio weights π^*_{St} and π^*_{Dt}. Applying Itô to (4.19) we find

$$\begin{aligned}
d\mathscr{W}_t^* &= re^{rt}w_0\tfrac{H_t}{H_0}\xi_t^{-1/\gamma}dt + e^{rt}w_0\tfrac{1}{H_0}\xi_t^{-1/\gamma}dH_t + e^{rt}w_0\tfrac{H_t}{H_0}d(\xi_t^{-1/\gamma}) + e^{rt}w_0\tfrac{1}{H_0}dH_t d(\xi_t^{-1/\gamma})\\
&= \text{drift} + e^{rt}w_0\tfrac{1}{H_0}\xi_t^{-1/\gamma}[\sigma_{1H}H_t dW_{1t} + \sigma_{2H}H_t dW_{2t}] - e^{rt}w_0\tfrac{H_t}{H_0}\tfrac{1}{\gamma}\xi_t^{-1/\gamma-1}d\xi_t\\
&= \text{drift} + \mathscr{W}_t^*[\sigma_{1H}dW_{1t} + \sigma_{2H}dW_{2t}] + e^{rt}w_0\tfrac{H_t}{H_0}\tfrac{1}{\gamma}\xi_t^{-1/\gamma-1}\xi_t[\lambda_1(v)dW_{1t} + \lambda_2(v)dW_{2t}]\\
&= \text{drift} + \mathscr{W}_t^*[\sigma_{1H}dW_{1t} + \sigma_{2H}dW_{2t}] + \mathscr{W}_t^*\tfrac{1}{\gamma}[\lambda_1(v)dW_{1t} + \lambda_2(v)dW_{2t}]\\
&= \text{drift} + \left[\sigma_{1H} + \tfrac{1}{\gamma}\lambda_1(v)\right]\mathscr{W}_t^* dW_{1t} + \left[\sigma_{2H} + \tfrac{1}{\gamma}\lambda_2(v)\right]\mathscr{W}_t^* dW_{2t},
\end{aligned} \tag{4.22}$$

where the first line uses the product rule, the second line makes use of (4.20) and the chain rule, the third line makes use of (4.19) and the definition of the Radon-Nikodym derivative (4.8), and the fourth line makes use of (4.19) again. Comparing (4.22) with our expression for the self-financing condition (4.11) we see that we have established two simultaneous equations from which π^*_{St} and π^*_{Dt} can be determined

$$\pi^*_{S,t}\sqrt{V_t} + \pi^*_{D,t}\sigma_{1D} = \sigma_{1H} + \tfrac{1}{\gamma}\lambda_1(v), \qquad \text{and} \qquad \pi^*_{D,t}\sigma_{2D} = \sigma_{2H} + \tfrac{1}{\gamma}\lambda_2(v).$$

Solving these we ultimately arrive at

$$\pi^*_{St} = \frac{1}{\sqrt{V_t}}\left\{\frac{\lambda_1(v)}{\gamma} - \frac{\sigma_{1D}}{\sigma_{2D}}\frac{\lambda_2(v)}{\gamma} + \sigma_{1H} - \frac{\sigma_{1D}\sigma_{2H}}{\sigma_{2D}}\right\}, \tag{4.23a}$$

$$\pi^*_{Dt} = \frac{1}{\sigma_{2D}}\left\{\frac{\lambda_2(v)}{\gamma} + \sigma_{2H}\right\}. \tag{4.23b}$$

Theorem 4.1. *It is well-worth summarising our findings in this section. Consider the control problem stated in (4.12). Defining the function*

$$H_t = H(t,v) = \mathbb{E}^{\mathbb{Q}_0}_{t,v}\left[\exp\left\{\frac{1-\gamma}{2\gamma^2}\int_t^T(\lambda_1^2(V_s) + \lambda_2^2(V_s))ds\right\}\right], \tag{4.24}$$

where $\mathbb{Q}_0$ is the measure defined through $\xi^0 \equiv d\mathbb{Q}_0/d\mathbb{P}$ where

$$d\xi_t^0 = -(1-1/\gamma)\xi_t^0(\lambda_1(v)dW_{1t} + \lambda_2(v)dW_{2t}), \tag{4.25}$$

the optimal wealth process can be written on the form

$$\mathscr{W}_t^* = e^{rt} w_0 \frac{H_t}{H_0} \xi_t^{-1/\gamma}. \tag{4.26}$$

Furthermore, the optimal controls which give rise to this maximal wealth process are of the form

$$\pi_{St}^* = \frac{1}{\sqrt{V_t}} \left\{ \frac{\lambda_1(v)}{\gamma} - \frac{\sigma_{1D}}{\sigma_{2D}} \frac{\lambda_2(v)}{\gamma} + \sigma_{1H} - \frac{\sigma_{1D}\sigma_{2H}}{\sigma_{2D}} \right\}, \tag{4.27a}$$

$$\pi_{Dt}^* = \frac{1}{\sigma_{2D}} \left\{ \frac{\lambda_2(v)}{\gamma} + \sigma_{2H} \right\}, \tag{4.27b}$$

where

$$\sigma_{1D} \equiv D^{-1}[\rho\beta\partial_v D + \sqrt{v}s\partial_s D], \qquad \sigma_{2D} \equiv D^{-1}[\sqrt{1-\rho^2}\beta\partial_v D], \tag{4.28}$$

$$\sigma_{1H} \equiv H^{-1}\rho\beta\partial_v H, \qquad \sigma_{2H} \equiv H^{-1}\sqrt{1-\rho^2}\beta\partial_v H. \tag{4.29}$$

4.4 Example: The Heston Model

Undoubtably, the most well-known of all stochastic volatility models is that proposed by Heston, [17]. The Heston model stands out for a number of reasons: first, the variance process is non-negative and mean-reverting, which harmonises with market data; secondly, the model is sufficiently parsimonious to allow for swift calibrations [27] [33]; thirdly, as exposed below, it famously admits comparatively simple[7] expressions for plain vanilla options; finally, said expressions yield implied volatilities which are found to fit the empirically observed volatility smile closely for a broad range[8] of medium-seized times to maturity [33].

Formally, the Heston model is a Cox-Ingersoll-Ross model for the variance process:

$$dV_t = \kappa(\theta - V_t)dt + \sigma_v\sqrt{V_t}(\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}), \tag{4.30}$$

where κ, θ, and σ_v are three parameters in $\mathbb{R}^+$ which signify the speed of mean reversion, the long term variance, and the volatility of variance respectively. Insofar as the Feller

[7] Contrary to what is sometimes claimed in the literature, these expressions can neither be considered *closed-form*, nor the more extensive *analytic* since they involve an integral.

[8] Matching the smile for very short or very long times to maturity proves more difficult. In particular, with regards to the former, the so-called volatility of variance, σ_v, tends to explode, which indicates that there's a jump effect neglected by the dynamics.

condition is satisfied, it can be shown that the variance process stays strictly positive at all times.[9] Moreover, the distribution of V_t under (4.30) is non-central χ^2, which in the asymptotic limit $t \to \infty$ tends towards a gamma distribution. This effectively disposes of one of the key shortfalls of classical GBM valuation as the resulting density function of log returns will be fatter (exponential) than the bell curve.

4.4.1 Vanilla Valuation

What really propelled the Heston model into the academic limelight is largely its ability to price European calls (and *ipso facto* European puts). For the reader's convenience we here briefly review the valuation formula, and tie it to the theory of pricing in incomplete markets alluded to in subsection 4.2.1. Specifically, the relevant pricing PDE (4.9) is of the form

$$\begin{aligned} 0 = \partial_t D + rs\partial_s D + \{\kappa(\theta - v) - \sigma_v\sqrt{v}[\rho\lambda_1 + \sqrt{1-\rho^2}\lambda_2]\}\partial_v D \\ + \tfrac{1}{2}vs^2\partial^2_{ss}D + \tfrac{1}{2}\sigma_v^2 v\partial^2_{vv}D + \rho\sigma_v vs\partial^2_{sv}D - rD, \end{aligned} \tag{4.31}$$

subject to $D(T',s) = [\phi(s-K)]^+$, where ϕ is a binary variable which takes on the value +1 if the option is a call, and -1 if the option is a put. Upon solving this equation, Heston crucially makes the assumption that the *market price of volatility risk*, λ_v, here defined as[10]

$$\lambda_v \equiv \sigma_v[\rho\lambda_1 + \sqrt{1-\rho^2}\lambda_2], \tag{4.32}$$

is proportional to $\sqrt{v}$, i.e.

$$\exists\bar{\lambda}_v \in \mathbb{R} \text{ s.t. } \lambda_v(v) = \bar{\lambda}_v\sqrt{v}. \tag{4.33}$$

We call this the **Heston assumption** and note that it constitutes a concrete instantiation of the weak Heston assumption explicated above. Nonetheless, based on our desire to solve the PDE for the H-function, it will in fact be convenient to assume something slightly stronger, viz.

Assumption 2 There exist constants $\bar{\lambda}_1$ and $\bar{\lambda}_2$ such that $\lambda_1(v) = \bar{\lambda}_1\sqrt{v}$ and $\lambda_2(v) = \bar{\lambda}_2\sqrt{v}$. We call this the **strong Heston assumption**.

---

[9] It is dubious that calibrated parameters actually satisfy this condition [27].

[10] The market price of volatilty risk (4.32) is concept which arises naturally insofar as the dynamical equations (4.3) have *not* had their random components decorellated through a Cholesky decomposition. Specifically, for the market price of risk vector

$$\lambda = \sigma^{-1}(\text{excess return vector}),$$

we would set $\sigma = [\sqrt{V_t}, 0; D^{-1}s\sqrt{v}\partial_s D + D^{-1}\rho\sigma_v\sqrt{v}\partial_v D, D^{-1}\sqrt{1-\rho^2}\sigma_v\sqrt{v}\partial_v D]$, whilst Heston sets $\sigma = \sigma' := [\sqrt{V_t}, 0; D^{-1}\sqrt{v}\partial_s D, D^{-1}\sigma_v\sqrt{v}\partial_v D]$ (the latter is related to the former through the multiplication of the lower triangular matrix $\mathbf{L} = [1,0;\rho,\sqrt{1-\rho^2}]$: $\sigma = \sigma'\mathbf{L}$). For convenience, Heston also absorbs the constant σ_v in his definition.

Remark 4.2. A partial motivation for (4.33) is provided through Breeden's consumption based model, $\lambda_v(V_t)dt = \gamma\mathbb{C}\mathrm{ov}[dV_t, dc_t/c_t]$, when the consumption process is chosen as in the (general equilibrium) Cox, Ingersoll, and Ross (1985) model [17]. Less generously, we might view it as a postulate detached from empirical evidence, which purposefully has been engineered in order to allow (4.31) to be solved. Be that as it may, under the proportionality assumption it can be shown that the ELMM, $\mathbb{Q}$, exists and that discounted asset prices are true martingales insofar as certain inequalities on the parameters are satisfied, [34]. This may be taken as a formal justification for Heston's well-known valuation formula:

Theorem 4.2. ***(Heston's Valuation Formula for European Vanillas)*** *The no-arbitrage price of a European vanilla option is given by*

$$\begin{aligned} D(t,s,v) &= \mathrm{HestonVanilla}(\kappa,\theta,\sigma_v,\rho,\bar{\lambda}_1,\bar{\lambda}_2,r,v,s,K,\tau',\phi) \\ &= \phi\{sQ_1(\phi) - e^{-r\tau'}KQ_2(\phi)\}, \end{aligned} \tag{4.34}$$

where $\phi = +1$ if D is a call, and $\phi = -1$ if D is a put, $\tau' \equiv T' - t$,

$$Q_j(\phi) \equiv \frac{1-\phi}{2} + \phi P_j(\ln s, v, \tau', \ln K), \tag{4.35}$$

for $j = 1, 2$, and we have defined

$$P_j(\ln s, v, \tau', \ln K) \equiv \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left\{\frac{e^{-i\varphi\ln K} f_j(\ln s, v, \tau', \varphi)}{i\varphi}\right\} d\varphi \tag{4.36a}$$

$$f_j(\ln s, v, \tau', \varphi) \equiv \exp\{C_j(\tau',\varphi) + D_j(\tau',\varphi)v + i\varphi\ln s\}, \tag{4.36b}$$

$$D_j(\tau',\varphi) \equiv \frac{b_j - \rho\sigma_v\varphi i + d_j}{\sigma_v^2}\left(\frac{1-e^{d_j\tau'}}{1-g_je^{d_j\tau'}}\right), \tag{4.36c}$$

$$C_j(\tau',\varphi) \equiv r\varphi i\tau' + \frac{a}{\sigma_v^2}\left\{(b_j - \rho\sigma_v\varphi i + d_j)\tau' - 2\ln\left(\frac{1-g_je^{d_j\tau'}}{1-g_j}\right)\right\}, \tag{4.36d}$$

where $d_j \equiv ((\rho\sigma_v\varphi i - b_j)^2 - \sigma_v^2(2u_j\varphi i - \varphi^2))^{1/2}$, $g_j \equiv (b_j - \rho\sigma_v\varphi i + d_j)/(b_j - \rho\sigma_v\varphi i - d_j)$, $a \equiv \kappa\theta$, $u_1 \equiv \frac{1}{2}$, $u_2 \equiv -\frac{1}{2}$, $b_1 \equiv \kappa + \bar{\lambda}_v - \rho\sigma_v$, and $b_2 \equiv \kappa + \bar{\lambda}_v$.

Remark 4.3. From the generalised option pricing formula (Björk [10], proposition 26.11)

$$D_t = \phi\{s\mathbb{S}(\phi S_{T'} \geq \phi K|\mathscr{F}_t) - e^{-r\tau'}K\mathbb{Q}(\phi S_{T'} \geq \phi K|\mathscr{F}_t)\}, \tag{4.37}$$

where $\mathbb{S}$ is the stock measure defined through the Radon-Nikodym derivative

$$\xi_u^s \equiv \frac{d\mathbb{S}}{d\mathbb{Q}} = e^{ru}\frac{S_u}{S_0},$$

it follows that Q_1 and Q_2 in (4.34) are risk-adjusted probabilities that the option expires in the money. Specifically,

$$Q_1 = \mathbb{S}_t(\phi S_{T'} \geq \phi K | S_t = s; V_t = v), \tag{4.38a}$$

$$Q_2 = \mathbb{Q}_t(\phi S_{T'} \geq \phi K | S_t = s; V_t = v). \tag{4.38b}$$

.

Remark 4.4. The formula stated in (4.34) is rather unconventionally expressed in terms of the market price of risk constants $\bar{\lambda}_i$, $i = 1, 2$ along with the $\mathbb{P}$-parameters of the variance process. More commonly, the valuation formula is specified directly in terms of the risk-neutral $\mathbb{Q}$-parameters: $\kappa^{\mathbb{Q}} \equiv \kappa + \sigma_v(\rho\bar{\lambda}_1 + \sqrt{1-\rho^2}\bar{\lambda}_2)$ and $\theta^{\mathbb{Q}} \equiv \theta\kappa/\kappa^{\mathbb{Q}}$ (diffusion parameters unchanged).

4.4.2 The Optimal Heston Controls

We assume the investor trades in a risk free money account, a stock and a European vanilla derivative in a market characterised by Hestonian stochastic volatility. From the generic optimal control functions (4.27) it follows that we must determine $\sigma_{1D}, \sigma_{2D}, \sigma_{1H}$, and σ_{2H} and thence the quantities $\partial_s D, \partial_v D, H$, and $\partial_v H$. We do this over the two subsequent lemmas.

Lemma 4.1. *The option* ***delta*** *is given by*

$$\Delta_t \equiv \partial_s D = \phi Q_1(\phi), \tag{4.39}$$

while the option ***vega*** *is given by*

$$\nu_t \equiv \partial_v D = s\nu_1(\ln s, v, \tau', \ln K) - Ke^{-r\tau'}\nu_2(\ln s, v, \tau', \ln K), \tag{4.40}$$

where

$$\nu_j(\ln s, v, \tau', \ln K) \equiv \frac{1}{\pi}\int_0^\infty \Re\left\{\frac{D_j(\tau', \varphi)e^{-i\varphi \ln K} f_j(\ln s, v, \tau', \varphi)}{i\varphi}\right\} d\varphi. \tag{4.41}$$

Proof. A hands-on differentiation of (4.34) with respect to the underlying, s, and subsequent algebraic simplification is a tedious exercise better avoided. A somewhat subtler but arguably simpler argument may be presented by noting that the valuation formula is first order homogenous in the variable pair (s, K), i.e.

$$\text{HestonVanilla}(\kappa,\theta,\sigma_v,\rho,\bar{\lambda}_1,\bar{\lambda}_2,r,v,as,aK,\tau',\phi) = \\ a\cdot\text{HestonVanilla}(\kappa,\theta,\sigma_v,\rho,\bar{\lambda}_1,\bar{\lambda}_2,r,v,s,K,\tau',\phi),$$

for any $a \in \mathbb{R}$, whence *Euler's Homogenous Function Theorem*[11] entails

$$D = s\partial_s D + K\partial_K D. \tag{4.42}$$

Comparing (4.34) with (4.42) it is tempting to deduce that $\partial_s D = \phi Q_1(\phi)$, yet some care must be taken here. Specifically, it is not immediately obvious that (4.34) is the so-called *natural form* of D [26]: with two terms present in the equation we can add any arbitrary component to one term, as long as we cancel it through a corresponding subtraction to the other term.[12] To establish that (4.34) is the natural one, we employ a well-known result from Breeden and Litzenberger [6] viz.

$$\begin{aligned}\partial_K D &= e^{-r\tau'}\partial_K \mathbb{E}^{\mathbb{Q}}_{t,s,v}[[\phi(S_{T'}-K)]^+] \\ &= e^{-r\tau'}\phi\partial_K \mathbb{E}^{\mathbb{Q}}_{t,s,v}[(S_{T'}-K)\mathbf{1}\{\phi S_{T'} \geq \phi K\}] \\ &= -e^{-r\tau'}\phi \mathbb{E}^{\mathbb{Q}}_{t,s,v}[\mathbf{1}\{\phi S_{T'} \geq \phi K\}] \\ &= -e^{-r\tau'}\phi \mathbb{Q}_{t,s,v}(\phi S_{T'} \geq \phi K).\end{aligned}$$

Comparing this with (4.37) the result follows.

Equation (4.40) follows immediately from differentiating (4.34) with respect to v. Note that the result is independent of ϕ.

Lemma 4.2. *Under the strong Heston assumption the H function has the exponential affine form*

$$H(t,v) = \exp\{A(\tau)+B(\tau)v\}, \tag{4.43}$$

where $\tau \equiv T-t$. *Here* $B : [0,T] \mapsto \mathbb{R}$ *is the function*

$$B(\tau) = \frac{1-\gamma}{\gamma^2}(\bar{\lambda}_1^2+\bar{\lambda}_2^2)\cdot\frac{e^{\omega\tau}-1}{(\omega+\alpha)(e^{\omega\tau}-1)+2\omega}, \tag{4.44}$$

where $\alpha \equiv \kappa + (1-1/\gamma)\bar{\lambda}_v$ *and*

---

[11] Recall that the function $g : \mathbb{R}^2 \mapsto \mathbb{R}$ is said to be homogenous of degree n if

$$g(ax_1, ax_2) = a^n g(x_1,x_2).$$

Let $x_1' = nx_1$ and $x_2' = nx_2$ then we find upon differentiating g with respect to a that $na^{n-1}g = \partial_{x_1'} g\partial_a x_1' + \partial_{x_2'} g\partial_a x_2' = x_1\partial_{ax_1} g + x_2\partial_{ax_2} g$. In particular, upon setting $a = 1$ we get Euler's result for homogenous functions:

$$ng = x_1\partial_{x_1} g + x_2\partial_{x_2} g.$$

[12] Let $g = x_1\partial_{x_1} g + x_2\partial_{x_2} g$ and $g = x_1 h_1(x_1,x_2) + x_2 h_2(x_1,x_2)$ then a necessary and sufficient condition for $\partial_{x_1} g = h_1(x_1,x_2)$ and $\partial_{x_2} g = h_2(x_1,x_2)$ is that $x_1^2\partial_{x_1} h_1 = x_2^2\partial_{x_2} h_2$ - see [26].

$$\omega \equiv \sqrt{\alpha^2 + \sigma_v^2 \frac{\gamma - 1}{\gamma^2}(\bar{\lambda}_1^2 + \bar{\lambda}_2^2)},$$

while $A : [0,T] \mapsto \mathbb{R}$ *is the function*

$$A(\tau) = \frac{\kappa\theta}{\alpha^2 - \omega^2}\left\{(\alpha+\omega)\tau + 2\ln\left|\frac{2\omega}{(\alpha+\omega)(e^{\omega\tau}-1)+2\omega}\right|\right\}. \qquad (4.45)$$

Proof. Substituting in the relevant parametric specifications (4.30), (4.33), into the governing PDE (4.21) we find

$$0 = -\partial_\tau H + [\kappa\theta - \{\kappa + (1-1/\gamma)\bar{\lambda}_v\}v]\partial_v H + \tfrac{1}{2}\sigma_v^2 v \partial_{vv}^2 H + \frac{1-\gamma}{2\gamma^2}(\bar{\lambda}_1^2 + \bar{\lambda}_2^2)vH, \quad (4.46)$$

subject to the initial condition $H(0,v) = 1$, where we have invoked the time transformation $t \mapsto \tau$. Since the coefficients are linear functions of v we form the ansatz that the solution is of an exponential affine form. Thus, upon substituting (4.43) into (4.46) and using the fact that the expression should hold for *any* value of v we find the coupled ODEs:

$$B'(\tau) = \tfrac{1}{2}\sigma_v^2 B^2(\tau) - \{\kappa + (1-1/\gamma)\bar{\lambda}_v\}B(\tau) + \frac{1-\gamma}{2\gamma^2}(\bar{\lambda}_1^2 + \bar{\lambda}_2^2), \qquad (4.47a)$$

$$A'(\tau) = \kappa\theta B(\tau), \qquad (4.47b)$$

subject to the boundary conditions $A(0) = B(0) = 0$, where $'$ denotes the derivative with respect to τ. The first equation is Riccatian, which readily allows us to extract the solution (4.44).[13] Note that ω is a real number insofar as $\gamma > 1$ which we henceforth assume to be the case. As for the function A we observe that (4.47b) can we written as $A(\tau) = \kappa\theta\int_0^\tau B(t)dt$. Performing this tedious integration we get the desired result.

Putting these results together we can finally state our theorem on the optimal (B,S,D)-portfolio weights in a Heston driven economy:

Theorem 4.3. *The optimal stock weight is given by*

$$\pi_{S,t}^* = \frac{\bar{\lambda}_1}{\gamma} - \frac{1}{\sqrt{1-\rho^2}}\left[\rho + \frac{s}{\sigma_v}\frac{\Delta_t}{v_t}\right]\frac{\bar{\lambda}_2}{\gamma} - s\frac{\Delta_t}{v_t}B(\tau), \qquad (4.48)$$

while the optimal vanilla option weight is

[13] Recall that the generic Ricatti equation $y'(x) = ay^2(x) + by(x) + c$ with $y(0) = 0$ has the solution $y(x) = [2c(e^{\delta x} - 1)]/[(\delta - b)(e^{\delta x} - 1) + 2\delta]$ where $\delta \equiv \sqrt{b^2 - 4ac}$ assuming $b^2 > 4ac$.

$$\pi^*_{D,t} = \frac{D_t \bar{\lambda}_2}{\gamma \sigma_v \sqrt{1-\rho^2} \nu_t} + \frac{D_t}{\nu_t} B(\tau), \tag{4.49}$$

where B is defined in (4.44), D_t is the option price given by (4.34), Δ_t is the option delta given in (4.39), and ν_t is the option vega given in (4.40). Note that the time parameter in B is τ (the investment horizon), while it for option quantities $\{D, \Delta, \nu\}$ is τ' (the maturity of the option).

We note that the first term in (4.48) is Merton's optimal stock weight in a simple (B,S)-economy with constant volatility. More generally, referencing standard results in the literature[14], we see that the first two terms in (4.48) and the first term in (4.49) constitute the optimal portfolio weights in a (B,S,D)-economy for a utility maximising investor who disregards stochastic fluctuations in the state variable v (otherwise known as the *myopic* or *1-period* strategy). Thus, the hedge against stochastic volatility is nested in the time-price-volatility dependent term $-s\frac{\Delta_t}{\nu_t}B(\tau)$ in (4.48) and $\frac{D_t}{\nu_t}B(\tau)$ in (4.49). In this connection we note that Δ is a function bounded by the interval $[0,1]$ for a call option ($[-1,0]$ for a put option), whilst $B(\tau)$ is a monotonically decreasing function bounded by the interval $((1-\gamma)(\bar{\lambda}_1^2+\bar{\lambda}_2^2)/(\gamma^2[\omega+\alpha]),0]$. s, D, and ν are all positive quantities unbounded from above. The signs of the volatility hedge corrections on the stock and the derivative are thus respectively positive and negative if D is a call option, and negative and negative if D is a put option. To appreciate the implications of this figure 4.1 plots the optimal (bank,stock,ATM call option)-weights for different times to maturity, when the tuple (s,v) is held constant at $(100,\theta)$. We assume that the risk free rate is 0.02, that the option expires at the end of the investment horizon ($\tau = \tau'$), and that the investor's level of risk aversion γ is 2. Other parameters are estimated from the S&P 500 index and are exhibited in table 4.1. Note here in particular that the market price of risk $\bar{\lambda}_2$ associated with W_2 is negative, corroborating standard empirical findings a la Bakshi and Kapadia [2].

Upon examining figure 4.1 we make the following observations: relative to the optimal Merton weight [**dash-dotted grey line**], Liu's volatility correction [**full grey line**] is barely noticeable, perturbing π^*_S at the order of magnitude 10^{-3}. By comparison, access to derivative trading prompts the investor to drastically increase her holding in the stock [**full red line**], by decreasing her long position in the money account [**full black line**], and shorting the call option [**full blue line**] at a rather modest level. This makes good sense: by shorting the call, the investor has a *negative* exposure to the risk endemic to the variance process thereby collecting positive risk premium [15]. Out of interest, we have also plotted the effect of including the volatility hedge terms in the optimal investment ratios [**Dash-dotted lines**]. As argued above, this respectively underestimates and overestimates the weights on stock and the derivative: here by as much as six percentage points for the stock and a single percentage point for the call (see the RHS figure for greater clarity). Volatility hedge corrections thus *seemingly* have the magnitude to perturb the terminal wealth of a rational investor by a measurable amount. Yet, this is in fact *not* the case when we Monte Carlo simulate the wealth process (4.11) of an investor trading in a Hestonian economy:

[14] See for example Munk [15] theorems 6.2 and 7.5

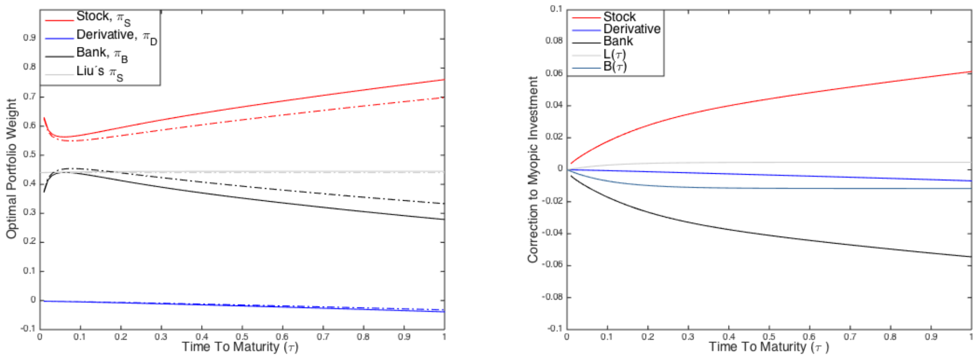

Fig. 4.1 **Left:** Optimal investment strategies in a (bank,stock,call option)-economy for different times to maturity with <u>s and v held constant</u>. Note that the investor shorts the derivative in order to enter into a significant long position in the underlying stock and deposit money in the bank. The grey line shows Liu's optimal portfolio weight on the stock when the investor disregards derivatives. The dash-dotted lines represent the corresponding strategies when we do not hedge stochastic variations in the state variable (volatility). For Liu's model this corresponds to the Merton weight λ_1/γ. **Right:** The size of the correction to the various portfolio weights brought about by hedging volatility. The figure also exhibits the magnitude of the deterministic functions $L(\tau)$ and $B(\tau)$. **NB:** as TTM approaches zero, so does v which creates problems with numerical instability in this region.

although the expected return is higher for someone who hedges volatility vis-à-vis one who does not, so is the associated variance. A Welch's t-test therefore cannot reject the null hypothesis that the two trading strategies have equal returns (p-value ≈ 0.65). This suggests that the real capital gains (if any) are to be garnered from access to the derivative security, and not the hedge corrections to volatility per se, in accordance with the findings by Liu and Pan [20].

4.4.3 Towards Higher Generality

A natural extension of the Heston model is to allow the model parameters to be time-dependent functions. This will not only allow for a financial landscape that changes dynamically over time, but also provide a more realistic pricing model the implied volatility surface of which calibrates much closer to market data. A tractable yet flexible functional form for this purpose is to specify $\{\kappa(t),\theta(t),\sigma_v(t),\rho(t),\bar{\lambda}_1(t),\bar{\lambda}_2(t)\}$ to be piecewise constant functions over some finite partition $t_0 < t_1 < \cdots < t_n$ of the market horizon $[0,T']$. That is, $\kappa(t) = \bar{\kappa}_1$ for $t \in (t_{n-1},t_n]$, $\kappa(t) = \bar{\kappa}_2$ for $t \in (t_{n-2},t_{n-1}]$ etc. with corresponding structures for the remaining parameters.[15] The key observation is then that we may solve all PDEs sequentially backwards in time. Firstly, we have that the optimal weights

[15] We shall still assume that the parameter-constants are specified within reasonable limits, i.e. that the governing dynamics (4.30) allows for a positive solution. Moreover, the Feller condition might be desired to be satisfied locally for each time interval of the partition.

(4.48)-(4.49) depends on the function B which solves the Riccati equation (4.47a). For the first subinterval $(t_{n-1}, t_n]$ expressed in backwards-time, $\tau \in [\tau_0, \tau_1) = [0, T - t_{n-1})$ with $\tau_k \equiv T - t_{n-k}$, we have the familiar boundary condition $B_1(0) = 0$ and solution B_1 as in equation (4.44) with $\{\bar{\kappa}_1, \bar{\theta}_1, \bar{\sigma}_{v1}, \bar{\rho}, \bar{\lambda}_{11}, \bar{\lambda}_{21}\}$. This gives us a value for $B_1(\tau_1)$ which will act as a (non-zero) boundary condition for B_2 on the next interval $[\tau_1, \tau_2)$. We may then proceed sequentially backwards in time[16] to obtain $B_2, \dots B_n$ over $[\tau_2, \tau_3), \dots, [\tau_{n-1}, \tau_n)$ to cover the whole of $[0, T] \subseteq [0, T']$. Secondly, we need to calculate Δ and ν over the same sequence of subintervals (extended to cover $[0, T']$) to complete the expression for our market weights. It turns out that this calculation is virtually identical to the one just considered: D_j and C_j of (4.36c)-(4.36d) solve a system of ODEs which is the same as (4.47a)-(4.47b) (see Mikhailov and Nögel [23] for details). Hence, we may compute the optimal weights for a time-dependent parameter specification of the Heston model as well (for reasons of brevity we exclude the technical details from this paper).

4.5 The Empirical Perspective

Based on our optimal portfolio weights in a Heston driven (B, S, D)-economy, we proceed to perform an empirical experiment which aims to measure the degree to which the inclusion of plain vanillas impacts the financial wealth of a utility maximising investor. In particular, we set out to perform a simple automated trading experiment where we let historical market prices from the S&P 500 index play the role of the fundamental risky security and where market prices of call options written on the same index constitute the derivative available in the economy. We will use market interest rates for the money account.

4.5.1 Market Data

For the tradable stock, we use 3,909 daily prices of the S&P 500 index from the period 2000-01-03 to 2015-08-31. The market price (sourced from Wharton Research Data Services[17]) is plotted in figure 4.2 together with the daily variance. The variance process is measured from high-frequency data with the realised volatility measure and we use pre-computed estimates from the Oxford-Man Institute's realised library.[18].

For the tradable derivative, we use daily mid-market prices of European call options on the S&P 500 index from the same time period, as shown in figure 4.3. The time period

[16] To this end we need the Ricatti equation with a *non-zero* initial condition i.e. $y'(x) = ay^2(x) + by(x) + c$, $y(0) = y_0$, which has the solution $y(x) = y_0 + [2(ay_0^2 + by_0 + c)(e^{\delta x} - 1)]/[(\delta - b - 2ay_0)(e^{\delta x} - 1) + 2\delta]$ where $\delta \equiv \sqrt{b^2 - 4ac}$ assuming $b^2 > 4ac$.

[17] https://wrds-web.wharton.upenn.edu/wrds/.

[18] The Realised Library version 0.2 by Heber, Gerd, Lunde, Shephard and Sheppard (2009) - see http://realized.oxford-man.ox.ac.uk For details on the realised volatility measure, see e.g. Andersen and Teräsvirta [1].

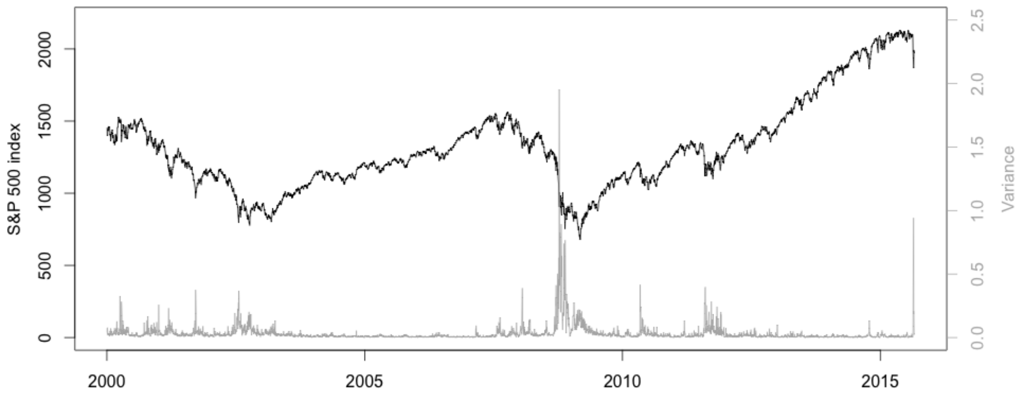

Fig. 4.2 Daily market prices and measured variance from the S&P 500 index. The price data is sourced from Wharton Research Data Services while the variance data is sourced from the Oxford-Man Institute's realised library.

covers prices of 23 call options with medium-sized times to maturity (in the range 18 to 36 months subject to data availability) and the strike-price of each option is selected to be ATM at initiation (or as close as possible thereto - subject to availability) thereby sowing the seed for high exposure to volatility risk. The strike-price and time-to-maturity structure is shown in figure 4.4. Note the varying TTMs for the options, which again are symptomatic of the data set available (we use the Option Metrics database sourced through Wharton Research Data Services).

For our last asset in the (B,S,D) economy, we use the daily short-term LIBOR rate for an interest to the risk-free money account. The LIBOR market-data is from the Option Metrics database as well.

4.5.2 Parameter Estimation

For our empirical experiment, we will trade according to the optimal portfolio weights $(\pi^*_{Bt}, \pi^*_{St}, \pi^*_{Dt})$ which are functions of the model parameters and the current stock price, call price and variance level at time t. To this end, we estimate the model parameters with the following approach. First, we estimate $(\kappa, \theta, \sigma_v, \rho)$ of the Cox-Ingersoll-Ross process from the daily variance data with a maximum likelihood method.[19] Secondly, based on these parameters, to determine the market prices of risk, $(\bar{\lambda}_1, \bar{\lambda}_2)$, we minimise the squares

[19] We use numerical optimisation of a Gaussian likelihood, from the method of Sørensen [30] based on estimating functions.

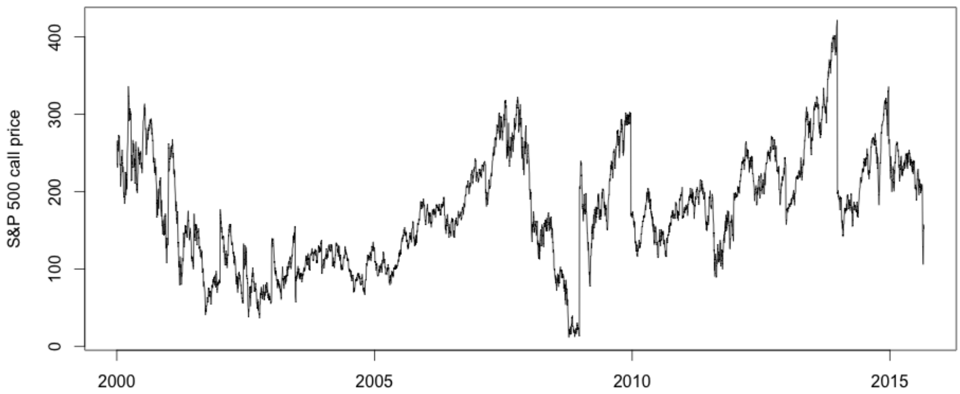

Fig. 4.3 Daily mid-market prices of European call options on the S&P 500 index sourced from Wharton Research Data Services (Option Metrics data). The corresponding strike prices and maturities are shown in figure 4.4.

error between daily observed S&P 500 call option prices, $\{\hat{C}_t\}$, and daily theoretical Heston prices $\{C_t^{He} = C_t^{He}(\bar{\lambda}_1, \bar{\lambda}_2)\}$, where the strike-maturity structure is chosen as in figure 4.4. Specifically, we solve the minimisation problem

$$(\bar{\lambda}_1, \bar{\lambda}_2) = \underset{(\bar{\lambda}_1, \bar{\lambda}_2) \in \mathbb{R}^{2+}}{\operatorname{argmin}} \sum_{t \in \mathbb{T}} \left(\hat{C}_t - C_t^{He}(\bar{\lambda}_1, \bar{\lambda}_2) \right)^2,$$

where $\mathbb{T} = \{$trading days between 2000-01-03 and 2015-08-31$\}$. The resulting parameters estimated from the market data are given in table 4.1.

	κ	θ	ε	ρ	$\bar{\lambda}_1$	$\bar{\lambda}_2$
Estimate	9.71	0.030	2.72	-0.17	0.88	-0.29

Table 4.1 Estimated model parameters with the two-step approach. The estimates are based on 3,909 daily observations of the stock price, the variance and the call price. Note that $\bar{\lambda}_1 > 0$ whilst $\bar{\lambda}_2 < 0$.

Remark 4.5. A few remarks on the above estimation procedure are in order here. First, note that when we will use the parameter estimates for the forthcoming trading experiment, we employ ex-ante estimates based on actual "future" market data. An alternative is to estimate parameters from historical data prior to the trading period. However, due to the amount of available data, this would impair the accuracy our estimates. Since we are primarily interested in the the efficiency of the trading strategy, (and not in the parameter estimation problem per se) we require as robust estimates as possible. Thus, we

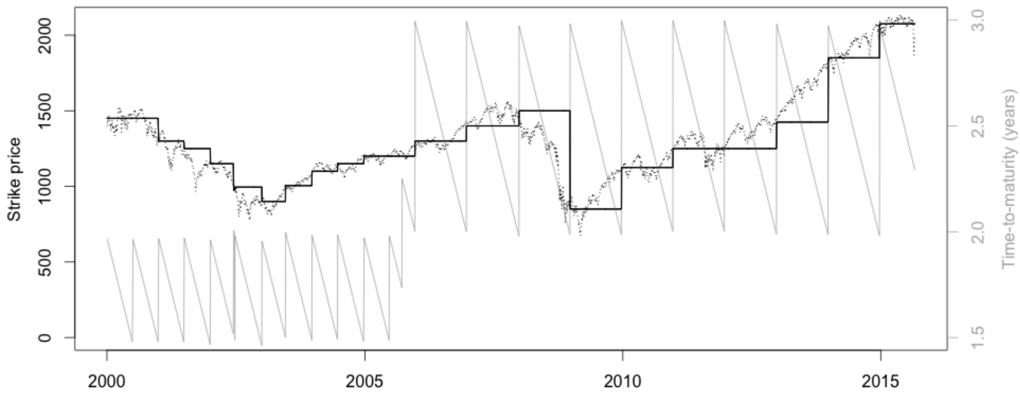

Fig. 4.4 The daily strike-price (black line) and time-to-maturity (grey line) for the call price in figure 4.3. The dotted line shows the S&P 500 index level.

opt for the former alternative. Secondly, we use a rather unconventional estimation procedure whereby we sequentially estimate the CIR-parameters under the statistical measure $\mathbb{P}$, followed by a mean square optimisation to back out the market prices of risk. A more commonplace approach is to formulate the pricing model under $\mathbb{Q}$ directly (see Remark 4.4), and to estimate the risk-neutral parameters from option data alone, again through mean square principles . This estimation approach is referred to as model-to-market calibration, and typically a whole surface of option prices is employed for day-to-day estimation of the parameters. Since we require statistical CIR-parameters along with the market price of risks separately for calculation of the optimal portfolio weights, we use the two-step approach in place of the calibration method.

4.5.3 Empirical Trading Experiment

With market data for the (B,S,D) economy from the S&P 500 index and the LIBOR rate, we set out to perform an empirical trading experiment. We intend to invest in a portfolio according to the optimal weights $(\pi^*_{Bt}, \pi^*_{St}, \pi^*_{Dt})$ and we trade dynamically with daily rebalancing as the time evolves during the period 2000-01-03 to 2015-08-31. Hence, if ΔX_{t_i} denotes the daily price-change from t_i to t_{i+1} of an asset in the economy, this means that we realise a daily change in the portfolio value

$$\Delta \mathscr{W}_{t_i} = \mathscr{W}_{t_i} \left(\pi^*_{Bt_i} \frac{\Delta B_{t_i}}{B_{t_i}} + \pi^*_{St_i} \frac{\Delta S_{t_i}}{S_{t_i}} + \pi^*_{Dt_i} \frac{\Delta D_{t_i}}{D_{t_i}} \right),$$

$$\mathscr{W}_{t_0} = w_0 \in \mathbb{R}^+,$$

for all dates $t_0, t_1, \ldots$ in the investment period and the realised daily wealth amounts to $\mathscr{W}_{t_{i+1}} = \mathscr{W}_{t_i} + \Delta \mathscr{W}_{t_i}$. Note that the portfolio is self-financing: an initial amount w_0 is invested at the initial time and there is no infusion or withdraw of capital from the portfolio during the investment period.

In addition to the parameters in table 4.1, with fix the risk aversion parameter to the arbitrary value $\gamma = 2$. For comparison purposes, we include a "naive" trading strategy with a constant equal weight invested in each asset, $(\pi_{Bt}, \pi_{St}, \pi_{Dt}) = (1/3, 1/3, 1/3)$. We also trade according to Liu's optimal investment strategy in the limited economy, that is, we invest in (B, S) with portfolio weights $(\pi_{Bt}^{Liu}, \pi_{St}^{Liu})$ (and $\pi_{Dt} = 0$ for the call option).

With this in mind, we conduct the following following two experiments for each of the three trading strategies:

1. **Trading throughout 2000-2015.** We set the investment period to be 2000-01-03 to 2015-08-31 and initialise the portfolios with a wealth $w_0 = 1,000$. We do not trade over the dates when there is a "new" option, i.e. every time there is a new expiry date since this would give false price moves of the call option due to changes in the strike price (see figure 4.3 and 4.4). The same rule pertains to Liu's strategy, even though there is no trading in the option. The realised wealth processes from trading according to the three strategies are shown in figure 4.5 while the optimal portfolio weights are shown in figure 4.7.
   To be able to compare the strategies from a financial perspective, we calculate realised Sharpe-ratios of daily returns as

$$\mathscr{S} = \frac{\text{Mean}(R_{t_i} - r_{t_i})}{\text{SD}(R_{t_i})},$$

   where $R_{t_i} = \log(\mathscr{W}_{t_i}) - \log(\mathscr{W}_{t_{i-1}})$, $t_1, t_2, \ldots$ are the daily returns of the investment portfolio and r_{t_i} is the daily returns of the money account (the daily return from the LIBOR rate). The results are shown in table 4.2.

Strategy	Mean return ($\mu(R)$)	Std. Dev. ($\sigma(R)$)	Sharpe-ratio ($\mathscr{S}$)	Sharpe-R. annual
$(\pi_{Bt}^*, \pi_{St}^*, \pi_{Dt}^*)$	0.011%	0.65%	0.39%	7.46%
$(\pi_{Bt}^{Liu}, \pi_{St}^{Liu}, 0)$	0.010%	0.56%	0.37%	7.16%
$(1/3, 1/3, 1/3)$	0.005%	2.27%	-0.14 %	-2.64%

Table 4.2 The Sharpe-ratio, mean and standard deviation of daily portfolio returns from the three strategies when trading throughout the whole period 2000-01-03 to 2015-08-31. The last column shows the annualised Sharpe-ratio. The daily mean-return of the money account is 0.0082%, which corresponds to an annualised return of 3.0%.

2. **Investment periods according to the option-expiry structure.** For our second empirical experiment, we reset our investment portfolio every time there is a "new" option, i.e. every time there is a new expiry date and strike-price (see figure 4.4). We set the investment period accordingly, i.e. to start when we reset the portfolio and to end at the

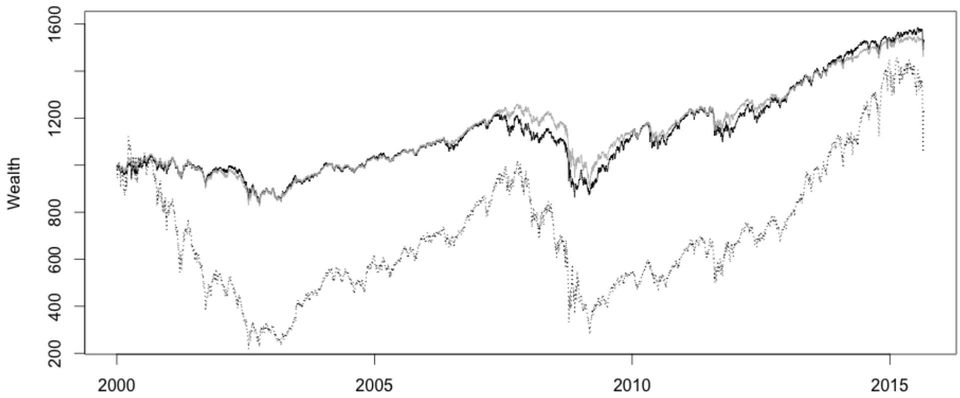

Fig. 4.5 Trading throughout 2000-01-03 to 2015-08-31: wealth processes from trading in {LIBOR,S&P500,Call} according to the naive strategy (dotted line) and optimal investment strategy (black line), and from trading in {LIBOR,S&P500} with Liu's optimal strategy (grey line).

date on which we will reset the portfolio the next time. The resulting realised portfolio value-processes from the optimal (B,S,D)- and (B,S) strategies are shown in figure 4.6 and the portfolio weights of the two strategies are plotted in figure 4.8. The Sharpe-ratios based on realised daily returns of the investment portfolios are collected in table 4.3, where the results for the naive strategy are included as well.

Strategy	Mean return ($\mu(R)$)	Std. Dev. ($\sigma(R)$)	Sharpe-ratio	Sharpe-R. annual
$(\pi^*_{Bt}, \pi^*_{St}, \pi^*_{Dt})$	0.010%	0.55%	0.37%	7.13%
$(\pi^{Liu}_{Bt}, \pi^{Liu}_{St}, 0)$	0.011%	0.65%	0.36%	6.93%
$(1/3, 1/3, 1/3)$	0.005%	2.28%	-0.14 %	-2.65%

Table 4.3 The Sharpe-ratio, mean and standard deviation of daily portfolio returns from the three strategies when trading according to the option-expiry structure. The last column shows the annualised Sharpe-ratio.

A note on the interpretation of these results is in order. First, whilst trading strategy (1) picks up a slightly higher mean return along the way for the (B,S,D) portfolio, the associated variance of returns is also higher (the risk averse investor might choke on this). Granted: the (B,S,D) strategy does come out victorious in the end, but this is largely happenstance: had we terminated our algorithm during the financial brouhaha, our conclusion would have been different. Indeed, the realised Sharpe ratio obtained vis-à-vis Liu's

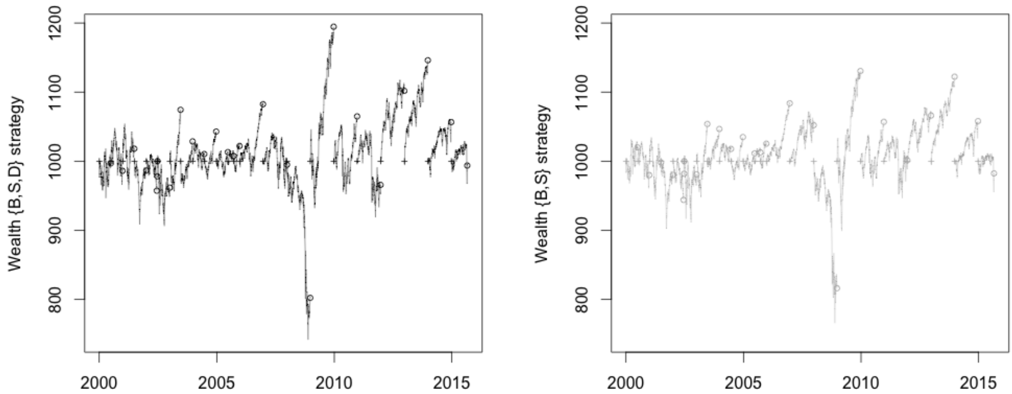

Fig. 4.6 Trading according to the option-expiry structure. **Left:** wealth process from trading in {LIBOR,S&P500,Call} according to the optimal investment strategy during investment periods that matches the expiry/strike structure of the call options. A cross indicates the beginning of an investment period (with initial wealth $w_0 = 1,000$) while a circle shows the terminal wealth at the end of the period. **Right:** wealth process from trading in {LIBOR,S&P500} according to Liu's optimal investment strategy during the same investment periods.

derivative-free trading strategy is of a very modest nature (0.3 percentage points in annualised Sharpe ratio).

As for strategy (2) the conclusion is largely invariant, only here the mean return and standard deviation for the (B,S,D) strategy are actually lower than Liu's strategy, jointly leading to a Sharpe ratio of (modest) superiority. We stress that this is about as far as we can go in our analysis here: while strategy (2) on first sight seems to warrant Welchian hypothesis testing (after all, we are seemingly performing the same experiment 23 times), this would be profoundly statistically flawed. Essentially, whenever we sample a wealth path we do so from a different space, where virtually every underlying parameter (bar the risk aversion) is different. Our conclusion is thus of a purely observational nature: we see that the utility maximising trading strategy including derivatives outperforms the one without; yet the financial benefit is *hardly* worth talking about. Furthermore, this statement is obviously contingent upon the overall setting of our experiments: from when we chose to terminate our algorithm, to which derivative product we decided to consider. Indeed, our analysis has shamelessly disregarded transactions costs, which will have a significant impact on our wealth paths. Nevertheless, as a first analysis, the results are certainly noteworthy.

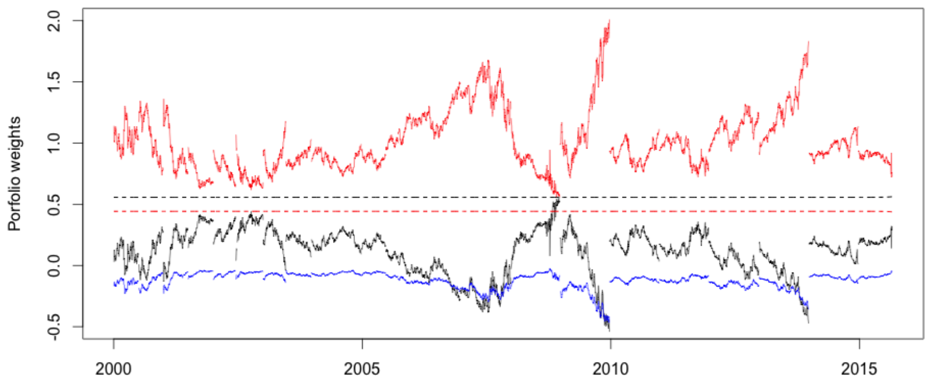

Fig. 4.7 Trading throughout 2000-01-03 to 2015-08-31: portfolio weights from the optimal {B,S,C}-strategy (solid lines) and Liu's {B,S}-strategy (dashed lines). Black lines show the weights in *B*, red lines the weights in *S* and blue line the weight in *C*. Notice that the weights of the {B,S,C}-strategy makes sudden jumps at the time-points where the expiry/strike of the option changes, and that we do not trade during these dates. Furthermore, observe that the derivative position is everywhere negative.

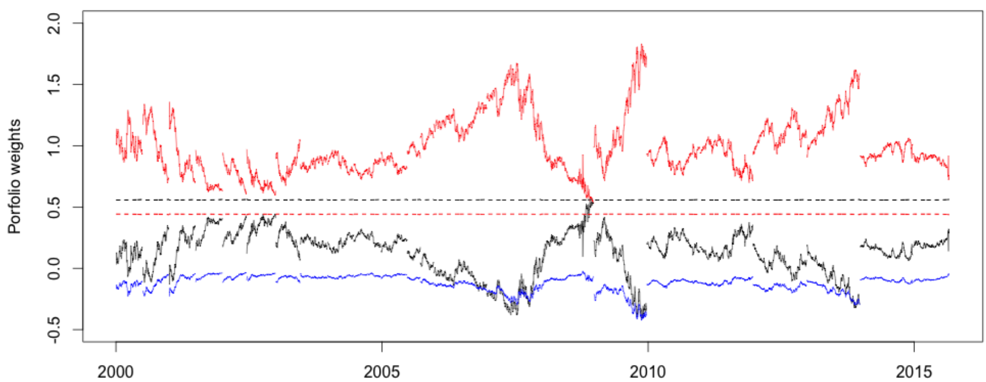

Fig. 4.8 Trading according to the option-expiry structure: portfolio weights from the optimal {B,S,C}-strategy (solid lines) and Liu's {B,S}-strategy (dashed lines). Black lines show the weights in *B*, red lines the weights in *S* and blue line the weight in *C*. Notice that the weights of the {B,S,C}-strategy makes sudden jumps at the time-points where the expiry/strike of the option changes, and that the investment portfolio is reset at these dates. Again, the derivative position is everywhere negative.

4.6 Conclusion

In the first part of this paper we derived optimal portfolio weights for a utility maximiser who trades in a (B,S,D)-economy in a *generic* stochastic volatility framework, thus extending the work by Liu and Pan. In the second part, we derived explicit expressions for the Heston model, which benefits by admitting closed form expressions for plain vanilla European options. Here, empirically based Monte Carlo simulations suggest that there is no tangible welfare benefit associated with hedging volatility per se: in other words, if our portfolio benefits from the inclusion of derivatives, it does so through shear myopic diversification. Liu and Pan are optimistic on this account: through quasi-empirical considerations they find considerable improvements[20] in the certainty equivalent wealth for investors who trade in derivatives. Our own findings, which arguably have a much firmer grounding in empirical data, are considerably more pessimistic: whilst we can extract a higher Sharpe ratio than Liu's (B,S)-strategy, this is of a very modest nature, and arguably a "fluke" brought about by the overall circumstances of our experiment. Obviously, we cannot rule out that more "cognisant" derivative strategies will have a greater impact upon the investor's wealth level: we will leave this for future research.

[20] The reported number is 14.2% for an investor who trades in a straddle position (long call, long put, same strike).

References

1. Andersen and Teräsvirta, *Handbook of Financial Time Series*, Springer, 2009, pp. 555-575.
2. Bakshi and Kapadia, *Delta-Hedged Gains and the Negative Market Volatility Risk Premium*, *Review of Financial Studies* 16, p. 527-566.
3. Björk, *Arbitrage Theory in Continuous Time*, Oxford University Press, 3rd edition.
4. Björk, Davis, and Landen, *Optimal investment under partial information*, *Mathematical Methods of Operations Research*, 71 (2010) pp. 371-399.
5. Branger and Hansis, *Asset Allocation: How Much Does Model Choice Matter?*, *Journal of Banking & Finance*, July 2012, Volume 36, Issue 7, pp. 1865-1882.
6. Breeden and Litzenberger, *Price of State-contingent Claims Implicit in Option Prices*, *The Journal of Business*, Vol. 51, No. 4 (Oct. 1978), pp. 621-651.
7. Brennan and Xia, *Stochastic Interest Rates and the Bond-Stock Mix*, *European Finance Review* (2000) 4 (2): 197-210.
8. Burden and Faires, *Numerical Analysis*. Brooks/Cole, 9th edition.
9. Carr and Sun, *A New Approach for Option Pricing under Stochastic Volatility*. In *Review of Derivatives Research*, May 2007, Volume 10, Issue 2, pp 87-150.
10. Chacko and Viceira, *Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets*. Rev. Financ. Stud. (Winter 2005) 18 (4): 1369-1402.
11. Cont and Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC, 2004.
12. Dimitroff, Lorenz and Szimayer, *A Parsimonious Multi-Asset Heston Model: Calibration and Derivative Pricing*. 2009, *Fraunhofer ITWM*.
13. Drimus, *Options on realized variance by transform methods: A non-affine stochastic volatility model*. 2011, *Quantitative Finance*.
14. Duffie, *Dynamic Asser Pricing Theory*. Princeton University Press, 3rd edition.
15. Ellersgaard and Jönsson, *Stochastic Volatility for Utility Maximisers Part I - The Bond-Stock Economy*. Unpublished manuscript. 2015.
16. Gatheral, Jaisson, Rosenbaum, *Volatility is Rough*, 2014. http://arxiv.org/pdf/1410.3394v1.pdf.
17. Heston, *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*, *The Review of Financial Studies*, 1993, Volume 6, number 2, pp. 327-343.
18. Kraft, *Optimal Portfolios and Heston's Stochastic Volatility Model: An Explicit Solution for Power Utility*, Quantitative Finance, Volume 5, Issue 3, 2005, pp. 303-313,
19. Liu, *Portfolio Selection in Stochastic Environments*, in *The Review of Financial Studies*, Vol. 20, No. 1 (Jan., 2007), pp. 1-39.
20. Liu and Pan, *Dynamic Derivative Strategies*, in *Journal of Financial Economics*, Volume 69, Issue 3, September 2003, pp. 401-430.
21. MacDonald, *Properties of the Confluent Hypergeometric Function*. Technical Report No. 84. 1948. Research Laboratory of Electronics, MIT. http://dspace.mit.edu/bitstream/handle/1721.1/4966/RLE-TR-084-14234239.pdf?sequence=1.
22. Merton, *Lifetime Portfolio Selection under Uncertainty: the Continuous-Time Case*. *The Review of Economics and Statistics* 51 (3) pp. 247?257 (1969).
23. Mikhailov and Nögel, *Heston?s Stochastic Volatility Model: Implementation, Calibration and some Extensions*, 2004, John Wiley and Sons.
24. Munk, *Dynamic Asset Allocation*. Unpublished lecture notes. December 18, 2013 edition http://mit.econ.au.dk/vip_htm/cmunk/noter/dynassal.pdf.
25. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications*. *Stochastic Modelling and Applied Probability 61*, Springer.
26. Reiss, and Wystup. *Computing Option Price Sensitivities Using Homogeneity*, *Journal of Derivatives*, 2001, 9(2): 41?53.
27. Ribeiro, and Poulsen, *Approximation Behooves Calibration*, *Quantitative Finance Letters*, 2013 vol. 1(1), pp. 36-40.
28. Ross, *Stochastic Control in Continuous Time*.
Unpublished lecture notes. http://www.swarthmore.edu/NatSci/kross1/Stat220notes.pdf.
29. Selby and Strickland, *Computing the Fong and Vasicek Pure Discount Bond Price Formula*. *The Journal of Fixed Income*, September 1995, Vol. 5, No. 2: pp. 78-84.

30. Sorensen, *On the Asymptotetics of Estimating Functions*. *Brazilian Journal of Probability and Statistics*, 1999, 13, pp. 111-136.
31. Vasicek and Fong, *Fixed-income Volatility Management*, *The Journal of Portfolio Management*, Volume 17, No. 4, Summer 1991, pages 41-46.
32. Vasicek and Fong, *Interest Rate Volatility as a Stochastic Factor*, 1991, Gifford Fing Associates working paper.
33. Weron and Wystup, *Heston's Model and the Smile*, in *Statistical Tools for Finance and Insurance*, pp. 161-181, Springer 2011.
34. Wong and Heyde, *On Changes of Measure in Stochastic Volatility Models*, *Journal of Applied Mathematics and Stochastic Analysis*, Volume 2006, Article ID 18130, pages 1-13.
35. Zariphopoulou, *A Solution Approach to Valuation with Unhedgeable Risks*, *Finance and Stochastics*, 5, pp. 61–82, 2001.

Chapter 5
Optimal Hedge Tracking Portfolios in a Limit Order Book
The Limit/Market Order Duality

Simon Ellersgaard

Abstract In this paper we develop a control theoretic solution to the manner in which a portfolio manager optimally should track a targeted Δ, given that he wishes to hedge a short position in European call options the underlying of which is traded in a limit order book. Specifically, we are interested in the interplay between posting limit and market orders respectively: when should the portfolio manager do what (and at what price)? To this end, we set up an Hamilton-Jacobi-Bellman quasi variational inequality which we can solve numerically. Our scheme is shown to be monotone, stable, and consistent and thence, modulo a comparison principle, convergent in the viscosity sense. Finally, we provide a concrete numerical study, comparing our algorithm with more naïve approaches to delta-hedging.

Key words: Delta Hedging, Limit Order Book, Hamilton-Jacobi-Bellman Equation.

The Department of Mathematical Sciences
Universitetsparken 5, 2100 Copenhagen Ø, Denmark
e-mail: ellersgaard@math.ku.dk.

5.1 Introduction

5.1.1 Mathematics of the Limit Order Book

In recent years the centralised trading platform known as the *limit order book* (LOB) has attracted considerable interest from the mathematical finance community. To a large extent this is a natural response to the evolution of the financial markets per se: more than half of the world's stock exchanges are now order driven, with many operating exclusively so (Hong Kong, Tokyo, Toronto etc.), whilst others have adopted a hybrid variant thereof (NYSE, NASDAQ and LSE) [25] [42]. From a modelling perspective we can also construe this paradigmatic shift as a considerable step towards axiomatic realism: gone are the days where modellers casually endorse Black-Scholes-Merton type assumptions; enter a world of de facto transaction costs, finite divisibility of assets, market impact, and a cacophonous conglomerate of price quotes. Finally, one cannot ignore the considerable appeal of capitalising on integrating the limit order book with algorithmic trading strategies (*high frequency trading*). For example, significant capital gains can be accrued simply by exploiting the upper bound on human comprehension speed: by having machine intelligence do our trading for us, we enter a domain of hitherto unexplored ultra-transient market inefficiencies (Lewis' *Flash Boys* [33] is a luminous account in this regard). Indeed, the seriousness of this business is forcefully cemented by noting that the market share of high frequency trades long has surpassed that of institutional investors. For instance, it is estimated that high frequency trades account for roughly fifty percent of all equity shares traded in the United States as of 2014 (TABB Group, [11]), with some sixty-odd percent for the futures market.[1]

Broadly speaking, the mathematical literature on the LOB falls within two non-mutually exclusive categories: on the one hand, the descriptivists aim to map so-called *stylised facts* (i.e. empirically consistent qualia) of the order book onto mathematical formalism. Key references here include Cont et al.'s queuing model, [5], Carmona and Webster's reinterpretation of the self-financing condition in high frequency markets, [12], and Donier et al.'s model of non-linear market impact, [18]. On the other hand we find the prescriptivists, who under fixed assumptions about market dynamics and utility preferences of so-called rational agents derive wealth optimising trading strategies using control theoretic arguments a la Merton, [35]. Here, the desire to formulate a solvable control problem invariably surpasses the need for model realism. Classical questions pertain to optimal trade execution (how should large orders optimally be partitioned such as to minimise their market impact?) and optimal limit order quotes (which prices should the market maker post to optimise his bequest?). As for the former, the key reference paper is indubitably the discrete time mean-variance optimisation by Almgren and Chriss [4] (and the continuous time extension by Almgren [3]). Other contributions worthy of mentioning include Forsyth [23], Gueant et al. [26], He and Mamaysky [29] and Obizhaeva and Wang [37]. As for the question of optimal limit order quotes, much research arguably takes its vantage point in the work by Avellaneda and Stoikov [6], which in turn is heavily inspired by a compara-

[1] Needless to say, these numbers are merely indicative: different sources invariably arrive at different estimates. Nonetheless, 50% is very much in the lower end of the scale, [17].

tively obscure paper by Ho and Stoll [30]. Good references here include the formalising paper by Fodra and Labadie [22], and the extension work by Cartea et al. [13].

In this paper we concern ourselves with an area which has thus far received relatively little attention, viz. derivative hedging in the limit order book using prescriptivist principles. To the best of our knowledge, it is only recently that a few models have been proposed on this matter, including Agliardi and Gencay's discrete time investigation[2] in which an explicit solution is found for an option hedger who aims to minimise illiquidity costs and the hedging error, [5]. In a somewhat similar vein, Li and Almgren [34] consider continuous time hedging in the presence of temporary and permanent market impact. A key insights here is that the portfolio manager no longer finds it tenable to be perfectly hedged or even within a fixed distance of being hedged. Rather, he may find himself arbitrarily mis-hedged, moving towards the Black-Scholes hedge ratio with a trading intensity proportional to the degree of mis-hedge and inversely proportional to illiquidity. Our paper, however, takes a somewhat different approach in the sense that we are less concerned with determining an optimal hedge ratio endogenously, and more with exploring the financial benefits offered by placing both limit and market orders when pursuing a pre-defined hedge strategy. In this sense, our work is much more akin to the optimal trade execution study by Cartea and Jaimungal [14], rather than the existing literature on limit order book hedging.

Remark 5.1. A plurality of sources provide highly readable accounts of the basic mechanisms and nomenclature of the limit order book, which prompts us to forgo a similar overview. Instead, we refer the reader to the survey paper by Gould et al. [25], or the more quantitatively oriented textbook by Foucault et al. [20].

5.1.2 Philosophy and Overview

Let it thus be made abundantly clear that our goal here is not to dictate which position in the underlying LOB-traded asset optimally hedges a derivative portfolio (although this surely is an important question). Instead, our interests lie with the duality inherent to order book trading: given an exogenously specified tolerance for "risk" when should a portfolio manager trade in limit orders and when should he trade in market orders? The Δ employed in our equations is for all practical purposes completely generic, meaning that our model readily is adaptable to more sophisticated LOB hedging strategies that might be proposed in the future. *Tracking* is thus the operative word of the title. From a slightly broader perspective, this paper is inevitably social science research of the pseudo-decreeing kind: based on a rudimentary model of the order book, and potentially mis-guided notions of how utility preferences can be formulated mathematically, we derive optimal limit order quotes and market order stopping times for a wealth maximising agent. At no point do we make such grandiose claims (otherwise so ubiquitous within our field) as to purport that these results have a direct impact or relevance to the actual financial markets. Nonetheless,

[2] This paper is heavily inspired by the model proposed in [37].

this does obviously not stop us from hoping that our work at least plants a seed for fruitful future research.

The structure of the rest of this paper is as follows: in section two we explicate the fundamental assumptions of the market and the portfolio manager and state the associated control problem. Section three is a survey of the exogenously specified hedge ratio: a careful analysis shows that the dynamics of the order book is weakly convergent towards the dynamics associated with the targeted delta. Section four exposes the Hamilton-Jacobi-Bellman quasi-variational inequality (HJB QVI) associated with the control problem, and suggests a dimensional reduction. Finding ourselves unable to extract an analytic solution thereto, section five sets us up for a numerical scheme which is shown to have the desired properties of monotonicity, stability, and consistency as dictated by Barles and Souganidis [8]. Finally, section six provides concrete numerical results, comparing the performance of our algorithm with a more naïve approach to Δ-hedging. Section seven concludes.

5.2 A Control Approach to Hedging in the LOB

5.2.1 Market Assumptions

As convention would have it, we consider a financial market formally captured by a filtered probability space $(\Omega, \mathscr{F}_T, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} \equiv \{\mathscr{F}_t\}_{t\in[0,T]}$ is the natural filtration generated by the stochastic processes $S_t, L_t^{\pm}$ and $M_t^{\pm}$, which will be defined shortly.[3] For simplicity, the market is assumed interest rate free, and of a singular non-dividend paying risky asset (a stock) which is traded in a limit order book. Modulo an important caveat explicated in remark 5.2 below, all prices in the order book are assumed to be integer multiples of the *tick size*, $\sigma_m \in \mathbb{Q}^+$. In particular, we assume that the dynamics of the *mid-price* obeys

$$dS_t = \sigma_m \int_{\mathbb{A}} z(J_1(dt \times dz) - J_2(dt \times dz)), \tag{5.1}$$

where $S_0 = s_0 \in \sigma_m \mathbb{N}^+$, $\mathbb{A} \subseteq \mathbb{R}^+ \equiv (0, \infty)$, and $J_i : \Omega \times [0,T] \times \mathbb{N}^+ \mapsto \mathbb{N}$ are independent Poisson random measures[4] of common intensity measure

$$\mu_i(dt \times dz) \equiv \mathbb{E}[J_i(dt \times dz)] = \gamma_t F_t(dz)dt,$$

for $i = 1,2$. Here γ_t is a non-negative process, which encodes the jump rate intensity, meaning that the probability that the stock price jumps upwards (resp. downwards) over the incremental time step $(t, t+dt]$ is $\lambda_t dt$. Furthermore, F_t is a cdf with integer support ($\subset \mathbb{A}$) which captures the conditional distribution of the number of tick points a process jumps, given that a jump occurs at time t. A couple of remarks are worth attaching to (5.1): first, the dynamics is form-invariant upon rewriting it in terms of the compensated Pois-

---

[3] For technical reasons $\mathbb{F}$ is augmented to satisfy the *usual conditions* i.e. $\mathscr{F}_t^{S,L^{\pm},M^{\pm}} = \sigma(\{S_u, L_u^{\pm}, M_u^{\pm} : 0 \leq u \leq t\} \cup \mathscr{N})$ where $\mathscr{N} \equiv \{O \subset \Omega : \exists F \in \mathscr{F} \text{ s.t. } O \subset F, \mathbb{P}(F) = 0\}$.

[4] For an introduction to this rich field see e.g. (in increasing order of complexity) Hanson [28], Cont and Tankov [16], and Last and Brandt [32].

son random measures $\tilde{J}_i(dt \times dz) \equiv J_i(dt \times dz) - \gamma_t F_t(dz)dt$. Hence, S_t is an $\mathbb{F}$-martingale. Secondly, there is a potential pathology nested in the dynamics in the sense that it admits negative price processes: nonetheless, we shall assume that the probability of this occurring is sufficiently low to be ignored (an otherwise very reasonable assumption over short temporal horizons / for low intensity processes).

The difference between the *best bid*, S_t^b, and the *best ask*, S_t^a, (the so-called *bid-ask spread*) is assumed time dependent and of magnitude

$$S_t^a - S_t^b = 2\Upsilon_t,$$

where $\Upsilon : [0,T] \mapsto \sigma_m \mathbb{N}^+$. Thus, investors who trade through market orders will minimally pay

$$S_t^a = S_t + \Upsilon_t$$

per share if they wish to buy the stock, and maximally earn

$$S_t^b = S_t - \Upsilon_t$$

per share if they wish to sell the stock, depending on the size of their orders vis-à-vis the number of shares available at the various price levels. At the aggregate level, we assume that market orders arrive in the limit order book, in a manner which can be modelled by an inhomogeneous Poisson process, M_t, with intensity rate

$$\lambda_t = \xi_t \exp\{\kappa_t \Upsilon_t\},$$

where ξ_t and κ_t are real valued positive functions. Clearly, whilst market orders are guaranteed instantaneous execution, the liquidity taking fees they incur may be less than appealing. This is not to say that one optimally should opt for the limit order alternative: despite superior prices, there is no guarantee of execution at all.

5.2.2 Portfolio Manager Assumptions

We consider the case of a portfolio manager who at time $t = 0$ sells off $\mathfrak{N}$ European call options of strike K and maturity T. Modulo an exogenously specified tolerance for risk, his aim is to keep his portfolio "delta neutral" throughout, whilst simultaneously maximising his terminal payoff at time T. Here, *delta neutrality* is to be understood as a pre-specified optimal inventory level $\Delta_t = \Delta(t, S_t)$ - the number of shares the portfolio manager *ideally* wishes to keep in his portfolio to hedge his short option position. If his inventory level at time t, Q_t, is found to be in excess of Δ_t, the portfolio manager is incentivised to offload some of his shares. Conversely, if Q_t falls short of Δ_t, he is incentivised to acquire shares. To this end, we assume that the portfolio manager places unit-sized limit orders and (if necessary) integer-sized market orders in the order book, but only one of the two at any given time. From this, one is prompted to ask the following questions, which will form the backbone of this paper:

Question 5.1. When should the portfolio manager trade in limit orders and when should he trade in market orders? In particular, regarding the former, which price quotes should he employ?

Whilst market sell (buy) orders are assumed to take place at the best bid (ask), we assume the portfolio manager is at liberty to decide how deep in the limit order book he places his limit orders. Specifically, let $S_t^- = S_t + \delta_t^-$ be the price level at which the portfolio manager places his ask quote at time t, and let $S_t^+ = S_t - \delta_t^+$ be the price level at which the portfolio manager places his bid quote at time t, where $\delta_t^\pm \geq 0$ is known as the *spread*. For mnemotechnical purposes we designate objects that give rise to a lower (higher) inventory by the superscript - (+).

Remark 5.2. It will be convenient to assume that $\delta_t^\pm$ can be chosen from all of $\mathbb{R}_+$, and not just $\sigma_m\mathbb{N}^+$. Although this clearly is incongruent with the discrete nature of the LOB, it will simplify the optimisation problem to be stated shortly.

To capture the execution risk inherent to limit orders, we assume that the probability of a sell (buy) limit order being lifted, given that a market order arrives, is of the form

$$\exp\{-\kappa_t \delta_t^-\},$$

($\exp\{-\kappa_t \delta_t^+\}$). Thus, from the market assumptions it follows that the portfolio manager's successful limit sell (buy) order executions are modellable as an inhomogeneous Poisson process L_t^- (L_t^+) with intensity

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}\{L_{t+\Delta t}^- - L_t^- = 1\} = \lambda_t \exp\{-\kappa_t \delta_t^-\}, \tag{5.2}$$

($\lim_{\Delta t \to 0} \mathbb{P}\{L_{t+\Delta t}^+ - L_t^+ = 1\}/\Delta t = \lambda_t \exp\{-\kappa_t \delta_t^+\}$).[5] The intuition here is clear: the further away from the mid-price the portfolio manager places his limit quotes, the less likely it is they will be executed. This disincentivises the portfolio manager from posting extreme limit order quotes, even though he obviously would benefit considerably from their execution. Insofar as a limit order placed at time t fails to be executed, we assume that the portfolio manager cancels it immediately, only to replace it with a (possibly) updated quote at time $t + dt$. Concordantly, the total *cash position* (the bank holding, B_t) of the portfolio manager obeys the jump formula

$$\begin{aligned} dB_t = &\mathbf{1}_{\{Q_t \geq \Delta_t\}}[(S_t + \delta_t^-)dL_t^- + (S_t - \Upsilon_t)dM_{0t}^-] \\ &- \mathbf{1}_{\{Q_t < \Delta_t\}}[(S_t - \delta_t^+)dL_t^+ + (S_t + \Upsilon_t)dM_{0t}^+], \end{aligned} \tag{5.3}$$

subject to the initial condition $B_0 = b_0$, where

$$M_{0t}^+ = \sum_{k=1} \mathbf{1}\{\tau_k^+ \leq t\}$$

codifies the number of buy market orders placed by the investor over up till time t at the *stopping times* $\{\tau_i^+ \in \mathbb{R}^+, i \in \mathbb{N}^+ | 0 < \tau_1^+ < \tau_2^+ < ... < t\}$, and

[5] This assumption finds its intellectual roots in the seminal work by Avellaneda and Stoikov [6].

$$M_{0t}^{-} = \sum_{k=1} \mathbf{1}\{\tau_k^{-} \leq t\}$$

captures the portfolio manager's sell market orders, where $\{\tau_i^{-} \in \mathbb{R}^{+}, i \in \mathbb{N}^{+} | 0 < \tau_1^{-} < \tau_2^{-} < ... < t\}$ are the stopping times at which sell market orders are placed.

At maturity we imagine that one of the following scenarios obtains: if the options are out of the money, the portfolio manager immediately liquidates his inventory using market orders. On the other hand, if the options are at/in the money the option holders exercise their right to buy the stock for the strike price: insofar as there is a mismatch between the portfolio managers inventory and the $\mathfrak{N}$ stocks due for delivery, he acquires/liquidates the difference, again using market orders. There is of course no a priori reason why we should opt for this terminal condition; indeed, it might seem more opportune go for for an optimal trade execution strategy à la Almgren-Chriss [4] in the manner proposed by Cartea and Jaimungal [14] at least if $S_T < K$. This is an obvious topic for future research.

Finally, suppose the portfolio manager derives *linear utility* from his level of *financial wealth*, but incurs a quadratic lifetime penalisation from any deviation from the target hedge portfolio. The control problem to be solved can thus be stated as follows

$$\begin{aligned} V(t,b,s,q) = \sup_{\{\{\delta_s^{\pm}\}_{s\in[t,T]},\tau^{\pm}\}\in\mathscr{A}(t,b,s,q)} \mathbb{E}_{t,b,s,q}\Bigg[& B_T + \mathbf{1}_{\{S_T<K\}} Q_T (S_T - \Upsilon_T) \\ & + \mathbf{1}_{\{S_T\geq K\}} \mathbf{1}_{\{Q_T\geq\mathfrak{N}\}} [\mathfrak{N}K + (Q_T - \mathfrak{N})(S_T - \Upsilon_T)] \\ & + \mathbf{1}_{\{S_T\geq K\}} \mathbf{1}_{\{Q_T<\mathfrak{N}\}} [\mathfrak{N}K - (\mathfrak{N} - Q_T)(S_T + \Upsilon_T)] \\ & - \eta \int_t^T (Q_u - \Delta_u)^2 du \Bigg], \end{aligned} \tag{5.4}$$

where $V : [0,T] \times \mathbb{R} \times \sigma_m\mathbb{Z} \times \mathbb{Z} \mapsto \mathbb{R}$, $\mathbb{E}_{t,b,s,q}[\cdot] \equiv \mathbb{E}[\cdot|\mathscr{F}_t]$ with $(B_t, S_t, Q_t) = (b,s,q)$, and the supremum runs over all admissible control functions: i.e. all non-negative $\mathscr{F}_t$-predictable limit order spreads and all $\mathscr{F}_t$-stopping times bounded above by T. Finally, $\eta \in \mathbb{R}^{+}$ is a parameter which captures the portfolio managers "readiness" to depart from the desired hedge strategy Δ (clearly, the greater the η, the less prone the portfolio manager will be to depart from the prescribed strategy). In practical terms, η may be seen as a function of compliance with external (regulatory) risk measures, as well as internal (company specific) risk management. Notice though, that analogous to more traditional risk aversion parameters, there is an inexorable nebulosity wrapped around this construct: whether real world risk preferences can be accurately mapped to this singular parameter, and if it can be done with a reasonable degree of empirical accuracy is at best questionable.[6]

[6] Tentatively, one might try to Monte Carlo simulate portfolio returns under various η specifications and study the associated portfolio return distributions.

5.3 The Question of the Δ

The question of an optimal hedge strategy for jump processes in a market with friction is one of considerable complexity which we shall pass over in silence. As suggested above, our main concern here lies with the duality offered by the limit order book trading strategies; hence, the Δ we will consider is largely illustrative in nature. Specifically, we shall suppose that the portfolio manager aims to track a hedge strategy as though he were trading in a driftless *Bachelierian* (arithmetic Brownian motion) economy, i.e. as though the market were frictionless with price dynamics[7]

$$dS_t = \sigma_t dW_t, \tag{5.5}$$

where W_t is a Wiener process and $\sigma : [0,T] \mapsto \mathbb{R}^+$ is a deterministic function. Using a standard no-arbitrage argument one may readily show that strike K maturity T call options should be priced according to

$$C_t^{K,T} = (S_t - K)\Phi(\delta_t) + \Sigma_t \phi(\delta_t), \tag{5.6}$$

where

$$\delta_t \equiv \delta(t, S_t) = \Sigma_t^{-1}(S_t - K), \quad \text{and} \quad \Sigma_t \equiv \sqrt{\int_t^T \sigma_u^2 du},$$

and we have introduced the usual functions: $\Phi(\cdot)$ as the standard normal cdf, and $\phi(\cdot)$ as the standard normal pdf.[8] Hence, from the net portfolio position $B_t + \Delta_t S_t - \mathfrak{N} C_t$, to hedge a short position in $\mathfrak{N}$ such call options, one should hold $\Delta_t = \mathfrak{N}\partial_S C_t$ or, equivalently,

$$\Delta_t = \mathfrak{N}\Phi(\delta_t), \tag{5.7}$$

units of the underlying asset, where we have used the standard identity $\phi'(x) = -x\phi(x)$. Viewed as a function of (t, S_t) we note that Δ_t has the obvious properties that $\Delta \to \mathfrak{N}$ when the options are deep in the money ($S_t \gg K$) and $\Delta \to 0$ when the options are deep out of the money ($S_t \ll K$). As $t \to T$ the transition between these two extremes becomes increasingly steep in it its rise, ultimately converging towards the step function $\mathfrak{N}\mathbf{1}\{S_T \geq K\}$ at expiry. Thus, the Δ of at the money calls is exceedingly sensitive to fluctuations in the price process near expiry, potentially resulting in the acquisition or decumulation of a large amount of shares in a short span of time. An illustration of this Δ is provided in figure 5.1 for constant parameters.

Again, we emphasise that this choice largely is to get the ball rolling: the reader is encouraged to experiment with alternative specifications. However, forbye the neglected

---

[7] Louis Bachelier is widely credited as the father of mathematical finance. In his doctoral dissertation *The Theory of Speculation* [7] Bachelier introduced Brownian motion in the modelling of stock prices. Specifically, he assumed arithmetic brownian motion with constant parameters: $dS_t = \mu dt + \sigma dW_t$.

[8] Specifically, from the martingale condition $C_t = \mathbb{E}_t[\max\{S_T - K, 0\}]$ where $S_T = S_T(Z) = S_t + \int_t^T \sigma_u dW_u = S_t + \Sigma_t Z$ with $Z \sim N(0,1)$ (to see the last equality it is helpful to recall the Itō isometry). Thus, upon evaluating the integral $C_t = \int_{\mathbb{R}} (S_T(z) - K)\mathbf{1}\{S_T(z) \geq K\}\phi(z)dz$ we get the desired result.

issue of market friction, we also note that the hedge ratio (5.7) is not altogether groundless; specifically, the jump dynamics (5.1) formally converges in distribution to a Bachelierian dynamics in the event that we let the jump rate intensity tend to infinity. To see this, we need the following proposition:

Proposition 5.1. *Let* $L_t = L_t^1 - L_t^2$, *where* $L_t^i \equiv L^i((0,t]\times\mathbb{A}) = \int_0^t\int_{\mathbb{A}} zJ_i(dt\times dz)$ *for* $i=1,2$. *Furthermore, let us introduce the function* $\zeta : [0,T]\mapsto\mathbb{R}^+$ *where*

$$\zeta_t \equiv \left(2\int_0^t \mathbb{E}[Z_s^2]\gamma_s ds\right)^{-1/2}, \tag{5.8}$$

then we have the following convergence in distribution as $\gamma_s\to\infty$*:*

$$\zeta_t L_t \xrightarrow{d} N(0,1), \tag{5.9}$$

where $N(0,1)$ *is the standard normal distribution.*

Proof. By Levy's Continuity Theorem, it suffices to show that the characteristic function $\zeta_t L_t$ converges pointwise to the characteristic function of $N(0,1)$.[9] To this end, suppose we partition the time interval into $n+1$ segments: $\mathscr{T}_n = \{t_i | t_0 = 0, t_1 = t_0 + \Delta t_0, t_2 = t_1 + \Delta t_1, ..., t_{n+1} = t = t_n + \Delta t_n\}$, where $\max_i \Delta t_i \to 0$ as $n\to\infty$. Similarly, let $\mathscr{A}_m = \{\Delta\mathbb{A}_1, \Delta\mathbb{A}_2, ..., \Delta\mathbb{A}_m | \cup_{j=1}^m \Delta\mathbb{A}_j = \mathbb{A}\}$ be a partition of the mark space over disjoint subsets.[10] From the elementary properties of Poisson random measures it follows that $J_i^{sq} \equiv J((t_s, t_s+\Delta t_s]\times\Delta\mathbb{A}_q)$ and $J_i^{ur} \equiv J((t_u, t_u+\Delta t_u]\times\Delta\mathbb{A}_r)$ are independent provided that $q\neq r$. Furthermore,

$$\mathbb{P}\left\{J_i^{sq} = k\right\} = e^{-\mu_i^{sq}}\frac{(\mu_i^{sq})^k}{k!},$$

where we have defined the discrete intensity measure $\mu_i^{sq} \equiv \mu_i((t_s, t_s+\Delta t_s], \Delta\mathbb{A}_q) = \gamma_s F_s(\Delta\mathbb{A}_q)\Delta t_s$. Thus, the characteristic function of $\zeta_t L_t$, i.e. $\varphi_{\zeta L}(a) \equiv \mathbb{E}[\exp\{ia\zeta_t L_t\}]$ where $i=\sqrt{-1}$ and $a\in\mathbb{R}$, can we decomposed as $\varphi_{\zeta L^1}(a)\varphi_{\zeta L^2}(-a)$ (by the independence of L^1 and L^2) whence

$$\begin{aligned}
\varphi_{\zeta L}(a) &= \mathbb{E}\left[\exp\left\{ia\zeta_t\int_0^t\int_{\mathbb{A}} zJ_1(ds\times dz)\right\}\right]\mathbb{E}\left[\exp\left\{-ia\zeta_t\int_0^t\int_{\mathbb{A}} zJ_2(ds\times dz)\right\}\right]\\
&= \lim_{m,n\to\infty}\mathbb{E}\left[\exp\left\{ia\zeta_{n+1}\sum_{s=0}^n\sum_{q=1}^m z_q J_1^{sq}\right\}\right]\mathbb{E}\left[\exp\left\{-ia\zeta_{n+1}\sum_{u=0}^n\sum_{r=1}^m z_r J_2^{ur}\right\}\right]
\end{aligned}$$

[9] Recall: if $\{X_n\}_{n\in\mathbb{N}^+}$ and X are random vectors in $\mathbb{R}^k$, and $\{\varphi_n\}_{n\in\mathbb{N}^+}$ and φ_X are the associated characteristic functions, then (I) X_n converges in distribution to X iff $\varphi_{X_n}(s)\to\varphi_X(s)$ for all $s\in\mathbb{R}^k$, and (II) if $\varphi_{X_n}(s)\to\varphi(s)$ pointwise for all $s\in\mathbb{R}^k$, and if φ is continuous on 0, then φ is the characteristic function of X and X_n converges in distribution to X.

[10] This discretisation idea is inspired by Hanson [28].

$$
\begin{aligned}
&= \lim_{m,n\to\infty} \prod_{s,u=0}^{n} \prod_{q,r=1}^{m} \mathbb{E}\left[\exp\{ia\zeta_{n+1}z_q J_1^{sq}\}\right] \mathbb{E}\left[\exp\{-ia\zeta_{n+1}z_r J_2^{ur}\}\right] \\
&= \lim_{m,n\to\infty} \prod_{s,u=0}^{n} \prod_{q,r=1}^{m} \sum_{k_1=0}^{\infty} \mathbb{P}\{J_i^{sq} = k_1\} \mathbb{E}\left[\exp\{ia\zeta_{n+1}z_q J_1^{sq}\} | J_1^{sq} = k_1\right] \cdot \\
&\qquad\qquad \sum_{k_2=0}^{\infty} \mathbb{P}\{J_i^{ur} = k_2\} \mathbb{E}\left[\exp\{-ia\zeta_{n+1}z_r J_2^{up}\} | J_2^{up} = k_2\right] \\
&= \lim_{m,n\to\infty} \prod_{s,u=0}^{n} \prod_{q,r=1}^{m} \sum_{k_1=0}^{\infty} e^{-\mu_1^{sq}} \frac{(\mu_1^{sq})^{k_1}}{k_1!} e^{ia\zeta_{n+1}z_q k_1} \sum_{k_2=0}^{\infty} e^{-\mu_2^{ur}} \frac{(\mu_2^{ur})^{k_2}}{k_2!} e^{-ia\zeta_{n+1}z_r k_2} \\
&= \lim_{m,n\to\infty} \prod_{s,u=0}^{n} \prod_{q,r=1}^{m} \exp\{\mu_1^{sq}(e^{ia\zeta_{n+1}z_q} - 1)\} \exp\{\mu_2^{ur}(e^{-ia\zeta_{n+1}z_r} - 1)\} \\
&= \lim_{m,n\to\infty} \exp\left\{ \sum_{s=0}^{n} \sum_{q=1}^{m} \mu_1^{sq}(e^{ia\zeta_{n+1}z_q} - 1) \right\} \exp\left\{ \sum_{u=0}^{n} \sum_{r=1}^{m} \mu_2^{ur}(e^{-ia\zeta_{n+1}z_r} - 1) \right\}, \\
&= \exp\left\{ \int_0^t \int_{\mathbb{A}} (e^{ia\zeta_t z} - 1)\gamma_s F_s(dz) ds \right\} \exp\left\{ \int_0^t \int_{\mathbb{A}} (e^{-ia\zeta_t z} - 1)\gamma_s F_s(dz) ds \right\}, \\
&= \exp\left\{ \int_0^t (\mathbb{E}[e^{ia\zeta_t Z_s}] + \mathbb{E}[e^{ia\zeta_t Z_s}]^* - 2)\gamma_s ds \right\},
\end{aligned}
$$

where $*$ designates the complex conjugate. Here, the fourth equality uses the *law of total expectation*, while the sixth equality uses the Taylor expansion of the exponential function. Finally, from the expansion

$$
\mathbb{E}[e^{ia\zeta_t Z_s}] = 1 + ia\zeta_t \mathbb{E}[Z_s^1] - \tfrac{1}{2}a^2\zeta_t^2 \mathbb{E}[Z_s^2] - \tfrac{1}{6}ia^3\zeta_t^3 \mathbb{E}[Z_s^3] + \mathscr{O}(a^4\zeta_t^4 \mathbb{E}[Z_s^4]),
$$

we find after internal cancellation that

$$
\varphi_{\zeta L}(a) = \exp\left\{ -a^2\zeta_t^2 \int_0^t \mathbb{E}[Z_s^2]\gamma_s ds + \mathscr{O}\left(a^4\zeta_t^4 \int_0^t \mathbb{E}[Z_s^4]\gamma_s ds\right) \right\}.
$$

Upon specifying ζ_t as in (5.8) and letting $\gamma_s \to \infty$ it follows that $\Phi_{\zeta L}(a) \to \exp\{-\frac{1}{2}s^2\}$, which is the characteristic function of the standard normal distribution. □

Hence, we readily deduce that

Corollary 5.1. *In the infinite intensity limit, the jump dynamics (5.1) converges in distribution to the Bachelier dynamics (5.5). In particular, for large γ_s we have that*

$$
\sigma_t \approx \sigma_m \sqrt{2\overline{\mathbb{E}[Z^2]\gamma}}, \tag{5.10}
$$

where the over-line designates the mean of the deterministic function over $[0,t]$.

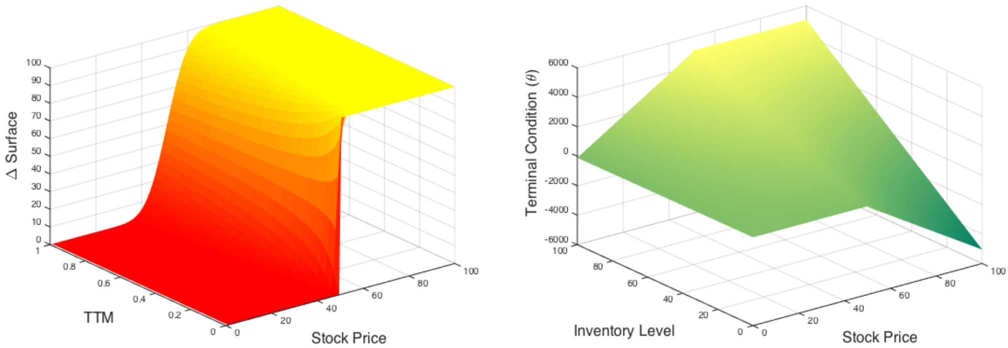

Fig. 5.1 Assume $N = 100$ and $K = 50.5$, with constant values for σ and Υ. (**Left:**) The targeted Bachelierian hedge strategy $\Delta(t, S_t) = \mathfrak{N}\Phi(\delta_t)$. Notice that the surface converges towards a step function at the expiry of the call options. (**Right:**) The terminal condition of the control problem disregarding the bank account. Observe that at the zeroth level of inventory we have the classic "hockey stick" pay-off structure associated with a short option position. At the other end, with a full level of inventory $Q = \mathfrak{N}$, the short option position is perfectly hedged which is reflected in the constant in-the-money pay-off.

Proof. From (5.9) we have for large γ_s the approximation $\zeta_t L_t \sim N(0,1)$. Hence,

$$\sigma_m \zeta_t L_t \sim (\sigma_m/\sqrt{t})N(0,t).$$

Rearranging this, the result follows. □

Remark 5.3. The results presented here may be viewed as an abstract generalisation of the textbook result that $\tilde{N}_t/\sqrt{\gamma}$ converges in distribution to $N(0,t)$ as $\gamma \to \infty$, where $\tilde{N}_t \equiv N_t - \gamma t$ and $N_t \sim \text{Pois}(\gamma t)$ cf. Cont and Tankov [5].[11]

5.4 The Hamilton-Jacobi-Bellman Formulation

Following standard results in Pham [18] and Fleming & Soner [21], it follows that the optimal value function given in (5.4) satisfies the Hamilton-Jacobi-Bellman quasi variational inequality (HJB QVI)

$$\begin{aligned} 0 = \max \Bigg\{ & \partial_t V(t,b,s,q) + \gamma_t \int_{\mathbb{A}} [V(t,b,s+\sigma_m z,q) - 2V(t,b,s,q) + V(t,b,s-\sigma_m z,q)] F_t(dz) \\ & + \sup_{\delta_t^- \in \mathbb{R}_+} \lambda_t e^{-\kappa_t \delta_t^-} [V(t,b+(s+\delta_t^-),s,q-1) - V(t,b,s,q)] \mathbf{1}_{\{q \geq \Delta_t\}} \end{aligned}$$

[11] We note that Cont and Tankov erroneously have omitted the square root in the first edition of their book.

$$
\begin{aligned}
&+\sup_{\delta_t^+\in\mathbb{R}_+} \lambda_t e^{-\kappa_t\delta_t^+}[V(t,b-(s-\delta_t^+),s,q+1)-V(t,b,s,q)]\mathbf{1}_{\{q<\Delta_t\}}-\eta(q-\Delta_t)^2\\
&\;;\;[V(t,b+(s-\Upsilon_t),s,q-1)-V(t,b,s,q)]\mathbf{1}_{\{q\geq\Delta_t\}}\\
&+[V(t,b-(s+\Upsilon_t),s,q+1)-V(t,b,s,q)]\mathbf{1}_{\{q<\Delta_t\}}\Bigg\},
\end{aligned}
\tag{5.11}
$$

subject to the terminal condition

$$
\begin{aligned}
V(T,b,s,q) = B_T &+ \mathbf{1}_{\{s<K\}}q(s-\Upsilon_T)+\mathbf{1}_{\{s\geq K\}}\mathbf{1}_{\{q\geq\mathfrak{N}\}}[\mathfrak{N}K+(q-\mathfrak{N})(s-\Upsilon_T)]\\
&+\mathbf{1}_{\{s\geq K\}}\mathbf{1}_{\{q<\mathfrak{N}\}}[\mathfrak{N}K-(\mathfrak{N}-q)(s+\Upsilon_T)].
\end{aligned}
\tag{5.12}
$$

A graphical representation of the latter is provided in figure 5.1 assuming $B_T = 0$. Notice that the step function nature of the terminal hedge hedge, Δ_T, clearly is reflected in the pay-off here: specifically, the value surface is constant along $Q = 0$ for $S_T < K$ and along $Q = \mathfrak{N}$ for $S \geq K$.

Sketch Proof. It is instructive to offer a *heuristic* argument for variational inequality (5.11). To this end let us focus on limit orders in the event the portfolio manager has a *sell incentive*, i.e. his inventory surpasses the targeted hedge ratio $Q_t \geq \Delta_t$. Using the *dynamic programming principle* as our vantage point it follows that

$$
0 = -\eta(q-\Delta_t)^2 dt + \sup_{\delta_t^-\in\mathbb{R}_+} \mathbb{E}_{t,b,s,q}[dV(t,b,s,q)]. \tag{5.13}
$$

cf. the appendix. Furthermore, using Itō's lemma for marked point processes [5], [28], [43], we have

$$
\begin{aligned}
dV(t,b,s,q) = \partial_t V(t,b,s,q)dt &+ \textstyle\int_{\mathbb{A}}[V(t,b,s+\sigma_m z,q)-V(t,b,s,q)]J_1(dt\times dz)\\
&+\textstyle\int_{\mathbb{A}}[V(t,b,s-\sigma_m z,q)-V(t,b,s,q)]J_2(dt\times dz)\\
&+[V(t,b+(s+\delta_t^-),s,q-1)-V(t,b,s,q)]dL_t^-,
\end{aligned}
\tag{5.14}
$$

where the first two jump terms are induced by jumps in the underlying process, (5.1), whilst the third jump in the bank/inventory is induced by a limit sell order being met. Recalling that $\mathbb{E}[J_1(dt\times dz)] = \mathbb{E}[J_2(dt\times dz)] = \gamma_t F_t(dz)dt$ and $\mathbb{E}[dL_t^-] = \lambda_t e^{-\kappa_t\delta_t^-}dt$ we find upon combing (5.13) and (5.14) that V satisfies the following HJB equation

$$
\begin{aligned}
0 = \partial_t V(t,b,s,q) &+ \gamma_t\int_{\mathbb{A}}[V(t,b,s+\sigma_m z,q)-2V(t,b,s,q)+V(t,b,s-\sigma_m z,q)]F_t(dz)\\
&+\sup_{\delta_t^-\in\mathbb{R}_+}\lambda_t e^{-\kappa_t\delta_t^-}[V(t,b+(s+\delta_t^-),s,q-1)-V(t,b,s,q)]-\eta(q-\Delta_t)^2.
\end{aligned}
\tag{5.15}
$$

We now add the possibility of market orders: if the inventory level Q_t is sufficiently greater than Δ_t the portfolio manager is incentivised to effectuate a guaranteed immediate sell through a market order. Thus (5.15) no longer applies (the righthand being strictly *less* than zero). Instead the value function jumps in accordance with the transaction costs associated with a sell market order:

$$0 = V(t, b + (s - \Upsilon_t), s, q - 1) - V(t, b, s, q). \tag{5.16}$$

On the other hand, if the portfolio manager is contended with limit orders, the righthand side of (5.16) should be strictly less than zero, being, as it were, the suboptimal strategy. Jointly, these considerations lead to the variational inequality (5.11) in the event $Q_t \geq \Delta_t$. Indeed, we may complete the picture by forming an analogous argument for the case where the portfolio manager is incentivised to acquire the underlying asset ($Q_t < \Delta_t$). □

Whilst searching for a closed form solution to (5.11) appears like an exercise in futility, it is immediately obvious that the problem at least offers a reduction in dimensionality. Specifically, since the money account B enters the terminal condition (5.12) linearly, and since the portfolio manager derives linear utility from his financial wealth, we have the ansatz

$$V(t, b, s, q) = b + \theta(t, s, q), \tag{5.17}$$

where $\theta : [0, T] \times \sigma_m \mathbb{Z} \times \mathbb{Z} \mapsto \mathbb{R}$. Hence, the HJB QVI reduces to

$$\begin{aligned}
0 = \max \Bigg\{ & \partial_t \theta(t, s, q) + \gamma_t \int_{\mathbb{A}} [\theta(t, s + \sigma_m z, q) - 2\theta(t, s, q) + \theta(t, s - \sigma_m z, q)] F_t(dz) \\
& + \sup_{\delta_t^- \in \mathbb{R}_+} \lambda_t e^{-\kappa_t \delta_t^-} [s + \delta_t^- + \theta(t, s, q - 1) - \theta(t, s, q)] \mathbf{1}_{\{q \geq \Delta_t\}} \\
& + \sup_{\delta_t^+ \in \mathbb{R}_+} \lambda_t e^{-\kappa_t \delta_t^+} [-(s - \delta_t^+) + \theta(t, s, q + 1) - \theta(t, s, q)] \mathbf{1}_{\{q < \Delta_t\}} - \eta (q - \Delta_t)^2 \\
& ;\ [s - \Upsilon_t + \theta(t, s, q - 1) - \theta(t, s, q)] \mathbf{1}_{\{q \geq \Delta_t\}} \\
& + [-(s + \Upsilon_t) + \theta(t, s, q + 1) - \theta(t, s, q)] \mathbf{1}_{\{q < \Delta_t\}} \Bigg\},
\end{aligned} \tag{5.18}$$

subject to the terminal condition

$$\begin{aligned}
\theta(T, s, q) = & \mathbf{1}_{\{s < K\}} q (s - \Upsilon_T) + \mathbf{1}_{\{s \geq K\}} \mathbf{1}_{\{q \geq \mathfrak{N}\}} [\mathfrak{N} K + (q - \mathfrak{N})(s - \Upsilon_T)] \\
& + \mathbf{1}_{\{s \geq K\}} \mathbf{1}_{\{q < \mathfrak{N}\}} [\mathfrak{N} K - (\mathfrak{N} - q)(s + \Upsilon_T)].
\end{aligned} \tag{5.19}$$

Upon solving the first order conditions

$$\delta_t^{*\pm} = \underset{\delta_t^\pm \in \mathbb{R}_+}{\operatorname{argmax}}\ \lambda_t e^{-\kappa_t \delta_t^\pm} [\pm s + \delta_t^\pm + \theta(t, s, q \pm 1) - \theta(t, b, s, q)], \tag{5.20}$$

(where the sign $\pm$ is chosen uniformly) and substituting these back in to (5.18) we obtain the following key result

Theorem 5.1. *The solution to the optimal value problem (5.4) is given by equation (5.17) where θ is a function which satisfies the variational inequality*

$$
\begin{aligned}
0 = \max \Bigg\{ & \partial_t \theta(t,s,q) + \gamma_t \int_{\mathbb{A}} [\theta(t,s+\sigma_m z,q) - 2\theta(t,s,q) + \theta(t,s-\sigma_m z,q)] F_t(dz) \\
& + \frac{\lambda_t}{\kappa_t} \left[e^{-\kappa_t \delta_t^{*-}} \mathbf{1}_{\{q \geq \Delta_t\}} + e^{-\kappa_t \delta_t^{*+}} \mathbf{1}_{\{q < \Delta_t\}} \right] - \eta (q - \Delta_t)^2 \\
& ; \ [s - \Upsilon_t + \theta(t,s,q-1) - \theta(t,s,q)] \mathbf{1}_{\{q \geq \Delta_t\}} \\
& + [-(s + \Upsilon_t) + \theta(t,s,q+1) - \theta(t,s,q)] \mathbf{1}_{\{q < \Delta_t\}} \Bigg\},
\end{aligned} \tag{5.21}
$$

where the optimal limit order controls are given by

$$
\delta_t^{*\pm} = \kappa_t^{-1} \pm s + \theta(t,s,q) - \theta(t,s,q \pm 1), \tag{5.22}
$$

and the terminal condition is of the form (5.19).

Remark 5.4. It is a well known empirical fact of limit order books that a subset of price fluctuations are of a transient nature, i.e. that the limit order mid-price has a certain proclivity towards reverting back to its former level after a jump. Recently, it has been argued that *squashed trawl processes* are opportune constructs to be used in the modelling of this so-called *temporary market impact* - see Shephard & Yang [44] or the trailblazing paper by Barndorff-Nielsen et al. [9]. In this note, we briefly wish to indicate that the framework developed here readily can be adapted to squashed trawls, assuming that the associated parameters are known to the portfolio manager. To this end, consider a set-up analogous to [44]; specifically, let the dynamics of the mid-price be

$$
dS_t = \sigma_m \int_{\mathbb{A}} \int_0^1 z(J_1(dt \times dz \times dx) - J_2(dt \times dz \times dx)), \tag{5.23}
$$

where X is an independent random variable which codifies the lifetime of a jump (duration of particle in the marked point process). We assume that $X \sim U[0,1]$ such that $\mathbb{E}[J_i(dt \times dz \times dx)] = \gamma_t F_t(dz) dx$ for $i = 1,2$. Furthermore, if x is a random variate below some threshold $b \in (0,1)$, then the associated point is remembered "in perpetuity" analogous to ordinary marked point processes on the real line. On the other hand, if x is a random variate in excess of b, then it is endowed with the finite lifetime $|d^{-1}(x)|$, where d is a monotonically increasing function which satisfies certain regularity conditions. Thus, whenever a particle p reaches its expiry a drop of equal but opposite magnitude to the one originally induced by p will be witnessed in the price process S. In steady state, it is intuitively obvious that the expected influx of particles (*births*) in the $(b,1]$ region precisely is balanced by the expected outflux of particles (*deaths*) from the same region. Transient memory particles thus have zero net expected contribution to changes in the value function (5.4). Thus, the HJB QVI associated with the midprice dynamics (5.23) is virtually equivalent to (5.11) only the intensity rate γ_t is suitably scaled to $\gamma_t b$ to restrict the scope to the permanent market impact.

5.5 Towards a Numerical Solution

Qua the dimensionality reduction offered by (5.17) we have a three-variable variational inequality, which with limited computational expenditure can be solved using an *explicit finite difference scheme*. Here we will elucidate how to set up the problem and demonstrate its convergence modulo the verisimilitude of a so-called comparison principle. To this end, we invoke a well-known viscosity convergence result demonstrated by Barles and Souganidis [8].[12] For highly readable accounts of finite difference methods in stochastic control theory we refer the reader to Tourin [24] and Forsyth and Labahn [24]. For a cursory overview of viscosity solutions we refer the reader to the appendix.

Now, following standard procedure, to approximate (5.18) in the finite difference sense, we introduce the bounded mesh

$$\begin{aligned}\mathscr{T}^{\Delta t} \times \mathscr{R}^{\sigma_m} \times \mathscr{R}^1 = \{0, \Delta t, ..., n\Delta t, ...N\Delta t = T\} \times \{s_{\min} = 0, \sigma_m, ..., i\sigma_m, ..., I\sigma_m = s_{\max}\} \\ \times \{q_{\min}, q_{\min} + 1, ..., q_{\min} + j, ..., q_{\min} + J = q_{\max}\},\end{aligned}$$

in lieu of the full state space $[0, T] \times \sigma_m\mathbb{Z} \times \mathbb{Z}$. Here, $s_{\min}, s_{\max} \in \sigma_m\mathbb{Z}$, and $q_{\min}, q_{\max} \in \mathbb{Z}$ are artificially imposed lower and upper boundaries which encapsulate the solution region of interest. Notice that the only discretisation which takes place is along the time axis (the stock and inventory dimensions are manifestly already discrete in the original problem). Furthermore, to handle the integral in (5.18) suppose that the number of ticks that the mid-price can jump at any given time is bounded from above by the constant K where $K \in \mathbb{N}^+$. Specifically, let $\int_{\mathbb{A}} F_s(dz) = \sum_{k=1}^{K} \mathbb{P}_s\{Z = k\} = 1$ and $K \ll I$ such that the price process surely stays in the mesh unless $S_t < K\sigma_m$ or $S_t > (I - K)\sigma_m$. To block the possibility that the price process jumps out of the mesh near the boundaries, a suitable rescaling of the probability weights are performed in those regions. Finally, the explicit approximation to the HJB QVI may then be stated as

$$\vartheta_{i,j}^n = \mathscr{G}[\vartheta^{n+1}] \equiv \max\{\mathscr{G}_{\Delta t}[\vartheta^{n+1}]\ ;\ \mathscr{G}_0[\vartheta^{n+1}]\}, \tag{5.24}$$

for $K \le i \le I - K$ and $q_{\min} + 1 \le j \le q_{\max} - 1$ where $\vartheta_{i,j}^n \equiv \theta_{\Delta t}(n\Delta t, i\sigma_m, q_{\min} + j)$ is a finite difference approximation to the dimensionally reduced optimal value function $\theta(n\Delta t, i\sigma_m, q_{\min} + j)$, $\vartheta^{n+1} \equiv (\vartheta_{i,j}^{n+1}, \vartheta_{i,j+1}^{n+1}, \vartheta_{i,j-1}^{n+1}, \{\vartheta_{i\pm s,j}^{n+1}\}_{s=1}^{K})^\intercal \in \mathbb{R}^{2K+3}$, and we have defined the finite difference operators

[12] We remark that this is not the only way in which convergence of HJB numerical schemes may be demonstrated. For a probabilistic proof we refer the reader to Kushner and Dupuis [31], whose seminal work on Markov chain approximations of the dynamic programming principle indubitably will resonate with some readers. For a viscosity approach to Markov chain convergence see Fleming and Soner [21].

$$\begin{aligned}
\mathscr{G}_{\Delta t}[\vartheta^{n+1}] \equiv \vartheta^{n+1}_{i,j} + \Delta t \Bigg\{ & \gamma_{n+1} \sum_{k=1}^{K} \mathbb{P}_{n+1}\{Z=k\}[\vartheta^{n+1}_{i+k,j} - 2\vartheta^{n+1}_{i,j} + \vartheta^{n+1}_{i-k,j}] \\
& + \sup_{\delta^-_{n+1}\in\mathbb{R}_+} \lambda_{n+1} e^{-\kappa_{n+1}\delta^-_{n+1}} [i\sigma_m + \delta^-_{n+1} + \vartheta^{n+1}_{i,j-1} - \vartheta^{n+1}_{i,j}] \mathbf{1}_{\{q_{\min}+j\geq\Delta_{n+1}\}} \\
& + \sup_{\delta^+_{n+1}\in\mathbb{R}_+} \lambda_{n+1} e^{-\kappa_{n+1}\delta^+_{n+1}} [-(i\sigma_m - \delta^+_{n+1}) + \vartheta^{n+1}_{i,j+1} - \vartheta^{n+1}_{i,j}] \mathbf{1}_{\{q_{\min}+j<\Delta_{n+1}\}} \\
& - \eta(q_{\min}+j-\Delta_{n+1})^2 \Bigg\}, \\
\mathscr{G}_0[\vartheta^{n+1}] \equiv [i\sigma_m - \Upsilon_{n+1} + \vartheta^{n+1}_{i,j-1}] & \mathbf{1}_{\{q_{\min}+j\geq\Delta_{n+1}\}} - [i\sigma_m + \Upsilon_{n+1} - \vartheta^{n+1}_{i,j+1}] \mathbf{1}_{\{q_{\min}+j<\Delta_{n+1}\}},
\end{aligned}$$

with analogous expressions near the boundaries. For completeness, note that the terminal condition of (5.24) takes on the form

$$\begin{aligned}
\vartheta^N_{i,j} = \mathbf{1}_{\{i\sigma_m<K\}}(q_{\min}+j)(i\,\sigma_m - \Upsilon_N) + \mathbf{1}_{\{i\,\sigma_m\geq K\}} \mathbf{1}_{\{q_{\min}+j\geq\mathfrak{N}\}}[\mathfrak{N}K + (q_{\min}+j-\mathfrak{N})\cdot \\
(i\sigma_m - \Upsilon_N)] + \mathbf{1}_{\{i\sigma_m\geq K\}} \mathbf{1}_{\{q_{\min}+j<\mathfrak{N}\}}[\mathfrak{N}K - (\mathfrak{N} - (q_{\min}+j))(i\sigma_m + \Upsilon_N)],
\end{aligned} \tag{5.25}$$

whilst the first order conditions become

$$\delta^{*\pm}_{n+1} = \kappa^{-1}_{n+1} \pm i\sigma_m + \vartheta^{n+1}_{i,j} - \vartheta^{n+1}_{i,j\pm1}.$$

Solving the control problem is thus algorithmically comparable to computing the no-arbitrage price of an American put option using finite difference methods: starting from the terminal condition and moving incrementally backwards in time, we must at each grid node decide whether a limit or a market order optimises our expected utility.

Proposition 5.2. *Assuming $2\gamma_n\Delta t \leq 1$ for $n = 1,2,...,N$, the numerical scheme (5.24), (5.25) is (i) monotone, (ii) stable and (iii) consistent. Thus, following Barles and Souganidis' Theorem 2.1 [8], if the scheme is also satisfying a comparison principle, then it converges locally uniformly to the unique viscosity solution of (5.18), (5.19).*

Proof. Following [8] and [21] we first recall the following definitions, suitably adapted to the problem at hand

1. **(Monotonicity)**: For all $\varepsilon \in \mathbb{R}_+$, and for all unit vectors $\{\hat{\mathbf{e}}_i\}_{i=1}^{2K+3}$ we require that

$$\mathscr{G}[\vartheta^{n+1}] \leq \mathscr{G}[\vartheta^{n+1} + \varepsilon\hat{\mathbf{e}}_i], \tag{5.26}$$

   where $[\hat{\mathbf{e}}_i]_j = \delta_{ij}$ where δ_{ij} is the Kronecker delta.
2. **(Stability)**:

$$||\vartheta^n||_\infty \leq C, \tag{5.27}$$

for some constant C independent of Δt, as $\Delta t \to 0$.[13]

3. **(Consistency)**: Finally, we require that the system satisfies the basic consistency condition

$$\lim_{\Delta t\to 0}\left|\frac{1}{\Delta t}\left(\mathscr{G}[\vartheta^{n+1}]-\vartheta^{n}_{i,j}\right)-\mathscr{H}\left[\theta((n+1)\Delta t, i\sigma_m, q_{\min}+j)\right]\right|=0, \tag{5.28}$$

where $\mathscr{H}[\cdot]$ is the HJB QVI, (5.18), centered at the coordinate

$$(t,s,q)=((n+1)\Delta t, i\sigma_m, q_{\min}+j).$$

We will work through these items systematically, focussing on the case where the limit order equation applies (the market order equation trivially satisfies 1, 2, and 3):

1. **(Monotonicity)**: This is trivially true for all elements of ϑ^{n+1} which only have a positive coefficient instantiation in $\mathscr{G}_{\Delta t}[\vartheta^{n+1}]$. In fact, the only non-trivial case is $\theta^{n+1}_{i,j}$. Specifically, upon eliminating all superfluous terms, the monotonicity condition requires:

$$\begin{aligned}
0 &\le \varepsilon+\Delta t\Big\{-2\gamma_{n+1}\varepsilon+\sup_{\delta^-_{n+1}\in\mathbb{R}_+}[-\varepsilon\lambda_{n+1}e^{-\kappa_{n+1}\delta^-_{n+1}}]\mathbf{1}_{\{q_{\min}+j\ge\Delta_{n+1}\}}\\
&\quad+\sup_{\delta^+_{n+1}\in\mathbb{R}_+}[-\varepsilon\lambda_{n+1}e^{-\kappa_{n+1}\delta^+_{n+1}}]\mathbf{1}_{\{q_{\min}+j<\Delta_{n+1}\}}\Big\}\\
&=\varepsilon-\Delta t\Big\{2\gamma_{n+1}\varepsilon+\varepsilon\lambda_{n+1}\inf_{\delta^-_{n+1}\in\mathbb{R}_+}[e^{-\kappa_{n+1}\delta^-_{n+1}}]\mathbf{1}_{\{q_{\min}+j\ge\Delta_{n+1}\}}\\
&\quad+\varepsilon\lambda_{n+1}\inf_{\delta^+_{n+1}\in\mathbb{R}_+}[e^{-\kappa_{n+1}\delta^+_{n+1}}]\mathbf{1}_{\{q_{\min}+j<\Delta_{n+1}\}}\Big\}.
\end{aligned}$$

   Since

$$\inf_{\delta^{\pm}_{n+1}} e^{-\kappa_{n+1}\delta^{\pm}_{n+1}}=0,$$

   we're left with the constraint

$$2\gamma_{n+1}\Delta t\le 1,\quad \text{for}\quad n=N-1,N-2,...,0. \tag{5.29}$$

   Indeed, numerical experiments corroborate that violating (5.29) leads to a failure in convergence of the algorithm. Notice that the inverse proportionality between γ and Δt in practice can lead to very fine grid spacings for realistic values of the jump rate intensity: hence, (5.29) significantly impedes the rapidity with which the algorithm can be executed.

2. **(Stability)**: Taking the L^∞-norm of the HJB equation and using the triangle inequality we obtain

[13] Per definition, $||\mathbf{x}||_\infty \equiv \max_i |x_i|$.

$$||\vartheta^n_{i,j}||_\infty \leq ||\vartheta^{n+1}_{i,j}||_\infty + \Delta t||\{...\}||_\infty,$$

where $\{...\}$ signifies the content of the curly brackets in the definition of $\mathscr{G}_{\Delta t}$. Setting $\Delta t \to 0$ and using the fact that the terminal condition $\vartheta^N_{i,j}$ by assumption is bounded from above, we obtain the desired result.

3. **(Consistency)**: This follows immediately upon substituting in the definitions of $\mathscr{G}_{\Delta t}$ and $\mathscr{H}$ in equation (5.28) and using the Taylor expansion

$$\vartheta^n_{i,j} = \vartheta^{n+1}_{i,j} - \Delta t \partial_t \vartheta^{n+1}_{i,j} + \mathscr{O}((\Delta t)^2).$$

To complete the proof we deploy similar arguments for the grid points near the boundaries (recall the rescaling of probabilities such as to avoid the price process leaving the grid). Since this effectively is a recapitulation of what has already been established we omit the details here. Finally, regarding the comparison principle, loosely put, what is required is that if θ_1 and θ_2 are two solutions of the HJB QVI with $\theta_1(T,s,q) \geq \theta_2(T,s,q)$ then we require $\theta_1(t,s,q) \geq \theta_2(t,s,q)$ for all t. For a more rigorous definition, see the viscosity references.

□

5.6 Example

5.6.1 The Compound Poisson Model

With a converging numerical scheme at hand, we are in a position to compute the optimal limit order book hedge strategy called for in question 5.1. For simplicity, we will assume that all parameters are constant in time. In particular, the dynamics of the mid-price is assumed to be the difference between two compound Poisson processes:

$$dS_t = \sigma_m(dY^1_t - dY^2_t),$$

with $S_0 = K$ and

$$Y^i_t = \sum_{j=1}^{N^i_t} Z^i_j,$$

for $i = 1,2$, where $\{N^1_t, N^2_t\}_{t\in[0,T]}$ are independent Poisson processes of common intensity γt, whilst the jump sizes $\{Z^i_j\}_{j\in\mathbb{N}^+}$ are i.i.d. random variables assumed to follow the geometric distribution[14]

$$\mathbb{P}\{Z^i_j = k\} = (1-p)^{k-1}p.$$

[14] Readers familiar with Lévy processes will notice that this essentially is a discrete symmetric version of the Kou model:

$$\nu(dz) = [p\lambda_+ e^{-\lambda_+ x}\mathbf{1}_{\{x>0\}} + (1-p)\lambda_- e^{-\lambda_- x}\mathbf{1}_{\{x<0\}}]dx,$$

where $\lambda_+, \lambda_- > 0$ and $p \in [0,1]$.

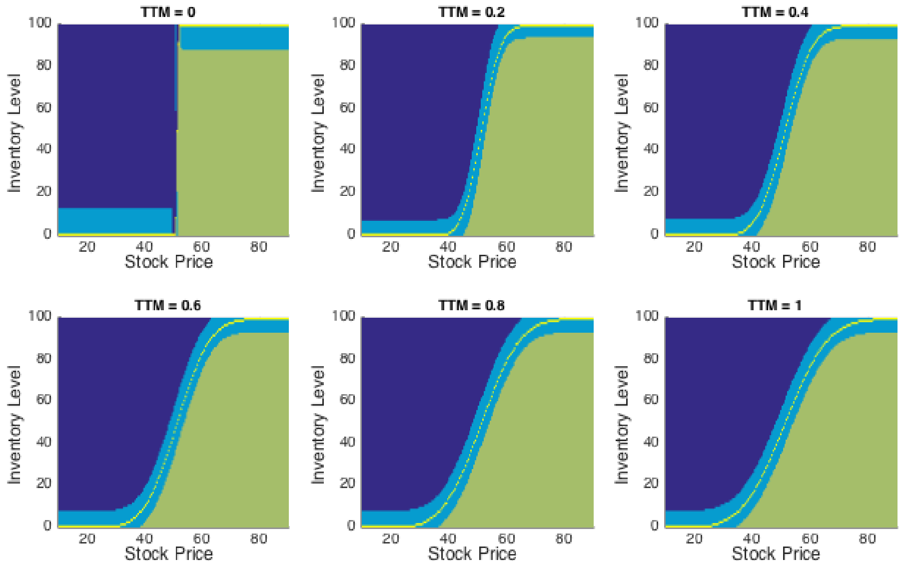

Fig. 5.2 The optimal limit/market order regions for the portfolio manager at different times to maturity. The yellow line which cuts each figure across diagonally is the targeted inventory level Δ_t for different values of the underlying stock. The light blue region surrounding it, is stock-inventory levels for which the portfolio manager is contented with placing limit orders in an attempt to attain delta neutrality. Finally, the dark blue region (top left) and the olive coloured region (bottom right) correspond to market orders, i.e. the case where the portfolio manager deems it necessary to trade immediately to bounce back into the limit order region. For the corresponding surface plots see figure 5.3.

Here $k \in \mathbb{N}^+$, $p \in (0,1]$, and the first and second moments are of the form

$$\mathbb{E}[Z_j^i] = \frac{1}{p}, \qquad \text{and} \qquad \mathbb{E}[(Z_j^i)^2] = \frac{2-p}{p^2}.$$

Jumps of unit tick size thus dominate the price process fluctuations, while larger jumps occur at a power decaying rate.[15] Finally, as a direct consequence of (5.7) alongside the approximation (5.10), the targeted Bachelierian hedge strategy must be of the form

$$\Delta_t = \mathfrak{N}\Phi\left(\frac{S_t - K}{\sigma_m\sqrt{2\gamma p^{-2}(2-p)(T-t)}}\right). \tag{5.30}$$

[15] In practical applications, we cut off the mass function at a level k such that the probability of a jump of size k occurring is much less than one over the total number of incremental time steps (N).

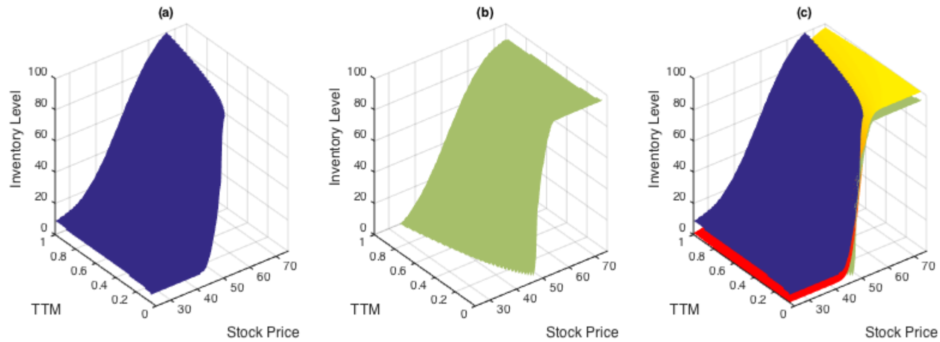

Fig. 5.3 A three dimensional representation of figure 5.2. **(a)** The optimal market sell surface. States which lie above this surface call for the immediate liquidation of inventory through market orders. **(b)** The optimal market buy surface. States which lie below this surface call for the immediate acquisition of inventory through market orders. **(c)** The two market surfaces together with the targeted hedge strategy.

5.6.2 Simulation

Definition 5.1. We designate the joint process $\{S_t, Q_t\}_{t\in[0,T]}$ (i.e. the price path traced out by the underlying asset over $[0,T]$, together with the associated inventory level held by the portfolio manager) the **stock-inventory path**.

Remark 5.5. Clearly, while the stock path is exogenously determined by market forces, the portfolio manager is at liberty to choose his associated inventory level.

Quantity	*Interpretation*	*Magnitude*	*Quantity*	*Interpretation*	*Magnitude*
K	strike	50.5	η	risk aversion	1
σ_m	tick-size	0.5	κ	limit decay	0.3
T	TTM	1	λ	limit intensity	100
$\mathfrak{N}$	no. of calls	100	γ	market intensity	200
Υ	half-spread	1	p	jump size prob.	0.9

Table 5.1 Parameter specifications used in simulation experiment. Some arbitrariness surrounds these numbers in the sense that no empirical calibration has been performed - nonetheless, they serve quite nicely for illustrative purposes.

Consider the parametric specifications listed in table 5.1. Upon solving the HJB QVI numerically, we find the optimal trade regions exhibited cross-sectionally in figure 5.2 for various times to maturity. Jointly, these temporal hypersurfaces form the surface plots exhibited in figure 5.3. The interpretation is as follows: the portfolio manager aims to keep his portfolio approximately delta-neutral understood in the sense of keeping his stock-inventory path on a par with the yellow hedge surface Δ_t (whatever the realised value of

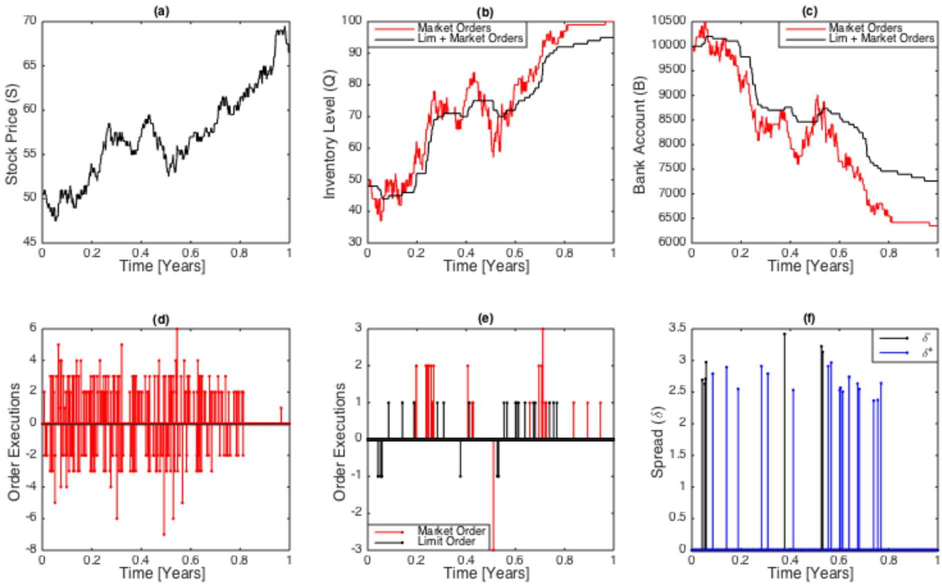

Fig. 5.4 Instantiation of a stock-inventory path. **(a)** The evolution of the mid-price, seeded at $S_0 = K$. **(b)** The evolution of the portfolio manager's inventory when he hedges using market orders only (red) or a combination of market and limit orders (black). **(c)** The evolution of the portfolio manager's bank account, again using market orders only (red) or a combination of market and limit orders (black). Notice the latter outperforms the former. The bank wealth is (arbitrarily) seeded at $B_0 = 10^4$. **(d)** The magnitude of the orders placed by the portfolio manager when he hedges with market orders only. **(e)** The magnitude of the orders placed by the portfolio manager when he hedges with a combination of market (red) and limit (black) orders. **(f)** The size of the spread posted by the portfolio manager in those cases where his limit orders are successful. Black stems are sells, while blue stems are buys.

the underlying, S_t, may be). However, he is contended with merely posting limit sell or buy orders (depending on whether he is above or below Δ_t) in the immediately adjacent region highlighted with a light blue colour in figure 5.2, thus allowing for minor deviations from the targeted hedge ratio. Should his stock-inventory path pierce either the dark blue surface or the olive coloured surface, his "risk aversion" kicks in, and he performs an immediate inventory alteration through unfavourable market orders (to the point where he re-enters the limit order region). Notice that as the market sell and the market buy surfaces both converge towards the strike K at maturity, it becomes increasingly difficult to avoid placing market orders with time if $S_t \approx K$.

The implications of this are forcefully demonstrated in figures 5.4 and 5.5 for the realisation of a single stock path.[16] Here we investigate the performance of the portfolio manager's combined limit-market order hedge strategy vis-à-vis the *naïve hedge strat-*

[16] In simulating a Poisson path, it is helpful to recall that the time spent between consecutive jumps, τ, is a random variable which follows the exponential distribution $\mathbb{P}\{\tau \leq t\} = 1 - \exp\{-\gamma t\}$, see [40] [45].

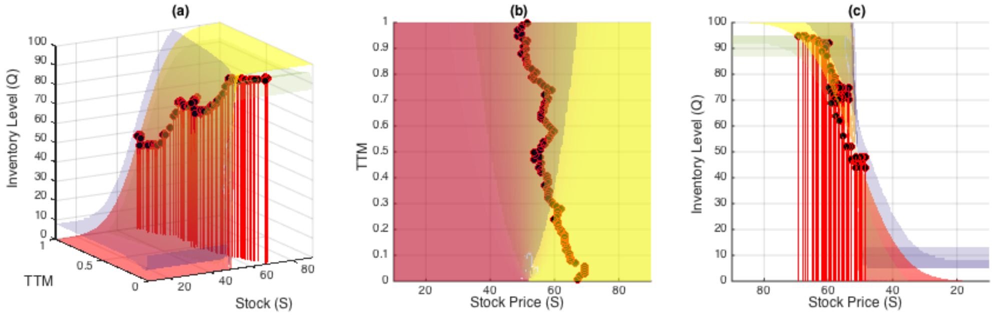

Fig. 5.5 **(a)** A three dimensional view of the stock-inventory path described in figure 5.4 when the portfolio manager uses a combination of limit and market orders. Importantly, the path does not remain embedded in the yellow hedge surface, but is free to bounce back and forth between the encompassing market surfaces discretely highlighted in blue and olive green. **(b)** The graph viewed from $Q = \mathfrak{N}$. Notice the symmetric nature of the market surfaces viewed from this angle. **(c)** The graph viewed from $t = 0$. Here it is quite apparent that the path moves within the pocket space defined by the surrounding market surfaces.

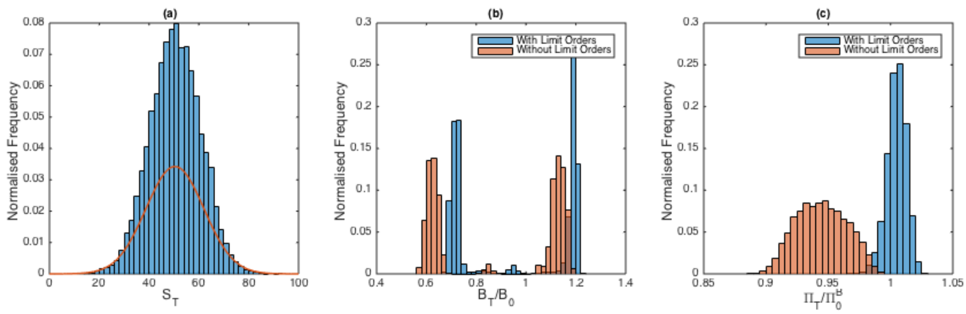

Fig. 5.6 **(a)** Histogram of the terminal stock price with Bachelierian normal fit. Clearly, the latter is inadequate except around the tails. **(b)** Histogram of the rate of return of the bank account. The "camel hump" shape is symptomatic of the binary "all or nothing" inventory position required for delta neutrality at maturity. **(c)** Histogram of the rate of return of the portfolio (bank + stock + options), where Π_0^B is the Bachelier value of the portfolio, defined below. Evidently (and unsurprisingly) the combined limit/market order strategy greatly outperforms the strategy which trades in market orders only.

egy which at all times utilises market orders to maintain delta neutrality. Unsurprisingly, since the former strategy permits all stock-inventory paths sandwiched between the optimal market surfaces, whilst the latter only allows those stock-inventory paths which are embedded[17] in the optimal hedge surface, $\{Q_t\}_{t\in[0,T]}$ and $\{B_t\}_{t\in[0,T]}$ are considerably less erratic in case of the former. Indeed, the lower frequency of trading, combined with the favourable prices associated with limit orders, yield a much more favourable cash position at the expiry of the options. Specifically, for the simulation at hand, the number of (not necessarily unit sized) market orders placed with the naïve strategy amounts to 220, while the number of market-limit orders placed with the control strategy come in at 27 & 24,

[17] Since we assume that the portfolio manager only buys and sells integer quantities of the underlying, it is clear that the embedding should be read as "correct to the nearest integer".

with an average spread for the limit orders of 2.75. As for the mean quadratic variation of the inventory path, $\overline{(\Delta Q)^2} \equiv \frac{1}{N}\sum_{i=1}^{N}(\Delta Q_i)^2$, the naïve strategy amounts to 1.664, while the control strategy comes in at 0.103. The analogous quantity, computed for the bank path yields 4964 vs. 334.

To develop an understanding of the distributive properties of the hedge strategies above, we repeat the experiment 10^4 times. The resulting histograms are exhibited in figure 5.6, while the estimators are given in see table 5.2. As expected, the distribution of the terminal stock value (figure 5.6(a)) falls symmetrically around the strike: however, it is worth noting that the empirical standard deviation (10.12) falls somewhat shy of the Bachelierian standard deviation (11.65). Indeed, the bell curve appears to be a poor fit to the simulation data except around the tails.[18] This highlights that the employ of the Bachelier dynamics in lieu of the jump processes is not altogether innocuous. As for the rate of return of the bank account, B_T/B_0, we refer to figure 5.6(b). Whilst it is clear that the combined limit-market order strategy is superior to the naïve market order strategy, the "camel hump" nature of the histogram might prima facie strike one as puzzling. This mystery is quickly deflated, however, upon realising that the terminal value of the targeted hedge strategy - the step function $\mathfrak{N}\mathbf{1}\{S_T \geq K\}$ - invariably will drive the portfolio manager to either liquidate or acquire ~ 50 shares relative to his initial inventory. The asymmetry between the negative and positive return humps is likely to be symptomatic of the asymmetric boundary condition (i.e. the operations executed given that the calls are in or out of the money). Finally, figure 5.6(c) exhibits the return of the hedge portfolio, Π_T/Π_0 where

$$\Pi_t = B_t + \Delta_t S_t - \mathfrak{N} \cdot C_t.$$

Blatantly, $\Pi_{t<T}$ is a model dependent quantity which prompts us to suggest the at-the-money Bachelier price of the call option,

$$C_{t,\mathrm{atm}}^{K,T} = \sigma_m \sqrt{\frac{\gamma(2-p)(T-t)}{\pi p^2}},$$

in accordance with equation (5.6). Again, while this is dynamically consistent with the choice of the Δ_t, (5.30), it is also in flagrant disregard for the jump nature of the underlying and the friction prevailing on the market. Hence, the Bachelier specification Π_0^B should not be seen as an absolute standard. Nevertheless, it is intriguing to note that it on average yields a positive return for the portfolio manager who uses a combination of limit and market orders to hedge his position, whilst the naïve hedge engenders a mean negative return. Indeed, a Welchian t-test shows that we can comfortably dismiss the null hypothesis of equal means at the 99% level. For more general values of Π_0 we simply observe that the mean of the control strategy return is higher than that of the naïve strategy, while the opposite is the case for the standard deviation.

[18] This is hardly surprising: e.g. it is a well-known fact that the Skellam distribution (the law of the difference between two Poisson random variables, $N_1 - N_2$) tends to the normal distribution if (i) N_1 and N_2 have common intensity and, importantly, (ii) $N_1 - N_2 = k$ is large. See Abramowitz and Stegun [1].

	B_T/B_0 LIMs	B_T/B_0 MOs	Π_T/Π_0^B LIMs	Π_T/Π_0^B MOs
Mean	0.9554	0.8868	1.0048	0.9445
Standard dev.	0.2345	0.2521	0.0077	0.0199
1% Quantile	0.6856	0.5828	0.9848	0.9044
25% Quantile	0.7183	0.6256	1.0000	0.9292
50% Quantile	0.9549	1.0574	1.0052	0.9438
75% Quantile	1.1926	1.1327	1.0102	0.9598
99% Quantile	1.2119	1.1763	1.0210	0.9843

Table 5.2 Estimators for figure see table 5.6. LIMs refers to the combined limit-market order strategy, while MOs refers to the naïve market order strategy.

5.7 Conclusion

The emergence of certain stylised properties of the otherwise hyper-complex limit order book beckons us to attempt mathematical modelling theoreof. In this paper we proposed a simple jump model of the mid-price, alongside a probability of having a limit order met which is an exponentially decaying function of the distance to the mid-price. It was shown that the jump model converges weakly to Bachelier's arithmetic Brownian motion for large values of the intensity. Furthermore, under the assumption that a portfolio manager wishing to Δ-hedge his short call position derives linear utility from his terminal wealth, but incurs quadratic lifetime penalisation from deviating from a targeted hedge strategy, we derived numerical values for his *optimal limit order quotes* and *stopping times* at which he should switch to pure market orders. Following the standard literature on convergence in the viscosity sense we proved that our scheme is monotone, stable, and consistent. We also exemplified the utility of our algorithm by comparing it to a naïve strategy which deploys nothing but market orders: a clear augmentation of the mean return and a clear reduction in the associated variance was established here. Finally, given the generic nature of our model, we reiterate that it is adaptable to more sophisticated expressions for the targeted hedge ratio in the limit order book.

5.7.1 Acknowledgements

The research underpinning this paper was conducted under the masterly guidance of Mark H.A. Davis. His many insightful suggestions have greatly enhanced the quality of my research. Obviously, any remaining errors and omissions are my responsibility alone.

References

1. Abramowitz and Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications. June 1965. pp. 374?378. ISBN 0486612724.
2. Aivaliotis and Palczewski, *Tutorial for Viscosity Solutions in Optimal Control of Diffusions*, October 2011. Available at SSRN: http://ssrn.com/abstract=1582548.
3. Almgren, *Optimal Trading with Stochastic Liquidity and Volatility*, *SIAM J. FINANCIAL MATH.*, 2012, Vol. 3, pp. 163-181.
4. Almgren and Chriss *Optimal Execution of Portfolio Transactions*, *J. Risk 3*, Winter 2000/2001, 5-39.
5. Agliardi and Gencay, *Hedging Through a Limit Order Book with Varying Liquidity*, *Journal of Derivatives*, Winter 2014, pp. 32-49.
6. Avellaneda and Stoikov, *High-frequency Trading in a Limit Order Book*, *Quantitative Finance*, Vol. 8, No. 3, April 2008, 217-224.
7. Bachelier, *Théorie de la Spéculation*, *Annales Scientifiques de l'É.N.S.*, 1900.
8. Barles and Souganidis, *Convergence of Approximation Schemes for Fully Nonlinear Second Order Equations*, *Asymptotic Analysis*, 4 (1991) 271-283.
9. Barndorff-Nielsen, Lunde, Shephard, and Veraart, *Integer-valued Trawl Processes: A Class of Stationary Infinitely Divisible Processes*, *Scandinavian Journal of Statistics*, 2014, 41, pp. 693-724.
10. Björk, *Arbitrage Theory in Continuous Time*, 2009, Oxford University Press, 3rd edition.
11. Bogard, *High Frequency Trading: An Important Conversation*, March 24, 2014, available online at: http://tabbforum.com/opinions/high-frequency-trading-an-important-conversation.
12. Carmona and Webster, *The Self-financing Condition in High-frequency Markets*, 2013, Unpublished, arXiv:1312.2302 [q-fin.TR].
13. Cartea, Jaimungal, and Ricci, *Buy Low, Sell High: A High Frequency Trading Perspective*, *SIAM Journal on Financial Mathematics*, 5.1 (2014): 415-444.
14. Cartea and Jaimungal, *Optimal Execution with Limit and Market Orders*, (November 20, 2014). Forthcoming: *Quantitative Finance*. Available at SSRN: http://ssrn.com/abstract=2397805 or http://dx.doi.org/10.2139/ssrn.2397805.
15. Cont, Stoikov, and Talreja, *A Stochastic Model for Order Book Dynamics*, *Operations Research*, Vol. 58, No. 3, May-June 2010, pp. 540-563.
16. Cont and Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC Financial Mathematics Series. 2004.
17. Demos, *'Real' Investors Eclipsed by Fast Trading*, newspaper article in *Financial Times*, April 24, 2012.
http://www.ft.com.iclibezp1.cc.ic.ac.uk/cms/s/0/da5d033c-8e1c-11e1-bf8f-00144feab49a.html#axzz3l4SFA7zo.
18. Donier, Bonart, Mastromatteo, and Bouchaud, *A Fully Consistent, Minimal Model for Non-linear Market Impact*, 2015, Working paper, http://arxiv.org/abs/1412.0141.
19. Duffie, *Dynamic Asset Pricing Theory*. 2001, Princeton University Press, 3rd edition.
20. Foucault, Pagano, and Roell, *Market Liquidity: Theory, Evidence, and Policy*, 2013, Oxford University Press.
21. Fleming and Soner, *Controlled Markov Processes and Viscosity Solutions*, 2005 ,Springer, 2nd edition.
22. Fodra and Labadie, *High-frequency Market-making with Inventory Constraints and Directional Bets*, 2012. Unpublished. arXiv:1206.4810 [q-fin.TR].
23. Forsyth, *A Hamilton Jacobi Bellman Approach to Optimal Trade Execution*, *Applied Numerical Mathematics*, Volume 61, Issue 2, February 2011, pp. 241-265.
24. Forsyth and Labahn, *Numerical methods for controlled Hamilton-Jacobi-Bellman PDEs in finance*, *Journal of Computational Finance*. 11:2 (2007/2008: Winter) 1-44.
25. Gould, Porter, Williams, McDonald, Fenn, and Howison, *Limit Order Books*, 2013, *Quantitative Finance*. Vol. 13, No. 11, pp. 1709-1742.
26. Gueant, Lehalle, and Fernandez Tapia, *Dealing with the Inventory Risk. A solution to the market making problem*, 2011. Unpublished. arXiv:1105.3115 [q-fin.TR].
27. Guilbaud and Pham, *Optimal High Frequency Trading with Limit and Market Orders*, (2011). Available at SSRN: http://ssrn.com/abstract=1871969 .

28. Hanson, *Applied Stochastic Processes and Control for Jump-Diffusions: Modeling, Analysis, and Computation (Advances in Design and Control)*, 2007.
29. He and Mamaysky, *Dynamic Trading Policies With Price Impact*, (October 2001). Yale ICF Working Paper No. 00-64. Available at SSRN: http://ssrn.com/abstract=289257.
30. Ho and Stoll, *Optimal dealer pricing under transactions and return uncertainty*, *J. Financ. Econ.* , 1981, 9, 47-73.
31. Kushner and Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2001, 2nd edition. Springer. ISBN 0-387-95139-3.
32. Last and Brandt, *Marked Point Processes on the Real Line: The Dynamical Approach (Probability and Its Applications)*, Springer, 1995 edition.
33. Lewis, *Flash Boys: Cracking the Money Code*, Penguin, 2015.
34. Li and Almgren, *Option Hedging with Smooth Market Impact*, Working paper. June 11, 2015. http://www.courant.nyu.edu/ almgren/papers/hedging.pdf.
35. Merton, *Lifetime Portfolio Selection under Uncertainty: the Continuous-Time Case*. *The Review of Economics and Statistics* 51 (3) pp. 247-257 (1969).
36. Munk, *Numerical Methods for Continuous-Time, Continuous-State Stochastic Control Problems*, Working paper. 1997. https://dl.dropboxusercontent.com/u/65974535/MyPapers/MCHAIN.pdf.
37. Obizhaeva and Wang, *Optimal Trading Strategy and Supply/Demand Dynamics*, *Journal of Financial Markets*, Volume 16, Issue 1, February 2013, Pages 1-32.
38. Øksendal and Sulem, *Applied Stochastic Control of Jump Diffusions*, (Universitext), 2007, Springer, 2nd edition.
39. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications*. *Stochastic Modelling and Applied Probability 61*, 2010, Springer. e-ISBN: 978-3-540-89500-8.
40. Privault, *Stochastic Calculus in Finance II*. Unpublished lecture notes. http://www.ntu.edu.sg/home/nprivault/indext.html.
41. Ross, *Stochastic Control in Continuous Time*. Unpublished lecture notes. http://www.swarthmore.edu/NatSci/kross1/Stat220notes.pdf.
42. Rosu, *Liquidity and information in order driven markets*. Working paper, SSRN eLibrary, 2010. Available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1286193.
43. Runggaldier, W. *Jump Diffusion Models*, in *Handbook of Heavy Tailed Distributions in Finance*, 2003 Elsevier Science B.V.
44. Shephard and Yang, *Continuous Time Analysis of Fleeting Discrete Price Moves*, Unpublished, 2015.
45. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models: v. 2*, Springer Finance, Corr. 2nd printing 2010 edition.
46. Tourin, *An introduction to Finite Difference methods for PDEs in Finance*, Lecture notes, 2010, Fields Institute, Toronto.

Appendix: The Dynamic Programming Principle (DPP)

An informal derivation of the DPP (5.13) runs along the following lines. Let X_T designate the total *financial wealth* of the portfolio manager, and let $\mathscr{T}_n = \{t_i | t_0 = 0, t_1 = t_0 + \Delta t_0, t_2 = t_1 + \Delta t_1, \ldots, t_{n+1} = T = t_n + \Delta t_n\}$ be a discrete temporal partitioning, then the optimal value function (5.4) for limit orders may be written as

$$V(t_i, b, s, q) = \lim_{n \to \infty} \sup_{\{\delta^+_{t_j}, \delta^-_{t_j}\}^n_{j=i} \geq 0} \mathbb{E}_{t_i, b, s, q}\left[X_{t_{n+1}} - \eta \sum_{j=i}^{n} (q - \Phi(\delta_{t_j}))^2 \Delta t_j\right].$$

where $(t, B_t, S_t, Q_t) = (t_i, b, s, q)$. Using the law of iterated expectations the discrete expression may be rewritten as

$$\begin{aligned}
V_{t_i} &= \sup_{\{\delta^+_{t_j}, \delta^-_{t_j}\}^n_{j=i} \geq 0} \mathbb{E}_{t_i}\left[X_{t_{n+1}} - \eta (q - \Phi(\delta_{t_i}))^2 \Delta t_i - \eta \sum_{j=i+1}^{n} (q - \Phi(\delta_{t_j}))^2 \Delta t_j\right] \\
&= \sup_{\{\delta^+_{t_j}, \delta^-_{t_j}\}^n_{j=i} \geq 0} \mathbb{E}_{t_i}\left[-\eta (q - \Phi(\delta_{t_i}))^2 \Delta t_i + \mathbb{E}_{t_{i+1}}\left[X_{t_{n+1}} - \eta \sum_{j=i+1}^{n} (q - \Phi(\delta_{t_j}))^2 \Delta t_j\right]\right] \\
&= \sup_{\delta^+_{t_i}, \delta^-_{t_i} \geq 0} \mathbb{E}_{t_i}\left[-\eta (q - \Phi(\delta_{t_i}))^2 \Delta t_i + \sup_{\{\delta^+_{t_j}, \delta^-_{t_j}\}^n_{j=i+1} \geq 0} \mathbb{E}_{t_{i+1}}\left[X_{t_{n+1}} - \eta \sum_{j=i+1}^{n} (q - \Phi(\delta_{t_j}))^2 \Delta t_j\right]\right] \\
&= \sup_{\delta^+_{t_i}, \delta^-_{t_i} \geq 0} \mathbb{E}_{t_i}\left[-\eta (q - \Phi(\delta_{t_i}))^2 \Delta t_i + V_{t_{i+1}}\right].
\end{aligned}$$

Subtracting V_{t_i} on both sides and using the $\mathscr{F}_{t_i}$-measurability of $\eta (q - \Phi(\delta_{t_i}))^2 \Delta t_i$ we arrive at the discretised DPP:

$$0 = -\eta (q - \Phi(\delta_{t_i}))^2 \Delta t_i + \sup_{\delta^+_{t_i}, \delta^-_{t_i} \geq 0} \mathbb{E}_{t_i}\left[V_{t_{i+1}} - V_{t_i}\right].$$

Setting $n \to \infty$ yields the desired result.

Appendix: A Brief Introduction to Viscosity Solutions

In this appendix we briefly sketch the bare essentials associated with the field of viscosity solutions. For a more thorough review the reader is referred to Aivaliotis and Palczewski [2], Fleming and Soner [21], and Pham [18], followed by the references in Barles and Souganidis [8]. For generality, consider the generic non-linear second order PDE

$$\mathscr{H}(\mathbf{x}, w(\mathbf{x}), \nabla w(\mathbf{x}), \nabla^2 w(\mathbf{x})) = 0, \qquad \mathbf{x} \in \bar{\mathbb{O}}, \tag{5.31}$$

where ∇ and ∇^2 respectively represent the gradient and the Hessian operators. Here $\mathscr{H}$: $\bar{\mathbb{O}} \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^n \mapsto \mathbb{R}$ and $w : \bar{\mathbb{O}} \mapsto \mathbb{R}$ are locally bounded *possibly discontinuous* functions, $\mathbb{S}^n$ is the space of symmetric $n \times n$ matrices, $\mathbb{O}$ is an open subset of $\mathbb{R}^n$, whilst $\bar{\mathbb{O}}$ is its closure.

Definition 5.2. The upper semi-continuous (usc) envelope and the lower semi-continuous (lsc) envelope of a function $\mathbf{f} : \bar{\mathbb{O}} \mapsto \mathbb{R}^n$ are respectively defined as

$$\mathbf{f}^*(\mathbf{x}) \equiv \limsup_{\mathbf{y} \to \mathbf{x}, \mathbf{y} \in \bar{\mathbb{O}}} \mathbf{f}(\mathbf{y}), \quad \text{and} \quad \mathbf{f}_*(\mathbf{x}) \equiv \liminf_{\mathbf{y} \to \mathbf{x}, \mathbf{y} \in \bar{\mathbb{O}}} \mathbf{f}(\mathbf{y}).$$

Definition 5.3. A locally bounded function $w : \bar{\mathbb{O}} \mapsto \mathbb{R}$ is a *viscosity subsolution* of (5.31) if $\forall \phi \in \mathscr{C}^2(\bar{\mathbb{O}})$ and $\forall \mathbf{x} \in \bar{\mathbb{O}}$ such that $u^* - \phi$ has a local maximum at $\mathbf{x}$, we have

$$\mathscr{H}(\mathbf{x}, u^*(\mathbf{x}), \nabla\phi(\mathbf{x}), \nabla^2\phi(\mathbf{x})) \leq 0.$$

On, the other hand, we say that w is a *viscosity supersolution* of (5.31) if $\forall \phi \in \mathscr{C}^2(\bar{\mathbb{O}})$ and $\forall \mathbf{x} \in \bar{\mathbb{O}}$ such that $u_* - \phi$ has a local minimum at $\mathbf{x}$, we have

$$\mathscr{H}(\mathbf{x}, u_*(\mathbf{x}), \nabla\phi(\mathbf{x}), \nabla^2\phi(\mathbf{x})) \geq 0.$$

If w is both a viscosity sub- and supersolution, it is said to be a *viscosity solution* of (5.31).

Part III
Appendices

Appendix A
Martingale Methods in Mathematical Finance

Abstract In this appendix we provide a cursory overview of standard results pertaining to martingale methods in mathematical finance. The focal point here is the power of probability measure transformations: in particular, we expose the Abstract Bayes' Theorem, Girsanov's theorem and the change of numeraire, the first and second fundamental theorems of asset pricing, and the concept of market price of risk. No in-text references are provided, but a full list of relevant "textbook sources" is provided at the end of this dissertation.

A.1 Martingales

Let $(\Omega, \mathscr{F}, \mathbb{P}, \mathbb{F} = \{\mathscr{F}\}_{t\geq 0})$ be a filtered probability space, and let $X_t : \Omega \times [0,\infty) \mapsto \mathbb{R}$ be a stochastic process defined upon it. We suppose X is integrable, meaning that $\mathbb{E}[|X_t|] < \infty$.

Definition A.1. We say that X_t is a **martingale** if $\mathbb{E}[X_t|\mathscr{F}_s] = X_s$ for all $s < t \in \mathbb{R}_+$. Furthermore, X_t is said to be a **submartingale** (resp. **supermartingale**) if $\mathbb{E}[X_t|\mathscr{F}_s] \geq X_s$ (resp. $\mathbb{E}[X_t|\mathscr{F}_s] \leq X_s$) for all $s < t \in \mathbb{R}_+$.

Furthermore, as an abstract generalisation we have the notion of a local martingale:

Definition A.2. We say that X_t is a **local martingale** if there exists a sequence of $\mathbb{F}$-stopping times $\tau_k : \Omega \mapsto [0,\infty)$ such that (i) the τ_k are almost surely increasing: $\mathbb{P}\{\tau_k < \tau_{k+1}\} = 1$; (ii) the τ_k diverge almost surely: $\mathbb{P}\{\tau_k \to \infty \text{ as } k \to \infty\} = 1$; (iii) the stopped process $X_t^{\tau_k} \equiv X_{\min\{t,\tau_k\}}$ is an $\mathbb{F}$-martingale for every k.

Generally, all martingales are local martingales, but the converse is not the case (the classic example of a process which is a local martingale but need not be martingale is the driftless diffusion process). However, if a local martingale M_t also satisfies (*) $\mathbb{E}[\sup_{s\in[0,t]} |M_s|] < \infty$ for every t, then M_t is indeed also a martingale. For example, if $M_t = f(t, W_t)$ where W_t is a Wiener process and $f \in \mathscr{C}^{1,2}([0,\infty) \times \mathbb{R})$ then we have that M_t is a local martingale provided the standard driftless condition

$$\{\partial_t + \tfrac{1}{2}\partial^2_{xx}\} f(t,x) = 0,$$

holds (cf. the Itō formula). In order to apply (*) it suffices that $\forall \varepsilon > 0\ \forall t\ \exists C = C(\varepsilon, t)$ such that $|f(s,x)| \le C \exp\{\varepsilon x^2\}$ for all $(s,x) \in [0,t] \times \mathbb{R}$.

A.2 Changing the Measure

Consider the probability space $(\Omega, \mathscr{F})$ then we may think of how different allocations of probabilities to events in this space are interconnected. We say that two probability measures $\mathbb{P}$ and $\mathbb{Q}$ are **equivalent** (labelled $\mathbb{P} \sim \mathbb{Q}$) on $\mathscr{F}$ just in case

$$\mathbb{P}(A) = 0 \Leftrightarrow \mathbb{Q}(A) = 0, \quad \forall A \in \mathscr{F}.$$

In particular, the **Radon-Nikodym theorem** instructs us that $\mathbb{P}(A) = 0 \Rightarrow \mathbb{Q}(A) = 0\ \forall A \in \mathscr{F}$ (i.e. $\mathbb{Q}$ is absolutely continuous w.r.t. $\mathbb{P}$ on $\mathscr{F}$: $\mathbb{Q} \ll \mathbb{P}$) if and only if there exists an $\mathscr{F}$-measurable mapping $\xi : \Omega \mapsto \mathbb{R}_+$ such that

$$\int_A d\mathbb{Q}(\omega) = \int_A \xi(\omega) d\mathbb{P}(\omega), \quad \forall A \in \mathscr{F}. \tag{A.1}$$

In the event that $A = \Omega$ the left-hand-side in this expression is unity (per definition of a probability measure). Likewise, the right-hand-side is defined as $\int_\Omega \xi d\mathbb{P} \equiv \mathbb{E}^{\mathbb{P}}[\xi]$. All in all, the quantity ξ is therefore a non-negative random variable with $\mathbb{E}^{\mathbb{P}}[\xi] = 1$. Since (A.1) infinitesimally can be written $\xi = d\mathbb{Q}/d\mathbb{P}$, ξ is commonly referred to as the **likelihood ratio** between $\mathbb{Q}$ and $\mathbb{P}$ or the **Radon-Nikodym derivative**. Three standard results surrounding ξ deserve mentioning:

1. For any random variable X on $L^1(\Omega, \mathscr{F}, \mathbb{Q})$: $\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}^{\mathbb{P}}[\xi X]$ and $\mathbb{E}^{\mathbb{Q}}[\xi^{-1} X] = \mathbb{E}^{\mathbb{P}}[X]$. Proof: obvious using definitions.
2. Assume $\mathbb{Q}$ is absolutely continuous w.r.t. $\mathbb{P}$ on $\mathscr{F}$ and that $\mathscr{G} \subseteq \mathscr{F}$, then the likelihood ratios $\xi^{\mathscr{F}}$ and $\xi^{\mathscr{G}}$ are related by $\xi^{\mathscr{G}} = \mathbb{E}^{\mathbb{P}}[\xi^{\mathscr{F}} | \mathscr{G}]$.
3. Finally, assume X is a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathbb{Q}$ be another measure on $(\Omega, \mathscr{F})$ with Radon-Nikodym derivative $\xi = d\mathbb{Q}/d\mathbb{P}$ on $\mathscr{F}$. Assume $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{G} \subseteq \mathscr{F}$ then
$$\mathbb{E}^{\mathbb{Q}}[X|\mathscr{G}] = \frac{\mathbb{E}^{\mathbb{P}}[\xi X|\mathscr{G}]}{\mathbb{E}^{\mathbb{P}}[\xi|\mathscr{G}]}, \quad \mathbb{Q}-\text{a.s.} \tag{A.2}$$
This result is sometimes referred to as the **Abstract Bayes' Theorem**.

Example: To get a feel for how these results are used in mathematical finance we consider the classical set-up: a filtered probability space $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}\}_{t\in[0,T]})$ on a compact interval $[0,T]$. Typically, we are interested in some stochastic process $\{X_t\}_{t\in[0,T]}$ (e.g. a stock price) such that Ω is the set of all possible paths of the process over $[0,T]$. Since all relevant uncertainty has been resolved at time T all (relevant) random variables will be known at time T. If we now consider the non-negative random variable ξ_T in $\mathscr{F}_T$, then provided

$\mathbb{E}^{\mathbb{P}}[\xi_T] = 1$ we may define a new probability measure $\mathbb{Q}$ on $\mathscr{F}_T$ by setting $d\mathbb{Q} = \xi_T d\mathbb{P}$. Per definition, ξ_T is a Radon-Nikodym derivative of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ on $\mathscr{F}_T$ so $\mathbb{Q} \ll \mathbb{P}$ on $\mathscr{F}_T$. Thus, we will also have $\mathbb{Q} \ll \mathbb{P}$ on $\mathscr{F}_t$ $\forall t \leq T$ so by the Radon-Nikodym Theorem there exists a random process $\{\xi_t\}_{t\in[0,T]}$ defined by $\xi_t = d\mathbb{Q}/d\mathbb{P}$ on $\mathscr{F}_t$, which we call the **likelihood process**. Item (2) above now immediately implies that the ξ-process is a $\mathbb{P}$-**martingale**:

$$\mathbb{E}^{\mathbb{P}}[\xi_{t'}|\mathscr{F}_t] = \xi_t, \;\; t' > t.$$

Using this fact alongside item (3) also gives us the result that:

$$\mathbb{E}^{\mathbb{Q}}[X_{t'}|\mathscr{F}_t] = \mathbb{E}^{\mathbb{P}}\left[\frac{\xi_{t'}}{\xi_t}X_{t'}\middle|\mathscr{F}_t\right], \tag{A.3}$$

which turns out to be extremely useful in option pricing upon jumping between different numeraires.

A.3 The First and Second Fundamental Theorems

We consider a market model consisting of the non-dividend paying asset price processes $S_0, S_1, ..., S_N$ on the time interval $[0,T]$.

Theorem A.1. ***The First Fundamental Theorem.*** *The market model is free of arbitrage if and only there exists a martingale measure, i.e. a measure $\mathbb{Q} \sim \mathbb{P}$ such that the processes*

$$\frac{S_{0t}}{S_{0t}}, \; \frac{S_{1t}}{S_{0t}}, \; \dots, \; \frac{S_{Nt}}{S_{0t}},$$

are (local) martingales under $\mathbb{Q}$.

Notice that we don't commit ourselves to the interpretation that the numeraire, S_0, is the risk free asset. However, if indeed $S_{0t} = B_t \equiv \exp(\int_0^t r_s ds)$ where r is a possibly stochastic short rate, and we assume all processes are Wiener driven, meaning that $dS_{it} = S_{it}\mu_{it}dt + S_{it}\boldsymbol{\sigma}_{it}^{\intercal} d\boldsymbol{W}_t^{\mathbb{P}}$, then a measure $\mathbb{Q} \sim \mathbb{P}$ (the so **risk-neutral measure** associated with the risk free numeraire) is a martingale measure if and only if

$$dS_{it} = S_{it}r_t dt + S_{it}\boldsymbol{\sigma}_{it}^{\intercal} d\boldsymbol{W}_t^{\mathbb{Q}}, \tag{A.4}$$

$\forall i \in \{0,1,...,N\}$, where $\boldsymbol{W}^{\mathbb{Q}}$ is a d-dimensional $\mathbb{Q}$-Wiener process. I.e. all assets have the short rate r as their local rates of return. Proof: apply Itō's lemma to S_{it}/S_{0t}. Just in case $\mu_{it} = r_t$ do we obtain a local martingale (i.e. vanishing drift).

Next, we consider what it takes for us to be able to replicate (synthesise) assets on the market using existing products:

Theorem A.2. ***The Second Fundamental Theorem.*** *Assuming absence of arbitrage, the market model is complete if and only if the martingale measure $\mathbb{Q}$ is unique.*

Remark A.1. This does clearly not say that there is only one martingale measure in existence. It only says that for this particular choice of numeraire (S_0) the measure is uniquely determined.

Theorem A.3. ***Pricing Contingent Claims.*** *Consider a contingent claim, X, that expires at time T. In order to avoid arbitrage we must price the claim according to*

$$X_t = S_{0t}\mathbb{E}^{\mathbb{Q}}\left[\frac{X_T}{S_{0T}}\middle|\mathscr{F}_t\right], \tag{A.5}$$

where $\mathbb{Q}$ is a martingale measure for $\{S_0, S_1, \ldots, S_N\}$ with S_0 as the numeraire. In particular, insofar as S_{0t} is the risk free asset $S_{0t} = \exp(\int_0^t r_s ds)$, then we obtain the classical pricing formula

$$X_t = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds}X_T\middle|\mathscr{F}_t\right]. \tag{A.6}$$

A.4 Lévy's Characterisation of Wiener Processes

A Wiener process (Brownian motion) W_t is a martingale with continuous paths and quadratic variation $[W,W](t) = t$. These properties actually suffice to characterise a Wiener process as demonstrated by Lévy:

Theorem A.4. ***Lévy's Theorem*** *Let $\boldsymbol{M}_t = (M_{1t}, M_{2t}, \ldots, M_{dt})$ be a martingale with respect to the filtration $\{\mathscr{F}_t\}, t \geq 0$. Assume that (i) $\forall i : M_{i0} = 0$, (ii) $\forall i : M_{it}$ has continuous paths, (iii) $[M_i, M_j](t) = \delta_{ij}t \ \forall t \geq 0$, then M_{1t}, M_{2t}, ..., M_{dt} are independent Wiener processes.*

Proof. We will prove this using characteristic functions. Consider the function $f(t, \boldsymbol{M}_t) = \exp\{\boldsymbol{h}^\top \boldsymbol{M} - \frac{1}{2}\boldsymbol{h}^\top \boldsymbol{h} t\}$. From Itō's lemma

$$\begin{aligned} df(t,\boldsymbol{M}_t) &= \left(\partial_t f(t,\boldsymbol{M}_t) + \tfrac{1}{2}\nabla_x^\intercal \nabla_x f(t,\boldsymbol{M}_t)\right) dt + \nabla_x f(t,\boldsymbol{M}_t)^\intercal d\boldsymbol{M}_t \\ &= \left(-\tfrac{1}{2}\boldsymbol{h}^\intercal \boldsymbol{h} f(t,\boldsymbol{M}_t) + \tfrac{1}{2}\boldsymbol{h}^\intercal \boldsymbol{h} f(t,\boldsymbol{M}_t)\right) dt + \nabla_x f(t,\boldsymbol{M}_t)^\intercal d\boldsymbol{M}_t \\ &= \nabla_x f(t,\boldsymbol{M}_t)^\intercal d\boldsymbol{M}_t. \end{aligned}$$

So f is clearly a martingale. It follows that

$$\mathbb{E}[f(t,\boldsymbol{M}_t)] = 1 \Leftrightarrow \mathbb{E}[e^{\boldsymbol{h}^\intercal \boldsymbol{M}_t}] = e^{\frac{1}{2}\boldsymbol{h}^\intercal \boldsymbol{h} t}.$$

The right-hand side is the moment generating function for independent normal random variables (mean 0 and variance t). The result follows. □

A.5 The Martingale Theorem and Girsanov's Theorem

Let $\boldsymbol{W}$ be a d-dimensional Wiener process and let X be a stochastic variable which is both $\mathscr{F}_T^{\boldsymbol{W}}$ measurable and L^1. Then there exists a uniquely determined $\mathscr{F}_T^{\boldsymbol{W}}$-adapted process $\boldsymbol{h} = (h_1, h_2, ..., h_d)$ such that X has the representation

$$X = \mathbb{E}[X] + \int_0^t \boldsymbol{h}_s^\intercal d\boldsymbol{W}_s. \tag{A.7}$$

Under the additional assumption that $\mathbb{E}[X^2] < \infty$ then $h_1, h_2, ..., h_d$ are in £^{2}.

We can use this lemma to prove the following

Theorem A.5. ***The Martingale Representation Theorem*** *Let* $\boldsymbol{W}$ *be a* d*-dimensional Wiener process, and assume that the filtration* $\{\mathscr{F}_t\}_{t\in[0,T]}$ *is defined as* $\mathscr{F}_t = \mathscr{F}_t^{\boldsymbol{W}}$ *for* $t \in [0,T]$. *Now let* M *be any* $\mathscr{F}_t$ *martingale. Then there exists a uniquely determined* $\mathscr{F}_t$ *adapted process* $\boldsymbol{h} = (h_1, h_2, ..., h_d)$ *such that* M *has the representation*

$$M_t = M_0 + \int_0^T \boldsymbol{h}_s^\intercal d\boldsymbol{W}_s, \quad t \in [0,T].$$

If the martingale M *is square integrable, then* $h_1, h_2, ..., h_d$ *are in* £^{2}.

Recall from section A.2 that the measure transformation $d\mathbb{Q} = \xi_T d\mathbb{P}$ on $\mathscr{F}_T$ (where ξ_T is a nonnegative random variable with $\mathbb{E}^{\mathbb{P}}[\xi_T] = 1$) generates a likelihood process $\{\xi_t\}_{t\in[0,T]}$ defined by $\xi_t \equiv d\mathbb{Q}/d\mathbb{P}$ on $\mathscr{F}_t$ which is a $\mathbb{P}$-martingale. It thus seems natural to define ξ_t as the solution to the SDE

$$d\xi_t = \phi_t \xi_t dW_t^{\mathbb{P}},$$

with initial condition $\xi_0 = 1$ for some choice of the process ϕ (the initial condition guarantees unitary expectation under $\mathbb{P}$). In fact, using this SDE we should be able to generate a host of natural measure transformations from $\mathbb{P}$ to the new measure $\mathbb{Q}$, which indeed also is the upshot of Girsanov's theorem:

Theorem A.6. ***Girsanov's Theorem*** *Let $\boldsymbol{W}^{\mathbb{P}}$ be a d-dimensional standard $\mathbb{P}$-Wiener process on $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}_{t\in[0,T]})$, and let ϕ be any d-dimensional adapted column vector process (referred to as the Girsanov kernel). Now define the process ξ on $[0,T]$ by*

$$d\xi_t = \xi_t \phi_t^{\intercal} d\boldsymbol{W}_t^{\mathbb{P}}, \quad \xi_0 = 1$$

or identically

$$\xi_t = \exp\left\{\int_0^t \phi_s^{\intercal} d\boldsymbol{W}_s^{\mathbb{P}} - \tfrac{1}{2}\int_0^t ||\phi_s||^2 ds\right\}.$$

*Now **assume** that $\mathbb{E}^{\mathbb{P}}[\xi_T] = 1$ (see the Novikov condition) and **define** the new probability measure $\mathbb{Q}$ on $\mathscr{F}_T$ by $d\mathbb{Q} = \xi_T d\mathbb{P}$ on $\mathscr{F}_T$ then*

$$d\boldsymbol{W}_t^{\mathbb{P}} = \phi_t dt + d\boldsymbol{W}_t^{\mathbb{Q}}, \tag{A.8}$$

where $\boldsymbol{W}_t^{\mathbb{Q}}$ is a $\mathbb{Q}$ Wiener process.

Proof. We will use Lévy's theorem to verify that $\boldsymbol{W}_t^{\mathbb{Q}}$ is indeed a $\mathbb{Q}$ Wiener process. Evidently (i) $\boldsymbol{W}_0^{\mathbb{Q}} = 0$, and (ii) $\boldsymbol{W}_t^{\mathbb{Q}}$ is continuous. Furthermore, it is cleat that (iii) $[W_i^{\mathbb{Q}}, W_j^{\mathbb{Q}}](t) = \delta_{ij} t$. The only thing left to show is that $\boldsymbol{W}_t^{\mathbb{Q}}$ is a $\mathbb{Q}$ martingale. To this end, it is already given that ξ_t is a $\mathbb{P}$ martingale with $\mathbb{E}^{\mathbb{P}}[\xi_T] = 1$. Applying Itō to the process $\xi_t \boldsymbol{W}_t^{\mathbb{Q}}$ we find that

$$\begin{aligned} d(\xi_t \boldsymbol{W}_t^{\mathbb{Q}}) &= \boldsymbol{W}_t^{\mathbb{Q}} d\xi_t + \xi_t d\boldsymbol{W}_t^{\mathbb{Q}} + d\xi_t d\boldsymbol{W}_t^{\mathbb{Q}} \\ &= \boldsymbol{W}_t^{\mathbb{Q}} \xi_t \phi_t^{\intercal} d\boldsymbol{W}_t^{\mathbb{P}} + \xi_t (d\boldsymbol{W}_t^{\mathbb{P}} - \phi_t dt) + (d\boldsymbol{W}_t^{\mathbb{P}} - \phi_t dt)\xi_t \phi_t^{\intercal} d\boldsymbol{W}_t^{\mathbb{P}} \\ &= \xi_t (\boldsymbol{W}_t^{\mathbb{Q}} \phi_t^{\intercal} + 1) d\boldsymbol{W}_t^{\mathbb{P}} \end{aligned}$$

which proves that $\xi_t \boldsymbol{W}_t^{\mathbb{Q}}$ is a $\mathbb{P}$ martingale. Now, applying the abstract Bayes' theorem we quickly deduce that

$$\mathbb{E}^{\mathbb{Q}}[\boldsymbol{W}_{t'}^{\mathbb{Q}} | \mathscr{F}_t] = \xi_t^{-1} \mathbb{E}^{\mathbb{P}}[\xi_{t'} \boldsymbol{W}_{t'}^{\mathbb{Q}} | \mathscr{F}_t] = \xi_t^{-1} \xi_t \boldsymbol{W}_t^{\mathbb{Q}} = \boldsymbol{W}_t^{\mathbb{Q}},$$

which was to be proven. □

- Assume that the Girsanov kernel ϕ is such that

$$\mathbb{E}^{\mathbb{P}}\left[e^{\frac{1}{2}\int_0^T ||\phi_t||^2 dt}\right] < \infty$$

  then ξ is a martingale and in particular $\mathbb{E}^{\mathbb{P}}[\xi_T] = 1$. This useful result is known as the **Novikov condition**.

- Girsonov's theorem holds in reverse. In particular, assume $\boldsymbol{W}^{\mathbb{P}}$ is a d-dimensional standard $\mathbb{P}$-Wiener process on $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}_{t\in[0,T]})$ and assume that $\mathscr{F}_t = \mathscr{F}_t^{\boldsymbol{W}}$ $\forall t$. Furthermore, assume there exists a measure $\mathbb{Q}$ such that $\mathbb{Q} \ll \mathbb{P}$ on $\mathscr{F}_T$ then there exists an adapted process ϕ such that the likelihood process ξ has the dynamics

$$d\xi_t = \xi_t \phi_t^{\intercal} d\boldsymbol{W}_t^{\mathbb{P}}, \quad \xi_0 = 1.$$

- SDEs of the form $dX_t = \mu_t dt + \boldsymbol{\sigma}_t d\boldsymbol{W}_t^{\mathbb{P}}$ transform as $dX_t = (\mu_t + \boldsymbol{\sigma}_t \phi)dt + \boldsymbol{\sigma}_t d\boldsymbol{W}_t^{\mathbb{Q}}$ under $\mathbb{Q}$, which means that the drift changes $\mu_t \mapsto \mu_t + \boldsymbol{\sigma}_t \phi$, but the diffusion remains unchanged.

Corollary A.1. ***The Correlated Girsanov Theorem*** *Let $\boldsymbol{W}^{\mathbb{P}}$ be a d-dimensional $\mathbb{P}$-Wiener process on $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}_{t\in[0,T]})$ with correlation matrix $\boldsymbol{\Sigma}$ (i.e. $d\boldsymbol{W}^{\mathbb{P}} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma} dt)$), and let ϕ be any d-dimensional adapted column vector process. Now define the process ξ on $[0,T]$ by $d\xi_t = \xi_t (\boldsymbol{\Sigma}^{-1}\phi_t)^{\intercal} d\boldsymbol{W}_t^{\mathbb{P}}$, where $\xi_0 = 1$, or identically*

$$\xi_t = \exp\left\{ \int_0^t (\boldsymbol{\Sigma}^{-1}\phi_s)^{\intercal} d\boldsymbol{W}_s^{\mathbb{P}} - \frac{1}{2} \int_0^t \phi_s^{\intercal} \boldsymbol{\Sigma}^{-1} \phi_s ds \right\}.$$

Again, assume that $\mathbb{E}^{\mathbb{P}}[\xi_T] = 1$ and ***define*** *the new probability measure $\mathbb{Q}$ on $\mathscr{F}_T$ by $d\mathbb{Q} = \xi_T d\mathbb{P}$ on $\mathscr{F}_T$ then*

$$d\boldsymbol{W}_t^{\mathbb{P}} = \phi_t dt + d\boldsymbol{W}_t^{\mathbb{Q}}, \tag{A.9}$$

where $\boldsymbol{W}_t^{\mathbb{Q}}$ is a $\mathbb{Q}$ Wiener process with correlation matrix $\boldsymbol{\Sigma}$.

Proof. The proof follows immediately by Cholesky decomposing the correlated Wiener vector $\boldsymbol{W}_t^{\mathbb{P}} = \boldsymbol{L}\bar{\boldsymbol{W}}_{t,}^{\mathbb{P}}$ where $\bar{\boldsymbol{W}}_{t,}^{\mathbb{P}}$ is a standard Wiener vector, and $\boldsymbol{L}\boldsymbol{L}^{\intercal} = \boldsymbol{\Sigma}$. □

A.6 The Market Price of Risk

Consider the case where we have N risky assets governed by the vector SDE system

$$d\boldsymbol{S}_t = \text{diag}(\boldsymbol{S}_t)[\boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\boldsymbol{W}_t^{\mathbb{P}}]$$

where $\boldsymbol{W}$ is a d-dimensional Wiener process with *independent components* and μ and $\boldsymbol{\sigma}$ respectively are N and $N \times d$ dimensional tensors adapted to the Wiener filtration. From equation (A.4) we know that under the risk free numeraire, S_0, $\mathbb{Q}$ is a martingale measure just if all tradable assets $\{S_0, S_1, ..., S_N\}$ have the short rate as their local rate of return:

$$d\boldsymbol{S}_t = \text{diag}(\boldsymbol{S}_t)[r_t \boldsymbol{\iota} dt + \boldsymbol{\sigma}_t d\boldsymbol{W}_t^{\mathbb{Q}}].$$

Girsanov's theorem informs us that the Wiener correlations are related by (A.9) so the question is, what is the kernel $\boldsymbol{\lambda}_t = -\boldsymbol{\phi}_t$ such that the drift changes as $\boldsymbol{\mu}_t \mapsto r_t\boldsymbol{\iota}$? From the last bullet point in the previous section, it is clear that $\boldsymbol{\lambda}_t$ must satisfy

$$\boldsymbol{\sigma}_t\boldsymbol{\lambda}_t = \boldsymbol{\mu}_t - r_t\boldsymbol{\iota}. \tag{A.10}$$

Clearly, the very existence of a risk neutral measure $\mathbb{Q}$ therefore necessitates that we can find a solution $\boldsymbol{\lambda}_t$ to this system. E.g, if $N < d$ then there are many solutions, one of which can be written as $\boldsymbol{\lambda}_t^* = \boldsymbol{\sigma}_t^{\intercal}(\boldsymbol{\sigma}_t\boldsymbol{\sigma}_t^{\intercal})^{-1}(\boldsymbol{\mu}_t - r_t\boldsymbol{\iota})$. On the other hand, if $N = d$ and $\boldsymbol{\sigma}$ is invertible then $\boldsymbol{\lambda}_t^* = \boldsymbol{\sigma}_t^{-1}(\boldsymbol{\mu}_t - r_t\boldsymbol{\iota})$ which is tantamount to the Sharpe ratio insofar as $\boldsymbol{\sigma}$ is the diagonal matrix $\mathrm{diag}(\sigma_1, ..., \sigma_N)$. In any case, we refer to $\boldsymbol{\lambda}$ as the **market price of risk** vector, which makes sense insofar that each λ_{jt} codifies the factor loading for the individual risk factor W_{jt}.

Theorem A.7. ***The Market Price of Risk***

- *Under absence of arbitrage, there will exist a market price of risk vector process $\boldsymbol{\lambda}_t$ satisfying $r_t\boldsymbol{\iota} = \boldsymbol{\mu}_t - \boldsymbol{\sigma}_t\boldsymbol{\lambda}_t$.*
- *The market price of risk $\boldsymbol{\lambda}_t$ is related to the Girsanov kernel through $\boldsymbol{\lambda}_t = -\boldsymbol{\phi}_t$ and thus to the risk neutral measure $\mathbb{Q}$ through*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left\{-\int_0^t \boldsymbol{\lambda}_s^{\intercal} d\boldsymbol{W}_s^{\mathbb{P}} - \tfrac{1}{2}\int_0^t ||\boldsymbol{\lambda}_s||^2 ds\right\}. \tag{A.11}$$

- *In a complete market, the market price of risk (or, alternatively, the martingale measure $\mathbb{Q}$) is uniquely determined and there is a unique price for every derivative.*
- *In an incomplete market there are several possible market prices of risk processes and several possible martingale measures which are consistent with no arbitrage.*
- *Thus, in an incomplete market $\{\boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbb{Q}\}$ are not determined by absence of arbitrage alone. Instead they will be determined by supply and demand on the market i.e. by the agents.*

Remark A.2. Take care to notice the condition that the components in $d\boldsymbol{W}^{\mathbb{P}}$ are independent. If this is not the case, i.e. if $d\boldsymbol{W}^{\mathbb{P}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}dt)$ for some $d \times d$ matrix $\boldsymbol{\Sigma}$, rewrite it as $d\boldsymbol{W}^{\mathbb{P}} = \boldsymbol{L}d\bar{\boldsymbol{W}}^{\mathbb{P}}$ where $d\bar{\boldsymbol{W}}^{\mathbb{P}}$ is a vector of i.i.d. Wiener increments and $\boldsymbol{L}$ is the lower triangular matrix arising from the Cholesky decomposetion $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^{\intercal}$. This has the effect that the market price of risk is defined through the equation $\boldsymbol{\sigma}_t\boldsymbol{L}_t\boldsymbol{\lambda}_t = \boldsymbol{\mu}_t - r_t\boldsymbol{\iota}$. In a complete market $N = d$ where $\boldsymbol{\sigma} = \mathrm{diag}(\sigma_1, ..., \sigma_N)$ this means that $\boldsymbol{\lambda}_t = \boldsymbol{L}^{-1}\boldsymbol{R}$ where $\boldsymbol{R}$ is the vector of Sharpe ratios: $([\mu_1 - r]/\sigma_1, ..., [\mu_N - r]/\sigma_N)$.

A.7 Changing the Numeraire

As it was strongly suggested in section A.3, there is no a priori reason why we should restrict ourselves to interpreting S_0 as the risk free asset in the First Fundamental Theorem as well as in the pricing equation (A.5). In fact, any **non-dividend paying tradeable asset** will do, although the martingale measures associated with each different numeraire will generally be distinct. To highlight this fact, we will write $\mathbb{Q}^0$ for a martingale measure under the numeraire S_0, $\mathbb{Q}^1$ for a martingale measure under the numeraire S_1 and so forth. We then have the following relationship between the different martingale measures

Theorem A.8. *Assume that $\mathbb{Q}^i$ is a martingale measure for the numeraire S_i on $\mathscr{F}_T$ and assume S_j is a positive asset price process such that S_{jt}/S_{it} is a true $\mathbb{Q}^i$ martingale (not just a local one). If we define $\mathbb{Q}^j$ on $\mathscr{F}_T$ by the likelihood process*

$$\xi_{it}^{j} = \frac{d\mathbb{Q}^j}{d\mathbb{Q}^i} = \frac{S_{i0}}{S_{j0}} \cdot \frac{S_{jt}}{S_{it}}, \quad 0 \le t \le T \tag{A.12}$$

then $\mathbb{Q}^j$ is a martingale measure for S_j.

Proof. The result follows by equation (A.3). Let X_t be an arbitrage free price process, then

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}^j}\left[\frac{X_{t'}}{S_{jt'}}\middle|\mathscr{F}_t\right] &= \mathbb{E}^{\mathbb{Q}^i}\left[\frac{\xi_{it'}^{j}}{\xi_{it}^{j}}\frac{X_{t'}}{S_{jt'}}\middle|\mathscr{F}_t\right] = \mathbb{E}^{\mathbb{Q}^i}\left[\frac{1}{\xi_{it}^{j}}\frac{S_{i0}}{S_{j0}}\frac{X_{t'}}{S_{it'}}\middle|\mathscr{F}_t\right] \\
&= \mathbb{E}^{\mathbb{Q}^i}\left[\frac{S_{j0}}{S_{i0}}\frac{S_{it}}{S_{jt}}\frac{S_{i0}}{S_{j0}}\frac{X_{t'}}{S_{it'}}\middle|\mathscr{F}_t\right] = \frac{S_{it}}{S_{jt}}\mathbb{E}^{\mathbb{Q}^i}\left[\frac{X_{t'}}{S_{it'}}\middle|\mathscr{F}_t\right] \\
&= \frac{S_{it}}{S_{jt}}\frac{X_t}{S_{it}} = \frac{X_t}{S_{jt}}.
\end{aligned}$$

So if $\mathbb{Q}^i$ is a martingale measure and $\mathbb{Q}^j$ is defined through ξ_i^j, then $\mathbb{Q}^j$ is a martingale measure. □

Theorem A.9. *Assume that the price processes obey the $\mathbb{Q}^i$ dynamics*

$$d\boldsymbol{S}_t = diag(\boldsymbol{S}_t)[\boldsymbol{\mu}_t^i dt + \boldsymbol{\sigma}_t d\boldsymbol{W}_t^{\mathbb{Q}^i}].$$

Then the $\mathbb{Q}^i$ dynamics of the likelihood process ξ_i^j is given by

$$d\xi_{it}^{j} = \xi_{it}^{j}(\boldsymbol{\sigma}_{jt}^{\intercal} - \boldsymbol{\sigma}_{it}^{\intercal})d\boldsymbol{W}_{it}.$$

In particular, the Girsanov kernel ϕ_i^j for the transition π^i to π^j is given by the volatility difference $\phi_{it}^j = \boldsymbol{\sigma}_{jt} - \boldsymbol{\sigma}_{it}$.

Proof. Apply Itō's lemma to (A.12) remembering that ξ_i^j is a $\mathbb{Q}^i$ martingale. □

Essentially, a numeraire change thus boils down to the following: we start out with the conventional pricing formula $X_t = \mathbb{E}_t^{\mathbb{Q}}[\frac{B_t}{B_T} X_T]$. Then we introduce the RN derivative $\xi_T = \frac{B_T}{B_0}\frac{S_0}{S_T}$, such that the pricing formula becomes (from the Abstract Bayes' Theorem):

$$X_t = \mathbb{E}_t^{\mathbb{Q}^S}[\tfrac{\xi_T}{\xi_t}\tfrac{B_t}{B_T} X_T] = \mathbb{E}_t^{\mathbb{Q}^S}[\tfrac{S_t}{S_T} X_T].$$

To get the dynamics of S_T under the $\mathbb{Q}^S$ measure we use Girsanov's Theorem. Specifically, since $\xi_T = d\mathbb{Q}/d\mathbb{Q}^S$ we know that $\eta_T := \xi_T^{-1} = d\mathbb{Q}^S/d\mathbb{Q}$ is such that $d\eta_t = \phi\eta_t dW_t^{\mathbb{Q}}$ (find this ϕ). Furthermore, $dW_t^{\mathbb{Q}} = \phi dt + dW_t^{\mathbb{Q}^S}$, which is what we need. For example, if we move from the bank ($B_t = e^{rt}$) to the stock numeraire ($S_t = S_0 \exp\{(r - \frac{1}{2}\sigma^2)t + \sigma W_t^{\mathbb{Q}}\}$), then we readily see $d\eta_t = \sigma\eta_t dW_t^{\mathbb{Q}}$, whence $dW_t^{\mathbb{Q}} = \sigma dt + dW_t^{\mathbb{Q}^S}$.

A.8 Dividend Paying Stocks

Consider the case where S_{nt} is the price process of a dividend paying asset, then we **cannot** use the First Fundamental Theorem to infer that S_{nt}/B_t is a martingale under the risk free measure $\mathbb{Q}$ (or more generally, that S_{nt}/S_{jt} is a martingale under the $\mathbb{Q}^j$ measure). It turns out that to generalise the martingale property, we must include the "sum" of all incremental changes in the deflated cumulative dividend, meaning:

Theorem A.10. ***Risk Neutral Valuation of Dividend Paying Assets*** *Let D_t be the cumulative dividend paid out by the asset S_n during the interval $[0,t]$. Then, under the risk neutral martingale measure $\mathbb{Q}$, the* ***normalised gain process***

$$G_t = \frac{S_{nt}}{B_t} + \int_0^t \frac{1}{B_s} dD_s$$

is a $\mathbb{Q}$-martingale.

Proof. We consider the dynamics of a self-financing portfolio which is long one unit of S_{nt} and where all dividends immediately are invested into the risk free bank account. Such a portfolio has the value process $\Pi_t = S_{nt} + X_t B_t$ where X_t denotes the instantaneous number of units of B_t. The point is, of course, that the portfolio can be viewed as a non-dividend paying asset, meaning that Π_t/B_t will be a $\mathbb{Q}$-martingale. Now, from Itō's lemma $d\Pi_t = dS_{nt} + X_t dB_t + B_t dX_t$. Combining this with the self-financing condition $d\Pi_t = dS_{nt} + dD_t + X_t dB_t$ we find that $dX_t = B_t^{-1} dD_t$. I.e.

$$\Pi_t = S_{nt} + \int_0^t B_s^{-1} B_t dD_s,$$

which will be a $\mathbb{Q}$ martingale upon being deflated by B_t. □

Theorem A.11. ***General Valuation of Dividend Paying Assets*** *Assume now S_{nt} is an asset associated with the cumulative dividend D_t, and let S_{jt} be the price process of a* ***non-dividend*** *paying asset. Assuming absence of arbitrage we denote the martingale measure for the numeraire S_j by $\mathbb{Q}^j$ then the following holds*

- *The normalised gain process G defined by*

$$G_t = \frac{S_{nt}}{S_{jt}} + \int_0^t \frac{1}{S_{js}} dD_s - \int_0^t \frac{1}{S_{js}^2} dD_s dS_{js}$$

  is a $\mathbb{Q}^j$ martingale.
- *If the dividend process D has no driving Wiener component (or more generally, if $dDdS_j = 0$) then the last term vanishes.*

Appendix B
PDE Methods in Mathematical Finance

Abstract In this appendix we briefly review some of the standard PDEs encountered in mathematical finance alongside their (probabilistic) solutions. Furthermore, we expose how a class of these PDEs may be solved numerically using finite difference methods. Particular emphasis is here laid on Llewellyn Thomas' tridiagonal matrix algorithm, which we generalise to higher dimensional (block matrix) cases.

B.1 From Martingales to PDEs

B.1.1 The Feynman-Kac Formula

Rather than solving the expected value problems inherent to much of mathematical finance, we may choose to reformulate them in deterministic terms qua partial differential equations. What bridges the gap between these fields is a fundamental result developed by Richard Feynman and Mark Kac:

Theorem B.1. ***The Feynman-Kac Formula*** *Suppose $f : \mathbb{R} \times [0,T] \mapsto \mathbb{R}$ satisfies the PDE*

$$\frac{\partial f}{\partial t}(x,t) + \mu^{\mathbb{Q}}(x,t)\frac{\partial f}{\partial x}(x,t) + \tfrac{1}{2}\sigma^2(x,t)\frac{\partial^2 f}{\partial x^2}(x,t) - r(x,t)f(x,t) + h(x,t) = 0,$$
$$\text{s.t.: } f(x,T) = \psi(x).$$

where $\mu^{\mathbb{Q}}, \sigma, r$ and h are known functions. Then the solution may be written as

$$f(x,t) = \mathbb{E}^{\mathbb{Q}}_{X_t=x}\left[\int_t^T D_t^u h(X_u,u)du + D_t^T \psi(x)\right], \tag{B.1}$$

where we have defined the discounting factor $D_t^{t'} := e^{-\int_t^{t'} r(X_s,s)ds}$ *and* X_t *is a stochastic process defined on the filtered probability space* $(\Omega, \mathscr{F}, \mathbb{Q}, \mathbb{F})$ *which follows the SDE*

$$dX_t = \mu^{\mathbb{Q}}(X,t)dt + \sigma(X_t,t)dW_t^{\mathbb{Q}}, \quad X_0 = x. \tag{B.2}$$

Proof. We imagine that f solves the PDE listed above. Furthermore, define the process

$$Y_v := \int_t^v D_t^u h(X_u,u)du + D_t^v f(X_v,v).$$

From Itō's lemma

$$dY_v = d\int_t^v D_t^u h(X_u,u)du + dD_t^v f(X_v,v) + D_t^v df(X_v,v) + dD_t^v df(X_v,v).$$

Now we make the following observations:

- $dD_t^v = -r(X_v,v)D_t^v dv$. In particular, the fourth term is zero.
- The first term can be written as $d\int_t^v D_t^u h(X_u,u)du = D_t^v h(X_v,v)dv$.
- From Itō's lemma $df = f_v dv + \mu^{\mathbb{Q}} f_x dX + \frac{1}{2}\sigma^2 f_{xx} dX^2$.

Bringing these insights together we obtain

$$\begin{aligned} dY_v =& D_t^v \Big[h(X_v,v) - r(X_v,v)f(X_v,v) + \frac{\partial f}{\partial v}(X_v,v) + \mu^{\mathbb{Q}}(X_v,v)\frac{\partial f}{\partial x}(X_v,v) \\ &+ \tfrac{1}{2}\sigma^2(X_v,v)\frac{\partial^2 f}{\partial x^2}(X_v,v)\Big] dv + D_t^v \sigma(X_v,v)\frac{\partial f}{\partial x}(X_v,v)dW_v^{\mathbb{Q}} \end{aligned}$$

By definition of f, the square bracket is zero. Hence, after integrating we are left with

$$Y_T = Y_t + \int_t^T D_t^v \sigma(X_v,v)\frac{\partial f}{\partial x}(X_v,v)dW_v^{\mathbb{Q}}.$$

Taking $\mathbb{Q}$ expectations conditional on $X_t = x$ and using the fact that the Itō integral is a martingale we get

$$\mathbb{E}^{\mathbb{Q}}_{X_t=x}[Y_T] = Y_t = f(x,t),$$

which is exactly what needed to be shown. □

B.1.2 The Kolmogorov Backward Equation

In a certain sense, the PDE in Theorem B.1 *is* the backward equation: i.e. given the dynamics (B.2) we may infer that the quantity (B.1) satisfies the listed PDE. Nonetheless, when discussing the Kolmogorov backward equation, one usually thinks of a governing equation of *transition probabilities*.

Theorem B.2. ***The Kolmogorov Backward Equation*** *Consider the filtered probability space $(\Omega, \mathscr{F}, \mathbb{Q}, \mathbb{F})$ and let X be a stochastic process on this space, which solves the SDE*

$$dX_t = \mu^{\mathbb{Q}}(X_t, t)dt + \sigma(X_t, t)dW_t^{\mathbb{Q}}.$$

Now consider the probability that X_s is within some set Y given that at time $t < s$ $X_t = x$, i.e.

$$\mathbb{Q}(X_s \in Y | X_t = x) = \int_{y \in Y} \rho(y, s|x, t)dy,$$

where ρ is the density function. Then ρ satisfies the PDE

$$\left\{ \frac{\partial}{\partial t} + \mu^{\mathbb{Q}}(x,t)\frac{\partial}{\partial x} + \tfrac{1}{2}\sigma^2(x,t)\frac{\partial^2}{\partial x^2} \right\} \rho(y,s|x,t) = 0,$$

for $(x,t) \in \mathbb{R} \times (0,s)$ and where $\rho(y,s|x,t) \to \delta(y)$ as $t \nearrow s$.

Proof. We could consider the boundary value problem

$$\frac{\partial f}{\partial t}(x,t) + \mu^{\mathbb{Q}}(x,t)\frac{\partial f}{\partial x}(x,t) + \tfrac{1}{2}\sigma^2(x,t)\frac{\partial^2 f}{\partial x^2}(x,t) = 0, \qquad (x,t) \in \mathbb{R} \times (0,s)$$

$$f(x,s) = \mathbf{1}\{x \in Y\}, \qquad x \in \mathbb{R}$$

where $\mathbf{1}$ is the indicator function. From Feynman-Kac we deduce the solution

$$f(x,t) = \mathbb{E}^{\mathbb{Q}}_{X_t = x}[\mathbf{1}\{X_s \in Y\}].$$

But this is, of course, just equal to

$$f(x,t) = \mathbb{Q}(X_s \in Y | X_t = x) = \int_{y \in Y} \rho(y,s|x,t)dy$$

Since this argument works in both directions, we have effectively already verified the Backward theorem. In particular, substituting the transition density form of the solution into the PDE we get

$$\int_{y\in Y}\left\{\frac{\partial}{\partial t}+\mu^{\mathbb{Q}}(x,t)\frac{\partial}{\partial x}+\tfrac{1}{2}\sigma^2(x,t)\frac{\partial^2}{\partial x^2}\right\}p(y,s|x,t)dy=0, \qquad (x,t)\in\mathbb{R}\times(0,s)$$

$$\int_{y\in Y}p(y,s|x,s)dy=\mathbf{1}\{x\in Y\}, \quad x\in\mathbb{R}.$$

But this holds for any set Y so we can drop the integrals (whence the indicator function becomes a Dirac delta function). □

B.1.3 The Kolmogorov Forward Equation

In the backward equation we observe that the differential operator operates on the variables upon which we condition (the "backward variables" if you will), i.e. (x,t). A corresponding result exists for the "forward variables" viz. (y,s):

Theorem B.3. ***The Kolmogorov Forward Equation (a.k.a.The Fokker-Planck Equation)*** *Consider the filtered probability space $(\Omega,\mathscr{F},\mathbb{Q},\mathbb{F})$ and let X be a stochastic process on this space, which solves the SDE*

$$dX_t=\mu^{\mathbb{Q}}(X_t,t)+\sigma(X_t,t)dW_t^{\mathbb{Q}}.$$

Now consider the probability that X_s is within some set Y given that at time $t<s$ $X_t=x$, i.e.

$$\mathbb{Q}(X_s\in Y|X_t=x)=\int_{y\in Y}\rho(y,s|x,t)dy$$

where ρ is the density function. Then ρ satisfies the PDE

$$\frac{\partial\rho}{\partial s}(y,s|x,t)+\frac{\partial}{\partial y}(\mu^{\mathbb{Q}}(y,s)\rho(y,s|x,t))-\tfrac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y,s)\rho(y,s|x,t))=0$$

for $(y,s)\in\mathbb{R}\times(0,T)$ and where $\rho(y,s|x,t)\to\delta(x)$ as $s\searrow t$.

Proof. Consider a test function $h(y,s)\in\mathscr{C}^{2,1}$ with compact support[1] in the set $\mathbb{R}\times(t,T)$ where $t<T$ are two fixed coordinates in time. From Itō's lemma,

$$\begin{aligned}h(X_T,T)=&h(X_t,t)+\int_t^T\left\{\frac{\partial}{\partial s}+\mu^{\mathbb{Q}}(y,s)\frac{\partial}{\partial y}+\tfrac{1}{2}\sigma^2(y,s)\frac{\partial^2}{\partial y^2}\right\}h(X_s,s)ds\\&+\int_t^T\frac{\partial h}{\partial y}(X_s,s)dW_s^{\mathbb{Q}}\end{aligned}$$

[1] I.e. the function is non-zero only in a closed, bounded region.

Applying the expectation operator $\mathbb{E}^{\mathbb{Q}}_{X_t=x}[\cdots]$ and using the compact support condition $(h(x,T)=h(x,t)=0)$ we obtain

$$\int_{\mathbb{R}}\int_t^T \rho(y,s|x,t)\left\{\frac{\partial}{\partial s}+\mu^{\mathbb{Q}}(y,s)\frac{\partial}{\partial y}+\tfrac{1}{2}\sigma^2(y,s)\frac{\partial^2}{\partial y^2}\right\}h(y,s)dsdy=0$$

Integrating by parts "in time" we find that

$$\begin{aligned}\int_{\mathbb{R}}\int_t^T \rho(y,s|x,t)\frac{\partial h}{\partial s}(y,s)dsdy &= \int_{\mathbb{R}}\left\{[\rho h]_t^T-\int_t^T\frac{\partial\rho}{\partial s}hds\right\}dy\\ &= -\int_{\mathbb{R}}\int_t^T h(y,s)\frac{\partial\rho}{\partial s}(y,s|x,t)dsdy\end{aligned}$$

Likewise, using Fubini's rule and integrating by parts in "state space"

$$\begin{aligned}&\int_{\mathbb{R}}\int_t^T \rho(y,s|x,t)\left\{\mu^{\mathbb{Q}}(y,s)\frac{\partial}{\partial y}+\tfrac{1}{2}\sigma^2(y,s)\frac{\partial^2}{\partial y^2}\right\}h(y,s)dsdy\\ &=\int_t^T\int_{\mathbb{R}}\rho(y,s|x,t)\mu^{\mathbb{Q}}(y,s)\frac{\partial h}{\partial y}(y,s)dyds+\tfrac{1}{2}\int_t^T\int_{\mathbb{R}}\rho(y,s|x,t)\sigma^2(y,s)\frac{\partial^2 h}{\partial y^2}(y,s)dyds\\ &=\int_t^T\left\{[\rho\mu^{\mathbb{Q}}h]_{-\infty}^{\infty}-\int_{\mathbb{R}}\frac{\partial(\rho\mu^{\mathbb{Q}})}{\partial y}hdy\right\}ds+\tfrac{1}{2}\int_t^T\left\{\left[\rho\sigma^2\frac{\partial h}{\partial y}\right]_{-\infty}^{\infty}-\int_{\mathbb{R}}\frac{\partial\rho\sigma^2}{\partial y}\frac{\partial h}{\partial y}dy\right\}ds\\ &=-\int_t^T\int_{\mathbb{R}}\frac{\partial(\rho\mu^{\mathbb{Q}})}{\partial y}hdyds-\tfrac{1}{2}\int_t^T\left\{\left[\frac{\partial\rho\sigma^2}{\partial y}h\right]_{-\infty}^{\infty}-\int_{\mathbb{R}}\frac{\partial^2\rho\sigma^2}{\partial y^2}hdy\right\}ds\\ &=-\int_{\mathbb{R}}\int_t^T h(y,s)\left\{\frac{\partial(\rho(y,s|x,t)\mu^{\mathbb{Q}}(y,s))}{\partial y}h-\tfrac{1}{2}\frac{\partial^2(\rho(y,s|x,t)\sigma^2(y,s))}{\partial y^2}\right\}dyds\end{aligned}$$

Combining the last three results, we have

$$\int_{\mathbb{R}}\int_t^T h(y,s)\left\{\frac{\partial\rho}{\partial s}(y,s|x,t)+\frac{\partial(\rho(y,s|x,t)\mu^{\mathbb{Q}}(y,s))}{\partial y}h-\tfrac{1}{2}\frac{\partial^2(\rho(y,s|x,t)\sigma^2(y,s))}{\partial y^2}\right\}dyds=0.$$

Indeed, since h was chosen as an arbitrary test function, the result now follows. □

B.2 Solving PDEs Through Finite Difference Methods

The story of how to set up and solve PDEs in option pricing using finite difference methods is one that has echoed into the furthest corners of quantitative finance. It is thus not altogether unreasonable to treat this subject laconically and in recapitulatory terms, under the assumption that the reader already has established a modicum of familiarity with the field. Those who seek a deeper understanding of the intricacies of finite difference methodologies are referred to the specialist literature for details.

B.2.1 Numerical Solutions to PDEs in One Spatial Dimension

Suppose we wish to price a contingent claim on some underlying quantity which can be modelled as a stochastic differential equation. Specifically, assume that the risk neutral dynamics of the underlying follows some generic local volatility model

$$dX_t = \mu(X_t,t)dt + \sigma(X_t,t)dW_t^{\mathbb{Q}}$$

for $t \in [0,T]$, where μ is the risk-neutral drift rate and σ denotes the volatility. Furthermore, let the terminal pay-off of the contingent claim be of the form $\Phi(X_T)$. If we let $f(x,t)$: $\mathbb{R}^+ \times [0,T] \mapsto \mathbb{R}$ denote the time t value of the contingent claim, where $X_t = x$, then a simple replicating argument shows that f obeys the PDE

$$\frac{\partial f}{\partial t}(x,t) + \mu(x,t)\frac{\partial f}{\partial x}(x,t) + \tfrac{1}{2}\sigma(x,t)^2\frac{\partial^2 f}{\partial x^2}(x,t) - r(x,t)f(x,t) = 0, \qquad \text{(B.3)}$$
$$(x,t) \in \mathbb{R}^+ \times [0,T),$$

with terminal condition $f(x,T) = \Phi(x)$, $\forall x \in \mathbb{R}_+$, where $r(x,t)$ has the interpretation of a local risk free rate. Generally, (B.3) does not admit analytical solutions, which prompts us to seek speedy and stable numerical solution algorithms instead. The basic idea is here to rewrite the governing PDE as a system of difference equations which can be solved iteratively backwards in time. To this end, we discretise the state-space $\mathbb{R}_+ \times [0,T]$ into a rectangular grid of J spatial coordinates $\mathbb{X} \equiv \{x_{\min} \equiv x_1, x_2, \ldots, x_{J-1}, x_{\max} \equiv x_J\}$ (s.t. $\forall j : x_{j+1} - x_j = \Delta x$) and $N+1$ temporal coordinates $\mathbb{T} \equiv \{t_0 \equiv 0, t_1, \ldots, t_{N-1}, t_N \equiv T\}$ (s.t. $\forall n : x_{n+1} - x_n = \Delta t$), where the upper and lower spatial boundaries on the grid are chosen to be plausible upper and lower bounds for the underlying process. To square the governing PDE (B.3) with this discretised space $\mathbb{X} \times \mathbb{T}$ we perform the obvious identifications $f_j^n \equiv f(x_j, t_n)$ (and similarly for μ_j^n, σ_j^n and r_j^n). It is somewhat less transparent what constitutes adequate discrete analogues for the first and second derivatives in space and time - indeed, the conventions for this vary between different numerical schemes. One particularly obvious choice is the **implicit method** in which

$$\frac{\partial f}{\partial t}(x_j,t_n) \approx \frac{f_j^{n+1} - f_j^n}{\Delta t}, \qquad \frac{\partial f}{\partial x}(x_j,t_n) \approx \frac{f_{j+1}^n - f_{j-1}^n}{2\Delta x}, \qquad \text{(B.4)}$$
$$\frac{\partial^2 f}{\partial x^2}(x_j,t_n) \approx \frac{f_{j+1}^n - 2f_j^n + f_{j-1}^n}{(\Delta x)^2}.$$

Nonetheless, our concern is not with these petty details, but rather the grander picture: what matters is that the classical schemes after some reshuffling of terms result in a discretised PDE of the form

$$a_j^n f_{j-1}^n + b_j^n f_j^n + c_j^n f_{j+1}^n = d_j^{n+1}, \qquad \text{(B.5)}$$

for $1 < j < J$ and $0 \le n < N$. Here, the coefficients a_j^n, b_j^n and c_j^n will generally be functions of $(r_j^n, \mu_j^n, \sigma_j^n, \Delta t, \Delta x)$. Analogously, the quantity d_j^{n+1} depends on the corresponding t_{n+1} counterparts as well as on $(f_{j-1}^{n+1}, f_j^{n+1}, f_{j+1}^{n+1})$.

Since the terminal condition of the PDE is known ($\forall j : f_j^N = \Psi(x_j)$) d_j^N can be computed. To determine the time $t_1 = 0$ value of $\mathbf{f} = (f_1, f_2, ..., f_J)$ we must therefore proceed along the following backwards iterative scheme: First, determine $\mathbf{f}^{N-1} = (f_1^{N-1}, f_2^{N-1}, ..., f_J^{N-1})$. From (B.5) we immediately have $J-2$ simultaneous linear equations involving the J unknowns "$\mathbf{f}^{N-1}$", viz.

$$\begin{aligned}
a_2^{N-1} f_1^{N-1} + b_2^{N-1} f_2^{N-1} + c_2^{N-1} f_3^{N-1} &= d_2^N, \\
a_3^{N-1} f_2^{N-1} + b_3^{N-1} f_3^{N-1} + c_3^{N-1} f_4^{N-1} &= d_3^N, \\
&\vdots \\
a_{J-1}^{N-1} f_{J-2}^{N-1} + b_{J-1}^{N-1} f_{J-1}^{N-1} + c_{J-1}^{N-1} f_J^{N-1} &= d_{J-2}^N.
\end{aligned}$$

To solve this system one must either fix two of the unknowns or add two additional equations, linearly independent of all the others. Typically, this involves thinking carefully about the nature of the external boundaries in the grid ($j = 0$ and $j = J+1$). E.g. financial derivatives which are deep in or out of the money might have obvious prices $\forall t < T$ or we can utilise the fact that their gammas will be close to zero. Either way, we will assume that we can meaningfully add the two extra conditions

$$\begin{aligned}
a_1^n f_0^n + b_1^n f_1^n + c_1^n f_2^n &= d_1^{n+1}, \\
a_J^n f_{J-1}^n + b_J^n f_J^n + c_J f_{J+1}^n &= d_J^{n+1},
\end{aligned} \tag{B.6}$$

$\forall n < N$, where f_0^n and f_{J+1}^n (in some form or the other) are known quantities. Thus we have obtained a full system of J simultaneous equations in J unknowns, wherefore we can solve for $\mathbf{f}^{N-1}$. Once this is done, the algorithm is repeated from the top: i.e. we determine $\mathbf{f}^{N-2}$ using equation (B.5) coupled with two boundary conditions and our knowledge of $\mathbf{f}^{N-1}$, ... and so forth till we reach the desired $\mathbf{f}^0$.

B.2.2 Thomas' Algorithm

System (B.5) coupled with the boundary equations (B.6) constitute a triangular linear matrix system:

$$\begin{pmatrix}
b_1^n & c_1^n & 0 & 0 & 0 & \cdots & 0 \\
a_2^n & b_2^n & c_2^n & 0 & 0 & \cdots & 0 \\
0 & a_3^n & b_3^n & c_3^n & 0 & \cdots & 0 \\
\vdots & & \ddots & \ddots & \ddots & & \vdots \\
0 & \cdots & 0 & 0 & a_{J-1}^n & b_{J-1}^n & c_{J-1}^n \\
0 & \cdots & 0 & 0 & 0 & a_J^n & b_J^n
\end{pmatrix}
\begin{pmatrix} f_1^n \\ f_2^n \\ f_3^n \\ \vdots \\ f_{J-1}^n \\ f_J^n \end{pmatrix}
=
\begin{pmatrix} d_1^{n+1} \\ d_2^{n+1} \\ d_3^{n+1} \\ \vdots \\ d_{J-1}^{n+1} \\ d_J^{n+1} \end{pmatrix}
-
\begin{pmatrix} a_1^n f_0^n \\ 0 \\ 0 \\ \vdots \\ 0 \\ c_J f_{J+1}^n \end{pmatrix},$$

which succinctly henceforth will be written

$$\mathbf{A}^n\mathbf{f}^n = \mathbf{d}_*^{n+1}, \tag{B.7}$$

where $\mathbf{A}^n \in \mathbb{R}^{J\times J}$ is the tridiagonal matrix, $\mathbf{f}^n \in \mathbb{R}^J$ is as defined above, and $\mathbf{d}_*^{n+1} \in \mathbb{R}^J$ encodes *both* of the vectors on the righthand side.

The nitty-gritty of this section may now be stated as follows: if we naïvely solve for $\mathbf{f}^n$ in (B.7) by computing $[\mathbf{A}^n]^{-1}$ using *Gaussian elimination with backward substitution* we will incur $\frac{1}{3}J^3 + J^2 - \frac{1}{3}J$ multiplications & divisions and $\frac{1}{3}J^3 + \frac{1}{2}J^2 - \frac{5}{6}J$ additions & subtractions [4]. The totality of binary operations would concordantly scale as $\mathscr{O}(J^3)$. Nonetheless, this is in fact computationally redundant as any (strictly diagonally dominant)[2] tridiagonal matrix system may be solved without computing the inverse matrix $[\mathbf{A}^n]^{-1}$ explicitly. Specifically, by adopting *Thomas' algorithm* for triangular matrix systems the number of binary operations may be brought down to scale as $\mathscr{O}(J)$.

The details of Thomas' scheme are straightforward: **first** an LU-decomposition of $\mathbf{A}^n$ is performed, i.e.

$$\mathbf{A}^n = \mathbf{L}^n\mathbf{U}^n,$$

where $\mathbf{L}^n$ is a lower triangular matrix and $\mathbf{U}^n$ is an upper triangular matrix:

$$\mathbf{L}^n = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ l_2^n & 1 & 0 & 0 & \dots & 0 \\ 0 & l_3^n & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & 0 & l_{J-1}^n & 1 & 0 \\ 0 & \dots & 0 & 0 & l_J^n & 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{U}^n = \begin{pmatrix} h_1^n & u_1^n & 0 & 0 & \dots & 0 \\ 0 & h_2^n & u_2^n & 0 & \dots & 0 \\ 0 & 0 & h_3^n & u_3^n & & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & & 0 & 0 & h_{J-1}^n & u_{J-1}^n \\ 0 & \dots & 0 & 0 & 0 & h_J^n \end{pmatrix}. \tag{B.8}$$

Notice that $\mathbf{A}^n$ and $\mathbf{L}^n\mathbf{U}^n$ have the same number of degrees of freedom, $3J-2$. By performing the multiplication explicitly

$$\mathbf{L}^n\mathbf{U}^n = \begin{pmatrix} h_1^n & u_1^n & 0 & 0 & \dots & 0 \\ l_2^n h_1^n & l_2^n u_1^n + h_2^n & u_2^n & 0 & \dots & 0 \\ 0 & l_3^n h_2^n & l_3^n u_2^n + h_3^n & u_3^n & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & & 0 & l_{J-1}^n h_{J-2}^n & l_{J-1}^n u_{J-2}^n + h_{J-1}^n & u_{J-1}^n \\ 0 & \dots & 0 & 0 & l_J^n h_{J-1}^n & l_J^n u_{J-1}^n + h_J^n \end{pmatrix},$$

and comparing element-wise with $\mathbf{A}^n$ we immediately see that $a_j^n = l_j^n h_{j-1}^n \ \forall j \in [2,J]$, $b_1^n = h_1^n$ and $b_j^n = l_j^n u_{j-1}^n + h_j^n \ \forall j \in [2,J]$ and finally $c_j^n = u_j^n \ \forall j \in [1, J-1]$. Thus, we may determine h_j^n, u_j^n and l_j^n by moving from left to right and downward through the triangular matrix, giving us the *first part* of Thomas' algorithm:

---

[2] A square matrix $\mathbf{M} \in \mathbb{R}^{J\times J}$ is strictly diagonally dominant if $\forall i : |M_{ii}| > \sum_{j\neq i} |M_{ij}|$. This is a sufficient condition [Burden & Faires, Theorem 6.21] for $\mathbf{M}$ to be non-singular.

Thomas' Algorithm, Part I: The LU decomposition:

Set	h_1^n	$:=$	b_1^n;
	u_1^n	$:=$	c_1^n;
For $j = 2,...,J-1$ set	l_j^n	$:=$	a_j^n/h_{j-1}^n;
	h_j^n	$:=$	$b_j^n - l_j^n u_{j-1}^n$;
	u_j^n	$:=$	c_j^n;
Set	l_J^n	$:=$	a_J^n/h_{J-1}^n;
	h_J^n	$:=$	$b_J^n - l_J^n u_{J-1}^n$;

The **second** and final step in the algorithm is to solve the system

$$\mathbf{L}^n\mathbf{U}^n\mathbf{f}^n = \mathbf{d}_*^{n+1},$$

which clearly does not require a full matrix inversion. Specifically, define $\mathbf{y}^n$ as the solution to the system $\mathbf{L}^n\mathbf{y}^n = \mathbf{d}_*^{n+1}$. Written in component form, it is evident that $y_1^n = d_{*1}^{n+1}$ and $l_j^n y_{j-1}^n + y_j^n = d_{*j}^{n+1}$ for $j \in [2,J]$, hence all components in $\mathbf{y}^n$ are determined simply by looking at the first equation in the system and then moving incrementally downwards. Now, by comparison, $\mathbf{y}^n = \mathbf{U}^n\mathbf{f}^n$ so we have, in fact, also facilitated an easy evaluation of $\mathbf{f}^n$. To see this, notice that $h_j^n f_j^n + u_j^n f_{j+1}^n = y_j^n$ for $j \in [1,J-1]$ and $h_J^n f_J^n = y_J^n$ so we must simply start with the last equation and work our way incrementally upwards. In algorithmic terms:

Thomas' Algorithm, Part II: Solving $\mathbf{L}\mathbf{y} = \mathbf{d}_*$; $\mathbf{U}\mathbf{x} = \mathbf{y}$:

Set	y_1^n	$:=$	d_{*1}^{n+1};
For $j = 2,...,J$ set	y_j	$:=$	$d_{*j}^{n+1} - l_j^n y_{j-1}^n$;
Set	f_J	$:=$	y_J^n/h_J^n;
For $j = J-1,...,1$ set	f_j^n	$:=$	$(y_j^n - u_j^n f_{j+1}^n)/h_j^n$;

As a complete specification of $\mathbf{f}^n$ has been obtained, this concludes the algorithm. Notice that $2J-2$ multiplications & divisions and $J-1$ additions & subtractions are done in Part I of the algorithm and that $3J-2$ multiplications & divisions and $2J-2$ additions & subtractions are done in Part II of the algorithm. The total number of binary operations needed therefore scales as $\mathscr{O}(J)$, which is a substantial improvement over a Gaussian procedure.

B.2.2.1 A Restatement

In practice, one often sees the algorithm expressed in the following (equivalent) form

Thomas' Algorithm:

$$
\begin{array}{lrcl}
\text{Set} & \gamma_1^n & := & c_1^n/b_1^n; \\
 & \delta_1^n & := & d_{*1}^{n+1}/b_1^n; \\
\text{For } j=2,...,J-1 \text{ set} & q_j^n & := & b_j^n - a_j^n \gamma_{j-1}^n \\
 & \gamma_j^n & := & c_j^n/q_j^n; \\
 & \delta_j^n & := & (d_{*j}^{n+1} - a_j^n \delta_{j-1}^n)/q_j^n; \\
\text{Set} & \delta_J^n & := & (d_{*J}^{n+1} - a_J^n \delta_{J-1}^n)/(b_J^n - a_J^n \gamma_{J-1}^n); \\
 & f_J^n & := & \delta_J^n; \\
\text{For } j=J-1,...,1 \text{ set} & f_j^n & := & \delta_j^n - \gamma_j^n f_{j+1}^n;
\end{array}
$$

Proof. To see the equivalence, notice that $f_1^n = (y_1^n - u_1^n f_2^n)/h_1^n = (d_{*1}^{n+1} - c_1^n f_2^n)/b_1^n = \delta_1^n - \gamma_1^n f_2^n$. Analogously, for $j = 2,...,J-1$,

$$
\begin{aligned}
f_j^n = \frac{y_j^n}{h_j^n} - \frac{u_j^n}{h_j^n} f_{j+1}^n &= \frac{d_{*j}^{n+1} - l_j^n y_{j-1}^n}{b_j^n - l_j^n u_{j-1}^n} - \frac{c_j^n}{b_j^n - l_j^n u_{j-1}^n} f_{j+1}^n \\
&\stackrel{\diamond}{=} \frac{d_{*j}^{n+1} - a_j^n [y_{j-1}^n/h_{j-1}^n]}{b_j^n - a_j^n [u_{j-1}^n/h_{j-1}^n]} - \frac{c_j^n}{b_j^n - a_j^n [u_{j-1}^n/h_{j-1}^n]} f_{j+1}^n
\end{aligned}
$$

We now make the obvious observation that when $j = 2$ then $[y_{j-1}^n/h_{j-1}^n] = \delta_1^n$ and $[u_{j-1}^n/h_{j-1}^n] = \gamma_1^n$. Comparing the coefficients in $\diamond$ with the definitions of δ_j^n and γ_j^n above, we therefore immediately find that $f_j^n = \delta_j^n - \gamma_j^n f_{j+1}^n$. Finally, in the event $j = J$: $f_J^n = y_J^n/h_J^n = (d_{*J}^{n+1} - l_J^n y_{J-1}^n)/(a_J^n - l_J^n u_{J-1}^n)$. Sub in $l_J^n = a_J^n/h_{J-1}^n$ to obtain $f_J^n = (d_{*J}^{n+1} - a_J^n [y_{J-1}^n/h_{J-1}])/(a_J^n - a_J^n [u_{J-1}^n/h_{J-1}^n]) = (d_{*J}^{n+1} - a_J^n \delta_{J-1}^n)/(a_J^n - a_J^n \gamma_{J-1}^n) = \delta_J^n$. $\square$

Remark: The computational complexity (binary count) is invariant under this rephrasing: we still find $5J-4$ multiplications & divisions and $3J-3$ additions & subtractions.

B.2.3 The Multi-dimensional Case

It turns out that a generalisation of the above is quite possible to multiple dimension. For simplicity we will illustrate this point by considering the case of two spatial dimensions. Specifically, suppose we have a contingent claim which depends on two, possibly correlated, underlying factors X and Y which can be modelled as stochastic differential equations. Specifically, let us assume a governing risk neutral dynamics of the form

$$
\begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} \mu_1(X_t,Y_t,t) \\ \mu_2(X_t,Y_t,t) \end{pmatrix} dt + \begin{pmatrix} \sigma_1(X_t,Y_t,t) & 0 \\ 0 & \sigma_2(X_t,Y_t,t) \end{pmatrix} \begin{pmatrix} dW_{1t}^{\mathbb{Q}} \\ dW_{2t}^{\mathbb{Q}} \end{pmatrix},
$$

where $dW_{1t}^{\mathbb{Q}}$ and $dW_{2t}^{\mathbb{Q}}$ have a Pearson correlation coefficient of ρ. Letting the terminal pay-off of the contingent claim be $\Psi(X_T, Y_T)$ then if $f(x,y,t) : \mathbb{R}_+ \times \mathbb{R}_+ \times [0,T] \mapsto \mathbb{R}$ denotes the time t value of the claim, it follows by a replicating argument that f satisfies

$$\begin{gathered}\frac{\partial f}{\partial t}(x,y,t) + \mu_1(x,y,t)\frac{\partial f}{\partial x}(x,y,t) + \mu_2(x,y,t)\frac{\partial f}{\partial y}(x,y,t) + \tfrac{1}{2}\left\{\sigma_1(x,y,t)^2\frac{\partial^2 f}{\partial x^2}(x,y,t)\right. \\ \left. + 2\rho\sigma_1(x,y,t)\sigma_2(x,y,t)\frac{\partial^2 f}{\partial x \partial y}(x,y,t) + \sigma_2(x,y,t)^2\frac{\partial^2 f}{\partial y^2}(x,y,t)\right\} - r(x,y,t)f(x,y,t) = 0, \\ (x,y,t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times [0,T),\end{gathered} \tag{B.9}$$

where $f(x,y,T) = \Psi(x,y)$, $\forall (x,y) \in \mathbb{R}_+ \times \mathbb{R}_+$.

We discretise the problem by considering a rectangular cuboid consisting of J x-coordinates, K y-coordinates and $N+1$ temporal coordinates *in lieu* of the full state space $\mathbb{R}_+ \times \mathbb{R}_+ \times [0,T]$. Again, we assume equidistance $\forall j : \Delta x = x_{j+1} - x_j$, $\forall k : \Delta y = y_{k+1} - y_k$ and $\forall n : \Delta t = t_{n+1} - t_n$ and that X and Y take on values greater than some upper boundaries and lower than some lower boundaries with negligible probability. Ergo, the discretised space is of the from $\mathbb{X} \times \mathbb{Y} \times \mathbb{T} = \{x_{\min} = x_1, x_2 = x_1 + \Delta x, ..., x_{\max} \equiv x_J\} \times \{y_{\min} = y_1, y_2 = y_1 + \Delta y, ..., y_{\max} \equiv y_K \times \{t_0 = 0, t_1 = t_0 + \Delta t, ..., t_N = t_0 + N\Delta t \equiv T\}$. Defining $f_{j,k}^n \equiv f(x_j, y_k, t_n)$ etc. the discretisation of the PDE runs almost exactly as in the single (spatial) variable case. The only thing novel about our problem is the existence of a cross second order spatial derivative, yet that can also be handled:

$$\frac{\partial^2 f}{\partial x \partial y}(x_j, y_k, t_n) \approx \frac{f_{j+1,k+1}^n - f_{j+1,k-1}^n - f_{j-1,k+1}^n + f_{j-1,k-1}^n}{4\Delta x \Delta y}. \tag{B.10}$$

Of course, our concern is still with the broader picture and as it happens the classical, reasonably stable, numerical schemes (the implicit method, Crank-Nicholson,...) now give rise to a discretised PDE of the form

$$\begin{gathered}a_{j,k}^n f_{j-1,k-1}^n + b_{j,k}^n f_{j-1,k}^n + c_{j,k}^n f_{j-1,k+1}^n + \hat{a}_{j,k}^n f_{j,k-1}^n + \hat{b}_{j,k}^n f_{j,k}^n + \\ \hat{c}_{j,k}^n f_{j,k+1}^n + \tilde{a}_{j,k}^n f_{j+1,k-1}^n + \tilde{b}_{j,k}^n f_{j+1,k}^n + \tilde{c}_{j,k}^n f_{j+1,k+1}^n = d_{j,k}^{n+1}\end{gathered} \tag{B.11}$$

for $1 < j < J$, $1 < k < K$ and $0 \leq n < N$. Again, the nine coefficients on the lefthand side will be functions of $(r_{j,k}^n, \mu_{j,k}^n, \sigma_{j,k}^n, \rho, \Delta t, \Delta x, \Delta y)$ as will the righthand side (albeit at at t_{n+1}). Moreover, $d_{j,k}^{n+1}$ will in general depend on $\{\ f_{a,b}^{n+1}\}$ where $a \in \{j-1, j, j+1\}$ and $b \in \{k-1, k, k+1\}$, so once again we have recovered a system which beckons a "backwards iterative solution approach" (recall that the time t_N face of the cuboid is known in accordance with the terminal condition $\forall j \forall k$: $f_{j,k}^N = \Psi(x_j, y_k)$).

Our only remaining problem is that (B.11) only gives us $(J-2) \times (K-2)$ linear equations in $J \times K$ unknowns. Thus, to recover the additional $2(K+J-2)$ linearly independent equations, we will assume that we can let $1 \leq j \leq J$, $1 \leq k \leq K$ in (B.11): all new instances of f involving zero$^{\text{th}}$, $(J+1)^{\text{th}}$ or $(K+1)^{\text{th}}$ subscripts are assumed obvious (e.g. by being relatable to neighbouring fs involving subscripts in one, J or K). In visual terms one may think of it as follows: upon inflating the cuboid uniformly by one grid point in all spatial dimensions, all additional $2(K+J-2)(N+1)$ grid points engulfed by the expansion are considered known.

B.2.3.1 A Block Tridiagonal Matrix

A traditional Gaussian elimination approach to system (B.11) incl. boundary equations would blatantly be stupendously computationally expensive (specifically, it requires $\mathscr{O}(J^3K^3)$ binary operations per time step). Fortunately, by carefully selecting the ordering of the set of two-tuples $\{(j,k)\}_{1\le j\le J, 1\le k\le K}$ [3] we may bring the linear equations into *block tridiagonal form*, which in turn is susceptible to a generalised multi-dimensional version of Thomas' algorithm. Two natural orderings come to mind, viz.

$$
\begin{aligned}
&(a): \quad (j,k)\in\{(1,1),(1,2),...,(1,K);(2,1),(2,2),...,(2,K);...;(J,1),(J,2),...,(J,K)\},\\
&(b): \quad (j,k)\in\{(1,1),(2,1),...,(J,1);(1,2),(2,2),...,(J,2);...;(1,K),(2,K),...,(J,K)\}.
\end{aligned}
$$

As it will shortly become obvious, choosing the former gives rise to a block tridiagonal matrix of bandwidth[4] $2K+2$, whilst the latter gives rise to a block tridiagonal matrix of bandwidth $2J+2$. Generally, it is favourable to minimise the bandwidth as it turns out to engender fewer binary operations in the solution algorithm.[5] Without loss of generality we will proceed as though $K\le J$ and thus opt for ordering (a).

The purported block tridiagonal nature now shows by writing out a few rows explicitly according to the chosen ordering. The details of this are fairly tedious and won't be reproduced here. Ultimately the picture that emerges is the following:

$$
\begin{pmatrix}
\hat{\mathbf{A}}_1^n & \tilde{\mathbf{A}}_1^n & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\
\mathbf{A}_2^n & \hat{\mathbf{A}}_2^n & \tilde{\mathbf{A}}_2^n & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\
\mathbf{0} & \mathbf{A}_3^n & \hat{\mathbf{A}}_3^n & \tilde{\mathbf{A}}_3^n & \mathbf{0} & \cdots & \mathbf{0}\\
\vdots & & \ddots & \ddots & \ddots & & \vdots\\
\mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{A}_{J-1}^n & \hat{\mathbf{A}}_{J-1}^n & \tilde{\mathbf{A}}_{J-1}^n\\
\mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_J^n & \hat{\mathbf{A}}_J^n
\end{pmatrix}
\begin{pmatrix}
\mathbf{f}_1^n\\ \mathbf{f}_2^n\\ \mathbf{f}_3^n\\ \vdots\\ \mathbf{f}_{J-1}^n\\ \mathbf{f}_J^n
\end{pmatrix}
=
\begin{pmatrix}
\boldsymbol{q}_1^{n+1}\\ \boldsymbol{q}_2^{n+1}\\ \boldsymbol{q}_3^{n+1}\\ \vdots\\ \boldsymbol{q}_{J-1}^{n+1}\\ \boldsymbol{q}_J^{n+1}
\end{pmatrix}
-
\begin{pmatrix}
\boldsymbol{\epsilon}_1^n\\ \boldsymbol{\epsilon}_2^n\\ \boldsymbol{\epsilon}_3^n\\ \vdots\\ \boldsymbol{\epsilon}_{J-1}^n\\ \boldsymbol{\epsilon}_J^n
\end{pmatrix},
\tag{B.12}
$$

where the matrix components are defined thusly:

[3] Where every two-tuple really should be read as shorthand notation for the coordinate $(j\Delta x, k\Delta y)$ in $\mathbb{X}\times\mathbb{Y}$.

[4] The bandwidth of a block tridiagonal matrix is defined as the maximal span of non-zero entries found in any given row. I.e. the number of elements starting with the first non-zero entry in a row and ending with the last non-zero entry in a row (but possibly including zero entries in between).

[5] The reader may realise this by inspecting the generalised version of Thomas' algorithm, which requires the inversion of J $\mathbb{R}^{K\times K}$ matrices (scheme (a)) or K $\mathbb{R}^{J\times J}$ matrices (scheme (b)).

$$ {}^x\mathbf{A}_j^n \equiv \begin{pmatrix} {}^x b_{j,1}^n & {}^x c_{j,1}^n & 0 & 0 & 0 & \cdots & 0 \\ {}^x a_{j,2}^n & {}^x b_{j,2}^n & {}^x c_{j,2}^n & 0 & 0 & \cdots & 0 \\ 0 & {}^x a_{j,3}^n & {}^x b_{j,3}^n & {}^x c_{j,3}^n & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 0 & {}^x a_{j,K-1}^n & {}^x b_{j,K-1}^n & {}^x c_{j,K-1}^n \\ 0 & \cdots & 0 & 0 & 0 & {}^x a_{j,K}^n & {}^x b_{j,K}^n \end{pmatrix} \in \mathbb{R}^{K\times K}, $$

where the superscript x represents a "[blank]", "hat" or "tilde" over the following object. Furthermore, $\mathbf{f}_j^n \equiv (f_{j,1}^n, f_{j,2}^n, f_{j,3}^n, ..., f_{j,K-1}^n, f_{j,K}^n) \in \mathbb{R}^K$, $\boldsymbol{q}_j^n \equiv (d_{j,1}^n, d_{j,2}^n, d_{j,3}^n, ..., d_{j,K-1}^n d_{j,K}^n) \in \mathbb{R}^K$ and

$$ \boldsymbol{\epsilon}_j^n \equiv \mathfrak{a}_{j,1}\hat{\mathbf{e}}_1 + \mathfrak{c}_{j,K}\hat{\mathbf{e}}_K + \delta_{j,1}\mathbf{A}_1^n\mathbf{f}_0^n + \delta_{j,J}\tilde{\mathbf{A}}_J^n\mathbf{f}_{J+1}^n \in \mathbb{R}^K, $$

where we have defined the coefficients $\mathfrak{a}_{j,1} \equiv a_{j,1}^n f_{j-1,0}^n + \hat{a}_{j,1}^n f_{j,0}^n + \tilde{a}_{j,1}^n f_{j+1,0}^n$ and $\mathfrak{c}_{j,K} \equiv c_{j,K}^n f_{j-1,K+1}^n + \hat{c}_{j,K}^n f_{j,K+1}^n + \tilde{c}_{j,K}^n f_{j+1,K+1}^n$. Lastly, $\hat{\mathbf{e}}_1$ is the unit vector $(1,0,0,...,0,0) \in \mathbb{R}^K$, $\hat{\mathbf{e}}_K$ is the unit vector $(0,0,0,...,0,1) \in \mathbb{R}^K$ and $\delta_{x,y}$ is the Kronecker delta.

B.2.3.2 Thomas' Generalised Algorithm

The mega-matrix in equation (B.12) is embedded in the space $\mathbb{R}^{J\cdot K\times J\cdot K}$ and will (for most practical purposes) be highly impractical to invert. However, we may proceed almost exactly as in subsection B.2.2 to extend Thomas' algorithm to matrices with matrix components. To this end, we introduce the lower and upper triangular matrices $\mathbf{L}^n$ and $\mathbf{U}^n$ as in (B.8) where each component now is a square matrix of dimensionality $k \times k$: $(l_i^n, h_j^n, u_j^n, 1, 0) \mapsto (\mathbf{L}_j^n, \mathbf{H}_j^n, \mathbf{U}_j^n, \mathbb{I}, \mathbf{0})$. Defining $\mathbf{d}_{*j} \equiv \mathbf{d}_j - \boldsymbol{\epsilon}_j$, $\forall j$, Thomas' generalised algorithm then boils down to the procedure exhibited below.

A few remarks should be wrapped around this algorithm. First of all, the non-commutativity of matrices must obviously be observed. The ordering below is correct and may be verified by following the original argument in detail. Secondly, whilst the notion might have some prima facie allure, notice that one *cannot* imbed a Thomas algorithm within Thomas' generalised algorithm to solve for the $\mathbf{L}_j^n$s and $\mathbf{y}_j^n$s (in the attempt to avoid an explicit inversion of the $\mathbf{H}_j^n$s). This follows from the simple fact that the $\mathbf{H}_j^n$s generally won't be tridiagonal beyond the case $j = 1$.

We may now compute the reduction in computational complexity as follows: from the J full matrix inversions of the $\mathbf{H}_j^n$s we incur from Gaussian elimination: $J \cdot (\frac{1}{3}K^3 + K^2 + \frac{1}{3}K)$ multiplications & divisions and $J \cdot (\frac{1}{3}K^3 + \frac{1}{2}K^2 - \frac{5}{6}K)$ additions & subtractions. The remaining multiplications & divisions in the algorithm amount to $5JK - 4K$, whilst the remaining additions & subtractions amount to $3JK - 3K$. Thus, under the assumption that $K \leq J$, the overall complexity of the algorithm scales as $\mathscr{O}(JK^3)$ - a substantial improvement over Gauss' $\mathscr{O}(J^3K^3)$.[6] Insofar as $J \leq K$ we would obtain the "mirror" result $\mathscr{O}(KJ^3)$, so one cannot exploit the difference in the powers of J and K beyond selecting the smallest possible bandwidth to begin with.

Thomas' Generalised Algorithm

Set	$\mathbf{H}_1^n$	$:=$	$\hat{\mathbf{A}}_1^n;$
	$\mathbf{U}_1^n$	$:=$	$\tilde{\mathbf{A}}_1^n;$
For $j = 2, ..., J-1$ set	$\mathbf{L}_j^n$	$:=$	$\mathbf{A}_j^n[\mathbf{H}_{j-1}^n]^{-1};$
	$\mathbf{H}_j^n$	$:=$	$\hat{\mathbf{A}}_j^n - \mathbf{L}_j^n\mathbf{U}_{j-1}^n;$
	$\mathbf{U}_j^n$	$:=$	$\tilde{\mathbf{A}}_j^n;$
Set	$\mathbf{L}_J^n$	$:=$	$\mathbf{A}_J^n[\mathbf{H}_{J-1}^n]^{-1};$
	$\mathbf{H}_J^n$	$:=$	$\hat{\mathbf{A}}_J^n - \mathbf{L}_J^n\mathbf{U}_{J-1}^n;$
Set	$\mathbf{y}_1^n$	$:=$	$\mathbf{d}_{*1}^{n+1};$
For $j = 2, ..., J$ set	$\mathbf{y}_j^n$	$:=$	$\mathbf{d}_{*j}^{n+1} - \mathbf{L}_j^n\mathbf{y}_{j-1}^n;$
Set	$\mathbf{f}_J^n$	$:=$	$[\mathbf{H}_J^n]^{-1}\mathbf{y}_J^n;$
For $j = J-1, ..., 1$ set	$\mathbf{f}_j^n$	$:=$	$[\mathbf{H}_j^n]^{-1}(\mathbf{y}_j^n - \mathbf{U}_j^n\mathbf{f}_{j+1}^n);$

Remark B.1. Although the generalised Thomas algorithm offers a significant reduction in computational speed, we do certainly *not* purport that it is the most efficient way to solve block-tridiagonal linear systems. It manifestly is not. Readers interested in diving into this rich field are referred to Saad [20].

[6] In actual implementations of Thomas' generalised algorithm one might find that it is appropriate to manipulate the indices, which will engender further binary operations. However, the totality of these prove to be considerably smaller than the leading order of magnitude.

Appendix C
Stochastic Control Methods in Mathematical Finance

Abstract In this appendix we briefly review the two basic paradigms of stochastic control theory, viz. the Hamilton-Jacobi-Bellman formalism and the martingale method. To emphasise the problem solving aspect of the two approaches, we shall use Merton's classical portfolio optimisation problem as our vantage point. The final section discusses important differences between the two approaches.

C.1 The Problem Posed

We consider the case of an investor who is assumed to live over a known temporal horizon $[0,T]$. His total wealth, $\mathscr{W}_t$, is modelled dynamically in time by a stochastic differential equation and is assumed to have the known initial value $\mathscr{W}_0 = w_0$. At any given instant the investor is faced with the choice of how much of his wealth to consume per unit time, c_t, and which proportion of his wealth, $\boldsymbol{\pi}_t = (\pi_{1t}, \pi_{2t}, \dots, \pi_{nt})^\intercal$, he should allocate to n risky asset, the price processes of which we codify by $\boldsymbol{S}_t = (S_{1t}, S_{2t}, \dots, S_{nt})^\intercal$. For simplicity, we assume a governing dynamics of the form

$$d\boldsymbol{S}_t = \operatorname{diag}(\boldsymbol{S}_t)[\boldsymbol{\mu} dt + \boldsymbol{\sigma} d\boldsymbol{W}_t^{\mathbb{P}}], \tag{C.1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\sigma} \in \mathbb{R}^{n\times n}$ are known constant tensors. The remaining wealth of the investor is to be placed (with proportion $1 - \boldsymbol{\pi}_t^\intercal \boldsymbol{\iota}$) in a riskless asset, B_t, which grows at the constant rate of interest r, whence $dB_t = rB_t dt$. Furthermore, we assume that consumption is everywhere non-negative, $c_t \geq 0$, whilst no such condition is placed on $\boldsymbol{\pi}_t$ (that is to say, we allow for short selling and leveraging). If u is the investors utility function, and δ is some subjective discount factor, then Merton's portfolio problem [12] is to find functions c_t^* and $\boldsymbol{\pi}_t^*$, $t \in [0,T]$, such that

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^T e^{-\delta t} u(c_t)dt + e^{-\delta T} u(\mathscr{W}_T)\right], \tag{C.2}$$

is maximised. I.e. our aim is to find a consumption-investment strategy such that the expected discounted utility of consumption over a life-time and the expected discounted utility of the bequest $\mathscr{W}_T$ is at its peak. To this end, let us assume we operate with utility of the constant relative risk aversion (CRRA) variety

$$u(x) = \frac{x^{1-\gamma}}{1-\gamma},$$

where $\gamma \in \mathbb{R}_+ \backslash \{1\}$ codifies the investor's risk aversion.

Finally, assume the overall portfolio dynamics is self-financing and that there are no monetary injections such as labour income. From the self-financing condition (Björk's Lemma 6.4 [10])[7] it follows that the wealth dynamics is of the form

$$d\mathscr{W}_t = \mathscr{W}_t[r + \boldsymbol{\pi}^\intercal \boldsymbol{\sigma}\boldsymbol{\lambda}]dt - c_t dt + \mathscr{W}_t \boldsymbol{\pi}^\intercal \boldsymbol{\sigma} d\boldsymbol{W}_t^{\mathbb{P}}, \tag{C.3}$$

where $\boldsymbol{\lambda} \equiv \boldsymbol{\sigma}^{-1}(\boldsymbol{\mu} - r\boldsymbol{\iota}) \in \mathbb{R}^n$ is the market price of risk vector.

Over the following pages, we expose standard solution methods to the Merton problem in a manner which allows for straightforward generalisation to more complicated scenarios. First up is the method most commonly encountered in the literature, viz. the Hamilton-Jacobi-Bellman (HJB) method, so named after the pioneering work on stochastic control theory by Richard Bellman in the 1950s alongside the celebrated (deterministic) calculus of variations result by William Hamilton and Carl Jacobi. Subsequently, we shall consider the so-called martingale method [6], which despite its probabilistic nature (however obfuscating it might appear at first), has much to offer over the traditional Bellmanian approach. Our level of pedagogy is somewhat on a par with Björk [10] and Munk [13]. Readers interested in more advanced surveys are referred to Karatzas and Shreve [10] and Pham [18].

Remark C.1. From a purely axiomatic perspective, Merton's problem of utility optimisation is a cacophony of dubious and overtly simplified assumptions not easily squared with real life investment-consumption processes of rational agents. Pitfalls include the the negligence of labour income and transactions costs, the constancy of $\delta, \boldsymbol{\mu}, \boldsymbol{\sigma}$, and r, the fixed lifetime of the investor and his simplified utility function. Nevertheless, we may rejoice in the fact that the relatively complex mathematical machinery of the problem above admits analytical solutions. Indeed, there is some solace to be sought in the more recent developments of the problem, which has addressed some of these issues and more.

---

[7] Here's the idea: let $\boldsymbol{P}_t = (P_{1t}, P_{2t}, \ldots, P_{mt})^\intercal \in \mathbb{R}^m$ be a pricing vector, and let $\boldsymbol{h}_t \in \mathbb{R}^m$ be the portfolio holding, such that the investor's total wealth at time t is $\mathscr{W}_t = \boldsymbol{h}_t^\intercal \boldsymbol{P}_t$. Suppose the investor last updated his portfolio at time $t - \Delta t$ (holding $\boldsymbol{h}_{t-\Delta t}$), then the value of his portfolio at t is $\mathscr{W}_t = \boldsymbol{h}_{t-\Delta t}^\intercal \boldsymbol{P}_t$. The cost of the new portfolio he buys at t is $\boldsymbol{h}_t^\intercal \boldsymbol{P}_t$. We allow for proceeds consumption of the magnitude $c_t \Delta t$ in the interval Δt i.e. all in all the self-financing condition is $\boldsymbol{h}_{t-\Delta t}^\intercal \boldsymbol{P}_t = \boldsymbol{h}_t^\intercal \boldsymbol{P}_t + c_t \Delta t$ or identically $\Delta \boldsymbol{h}_t^\intercal \boldsymbol{P}_t + c_t \Delta t = 0$. Adding and subtracting $\Delta \boldsymbol{h}_t^\intercal \boldsymbol{P}_{t-\Delta t}$ and letting $\Delta t \to 0$ we get the budget equation $\boldsymbol{P}_t^\intercal d\boldsymbol{h}_t + d\boldsymbol{P}_t^\intercal d\boldsymbol{h}_t + c_t dt = 0$. But applying Itô to $\mathscr{W}_t = \boldsymbol{h}_t^\intercal \boldsymbol{P}_t$ we get $d\mathscr{W}_t = \boldsymbol{h}_t^\intercal d\boldsymbol{P}_t + \boldsymbol{P}_t^\intercal d\boldsymbol{h}_t + d\boldsymbol{P}_t^\intercal d\boldsymbol{h}_t$, which combined with our budget constraint gives us the self-financing condition $d\mathscr{W}_t = \boldsymbol{h}_t^\intercal d\boldsymbol{P}_t - c_t dt$ or identically $d\mathscr{W}_t = \mathscr{W}_t \sum_i \pi_{it} dP_{it}/P_{it} - c_t dt$ where we have defined the weight $\pi_{it} \equiv P_{it} h_{it}/\mathscr{W}_t$. Clearly, $\sum_i \pi_{it} = 1$ so the nomenclature *weight* is appropriate.

C.2 The Hamilton-Jacobi-Bellman Method

Theory: The basic point of this approach is to translate the optimal control problem into an equivalent non-linear partial differential equation known as the Hamilton-Jacobi-Bellman (HJB) equation. Insofar as a solution to the latter can be found (which is far from given), one can immediately deduce closed form expressions for the optimal controls.

To this end, we start out by defining the function for which we desire a governing PDE, viz. the so-called **optimal value function** (or **indirect utility function**) $J(s,w) : [0,T] \times \mathbb{R} \mapsto \mathbb{R}$, defined as

$$V(s,w) \equiv \sup_{\{\pi_t, c_t\}_{t\in[s,T]}} I(\pi_t, c_t | s, w), \qquad \text{where} \tag{C.4}$$

$$I(\pi_t, c_t | s, w) \equiv \mathbb{E}^{\mathbb{P}}_{s,w}\left[\int_s^T e^{-\delta(t-s)} u(c_t)dt + e^{-\delta(T-s)} u(\mathscr{W}_T)\right],$$

which, of course, is nothing but our original problem with a generic starting point $s \in [0,T]$ and wealth w (think of (C.4) as the scenario where we have to solve the Merton problem for an investor who has already lived for s years). Furthermore, we must make the following important assumptions:

1. There are optimal Markov control functions $\boldsymbol{\pi}^*_t = \boldsymbol{\pi}^*(t, \mathscr{W}_t) : [s,T] \times \mathbb{R} \mapsto \mathbb{R}$ and $c^*_t = c^*(t, \mathscr{W}_t) : [s,T] \times \mathbb{R} \mapsto \mathbb{R}_+$ such that the supremum is attained, i.e. s.t. $V(s,w) = I(\boldsymbol{\pi}^*_t, c^*_t | s, w)$. In particular, these controls must be **admissible** meaning that they satisfy the constraints enunciated above at all times. We say that there is an optimal admissible **feedback control law**, $\mathscr{L}^* : \{\boldsymbol{\pi}^*_t, c^*_t\}$. This is an existence claim, but it is not a uniqueness claim.
2. $V \in \mathscr{C}^{1,2}$. I.e. the first order temporal derivative, and the first and second order wealth derivatives of V all exist. This is non-obvious!
3. A number of limiting procedures in the following arguments can be justified.

Given these assumptions, a PDE to which $V(s,w)$ is a solution can be derived by following these standard steps in dynamic programming:

1. Fix the coordinate $(s,w) \in [0,T] \times \mathbb{R}$ and consider the following two strategies over the interval $[s,T]$: **Strategy I** use the optimal control law $\mathscr{L}^* : \{\boldsymbol{\pi}^*_t, c^*_t\}$. **Strategy II** Use the (sub)-optimal control law $\mathscr{L}' : \{\boldsymbol{\pi}'_t, c'_t\}$ where

$$\mathscr{L}' : \{\boldsymbol{\pi}'_t, c'_t\} \equiv \begin{cases} \mathscr{L} : \{\boldsymbol{\pi}_t, c_t\}, & \text{for } (t, \mathscr{W}_t) \in [s, s+\Delta s] \times \mathbb{R} \\ \mathscr{L}^* : \{\boldsymbol{\pi}^*_t, c^*_t\}, & \text{for } (t, \mathscr{W}_t) \in (s+\Delta s, T] \times \mathbb{R}, \end{cases}$$

   where Δs is some incremental time step. Notice that it the optimal control is used over the latter time interval $(s+\Delta s, T]$.
2. Compute the Merton expectation $I(\boldsymbol{\pi}_t, c_t | s, w)$ for both strategies.
3. Evidently, strategy I has to be at least as good as strategy II vis-a-vis the Merton expectation. Using this, and letting $\Delta s \to 0$ we obtain the HJB PDE.

From assumption (1) the first strategy is trivially $I(\boldsymbol{\pi}_t^*, c_t^*|s,w) = V(s,w)$. For the second strategy we observe that we switch from a random control ($\mathscr{L}$) to an optimal control ($\mathscr{L}^*$) after Δs amounts of time. The wealth will therefore evolve to the stochastic state $\mathscr{W}_{s+\Delta s}^{\mathscr{L}}$ at $s+\Delta s$ and thence to its terminal value $\mathscr{W}_T^{\mathscr{L}^*}$ at T. Thus, $I(\pi_t', c_t'|s,w)$ can be written as

$$\begin{aligned}
&\mathbb{E}_{s,w}^{\mathbb{P}}\left[\int_s^{s+\Delta s} e^{-\delta(t-s)}u(c_t)dt + \int_{s+\Delta s}^{T} e^{-\delta(t-s)}u(c_t^*)dt + e^{-\delta(T-s)}u(\mathscr{W}_T^{\mathscr{L}^*})\right]\\
&= \mathbb{E}_{s,w}^{\mathbb{P}}\Bigg[\int_s^{s+\Delta s} e^{-\delta(t-s)}u(c_t)dt + e^{-\delta\Delta s}\cdot\\
&\quad \mathbb{E}_{s+\Delta s,\mathscr{W}_{s+\Delta s}^{\mathscr{L}}}^{\mathbb{P}}\left[\int_{s+\Delta s}^{T} e^{-\delta(t-(s+\Delta s))}u(c_t^*)dt + e^{-\delta(T-(s+\Delta s))}u(\mathscr{W}_T^{\mathscr{L}^*})\right]\Bigg]\\
&= \mathbb{E}_{s,w}^{\mathbb{P}}\left[\int_s^{s+\Delta s} e^{-\delta(t-s)}u(c_t)dt + e^{-\delta\Delta s}V(s+\Delta s, \mathscr{W}_{s+\Delta s}^{\mathscr{L}})\right],
\end{aligned}$$

where the second line uses the *law of iterated expectation*s, and the final line the definition of the optimal value function. Hence, using the first insight from step (3) we have that strategies I and II compare as

$$V(s,w_s) \geq \mathbb{E}_{s,w}^{\mathbb{P}}\left[\int_s^{s+\Delta s} e^{-\delta(t-s)}u(c_t)dt + e^{-\delta\Delta s}V(s+\Delta s, \mathscr{W}_{s+\Delta s}^{\mathscr{L}})\right]. \tag{C.5}$$

Now using assumption (2) we can use Itō's formula to write

$$\begin{aligned}
V(s+\Delta s, \mathscr{W}_{s+\Delta s}^{\mathscr{L}}) =& V(s,w_s) + \int_s^{s+\Delta s}\Big\{\partial_s V(t,\mathscr{W}_t^{\mathscr{L}})dt\\
&+ \partial_w V(t,\mathscr{W}_t^{\mathscr{L}})d\mathscr{W}_t^{\mathscr{L}} + \tfrac{1}{2}\partial_{ww}^2 V(t,\mathscr{W}_t^{\mathscr{L}})(d\mathscr{W}_t^{\mathscr{L}})^2\Big\},
\end{aligned}$$

which combined with our wealth dynamics (C.3) becomes

$$\begin{aligned}
V(s+\Delta s, \mathscr{W}_{s+\Delta s}^{\mathscr{L}}) =& V(s,w_s) + \int_s^{s+\Delta s}\Big\{\partial_s V(t,\mathscr{W}_t^{\mathscr{L}}) + \mathbf{A}^{\mathscr{L}}V(t,\mathscr{W}_t^{\mathscr{L}})\Big\}dt\\
&+ \int_s^{s+\Delta s}\mathscr{W}_t^{\mathscr{L}}\partial_w V(t,\mathscr{W}_t^{\mathscr{L}})\boldsymbol{\pi}_t^{\intercal}\boldsymbol{\sigma}d\mathbf{W}_t^{\mathbb{P}},
\end{aligned} \tag{C.6}$$

where we have defined the differential operator

$$\mathbf{A}^{\mathscr{L}} \equiv \mathscr{W}_t^{\mathscr{L}}[r+\boldsymbol{\pi}_t^{\intercal}\boldsymbol{\sigma}\boldsymbol{\lambda}]\partial_w - c_t\partial_w + \tfrac{1}{2}(\mathscr{W}_t^{\mathscr{L}})^2\boldsymbol{\pi}_t^{\intercal}\boldsymbol{\sigma}\boldsymbol{\sigma}^{\intercal}\boldsymbol{\pi}\partial_{ww}^2.$$

Substituting (C.6) into inequality (C.5) and assuming square integrability in order for the stochastic integral to vanish, we obtain after a bit or rearranging

$$(e^{\delta\Delta s}-1)V(s,w) \geq \mathbb{E}_{s,w}^{\mathbb{P}}\left[\int_s^{s+\Delta s}\{e^{\delta\Delta s-\delta(t-s)}u(c_t) + \partial_s V(t,\mathscr{W}_t^{\mathscr{L}}) + \mathbf{A}^{\mathscr{L}}V(t,\mathscr{W}_t^{\mathscr{L}})\}dt\right]. \tag{C.7}$$

Suppose now we divide through on both sides by Δs and take the limit as $\Delta s \to 0$. If our expression exhibits sufficient regularity we can justify interchanging the limit and the expectation operator. Thus,

$$\delta V(s,w) \geq u(c_s) + \partial_s V(s,w) + \mathbf{A}^{\mathscr{L}} V(s,w), \tag{C.8}$$

where we have used the fact that $(e^{\delta \Delta S} - 1)/\Delta s \to \delta$ as $\Delta s \to 0$.Notice that our functions $V, \boldsymbol{\pi}_t, c_t$ (and consequently also $\mathbf{A}^{\mathscr{L}}$) here are evaluated at the initial coordinate (s,w). However, whilst (s,w) hitherto has been treated as fixed, it was arbitrarily chosen and thence equation (C.8) must hold true for all $(s,w_s) \in [0,T] \times \mathbb{R}$, with equality holding for the optimal control $\mathscr{L}^*$ only. Hence, we arrive at the theorem:

Theorem C.1. ***The Hamilton-Jacobi-Bellman Equation for Merton's Problem.*** *Consider a wealth process (C.3). Let $V(s,w)$ be defined as in (C.4), and assume it satisfies assumptions (1)-(3) declared above, then $V(s,w)$ satisfies the HJB equation*

$$\delta V(s,w) = \partial_s V(s,w) + \sup_{c_s \in \mathbb{R}_+, \boldsymbol{\pi}_s \in \mathbb{R}^n} \left\{ u(c_s) + \mathbf{A} V(s,w) \right\}, \tag{C.9}$$

$\forall (s,w) \in (0,T) \times \mathbb{R}$, *where*

$$\mathbf{A} \equiv w[r + \boldsymbol{\pi}_t^\intercal \boldsymbol{\sigma} \boldsymbol{\lambda}]\partial_w - c_t \partial_w + \tfrac{1}{2} w^2 \boldsymbol{\pi}_t^\intercal \boldsymbol{\sigma} \boldsymbol{\sigma}^\intercal \boldsymbol{\pi} \partial^2_{ww},$$

and we have the obvious boundary condition $V(T,w) = u(w), \forall w \in \mathbb{R}$ (if we start the Merton problem when the investor dies there's nothing but the bequest). For each $(s,w) \in [0,T] \times \mathbb{R}$ the supremum is attained by $c_s^*, \boldsymbol{\pi}_s^*$.

Remark C.2. Importantly, the HJB equation (C.9), whilst highly non-linear, only involves the supremum over all admissible consumptions and holdings of risky assets at time s, and *not* the supremum over the entire process as we saw it in (C.4).

Do notice that the theorem above only has the form of a **necessary** condition: i.e. if V is an optimal value function and $\mathscr{L}^*$ an optimal control, then V satisfies the HJB equation with $\mathscr{L}^*$ giving rise to the supremum. Proving that the HJB equation is also **sufficient** for optimality can for more general problems prove considerably more arduous. Nevertheless, for our present purposes, this so-called **verification steps** runs as follows:

Theorem C.2. ***The Verification Theorem for Merton's Problem.*** *Suppose we have the functions $\varphi(s,w)$, $\boldsymbol{\pi}^*(s,w)$ and $c^*(s,w)$ such that*

- *φ is sufficiently integrable (see above) and solves the HJB equation*

$$\delta \varphi(s,w) = \partial_s \varphi(s,w) + \sup_{c_s \in \mathbb{R}_+, \boldsymbol{\pi}_s \in \mathbb{R}^n} \left\{ u(c_s) + \mathbf{A} \varphi(s,w) \right\},$$

$\forall (s,w) \in (0,T) \times \mathbb{R}$, *with the terminal condition* $\varphi(T,w) = u(w), \forall w \in \mathbb{R}$.

- $\boldsymbol{\pi}^*(s,w) : [0,T] \times \mathbb{R} \mapsto \mathbb{R}^n$ *and* $c^*(s,w) : [0,T] \times \mathbb{R} \mapsto \mathbb{R}_+$ *- that is,* $\boldsymbol{\pi}^*$ *and* c^* *are admissible control laws (they satisfy the pre-specified function constraints).*
- *For each fixed* (s,w) *the supremum in the expression*

$$\sup_{c_s \in \mathbb{R}_+, \boldsymbol{\pi}_s \in \mathbb{R}^n} \left\{ u(c_s) + \mathbf{A}\varphi(s,w) \right\},$$

is attained by the choice $\boldsymbol{\pi}_s = \boldsymbol{\pi}^*(s,w)$, $c_s = c^*(s,w)$.

Then it holds that

1. *The optimal value function (C.4) to Merton's control problem is given by* $V(s,w) = \varphi(s,w)$.
2. *There exist an optimal control law, viz.* $\{\boldsymbol{\pi}^*(s,w), c^*(s,w)\}$.

Proof. Let functions $\varphi, \boldsymbol{\pi}^*$ and c^* be given as above. Select the arbitrary admissible control law $\mathscr{L} : \{\boldsymbol{\pi}_t, c_t\}$ and fix a coordinate (s,w). If we define the dynamics of the wealth process $\mathscr{W}_t^{\mathscr{L}}$ as in (C.3) with boundary $\mathscr{W}_s^{\mathscr{L}} = w$, then an straight-forward application of Itō's formula to $e^{-\delta(t-s)}\varphi(t,\mathscr{W}_t^{\mathscr{L}})$ gives

$$e^{-\delta(T-s)}\varphi(T,\mathscr{W}_T^{\mathscr{L}}) = \varphi(s,w) + \int_s^T e^{-\delta(t-s)} \left\{ [\partial_t - \delta]\varphi(t,\mathscr{W}_t^{\mathscr{L}}) + \mathbf{A}\varphi(t,\mathscr{W}_t^{\mathscr{L}}) \right\} dt + \int_s^T e^{-\delta(t-s)} \mathscr{W}_t \partial_w \varphi(t,\mathscr{W}_t^{\mathscr{L}}) \boldsymbol{\pi}_t^{\intercal} \boldsymbol{\sigma} d\mathbf{W}_t^{\mathbb{P}}.$$

Using our assumptions that φ satisfies the HJB equation and has the terminal value $\varphi(T,\mathscr{W}_T^{\mathscr{L}}) = u(\mathscr{W}_T^{\mathscr{L}})$ we get

$$\varphi(s,w) \geq \int_s^T e^{-\delta(t-s)} u(c_t) dt + e^{-\delta(T-s)} u(\mathscr{W}_T^{\mathscr{L}}) - \int_s^T e^{-\delta(t-s)} \mathscr{W}_t \partial_w \varphi(t,\mathscr{W}_t^{\mathscr{L}}) \boldsymbol{\pi}_t^{\intercal} \boldsymbol{\sigma} d\mathbf{W}_t^{\mathbb{P}}.$$

Applying the (t,w) conditional expectation to this equation, and using the integrability assumption:

$$\varphi(s,w_s) \geq \mathbb{E}_{t,w}^{\mathbb{P}} \left[\int_s^T e^{-\delta(t-s)} u(c_t) dt + e^{-\delta(T-s)} u(\mathscr{W}_T^{\mathscr{L}}) \right] \equiv I(\boldsymbol{\pi}_t, c_t | s, w_s). \tag{C.10}$$

This inequality is true for arbitrary control laws - also in the event that we selected the supremal control law. Hence, from the definition of $\mathscr{V}$, (C.4):

$$\varphi(s,w) \geq V(s,w). \tag{C.11}$$

Had we opted for using the functions $\boldsymbol{\pi}^*, c^*$ it is clear that we would have obtained a strict equality in equation (C.10) viz. $\varphi(s,w) = I(\boldsymbol{\pi}_t^*, c_t^*|s,w)$. If we substitute this into the trivial inequality $V(s,w) \geq I(\boldsymbol{\pi}_t^*, c_t^*|s,w)$ we get:

$$V(s,w_s) \geq \varphi(s,w). \tag{C.12}$$

Evidently, (C.11) and (C.12) jointly imply $\varphi = V$ and that $\{\boldsymbol{\pi}_t^*, c_t^*\}$ is an optimal control.
□

Extracting the Solution: Having formally established the equivalence between solving the control problem (C.4) and the HJB equation (C.9), we are now left with the task of finding an explicit solution to the latter. Differentiating the HJB equation partially with respect to c_s and $\boldsymbol{\pi}_s$ and equating to zero, we find that the optimal feedback controls are of the form

$$c_s^* = (u')^{-1}(\partial_w V(s,w)), \qquad \text{and} \qquad \boldsymbol{\pi}_s^* = -\frac{\partial_w V(s,w)}{w\partial_{ww}^2 V(s,w)}(\boldsymbol{\sigma}^\intercal)^{-1}\boldsymbol{\lambda}, \tag{C.13}$$

where $(u')^{-1}(\cdot) = (\cdot)^{-1/\gamma}$ is the inverse marginal utility function (for CRRA). Substituting these functions back into (C.9) reveals that the governing dynamics is, in fact, a highly **non-linear** PDE, viz.

$$\delta V(s,w) = \partial_s V(s,w) + \frac{\gamma}{1-\gamma}(\partial_w V(s,w))^{1-\frac{1}{\gamma}} + rw\partial_w V(s,w) - \tfrac{1}{2}||\boldsymbol{\lambda}||^2 \frac{(\partial_w V(s,w))^2}{\partial_{ww}^2 V(s,w)}, \tag{C.14}$$

subject to the terminal condition $V(T,w) = w^{1-\gamma}/(1-\gamma)$. Generally, non-linearity can have a rather devastating effect on our ability to extract a solution (analytically or numerically), yet this expressions admits a surprisingly simple solution based on the ansatz that V is separable in time and wealth. Specifically, from the dynamics of the wealth (C.3) it can be argued that if $\{\boldsymbol{\pi}_t^*, c_t^*\}$ is the optimal control plan for an investor with initial wealth w_0, then $\{\boldsymbol{\pi}_t^*, kc_t^*\}$ should be the optimal control plan for an investor with initial wealth kw_0. I.e. consumption is suitably scaled to allow for the extra cash, whilst the proportion invested in each asset remains unchanged. Using this fact it can easily be shown that the optimal value function is homogenous of degree $1-\gamma$: i.e. $V(s,kw) = k^{1-\gamma}V(s,w)$ [13]. Indeed, by setting $k = w^{-1}$ we obtain the separability ansatz

$$V(s,w) = h(s)^\gamma \frac{w^{1-\gamma}}{1-\gamma}, \tag{C.15}$$

where $h : [0,T] \mapsto \mathbb{R}$. Combining this with equation (C.14) we find after a bit of manipulation that

$$\left[\left(\frac{\delta}{1-\gamma} - r - \frac{1}{2\gamma}||\boldsymbol{\lambda}||^2\right)h(s) - \frac{\gamma}{1-\gamma}(1+\partial_t h(s))\right] h(s)^{\gamma-1} w^{1-\gamma} = 0,$$

subject to $g(T) = 1$. Indeed, since this equation holds for all t we conclude that the expression in the square bracket is nil, which readily enables us to deduce that

$$h(s) = \zeta^{-1}(1+[\zeta-1]e^{-\zeta(T-s)}), \qquad \text{where} \qquad \zeta \equiv \frac{\delta - r(1-\gamma)}{\gamma} - \tfrac{1}{2}\frac{1-\gamma}{\gamma^2}||\boldsymbol{\lambda}||^2. \tag{C.16}$$

Inserting the separability ansatz (C.15) into the optimal control expressions (C.13) we thus conclude

$$c_s^* = \frac{w}{h(s)}, \qquad \text{and} \qquad \boldsymbol{\pi}_s^* = \frac{(\boldsymbol{\sigma}^\intercal)^{-1}\boldsymbol{\lambda}}{\gamma}, \tag{C.17}$$

which is to say that the rate of consumption is directly proportional to the instantaneous level of wealth (but exponentially decaying in time), while the investment strategy is constant and equal to a mean-variance optimisation a la Markowitz.[8]

C.3 The Martingale Method

As we shall shortly argue, there are considerable problems of a rather technical nature associated with the HJB approach. Nonetheless, the relative ease with which we can write down the HJB equation and in some scenarios solve it, makes it the methodology of choice in much of the financial literature. The fact that closed form expressions for the controls can be found is somehow used to justify glossing over important technical details along the way. Here we provide an alternative route based on martingale considerations, which (i) saves us from the pitfalls of the HJB approach, (ii) provides a framework for handling problems of a more general nature.

The point is here that rather than solving (C.4) dynamically as above, we solve it statically subject to the constraint that the portfolio must be be self-financing (in other words, we treat the control problem as a Lagrange multiplier problem). Specifically, from elementary martingale theory it follows that $\mathbb{Q}$ discounted "cashflows" should be equal to the present wealth level:

$$w_0 = \mathbb{E}^{\mathbb{Q}}\left[\int_0^T e^{-rt}c_t dt + e^{-rT}\mathscr{W}_T\right], \tag{C.18}$$

(we refer to this as the **budget equation**).[9] Thus, we are faced with the following Lagrange multiplier problem

$$\mathfrak{L} = \mathbb{E}^{\mathbb{P}}\left[\int_0^T e^{-\delta t}u(c_t)dt + e^{-\delta T}u(\mathscr{W}_T) - \eta\left\{\int_0^T e^{-rt}\xi_t c_t dt + e^{-rT}\xi_T\mathscr{W}_T\right\}\right],$$

where η is the multiplier, and $\xi_t = d\mathbb{Q}/d\mathbb{P}(t)$ is the Radon-Nikodym derivative defined as $d\xi_t = -\xi_t\boldsymbol{\lambda}^\intercal d\boldsymbol{W}_t^{\mathbb{P}}$ with $\boldsymbol{\lambda} = \boldsymbol{\sigma}^{-1}(\boldsymbol{\mu} - r\boldsymbol{\iota})$. Differentiating partially with respect to c_s and $\mathscr{W}_T$ and equating to zero, we find the following expressions for the optimal rate of consumption and the optimal terminal wealth

8 Specifically, $\boldsymbol{\pi}^* = \underset{\boldsymbol{\pi}}{\operatorname{argmax}}\{\boldsymbol{\pi}^\intercal\boldsymbol{\mu} + (1-\boldsymbol{\pi}^\intercal\boldsymbol{\iota})r - \frac{\gamma}{2}\boldsymbol{\pi}^\intercal\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal\boldsymbol{\pi}\}$ where $\boldsymbol{\pi}^\intercal\boldsymbol{\mu} + (1-\boldsymbol{\pi}^\intercal\boldsymbol{\iota})r$ is the expected value of the portfolio, and $\boldsymbol{\pi}^\intercal\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal\boldsymbol{\pi}$ is its covariance matrix.

9 For a formal proof of this result see [13].

$$c_s^* = (u')^{-1}(\eta e^{(\delta-r)s}\xi_s), \qquad \text{and} \qquad \mathscr{W}_T^* = (u')^{-1}(\eta e^{(\delta-r)T}\xi_T), \tag{C.19}$$

where we reiterate that $(u')^{-1}(\cdot) = (\cdot)^{-1/\gamma}$ is the inverse marginal utility function. To rid our expressions of the multiplier η we first substitute (C.19) into the budget constraint (C.18) to find

$$\eta^{-1/\gamma} = \frac{w_0}{\int_0^T e^{-(r+q)t}\mathbb{E}^{\mathbb{P}}[\xi_t^{1-1/\gamma}]dt + e^{-(r+q)T}\mathbb{E}^{\mathbb{P}}[\xi_T^{1-1/\gamma}]},$$

where we have used Fubini's theorem alongside the shorthand notation $q \equiv (\delta - r)/\gamma$. Substituting this expression back into (C.19) we thus have

$$c_s^* = \frac{e^{-qs}\xi_s^{-1/\gamma}w_0}{h(0)}, \qquad \text{and} \qquad \mathscr{W}_T^* = \frac{e^{-qT}\xi_T^{-1/\gamma}w_0}{h(0)}, \tag{C.20}$$

where we for for convenience have introduced the function $h : [0,T] \mapsto \mathbb{R}$

$$h(s) = \int_s^T e^{-(r+q)(t-s)}\mathbb{E}_s^{\mathbb{P}}\left[\left(\frac{\xi_t}{\xi_s}\right)^{1-1/\gamma}\right]dt + e^{-(r+q)(T-s)}\mathbb{E}_s^{\mathbb{P}}\left[\left(\frac{\xi_T}{\xi_s}\right)^{1-1/\gamma}\right]. \tag{C.21}$$

It is no coincidence that we have recycled the notation $h(\cdot)$ for the function above - in fact, we shall now show that (C.16) and (C.21) are two sides of the same coin. To this end, let us first evaluate the expectations. While this is a straightforward exercise under the $\mathbb{P}$-measure (see [13]), a measure transformation argument is more interesting. Let $\mathbb{Q}_0$ be the measure defined through $\xi_t^0 = d\mathbb{Q}^0/d\mathbb{P}(t)$ where $d\xi_t^0 = -(1-1/\gamma)\xi_t^0\boldsymbol{\lambda}^\intercal d\boldsymbol{W}_t^{\mathbb{P}}$ then

$$\begin{aligned}
\mathbb{E}_s^{\mathbb{P}}\left[\left(\frac{\xi_t}{\xi_s}\right)^{1-1/\gamma}\right] &= \mathbb{E}_s^{\mathbb{P}}\left[e^{-(1-1/\gamma)\int_s^t \boldsymbol{\lambda}^\intercal d\boldsymbol{W}_u^{\mathbb{P}} - \frac{1}{2}(1-1/\gamma)\int_s^t ||\boldsymbol{\lambda}||^2 du}\right] \\
&= \mathbb{E}_s^{\mathbb{P}}\left[e^{-(1-1/\gamma)\int_s^t \boldsymbol{\lambda}^\intercal d\boldsymbol{W}_u^{\mathbb{P}} - \frac{1}{2}(1-1/\gamma)^2\int_s^t ||\boldsymbol{\lambda}||^2 du} e^{\frac{1-\gamma}{2\gamma^2}\int_s^t ||\boldsymbol{\lambda}||^2 du}\right] \\
&= \mathbb{E}_s^{\mathbb{P}}\left[\frac{\xi_t^0}{\xi_s^0} e^{\frac{1-\gamma}{2\gamma^2}||\boldsymbol{\lambda}||^2(t-s)}\right] \\
&= \mathbb{E}_s^{\mathbb{Q}_0}\left[e^{\frac{1-\gamma}{2\gamma^2}||\boldsymbol{\lambda}||^2(t-s)}\right] = e^{\frac{1-\gamma}{2\gamma^2}||\boldsymbol{\lambda}||^2(t-s)},
\end{aligned} \tag{C.22}$$

where the last line makes use of the abstract Bayes' formula (A.2). Inserting this into (C.21) and noting that $r + q - \frac{1-\gamma}{2\gamma^2}||\boldsymbol{\lambda}||^2 = \zeta$ we find after a few manipulations that $h(s)$ is indeed equal to equation (C.16).

To establish equivalence between the optimal consumption rate given in (C.20) and the expression given in (C.17) we need now only derive an expression for the optimal wealth level $\mathscr{W}_s^*$ for arbitrary $s \in [0,T]$. Upon substituting this expression into $c_s^* = e^{-qs}\xi_t^{-1/\gamma}w_0/h(0)$ the result follows immediately. Our starting point is the budget equation evaluated at time s, which we shall subject to the measure transformation $\mathbb{Q} \mapsto \mathbb{P}$ using the abstract Bayes' formula:

$$
\begin{aligned}
\mathscr{W}_s^* &= \mathbb{E}_s^{\mathbb{Q}}\left[\int_s^T e^{-r(t-s)} c_t^* dt + e^{-r(T-s)} \mathscr{W}_T^*\right] \\
&= \frac{1}{\xi_s}\mathbb{E}_s^{\mathbb{P}}\left[\int_s^T e^{-r(t-s)} \xi_t c_t^* dt + e^{-r(T-s)} \xi_T \mathscr{W}_T^*\right] \\
&= \frac{1}{\xi_s}\mathbb{E}_s^{\mathbb{P}}\left[\int_s^T e^{-r(t-s)} \xi_t \frac{e^{-qt}\xi_t^{-1/\gamma} w_0}{h(0)} dt + e^{-r(T-s)} \xi_T \frac{e^{-qT}\xi_T^{-1/\gamma} w_0}{h(0)}\right] \\
&= \frac{w_0 e^{-qs}\xi_s^{-1/\gamma}}{h(0)}\mathbb{E}_s^{\mathbb{P}}\left[\int_s^T e^{-(r+q)(t-s)} \left(\frac{\xi_t}{\xi_s}\right)^{1-1/\gamma} dt + e^{-(r+q)(T-s)} \left(\frac{\xi_T}{\xi_s}\right)^{1-1/\gamma}\right] \\
&= \frac{w_0 e^{-qs}\xi_s^{-1/\gamma} h(s)}{h(0)}. \qquad \text{(C.23)}
\end{aligned}
$$

Substituting this into c_s^* we obtain the desired expression

$$
c_s^* = \frac{\mathscr{W}_s^*}{h(s)}. \qquad \text{(C.24)}
$$

Finally, we are left with the question of the optimal investment ratio $\boldsymbol{\pi}_s^*$. Applying Itō to (C.23) and using the fact that $d\xi_s = -\xi_s \boldsymbol{\lambda}^\intercal d\boldsymbol{W}_s^{\mathbb{P}}$ we get

$$
d\mathscr{W}_s^* = \text{drift} + \mathscr{W}_s^* \tfrac{1}{\gamma}\boldsymbol{\lambda}^\intercal d\boldsymbol{W}_s^{\mathbb{P}}.
$$

At the same time we remind the reader of the (optimal) self-financing condition

$$
d\mathscr{W}_s^* = \mathscr{W}_s^*[r + (\boldsymbol{\pi}^*)^\intercal \boldsymbol{\sigma}\boldsymbol{\lambda}]ds - c_s^* ds + \mathscr{W}_s^*(\boldsymbol{\pi}^*)^\intercal \boldsymbol{\sigma} d\boldsymbol{W}_s^{\mathbb{P}}.
$$

Comparing the diffusion terms we immediately see that

$$
\boldsymbol{\pi}_s^* = \frac{(\boldsymbol{\sigma}^\intercal)^{-1}\boldsymbol{\lambda}}{\gamma}, \qquad \text{(C.25)}
$$

as desired.

C.4 Discussion

The Hamilton-Jacobi-Bellman equation and the martingale method thus offer complementary approaches to extracting optimal controls to stochastic control problems. For the Merton problem treated here, there is no significant difference in the computational intensity between the two approaches; nor any difference between the level of rigour, yet considerable differences can arise for more general problems. Here, we briefly summarise important points to keep in mind when tackling more general problems:

- Writing down the HJB (and in some cases: solving it) is the easy part. Formally verifying that the HJB equation is also sufficient for optimality usually requires a higher level of mathematical sophistication. E.g. it is worth noting that the standard body of theory for which verification is considered trivial excludes important financial models such as the CIR process qua the requirement that the Lipschitz condition holds.
- Assuming that the optimal value function V is smooth ($\mathscr{C}^{1,2}$) is generally not innocuous. In fact, it is simply not true even for relatively simple examples (see e.g. Pham section 3.7 [18]). To mitigate this deplorable situation Crandall and Lions introduced the concept of **viscosity solutions** in the 1980s thus providing us with a way to formulate HJB equations rigorously whilst only assuming local boundedness of the optimal value functions.
- While the martingale approach circumnavigates these issues, it is correspondingly more difficult to get explicit expressions for when the market is **incomplete**. For an introduction to so-called **convex duality** methods in incomplete markets the reader is referred to Schachermayer [21].
- An important assumption when using the HJB formalism is the Markovian structure on the dynamical equations (e.g. $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ have to be of the form $\boldsymbol{\mu}(t,\boldsymbol{S}_t)$ and $\boldsymbol{\sigma}(t,\boldsymbol{S}_t)$ (*)). No such assumption prevails for the martingale approach, which makes it more general ($\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are allowed to be arbitrary adapted path dependent processes).
- Nonetheless, a Markovian structure is advantageous to work with even for the martingale approach. E.g. it can enable us to write down nice linear PDEs for the optimal controls (or the components thereof at least) while the HJB formalism is intrinsically non-linear. From a computational perspective (be it analytically or numerically) the former is obviously to be preferred. To appreciate this point to a greater extent suppose we consider the local volatility model (*) above *sans* consumption then (C.22) reads

$$\mathbb{E}_s^{\mathbb{Q}_0}\left[e^{\frac{1-\gamma}{2\gamma^2}\int_s^T ||\boldsymbol{\lambda}(u,\boldsymbol{S}_u)||^2 du}\right] =: g(s,\boldsymbol{S}_s). \tag{C.26}$$

Recalling that $\xi_s^0 = d\mathbb{Q}_0/d\mathbb{P}(s)$ where $d\xi_s^0 = -(1-1/\gamma)\xi_s^0\boldsymbol{\lambda}^\intercal d\boldsymbol{W}_s^{\mathbb{P}}$ it follows form Girsanov's theorem (A.9) that

$$d\boldsymbol{W}_s^{\mathbb{P}} = -(1-1/\gamma)\boldsymbol{\lambda}ds + d\boldsymbol{W}_s^{\mathbb{Q}_0},$$

whence

$$d\boldsymbol{S}_s = \mathrm{diag}(\boldsymbol{S}_s)[\{\gamma^{-1}\boldsymbol{\mu}(t,\boldsymbol{S}_s) + (1-\gamma^{-1})r\boldsymbol{\iota}\}ds + \boldsymbol{\sigma}(t,\boldsymbol{S}_s)d\boldsymbol{W}_s^{\mathbb{Q}_0}].$$

Thus, from Feynman-Kac (B.1) we deduce that (C.26) is a solution of

$$\begin{aligned}0 =& \partial_s g(s,\boldsymbol{x}) + [\mathrm{diag}(\boldsymbol{x})\{\gamma^{-1}\boldsymbol{\mu}(t,\boldsymbol{x}) + (1-\gamma^{-1})r\boldsymbol{\iota}\}]^\intercal \nabla_{\boldsymbol{x}} g(s,\boldsymbol{x}) + \\ &\frac{1}{2}\mathrm{tr}[\mathrm{diag}(\boldsymbol{x})\boldsymbol{\sigma}(t,\boldsymbol{x})\boldsymbol{\sigma}(t,\boldsymbol{x})^\intercal \mathrm{diag}(\boldsymbol{x})\nabla^2_{\boldsymbol{x}\boldsymbol{x}} g(s,\boldsymbol{x})] + \frac{1-\gamma}{2\gamma^2}||\boldsymbol{\lambda}(u,\boldsymbol{x})||^2 g(s,\boldsymbol{x}),\end{aligned} \tag{C.27}$$

with terminal condition $g(T,\boldsymbol{x}) = 1$.

References

1. Baltas, N., *Stochastic Calculus for Finance*, Unpublished lecture notes, Imperial College Business School, 2011.
2. Björk, T., *Arbitrage Theory in Continuous Time*, Oxford University Press, 2009. 3rd Edition.
3. Björk, T., *Equilibrium Theory in Continuous Time*, Unpublished lecture notes, 2012. http://www.math.kth.se/matstat/fofu/Equilibrium%20Course%20Lecture%20Notes.pdf.
4. Burden, R. and Faires, J., *Numerical Analysis*. Brooks/Cole, 9th edition.
5. Cont, R. (ed), *Encyclopedia of Quantitative Finance*, John Wiley & Sons; IV Volume Set edition (23 Feb. 2010).
6. Davis, M.H.A., *Martingale Methods in Stochastic Control*, Jan. 1979. https://dspace.mit.edu/bitstream/handle/1721.1/891/P-0874-15596929.pdf?sequence=1.
7. Duffie, D., *Dynamic Asset Pricing Theory*. 2001, Princeton University Press, 3rd edition.
8. Hull, J., *Options, Futures, and Other Derivatives*. Pearson, 7th edition.
9. Ingersoll, J., *Financial Decision Making*, Rowman and Littlefield Studies in Financial Economics, 1987.
10. Karatzas, I. and Shreve, S., *Methods of Mathematical Finance*, 1998 Springer-Verlag New York, Inc.
11. Klebaner, F., *Introduction to Stochastic Calculus with Applications*, 2nd edition, 2005, Imperial College Press.
12. Merton, R., *Continuous-time Finance*, Blackwell Publishing, Revised Edition, 6. Aug. 1992.
13. Munk, C., *Dynamic Asset Allocation*, Unpublished lecture notes (18 Dec. 2013).
14. Munk, C., *Financial Asset Pricing Theory*, OUP UK, reprint edition (12 Feb. 2015).
15. Munk, C., *Fixed Income Modelling*, OUP UK (1 Sept. 2011).
16. Øksendal, B., *Stochastic Differential Equations: An Introduction with Applications*, (Universitext), Springer; 2003. Corr. 5th edition (4 Feb. 2014).
17. Østerby, O., *Numerical Solutions of Parabolic Equations*, Unpublished lecture notes. http://faculty.ksu.edu.sa/rizwanbutt/Documents/pde.pdf.
18. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications*. *Stochastic Modelling and Applied Probability 61*, 2010, Springer. e-ISBN: 978-3-540-89500-8.
19. Privault, N., *Notes on Stochastic Finance*, 2013. Lecture Notes FE6516, http://www.ntu.edu.sg/home/nprivault/indext.html.
20. Saad, Y., *Iterative Methods for Sparse Matrices*, Society for Industrial and Applied Mathematics; 2 edition (April 30, 2003).
21. Schachermayer, W., *Optimal Investment in Incomplete Financial Markets*. *Mathematical Finance?Bachelier Congress 2000*, H. Geman, Ed. Springer-Verlag, Berlin Heidelberg New York, 2002.
22. Seydel, R., *Tools for Computational Finance*. Springer; 2012 edition (14 Mar. 2012).
23. Shreve, S., *Stochastic Calculus for Finance II*, 2008. Springer.
24. Tourin, *An introduction to Finite Difference methods for PDEs in Finance*, Lecture notes, 2010, Fields Institute, Toronto.
25. Wilmott, P., *Paul Wilmott Introduces Quantitative Finance*, 2nd edition, John Wiley & Sons, Ltd.
26. Zheng, H., *Numerical Finance*, Unpublished lecture notes, Imperial College Business School, 2012.