

Essays in Capital Markets

by

Leonid Kogan

Ph.D., Theoretical and Applied Mechanics, Cornell University, 1995

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1999

© Massachusetts Institute of Technology 1999. All rights reserved.

Author ...

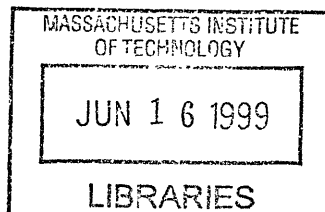
.....
Sloan School of Management
May 25, 1999

Certified by

Andrew Lo
Andrew Lo
Harris & Harris Group Professor
Thesis Supervisor

Accepted by

Birger Wernerfelt
Chairman, Department Committee on Graduate Students



ARCHIVES

Essays in Capital Markets

by
Leonid Kogan

Submitted to the Sloan School of Management
on May 25, 1999, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis consists of three essays in capital markets.

In the first essay, given a European derivative security with an arbitrary payoff function and a corresponding set of underlying securities on which the derivative security is based, we solve the optimal-replication problem: find a self-financing dynamic portfolio strategy—involving only the underlying securities—that most closely approximates the payoff function at maturity. By applying stochastic dynamic programming to the minimization of a mean-squared-error loss function under Markov state-dynamics, we derive recursive expressions for the optimal-replication strategy that are readily implemented in practice. The approximation error or “ ϵ ” of the optimal-replication strategy is also given recursively and may be used to quantify the “degree” of market incompleteness. To investigate the practical significance of these ϵ -arbitrage strategies, we consider several numerical examples including path-dependent options and options on assets with stochastic volatility and jumps.

In the second essay we study the tracking error, resulting from the discrete-time application of continuous-time delta-hedging procedures for European options. We characterize the asymptotic distribution of the tracking error as the number of discrete time periods increases, and its joint distribution with other assets. We introduce the notion of temporal granularity of the continuous time stochastic model that enables us to characterize the degree to which discrete time approximations of continuous time models track the payoff of the option. We derive closed form expressions for the granularity for a put or call option on a stock that follows a geometric Brownian motion and a mean-reverting process. These expressions offer insight into the tracking error involved in applying continuous-time delta-hedging in discrete time. We also introduce alternative measures of the tracking error and analyze their properties.

The third essay presents a general equilibrium model of financial asset prices with irreversible real investment. The focus is on the effects of the irreversibility of real investment on financial asset prices. The model shows how this irreversibility leads to time variation in volatility and systematic risk of stock returns. Changes in these variables are driven by real economic activity, in particular, by firms' investment decisions. Thus, systematic risk of stock returns and their volatility are affected by economy-wide and industry-specific shocks. Firm-specific variables, particularly market-to-book ratios, are linked to real activity and contain information about the dynamic behavior of stock returns. The model of this paper also provides a framework for analyzing futures prices. A comparison between the economy with irreversible investment and an identical economy without the irreversibility shows that all of these results should be attributed to the irreversibility of real investment.

Thesis Supervisor: Andrew Lo
Title: Harris & Harris Group Professor

Acknowledgments

I would like to thank my thesis advisor Andrew Lo for his guidance, support and the invariable enthusiasm he brought into the research process. I am also grateful to my committee members Jiang Wang and Stephen Ross for many exciting discussions and their insightful comments.

I am grateful to Marina for her love and devotion. I also owe a lot to my fellow Ph.D. students who have helped make my life at MIT more enjoyable. I thank Harry for many academic and not-so-academic discussions and for exchanging stock tips with me (particularly for not taking my tips too seriously).

I thank Dimitris Bertsimas and Andrew Lo for supporting me as a Research Assistant and Lehman Brothers for awarding me the Lehman Brothers Fellowship.

Contents

1 Hedging Derivative Securities and Incomplete Markets: An ϵ-Arbitrage Approach (joint with Dimitris Bertsimas and Andrew Lo)	17
1.1 Abstract	17
1.2 Introduction	19
1.3 ϵ -Arbitrage Strategies	24
1.3.1 The Optimal Replication Problem	24
1.3.2 Examples	28
1.3.3 ϵ -Arbitrage in Discrete Time	30
1.3.4 ϵ -Arbitrage in Continuous Time	34
1.3.5 Interpreting ϵ^* and V_0^*	37
1.4 Illustrative Examples	40
1.4.1 State-Independent Returns	41
1.4.2 Geometric Brownian Motion	42
1.4.3 Jump-Diffusion Models	45
1.4.4 Stochastic Volatility	51
1.5 Numerical Analysis	53
1.5.1 The Numerical Procedure	53
1.5.2 Geometric Brownian Motion	54
1.5.3 Jump-Diffusion Models	56
1.5.4 Stochastic Volatility	61
1.5.5 Path-Dependent Options	66
1.6 Measuring the Degree of Market Incompleteness	71
1.6.1 Calibrating the Stochastic Processes	71
1.6.2 Numerical Results	73
1.7 Conclusion	77
1.8 Appendix	79
1.8.1 Proof of Theorem 1.1	79
1.8.2 Proof of Theorem 1.2	80

2	When Is Time Continuous? (joint with Dimitris Bertsimas and Andrew Lo)	86
2.1	Abstract	86
2.2	Introduction	87
2.3	Defining Temporal Granularity	90
2.3.1	Delta-Hedging in Complete Markets	93
2.3.2	Asymptotic Behavior of the Tracking Error and RMSE	97
2.3.3	Interpretation of Granularity	101
2.4	Applications	103
2.4.1	Granularity of Geometric Brownian Motion	103
2.4.2	Granularity of a Mean-Reverting Process	105
2.5	Monte Carlo Analysis	106
2.5.1	Accuracy of the Asymptotics	106
2.5.2	Qualitative Behavior of the RMSE	111
2.6	Extensions and Generalizations	114
2.6.1	Sample-Path Properties of Tracking Errors	114
2.6.2	Joint Distributions of Tracking Errors and Prices	116
2.6.3	A PDE Characterization of the Tracking Error	118
2.6.4	Alternative Measures of the Tracking Error	119
2.7	Conclusions	122
2.8	Appendix	124
2.8.1	Proof of Theorem 1(a)	124
2.8.2	Proof of Theorem 1(b)	127
2.8.3	Proof of Theorem 1(c)	129
2.8.4	Proof of Theorem 2(a)	129
2.8.5	Proof of Theorem 2(b)	134
2.8.6	Proof of Theorem 3	136
2.8.7	Proof of Theorem 4	137
3	Asset Prices and Irreversible Investment	142
3.1	Abstract	142
3.2	Introduction	143
3.3	Literature Review	145
3.4	Irreversible Investment in General Equilibrium	149
3.4.1	The Central Planner's Problem	152
3.4.2	The Solution of the Central Planner's Problem	155
3.4.3	The Competitive Equilibrium	162
3.4.4	Closed-Form Approximations	167
3.5	Reversible Investment in General Equilibrium	171
3.5.1	The Central Planner's Problem	171
3.5.2	The Solution of the Central Planner's Problem	173

3.5.3	The Competitive Equilibrium	174
3.6	Irreversibility and Stock Returns	175
3.6.1	Volatility	177
3.6.2	Systematic Risk	180
3.7	Extensions and Generalizations	185
3.8	Conclusion	191
3.9	Appendix	192
3.9.1	Futures Prices	192
3.9.2	Proof of Proposition 3.3	202
3.9.3	Numerical Solution of (3.16–3.18, 3.24, 3.25)	205
3.9.4	Proof of Proposition 3.4	208
3.9.5	Proof of Proposition 3.6	217
3.9.6	Perturbation Analysis, Section 3.4.4	222
3.9.7	Proof of Proposition 3.5.2	228

List of Figures

- 1-1 The difference between the replication cost and the intrinsic value of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows a geometric Brownian motion with parameter values $\mu = 0.07$ and $\sigma = 0.13$ corresponding to the solid line. In Panel (a), μ is varied and σ is fixed; in Panel (b), σ is varied and μ is fixed. In both cases, the variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line) and 0.5 (pluses). 55
- 1-2 The replication error of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows a geometric Brownian motion with parameter values $\mu = 0.07$ and $\sigma = 0.13$ corresponding to the solid line. In Panel (a), μ is varied and σ is fixed; in Panel (b), σ is varied and μ is fixed. In both cases, the variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line) and 0.5 (pluses). 55
- 1-3 The relative replication error of a six-month maturity European put option (relative to the replication cost), plotted as a function of the initial stock price. The stock price follows a geometric Brownian motion with parameter values $\mu = 0.07$ and $\sigma = 0.13$ corresponding to the solid line. In Panel (a), μ is varied and σ is fixed; in Panel (b), σ is varied and μ is fixed. In both cases, the variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line) and 0.5 (pluses). 57
- 1-4 The difference between the replication cost and the intrinsic value of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process given in (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses). 58

- 1-5 The replication error of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process given in (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses). 59
- 1-6 The relative replication error of a six-month maturity European put option (relative to the replication cost), plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses). 60
- 1-7 The difference between the replication cost and the intrinsic value of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows the stochastic volatility model (1.84)–(1.85) with parameter values $\mu = 0.07$, $\zeta = 0.153$, $\delta = 2$, $\kappa = 0.4$, and $\sigma_0 = 0.13$ corresponding to the solid line. In Panels (a)–(d), ζ , δ , κ , and σ_0 are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses). 62
- 1-8 The replication error of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows the stochastic volatility model (1.84)–(1.85) with parameter values $\mu = 0.07$, $\zeta = 0.153$, $\delta = 2$, $\kappa = 0.4$, and $\sigma_0 = 0.13$ corresponding to the solid line. In Panels (a)–(d), ζ , δ , κ , and σ_0 are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses). 63

- 1-9 The relative replication error of a six-month maturity European put option (relative to the replication cost), plotted as a function of the initial stock price. The stock price follows the stochastic volatility model (1.84)–(1.85) with parameter values $\mu = 0.07$, $\zeta = 0.153$, $\delta = 2$, $\kappa = 0.4$, and $\sigma_0 = 0.13$ corresponding to the solid line. In Panels (a)–(d), ζ , δ , κ , and σ_0 are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses). 64
- 1-10 The replication cost of a six-month maturity European option to “sell at the high”, plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $m = 1$, $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses). 67
- 1-11 The replication error of a six-month maturity European option to “sell at the high”, plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $m = 1$, $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses). 69
- 1-12 The relative replication error of a six-month maturity European option to “sell at the high” (relative to the replication cost), plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $m = 1$, $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses). 70
- 1-13 The difference between the replication cost and the intrinsic value of a six-month maturity European put option, plotted as a function of the initial stock price. Several processes for the stock price are plotted: geometric Brownian motion (1.41)–(1.43) (solid line); the mixed jump-diffusion model (1.48)–(1.51), (1.54), (1.82), and (1.83) (dashed line); and the stochastic volatility model (1.84)–(1.85) (dashed-dotted line). The parameter values are given by (1.87), (1.88), and (1.89). 72

1-14	The replication error of a six-month maturity European put option, plotted as a function of the initial stock price. Several processes for the stock price are plotted: geometric Brownian motion (1.41)–(1.43) (solid line); the mixed jump-diffusion model (1.48)–(1.51), (1.54), (1.82), and (1.83) (dashed line); and the stochastic volatility model (1.84)–(1.85) (dashed-dotted line). The parameter values are given by (1.87), (1.88), and (1.89).	75
1-15	The relative replication error of a six-month maturity European put option (relative to the replication cost), plotted as a function of the initial stock price. Several processes for the stock price are plotted: geometric Brownian motion (1.41)–(1.43) (solid line); the mixed jump-diffusion model (1.48)–(1.51), (1.54), (1.82), and (1.83) (dashed line); and the stochastic volatility model (1.84)–(1.85) (dashed-dotted line). The parameter values are given by (1.87), (1.88), and (1.89).	76
2-1	Empirical probability density functions of (a) the tracking error and (b) the normalized tracking error (dashed line), are plotted for different values of the trading frequency N . Figure (b) also shows the empirical probability density function of the asymptotic distribution (2.17) (solid line). The stock price process is given by (2.19) with parameters $\mu = 0.1$, $\sigma = 0.3$, $P_0 = 1.0$. The option is a European call (put) option with strike price $K = 1$	107
2-2	(a) The logarithm of the root-mean-squared error $\log_{10}(\text{RMSE}^{(N)})$ is plotted as a function of the logarithm of the trading frequency $\log_{10}(N)$. The option is a European call (put) option with the strike price $K = 1$. The stock price process is given by (2.19) with parameters $\mu = 0.1$, $P_0 = 1.0$. The diffusion coefficient of the stock price process takes values $\sigma = 0.3$ (x's), $\sigma = 0.2$ (o's) and $\sigma = 0.1$ (+'s). (b) The root-mean-squared error RMSE is plotted as a function of the initial stock price P_0 . The option is a European put option with the strike price $K = 1$. Parameters of the stock price process are $\mu = 0.1$, $\sigma = 0.3$	111
2-3	The granularity g is plotted as a function of σ and μ . The option is a European call (put) option with strike price $K = 1$. The stock price process is geometric Brownian motion and initial stock price $P_0 = 1$	112
2-4	Granularity g is plotted as a function of P_0 and α . The option is a European call (put) option with the strike price $K = 1$. Parameters of the stock price process are $\sigma = 0.2$, $\mu = 0.05$. The stock price process is given by (2.22). Mean-reversion parameter γ takes two values: (a) $\gamma = 0.1$ and (b) $\gamma = 3.0$. . .	113

- 3-1 The optimal consumption policy $\tilde{c}_1^*(\Omega)$ (solid), the first-order approximation (dash) and the second-order approximation (dash-dot) are plotted against the ratio of the capital stocks $\Omega = K_2/K_1$. The following set of parameter values is used: $b = 0.05$, $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$ 170
- 3-2 The optimal consumption policy $\tilde{c}_1^*(\Omega)$ (solid), the first-order approximation (dash) and the second-order approximation (dash-dot) are plotted against the ratio of the capital stocks $\Omega = K_2/K_1$. The following set of parameter values is used: $b = 0.1$, $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$ 171
- 3-3 The instantaneous standard deviation of the industry portfolio is plotted as a function of the industry's q . The following set of parameter values is used: $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.01$, $\sigma_1 = 0.124$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.08$. Four different values of σ_{21} are used: 0.0, 0.02, 0.04, 0.06. 180
- 3-4 The instantaneous standard deviation of the industry portfolio is plotted as a function of the industry's q . The following set of parameter values is used: $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$. Four different values of σ_{21} are used: 0.0, 0.06, 0.12, 0.18. 181
- 3-5 The beta of the industry portfolio is plotted as a function of the industry's q . The following set of parameter values is used: $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$. Four different values of σ_{21} are used: 0.0, 0.06, 0.12, 0.18. 183
- 3-6 The beta of the industry portfolio is plotted as a function of the industry's q . The following set of parameter values is used: $\gamma = 2$, $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.06$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$, $i_{\max} = 1$ 189
- 3-7 q is plotted against its argument, $\xi - \xi^*$. The following set of parameter values is used: $\gamma = 2$, $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.06$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$, $i_{\max} = 1$ 190
- 3-8 A time series of the spot price. The following set of parameter values is used: $\rho = 0.05$, $\gamma = 2$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$ 194
- 3-9 The term structure of futures prices is plotted as a function of q . The following set of parameter values is used: $\rho = 0.05$, $\gamma = 2$, $\alpha = 0.067$, $\delta = -0.02$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$ 199
- 3-10 The term structure of volatility of futures prices is plotted as a function of q . The following set of parameter values is used: $\rho = 0.05$, $\gamma = 2$, $\alpha = 0.067$, $\delta = -0.02$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$ 200

List of Tables

1.1	Comparison of optimal-replication strategy and Black-Scholes delta-hedging strategy for replicating an at-the-money put option on 1,000 shares of a \$40-stock over 25 trading periods for two simulated sample paths of a geometric Brownian motion with parameters $\mu = 0.07$ and $\sigma = 0.13$	23
1.2	Monte Carlo simulation of the optimal-replication strategy θ^* for replicating a six-month at-the-money European put-option, for six sets of parameter values of the stochastic volatility model (1.84)–(1.85), including the set of parameter values that yields a geometric Brownian motion (last row). For each set of parameter values, 1,000 independent sample paths were simulated, each path containing 25 periods, and $P_0 = 1$	65
1.3	Comparison of replication costs and errors of the optimal-replication strategy for replicating a six-month European put option under competing specifications of price dynamics: geometric Brownian motion (1.41)–(1.43); the mixed jump-diffusion model (1.48)–(1.51), (1.54), (1.82), and (1.83); and the stochastic volatility model (1.84)–(1.85). The parameter values are given by (1.87), (1.88), and (1.89).	74
2.1	The sensitivity of the RMSE as a function of the initial price P_0 . The RMSE is estimated using Monte Carlo simulation. Options are European calls and puts with strike price $K = 1$. 250,000 simulations are performed for every set of parameter values. The stock price follows a geometric Brownian motion (2.19). The drift and diffusion coefficients of the stock price process are $\mu = 0.1$ and $\sigma = 0.3$. $\text{RMSE}^{(N)}$ is compared to the asymptotic approximation $gN^{-1/2}$ in (2.13)–(2.20). The relative error (RE) of the asymptotic approximation is defined as $ gN^{-1/2} - \text{RMSE}^{(N)} /\text{RMSE}^{(N)} \times 100\%$	108

- 2.2 The sensitivity of the RMSE as a function of volatility σ . The RMSE is estimated using Monte Carlo simulation. Options are European calls and puts with strike price $K = 1$. 250,000 simulations are performed for every set of parameter values. The stock price follows a geometric Brownian motion (2.19). The drift coefficient of the stock price process is $\mu = 0.1$, and the initial stock price is $P_0 = 1.0$. $\text{RMSE}^{(N)}$ is compared to the asymptotic approximation $gN^{-1/2}$ in (2.13)–(2.20). The relative error (RE) of the asymptotic approximation is defined as $|gN^{-1/2} - \text{RMSE}^{(N)}|/\text{RMSE}^{(N)} \times 100\%$. 109
- 2.3 The sensitivity of the RMSE as a function of the drift μ . The RMSE is estimated using Monte Carlo simulation. Options are European calls and puts with strike price $K = 1$. 250,000 simulations are performed for every set of parameter values. The stock price follows a geometric Brownian motion (2.19). The diffusion coefficient of the stock price process is $\sigma = 0.3$, and the initial stock price is $P_0 = 1.0$, the number of trading periods is $N = 20$. $\text{RMSE}^{(N)}$ is compared to the asymptotic approximation $gN^{-1/2}$ in (2.13)–(2.20). The relative error (RE) of the asymptotic approximation is defined as $|gN^{-1/2} - \text{RMSE}^{(N)}|/\text{RMSE}^{(N)} \times 100\%$ 110
- 3.1 The optimal investment threshold Ω^* , the first-order and the second-order approximations are tabulated for different values of the risk-aversion parameter γ and b . The following set of parameter values is used: $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$ 172
- 3.2 The first two unconditional moments of the beta of the industry portfolio are tabulated for several sets of parameter values. The following parameter values are used throughout the table: $\rho = 0.05$, $\alpha = 0.067$, $\sigma_1 = 0.124$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$ 186

Chapter 1

Hedging Derivative Securities and Incomplete Markets: An ϵ -Arbitrage Approach (joint with Dimitris Bertsimas

and Andrew Lo)

1.1 Abstract

Given a European derivative security with an arbitrary payoff function and a corresponding set of underlying securities on which the derivative security is based, we solve the optimal-replication problem: find a self-financing dynamic portfolio strategy—involving only the underlying securities—that most closely approximates the payoff function at maturity. By applying stochastic dynamic programming to the minimization of a mean-squared-error loss function under Markov state-dynamics, we derive recursive expressions for the optimal-replication strategy that are readily implemented in practice. The approximation error or

“ ϵ ” of the optimal-replication strategy is also given recursively and may be used to quantify the “degree” of market incompleteness. To investigate the practical significance of these ϵ -arbitrage strategies, we consider several numerical examples including path-dependent options and options on assets with stochastic volatility and jumps.

1.2 Introduction

One of the most important breakthroughs in modern financial economics is Merton's (1973) insight that under certain conditions the frequent trading of a small number of long-lived securities can create new investment opportunities that would otherwise be unavailable to investors. These conditions—now known collectively as *dynamic spanning* or *dynamically complete markets*—and the corresponding asset-pricing models on which they are based have generated a rich literature and an even richer industry in which complex financial securities are synthetically replicated by sophisticated trading strategies involving considerably simpler instruments.¹ This approach is the basis of the celebrated Black and Scholes (1973) and Merton (1973) option-pricing formula, the arbitrage-free method of pricing and, more importantly, hedging other derivative securities, and the martingale characterization of prices and dynamic equilibria.

The essence of dynamic spanning is the ability to replicate exactly the payoff of a complex security by a dynamic portfolio strategy of simpler securities which is *self-financing*, i.e., no cash inflows or outflows except at the start and at the end. If such a *dynamic-hedging* strategy exists, then the initial cost of the portfolio must equal the price of the complex security, otherwise an arbitrage opportunity exists. For example, under the assumptions of Black and Scholes (1973) and Merton (1973), the payoff of a European call-option on a non-dividend-paying stock can be replicated exactly by a dynamic-hedging strategy involving only stocks and riskless borrowing and lending.

But the conditions that guarantee dynamic spanning are nontrivial restrictions on market structure and price dynamics (see, for example, Duffie and Huang [1985]), hence there

¹In addition to Merton's seminal paper, several other important contributions to the finance literature are responsible for our current understanding of dynamic spanning. In particular, see Cox and Ross (1976), Duffie (1985), Duffie and Huang (1985), Harrison and Kreps (1979), and Huang (1985a,b).

are situations in which exact replication is impossible.² These instances of *market incompleteness* are often attributable to institutional rigidities and market frictions—transactions costs, periodic market closures, and discreteness in trading opportunities and prices—and while the pricing of complex securities can still be accomplished in some cases via equilibrium arguments,³ this still leaves the question of optimal replication unanswered. Perfect replication is impossible in dynamically incomplete markets, but how close can one come, and what does the optimal-replication strategy look like?

In this paper we answer these questions by applying optimal control techniques to the optimal-replication problem: given an arbitrary payoff function and a set of fundamental securities, find a self-financing dynamic portfolio strategy involving only the fundamental securities that most closely approximates the payoff in a mean-squared sense. The initial cost of such an optimal strategy can be viewed as the “production cost” of the option, i.e., it is the cost of the best dynamic approximation to the payoff function given the set of fundamental securities traded. Such an interpretation is more than a figment of economic imagination—the ability to synthesize options via dynamic trading strategies is largely responsible for the growth of the multi-trillion-dollar over-the-counter derivatives market.⁴

Of course, the nature of the optimal-replication strategy depends on how we measure the closeness of the payoff and its approximation. For tractability and other reasons (see Section 1.3.5), we choose a mean-squared-error loss function and we denote by ϵ the root-

²Suppose, for example, that stock price volatility σ in the Black and Scholes (1973) framework is stochastic.

³Examples of continuous-time incomplete-markets models include Breeden (1979), Duffie (1987), Duffie and Shafer (1985, 1986), Föllmer and Sonderman (1986), and He and Pearson (1991). Examples of discrete-time incomplete-markets models include Aiyagari (1994), Aiyagari and Gertler (1991), He and Modest (1995), Heaton and Lucas (1992, 1996), Lucas (1994), Scheinkman and Weiss (1986), Telmer (1993), and Weil (1992).

⁴In contrast to exchange-traded options such as equity puts and calls, over-the-counter derivatives are considerably more illiquid. If investment houses were unable to synthesize them via dynamic trading strategies, they would have to take the other side of every option position that their clients' wish to take (net of offsetting positions among the clients themselves). Such risk exposure would dramatically curtail the scope of the derivatives business, limiting both the size and type of contracts available to end users.

mean-squared-error of an optimal-replication strategy. In a dynamically complete market, the approximation error ϵ is identically zero, but when the market is incomplete, we propose ϵ as a measure of the “degree” of incompleteness. Although from a theoretical point of view dynamic spanning either holds or does not hold, a gradient for market completeness seems more natural from an empirical and a practical point of view. We provide examples of stochastic processes that imply dynamically incomplete markets, e.g., stochastic volatility, and yet still admit ϵ -arbitrage strategies for replicating options to within ϵ , where ϵ can be evaluated numerically.

In this respect, our contributions complement the results of Schweizer (1992, 1995) in which the optimal-replication problem is also solved for a mean-squared-error loss function. Schweizer considers more general stochastic processes than we do—we focus only on vector-Markov price processes—and uses variational principles to characterize the optimal-replication strategy. Although our approach can be viewed as a special case of his, the Markov assumption allows us to obtain considerably sharper results and yields an easily implementable numerical procedure (via dynamic programming) for determining the optimal-replication strategy and the replication error ϵ in practice.

Our results also complement the burgeoning literature on option pricing with transactions costs, e.g., Leland (1985), Hodges and Neuberger (1989), Bensaid, et al. (1992), Boyle and Vorst (1992), Davis, Panas, and Zariphopoulou (1993), Edirisinghe, Naik, and Uppal (1993), Henrotte (1993), Avellaneda and Paras (1994), Neuberger (1994), Whalley and Wilmott (1994), Grannan and Swindle (1996), and Toft (1996) (see, also, the related papers by Hutchinson, Lo, and Poggio (1994), Brandt (1998), and Bertsimas, Kogan, and Lo (1999)). In these studies, the existence of transactions costs induces discrete trading intervals, and the optimal replication is solved for some special cases, e.g., call and put options

on stocks with geometric Brownian motion or constant-elasticity-of-variance price dynamics. In this paper, we solve the more general problem of optimally replicating an arbitrary derivative security where the underlying asset is driven by a vector Markov process.

To demonstrate the practical relevance of our optimal-replication strategy, even in the simplest case of the Black and Scholes (1973) model where an explicit optimal-replication strategy is available, Table 1.2 presents a comparison of our optimal-replication strategy with the standard Black-Scholes “delta-hedging” strategy for replicating an at-the-money put option on 1,000 shares of a \$40-stock over 25 trading periods for two simulated sample paths of a geometric Brownian motion with drift $\mu = 0.07$ and diffusion coefficient $\sigma = 0.13$ (rounded to the nearest \$0.125).

V_t^* denotes the period- t value of the optimal replicating portfolio, θ_t^* denotes the number of shares of stock held in that portfolio, and V_t^{BS} and θ_t^{BS} are defined similarly for the Black-Scholes strategy.

Despite the fact that both sample paths are simulated geometric Brownian motions with identical parameters, the optimal-replication strategy has a higher replication error than the Black-Scholes strategy for path A and a lower replication error than Black-Scholes for path B.⁵ That the optimal-replication strategy underperforms the Black-Scholes strategy for path A is not surprising since the optimal-replication strategy is optimal only in a mean-squared sense (see Section 1.3.1), not path by path.⁶ That the Black-Scholes strategy underperforms the optimal-replication strategy for path B is also not surprising since the former is designed to replicate the option with continuous trading whereas the optimal-replication strategy is

⁵Specifically, $V_{25}^* - 1000 \times \text{Max}[0, \$40 - P_{25}] = \$199.1$ and $V_{25}^{\text{BS}} - 1000 \times \text{Max}[0, \$40 - P_{25}] = \$172.3$ for path A, and $V_{25}^* - 1000 \times \text{Max}[0, \$40 - P_{25}] = -\$40.3$ and $V_{25}^{\text{BS}} - 1000 \times \text{Max}[0, \$40 - P_{25}] = -\$299.2$ for path B.

⁶These two sample paths were chosen to be illustrative, not conclusive. In a more extensive simulation study in which 250,000 sample paths were generated, the average replication error of the Black-Scholes strategy is \$248.0 and the average error of the optimal-replication strategy is \$241.2.

designed to replicate the option with 25 trading periods.

t	Sample Path A					Sample Path B				
	P_t	V_t^*	θ_t^*	V_t^{BS}	θ_t^{BS}	P_t	V_t^*	θ_t^*	V_t^{BS}	θ_t^{BS}
0	40.000	1461.0	-474.8	1466.4	-481.7	40.000	1461.0	-474.8	1466.4	-481.7
1	40.750	1104.9	-394.2	1105.1	-400.8	39.875	1520.4	-491.8	1526.6	-495.9
2	42.125	562.9	-252.1	554.0	-264.0	40.250	1336.0	-452.4	1340.6	-454.3
3	41.375	751.9	-318.6	752.0	-331.8	36.500	3032.3	-762.4	3044.3	-845.8
4	42.000	552.8	-256.1	544.7	-267.2	36.875	2746.4	-744.1	2727.1	-822.1
5	43.125	264.7	-157.8	244.0	-169.5	36.500	3025.4	-777.9	3035.4	-858.3
6	43.250	245.0	-145.6	222.8	-155.1	37.000	2636.5	-750.1	2606.3	-824.5
7	42.250	390.6	-216.7	377.9	-229.5	39.875	479.9	-407.8	235.8	-500.5
8	43.000	228.2	-149.7	205.8	-160.6	39.875	479.9	-412.1	235.8	-501.4
9	41.750	415.2	-250.3	406.6	-267.9	40.125	376.9	-384.8	110.4	-468.4
10	42.000	352.7	-221.1	339.6	-235.5	39.500	617.3	-466.9	403.2	-556.1
11	42.625	214.5	-157.2	192.4	-168.9	41.250	-199.7	-227.2	-570.0	-315.0
12	41.750	352.1	-233.8	340.2	-248.5	40.625	-57.7	-300.8	-373.1	-394.7
13	41.500	410.5	-258.4	402.3	-271.0	39.875	167.9	-403.7	-77.0	-506.9
14	42.625	119.8	-128.5	97.5	-141.7	39.375	369.8	-482.4	176.4	-590.1
15	42.875	87.7	-100.5	62.1	-110.7	39.625	249.2	-452.1	28.9	-552.9
16	42.875	87.7	-91.5	62.1	-99.2	39.750	192.7	-439.2	-40.2	-534.3
17	43.125	64.8	-65.5	37.3	-70.4	39.250	412.3	-533.4	226.9	-632.3
18	43.000	73.0	-62.3	46.1	-65.4	39.500	278.9	-500.5	68.8	-592.6
19	43.000	73.0	-50.8	46.1	-51.7	39.750	153.8	-461.4	-79.3	-546.5
20	41.875	130.2	-121.8	104.3	-128.2	39.750	153.8	-472.5	-79.3	-552.5
21	41.125	221.5	-209.8	200.4	-219.8	39.875	94.8	-452.8	-148.4	-526.6
22	41.375	169.1	-137.5	145.5	-140.7	39.625	208.0	-538.7	-16.7	-610.2
23	40.625	272.2	-263.5	251.0	-271.2	39.875	73.3	-476.8	-169.3	-542.8
24	40.000	436.9	-475.7	420.5	-496.3	40.000	13.7	-432.2	-237.1	-496.3
25	40.500	199.1	0.0	172.3	0.0	40.125	-40.3	0.0	-299.2	0.0

Table 1.1: Comparison of optimal-replication strategy and Black-Scholes delta-hedging strategy for replicating an at-the-money put option on 1,000 shares of a \$40-stock over 25 trading periods for two simulated sample paths of a geometric Brownian motion with parameters $\mu = 0.07$ and $\sigma = 0.13$.

For sample path A, the differences between the optimal-replication strategy and the Black-Scholes are not great— V_t^* and θ_t^* are fairly close to their Black-Scholes counterparts. However, for sample path B, where there are two large price movements, the differences between the two replication strategies and the replication errors are substantial. Even in

such an idealized setting, the optimal-replication strategy can still play an important role in the dynamic hedging of risks.

In Section 1.3 we introduce the optimal-replication problem and propose a solution based on stochastic dynamic programming. The scope of the ϵ -arbitrage approach is illustrated in Sections 1.4 and 1.5 analytically and numerically for several examples including path-dependent options and options on assets with mixed jump-diffusion and stochastic-volatility price dynamics. The sensitivity of the replication error to price dynamics is studied in Section 1.6, and we conclude in Section 1.7.

1.3 ϵ -Arbitrage Strategies

In this section, we formulate and propose a solution approach for the problem of replicating a derivative security in incomplete markets. In Section 1.3.1 we introduce the optimal-replication problem and the principle of ϵ -arbitrage, and provide examples in Section 1.3.2 of the types of incompleteness that our framework can accommodate. In Sections 1.3.3 and 1.3.4 we propose stochastic dynamic programming algorithms in discrete and continuous time, respectively, that solve the optimal-replication problem.

1.3.1 The Optimal Replication Problem

Consider an asset with price P_t at time t where $0 \leq t \leq T$ and let $F(P_T, \mathbf{Z}_T)$ denote the payoff of some European derivative security at maturity date T which is a function of P_T and other variables \mathbf{Z}_T (see below). For expositional convenience, we shall refer to the asset as a *stock* and the derivative security as an *option* on that stock, but our results are considerably more general.

As suggested by Merton's (1973) derivation of the Black-Scholes formula, the optimal-

replication problem is to find a dynamic portfolio strategy—purchases and sales of stock and riskless borrowing and lending—on $[0, T]$ that is self-financing and comes as close as possible to the payoff $F(P_T, \mathbf{Z}_T)$ at T . To formulate the optimal-replication problem more precisely, we begin with the following assumptions:

(A1) *Markets are frictionless, i.e., there are no taxes, transactions costs, shortsales restrictions, and borrowing restrictions.*

(A2) *The riskless borrowing and lending rate is 0.*⁷

(A3) *There exists a finite-dimensional vector \mathbf{Z}_t of state variables whose components are not perfectly correlated with the prices of any traded securities, and $[P_t \ \mathbf{Z}_t]'$ is a vector Markov process*

(A4) *Trading takes place at known fixed times $t \in \mathcal{T}$. If $\mathcal{T} = \{t_0, t_1, \dots, t_N\}$, trading is said to be discrete. If $\mathcal{T} = [0, T]$, trading is said to be continuous.*

At time 0 consider forming a portfolio of stocks and riskless bonds at a cost V_0 and as time progresses, let θ_t , B_t , and V_t denote the number of shares of the stock held, the dollar value of bonds held, and the market value of the portfolio at time t , respectively, $t \in \mathcal{T}$, hence:

$$V_t = \theta_t P_t + B_t . \tag{1.1}$$

In addition, we impose the condition that after time 0, the portfolio is *self-financing*, i.e., all long positions in one asset are completely financed by short positions in the other asset

⁷This entails no loss of generality since we can always renormalize all prices by the price of a zero-coupon bond with maturity at time T (see, for example, Harrison and Kreps [1979]).

so that the portfolio experiences no cash inflows or outflows:

$$P_{t_{i+1}}(\theta_{t_{i+1}} - \theta_{t_i}) + B_{t_{i+1}} - B_{t_i} = 0 \quad , \quad 0 < t_i < t_{i+1} \leq T . \quad (1.2)$$

This implies that:

$$V_{t_{i+1}} - V_{t_i} = \theta_{t_i}(P_{t_{i+1}} - P_{t_i}) \quad (1.3)$$

and, in continuous time,

$$dV_t = \theta_t dP_t . \quad (1.4)$$

We seek a self-financing portfolio strategy $\{\theta_t\}$, $t \in \mathcal{T}$, such that the terminal value V_T of the portfolio is as close as possible to the option's payoff $F(P_T, \mathbf{Z}_T)$. Of course, there are many ways of measuring "closeness", each giving rise to a different optimal-replication problem. For reasons that will become clear shortly (see Section 1.3.5, we choose a mean-squared-error loss function, hence our version of the optimal-replication problem is:⁸

$$\min_{\{\theta_t\}} \mathbb{E} \left[V_T - F(P_T, \mathbf{Z}_T) \right]^2 \quad (1.5)$$

subject to self-financing condition (1.3) or (1.4), the dynamics of $[P_t \ \mathbf{Z}_t]'$, and the initial wealth V_0 , where the expectation \mathbb{E} is conditional on information at time 0.⁹

⁸Other recent examples of the use of mean-squared-error loss functions in related dynamic-trading problems include Duffie and Jackson (1990), Duffie and Richardson (1991), Schäl (1994), and Schweizer (1992, 1995).

⁹Note that we have placed no constraints on $\{\theta_t\}$, hence it is conceivable that for certain replication strategies, V_T is negative with positive probability. Imposing constraints on $\{\theta_t\}$ to ensure the non-negativity of V_T would render the optimal-replication problem (1.5) intractable. However, negative values for V_T is not nearly as problematic in the context of the optimal-replication problem as it is for the optimal consumption and portfolio problem of, for example, Merton (1971). In particular, V_T does not correspond

A natural measure of the success of the optimal-replication strategy is the square root of the mean-squared replication error (1.5) evaluated at the optimal $\{\theta_t\}$, hence we define

$$\epsilon(V_0) \equiv \sqrt{\min_{\{\theta_t\}} E \left\{ [V_T - F(P_T, \mathbf{Z}_T)]^2 \right\}}. \quad (1.6)$$

We shall show below that $\epsilon(V_0)$ can be minimized with respect to the initial wealth V_0 to yield the *least-cost* optimal-replication strategy and a corresponding measure of the minimum replication error ϵ^* :

$$\epsilon^* \equiv \min_{\{V_0\}} \epsilon(V_0). \quad (1.7)$$

In the case of Black and Scholes (1973) and Merton (1973), there exists optimal-replication strategies for which $\epsilon^* = 0$, hence we say that *perfect* arbitrage pricing holds.

But there are situations—dynamically incomplete markets, for example—where perfect arbitrage pricing does not hold. In particular, assumption (A3), the presence of state variables \mathbf{Z}_t that are not perfectly correlated with the prices of any traded securities, is the source of market incompleteness in our framework. While this captures only one potential source of incompleteness—and does so only in a “reduced-form” sense—nevertheless, it is a particularly relevant source of incompleteness in financial markets. Of course, we recognize that the precise nature of incompleteness, e.g., institutional rigidities, transactions costs, technological constraints, will affect the pricing and hedging of derivative securities in complex ways.¹⁰ Nevertheless, how well one security can be replicated by sophisticated

to an individual's wealth, but is the terminal value of a portfolio designed to replicate a particular payoff function. See Dybvig and Huang (1988) and Merton (1992, Chapter 6) for further discussion.

¹⁰For more “structural” models in which institutional sources of market incompleteness are studied, e.g., transactions costs, shortsales constraints, undiversifiable labor income, see Aiyagari (1994), Aiyagari and Gertler (1991), He and Modest (1995), Heaton and Lucas (1992, 1996), Lucas (1994), Scheinkman and Weiss (1986), Telmer (1993), and Weil (1992). See Magill and Quinzii (1996) for a comprehensive analysis

trading in other securities does provide one measure of the degree of market incompleteness even if it does not completely characterize it. In much the same way that the Black and Scholes (1973) and Merton (1973) models focus on the *relative* pricing of options—relative to the exogenously specified price dynamics for the underlying asset—we hope to capture the degree of *relative* incompleteness, relative to an exogenously specified set of Markov state variables that are not completely hedgeable.

In some of these cases, we shall show in Sections 1.3.3 and 1.3.4 that ϵ -arbitrage pricing is possible, i.e., it is possible to derive a mean-square-optimal replication strategy that is able to approximate the terminal payoff $F(P_T, \mathbf{Z}_T)$ of an option to within ϵ^* . But before turning to the solution of the optimal-replication problem, we provide several illustrative examples that delineate the scope of our framework.

1.3.2 Examples

Despite the restrictions imposed by assumptions (A1)–(A4), our framework can accommodate many kinds of market incompleteness and various types of derivative securities as the following examples illustrate:

(a) Stochastic Volatility. Consider a stock price process that follows a diffusion process with stochastic volatility, e.g., Hull and White (1987) and Wiggins (1987). The stock price and stock-price volatility are assumed to be governed by the following pair of stochastic differential equations:

$$dP_t = \mu P_t dt + \sigma_t P_t dW_{P_t}$$

$$d\sigma_t = g(\sigma_t) dt + \kappa \sigma_t dW_{\sigma_t}$$

of market incompleteness.

where W_{P_t} and W_{σ_t} are Brownian motions with mutual variation $dW_{P_t} dW_{\sigma_t} = \rho dt$.

This stochastic volatility model is included in our framework by defining $Z_t = \sigma_t$.

Then, clearly the vector process $[P_t \ Z_t]'$ is Markov.

(b) Options on the Maximum. In this and the next two examples we assume that $\mathcal{T} = \{t_0, t_1, \dots, t_N\}$ and that the stock price P_t process is Markov for expositional simplicity. The payoff of the option on the maximum stock price is given by

$$F\left(\max_{i=0, \dots, N} P_{t_i}\right). \quad (1.8)$$

Define the state variable

$$Z_{t_i} \equiv \max_{k=0, \dots, i} P_{t_k}.$$

The process $[P_{t_i} \ Z_{t_i}]'$ is Markov since the distribution of $P_{t_{i+1}}$ depends only on P_{t_i} and

$$Z_{t_{i+1}} = \max[Z_{t_i}, P_{t_{i+1}}] \quad , \quad Z_0 = P_0.$$

The payoff of the option can be expressed in terms of the terminal value of the state variables (P_T, Z_T) as $F(Z_T)$.

(c) Asian Options. The payoff of “Asian” or “average-rate” options is given by

$$F\left(\frac{1}{N+1} \sum_{i=0}^N P_{t_i}\right).$$

Let Z_{t_i} be the following state variable

$$Z_{t_i} \equiv \frac{1}{i+1} \sum_{k=0}^i P_{t_k}$$

and observe that the process $[P_{t_i} \ Z_{t_i}]'$ is Markov since the distribution of $P_{t_{i+1}}$ depends only on P_{t_i} and

$$Z_{t_{i+1}} = \frac{Z_{t_i}(i+1) + P_{t_{i+1}}}{(i+2)}, \quad Z_0 = P_0.$$

As before, the payoff of the option can be written as $F(Z_T)$.

(d) Knock-Out Options. Given a knock-out price \bar{P} , the payoff of a knock-out option is $\beta_T h(P_T)$, where $h(\cdot)$ is a function of the terminal stock price and

$$\beta_T = I_{[\max_{i=0, \dots, N} P_{t_i} \leq \bar{P}]}.$$

Define the state variable Z_t :

$$\begin{aligned} Z_0 &= I_{[P_0 \leq \bar{P}]}, \\ Z_{t_{i+1}} &= I_{[P_{t_{i+1}} \leq \bar{P}]} \cdot Z_{t_i}. \end{aligned}$$

It is easy to see that resulting process $[P_{t_i} \ Z_{t_i}]'$ is Markov, $Z_T = \beta_T$. The payoff of the option is given by $F(P_T, Z_T) = Z_T h(P_T)$.

1.3.3 ϵ -Arbitrage in Discrete Time

In this section, we propose a solution for the optimal-replication problem (1.5) in discrete time via stochastic dynamic programming. To simplify notation, we adopt the following

convention for discrete-time quantities: time subscripts t_i are replaced by i , e.g., the stock price P_{t_i} will be denoted as P_i and so on. Under this convention, we can define the usual *cost-to-go* or *value* function J_i as:

$$J_i(V_i, P_i, \mathbf{Z}_i) \equiv \min_{\theta^{(k, V_k, P_k, \mathbf{Z}_k)}, i \leq k \leq N-1} \mathbb{E} \left[[V_N - F(P_N, \mathbf{Z}_N)]^2 \mid V_i, P_i, \mathbf{Z}_i \right] \quad (1.9)$$

where V_i , P_i , and \mathbf{Z}_i comprise the state variables, θ_i is the control variable, and the self-financing condition (1.3) and the Markov property (A3) comprise the law of motion for the state variables. By applying Bellman's principle of optimality recursively (see, for example, Bertsekas [1995]):

$$J_N(V_N, P_N, \mathbf{Z}_N) = [V_N - F(P_N, \mathbf{Z}_N)]^2 \quad (1.10)$$

$$J_i(V_i, P_i, \mathbf{Z}_i) = \min_{\theta^{(i, V_i, P_i, \mathbf{Z}_i)}} \mathbb{E} \left[J_{i+1}(V_{i+1}, P_{i+1}, \mathbf{Z}_{i+1}) \mid V_i, P_i, \mathbf{Z}_i \right] \\ i = 0, \dots, N-1 \quad (1.11)$$

the optimal-replication strategy $\theta^*(i, V_i, P_i, \mathbf{Z}_i)$ can be characterized and computed.¹¹ In particular, we have:¹²

Theorem 1.1 *Under Assumptions (A1)–(A4) and (1.3), the solution of the optimal-replication problem (1.5) for $\mathcal{T} = \{t_0, t_1, \dots, t_N\}$ is characterized by the following:*

(a) *The value function $J_i(V_i, P_i, \mathbf{Z}_i)$ is quadratic in V_i , i.e., there are functions $a_i(P_i, \mathbf{Z}_i)$,*

¹¹Schweizer (1995) provides sufficient conditions for the existence of the optimal hedging strategy.

¹²Proofs are relegated to the Appendix.

$b_i(P_i, \mathbf{Z}_i)$, and $c_i(P_i, \mathbf{Z}_i)$ such that

$$J_i(V_i, P_i, \mathbf{Z}_i) = a_i(P_i, \mathbf{Z}_i) \cdot [V_i - b_i(P_i, \mathbf{Z}_i)]^2 + c_i(P_i, \mathbf{Z}_i), \quad i = 0, \dots, N. \quad (1.12)$$

(b) The optimal control $\theta^*(i, V_i, P_i, \mathbf{Z}_i)$ is linear in V_i , i.e.,

$$\begin{aligned} \theta^*(i, V_i, P_i, \mathbf{Z}_i) &= p_i(P_i, \mathbf{Z}_i) - V_i q_i(P_i, \mathbf{Z}_i) \\ &= (p_i - q_i b_i) - q_i (V_i - b_i) \end{aligned} \quad (1.13)$$

(c) The functions $a_i(\cdot)$, $b_i(\cdot)$, $c_i(\cdot)$, $p_i(\cdot)$, and $q_i(\cdot)$, are defined recursively as

$$a_N(P_N, \mathbf{Z}_N) = 1 \quad (1.14)$$

$$b_N(P_N, \mathbf{Z}_N) = F(P_N, \mathbf{Z}_N) \quad (1.15)$$

$$c_i(P_N, \mathbf{Z}_N) = 0 \quad (1.16)$$

and for $i = N-1, \dots, 0$

$$p_i(P_i, \mathbf{Z}_i) = \frac{\mathbb{E}[a_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) \cdot b_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) \cdot (P_{i+1} - P_i) | P_i, \mathbf{Z}_i]}{\mathbb{E}[a_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) \cdot (P_{i+1} - P_i)^2 | P_i, \mathbf{Z}_i]} \quad (1.17)$$

$$q_i(P_i, \mathbf{Z}_i) = \frac{\mathbb{E}[a_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) \cdot (P_{i+1} - P_i) | P_i, \mathbf{Z}_i]}{\mathbb{E}[a_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) \cdot (P_{i+1} - P_i)^2 | P_i, \mathbf{Z}_i]} \quad (1.18)$$

$$a_i(P_i, \mathbf{Z}_i) = \mathbb{E}[a_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) \cdot (1 - q_i(P_i, \mathbf{Z}_i)(P_{i+1} - P_i))^2 | P_i, \mathbf{Z}_i] \quad (1.19)$$

$$b_i(P_i, \mathbf{Z}_i) = \frac{1}{a_i(P_i, \mathbf{Z}_i)} \mathbb{E}[a_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) \cdot (b_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) - p_i(P_i, \mathbf{Z}_i)(P_{i+1} - P_i)) \cdot (1 - q_i(P_i, \mathbf{Z}_i)(P_{i+1} - P_i)) | P_i, \mathbf{Z}_i] \quad (1.20)$$

$$c_i(P_i, \mathbf{Z}_i) = \mathbb{E}[c_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) | P_i, \mathbf{Z}_i] + \mathbb{E}[a_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) \cdot (b_{i+1}(P_{i+1}, \mathbf{Z}_{i+1}) - p_i(P_i, \mathbf{Z}_i)(P_{i+1} - P_i))^2 | P_i, \mathbf{Z}_i] - a_i(P_i, \mathbf{Z}_i) \cdot b_i(P_i, \mathbf{Z}_i)^2 \quad (1.21)$$

(d) Under the optimal-replication strategy θ^* , the minimum replication error as a function of the initial wealth V_0 is

$$J_0(V_0, P_0, \mathbf{Z}_0) = a_0(P_0, \mathbf{Z}_0) \cdot [V_0 - b_0(P_0, \mathbf{Z}_0)]^2 + c_0(P_0, \mathbf{Z}_0). \quad (1.22)$$

hence the initial wealth that minimizes the replication error is $V_0^* = b_0(P_0, \mathbf{Z}_0)$, the least-cost optimal-replication strategy is the $\{\theta^*(i, V_i, P_i, \mathbf{Z}_i)\}$ that corresponds to this initial wealth, and the minimum replication error over all V_0 is:¹³

$$\epsilon^* = \sqrt{c_0(P_0, \mathbf{Z}_0)}. \quad (1.23)$$

¹³It is simple to show by induction that $a_i(P_i, \mathbf{Z}_i) > 0$ and $c_i(P_i, \mathbf{Z}_i) \geq 0$.

Theorem 1.1 shows that the optimal-replication strategy θ^* has a particularly simple structure (1.13) which can be related to the well-known “delta-hedging” strategy of the continuous-time Black-Scholes/Merton option-pricing model. In particular, if prices P_t follow a geometric Brownian motion, it can be shown that the first term of the right side of (1.13) corresponds to the derivative security’s “delta” (the partial derivative of the security’s price with respect to P_t), and the second term vanishes in the continuous-time limit (see Section 1.4.2 for further discussion). This underscores the fact that standard delta-hedging strategies, which are derived from continuous-time models, need not be optimal when applied in discrete time, and the extent to which the continuous-time and discrete-time replication strategies differ is captured by the second term of (1.13).

The fact that both the optimal control (1.13) and the value function (1.12) are defined recursively is especially convenient for computational purposes. Finally, because the value function is quadratic in V_t , it possesses a global minimum as a function of the initial wealth V_0 , and this global minimum and the initial wealth that attains it can be computed easily.

1.3.4 ϵ -Arbitrage in Continuous Time

For the continuous-time case $\mathcal{T} = [0, T]$, let $[P_t \ \mathbf{Z}_t]'$ follow a vector Markov diffusion process

$$dP_t = \mu_0(t, P_t, \mathbf{Z}_t)P_t dt + \sigma_0(t, P_t, \mathbf{Z}_t)P_t dW_{0t} \quad (1.24)$$

$$dZ_{jt} = \mu_j(t, P_t, \mathbf{Z}_t)Z_{jt} dt + \sigma_j(t, P_t, \mathbf{Z}_t)Z_{jt}dW_{jt}, \quad j = 1, \dots, J \quad (1.25)$$

where W_{jt} , $j = 0, \dots, J$ are Wiener processes with mutual variation

$$dW_{jt} dW_{kt} = \rho_{jk}(t, P_t, \mathbf{Z}_t) dt .$$

The continuous-time counterpart of the Bellman recursion is the Hamilton-Jacobi-Bellman equation (see, for example, Fleming and Rishel [1975]), and this yields the following:

Theorem 1.2 *Under Assumptions (A1)-(A4) and (1.4), the solution of the optimal-replication problem (1.5) for $\mathcal{T} = [0, T]$ is characterized by the following:*

(a) *The value function $J(t, V_t, P_t, \mathbf{Z}_t)$ is quadratic in V_t , i.e., there are functions $a(t, P_t, \mathbf{Z}_t)$, $b(t, P_t, \mathbf{Z}_t)$, and $c(t, P_t, \mathbf{Z}_t)$ such that*

$$J(t, V_t, P_t, \mathbf{Z}_t) = a(t, P_t, \mathbf{Z}_t) \cdot [V_t - b(t, P_t, \mathbf{Z}_t)]^2 + c(t, P_t, \mathbf{Z}_t) , \quad 0 \leq t \leq T .$$

(1.26)

(b) *For $t \in [0, T]$ the functions $a(t, P_t, \mathbf{Z}_t)$, $b(t, P_t, \mathbf{Z}_t)$, and $c(t, P_t, \mathbf{Z}_t)$ satisfy the following*

system of partial differential equations:¹⁴

$$\begin{aligned} \frac{\partial a}{\partial t} + \sum_{j=0}^J \mu_j Z_j \frac{\partial a}{\partial Z_j} + \frac{1}{2} \sum_{i,j=0}^J \sigma_i \sigma_j Z_i Z_j \rho_{ij} \frac{\partial^2 a}{\partial Z_i \partial Z_j} = \\ \left(\frac{\mu_0}{\sigma_0} \right)^2 a + 2 \sum_{j=0}^J \frac{\sigma_j}{\sigma_0} \mu_0 \rho_{0j} Z_j \frac{\partial a}{\partial Z_j} + \\ \frac{1}{a} \sum_{i,j=0}^J \sigma_i \sigma_j Z_i Z_j \rho_{0i} \rho_{0j} \frac{\partial a}{\partial Z_i} \frac{\partial a}{\partial Z_j} \end{aligned} \quad (1.27)$$

$$\begin{aligned} \frac{\partial b}{\partial t} + \sum_{j=0}^J \mu_j Z_j \frac{\partial b}{\partial Z_j} + \frac{1}{2} \sum_{i,j=0}^J \sigma_i \sigma_j Z_i Z_j \rho_{ij} \frac{\partial^2 b}{\partial Z_i \partial Z_j} = \\ \sum_{j=0}^J \frac{\partial b}{\partial Z_j} Z_j \left(\frac{\sigma_j}{\sigma_0} \mu_0 \rho_{0j} - \frac{1}{2} \sum_{i=0}^J \frac{\sigma_i \sigma_j Z_i}{a} (\rho_{0i} \rho_{0j} - \rho_{ij}) \frac{\partial a}{\partial Z_i} \right) \end{aligned} \quad (1.28)$$

$$\begin{aligned} \frac{\partial c}{\partial t} + \sum_{j=0}^J \mu_j Z_j \frac{\partial c}{\partial Z_j} + \frac{1}{2} \sum_{i,j=0}^J \sigma_i \sigma_j Z_i Z_j \rho_{ij} \frac{\partial^2 c}{\partial Z_i \partial Z_j} = \\ a \sum_{i,j=0}^J \sigma_i \sigma_j Z_i Z_j \frac{\partial b}{\partial Z_i} \frac{\partial b}{\partial Z_j} (\rho_{0i} \rho_{0j} - \rho_{ij}) . \end{aligned} \quad (1.29)$$

with boundary conditions:

$$a(T, P_T, \mathbf{Z}_T) = 1 \quad , \quad b(T, P_T, \mathbf{Z}_T) = F(P_T, \mathbf{Z}_T) \quad , \quad c(T, P_T, \mathbf{Z}_T) = 0 \quad (1.30)$$

where Z_i denotes the i -th component of \mathbf{Z}_t and $Z_0 \equiv P_t$.

(c) The optimal control $\theta^*(t, V_t, P_t, \mathbf{Z}_t)$ is linear in V_t and is given by:

$$\theta^*(t, V_t, P_t, \mathbf{Z}_t) = \sum_{j=0}^J \rho_{0j} \frac{\sigma_j Z_j}{\sigma_0 Z_0} \frac{\partial b}{\partial Z_j} - (V_t - b) \left(\sum_{j=0}^J \rho_{0j} \frac{\sigma_j Z_j}{a \sigma_0 Z_0} \frac{\partial a}{\partial Z_j} - \frac{\mu_0}{\sigma_0^2 Z_0} \right) . \quad (1.31)$$

¹⁴We omit the arguments of $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ in (1.27)–(1.29) to economize on notation.

(d) *Under the optimal-replication strategy θ^* , the minimum replication error as a function of the initial wealth V_0 is*

$$J(0, V_0, P_0, \mathbf{Z}_0) = a(0, P_0, \mathbf{Z}_0)[V_0 - b(0, P_0, \mathbf{Z}_0)]^2 + c(0, P_0, \mathbf{Z}_0) \quad (1.32)$$

hence the initial wealth that minimizes the replication error is $V_0^ = b(0, P_0, \mathbf{Z}_0)$, the least-cost optimal-replication strategy is the $\{\theta^*(t, V_t, P_t, \mathbf{Z}_t)\}$ that corresponds to this initial wealth, and the minimum replication error over all V_0 is:¹⁵*

$$\epsilon^* = \sqrt{c(0, P_0, \mathbf{Z}_0)} . \quad (1.33)$$

1.3.5 Interpreting ϵ^* and V_0^*

Theorems 1.1 and 1.2 show that the optimal-replication problem (1.5) can be solved for a mean-squared-error measure of replication error under Markov state dynamics. In particular, the optimal-replication strategy $\theta^*(\cdot)$ is a dynamic trading strategy that yields the minimum mean-squared replication error $\epsilon(V_0)$ for an initial wealth V_0 . The fact that $\epsilon(V_0)$ depends on V_0 should come as no surprise, and the fact that $\epsilon(V_0)$ is quadratic in V_0 underscores the possibility that delta-hedging strategies can be under- or over-capitalized, i.e., there exists a unique V_0^* that minimizes the mean-squared replication error. One attractive feature of our approach is the ability to quantify the impact of capitalization V_0 on the replication error $\epsilon(V_0)$.

¹⁵It can be shown that $a(t, P_t, \mathbf{Z}_t) > 0$ and $c(t, P_t, \mathbf{Z}_t) \geq 0$.

V_0^* Is Not a Price

In this sense, V_0^* may be viewed as the minimum production cost of replicating the payoff $F(P_T, \mathbf{Z}_T)$ as closely as possible, to within ϵ^* . However, because we have assumed that markets are dynamically incomplete—otherwise ϵ^* is 0 and perfect replication is possible— V_0^* cannot be interpreted as the price of a derivative security with payoff $F(P_T, \mathbf{Z}_T)$ unless additional economic structure is imposed. In particular, in dynamically incomplete markets derivatives cannot be priced by arbitrage considerations alone—we must resort to an equilibrium model in which the prices of all traded assets are determined by supply and demand.

To see why V_0^* cannot be interpreted as a price, observe that two investors with different risk preferences may value $F(P_T, \mathbf{Z}_T)$ quite differently, and will therefore place different valuations on the replication error ϵ^* . While both investors may agree that V_0^* is the minimum cost for the optimal-replication strategy $\theta^*(\cdot)$, they may differ in their willingness to pay such a cost for achieving the replication error ϵ^* .¹⁶ Moreover, some investors' preferences may not be consistent with a symmetric loss function, e.g., they may value negative replication errors quite differently than positive replication errors.

More to the point, an asset's price is the outcome of a market equilibrium in which investors' preferences, budget dynamics, and information structure interact through the imposition of market-clearing conditions, i.e., supply equals demand. In contrast, V_0^* is the solution to a simple dynamic optimization problem that does not typically incorporate any notion of economic equilibrium.

¹⁶See Duffie and Jackson (1990) and Duffie and Richardson (1991) for examples of replication strategies under specific preference assumptions.

Why Mean-Squared Error?

In fact, there are many possible loss functions, each giving rise to a different set of optimal-replication strategies, hence a natural question to ask in interpreting Theorems 1.1 and 1.2 is why use mean-squared error?

An obvious motivation is, of course, tractability. We showed in Sections 1.3.3 and 1.3.4 that the optimal-replication problem can be solved via stochastic dynamic programming for a mean-squared-error loss function and Markov state dynamics, and that the solution can be implemented as an exact and efficient recursive algorithm. In Sections 1.4 and 1.5, we apply this algorithm to a variety of derivative securities in incomplete markets and demonstrate its practical relevance analytically and numerically.

Another motivation is that a symmetric loss function is the most natural choice when we have no prior information about whether the derivative to be replicated is being purchased or sold. Indeed, when a derivatives broker is asked by a client to provide a price quote, the client typically does not reveal whether he is a buyer or seller until after receiving both bid and offer prices. In such cases, asymmetric loss functions are inappropriate since positive replication errors for a long position become negative replication errors for the short position.

Of course, in more structured applications such as Duffie and Jackson (1990) in which investors' preferences, budget dynamics, and information sets are specified, it is not apparent that mean-squared-error optimal-replication strategies are optimal from a particular investor's point of view. However, even in these cases, a slight modification of the mean-squared-error loss function yields optimal-replication strategies that have natural economic interpretations. In particular, by defining mean-squared error with respect to an equivalent martingale measure, the minimum production cost V_0^* associated with this loss function can

be interpreted as an equilibrium market price which, by definition, incorporates all aspects of the economic environment in which the derivative security is traded.

The difficulty with such an interpretation is the multiplicity of equivalent martingale measures in incomplete markets—it is only when markets are dynamically complete that the equivalent martingale measure is unique (see Harrison and Kreps [1979]). It may be possible to compute upper and lower bounds for ϵ^* over the entire set of equivalent martingale measures, but without additional structure these bounds are likely to be extremely wide and of little practical relevance.

Nevertheless, ϵ^* is a useful metric for the degree of market incompleteness, providing an objective measure of the difficulty in replicating a derivative security. For example, we shall see in Section 1.6 that although stochastic volatility and mixed jump-diffusion processes both imply market incompleteness, our ϵ -arbitrage strategy shows that for certain parameter values, the former is a more difficult type of incompleteness to hedge than the latter.

1.4 Illustrative Examples

To illustrate the scope of the ϵ -arbitrage approach to the optimal-replication problem, we apply the results of Section 1.3 to four specific cases for the return-generating process: state-independent returns (Section 1.4.1), geometric Brownian motion (Section 1.4.2), a jump-diffusion model (Section 1.4.3), and a stochastic volatility model (1.4.4).

1.4.1 State-Independent Returns

Suppose that stock returns are state-independent so that

$$P_i = P_{i-1}(1 + \phi_{i-1}) \quad (1.34)$$

where ϕ_{i-1} is independent of the current stock price and all other state variables. This, together with the Markov assumption (A3), implies that returns are statistically independent (but not necessarily identically distributed) through time. Also, let the payoff of the derivative security $F(P_T)$ depend only on the price of the risky asset at time T .

In this case, there is no need for additional state variables \mathbf{Z}_i and the expressions in Theorem 1.1 simplify to:

$$a_N = 1 \quad , \quad b_N(P_N) = F(P_N) \quad , \quad c_N(P_N) = 0 \quad (1.35)$$

and for $i = N-1, \dots, 0$,

$$a_i = a_{i+1} \frac{\sigma_i^2}{\sigma_i^2 + \mu_i^2} \quad (1.36)$$

$$b_i(P_i) = E[b_{i+1}(P_i(1 + \phi_i))|P_i] - \frac{\mu_i}{\sigma_i^2} \text{Cov}[\phi_i, b_{i+1}(P_i(1 + \phi_i))|P_i] \quad (1.37)$$

$$c_i(P_i) = E[c_{i+1}(P_i(1 + \phi_i))|P_i] + \frac{a_{i+1}}{\sigma_i^2} \left\{ \sigma_i^2 \text{Var}[b_{i+1}(P_i(1 + \phi_i))|P_i] - \text{Cov}[\phi_i, b_{i+1}(P_i(1 + \phi_i))|P_i]^2 \right\} \quad (1.38)$$

$$p_i(P_i) = \frac{E[\phi_i b_{i+1}(P_i(1 + \phi_i))|P_i]}{(\sigma_i^2 + \mu_i^2)P_i} \quad (1.39)$$

$$q_i(P_i) = \frac{\mu_i}{(\sigma_i^2 + \mu_i^2)P_i} \quad (1.40)$$

where $\mu_i = E[\phi_i]$ and $\sigma_i^2 = \text{Var}[\phi_i]$.

1.4.2 Geometric Brownian Motion

Let the stock price process follow the geometric Brownian motion of Black and Scholes (1973) and Merton (1973). We show that the ϵ -arbitrage approach yields the Black-Scholes/Merton results in the limit of continuous time, but in discrete time there are important differences between the optimal-replication strategy of Theorem 1.1 and the standard Black-Scholes/Merton delta-hedging strategy.

For notational convenience, let all discrete time intervals $[t_i, t_{i+1})$ be of equal length $t_{i+1} - t_i = \Delta t$. The assumption of geometric Brownian motion then implies:

$$P_{i+1} = P_i \cdot (1 + \phi_i) \tag{1.41}$$

$$\log(1 + \phi_i) = \left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t}z_i \tag{1.42}$$

$$z_i \sim \mathcal{N}(0, 1) . \tag{1.43}$$

Recall that for $\Delta t \ll 1$ (a large number of time increments in $[0, T]$), the following approximation holds (see, for example, Merton [1992, Chapter 3]):

$$\phi_i \sim \mathcal{N}(\mu\Delta t, \sigma^2\Delta t) + O(\Delta t^{3/2})$$

This, and Taylor's theorem, imply the following approximations for the recursive relations

(1.36)–(1.38) of Section 1.4.1:

$$\begin{aligned} \text{Var}[b_{i+1}(P_i(1 + \phi_i))|P_i] &= b'_{i+1}(P_i)\sigma^2 P_i^2 \Delta t + O(\Delta t^2) \\ \text{Cov}[\phi_i, b_{i+1}(P_i(1 + \phi_i))|P_i] &= b'_{i+1}(P_i)\sigma^2 P_i \Delta t + O(\Delta t^2) \\ \text{E}[b_{i+1}(P_i(1 + \phi_i))|P_i] &= b_{i+1}(P_i) + b'_{i+1}(P_{i-1})\mu P_i \Delta t + \\ &\quad b''_{i+1}(P_i)\frac{\sigma^2}{2}P_i^2 \Delta t + O(\Delta t^2) \\ \text{E}[c_{i+1}(P_i(1 + \phi_i))|P_i] &= c_i(P_i) + c'_{i+1}(P_{i-1})\mu P_i \Delta t + \\ &\quad c''_{i+1}(P_i)\frac{\sigma^2}{2}P_i^2 \Delta t + O(\Delta t^2) . \end{aligned}$$

We can then rewrite (1.37)–(1.38) as

$$\begin{aligned} b_i(P_i) &= b_{i+1}(P_i) + b''_{i+1}(P_i)\frac{\sigma^2}{2}P_i^2 \Delta t + O(\Delta t^2) \\ c_i(P_i) &= c_{i+1}(P_i) + c'_{i+1}(P_{i-1})\mu P_i \Delta t + c''_{i+1}(P_i)\frac{\sigma^2}{2}P_i^2 \Delta t + O(\Delta t^2) \end{aligned}$$

and conclude that the system (1.37)–(1.38) approximates the following system of PDE's

$$\frac{\partial b(t, P)}{\partial t} = - \frac{\sigma^2 P^2}{2} \frac{\partial^2 b(t, P)}{\partial P^2} \quad (1.44)$$

$$\frac{\partial c(t, P)}{\partial t} = - \mu P \frac{\partial c(t, P)}{\partial P} - \frac{\sigma^2 P^2}{2} \frac{\partial^2 c(t, P)}{\partial P^2} \quad (1.45)$$

up to $O(\Delta t)$ terms. But (1.44) is the Black and Scholes (1973) PDE, hence we see that in the limit of continuous trading, i.e., as $N \rightarrow \infty$ and $\Delta t \rightarrow 0$ for a fixed $T \equiv N\Delta t$, the discrete-time

optimal-replication strategy of Theorem 1.1 characterizes the Black and Scholes (1973) and Merton (1973) models.

Moreover, the expression for $c(t, P)$, (1.45), is homogeneous, hence $c(t, P) \equiv 0$ due to the boundary condition $c(T, \cdot) = 0$. This is consistent with the fact that in the Black-Scholes case it is possible to replicate the option exactly, so that the replication error vanishes in the continuous-time limit.

In particular, it can be shown that the components of the discrete-time optimal-replication strategy (1.13) converge to the following continuous-time counterparts:

$$p_i - q_i b_i \rightarrow \frac{\partial b}{\partial P} \quad , \quad q_i \rightarrow \frac{\mu}{\sigma^2 P}$$

hence the continuous-time limit of the optimal-replication strategy $\theta^*(\cdot)$ is given by:

$$\theta^*(t, V_t, P_t) = \frac{\partial b(t, P_t)}{\partial P_t} - \frac{\mu}{\sigma^2 P_t} [V_t - b(t, P_t)] \quad . \quad (1.46)$$

Now at time $t = 0$, and for the minimum production-cost initial wealth V_0^* , this expression reduces to

$$\theta^*(0, V_0^*, P_0) = \frac{\partial b(0, P_0)}{\partial P_0}$$

since $V_0^* = b(0, P_0)$. But exact replication is possible in this case, hence the value of the replicating portfolio is always equal to $b(t, P_t)$ for every realization of the stock price process, i.e.,

$$V_t = b(t, P_t)$$

for all $t \in [0, T]$, which implies that

$$\theta^*(t, V_t, P_t) = \frac{\partial b(t, P_t)}{\partial P_t}. \quad (1.47)$$

As expected, for every realization of the stock price process the optimal-replication strategy coincides with the delta-hedging strategy given by the Black-Scholes hedge ratio. However, note that the functional form of (1.46) is different from the Black-Scholes hedging formula—the optimal-replication strategy depends explicitly on the value of the replicating portfolio V_t .

1.4.3 Jump-Diffusion Models

In this section, we apply results of Section 1.3 to the replication and pricing of options on a stock with mixed jump-diffusion price dynamics. As before, we assume that all time intervals $t_{i+1} - t_i = \Delta t$ are regularly spaced. Following Merton (1976), we assume the following model for the stock price process:

$$P_{i+1} = P_i(1 + \phi_i) \quad (1.48)$$

$$\log(1 + \phi_i) = \left(\mu - \lambda k - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t}z_i + \sum_{j=0}^{n_i} \log Y_j \quad (1.49)$$

$$z_i \sim \mathcal{N}(0, 1) \quad (1.50)$$

$$k = \mathbb{E}[Y_j - 1] \quad (1.51)$$

$$\text{Prob}(n_i = m) = e^{-\lambda\Delta t} \frac{(\lambda\Delta t)^m}{m!} \quad (1.52)$$

where the jump magnitudes $\{Y_j\}$ are independently and identically distributed random variables and jump arrivals follow a Poisson process with constant arrival rate λ .

We consider two types of jumps: jumps of deterministic magnitude, and jumps with

lognormally distributed jump magnitudes. In the first case:

$$Y_j = 1 + \delta . \tag{1.53}$$

If we set $\sigma = 0$ in (1.48), this model corresponds to the continuous-time jump process considered by Cox and Ross (1976). In the second case:

$$\log Y_i \sim \mathcal{N}(0, \delta^2) . \tag{1.54}$$

There are two methods of calculating the optimal-replication strategy for the mixed jump-diffusion model. One method is to begin with the solutions of the dynamic programming problem given in Sections 1.3.3 and 1.3.4, derive a limiting system of partial differential equations as in Section 1.4.2, and solve this system numerically, using one of the standard finite difference schemes.

The second method is to implement the solution of the dynamic programming problem directly, without the intermediate step of reducing it to a system of PDE's.

The advantage of the second method is that it treats a variety of problems in a uniform fashion, the only problem-dependent part of the approach being the specification of the stochastic process. On the other hand, the first approach yields a representation of the solution as a system of PDE's, which can often provide some information about the qualitative properties of the solution even before a numerical solution is obtained.

With these considerations in mind, we shall derive a limiting system of PDE's for the deterministic-jump-magnitude specification (1.53) and use it to find conditions on the parameters of the stochastic process which allow exact replication of the option's payoff, or, equivalently, arbitrage pricing. For the lognormal-jump-magnitude specification (1.54), we

shall obtain numerical solutions directly from the dynamic programming algorithm of Theorem 1.1.

The Continuous-Time Limit

To derive the continuous-time limit of (1.36)–(1.38) we follow the same procedure as in Section 1.4.2 which yields the following system of PDE's:

$$\begin{aligned} \frac{\partial b(t, P)}{\partial t} = & -\lambda [b(t, P(1 + \delta)) - b(t, P)] + \lambda \delta P \frac{\partial b(t, P)}{\partial P} - \frac{\sigma^2 P^2}{2} \frac{\partial^2 b(t, P)}{\partial P^2} - \\ & \frac{\mu \lambda \delta}{\lambda \delta^2 + \sigma^2} \left[\delta P \frac{\partial b(t, P)}{\partial P} - [b(t, P(1 + \delta)) - b(t, P)] \right] \end{aligned} \quad (1.55)$$

$$\begin{aligned} \frac{\partial c(t, P)}{\partial t} = & -\lambda [c(t, P(1 + \delta)) - c(t, P)] - (\mu - \lambda \delta) P \frac{\partial c(t, P)}{\partial P} - \frac{\sigma^2 P^2}{2} \frac{\partial^2 c(t, P)}{\partial P^2} - \\ & a(t) \frac{\lambda \sigma^2}{\lambda \delta^2 + \sigma^2} \left[\delta P \frac{\partial b(t, P)}{\partial P} - [b(t, P(1 + \delta)) - b(t, P)] \right]^2 \end{aligned} \quad (1.56)$$

$$\frac{da(t)}{dt} = \frac{\mu^2}{\lambda \delta^2 + \sigma^2} a(t) \quad (1.57)$$

with boundary conditions:

$$a(T) = 1 \quad (1.58)$$

$$c(T, P) = 0 \quad (1.59)$$

$$b(T, P) = F(P) . \quad (1.60)$$

We can use the boundary conditions to solve (1.57):

$$a(t) = \exp \left[\frac{\mu^2}{\lambda \delta^2 + \sigma^2} (t - T) \right] . \quad (1.61)$$

The optimal-replication strategy is given by:

$$\theta^*(t, V_t, P_t) = \frac{\partial b(t, P_t)}{\partial P_t} - \frac{\mu}{(\lambda\delta^2 + \sigma^2)P_t} [V_t - b(t, P_t)] - \frac{\lambda\delta^2}{\lambda\delta^2 + \sigma^2} \frac{\partial b(t, P_t)}{\partial P_t} + \frac{\lambda\delta}{(\lambda\delta^2 + \sigma^2)P_t} [b(t, P_t(1 + \delta)) - b(t, P_t)] . \quad (1.62)$$

For exact replication to be possible, $c(t, P) \equiv 0$ must be a solution of (1.56). This implies that (1.56) is homogeneous, i.e.,

$$\frac{\lambda\sigma^2}{\lambda\delta^2 + \sigma^2} \left\{ \delta P \frac{\partial b(t, P)}{\partial P} - [b(t, P(1 + \delta)) - b(t, P)] \right\}^2 = 0 \quad (1.63)$$

for all $b(t, P)$ satisfying (1.55), which is equivalent to

$$\lambda\delta\sigma^2 = 0 . \quad (1.64)$$

Condition (1.64) is satisfied if at least one of the following is true:

- Jumps occur with zero probability.
- Jumps have zero magnitude.
- The diffusion coefficient is equal to zero, i.e., stock price follows a pure jump process.

But these are precisely the conditions for the arbitrage-pricing of options on mixed jump-diffusion assets, e.g., Merton (1976).

Perturbation Analysis with Small Jump Amplitudes

Consider the behavior of $b(t, P)$ and $c(t, P)$ when the jump magnitude is small, i.e., $\delta \ll 1$. In this case the market is “almost complete” and solution of the option replication problem is obtained as a perturbation of the complete-markets solution of Black and Scholes (1973)

and Merton (1973). In particular, we treat the amplitude of stock price jumps as a small parameter and look for a solution of (1.55)–(1.60) of the following form:

$$b(t, P) = b_0(t, P) + \delta b_1(t, P) + \delta^2 b_2(t, P) + \dots \quad (1.65)$$

$$c(t, P) = c_0(t, P) + \delta c_1(t, P) + \delta^2 c_2(t, P) + \delta^3 c_3(t, P) + \delta^4 c_4(t, P) + \dots \quad (1.66)$$

After substituting this expansion into (1.56)–(1.60), it is apparent that the functions $b_0(t, P)$, $b_2(t, P)$, and $c_4(t, P)$ must satisfy the following system of partial differential equations:

$$\frac{\partial b_0(t, P)}{\partial t} = -\frac{\sigma^2 P^2}{2} \frac{\partial^2 b_0(t, P)}{\partial P^2} \quad (1.67)$$

$$\frac{\partial b_2(t, P)}{\partial t} = -\frac{\lambda P^2}{2} \frac{\partial^2 b_0(t, P)}{\partial P^2} \quad (1.68)$$

$$\frac{\partial c_4(t, P)}{\partial t} = -\mu P \frac{\partial c_4(t, P)}{\partial P} - \frac{\sigma^2 P^2}{2} \frac{\partial^2 c_4(t, P)}{\partial P^2} - a(t) \frac{\lambda P^4}{4} \left(\frac{\partial^2 b_0(t, P)}{\partial P^2} \right)^2 \quad (1.69)$$

with boundary conditions:

$$b_0(T, P) = F(P) \quad (1.70)$$

$$b_2(T, P) = 0 \quad (1.71)$$

$$c_4(T, P) = 0 \quad (1.72)$$

and

$$b_1 = c_1 = c_2 = c_3 = 0.$$

System (1.67)–(1.72) can be solved to yield:

$$b(t, P) = b_0(t, P) + \frac{\lambda\delta^2}{\sigma^2}[b_0(t, P) - F(P)] + O(\delta^3) \quad (1.73)$$

where $b_0(t, P)$ is the option price in the absence of a jump component, i.e., the Black-Scholes formula in the case of put and call options. Observe that for an option with a convex payoff function $b_0(t, P) \geq F(P)$, which implies that $b(t, P) \geq b_0(t, P)$, i.e., the addition of a small jump component to geometric Brownian motion increases the price of the option. This qualitative behavior of the option price is consistent with the results in Merton (1976) which were obtained with equilibrium arguments.

The optimal-replication strategy (1.62) is given by:

$$\begin{aligned} \theta^*(t, V_t, P_t) = & \frac{\partial b_0(t, P_t)}{\partial P_t} + \frac{\mu}{\sigma^2 P_t} [b_0(t, P_t) - V_t] + \\ & \frac{\lambda\delta^2}{\sigma^2} \left[\frac{\partial b_0(t, P_t)}{\partial P_t} - \frac{\partial F(P_t)}{\partial P_t} + V_t - F(P_t) \right] + O(\delta^3). \end{aligned} \quad (1.74)$$

and the corresponding replication error is:

$$c(t, P) = \delta^4 c_4(t, P) + O(\delta^6) = O(\delta^4) \quad (1.75)$$

where $c_4(t, P)$ solves (1.69) and (1.72).

Equations (1.73) and (1.74) provide closed-form expressions for the replication cost and the optimal-replication strategy when the amplitude of jumps is small, i.e., when markets are almost complete, and (1.75) describes the dependence of the replication error on the jump magnitude.

1.4.4 Stochastic Volatility

Let stock prices follow a diffusion process with stochastic volatility as in Hull and White (1987) and Wiggins (1987):

$$dP_t = \mu P_t dt + \sigma_t P_t dW_{P_t} \quad (1.76)$$

$$d\sigma_t = g(\sigma_t) dt + \kappa \sigma_t dW_{\sigma_t} \quad (1.77)$$

where W_{P_t} and W_{σ_t} are Brownian motions with mutual variation $dW_{P_t} dW_{\sigma_t} = \rho dt$.

The Continuous-Time Solution

Although applying the results of Section 1.3 to (1.76)–(1.77) is conceptually straightforward, the algebraic manipulations are quite involved in this case. A simpler alternative to deriving a system of PDE's as the continuous-time limit of the solution in Theorem 1.1 is to formulate the problem in continuous time at the outset and solve it using continuous-time stochastic control methods. This approach simplifies the calculations considerably.

Specifically, the pair of stochastic processes (P_t, σ_t) satisfies assumptions of Section 1.3.4, therefore results of this section can be used to derive the optimal-replication strategy, the minimum production-cost of optimal replication, and the replication error. In particular, the application of the results of Section 1.3.4 to (1.76)–(1.77) yields the following system of

PDE's:

$$\begin{aligned} \frac{\partial a(t, \sigma)}{\partial t} &= \frac{\mu^2}{\sigma^2} a(t, \sigma) - (g(\sigma) + 2\rho\kappa\mu) \frac{\partial a(t, \sigma)}{\partial \sigma} + \\ &\quad \frac{1}{a(t, \sigma)} \left(\rho\kappa\sigma \frac{\partial a(t, \sigma)}{\partial \sigma} \right)^2 - \frac{1}{2} \kappa^2 \sigma^2 \frac{\partial^2 a(t, \sigma)}{\partial \sigma^2} \end{aligned} \quad (1.78)$$

$$\begin{aligned} \frac{\partial b(t, P, \sigma)}{\partial t} &= - (g(\sigma) - \rho\mu\kappa) \frac{\partial b(t, P, \sigma)}{\partial \sigma} - \frac{\kappa^2 \sigma^2}{2} \frac{\partial^2 b(t, P, \sigma)}{\partial \sigma^2} - \\ &\quad \frac{\sigma^2 P^2}{2} \frac{\partial^2 b(t, P, \sigma)}{\partial P^2} - \rho\kappa\sigma^2 P \frac{\partial^2 b(t, P, \sigma)}{\partial P \partial \sigma} - \\ &\quad (1 - \rho^2) \frac{\kappa^2 \sigma^2}{a(t, \sigma)} \frac{\partial a(t, \sigma)}{\partial \sigma} \frac{\partial b(t, P, \sigma)}{\partial \sigma} \end{aligned} \quad (1.79)$$

$$\begin{aligned} \frac{\partial c(t, P, \sigma)}{\partial t} &= - g(\sigma) \frac{\partial c(t, P, \sigma)}{\partial \sigma} - \mu P \frac{\partial c(t, P, \sigma)}{\partial P} - \frac{\kappa^2 \sigma^2}{2} \frac{\partial^2 c(t, P, \sigma)}{\partial \sigma^2} - \\ &\quad \rho\kappa\sigma^2 P \frac{\partial^2 c(t, P, \sigma)}{\partial P \partial \sigma} - \frac{\sigma^2 P^2}{2} \frac{\partial^2 c(t, P, \sigma)}{\partial P^2} + \\ &\quad a(t, \sigma) \kappa^2 \sigma^2 (\rho^2 - 1) \left(\frac{\partial b(t, P, \sigma)}{\partial \sigma} \right)^2 \end{aligned} \quad (1.80)$$

with boundary conditions:

$$a(T, \sigma) = 1 \quad , \quad b(T, P, \sigma) = F(P, \sigma) \quad , \quad c(T, P, \sigma_T) = 0 \quad .$$

The optimal-replication strategy is given by:

$$\begin{aligned} \theta^*(t, V_t, P_t, \sigma_t) &= \frac{\partial b(t, P_t, \sigma_t)}{\partial P_t} + \frac{\rho\kappa}{P_t} \frac{\partial b(t, P_t, \sigma_t)}{\partial \sigma_t} - \frac{V_t - b(t, P_t, \sigma_t)}{a(t, \sigma_t)} \times \\ &\quad \frac{\rho\kappa}{P_t} \frac{\partial a(t, \sigma_t)}{\partial \sigma_t} - \frac{\mu}{\sigma_t^2 P_t} [V_t - b(t, P_t, \sigma_t)] \quad . \end{aligned} \quad (1.81)$$

Exact replication is possible when the following equation is satisfied:

$$\kappa^2 (\rho^2 - 1) = 0 \quad .$$

and this corresponds to the following special cases:

- Volatility is a deterministic function of time.
- The Brownian motions driving stock prices and volatility are perfectly correlated.

Both of these conditions yield well-known special cases where arbitrage-pricing is possible (see, for example, Geske [1979] and Rubinstein [1983]). If we set $\kappa = g(\sigma) = 0$, (1.79) reduces to the Black and Scholes (1973) PDE.

1.5 Numerical Analysis

The essence of the ϵ -arbitrage approach to the optimal-replication problem is the recognition that although perfect replication may not be possible in some situations, the optimal-replication strategy of Theorem 1.1 may come very close. How close is, of course, an empirical matter hence in this section we present several numerical examples that complement the theoretical analysis of Section 1.4.

In Section, 1.5.1 we describe our numerical procedure and apply it to the case of geometric Brownian motion in Section 1.5.2, a mixed jump-diffusion model with a lognormal jump magnitude in Section 1.5.3, a stochastic volatility model in Section 1.5.4, and to a path-dependent option to “sell at the high” in Section 1.5.5.

1.5.1 The Numerical Procedure

To implement the solution (1.17)–(1.21) of the optimal-replication problem numerically, we begin by representing the functions $a_i(P, \mathbf{Z})$, $b_i(P, \mathbf{Z})$, and $c_i(P, \mathbf{Z})$ by their values over a spatial grid $\{(P^j, \mathbf{Z}^k) : j = 1, \dots, J, k = 1, \dots, K\}$. For any given (P, \mathbf{Z}) , values $a_i(P, \mathbf{Z})$, $b_i(P, \mathbf{Z})$, and $c_i(P, \mathbf{Z})$ are obtained from $a_i(P^j, \mathbf{Z}^k)$, $b_i(P^j, \mathbf{Z}^k)$, and $c_i(P^j, \mathbf{Z}^k)$

using a piece-wise quadratic interpolation. This procedure provides an accurate representation of $a_i(P, \mathbf{Z})$, $b_i(P, \mathbf{Z})$, and $c_i(P, \mathbf{Z})$ with a reasonably small number of sample points. The values $a_i(P^j, \mathbf{Z}^k)$, $b_i(P^j, \mathbf{Z}^k)$, and $c_i(P^j, \mathbf{Z}^k)$ are updated according to the recursive procedure (1.17)–(1.19).

We evaluate the expectations in (1.17)–(1.19) by replacing them with the corresponding integrals. For all the models considered in this paper, these integrals involve Gaussian kernels. We use Gauss-Hermite quadrature formulas (see, for example, Stroud [1971]) to obtain efficient numerical approximations of these integrals.

In all cases except for the path-dependent options, we perform numerical computations for a European put option with a unit strike price ($K = 1$), i.e., $F(P_T) = \max[0, K - P_T]$, and a six-month maturity. It is apparent from (1.17)–(1.21) that for a call option with the same strike price K , the replication error $c_i(\cdot)$ is the same as that of a put option, and the replication cost $b_i(\cdot)$ satisfies the put-call parity relation. We assume 25 trading periods, defined by $t_0 = 0, t_{i+1} - t_i = \Delta t = 1/50$.

1.5.2 Geometric Brownian Motion

Let stock prices follow a geometric Brownian motion, which implies that returns are log-normally distributed as in (1.41)–(1.43). We set $\mu = 0.07$ and $\sigma = 0.13$, and to cover a range of empirically plausible parameter values, we vary each parameter by increasing and decreasing them by 25% and 50% while holding the values of other parameter fixed. Figure 1-1 displays the minimum replication cost V_0^* minus the intrinsic value $F(P_0)$, for the above range of parameter values, as a function of the stock price at time 0.

Figure 1-1 shows that V_0^* is not sensitive to changes in μ and increases with σ . This is not surprising given that V_0^* approximates the Black-Scholes option pricing formula.

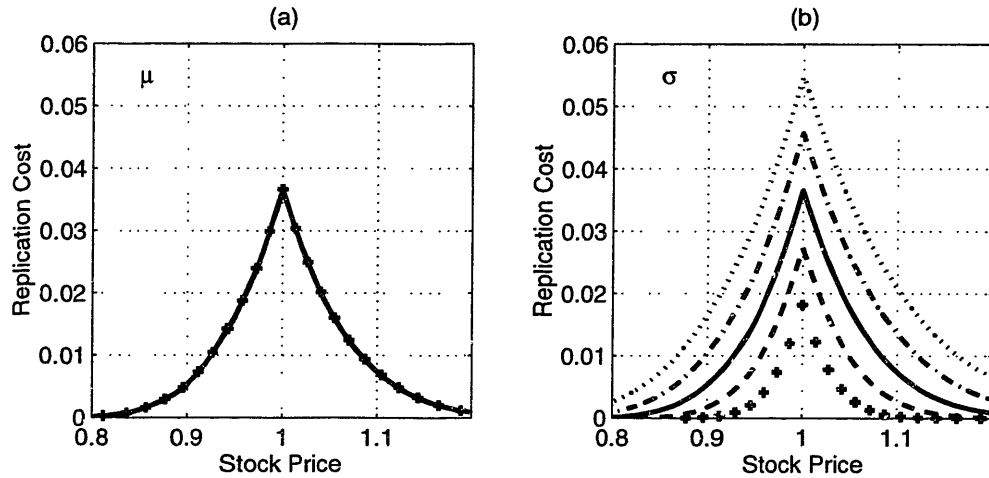


Figure 1-1: The difference between the replication cost and the intrinsic value of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows a geometric Brownian motion with parameter values $\mu = 0.07$ and $\sigma = 0.13$ corresponding to the solid line. In Panel (a), μ is varied and σ is fixed; in Panel (b), σ is varied and μ is fixed. In both cases, the variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line) and 0.5 (pluses).

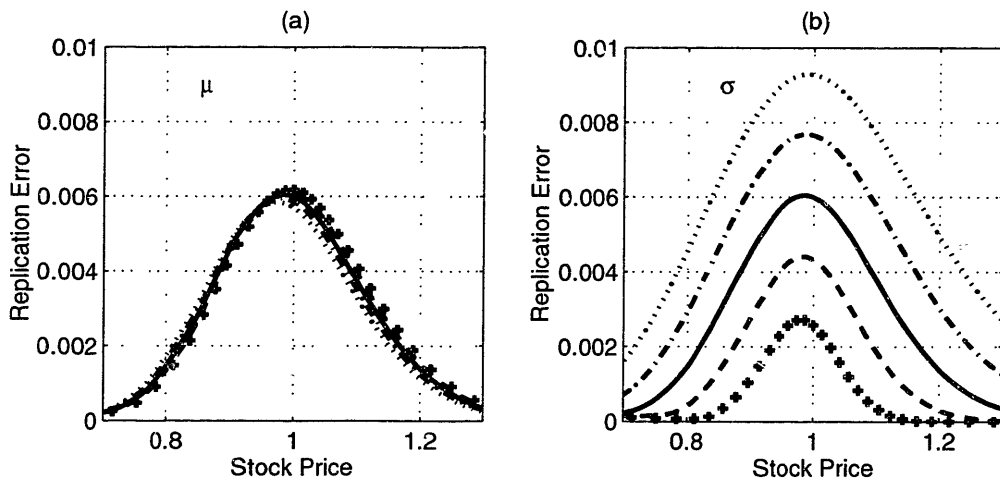


Figure 1-2: The replication error of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows a geometric Brownian motion with parameter values $\mu = 0.07$ and $\sigma = 0.13$ corresponding to the solid line. In Panel (a), μ is varied and σ is fixed; in Panel (b), σ is varied and μ is fixed. In both cases, the variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line) and 0.5 (pluses).

Figure 1-2 shows the dependence of the replication error ϵ^* on the initial stock price. Again we observe low sensitivity to the drift μ but, as in Figure 1-1, the replication error tends to increase with the volatility. Also, the replication error is highest when the stock price is close to the strike price.

Another important characteristic of the replication process is the ratio of the replication error to the replication cost ϵ^*/V_0^* , which we call the relative replication error. This ratio is more informative than the replication error itself since it describes the replication error per dollar spent, as opposed to the error of replicating a single option contract.

The dependence of the relative replication error on the initial stock price is displayed in Figure 1-3. This figure shows that the relative replication error is an increasing function of the initial stock price, i.e., it is higher for out-of-the-money options. Also, the relative replication error decreases with volatility for out-of-the-money options. This is not surprising given that it was defined as a ratio of the replication error to the hedging cost, both of which are increasing functions of volatility. According to this definition, the dependence of the relative replication error on volatility is determined by the tradeoff between increasing replication error and increasing replication cost.

1.5.3 Jump-Diffusion Models

In our numerical implementation of the jump-diffusion model (1.48)–(1.52) and (1.54), we restrict the number of jumps over a single time interval to be no more than three, which amounts to modifying the distribution of n_i in (1.49), originally given by (1.52).¹⁷

¹⁷This “truncation problem” is a necessary evil in the estimation of jump-diffusion models. See Ball and Torous (1985) for further discussion.

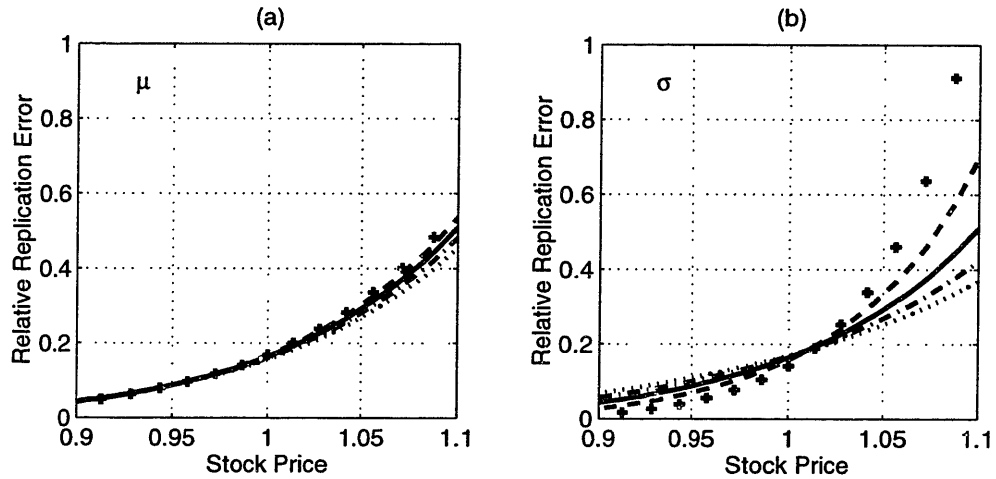


Figure 1-3: The relative replication error of a six-month maturity European put option (relative to the replication cost), plotted as a function of the initial stock price. The stock price follows a geometric Brownian motion with parameter values $\mu = 0.07$ and $\sigma = 0.13$ corresponding to the solid line. In Panel (a), μ is varied and σ is fixed; in Panel (b), σ is varied and μ is fixed. In both cases, the variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line) and 0.5 (pluses).

Specifically, we replace (1.52) with

$$\text{Prob}[n_i = m] = e^{-\lambda\Delta t} \frac{(\lambda\Delta t)^m}{m!}, \quad m = 1, 2, 3 \quad (1.82)$$

$$\text{Prob}[n_i = 0] = 1 - \sum_{m=1}^3 \text{Prob}[n_i = m]. \quad (1.83)$$

Besides this adjustment in the distribution of returns, our numerical procedure is exactly the same as in Section 1.4.2. We start with the following parameter values:

$$\mu = 0.07, \quad \sigma = 0.106, \quad \lambda = 25, \quad \delta = 0.015.$$

Then we study sensitivity of the solution to the parameter values by increasing and decreasing them by 25% and 50% while holding the other parameter values fixed. Our numerical

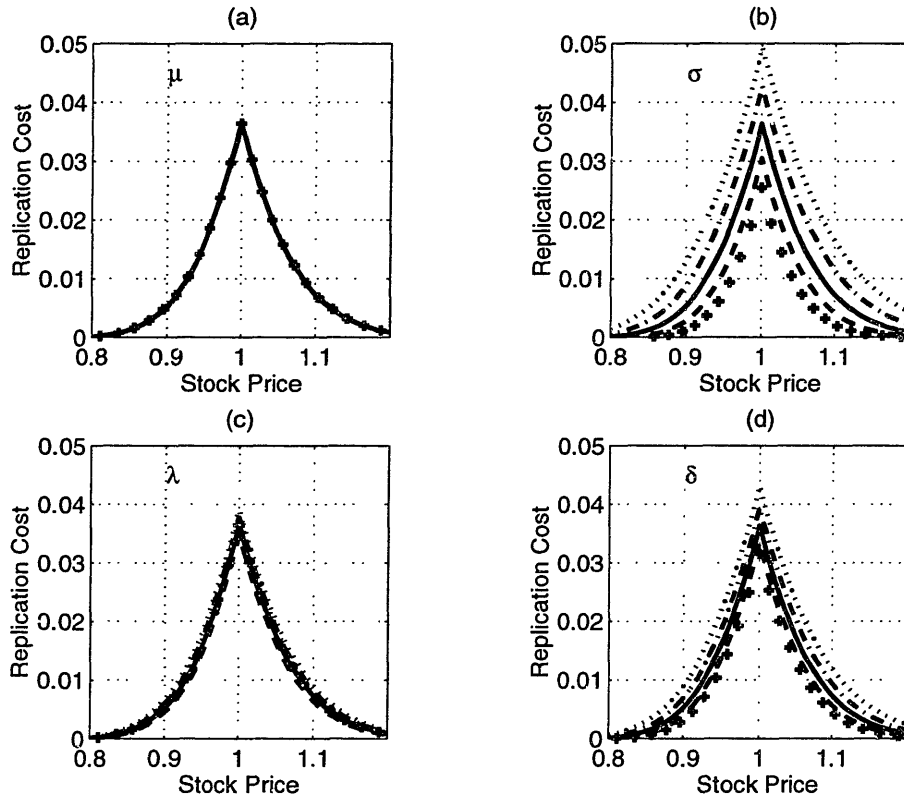


Figure 1-4: The difference between the replication cost and the intrinsic value of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process given in (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses).

results are summarized in Figures 1-4, 1-5, and 1-6.

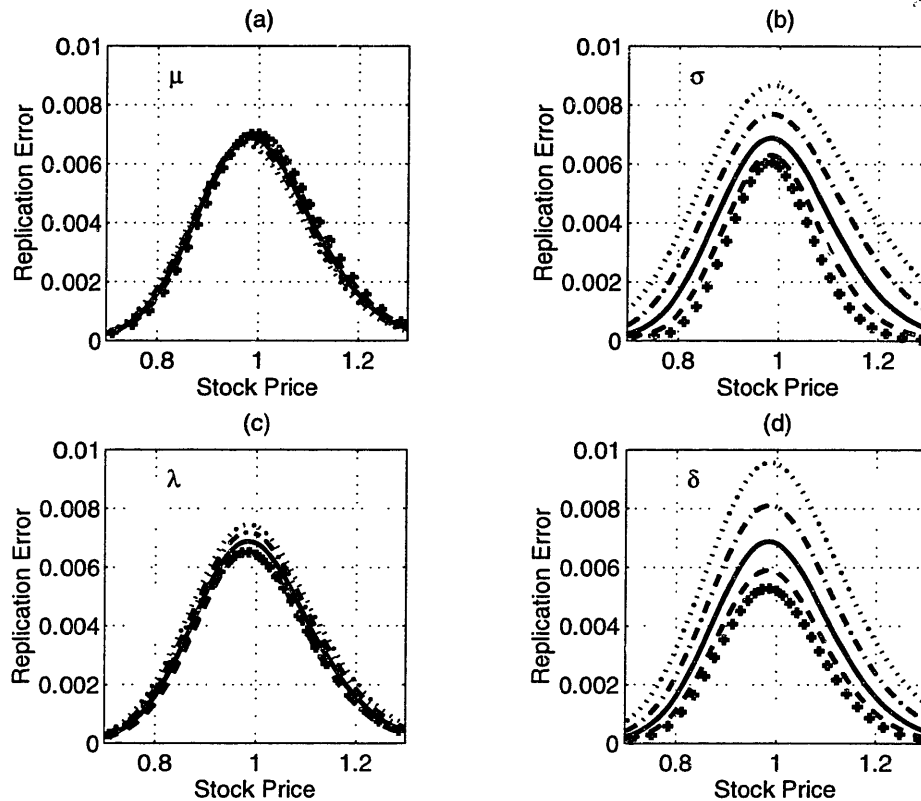


Figure 1-5: The replication error of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process given in (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses).

Figure 1-4 shows that the replication cost V_0^* is not sensitive to the drift rate μ and is increasing in volatility σ , the jump intensity λ , and the standard deviation δ of the jump magnitude. It is most sensitive to σ . Figure 1-5 shows that the replication error ϵ^* is not sensitive to μ and increases with all other parameters, with the highest sensitivity to δ . Finally, Figure 1-6 shows that the relative replication error ϵ^*/V_0^* is sensitive only to σ and it decreases as a function of σ for out-of-the-money options.

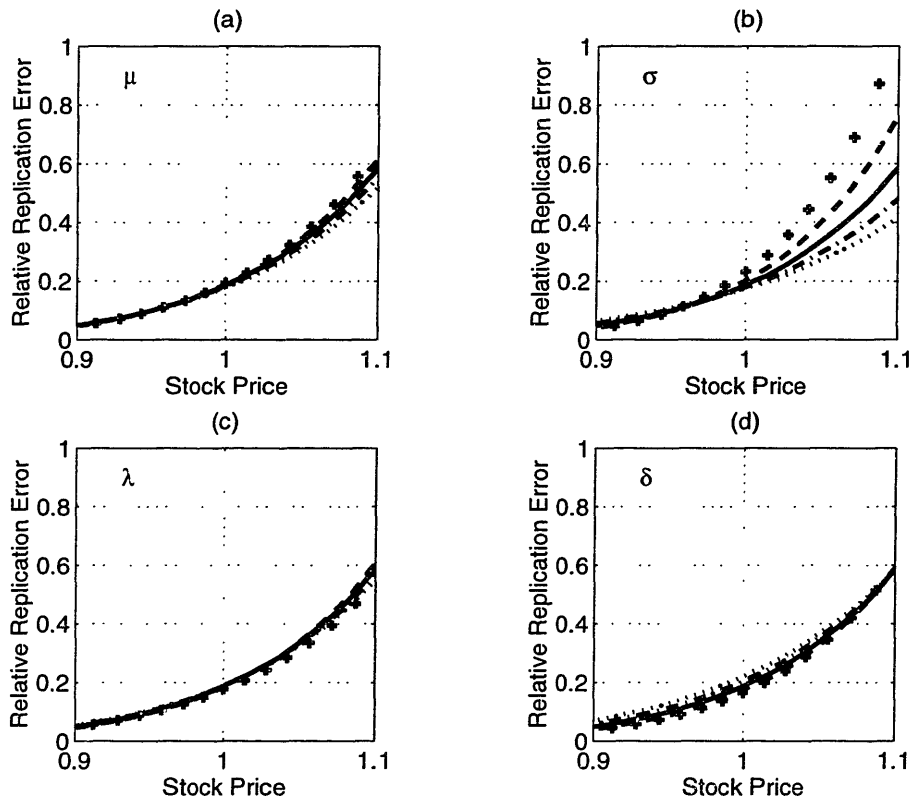


Figure 1-6: The relative replication error of a six-month maturity European put option (relative to the replication cost), plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses).

1.5.4 Stochastic Volatility

We begin by assuming a particular functional form for $g(\sigma)$ in (1.77):

$$g(\sigma) = -\delta\sigma(\sigma - \zeta) .$$

We also assume that the Brownian motions driving the stock price and volatility are uncorrelated. Since the closed-form expressions for the transition probability density of the diffusion process with stochastic volatility are not available, we base our computations on the discrete-time approximations of this process.¹⁸ The dynamics of stock prices and volatility are described by

$$P_{i+1} = P_i \exp \left((\mu - \sigma_i^2/2)\Delta t + \sigma_i \sqrt{\Delta t} z_{P_i} \right) \quad (1.84)$$

$$\sigma_{i+1} = \sigma_i \exp \left((-\delta(\sigma_i - \zeta) - \kappa^2/2)\Delta t + \kappa \sqrt{\Delta t} z_{\sigma_i} \right) \quad (1.85)$$

where $z_{P_i}, z_{\sigma_i} \sim \mathcal{N}(0, 1)$ and $E[z_{P_i} z_{\sigma_i}] = 0$. The parameters of the model are chosen to be

$$\mu = 0.07 \quad , \quad \zeta = 0.153 \quad , \quad \delta = 2 \quad , \quad \kappa = 0.4 . \quad (1.86)$$

We also assume that at time $t = 0$ volatility σ_0 is equal to 0.13. As before, we study sensitivity of the solution to parameter values. Our findings are summarized in Figures 1-7, 1-8, and 1-9.

We do not display the dependence on μ in these figures since the sensitivity to this

¹⁸This is done mostly for convenience, since we can approximate the transition probability density using Monte Carlo simulations. While the discrete-time approximations lead to significantly more efficient numerical algorithms, they are also consistent with many estimation procedures that replace continuous-time processes with their discrete-time approximations (see, for example, Ball and Torous [1985] and Wiggins [1987]).

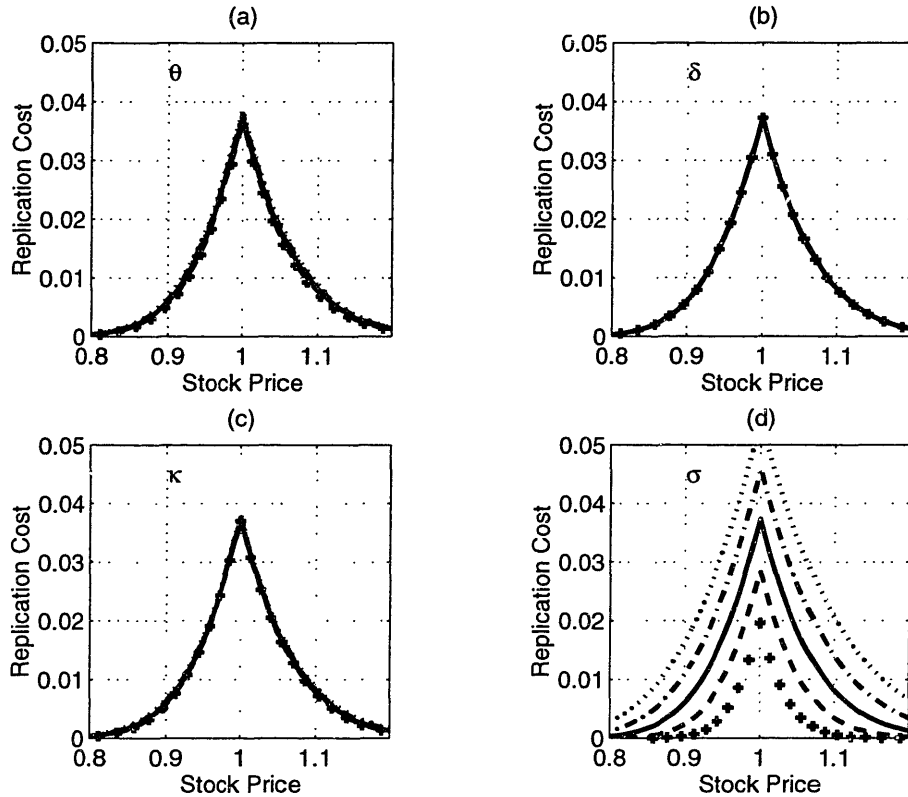


Figure 1-7: The difference between the replication cost and the intrinsic value of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows the stochastic volatility model (1.84)–(1.85) with parameter values $\mu = 0.07$, $\zeta = 0.153$, $\delta = 2$, $\kappa = 0.4$, and $\sigma_0 = 0.13$ corresponding to the solid line. In Panels (a)–(d), ζ , δ , κ , and σ_0 are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses).

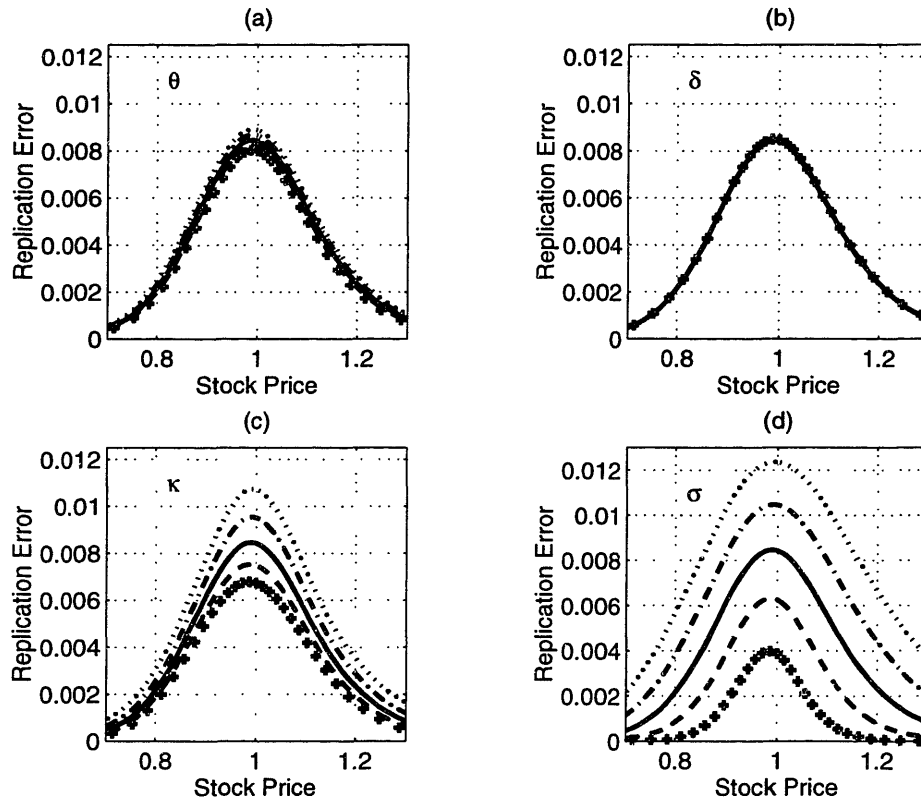


Figure 1-8: The replication error of a six-month maturity European put option, plotted as a function of the initial stock price. The stock price follows the stochastic volatility model (1.84)–(1.85) with parameter values $\mu = 0.07$, $\zeta = 0.153$, $\delta = 2$, $\kappa = 0.4$, and $\sigma_0 = 0.13$ corresponding to the solid line. In Panels (a)–(d), ζ , δ , κ , and σ_0 are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses).

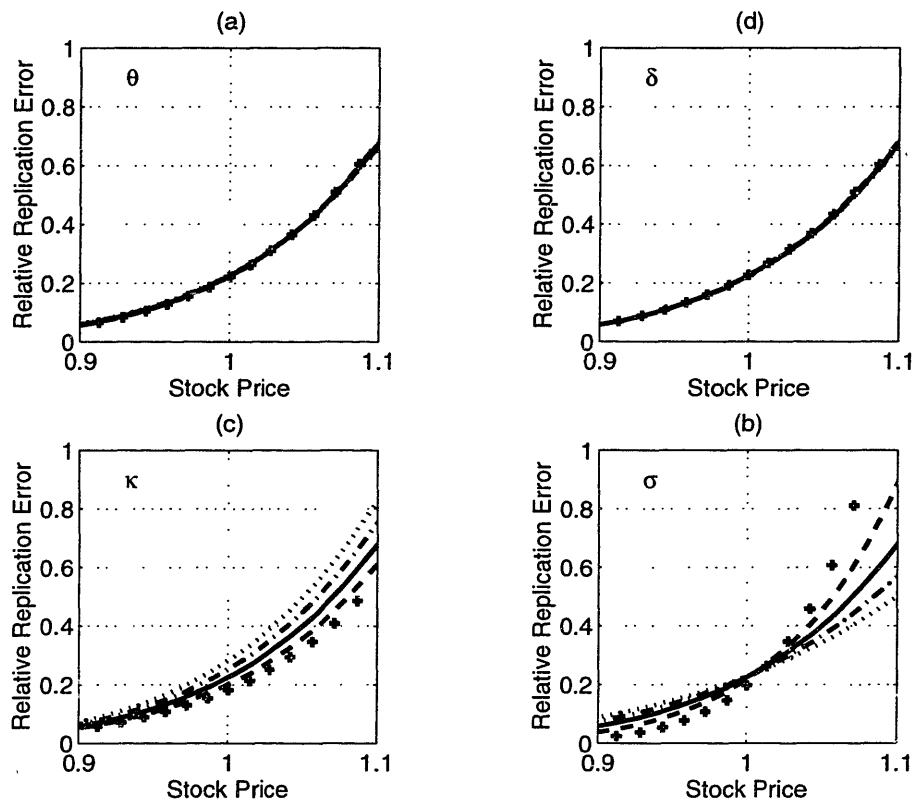


Figure 1-9: The relative replication error of a six-month maturity European put option (relative to the replication cost), plotted as a function of the initial stock price. The stock price follows the stochastic volatility model (1.84)–(1.85) with parameter values $\mu = 0.07$, $\zeta = 0.153$, $\delta = 2$, $\kappa = 0.4$, and $\sigma_0 = 0.13$ corresponding to the solid line. In Panels (a)–(d), ζ , δ , κ , and σ_0 are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses).

parameter is so low. Figure 1-7 shows that the replication cost is sensitive only to the initial value of volatility σ_0 and, as expected, the replication cost increases with σ_0 . Figure 1-8 shows that the replication error is sensitive to κ and σ_0 and is increasing in both of these parameters. According to Figure 1-9, the relative replication error is increasing in κ . It also increases in σ_0 for in-the-money options and decreases for out-of-the-money options.

In addition to its empirical relevance, the stochastic volatility model (1.76)–(1.77) also provides a clear illustration of the use of ϵ^* as a quantitative measure of dynamic market-incompleteness. Table 1.5.4 reports the results of Monte Carlo experiments in which the optimal-replication strategy is implemented for six sets of parameter values for the stochastic volatility model, including the set that yields geometric Brownian motion.

Model				Performance of Optimal Replication Strategy				
σ_0	ζ	δ	κ	V_0^*	ϵ^*	$\hat{\epsilon}^*$	θ_0^*	$\bar{\theta}^*$
0.13	0.153	2.00	0.400	0.0374	0.0084	0.0086	-0.474	-0.44
0.13	0.137	1.50	0.200	0.0367	0.0067	0.0068	-0.475	-0.43
0.13	0.133	1.00	0.100	0.0366	0.0062	0.0063	-0.475	-0.42
0.13	0.131	0.50	0.050	0.0366	0.0060	0.0059	-0.475	-0.41
0.13	0.130	0.25	0.025	0.0365	0.0060	0.0060	-0.475	-0.43
0.13	0.130	0.00	0.000	0.0365	0.0060	0.0062	-0.475	-0.42

Table 1.2: Monte Carlo simulation of the optimal-replication strategy θ^* for replicating a six-month at-the-money European put-option, for six sets of parameter values of the stochastic volatility model (1.84)–(1.85), including the set of parameter values that yields a geometric Brownian motion (last row). For each set of parameter values, 1,000 independent sample paths were simulated, each path containing 25 periods, and $P_0 = 1$.

For each set of parameter values, 1,000 independent sample paths of the stock price are simulated, each sample path containing 25 observations, and for each path the optimal-

replication strategy is implemented. The averages (over the 1,000 sample paths) of the minimum production cost V_0^* , the realized replication error $\hat{\epsilon}^*$, the initial optimal stock holdings θ_0^* , and the average optimal stock holdings $\bar{\theta}^*$ (over the 25 periods), is reported in each row. For comparison, the theoretical replication error ϵ^* is also reported.

Since stochastic volatility implies dynamically incomplete markets whereas geometric Brownian motion implies the opposite, these six sets of simulations comprise a sequence of models that illustrate the fact that market completeness need not be a binary characteristic. In particular, Table 1.5.4 shows that as the parameter values move closer to geometric Brownian motion, the average replication error ϵ^* decreases from 0.0086 to 0.0060. Moreover, the decrease between the first and second rows is considerably larger than the decrease between the second and third rows—the second and third rows imply price processes that are closer to each other in their degree of market completeness than that of the first row. Such specific rank orderings and sharp numerical comparisons are simply unavailable from standard dynamic equilibrium models that have been used to model market incompleteness.

Of course, ϵ^* is only one of many possible measures of market incompleteness—a canonical measure seems unlikely to emerge from the current literature—nevertheless it is an extremely useful measure given the practical implications that it contains for dynamically hedging risks.

1.5.5 Path-Dependent Options

We consider the option to “sell at the high” as described by Goldman et al. (1979), under the assumption that the stock price follows the mixed jump-diffusion process (1.48)–(1.51),

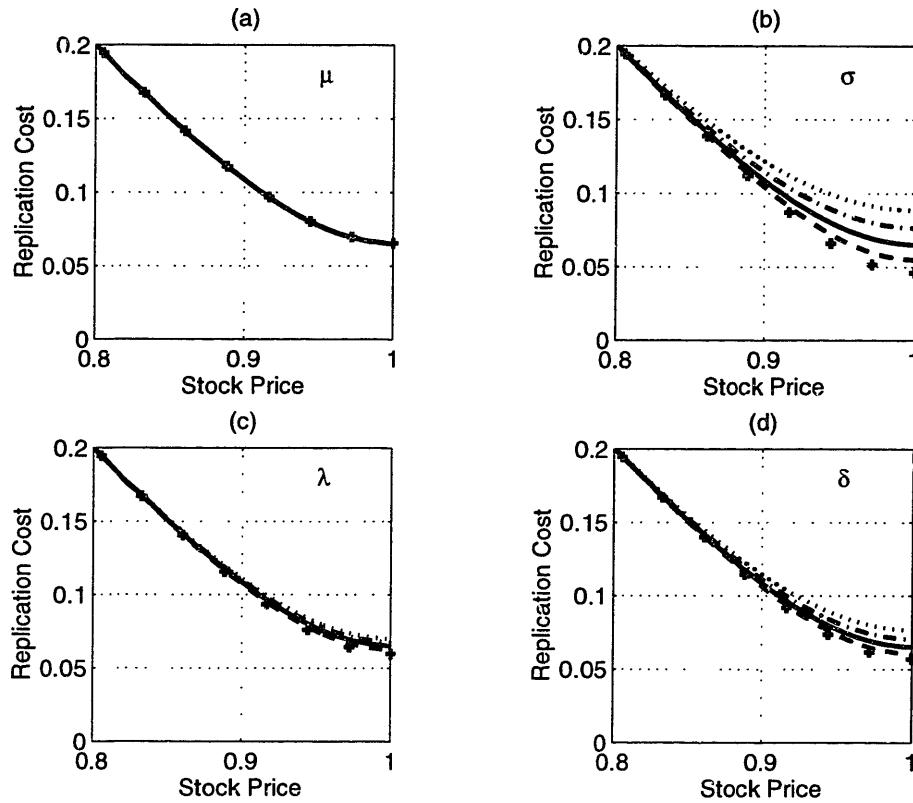


Figure 1-10: The replication cost of a six-month maturity European option to “sell at the high”, plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $m = 1$, $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses).

(1.54), (1.82), (1.83). We define the state variable Z :

$$Z_0 = m \geq P_0$$

$$Z_{i+1} = \max[Z_i, P_{i+1}] .$$

According to this definition, Z_i is the running maximum of the stock price process at time t_i . The initial value of Z_i is m , i.e., we assume that at time 0 the running maximum is equal to m .

The payoff of the option is given by

$$F(P_T, Z_T) = Z_T - P_T .$$

In our numerical analysis we set $m = 1$ as a convenient normalization. Note that this convention is just a change of scale and does not lead to any loss of generality.

The parameters for the stock price process are taken to be the same as in Section (1.4.3). The sensitivity of the replication cost and replication error on the initial stock price and parameters of the stock price process are reported in Figures 1-10, 1-11, and 1-12.

The qualitative behavior of the replication cost as a function of the initial stock price is similar to that of the option price as described in Goldman et al. (1979).¹⁹ Figure 1-10 shows that the replication cost V_0 is not sensitive to the drift rate μ and is increasing in volatility σ , the jump intensity jumps λ , and the standard deviation δ of the jump magnitude. It is most sensitive to σ . These observations are consistent with the behavior of the replication error of

¹⁹The difference between our model and that in Goldman et al. (1979) is that the latter assumes that the stock price follows a geometric Brownian motion and that continuous-time trading is allowed. Also the running maximum of the stock price process is calculated continuously, not over a discrete set of time moments, as in our case.

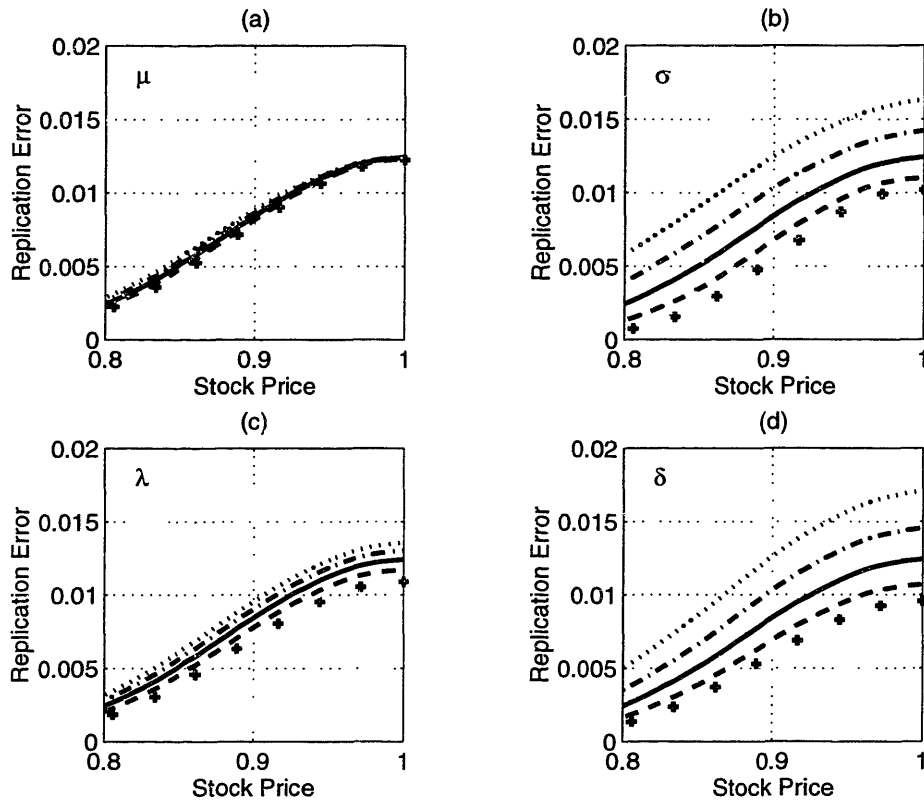


Figure 1-11: The replication error of a six-month maturity European option to “sell at the high”, plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $m = 1$, $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses).

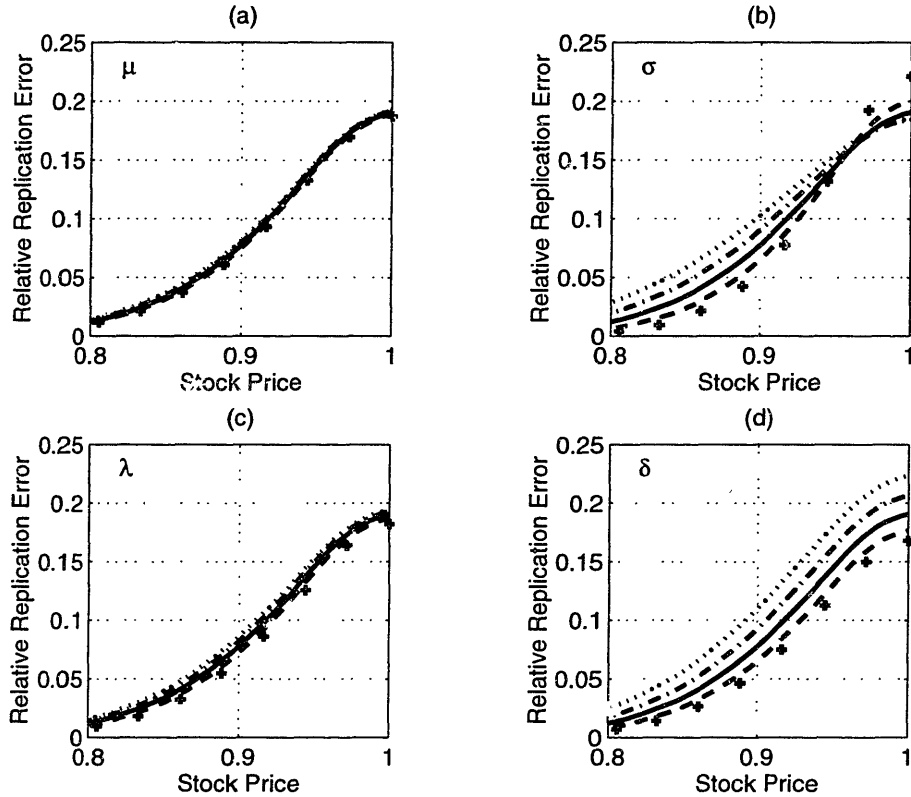


Figure 1-12: The relative replication error of a six-month maturity European option to “sell at the high” (relative to the replication cost), plotted as a function of the initial stock price. The stock price follows the mixed jump-diffusion process (1.48)–(1.51), (1.54), (1.82), and (1.83) with parameter values $m = 1$, $\mu = 0.07$, $\sigma = 0.106$, $\lambda = 25$, and $\delta = 0.015$ corresponding to the solid line. In Panels (a)–(d), μ , σ , λ , and δ are each varied, respectively, while the other parameter values are held fixed. The variation in each parameter is obtained by multiplying its original value by 1.25 (dashed-dotted line), 1.5 (dots), 0.75 (dashed line), and 0.5 (pluses).

the European put option in Section (1.4.3). According to Figure 1-11, the replication error ϵ^* is not sensitive to μ and is increasing in all other parameters with the highest sensitivity to δ and σ . Figure 1-6 shows that the relative replication error ϵ^*/V_0 is sensitive to σ and δ . It is an increasing function of δ , while the sign of the change of ϵ^*/V_0 with σ depends on the initial stock price P_0 .

1.6 Measuring the Degree of Market Incompleteness

In this section, we propose to measure the degree of market incompleteness by exploring the sensitivity of the replication error and the replication cost of a particular option contract to the specification of the stock-price dynamics. Specifically, we compare the following models: geometric Brownian motion, a mixed jump-diffusion process, and a diffusion process with stochastic volatility. The parameters of these models are calibrated to give rise to identical values of the expected instantaneous rate of return and volatility, hence we can view these three models as competing specifications of the same data-generating process.

1.6.1 Calibrating the Stochastic Processes

We consider a European put option with a unit strike price ($K = 1$) and a six-month maturity, i.e., $F(P_T) = \max[0, K - P_T]$. There are 25 trading periods, defined by $t_{i+1} - t_i = \Delta t = 1/50$. Since the closed-form expressions for the transition probability density of the mixed jump-diffusion process and the process with stochastic volatility are not available, we base our computations on the discrete-time approximations of these processes. The model specifications and corresponding parameter values are:

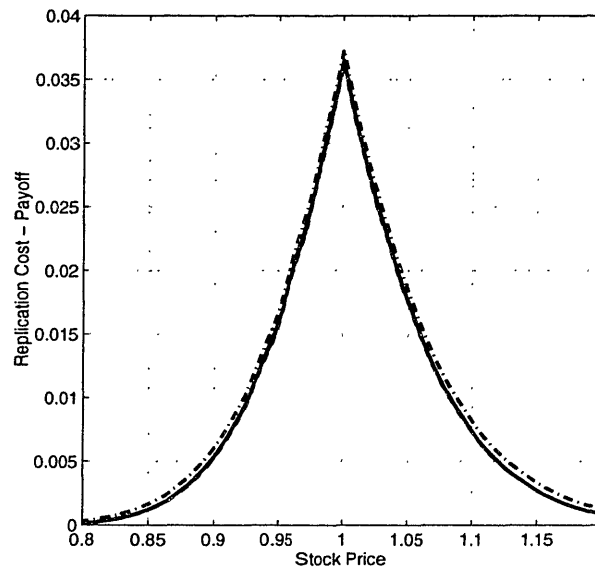


Figure 1-13: The difference between the replication cost and the intrinsic value of a six-month maturity European put option, plotted as a function of the initial stock price. Several processes for the stock price are plotted: geometric Brownian motion (1.41)–(1.43) (solid line); the mixed jump-diffusion model (1.48)–(1.51), (1.54), (1.82), and (1.83) (dashed line); and the stochastic volatility model (1.84)–(1.85) (dashed-dotted line). The parameter values are given by (1.87), (1.88), and (1.89).

1. **Geometric Brownian Motion.** Returns on the stock are lognormal, given by (1.41)–(1.43). We use the following parameter values:

$$\mu = 0.07 \quad , \quad \sigma = 0.13 . \quad (1.87)$$

2. **Mixed Jump-Diffusion.** The distribution of returns on the stock is given by (1.48)–(1.51), (1.54), (1.82), and (1.83). We use the following parameter values:

$$\mu = 0.07 \quad , \quad \sigma = 0.106 \quad , \quad \lambda = 25 \quad , \quad \delta = 0.015 . \quad (1.88)$$

3. **Diffusion with Stochastic Volatility.** Stock-price and volatility dynamics are given by (1.84)–(1.85). We assume that at time $t = 0$, volatility σ_0 is equal to 0.13, and the other parameter values are:

$$\mu = 0.07 \quad , \quad \zeta = 0.153 \quad , \quad \delta = 2 \quad , \quad \kappa = 0.4 . \quad (1.89)$$

1.6.2 Numerical Results

Figures 1-13–1-15 and Table 1.6.2 summarize our numerical results. Figure 1-13 presents the replication cost V_0^* minus the intrinsic value $F(P_0)$ for the three models as a function of the stock price at time $t = 0$. The hedging costs for the first two models are practically identical, while the stochastic volatility model can give rise to a significantly higher hedging costs for a deep-out-of-money option. Figure 1-14 and Table 1.6.2 shows the dependence of the replication error ϵ^* on the initial stock price.

All three models exhibit qualitatively similar behavior: the replication error is highest

Price Dynamics	Initial Stock Price P_0				
	0.90	0.95	1.00	1.05	1.10
Replication Cost Minus Intrinsic Value ($V_0^* - F(P_0)$)					
Geometric Brownian Motion	0.0054	0.0161	0.0365	0.0176	0.0074
Jump/Diffusion	0.0053	0.0159	0.0364	0.0175	0.0074
Stochastic Volatility	0.0060	0.0168	0.0374	0.0185	0.0082
Replication Error (ϵ^*)					
Geometric Brownian Motion	0.0046	0.0058	0.0060	0.0052	0.0038
Jump/Diffusion	0.0051	0.0066	0.0068	0.0059	0.0043
Stochastic Volatility	0.0061	0.0079	0.0084	0.0074	0.0056
Relative Replication Error (ϵ^*/V_0^*)					
Geometric Brownian Motion	0.043	0.088	0.164	0.292	0.509
Jump/Diffusion	0.049	0.100	0.187	0.335	0.585
Stochastic Volatility	0.058	0.119	0.226	0.400	0.679

Table 1.3: Comparison of replication costs and errors of the optimal-replication strategy for replicating a six-month European put option under competing specifications of price dynamics: geometric Brownian motion (1.41)–(1.43); the mixed jump-diffusion model (1.48)–(1.51), (1.54), (1.82), and (1.83); and the stochastic volatility model (1.84)–(1.85). The parameter values are given by (1.87), (1.88), and (1.89).

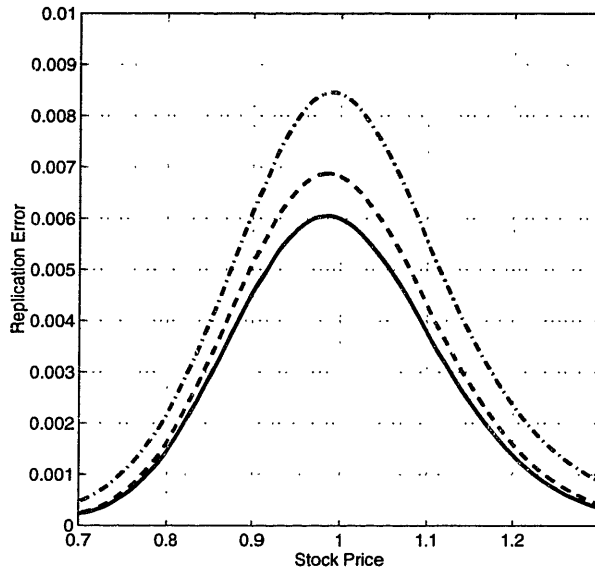


Figure 1-14: The replication error of a six-month maturity European put option, plotted as a function of the initial stock price. Several processes for the stock price are plotted: geometric Brownian motion (1.41)–(1.43) (solid line); the mixed jump-diffusion model (1.48)–(1.51), (1.54), (1.82), and (1.83) (dashed line); and the stochastic volatility model (1.84)–(1.85) (dashed-dotted line). The parameter values are given by (1.87), (1.88), and (1.89).

close to the strike price. For our choice of parameter values the replication error is highest for the stochastic volatility model and lowest for geometric Brownian motion. However, this need not hold in general. As we demonstrate in Section 1.4.3, the replication error of the mixed jump-diffusion process depends critically on δ and λ in (1.52, 1.54), thus, by varying these parameters, one can reverse the order of the curves in Figure 1-14 without changing the annualized volatility of the mixed jump-diffusion process.

The dependence of the relative replication error on the initial stock price is captured in Figure 1-15. As in Figure 1-13, the relative replication errors for the first two models are practically identical, while the stochastic volatility model can exhibit considerably higher errors. Also, while the relative replication error can be significant, particularly for an out-of-the-money option, the variation across the models is not as significant as one would expect. When continuous-time trading is allowed, the replication error for the geometric Brownian

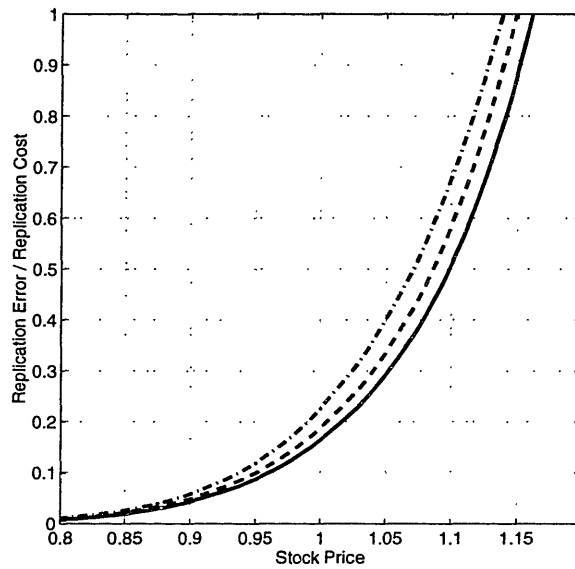


Figure 1-15: The relative replication error of a six-month maturity European put option (relative to the replication cost), plotted as a function of the initial stock price. Several processes for the stock price are plotted: geometric Brownian motion (1.41)–(1.43) (solid line); the mixed jump-diffusion model (1.48)–(1.51), (1.54), (1.82), and (1.83) (dashed line); and the stochastic volatility model (1.84)–(1.85) (dashed-dotted line). The parameter values are given by (1.87), (1.88), and (1.89).

motion model is zero, while the other two models give rise to strictly positive replication errors. This is an implication of the fact that the first model describes a dynamically complete market, while the other two correspond to markets which are dynamically incomplete (due to the absence of a sufficient number of traded instruments).

Nevertheless, as Figure 1-15 illustrates, the transition from continuous- to discrete-time trading can smear the differences between these models, leading to relative replication errors of comparable magnitude. This shows that the inability to trade continuously is just as important a source of market incompleteness as the absence of traded instruments.

1.7 Conclusion

We have proposed a method for replicating derivative securities in dynamically incomplete markets. Using stochastic dynamic programming, we construct a self-financing dynamic portfolio strategy that best approximates an arbitrary payoff function in a mean-squared sense. When markets are dynamically complete, as in the Black and Scholes (1973) and Merton (1973) models, our optimal-replication strategy coincides with the delta-hedging strategies of arbitrage-based models. Moreover, we provide an explicit algorithm for computing such strategies, which can be a formidable challenge in spite of market completeness, e.g., path-dependent derivatives such as “look-back” options.

When markets are not dynamically complete, as in the case of options on assets with stochastic volatility or with jump components, our approach yields the minimum production cost of a self-financing portfolio strategy with a terminal value that comes as close as possible (in a mean-squared sense) to the option’s payoff. This is the essence of the ϵ -arbitrage approach to synthetically replicating a derivative security.

We also argue that the replication error of the optimal-replication strategy can be used

as a quantitative measure for the degree of market incompleteness. Despite the difficulties in making welfare comparisons between markets with different types of incompleteness (see, for example, Duffie [1987], Duffie and Shafer [1985, 1986], and Hart [1974]), the minimum replication error of an ϵ -arbitrage strategy does provide one practical metric by which market completeness can be judged. After all, if it is possible to replicate the payoff of a derivative security to within some small error ϵ , the market for that security may be considered complete for all practical purposes even if ϵ is not zero.

Of course, this is only one of many possible measures of market completeness and we make no claims of generality here. Instead, we hope to have shown that Merton's (1973) seminal idea of dynamic replication has far broader implications than the dynamically-complete-markets setting in which it was originally developed. We plan to explore other implications in future research.

1.8 Appendix

The proofs of Theorems 1.2 and 1.1 are conceptually straightforward but notationally quite cumbersome. Therefore, we present only a brief sketch of the proofs below—interested readers can contact the authors for the more detailed mathematical appendix.

1.8.1 Proof of Theorem 1.1

The proof of Theorem 1.1 follows from dynamic programming. For $i = N$, (1.14)–(1.16) are clearly true, given (1.10). We now show that (1.17)–(1.21) describe the solution of the optimization problem in (1.9). First, as we observed in Section 1.3.3, the functions $a_i(\cdot, \cdot)$ are positive. Together with (1.3) this implies that

$$\mathbb{E} \left[J_{i+1} \left(V_i + \theta_i (P_{i+1} - P_i), P_{i+1}, \mathbf{Z}_{i+1} \right) \middle| V_i, P_i, \mathbf{Z}_i \right]$$

is a convex function of θ_i . Therefore, we can use the first-order condition to solve the optimization problem in (1.11):

$$\frac{d}{d\theta_i} \mathbb{E} \left[J_{i+1} \left(V_i + \theta_i (P_{i+1} - P_i), P_{i+1}, \mathbf{Z}_{i+1} \right) \middle| V_i, P_i, \mathbf{Z}_i \right] = 0, \quad (1.90)$$

where $J_{i+1}(\cdot, \cdot, \cdot)$ is given by (1.12). Equation (1.90) is a linear equation in θ_i and it is straightforward to check that its solution, $\theta^*(i, V_i, P_i, \mathbf{Z}_i)$, is given by (1.13), (1.17), and (1.18). We now substitute (1.13) into (1.3) and use (1.11) to calculate

$$\begin{aligned} J_i(V_i, P_i, \mathbf{Z}_i) &= \\ \mathbb{E} \left[J_{i+1} \left(V_i + \theta^*(i, V_i, P_i, \mathbf{Z}_i) \cdot (P_{i+1} - P_i), P_{i+1}, \mathbf{Z}_{i+1} \right) \middle| V_i, P_i, \mathbf{Z}_i \right]. \end{aligned} \quad (1.91)$$

Equations (1.19)–(1.21) are obtained by rearranging terms in (1.91). ■

1.8.2 Proof of Theorem 1.2

The more tedious algebraic manipulations of this proof were carried out using the symbolic algebra program *Maple*. Therefore, we shall outline the main ideas of the proof without reporting all of the details.

The cost-to-go function $J(t, V_t, P_t, \mathbf{Z}_t)$ satisfies the dynamic programming equation

$$\frac{\partial J}{\partial t} + \min_{\theta_t} \left\{ \left[\sum_{j=0}^J \mu_j Z_j \frac{\partial}{\partial Z_j} + \frac{1}{2} \sum_{i,j=0}^J \sigma_i \sigma_j Z_i Z_j \rho_{ij} \frac{\partial^2}{\partial Z_i \partial Z_j} + \theta_t \mu_0 Z_0 \frac{\partial}{\partial W} + \frac{1}{2} (\theta_t \sigma_0 Z_0)^2 \frac{\partial^2}{\partial W^2} + \theta_t \sum_{j=0}^J \sigma_0 \sigma_j \rho_{0j} Z_0 Z_j \frac{\partial^2}{\partial W \partial Z_j} \right] J \right\} = 0 \quad (1.92)$$

with boundary condition:

$$J(T, V_T, P_T, \mathbf{Z}_T) = [V_T - F(P_T, \mathbf{Z}_T)]^2 \quad (1.93)$$

where some of the functional dependencies were omitted to simplify the notation.

We must now check that the function $J(t, V_t, P_t, \mathbf{Z}_t)$, given by (1.26), (1.27)–(1.30), and the optimal control (1.31), satisfies (1.92)–(1.93). Boundary conditions (1.30) immediately imply (1.93). Next we substitute (1.26) into (1.92). It is easy to check, using equation (1.27), that function $a(\cdot)$ is positive. Therefore, the first-order condition is sufficient for the minimum in (1.92). This condition is a linear equation in θ_t which is solved by (1.31). It is now straightforward to verify that, whenever functions $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ satisfy (1.27)–(1.29), (1.92) is satisfied as well. ■

References

- Ait-Sahalia, Y. and A. Lo, 1996, "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices", to appear in *Journal of Finance*.
- Aiyagari, R., 1994, "Uninsured Idiosyncratic Risk and Aggregate Saving", *Quarterly Journal of Economics* 109, 659–684.
- Aiyagari, R. and M. Gertler, 1991, "Asset Returns with Transactions Costs and Uninsured Individual Risk", *Journal of Monetary Economics* 27, 311–331.
- Amin, K., 1993, "Jump Diffusion Option Valuation in Discrete Time", *Journal of Finance* 48, 1833–1863.
- Amin, K. and V. Ng, 1993, "Option Valuation With Systematic Stochastic Volatility", *Journal of Finance* 48, 881–910.
- Avellaneda, M. and A. Paras, 1994, "Dynamic Hedging Portfolios for Derivative Securities in the Presence of Large Transactions Costs", *Applied Mathematical Finance* 1, 165–193.
- Ball, C. and W. Torous, 1985, "On Jumps in Common Stock Prices and Their Impact on Call Option Pricing", *Journal of Finance* 40, 155–173.
- Banz, R. and M. Miller, 1978, "Prices for State- Contingent Claims: Some Estimates and Applications", *Journal of Business* 51, 653–672.
- Bensaid, B., Lesne, J., Pages, H. and J. Scheinkman, 1992, "Derivative Asset Pricing with Transaction Costs", *Mathematical Finance* 2, 63–86.
- Bertsekas, D., 1995, *Dynamic Programming and Optimal Control*, Vol. I. Belmont, MA: Athena Scientific.
- Bertsimas, D., Lo, A. and L. Kogan, 1999, "When Is Time Continuous?", to appear in *Journal of Financial Economics*.
- Bick, A., 1995, "Quadratic-Variation-Based Dynamic Strategies", *Management Science*, 41(4), 722–732.
- Black, F. and M. Scholes, 1973, "Pricing of Options and Corporate Liabilities", *Journal of Political Economy* 81, 637–654.
- Boyle, P. and D. Emanuel, 1980, "Discretely Adjusted Option Hedges", *Journal of Financial Economics* 8, 259–282.
- Boyle, P. and T. Vorst, 1992, "Option Replication in Discrete Time with Transaction Costs", *Journal of Finance* 47, 271–294.
- Brandt, M., 1998, "Multiperiod Hedging Contingent Claims", Working Paper, Wharton School, University of Pennsylvania.

- Breeden, D. 1979, "An Intertemporal Capital Pricing Model with Stochastic Investment Opportunities", *Journal of Financial Economics* 7, 265–296.
- Breeden, D. and R. Litzenberger, 1978, "State contingent Prices Implicit in Option Prices", *Journal of Business* 51, 621–651.
- Brennan, M., 1979, "The Pricing of Contingent claims in Discrete-Time Models", *Journal of Finance* 34, 53–68.
- Cox, J. and S. Ross, 1976, "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics* 3, 145–166.
- Cox, J., J. Ingersoll and S. Ross, 1985, "An Intertemporal General Equilibrium Model of Asset Prices", *Econometrica* 53, 363–384.
- Davis, M., Panas, V. and T. Zariphopoulou, 1993, "European Option Pricing with Transaction Costs", *SIAM Journal of Control and Optimization* 31, 470–493.
- Derman, E. and I. Kani, 1994, "Riding on the Smile", *RISK* 7 (February), 32–39.
- Duffie, D., 1996, *Dynamic Asset Pricing Theory*, Second Edition. Princeton, NJ: Princeton University Press.
- Duffie, D., 1987, "Stochastic Equilibria with Incomplete Financial Markets", *Journal of Economic Theory* 41, 405–416.
- Duffie, D. and C. Huang, 1985, "Implementing Arrow-Debreu Equilibria by Continuous Trading of Few Long-Lived Securities", *Econometrica* 53, 1337–1356.
- Duffie, D. and M. Jackson, 1990, "Optimal Hedging and Equilibrium in a Dynamic Futures Market", *Journal of Economic Dynamics and Control* 14, 21–33.
- Duffie, D. and W. Shafer, 1985, "Equilibrium in Incomplete Markets I: A Basic Model of Generic Existence", *Journal of Mathematical Economics* 14, 285–300.
- Duffie, D. and W. Shafer, 1986, "Equilibrium in Incomplete Markets II: Generic Existence in Stochastic Economies", *Journal of Mathematical Economics* 15, 199–216.
- Duffie, D. and H. Richardson, 1991, "Mean-Variance Hedging in Continuous Time", *Annals of Applied Probability* 1, 1–15.
- Duffie, D. and W. Zame, 1987, "The Consumption-Based Capital Asset Pricing Model", GSB Research Paper #922, Stanford University.
- Dybvig, P. and C. Huang, 1988, "Nonnegative Wealth, Absence of Arbitrage, and Feasible Consumption Plans", *Review of Financial Studies* 1, 377–401.
- Dumas, B., J. Flemming and R.E. Whaley, 1995, "Implied Volatility Functions: Empirical Tests", working paper, Fuqua School of Business, Duke University.
- Dupire, B., 1994, "Pricing with a Smile", *RISK* 7 (January), 18–20.
- Edirisinghe, C., Naik, V. and R. Uppal, 1993, "Optimal Replication of Options with Transaction Costs and Trading Restrictions", *Journal of Financial and Quantitative Analysis* 28, 117–138.

- Fleming W., R. Rishel, 1975, *Deterministic and Stochastic Optimal Control*. New York: Springer-Verlag.
- Föllmer, H. and D. Sonderman, 1986, "Hedging of Non-Redundant Contingent-Claims", in W. Hillebrand and A. Mas-Colell, eds., *Contributions to Mathematical Economics, in Honor of Gérard Debreu*. Amsterdam: North-Holland.
- Garman, M., 1976, "A General Theory of Asset Valuation under Diffusion State Processes", Working Paper No. 50, University of California Berkeley.
- Geske, R., 1979, "The Valuation of Compound Options", *Journal of Financial Economics* 7, 63–81.
- Grannan, E. and G. Swindle, 1996, "Minimizing Transaction Costs of Option Hedging Strategies", *Mathematical Finance* 6, 341–364.
- Harrison, J. and D. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory* 20, 381–408.
- Hart, O., 1974, "On the Existence of Equilibrium in a Securities Model", *Journal of Economic Theory* 9, 293–311.
- He, H. and D. Modest, 1995, "Market Frictions and Consumption-Based Asset Pricing", *Journal of Political Economy* 103, 94–117.
- Heaton, J. and D. Lucas, 1992, "The Effects of Incomplete Insurance Markets and Trading Costs in a Consumption-Based Asset Pricing Model", *Journal of Economic Dynamics and Control* 16, 601–620.
- Heaton, J. and D. Lucas, 1996, "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing", *Journal of Political Economy* 104, 443–487.
- He, H. and N. Pearson, 1991, "Consumption and Portfolio Policies with Incomplete Markets and Short-Sale Constraints", *Journal of Economic Theory* 54, 259–304.
- Henrotte, P., 1993, "Transactions Costs and Duplication Strategies", unpublished working paper, Stanford University.
- Hodges, S. and A. Neuberger, 1989, "Optimal Replication of Contingent Claims Under Transaction Costs", *Review of Futures Markets* 8, 222–239.
- Huang, C., 1987, "An Intertemporal General Equilibrium Asset Pricing Model: The Case of Diffusion Information", *Econometrica* 55, 117–142.
- Hull, J. and A. White, 1987, "The Pricing of Options on Assets With Stochastic Volatilities", *Journal of Finance* 42, 281–300.
- Hutchinson, J., Lo, A. and T. Poggio, 1994, "A Nonparametric Approach to the Pricing and Hedging of Derivative Securities Via Learning Networks", *Journal of Finance* 49, 851–889.
- Jarrow, R. and A. Rudd, 1982, "Approximate Option Valuation for Arbitrary Stochastic Processes", *Journal of Financial Economics* 10, 347–369.

- Johnson H. and D. Shanno, 1987, "Option Pricing When the Variance is Changing", *Journal of Financial and Quantitative Analysis* 22, 143–151.
- Leland, H., 1985, "Option Pricing and Replication with Transaction Costs", *Journal of Finance*, 40, 1283–1301.
- Longstaff, F., 1992, "An Empirical Examination of the Risk-Neutral Valuation Model", working paper, College of Business, Ohio State University, and the Anderson Graduate School of Management, UCLA.
- Longstaff, F., 1995, "Option Pricing and the Martingale Restriction", *Review of Financial Studies* 8, 1091–1124.
- Lucas, R., 1978, "Asset Prices in an Exchange Economy", *Econometrica* 46, 1426–1446.
- Lucas, D., 1994, "Asset Pricing With Undiversifiable Income Risk and Short Sales Constraints: Deepening the Equity Premium Puzzle", *Journal of Monetary Economics* 34, 325–341.
- Magill, M. and M. Quinzii, 1996, *Theory of Incomplete Markets*. Cambridge, MA: MIT Press.
- Merton, R., 1971, "Optimum Consumption and Portfolio Rules in a Continuous-Time Model", *Journal of Economic Theory* 3, 373–413.
- Merton, R., 1973, "An Intertemporal Capital Asset Pricing Model", *Econometrica* 41, 867–887.
- Merton, R., 1976, "Option Pricing when Underlying Stock Returns are Discontinuous", *Journal of Financial Economics* 3, 125–144.
- Merton, R., 1992, *Continuous-Time Finance*. Cambridge, MA: Blackwell Publishers.
- Neuberger, A., 1994, "Option Replication with Transaction Costs: An Exact Solution for the Pure Jump Process", in D. Chance and R. Trippi, eds., *Advances in Futures and Options Research*, Volume 7. Greenwich, CT: JAI Press.
- Rady, S., 1994, "State Prices Implicit in Valuation Formulae for Derivative Securities: A Martingale Approach", Discussion Paper No. 181, LSE Financial Markets Group, London, UK.
- Rubinstein, M., 1976, "The Valuation of Uncertain Income Streams and the Pricing of Options", *Bell Journal of Economics* 7, 407–425.
- Rubinstein, M., 1983, "Displaced Diffusion Option Pricing", *Journal of Finance* 38, 213–217.
- Rubinstein, M., 1994, "Implied Binomial Trees", *Journal of Finance* 49, 771–818.
- Schäl, M., 1994, "On Quadratic Cost Criteria for Option Hedging" *Mathematics of Operations Research* 19, 121–131.
- Scheinkman, J. and L. Weiss, 1986, "Borrowing Constraints and Aggregate Economic Activity", *Econometrica* 54, 23–45.

- Schweizer, M., 1992, "Mean-Variance Hedging for General Claims", *Annals of Applied Probability* 2, 171–179.
- Schweizer, M., 1995, "Variance-Optimal Hedging in Discrete Time", *Mathematics of Operations Research* 20, 1–31.
- Shimko, D., 1993, "Bounds of Probability", *RISK* 6, 33–37.
- Stroud, A., 1971, *Approximate Calculation of Multiple Integrals*. Englewood Cliffs, NJ: Prentice-Hall.
- Stutzer, M., 1995, "A Simple Nonparametric Approach to Derivative Security Valuation", forthcoming in the *Journal of Finance*.
- Telmer, C., 1993, "Asset-Pricing Puzzles and Incomplete Markets", *Journal of Finance* 48, 1803–1832.
- Toft, K., 1996, "On the Mean-Variance Tradeoff in Option Replication with Transactions Costs", *Journal of Financial and Quantitative Analysis* 31, 233–263.
- Weil, P., 1992, "Equilibrium Asset Prices with Undiversifiable Labor Income Risk", *Journal of Economic Dynamics and Control* 16, 769–790.
- Whalley, E. and P. Wilmott, 1994, "Hedge with an Edge", *RISK* 7.
- Wiggins, J., 1987, "Option Values Under Stochastic Volatility: Theory and Empirical Estimates", *Journal of Financial Economics* 5, 351–372.

Chapter 2

When Is Time Continuous? (joint with

Dimitris Bertsimas and Andrew Lo)

2.1 Abstract

We study the tracking error, resulting from the discrete-time application of continuous-time delta-hedging procedures for European options. We characterize the asymptotic distribution of the tracking error as the number of discrete time periods increases, and its joint distribution with other assets. We introduce the notion of temporal granularity of the continuous time stochastic model that enables us to characterize the degree to which discrete time approximations of continuous time models track the payoff of the option. We derive closed form expressions for the granularity for a put or call option on a stock that follows a geometric Brownian motion and a mean-reverting process. These expressions offer insight into the tracking error involved in applying continuous-time delta-hedging in discrete time. We also introduce alternative measures of the tracking error and analyze their properties.

2.2 Introduction

Since Wiener's (1923) pioneering construction of Brownian motion and Itô's (1951) theory of stochastic integrals, continuous-time stochastic processes have become indispensable to many disciplines ranging from chemistry and physics to engineering to biology to financial economics. In fact, the application of Brownian motion to financial markets pre-dates Wiener's contribution by almost a quarter century (see Bachelier (1900)), and Merton's (1973) seminal derivation of the Black and Scholes (1973) option-pricing formula in continuous time—and, more importantly, his notion of delta hedging and dynamic replication—is often cited as the foundation of today's multi-trillion dollar derivatives industry.

Indeed, the mathematics and statistics of Brownian motion have become so intertwined with so many scientific theories that we often forget the fact that continuous-time processes are only approximations to physically realizable phenomena. In fact, for the more theoretically inclined, Brownian motion may seem more “real” than discrete-time discrete-valued processes. Of course, whether time is continuous or discrete is a theological question best left for philosophers. But a more practical question remains: under what conditions are continuous-time models good approximations to specific physical phenomena, i.e., when does time seem “continuous” and when does it seem “discrete”?

In this paper, we provide a concrete answer to this question in the context of continuous-time derivative-pricing models, e.g., Merton (1973), by characterizing the replication errors that arise from delta hedging derivatives in discrete time.

Delta-hedging strategies play a central role in the theory of derivatives and in our understanding of dynamic notions of spanning and market completeness. In particular, delta-hedging strategies are recipes for replicating the payoff of a complex security by sophisticated dynamic trading of simpler securities. When markets are dynamically complete (see, for

example, Harrison and Kreps [1979] and Duffie and Huang [1985]) and continuous trading is feasible, it is possible to replicate certain derivative securities perfectly. However, when markets are not complete or when continuous trading is not feasible, e.g., trading frictions or periodic market closings, perfect replication is not possible and the usual delta-hedging strategies exhibit *tracking* errors. These tracking errors comprise the focus of our attention.

Specifically, we characterize the asymptotic distribution of the tracking errors of delta-hedging strategies using continuous-record asymptotics, i.e., we implement these strategies in discrete time and let the number of time periods increase while holding the time span fixed. Since the delta-hedging strategies we consider are those implied by continuous-time models like Merton (1973), it is not surprising that tracking errors arise when such strategies are implemented in discrete time, nor is it surprising that these errors disappear in the limit of continuous time. However, by focusing on the continuous-record asymptotics of the tracking error, we can quantify the discrepancy between the discrete-time hedging strategy and its continuous-time limit, answering the question “When is time continuous?” in the context of replicating derivative securities.

We show that the normalized tracking error converges weakly to a particular stochastic integral and that the root-mean-squared tracking error is of order $N^{-1/2}$ where N is the number of discrete time periods over which the delta hedging is performed. This provides a natural definition for *temporal granularity*: it is the coefficient that corresponds to the $O(N^{-1/2})$ term. We derive a closed-form expression for the temporal granularity of a diffusion process paired with a derivative security, and propose this as a measure of the “continuity” of time. The fact that granularity is defined with respect to a derivative-security/price-process pair underscores the obvious: a need for specificity in quantifying the approximation errors of continuous-time processes. It is impossible to tell how good

an approximation a continuous-time process is to a physical process without specifying the nature of the physical process.

In addition to the general usefulness of a measure of temporal granularity for continuous-time stochastic processes, our results have other, more immediate applications. For example, for a broad class of derivative securities and price processes, our measure of granularity provides a simple method for determining the approximate number of hedging intervals N^* needed to achieve a target root-mean-squared-error δ : $N^* = g^2/\delta^2$ where g is the granularity coefficient of the derivative-security/price-process pair. This expression shows that to halve the root-mean-squared-error of a typical delta-hedging strategy, the number of hedging intervals must be increased approximately fourfold.

Moreover, for some special cases, e.g., the Black-Scholes case, the granularity coefficient can be obtained in closed form, and these cases shed considerable light on several aspects of derivatives replication. For example, in the Black-Scholes case, does an increase in volatility make it easier or more difficult to replicate a simple call option? Common intuition suggests that the tracking error increases with volatility, but the closed-form expression for granularity (2.20) shows that the granularity achieves a maximum as a function of σ and that beyond this point, it becomes a decreasing function of σ . The correct intuition is that at lower levels of volatility, tracking error is an increasing function of volatility because an increase in volatility implies more price movements and a greater likelihood of hedging errors in each hedging interval. But at higher levels of volatility, price movements are so extreme that an increase in volatility in this case implies that prices are less likely to fluctuate near the strike price where delta-hedging errors are the largest, hence granularity is a decreasing function of σ . In other words, at sufficiently high levels of volatility, the nonlinear payoff function of a call option “looks” approximately linear and is therefore easier

to hedge. Similar insights can be gleaned from other closed-form expressions of granularity (see, for example, Section 2.4.2).

In Section 2.3, we provide a complete characterization of the asymptotic behavior of the tracking error for delta hedging an arbitrary derivative security, and formally introduce the notion of granularity. To illustrate the practical relevance of granularity, in Section 2.4 we obtain closed-form expressions for granularity in two specific cases: call options under geometric Brownian motion, and under a mean-reverting process. In Section 2.5 we check the accuracy of our continuous-record asymptotic approximations by presenting Monte Carlo simulation experiments for the two examples of Section 2.4 and comparing them to the corresponding analytical expressions. We present other extensions and generalizations in Section 2.6 such as a characterization of the sample-path properties of tracking errors, the joint distributions of tracking errors and prices, a PDE characterization of the tracking error, and more general loss functions than root-mean-squared tracking error. We conclude in Section 2.7.

2.3 Defining Temporal Granularity

The relationship between continuous-time and discrete-time models in economics and finance has been explored in a number of studies. One of the earliest examples is Merton (1969), in which the continuous-time limit of the budget equation of a dynamic portfolio choice problem is carefully derived from discrete-time considerations (see also Merton [1975, 1982b]). Foley's (1975) analysis of "beginning-of-period" versus "end-of-period" models in macroeconomics is similar in spirit, though quite different in substance.

More recent interest in this issue stems primarily from two sources. On the one hand, it is widely recognized that continuous-time models are useful and tractable approximations

to more realistic discrete-time models. Therefore, it is important to establish that key economic characteristics of discrete-time models converge properly to the characteristics of their continuous-time counterparts. A review of recent research along these lines can be found in Duffie and Protter (1992).

On the other hand, while discrete-time and discrete-state models such as those based on binomial and multinomial trees, e.g., Cox, Ross, and Rubinstein (1979), He (1990, 1991), and Rubinstein (1994), may not be realistic models of actual markets, nevertheless they are convenient computational devices for analyzing continuous-time models. Willinger and Taqqu (1991) formalize this notion and provide a review of this literature.

For derivative-pricing applications, the distinction between discrete-time and continuous-time models is a more serious one. For all practical purposes, trading takes place at discrete intervals, and a discrete-time implementation of Merton's (1973) continuous-time delta-hedging strategy cannot perfectly replicate an option's payoff. The tracking error that arises from implementing a continuous-time hedging strategy in discrete time has been studied by several authors.

One of the first studies was conducted by Boyle and Emanuel (1980), who consider the statistical properties of "local" tracking errors. At the beginning of a sufficiently small time interval, they form a hedging portfolio comprised of options and stock according to the continuous-time Black-Scholes/Merton delta-hedging formula. The composition of this hedging portfolio is held fixed during this time interval, which gives rise to a tracking error (in continuous time, the composition of this portfolio would be adjusted continuously to keep its dollar value equal to zero). The dollar-value of this portfolio at the end of the interval is then used to quantify the tracking error.

More recently, Toft (1996) shows that a closed-form expression for the variance of the

cash flow from a discrete-time delta-hedging strategy can be obtained for a call or put option in the special case of geometric Brownian motion. However, he observes that this expression is likely to span several pages and is therefore quite difficult to analyze.

But perhaps the most relevant literature for our purposes is Leland's (1985) investigation of discrete-time delta-hedging strategies motivated by the presence of transactions costs, an obvious but important motivation (why else would one trade discretely?) that spurred a series of studies on option pricing with transactions costs, e.g., Figlewski (1989), Hodges and Neuberger (1989), Bensaid et al. (1992), Boyle and Vorst (1992), Ederisinghe, Naik, and Uppal (1993), Henrotte (1993), Avellaneda and Paras (1994), Neuberger (1994), and Grannan and Swindle (1996). This strand of the literature provides compelling economic motivation for discrete delta-hedging: trading continuously would generate infinite transactions costs. However, the focus of these studies is primarily the *tradeoff* between the magnitude of tracking errors and the cost of replication. Since we focus on only one of these two issues—the approximation errors that arise from applying continuous-time models discretely—we are able to characterize the statistical behavior of tracking errors much more generally, i.e., for large classes of price processes, payoff functions, and state variables.

Specifically, we investigate the discrete-time implementation of continuous-time delta-hedging strategies and derive the asymptotic distribution of the tracking error in considerable generality by appealing to continuous-record asymptotics. We introduce the notion of temporal granularity which is central to the issue of when time may be considered continuous, i.e., when continuous-time models are good approximations to discrete-time phenomena. In Section 2.3.1, we describe the framework in which our delta-hedging strategy will be implemented and define the tracking error and related quantities. In Section 2.3.2, we characterize the continuous-record asymptotic behavior of the tracking error and define

the notion of temporal granularity. We provide an interpretation of granularity in Section 2.3.3 and discuss its implications.

2.3.1 Delta-Hedging in Complete Markets

We begin by specifying the market environment. For simplicity, we assume that there are only two traded securities: a riskless asset (bond) and a risky asset (stock). Time t is normalized to the unit interval so that trading takes place from $t=0$ to $t=1$. In addition, we assume

(A5) *Markets are frictionless, i.e., there are no taxes, transactions costs, shortsales restrictions, or borrowing restrictions.*

(A6) *The riskless borrowing and lending rate is 0.¹*

(A7) *The price P_t of the risky asset follows a diffusion process*

$$\frac{dP_t}{P_t} = \mu(t, P_t) dt + \sigma(t, P_t) dW_t \quad , \quad \sigma(t, P_t) \geq \sigma_0 > 0 \quad (2.1)$$

where the coefficients μ and σ satisfy standard regularity conditions that guarantee existence and uniqueness of the strong solution of (2.1) and market completeness (see Duffie [1996]).

We now introduce a European derivative security on the stock that pays $F(P_1)$ dollars at time $t=1$. We will call $F(\cdot)$ the payoff function of the derivative. The equilibrium price of the derivative, $H(t, P_t)$, satisfies the following partial differential equation (PDE) (see,

¹This entails little loss of generality since we can always renormalize all prices by the price of a zero-coupon bond with maturity at time 1 (see, for example, Harrison and Kreps [1979]). However, this assumption does rule out the case of a stochastic interest rate.

for example, Cox, Ingersoll, and Ross [1985]):

$$\frac{\partial H(t, x)}{\partial t} + \frac{1}{2} \sigma^2(t, x) x^2 \frac{\partial^2 H(t, x)}{\partial x^2} = 0 \quad (2.2)$$

with the boundary condition

$$H(1, x) = F(x) . \quad (2.3)$$

This is a generalization of the standard Black-Scholes model which can be obtained as a special case when the coefficients of the diffusion process (2.1) are constant, i.e., $\mu(t, P_t) = \mu$, $\sigma(t, P_t) = \sigma$, and the payoff function $F(P_1)$ is given by $\text{Max}[P_1 - K, 0]$ or $\text{Min}[P_1, K]$.

The delta-hedging strategy was introduced by Black and Scholes (1973) and Merton (1973) and when implemented continuously on $t \in [0, 1]$, the payoff of the derivative at expiration can be replicated perfectly by a portfolio of stocks and riskless bonds. This strategy consists of forming a portfolio at time $t = 0$ containing only stocks and bonds with an initial investment of $H(0, P_0)$, and rebalancing it continuously in a *self-financing* manner—all long positions are financed by short positions and no money is withdrawn or added to the portfolio—so that at all times $t \in [0, 1]$ the portfolio contains $\partial H(t, P_t) / \partial P_t$ shares of the stock. The value of such a portfolio at time $t = 1$ is exactly equal to the payoff, $F(P_1)$, of the derivative. Therefore, the price, $H(t, P_t)$, of the derivative can also be considered the production cost of replicating the derivative's payoff $F(P_1)$ starting at time t .

Such an interpretation becomes important when continuous-time trading is not feasible. In this case, $H(t, P_t)$ can no longer be viewed as the equilibrium price of the derivative. However, the function $H(t, P_t)$, defined formally as a solution of (2.2)–(2.3), can still be

viewed as the production cost $H(0, P_0)$ of an approximate replication of the derivative's payoff, and may be used to define the production process itself (we formally define a discrete-time delta-hedging strategy below).² Therefore, when we refer to $H(t, P_t)$ as the derivative's "price" below, we shall have in mind this more robust interpretation of production cost and approximate replication strategy.³

More formally, we assume:

(A8) *Trading takes place only at N regularly spaced times t_i , $i = 1, \dots, N$, where*

$$t_i \in \left\{ 0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N} \right\}.$$

Under (A8), the difference between the payoff of the derivative and the end-of-period dollar-value of the replicating portfolio—the *tracking error*—will be nonzero.

Following Hutchinson, Lo, and Poggio (1994), let $V_{t_i}^{(N)}$ be the value of the replicating portfolio at time t_i . Since the replicating portfolio consists of shares of the stock and the bond, we can express $V_{t_i}^{(N)}$ as

$$V_{t_i}^{(N)} = V_{S,t_i}^{(N)} + V_{B,t_i}^{(N)} \quad (2.4)$$

where $V_{S,t_i}^{(N)}$ and $V_{B,t_i}^{(N)}$ denote the dollar amount invested in the stock and the bond, respectively, in the replicating portfolio at time t_i . At time $t=0$ the total value of the replicating

²The term "approximate replication" indicates the fact that when continuous trading is not feasible, the difference between the payoff of the derivative and the end-of-period dollar-value of the replicating portfolio will be nonzero. See Bertsimas, Kogan, and Lo (1997) for a discussion of derivative replication in discrete time and the distinction between production cost and equilibrium price.

³Alternatively, we can conduct the following equivalent thought experiment: while some market participants can trade costlessly and continuously in time and thus ensure that the price of the derivative is given by the solution of (2.2)–(2.3), we will focus our attention on other market participants who can trade only a finite number of times.

portfolio is equal to the price (production cost) of the derivative

$$V_0^{(N)} = H(0, P_0) \quad (2.5)$$

and its composition is given by

$$V_{S,0}^{(N)} = \left. \frac{\partial H(t, P_t)}{\partial P_t} \right|_{t=0} P_0, \quad V_{B,0}^{(N)} = V_0^{(N)} - V_{S,0}^{(N)}, \quad (2.6)$$

hence the portfolio contains $\partial H(t, P_t)/\partial P_t|_{t=0}$ shares of stock. The replicating portfolio is rebalanced at time periods t_i so that

$$V_{S,t_i}^{(N)} = \left. \frac{\partial H(t, P_t)}{\partial P_t} \right|_{t=t_i} P_{t_i}, \quad V_{B,t_i}^{(N)} = V_{t_i}^{(N)} - V_{S,t_i}^{(N)}. \quad (2.7)$$

Between time periods t_i and t_{i+1} , the portfolio composition remains unchanged. This gives rise to non-zero tracking errors $\epsilon_{t_i}^{(N)}$:

$$\epsilon_{t_i}^{(N)} \equiv H(t_i, P_{t_i}) - V_{t_i}^{(N)}. \quad (2.8)$$

The value of the replicating portfolio at time $t = 1$ is denoted by $V_1^{(N)}$ and the end-of-period tracking error is denoted by $\epsilon_1^{(N)}$.

The sequence of tracking errors contains a great deal of information about the approximation errors of implementing a continuous-time hedging strategy in discrete time, and in Sections 2.3.2 and 2.6 we provide a complete characterization of the continuous-time limiting distribution of $\epsilon_1^{(N)}$ and $\{\epsilon_{t_i}\}$. However, because tracking errors also contain noise, we also investigate the properties of the root-mean-squared-error (RMSE) of the end-of-period

tracking error ϵ_1 (see Hutchinson, Lo, and Poggio [1994] for other alternatives):

$$\text{RMSE}^{(N)} = \sqrt{\mathbf{E}_0[(\epsilon_1^{(N)})^2]} . \quad (2.9)$$

where $\mathbf{E}_0[\cdot]$ denotes the conditional expectation, conditional on information available at time $t=0$. Whenever exact replication of the derivative's payoff is impossible, $\text{RMSE}^{(N)}$ is positive.

Of course, root-mean-squared-error is only one of many possible summary statistics of the tracking error. A more general specification is the expected loss of the tracking error

$$\mathbf{E}_0 \left[U(\epsilon_1^{(N)}) \right]$$

where $U(\cdot)$ is a general *loss function*, and we consider this case explicitly in Section 2.6.4.

2.3.2 Asymptotic Behavior of the Tracking Error and RMSE

We characterize analytically the asymptotic behavior of the tracking error and RMSE by appealing to continuous-record asymptotics, i.e., by letting the number of trading periods N increase without bound while holding the time span fixed. This characterization provides several important insights into the behavior of the tracking error of general European derivative securities that previous studies have only hinted at indirectly (and only for simple put and call options).⁴ A by product of this characterization is a useful definition for the temporal granularity of a continuous-time stochastic process (relative to a specific derivative security).

We begin with the case of smooth payoff functions $F(P_1)$:

⁴See, for example, Boyle and Emanuel (1980) Hutchinson, Lo, and Poggio (1994), Leland (1985), and Toft (1996).

Theorem 1 *Let the derivative's payoff function $F(x)$ in (2.3) be six times continuously differentiable and all of its derivatives be bounded, and suppose there exists a positive constant K such that functions $\mu(\tau, x)$ and $\sigma(\tau, x)$ in (2.1) satisfy*

$$\left| \frac{\partial^{\beta+\gamma}}{\partial \tau^\beta \partial x^\gamma} \mu(\tau, x) \right| + \left| \frac{\partial^{\beta+\gamma}}{\partial \tau^\beta \partial x^\gamma} \sigma(\tau, x) \right| + \left| \frac{\partial^\alpha}{\partial x^\alpha} (x\sigma(\tau, x)) \right| \leq K \quad (2.10)$$

where $(\tau, x) \in [0, 1] \times [0, \infty)$, $1 \leq \alpha \leq 6$, $0 \leq \beta \leq 1$, $0 \leq \gamma \leq 3$, and all partial derivatives are continuous. Then under Assumptions (A5)–(A8):

(a) *The RMSE of the discrete-time delta-hedging strategy (2.7) satisfies*

$$\text{RMSE}^{(N)} = \mathbf{O}\left(\frac{1}{\sqrt{N}}\right). \quad (2.11)$$

(b) *The normalized tracking error satisfies:*

$$\sqrt{N} \epsilon_1^{(N)} \Rightarrow G$$

where

$$G \equiv \frac{1}{\sqrt{2}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW_t', \quad (2.12)$$

W_t' is a Wiener process independent of W_t , and “ \Rightarrow ” denotes convergence in distribution.

(c) *The RMSE of the discrete-time delta-hedging strategy (2.7) satisfies*

$$\text{RMSE}^{(N)} = \frac{g}{\sqrt{N}} + \mathbf{O}\left(\frac{1}{N}\right) \quad (2.13)$$

where

$$g = \sqrt{\mathbf{E}_0[\mathcal{R}]}, \quad (2.14)$$

$$\mathcal{R} = \frac{1}{2} \int_0^1 \left(\sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} \right)^2 dt. \quad (2.15)$$

Proof: See the Appendix.

Theorem 1 shows that the tracking error is asymptotically equal in distribution to G/\sqrt{N} (up to $\mathbf{O}(N^{-1})$ terms), where G is a random variable given by (2.12). The expected value of G is zero by the martingale property of stochastic integrals. Moreover, the independence of the Wiener processes W'_t and W_t implies that the asymptotic distribution of the normalized tracking error is symmetric, i.e., in the limit of frequent trading, positive values of the normalized tracking error are just as likely as negative values of the same magnitude.

This result might seem somewhat counterintuitive at first, especially in light of Boyle and Emanuel's (1980) finding that in the Black-Scholes framework the distribution of the local tracking error over a short trading interval is significantly skewed. However, Theorem 1(b) describes the asymptotic distribution of the tracking error over the *entire life* of the derivative, not over short intervals. Such an aggregation of local errors leads to a symmetric asymptotic distribution, just as a normalized sum of random variables will have a Gaussian distribution asymptotically under certain conditions, e.g., the conditions for a functional central limit theorem to hold.

Note that Theorem 1 applies to a wide class of diffusion processes (2.1) and to a variety of derivative payoff functions $F(P_1)$. In particular, it holds when the stock price follows a diffusion process with constant coefficients, as in Black and Scholes (1973).⁵ However,

⁵For the Black-Scholes case, the formula for the RMSE (2.14)–(2.15) was first derived by Grannan and Swindle (1996). Our results provide a more complete characterization of the tracking error in their framework—we derive the asymptotic distribution—and our analysis applies to more general trading strategies than theirs, e.g., they consider strategies obtained by deterministic time deformations; our framework can accommodate deterministic and stochastic time deformations.

the requirement that the payoff function $F(P_1)$ is smooth—six times differentiable with bounded derivatives—is violated by the most common derivatives of all: simple puts and calls. In the next theorem, we extend our results to cover this most basic set of payoff functions.

Theorem 2 *Let the payoff function $F(P_1)$ be continuous and piecewise linear, and suppose (2.10) holds. In addition, let*

$$\left| x^2 \frac{\partial^\alpha \sigma(\tau, x)}{\partial x^\alpha} \right| \leq K_2 \quad (2.16)$$

for $(\tau, x) \in [0, 1] \times [0, \infty)$, $2 \leq \alpha \leq 5$, and some positive constant K_2 . Then under Assumptions (A5)–(A8):

(a) *The RMSE of the discrete-time delta-hedging strategy (2.7) satisfies*

$$\text{RMSE}^{(N)} = \frac{g}{\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right)$$

where g is given by (2.14)–(2.15).

(b) *The normalized tracking error satisfies*

$$\sqrt{N} \epsilon_1^{(N)} \Rightarrow \frac{1}{\sqrt{2}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW_t' \quad (2.17)$$

where W_t' is a Wiener process independent of W_t .⁶

Proof: See the Appendix.

⁶It is easy to show, using Hölder's inequality, that $g < \infty$ and this implies that the stochastic integral in (2.17) is well defined. See Bertsimas, Kogan, and Lo (1998) for further details.

By imposing an additional smoothness condition (2.16) on the diffusion coefficient $\sigma(\tau, x)$, Theorem 2 assures us that the conclusions of Theorem 1 also hold for the most common types of derivatives, those with piecewise linear payoff functions.

Theorems 1 and 2 allows us to define the coefficient of *temporal granularity* g for any combination of continuous-time process $\{P_t\}$ and derivative payoff function $F(P_1)$ —it is the constant associated with the leading term of the RMSE’s continuous-record asymptotic expansion:

$$g \equiv \sqrt{\frac{1}{2} \mathbf{E}_0 \left[\int_0^1 \left(\sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} \right)^2 dt \right]} \quad (2.18)$$

where $H(t, P_t)$ satisfies (2.2) and (2.3).

2.3.3 Interpretation of Granularity

The interpretation for temporal granularity is clear: it is a measure of the approximation errors that arise from implementing a continuous-time delta-hedging strategy in discrete time. A derivative-pricing model—recall that this is comprised of a payoff function $F(P_1)$ and a continuous-time stochastic process for P_t —with high granularity requires a larger number of trading periods to achieve the same level of tracking error as a derivative-pricing model with low granularity. In the former case, time is “grainier”, calling for more frequent hedging activity than the latter case. More formally, according to Theorems 1 and 2, to a first-order approximation the RMSE of an N -trade delta-hedging strategy is g/\sqrt{N} . Therefore, if we desire the RMSE to be within some small value δ , we require

$$N \approx \frac{g^2}{\delta^2}$$

trades in the unit interval. For a fixed error δ , the number of trades needed to reduce the RMSE to within δ grows quadratically with granularity. If one derivative-pricing model has twice the granularity of another, it would require four times as many delta-hedging transactions to achieve the same RMSE tracking error.

From (2.18) it is clear that granularity depends on the derivative-pricing formula $H(t, P_t)$ and the price dynamics P_t in natural ways. Equation (2.18) formalizes the intuition that derivatives with higher volatility and higher “gamma” risk (large second derivative with respect to stock price) are more difficult to hedge, since these cases imply larger values for the integrand in (2.18). Accordingly, derivatives on less volatile stocks are easier to hedge. Consider a stock price process which is almost deterministic, i.e., $\sigma(t, P_t)$ is very small. This implies a very small value for g , hence derivatives on such a stock can be replicated almost perfectly, even if continuous trading is not feasible. Alternatively, such derivatives require relatively few rebalancing periods N to maintain small tracking errors.

Also, a derivative with a particularly simple payoff function should be easier to hedge than derivatives on the same stock with more complicated payoffs. For example, consider a derivative with the payoff function $F(P_1) = P_1$. This derivative is identical to the underlying stock, and can always be replicated perfectly by buying a unit of the underlying stock at time $t=0$ and holding it until expiration. The tracking error for this derivative is always equal to zero, no matter how volatile the underlying stock is. This intuition is made precise by Theorem 1, which describes exactly how the error depends on the properties of the stock price process and the payoff function of the derivative: it is determined by the behavior of the integral \mathcal{R} , which tends to be large when stock prices “spend more time” in regions of the domain that imply high volatility and high convexity or gamma of the derivative.

We will investigate the sensitivity of g to the specification of the stock price process in

Sections 2.4 and 2.5.

2.4 Applications

To develop further intuition for our measure of temporal granularity, in this section we derive closed-form expressions for g in two important special cases: the Black-Scholes option pricing model with geometric Brownian motion, and the Black-Scholes model with a mean-reverting (Ornstein-Uhlenbeck) process.

2.4.1 Granularity of Geometric Brownian Motion

Suppose that stock price dynamics are given by:

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t. \quad (2.19)$$

where μ and σ are constants. Under this assumption we obtain the following explicit characterization of the granularity g .

Theorem 3 *Under Assumptions (A5)-(A8), stock price dynamics (2.19), and the payoff function of simple call and put options, the granularity g in (2.13) is given by*

$$g = K\sigma \left(\int_0^1 \frac{\exp \left[-\frac{[\mu t + \ln(\frac{P_0}{K}) - \sigma^2/2]^2}{\sigma^2(1+t)} \right]}{4\pi\sqrt{1-t^2}} dt \right)^{\frac{1}{2}} \quad (2.20)$$

where K is the option's strike price.

Proof: See the Appendix.

It is easy to see that $g = 0$ if $\sigma = 0$ and g increases with σ in the neighborhood of zero.

When σ increases without bound, the granularity g decays to zero, which means that it has

at least one local maximum as a function of σ . The granularity g also decays to zero when P_0/K approaches zero or infinity. In the important special case of $\mu = 0$, we conclude by direct computation that g is a unimodal function of P_0/K , that achieves its maximum at $P_0/K = \exp(\sigma^2/2)$.

The fact that granularity is not monotone increasing in σ may seem counterintuitive at first—after all, how can delta-hedging errors become smaller for larger values of σ ? The intuition follows from the fact that at small levels of σ , an increase in σ leads to larger granularity because there is a greater chance that the stock price will fluctuate around regions of high gamma (where $\partial^2 H(t, P_t)/\partial P_t^2$ is large, i.e., near the money), leading to greater tracking errors. However, at very high levels of σ , prices fluctuate so wildly that an increase in σ will decrease the probability that the stock price stays in regions of high gamma for very long—in these extreme cases, the payoff function “looks” approximately linear hence granularity becomes a decreasing function of σ .

Also, we show below that g is not very sensitive to changes in μ when σ is sufficiently large. This implies that, for an empirically relevant range of parameter values, g , as a function of the initial stock price, achieves its maximum close to the strike price, i.e., at $P_0/K \approx 1$. These observations are consistent with the behavior of the tracking error for finite values of N that we see in the Monte Carlo simulations of Section 2.5.

When stock prices follow a geometric Brownian motion, expressions similar to (2.20) can be obtained for derivatives other than simple puts and calls. For example, for a straddle, consisting of one put and one call option with the same strike price K , the constant g is twice as large as for the put or call option alone.

2.4.2 Granularity of a Mean-Reverting Process

Let $p_t \equiv \ln(P_t)$ and suppose

$$dp_t = \left(-\gamma(p_t - (\alpha + \beta t)) + \beta \right) dt + \sigma dW_t \quad (2.21)$$

where $\beta = \mu - \sigma^2/2$ and α is a constant. This is an Ornstein-Uhlenbeck process with a linear time trend, and the solution of (2.21) is given by

$$p_t = (p_0 - \alpha)e^{-\gamma t} + (\alpha + \beta t) + \sigma \int_0^t e^{-\gamma(t-s)} dW_s. \quad (2.22)$$

Under these price dynamics, we have:

Theorem 4 *Under Assumptions (A5)–(A8), stock price dynamics (2.21), and the payoff function of simple call and put options, the granularity g in (2.13) is given by*

$$g = K\sigma \left(\int_0^1 \frac{\sqrt{\gamma} \exp \left[\frac{-\gamma[\alpha + \mu t + (\ln(\frac{P_0}{K}) - \alpha) \exp(-\gamma t) - \sigma^2/2]^2}{\sigma^2[\gamma(1-t) + 1 - \exp(-2\gamma t)]} \right]}{4\pi\sqrt{1-t}\sqrt{\gamma(1-t) + 1 - \exp(-2\gamma t)}} dt \right)^{\frac{1}{2}} \quad (2.23)$$

where K is the option's strike price.

Proof: See Bertsimas, Kogan, and Lo (1998).

Expression (2.23) is a direct generalization of (2.20): when the mean-reversion parameter γ is set to zero, the process (2.21) becomes a geometric Brownian motion and (2.23) reduces to (2.20). Theorem 4 has some interesting qualitative implications for the behavior of the tracking error in presence of mean-reversion. We will discuss them in detail in the next section.

2.5 Monte Carlo Analysis

Since our analysis of granularity is based entirely on continuous-record asymptotics, we must check the quality of these approximations by performing Monte Carlo simulation experiments for various values of N . The results of these Monte Carlo simulations are reported in Section 2.5.1. We also use Monte Carlo simulations to explore the qualitative behavior of the RMSE for various parameter values of the stock price process, and these simulations are reported in Section 2.5.2.

2.5.1 Accuracy of the Asymptotics

We begin by investigating the distribution of the tracking error $\epsilon^{(N)}$ for various values of N . We do this by simulating the hedging strategy of Section 2.3.2 for call and put options,⁷ assuming that price dynamics are given by a geometric Brownian motion (2.19).⁸ We set the parameters of the stock price process to $\mu = 0.1$, $\sigma = 0.3$, $P_0 = 1.0$ and let the strike price be $K = 1$. We consider $N = 10, 20, 50, 100$, and simulate the hedging process 250,000 times for each value of N .

Figure 2-1a shows the empirical probability density function (PDF) of $\epsilon_1^{(N)}$ for each N . As expected, the distribution of the tracking error becomes tighter as the trading frequency increases. It is also apparent that the tracking error can be significant even for $N = 100$. Figure 2-1b contains the empirical PDFs of the normalized tracking error, $\sqrt{N}\epsilon_1^{(N)}$, for the same values of N . These PDFs are compared to the PDF of the asymptotic distribution

⁷According to Theorem 1, the asymptotic expressions for the tracking error and the RMSE are the same for put and call options since these options have the same second partial derivative of the option price with respect to the current stock price. Moreover, it is easy to verify, using the put-call parity relation, that these options give rise to identical tracking errors.

⁸When the stock price process P_t follows a geometric Brownian motion, the stock price at time t_{i+1} is distributed (conditional on the stock price at time t_i) as $P_{t_i} \exp((\mu - \sigma^2/2)\Delta t + \sigma\sqrt{\Delta t}\eta)$, where $\eta \sim \mathcal{N}(0, 1)$. We use this relation to simulate the delta-hedging strategy.

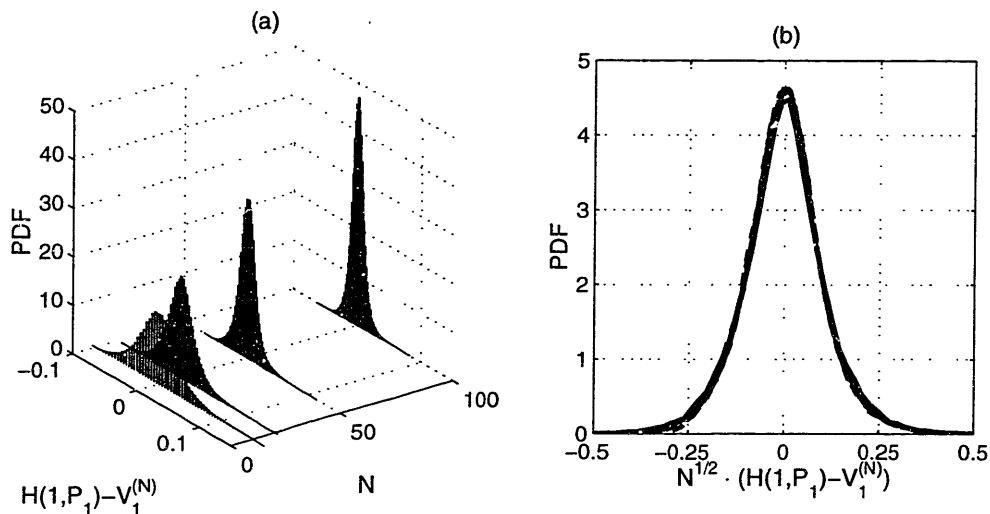


Figure 2-1: Empirical probability density functions of (a) the tracking error and (b) the normalized tracking error (dashed line), are plotted for different values of the trading frequency N . Figure (b) also shows the empirical probability density function of the asymptotic distribution (2.17) (solid line). The stock price process is given by (2.19) with parameters $\mu = 0.1$, $\sigma = 0.3$, $P_0 = 1.0$. The option is a European call (put) option with strike price $K = 1$.

(2.17), which is estimated by approximating the integral in (2.17) using a first-order Euler scheme. The functions in Figure 2-1b are practically identical and indistinguishable, which suggests that the asymptotic expression for the distribution of $\sqrt{N}\epsilon_1^{(N)}$ in Theorem 1(b) is an excellent approximation to the finite-sample PDF for values of N as small as 10.

To evaluate the accuracy of the asymptotic expression g/\sqrt{N} for finite values of N , we compare g/\sqrt{N} to the actual RMSE from Monte Carlo simulations of the delta-hedging strategy of Section 2.3.2. Specifically, we simulate the delta-hedging strategy for a set of European put and call options with strike price $K = 1$ under geometric Brownian motion (2.19) with different sets of parameter values for (σ, μ, P_0) . The tracking error is tabulated as a function of these parameters and the results are summarized in Tables 2.1, 2.2, and 2.3.

Tables 1-3 show that g/\sqrt{N} is an excellent approximation to the RMSE across a wide

<i>Parameters</i>			<i>Call Option</i>				<i>Put Option</i>	
N	P_0	$gN^{-1/2}$	$H(0, P_0)$	$\frac{\text{RMSE}^{(N)}}{H(0, P_0)}$	$\text{RMSE}^{(N)}$	RE	$H(0, P_0)$	$\frac{\text{RMSE}^{(N)}}{H(0, P_0)}$
10	0.50	.0078	7e-4	9.64	.0071	8.9 %	.501	.014
20	0.50	.0055	7e-4	6.88	.0052	7.9 %	.501	.010
50	0.50	.0035	7e-4	4.43	.0033	5.9 %	.501	.007
100	0.50	.0025	7e-4	3.22	.0024	3.1 %	.501	.005
10	0.75	.0259	.023	1.08	.0248	3.8 %	.273	.091
20	0.75	.0183	.023	.760	.0177	2.9 %	.273	.065
50	0.75	.0116	.023	.490	.0113	2.6 %	.273	.041
100	0.75	.0082	.023	.345	.0082	2.3 %	.273	.029
10	1.00	.0334	.119	.269	.0327	4.1 %	.119	.269
20	1.00	.0236	.119	.192	.0227	3.4 %	.119	.192
50	1.00	.0149	.119	.122	.0145	2.3 %	.119	.122
100	1.00	.0106	.119	.087	.0104	1.9 %	.119	.087
10	1.25	.0275	.294	.088	.0263	5.8 %	.044	.588
20	1.25	.0194	.294	.064	.0187	3.9 %	.044	.423
50	1.25	.0123	.294	.041	.0120	2.7 %	.044	.271
100	1.25	.0087	.294	.029	.0087	1.8 %	.044	.195
10	1.50	.0181	.515	.033	.0169	7.7 %	.015	1.13
20	1.50	.0128	.515	.024	.0122	5.3 %	.015	.816
50	1.50	.0081	.515	.015	.0076	2.9 %	.015	.528
100	1.50	.0057	.515	.011	.0056	3.0 %	.015	.373

Table 2.1: The sensitivity of the RMSE as a function of the initial price P_0 . The RMSE is estimated using Monte Carlo simulation. Options are European calls and puts with strike price $K = 1$. 250,000 simulations are performed for every set of parameter values. The stock price follows a geometric Brownian motion (2.19). The drift and diffusion coefficients of the stock price process are $\mu = 0.1$ and $\sigma = 0.3$. $\text{RMSE}^{(N)}$ is compared to the asymptotic approximation $gN^{-1/2}$ in (2.13)–(2.20). The relative error (RE) of the asymptotic approximation is defined as $|gN^{-1/2} - \text{RMSE}^{(N)}|/\text{RMSE}^{(N)} \times 100\%$.

<i>Parameters</i>			<i>Call (Put) Option</i>			
N	σ	$gN^{-1/2}$	$H(0, P_0)$	$\frac{\text{RMSE}^{(N)}}{H(0, P_0)}$	$\text{RMSE}^{(N)}$	RE
10	0.3	.0334	.119	.269	.0327	4.1 %
20	0.3	.0236	.119	.192	.0227	3.4 %
50	0.3	.0149	.119	.122	.0145	2.3 %
100	0.3	.0106	.119	.087	.0104	1.9 %
10	0.2	.0219	.080	.266	.0212	3.4 %
20	0.2	.0155	.080	.189	.0151	3.0 %
50	0.2	.0098	.080	.121	.0096	2.1 %
100	0.2	.0069	.080	.086	.0068	1.7 %
10	0.1	.0100	.040	.255	.0102	1.6 %
20	0.1	.0071	.040	.177	.0071	.04 %
50	0.1	.0045	.040	.111	.0044	1.1 %
100	0.1	.0032	.040	.078	.0031	0.9 %

Table 2.2: The sensitivity of the RMSE as a function of volatility σ . The RMSE is estimated using Monte Carlo simulation. Options are European calls and puts with strike price $K = 1$. 250,000 simulations are performed for every set of parameter values. The stock price follows a geometric Brownian motion (2.19). The drift coefficient of the stock price process is $\mu = 0.1$, and the initial stock price is $P_0 = 1.0$. $\text{RMSE}^{(N)}$ is compared to the asymptotic approximation $gN^{-1/2}$ in (2.13)–(2.20). The relative error (RE) of the asymptotic approximation is defined as $|gN^{-1/2} - \text{RMSE}^{(N)}|/\text{RMSE}^{(N)} \times 100\%$.

<i>Parameters</i>		<i>Call (Put) Option</i>			
μ	$gN^{-1/2}$	$H(0, P_0)$	$\frac{\text{RMSE}^{(N)}}{H(0, P_0)}$	RMSE ^(N)	RE
0.0	.0235	.119	.189	.0226	4.3 %
0.1	.0236	.119	.192	.0229	3.4 %
0.2	.0230	.119	.190	.0226	1.7 %
0.3	.0218	.119	.184	.0220	.96 %

Table 2.3: The sensitivity of the RMSE as a function of the drift μ . The RMSE is estimated using Monte Carlo simulation. Options are European calls and puts with strike price $K = 1$. 250,000 simulations are performed for every set of parameter values. The stock price follows a geometric Brownian motion (2.19). The diffusion coefficient of the stock price process is $\sigma = 0.3$, and the initial stock price is $P_0 = 1.0$, the number of trading periods is $N = 20$. $\text{RMSE}^{(N)}$ is compared to the asymptotic approximation $gN^{-1/2}$ in (2.13)–(2.20). The relative error (RE) of the asymptotic approximation is defined as $|gN^{-1/2} - \text{RMSE}^{(N)}|/\text{RMSE}^{(N)} \times 100\%$.

range of parameter values for (μ, σ, P_0) , even for as few as $N = 10$ delta-hedging periods.

2.5.2 Qualitative Behavior of the RMSE

The Monte Carlo simulations of Section 2.5.1 show that RMSE increases with the diffusion coefficient σ in an empirically relevant range of parameter values (see Table 2.2), and that the RMSE is not very sensitive to the drift rate μ of the stock price process when σ is sufficiently large (see Table 2.3). These properties are illustrated in Figures 2-2a and 2-3. In Figure 2-2a the logarithm of RMSE is plotted against the logarithm of trading periods N for $\sigma = 0.1, 0.2, 0.3$ —as σ increases, the locus of points shifts upward. Figure 2-3 shows that gra thout
bound.

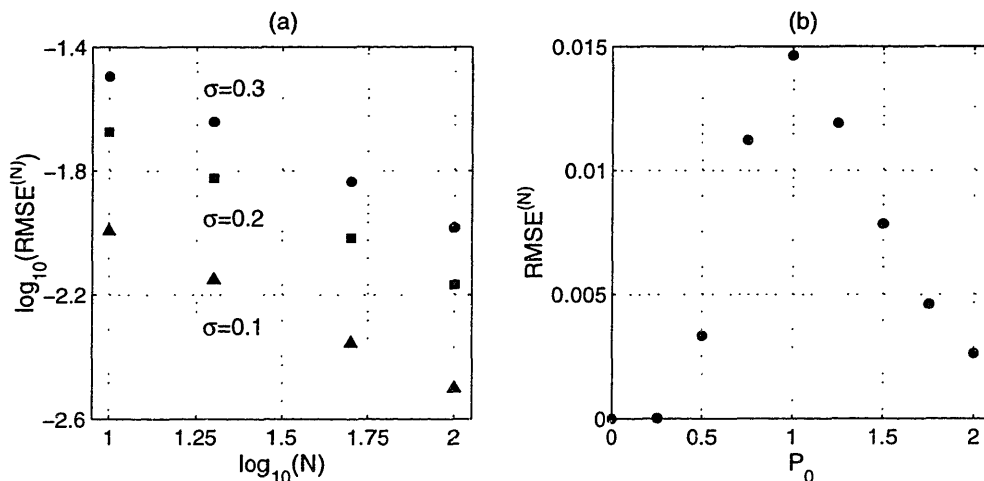


Figure 2-2: (a) The logarithm of the root-mean-squared error $\log_{10}(\text{RMSE}^{(N)})$ is plotted as a function of the logarithm of the trading frequency $\log_{10}(N)$. The option is a European call (put) option with the strike price $K = 1$. The stock price process is given by (2.19) with parameters $\mu = 0.1$, $P_0 = 1.0$. The diffusion coefficient of the stock price process takes values $\sigma = 0.3$ (x's), $\sigma = 0.2$ (o's) and $\sigma = 0.1$ (+s). (b) The root-mean-squared error $\text{RMSE}^{(N)}$ is plotted as a function of the initial stock price P_0 . The option is a European put option with the strike price $K = 1$. Parameters of the stock price process are $\mu = 0.1$, $\sigma = 0.3$.

Figure 2-2b plots the RMSE as a function of the initial stock price P_0 . RMSE is a

unimodal function of P_0/K (recall that the strike price has been normalized to $K = 1$ in all our calculations), achieving its maximum around 1 and decaying to zero as P_0/K approaches zero or infinity (see Table 2.1). This confirms the common intuition that close-to-the-money options are the most difficult to hedge—they exhibit the largest RMSE.

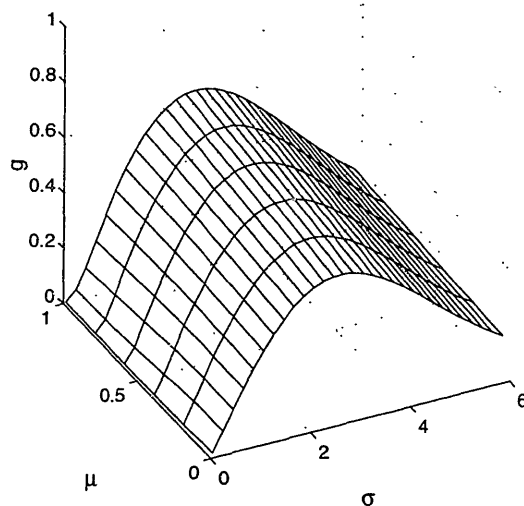


Figure 2-3: The granularity g is plotted as a function of σ and μ . The option is a European call (put) option with strike price $K = 1$. The stock price process is geometric Brownian motion and initial stock price $P_0 = 1$.

Finally, the relative importance of the RMSE can be measured by the ratio of the RMSE to the option price: $\text{RMSE}^{(N)}/H(0, P_0)$. This quantity is the root-mean-squared error per dollar invested in the option. Table 2.1 shows that this ratio is highest for out-of-the-money options, despite the fact that the RMSE is highest for close-to-the-money options. This is due to the fact that the option price decreases faster than the RMSE as the stock moves away from the strike.

Now consider the case of mean-reverting stock price dynamics (2.21). Recall that under

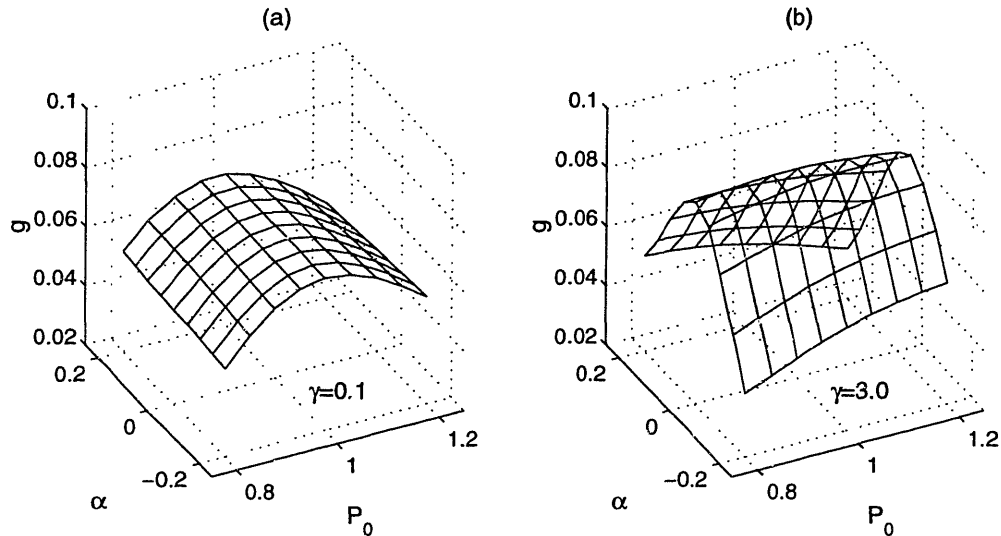


Figure 2-4: Granularity g is plotted as a function of P_0 and α . The option is a European call (put) option with the strike price $K = 1$. Parameters of the stock price process are $\sigma = 0.2$, $\mu = 0.05$. The stock price process is given by (2.22). Mean-reversion parameter γ takes two values: (a) $\gamma = 0.1$ and (b) $\gamma = 3.0$.

these dynamics, the Black-Scholes formula still holds.⁹ Nevertheless, the behavior of granularity and RMSE is quite different in this case. Figure 2-4 plots the granularity g of call and put options for the Ornstein-Uhlenbeck process (2.21) as a function of α and P_0 . Figure 2-4a assumes a value of 0.1 for the mean-reversion parameter γ and Figure 2-4b assumes a value of 3.0. It is clear from these two plots that the degree of mean reversion γ has an enormous impact on granularity. When γ is small, Figure 2-4a shows that the RMSE is highest when P_0 is close to the strike price and is not sensitive to α . But when γ is large, Figure 2-4b suggests that the RMSE is highest when $\exp(\alpha)$ is close to the strike price and is not sensitive to P_0 .

The influence of γ on the granularity can be understood by recalling that granularity is closely related to the option's gamma (see Section 2.3.3). When γ is small, the stock price is more likely to spend time in the neighborhood of the strike price—the region with the

⁹However, the numerical value for σ may be different than that of a geometric Brownian motion because the presence of mean-reversion can affect conditional volatility, holding unconditional volatility fixed. See Lo and Wang (1995) for further discussion.

highest “gamma” or $\partial^2 H(t, P_t)/\partial P_t^2$ —when P_0 is close to K . However, when γ is large, the stock price is more likely to spend time in the neighborhood of $\exp(\alpha)$, thus g is highest when $\exp(\alpha)$ is close to K .

2.6 Extensions and Generalizations

The analysis of Section 2.3 can be extended in a number of directions, and we briefly outline four of the most important of these extensions here. In Section 2.6.1, we show that the normalized tracking error converges in a much stronger sense than simply in distribution, and that this stronger “sample-path” notion of convergence—called, ironically, “weak” convergence—can be used to analyze the tracking error of American-style derivative securities. In Section 2.6.2 we characterize the asymptotic joint distributions of the normalized tracking error and asset prices, a particularly important extension for investigating the tracking error of delta hedging a portfolio of derivatives. In Section 2.6.3, we provide another characterization of the tracking error, one that relies on PDE’s, that offers important computational advantages. And in Section 2.6.4, we consider alternatives to mean-squared-error loss functions and show that for quite general loss functions, the behavior of the expected loss of the tracking error is characterized by the same stochastic integral (2.17) as in the mean-squared-error case.

2.6.1 Sample-Path Properties of Tracking Errors

Recall that the normalized tracking error *process* is defined as:

$$\sqrt{N} \epsilon_t^{(N)} = \sqrt{N} (H(t, P_t) - V_t^{(N)}) \quad , \quad t \in [0, 1] .$$

It can be shown that $\sqrt{N}\epsilon_t^{(N)}$ converges weakly to the stochastic process G_t , characterized by the stochastic integral in (2.12) as a function of its upper limit:¹⁰

$$G_t = \frac{1}{\sqrt{2}} \int_0^t \sigma^2(t, P_s) P_s^2 \frac{\partial^2 H(s, P_s)}{\partial P_s^2} dW'_s.$$

This stronger notion of convergence yields stronger versions of Theorem 1 and 2 that can be used to analyze a number of sample-path properties of the tracking error by appealing to the Continuous Mapping Theorem (see Billingsley [1968]). This well-known result shows that the asymptotic distribution of any continuous functional $\xi(\cdot)$ of the normalized tracking error is given by $\xi(G_t)$. For example, the maximum of the normalized tracking error over the entire life of the derivative security, $\max_t \sqrt{N}\epsilon_t^{(N)}$, is distributed as $\max_t G_t$ asymptotically.

These results can be applied to the normalized tracking errors of American-style derivatives in a straightforward manner. Such derivatives differ from European derivatives in one respect: they can be exercised prematurely. Therefore, the valuation of these derivatives consists of both computing the derivative price function $H(t, P_t)$ and the optimal exercise schedule, which can be represented as a stopping time τ . Then the tracking error at the moment when the derivative is exercised behaves asymptotically as G_τ/\sqrt{N} .¹¹ The tracking error, conditional on the derivative not being exercised prematurely, is distributed asymptotically as $(G_1/\sqrt{N} | \tau = 1)$.

¹⁰The proof of this result consists of two steps. The first step is to establish that the sequence of measures induced by $\sqrt{N}\epsilon_t^{(N)}$ is tight (relatively compact). This can be done by verifying local inequalities for the moments of processes $\sqrt{N}\epsilon_t^{(N)}$ using the machinery developed in the proof of Theorem 2 (we must use Burkholder's inequality instead of the isometric property and Hölder's inequality instead of Schwarz's inequality throughout—see Bertsimas, Kogan, and Lo [1998] for further details). The second step is to characterize the limiting process. Such a characterization follows from the proof in Appendix 2.8.2 and the fact that the results in Duffie and Protter (1992) guarantee weak convergence of stochastic processes, not just convergence of their one-dimensional marginal distributions.

¹¹Some technical regularity conditions, e.g., the smoothness of the exercise boundary, are required to ensure convergence. See, for example, Kushner and Dupuis (1992).

2.6.2 Joint Distributions of Tracking Errors and Prices

Theorems 1 and 2 provide a complete characterization of the tracking error and RMSE for individual derivatives, but what is often of more practical interest is the behavior of a portfolio of derivatives. Delta-hedging a portfolio of derivatives is typically easier because of the effects of diversification—as long as tracking errors are not perfectly correlated across derivatives, the portfolio tracking error will be less volatile than the tracking error of individual derivatives.

To address portfolio issues, we require the joint distribution of tracking errors for multiple stocks, as well as the joint distribution of tracking errors and prices. Consider another stock with price $P_t^{(2)}$ governed by the diffusion equation

$$\frac{dP_t^{(2)}}{P_t^{(2)}} = \mu^{(2)}(t, P_t^{(2)}) dt + \sigma^{(2)}(t, P_t^{(2)}) dW_t^{(2)} \quad (2.24)$$

where $W_t^{(2)}$ can be correlated with W_t . According to the proof of Theorem 1(b) (see Appendix 2.8.2), since the random variables $(W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)$ and $W_{t_{i+1}}^{(2)} - W_{t_i}^{(2)}$ are uncorrelated,¹² the Wiener processes W_t' and $W_t^{(2)}$ are independent. Therefore, as N increases without bound the pair of random variables $(\sqrt{N} \epsilon_1^{(N)}, P_1^{(2)})$ converges in distribution to:

$$\left(\sqrt{N} \epsilon_1^{(N)}, P_1^{(2)} \right) \Rightarrow \left(\frac{1}{\sqrt{2}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW_t', P_1^{(2)} \right) \quad (2.25)$$

where W_t' is independent of W_t and $W_t^{(2)}$.

An immediate corollary of this result is that the normalized tracking error is uncorrelated

¹²This follows from the fact that, for every pair of standard normal random variables X and Y with correlation ρ , $X = \rho Y + (1 + \rho^2)^{1/2} Z$, where Z is a standard normal random variable, independent of Y . Thus X and $Y^2 - 1$ are uncorrelated.

with any asset in the economy. This follows easily from (2.25) since, conditional on the realization of P_t and $P_t^{(2)}$, $t \in [0, 1]$, the normalized tracking error has zero expected value asymptotically. However, this does not imply that the asymptotic joint distribution of $(\sqrt{N} \epsilon_1^{(N)}, P_1^{(2)})$ does not depend on the correlation between W_t and $W_t^{(2)}$ —it does, since this correlation determines the joint distribution of P_t and $P_t^{(2)}$.

The above argument applies without change when the price of the second stock follows a diffusion process different from (2.24), and can also easily be extended to the case of multiple stocks.

To derive the joint distribution of the normalized tracking errors for multiple stocks, we consider the case of two stocks since the generalization to multiple stocks is obvious. Let W_t and $W_t^{(2)}$ have mutual variation $dW_t dW_t^{(2)} = \rho(t, P_t, P_t^{(2)}) dt$, where $\rho(\cdot)$ is a continuously differentiable function with bounded first-order partial derivatives. We have already established that the asymptotic distribution of the tracking error is characterized by the stochastic integral (2.12). To describe the asymptotic joint distribution of two normalized tracking errors, it is sufficient to find the mutual variation of the Wiener processes in the corresponding stochastic integrals. According to the proof of Theorem 1(b) (Appendix 2.8.2), this amounts to computing the expected value of the product

$$\left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right) \left((W_{t_{i+1}}^{(2)} - W_{t_i}^{(2)})^2 - (t_{i+1} - t_i) \right).$$

Using Itô's formula, it is easy to show that the expected value of the above expression is equal to

$$\mathbf{E}_0 \left[2\rho^2(t, P_{t_i}, P_{t_i}^{(2)}) \right] (\Delta t)^2 + \mathbf{O} \left((\Delta t)^{\frac{5}{2}} \right).$$

This implies that $\rho^2(t, P_t, P_t^{(2)})$ is the mutual variation of the two Wiener processes in the stochastic integrals (2.12) that describe the asymptotic distributions of the normalized tracking errors of the two stocks. Together with Theorem 1(b), this completely determines the asymptotic joint distribution of the two normalized tracking errors.¹³

Note that the correlation of two Wiener processes describing the asymptotic behavior of two normalized tracking errors is always nonnegative, regardless of the sign of the mutual variation of the original Wiener processes W_t and $W_t^{(2)}$. In particular, when two derivatives have convex price functions, this means that even if the returns on the two stocks are negatively correlated, the tracking errors resulting from delta hedging derivatives on these stocks are asymptotically positively correlated.

2.6.3 A PDE Characterization of the Tracking Error

It is possible to derive an alternative characterization of the tracking error using the intimate relationship between diffusion processes and PDE's. Although this may seem superfluous given the analytical results of Theorems 1 and 2, the numerical implementation of a PDE representation is often computationally more efficient.

To illustrate our approach, we begin with the RMSE. According to Theorem 1(c), the RMSE can be completely characterized asymptotically if g is known (see (2.14)). Using the Feynman-Kac representation of the solutions of PDE's (see Karatzas and Shreve [1991, Proposition 4.2.]), we conclude that $g^2 = u(0, P_0)$, where $u(t, x)$ solves the following:

$$\left[\frac{\partial}{\partial t} + \mu(t, x)x \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2(t, x)x^2 \frac{\partial^2}{\partial x^2} \right] u(t, x) + \frac{1}{2} \left(\sigma^2(t, x)x^2 \frac{\partial^2 H(t, x)}{\partial x^2} \right)^2 = 0 \quad (2.26)$$

¹³This result generalizes the findings of Boyle and Emanuel (1980).

$$u(1, x) = 0, \forall x. \quad (2.27)$$

The PDE (2.26)–(2.27) is of the same degree of difficulty as the fundamental PDE (2.2)–(2.3) that must be solved to obtain the derivative-pricing function $H(t, P_t)$. This new representation of the RMSE can be used to implement an efficient numerical procedure for calculating RMSE without resorting to Monte Carlo simulation.¹⁴

Summary measures of the tracking error with general loss functions can also be computed numerically along the same lines, using the Kolmogorov backward equation. The probability density function of the normalized tracking error $\sqrt{N}\epsilon_1^{(N)}$ can be determined numerically as a solution of the Kolmogorov forward equation (see, for example, Karatzas and Shreve [1991, pp. 368–369]).

2.6.4 Alternative Measures of the Tracking Error

As we observed in Section 2.3.2, the root-mean-squared error is only one of many possible summary measures of the tracking error. An obvious alternative is the L_p -norm:

$$\mathbf{E}_0 \left[\left| \epsilon_1^{(N)} \right|^p \right]^{\frac{1}{p}} \quad (2.28)$$

where p is chosen so that the expectation is finite (otherwise the measure will not be particularly informative). More generally, the tracking error can be summarized by

$$\mathbf{E}_0 \left[U(\epsilon_1^{(N)}) \right] \quad (2.29)$$

where $U(\cdot)$ is an arbitrary loss function.

¹⁴Results of some preliminary numerical experiments provide encouraging evidence of the practical value of this new representation.

Consider the set of measures (2.28) first and assume for simplicity that $p \in [1, 2]$. From (2.17), it follows that

$$\mathbf{E}_0 \left[\left| \epsilon_1^{(N)} \right|^p \right]^{\frac{1}{p}} \sim N^{-1/2} \mathbf{E}_0 \left[\left| \frac{1}{\sqrt{2}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW_t' \right|^p \right]^{\frac{1}{p}}. \quad (2.30)$$

hence the moments of the stochastic integral in (2.17) describe the asymptotic behavior of the moments of the tracking error. Conditional on the realization of $\{P_t\}$, $t \in [0, 1]$, the stochastic integral on the right side of (2.30) is normally distributed with zero mean and variance

$$\int_0^1 \left(\sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} \right)^2 dt$$

which follows from Hull and White (1987). The intuition is that, conditional on the realization of the integrand, the stochastic integral behaves as an integral of a deterministic function with respect to the Wiener process which is a normal random variable. Now let m_p denote an L_p -norm of the standard normal random variable.¹⁵ Then (2.30) can be rewritten as:

$$\mathbf{E}_0 \left[\left| \epsilon_1^{(N)} \right|^p \right]^{\frac{1}{p}} \sim \frac{m_p}{\sqrt{N}} \mathbf{E}_0 \left[\mathcal{R}^{\frac{p}{2}} \right]^{\frac{1}{p}} \quad (2.31)$$

where \mathcal{R} is given by (2.15).

As in the case of a quadratic loss function, \mathcal{R} plays a fundamental role here in describing the behavior of the tracking error. When $p = 2$, \mathcal{R} enters (2.31) linearly and closed-form expressions can be derived for special cases. However, even when $p \neq 2$, the qualitative

¹⁵If X is a standard normal random variable, then $m_p = \mathbf{E}[|X|^p]^{1/p}$.

impact of \mathcal{R} on the tracking error is the same as for $p = 2$ and our discussion of the qualitative behavior of the tracking error applies to this case as well.

For general loss functions $U(\cdot)$ that satisfy certain growth conditions and are sufficiently smooth near the origin, the delta-method can be applied and we obtain:

$$\mathbf{E}_0 \left[U \left(\epsilon_1^{(N)} \right) \right] \sim \frac{1}{N} \left| U''(0) \right| g^2 = \frac{1}{N} \left| U''(0) \right| \mathbf{E}_0[\mathcal{R}] . \quad (2.32)$$

When $U(\cdot)$ is not differentiable at 0, the delta method cannot be used. However, we can use the same strategy as in our analysis of L_p -norms to tackle this case. Suppose that $U(\cdot)$ is dominated by a quadratic function. Then

$$\mathbf{E}_0 \left[U \left(\epsilon_1^{(N)} \right) \right] \approx \mathbf{E}_0 \left[U \left(\frac{1}{\sqrt{2N}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW_t' \right) \right] . \quad (2.33)$$

Now let

$$m_U(x) = \mathbf{E}[U(x\eta)] \quad , \quad \eta \sim \mathcal{N}(0, 1) .$$

Then

$$\mathbf{E}_0 \left[U \left(\epsilon_1^{(N)} \right) \right] \approx \mathbf{E}_0 \left[m_U \left(\sqrt{\mathcal{R}/N} \right) \right] . \quad (2.34)$$

When the loss function $U(\cdot)$ is convex, $m_U(\cdot)$ is an increasing function (by second-order stochastic dominance). Therefore, the qualitative behavior of the measure (2.29) is also determined by \mathcal{R} and is the same as that of the RMSE.

2.7 Conclusions

We have argued that continuous-time models are meant to be approximations to physical phenomena, and as such, their approximation errors should be better understood. In the specific context of continuous-time models of derivative securities, we have quantified the approximation error through our definition of temporal granularity. The combination of a specific derivative security and a stochastic process for the underlying asset's price dynamics can be associated with a measure of how “grainy” the passage of time is. This measure is related to the ability to replicate the derivative security through a delta-hedging strategy implemented in discrete time. Time is said to be very granular if the replication strategy does not work well—in such cases, time is not continuous. If, however, the replication strategy is very effective, time is said to be very smooth or continuous.

Under the assumption of general Markov diffusion price dynamics, we show that that the tracking errors for derivatives with sufficiently smooth or continuous piecewise linear payoff functions behave asymptotically (in distribution) as G/\sqrt{N} . We characterize the distribution of the random variable G as a stochastic integral, and also obtain the joint distribution of G with prices of other assets and with other tracking errors. We demonstrate that the root-mean-squared error behaves asymptotically as g/\sqrt{N} , where the constant g is what we call the coefficient of *temporal granularity*. For two special cases—call or put options on geometric Brownian motion and on an Ornstein-Uhlenbeck process—we are able to evaluate the coefficient of granularity explicitly.

We also consider a number of extensions of our analysis, including an extension to alternative loss functions, a demonstration of the weak convergence of the tracking error process, a derivation of the joint distribution of tracking errors and prices, and an alternative characterization of the tracking error in terms of PDE's that can be used for efficient

numerical implementation.

Because these results depend so heavily on continuous-record asymptotics, we perform Monte Carlo simulations to check the quality of our asymptotics. For the case of European puts and calls with geometric Brownian motion price dynamics, our asymptotic approximations are excellent, providing extremely accurate inferences over the range of empirically relevant parameter values, even with a small number of trading periods.

Of course, our definition of granularity is not invariant to the derivative security, the underlying asset's price dynamics, and other variables. But we regard this as a positive feature of our approach, not a drawback. After all, any plausible definition of granularity must be a relative one, balancing the coarseness of changes in the time domain against the coarseness of changes in the "space" or price domain. Although the title of this paper suggests that time is the main focus of our analysis, it is really the relation between time and price that determines whether or not continuous-time models are good approximations to physical phenomena. It is our hope that the definition of granularity proposed in this paper is one useful way of tackling this very complex issue.

2.8 Appendix

The essence of these proofs involves the relation between the delta-hedging strategy and mean-square approximations of solutions of systems of stochastic differential equations described in Milstein (1974, 1987, 1995). Readers interested in additional details and intuition should consult these references directly. We present the proof of Theorem 1 only, and refer readers to Bertsimas, Kogan, and Lo (1998) for the others.

2.8.1 Proof of Theorem 1(a)

First we observe that the regularity conditions (2.10) imply the existence of a positive constant K_1 such that

$$\left| \frac{\partial^{\beta+\gamma}}{\partial \tau^\beta \partial x^\gamma} H(\tau, x) \right| \leq K_1 \quad (2.35)$$

for $(\tau, x) \in [0, 1] \times [0, \infty)$, $0 \leq \beta \leq 1$, $1 \leq \gamma \leq 4$, and all partial derivatives are continuous.¹⁶

Next, by Itô's formula,

$$\begin{aligned} H(1, P_1) = H(0, P_0) &+ \int_0^1 \left(\frac{\partial H(t, P_t)}{\partial t} + \frac{1}{2} \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} \right) dt + \\ &\int_0^1 \frac{\partial H(t, P_t)}{\partial P_t} dP_t . \end{aligned} \quad (2.36)$$

¹⁶Since the price of the derivative $H(\tau, x)$ is defined as a solution of (2.2), it is equal to the expectation of $F(P_1)$ with respect to the equivalent martingale measure (see Duffie [1996]), i.e.,

$$H(\tau, x) = \mathbf{E}_{(t=\tau, P_t^*=x)}[F(P_1^*)]$$

where

$$\frac{dP_t^*}{P_t^*} = \sigma(t, P_t^*) dW_t^* .$$

and W_t^* is a Brownian motion under the equivalent martingale measure. Equation (2.35) now follows from Friedman (1975; Theorems 5.4 and 5.5, p. 122). The same line of reasoning is followed in He (1989, p. 68). Of course, one could derive (2.35) using purely analytic methods, e.g. Friedman (1964; Theorem 10, p. 72, Theorem 11, p. 24; and Theorem 12, p. 25).

According to (2.2), the first integral on the right-hand side of (2.36) is equal to zero. Thus,

$$H(1, P_1) = H(0, P_0) + \int_0^1 \frac{\partial H(t, P_t)}{\partial P_t} dP_t \quad (2.37)$$

which implies that $H(t, P_t)$ can be characterized as a solution of the system of stochastic differential equations

$$\begin{cases} dX_t = \frac{\partial H(t, P_t)}{\partial P_t} \mu(t, P_t) P_t dt + \frac{\partial H(t, P_t)}{\partial P_t} \sigma(t, P_t) P_t dW_t, \\ dP_t = \mu(t, P_t) P_t dt + \sigma(t, P_t) P_t dW_t. \end{cases} \quad (2.38)$$

At the same time, $V_1^{(N)}$ is given by

$$V_1^{(N)} = H(0, P_0) + \sum_{i=0}^{N-1} \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} (P_{t_{i+1}} - P_{t_i}), \quad (2.39)$$

which can be interpreted as a solution of the following approximation scheme of (2.38) (as defined in Milstein [1987]):

$$\begin{cases} \bar{X}_{t_{i+1}} - X_{t_i} = \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} (P_{t_{i+1}} - P_{t_i}), \\ \bar{P}_{t_{i+1}} - P_{t_i} = P_{t_{i+1}} - P_{t_i}, \end{cases} \quad (2.40)$$

where \bar{X} and \bar{P} denote approximations to X and P , respectively. We now compare (2.40) to the Euler approximation scheme in Milstein (1995)

$$\left\{ \begin{array}{l} \bar{X}_{t_{i+1}} - X_{t_i} = \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} \mu(t_i, P_{t_i}) P_{t_i} (t_{i+1} - t_i) + \\ \qquad \qquad \qquad \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} \sigma(t_i, P_{t_i}) P_{t_i} (W_{t_{i+1}} - W_{t_i}), \\ \bar{P}_{t_{i+1}} - P_{t_i} = \mu(t_i, P_{t_i}) (t_{i+1} - t_i) + \\ \qquad \qquad \qquad \sigma(t_i, P_{t_i}) (W_{t_{i+1}} - W_{t_i}). \end{array} \right. \quad (2.41)$$

Regularity conditions (2.10) and (2.35) allow us to conclude (see Milstein [1995, Theorem 2.1]) that a one-step version of the approximation scheme (2.41) has order-of-accuracy 2 in expected deviation and order-of-accuracy 1 in mean-square deviation (see Milstein (1987), Milstein (1995) for definitions and discussion). It is easy to check that the approximation scheme (2.40) exhibits this same property. Milstein (1995, Theorem 1.1) relates the one-step order-of-accuracy of the approximation scheme to its order-of-accuracy on the whole interval (see also Milstein [1987]). We use this theorem to conclude that (2.40) has mean-square order-of-accuracy 1/2, i.e.,

$$\sqrt{\mathbf{E}_0 \left[(X(1, P_1) - \bar{X}(1, P_1))^2 \right]} = \mathbf{O} \left(\frac{1}{\sqrt{N}} \right). \quad (2.42)$$

We now recall that $X(t, P_t) = H(t, P_t)$ and $\bar{X}(1, P_1) = V_1^{(N)}$ and conclude that

$$\sqrt{\mathbf{E}_0 \left[(H(1, P_1) - V_1^{(N)})^2 \right]} = \mathbf{O} \left(\frac{1}{\sqrt{N}} \right). \quad (2.43)$$

which completes the proof. ■

2.8.2 Proof of Theorem 1(b)

We follow the same line of reasoning as in the proof of Theorem 1(a), but we use the Milstein approximation scheme for (2.38) instead of the Euler scheme:

$$\left\{ \begin{array}{l} \bar{X}_{t_{i+1}} - X_{t_i} = \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} \mu(t_i, P_{t_i}) P_{t_i} (t_{i+1} - t_i) + \\ \quad \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} \sigma(t_i, P_{t_i}) P_{t_i} (W_{t_{i+1}} - W_{t_i}) + \\ \quad \left(\frac{\partial^2 H(t, P_t)_{t=t_i}}{\partial P_t^2} \sigma(t_i, P_{t_i}) P_{t_i} + \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} \frac{\partial(\sigma(t, P_t) P_t)_{t=t_i}}{\partial P_t} \right) \times \\ \quad \frac{1}{2} \sigma(t_i, P_{t_i}) P_{t_i} \left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right), \\ \bar{P}_{t_{i+1}} - P_{t_i} = \mu(t_i, P_{t_i}) P_{t_i} (t_{i+1} - t_i) + \sigma(t_i, P_{t_i}) P_{t_i} (W_{t_{i+1}} - W_{t_i}) + \\ \quad \frac{1}{2} \sigma(t_i, P_{t_i}) P_{t_i} \frac{\partial(\sigma(t, P_t) P_t)_{t=t_i}}{\partial P_t} \left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right). \end{array} \right. \quad (2.44)$$

According to Milstein (1974) (see also Milstein [1995, Theorem 2.1]), this one-step scheme has order-of-accuracy 2 in expected deviation and 1.5 in mean-square deviation. It is easy to check by comparison that the scheme

$$\left\{ \begin{array}{l} \bar{X}_{t_{i+1}} - X_{t_i} = \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} (P_{t_{i+1}} - P_{t_i}) + \\ \quad \frac{1}{2} \sigma(t_i, P_{t_i})^2 P_{t_i}^2 \frac{\partial^2 H(t, P_t)_{t=t_i}}{\partial P_t^2} \left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right), \\ \bar{P}_{t_{i+1}} - P_{t_i} = P_{t_{i+1}} - P_{t_i} \end{array} \right. \quad (2.45)$$

has the same property. We now use Milstein (1995, Theorem 1.1) to conclude that

$$H(1, P_1) - V_1^{(N)} = \sum_{i=0}^{N-1} \frac{1}{2} \sigma(t_i, P_{t_i})^2 P_{t_i}^2 \frac{\partial^2 H(t, P_t)_{t=t_i}}{\partial P_t^2} \times \left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right) + \mathbf{O}\left(\frac{1}{N}\right), \quad (2.46)$$

where $f = \mathbf{O}\left(\frac{1}{N}\right)$ means that $\lim_{N \rightarrow \infty} N \sqrt{\mathbf{E}_{t=0}[f^2]} < \infty$. By Slutsky's theorem, we can ignore the $\mathbf{O}\left(\frac{1}{N}\right)$ term in considering the convergence in distribution of $\sqrt{N}(H(1, P_1) - V_1^{(N)})$, since $\sqrt{N} \mathbf{O}\left(\frac{1}{N}\right)$ converges to zero in mean-squared and, therefore, also in probability. Observe now that, since $W_{t_{i+1}} - W_{t_i}$ and $W_{t_{j+1}} - W_{t_j}$ are independent for $i \neq j$, $(W_{t_{i+1}} - W_{t_i})^2$ and $W_{t_{i+1}} - W_{t_i}$ are uncorrelated, $\mathbf{E}_0[(W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)] = 0$ and $\mathbf{E}_0[((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i))^2] = 2/(t_{i+1} - t_i)^2$, by the functional central limit theorem (see Ethier and Kurtz [1986]), a piecewise constant martingale

$$\sqrt{N/2} \sum_{i=0}^{[Nt]-1} \left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right) \quad (2.47)$$

converges weakly on $[0, 1]$ to a standard Brownian motion W'_t , which is independent of W_t .¹⁷

We complete the proof by applying Duffie and Protter (1992, Lemma 5.1 and Corollary 5.1). ■

¹⁷The notation $[Nt]$ denotes the integer part of Nt and we use the convention $\Sigma_{\mathbb{C}}^{-1} = 0$.

2.8.3 Proof of Theorem 1(c)

Equation (2.13) follows immediately from Theorem 1(a) and the proof of Theorem 1(b).¹⁸

Combined with Theorem 1(b), (2.13) implies that

$$g = \sqrt{\frac{1}{2} \mathbf{E}_0 \left[\left(\int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW_t' \right)^2 \right]} . \quad (2.48)$$

Equation (2.14) follows from (2.48) using the isometric property of stochastic integrals. ■

2.8.4 Proof of Theorem 2(a)

Before we present the proof, we establish the following result:

$$\mathbf{E}_{t=0} \left[\int_0^1 \left| \frac{\partial^2 H(t, P_t)}{\partial P_t^2} \right|^{2p} dt \right] < \infty \quad (2.49)$$

for some $p > 1$. First, since $F(x)$ is a continuous, piece-wise linear function, it suffices to establish (2.49) for $F(x) = \max(0, x - y)$, $y \geq 0$. Using this definition of $F(x)$ and the fact that one can differentiate (2.2) with respect to P_t (see Theorem 10, p. 72, Friedman (1964)), $\partial^2 H(\tau, x)/\partial x^2$ is equal to the fundamental solution of the Cauchy problem for the following partial differential equation:

$$\frac{\partial u(\tau, x)}{\partial \tau} + \frac{1}{2} \frac{\partial^2}{\partial x^2} [\sigma^2(\tau, x) x^2 u(\tau, x)] = 0.$$

After a logarithmic change of variables, conditions (2.10, 2.16) and $\sigma(\tau, x) \geq \sigma_0 > 0$ allow us to apply Theorem 4.5, p. 141, Friedman (1975), from which we conclude that there exist

¹⁸Relation (2.46), established as a part of the proof of Theorem 1(b), guarantees that convergence in (2.12) occurs not only in distribution, but also in mean-squared sense.

positive constants K_4 and K_5 , such that

$$\left| \frac{\partial^2 H(\tau, x)}{\partial x^2} \right| = |u(\tau, x)| \leq K_4(1 - \tau)^{-\frac{1}{2}} \exp \left[-K_5 \frac{|\ln(x) - \ln(y)|^2}{1 - \tau} \right].$$

By the same theorem and Theorem 5.4, p. 149, Friedman (1975), distribution of P_t has a density, which is bounded above by $K_6 t^{-1/2}$, $K_6 > 0$. Now, by direct computation, we find that there exists a positive constant K_7 , such that

$$\mathbf{E}_{t=0} \left[\left| \frac{\partial^2 H(t, P_t)}{\partial P_t^2} \right|^{2p} \right] \leq K_7(1 - t)^{-p + \frac{1}{2}}.$$

Condition (2.49) now follows by Fubini's theorem.

To prove the statement of part (a), we use Itô's formula and (2.2) to establish that

$$\begin{aligned} H(1, P_1) - V_1^{(N)} &= H(1, P_1) - H(0, P_0) - \sum_{i=0}^{N-1} \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} (P_{t_{i+1}} - P_{t_i}) \\ &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\frac{\partial H(t, P_t)}{\partial P_t} - \frac{\partial H(t, P_t)_{t=t_i}}{\partial P_t} \right) dP_t \\ &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} dP_\tau \right) dP_t. \end{aligned}$$

We now use (2.1) to rewrite the last expression as a sum of four terms:

$$\begin{aligned}
& \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} \mu(\tau, P_\tau) P_\tau d\tau \right) \mu(t, P_t) P_t dt \\
& + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} \mu(\tau, P_\tau) P_\tau d\tau \right) \sigma(t, P_t) P_t dW_t \\
& + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} \sigma(\tau, P_\tau) P_\tau dW_\tau \right) \mu(t, P_t) P_t dt \\
& + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} \sigma(\tau, P_\tau) P_\tau dW_\tau \right) \sigma(t, P_t) P_t dW_t \\
& \equiv I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

Now we will show that, for $k = 1, 2, 3$, $\lim_{N \rightarrow \infty} N \mathbf{E}_{t=0}[I_k^2] = 0$.

Consider the term I_1 first. Using Schwartz's inequality, we conclude that

$$\begin{aligned}
& N \mathbf{E}_{t=0}[I_1^2] \\
& \leq N^2 \mathbf{E}_{t=0} \left[\Delta t \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \left(\frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} \mu(\tau, P_\tau) P_\tau \right)^2 d\tau \cdot (t - t_i) \right) \mu^2(t, P_t) P_t^2 dt \right] \\
& \leq \mathbf{E}_{t=0} \left[\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \left(\frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} \mu(\tau, P_\tau) P_\tau \right)^2 d\tau \right) \mu^2(t, P_t) P_t^2 dt \right] \\
& = \Delta t \mathbf{E}_{t=0} \left[\frac{1}{2} \sum_{i=0}^{N-1} \left(\frac{\partial^2 H(t, P_t)_{t=\tau_i^*}}{\partial P_t^2} \right)^2 \mu^2(\tau_i^*, P_{\tau_i^*}) P_{\tau_i^*}^2 \mu^2(t_i^*, P_{t_i^*}) P_{t_i^*}^2 \Delta t \right],
\end{aligned}$$

where $t_{i+1} - t_i = \Delta t = 1/N$. Random times τ_i^* and t_i^* satisfy $t_i \leq \tau_i^* \leq t_i^* \leq t_{i+1}$ and depend on a particular realization of the stock price process. Their existence is guaranteed by the mean-value theorem. The sum under the expectation sign converges to the integral

$$\frac{1}{2} \int_0^1 \left(\frac{\partial^2 H(t, P_t)}{\partial P_t^2} \mu^2(t, P_t) P_t^2 \right)^2 dt,$$

so we just need to justify the passage to the limit under the expectation sign. This can be done using the Lebesgue dominated convergence theorem. Observe that the integral sum is bounded above by

$$\begin{aligned} & \frac{1}{\Delta t} \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^{t_{i+1}} \left(\frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} P_\tau \right)^2 d\tau \right) \max_{(t,S) \in [0,1] \times [0,\infty]} \mu^4(t, S) \max_{t \in [0,1]} P_t^2 dt \\ &= \max_{(t,S) \in [0,1] \times [0,\infty]} \mu^4(t, S) \int_0^1 \left(\frac{\partial^2 H(t, P_t)}{\partial P_t^2} P_t \right)^2 \max_{t \in [0,1]} P_t^2 dt. \end{aligned}$$

By Hölder's inequality, expected value of the integral in the last expression is bounded above by

$$\mathbf{E}_{t=0} \left[\int_0^1 \left(\frac{\partial^2 H(t, P_t)}{\partial P_t^2} \right)^{2p} dt \right]^{\frac{1}{p}} \mathbf{E}_{t=0} \left[\max_{t \in [0,1]} P_t^{\frac{4p}{p-1}} \right]^{\frac{p-1}{p}},$$

for any $p > 1$. Regularity conditions on functions $\mu(t, S)$ and $\sigma(t, S)$ imply that the second term is finite (see Friedman (1975), Theorem 2.3, p. 107). The first term is finite by (2.49).

This allows us to apply the Lebesgue dominated convergence theorem.

For $k = 2$ we use the isometric property of the stochastic integral first and then apply

Schwartz's inequality. As a result we obtain

$$\begin{aligned}
N\mathbf{E}_{t=0}[I_2^2] &= N\mathbf{E}_{t=0} \left[\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} \mu(\tau, P_\tau) P_\tau d\tau \right)^2 \sigma^2(t, P_t) P_t^2 dt \right] \\
&\leq \mathbf{E}_{t=0} \left[\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \left(\frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} \mu(\tau, P_\tau) P_\tau \right)^2 d\tau \right) \sigma^2(t, P_t) P_t^2 dt \right].
\end{aligned}$$

The last expression converges to zero, by the same argument as in case $k = 1$. Similar argument (using Schwartz's inequality first and then the isometric property) proves the statement for $k = 3$.

We now consider the last term I_4 . Using the isometric property of the stochastic integral,

$$\begin{aligned}
\mathbf{E}_{t=0} [I_4^2] &= \mathbf{E}_{t=0} \left[\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^t \left(\frac{\partial^2 H(\tau, P_\tau)}{\partial P_\tau^2} \right)^2 \sigma^2(\tau, P_\tau) P_\tau^2 d\tau \right) \sigma^2(t, P_t) P_t^2 dt \right] \\
&= \mathbf{E}_{t=0} \left[\sum_{i=0}^{N-1} \left(\frac{\partial^2 H(t, P_t)_{t=\tau_i^*}}{\partial P_t^2} \right)^2 \sigma^2(\tau_i^*, P_{\tau_i^*}) P_{\tau_i^*}^2 \sigma^2(t_i^*, P_{t_i^*}) P_{t_i^*}^2 \right] \frac{(\Delta t)^2}{2},
\end{aligned}$$

for some $t_i \leq \tau_i^* \leq t_i^* \leq t_{i+1}$. Now, by the same argument as in case $k = 1$,

$$\begin{aligned}
\lim_{N \rightarrow \infty} N\mathbf{E}_{t=0} [I_4^2] &= \lim_{N \rightarrow \infty} \mathbf{E}_{t=0} \left[\sum_{i=0}^{N-1} \left(\frac{\partial^2 H(t, P_t)_{t=\tau_i^*}}{\partial P_t^2} \right)^2 \sigma^2(\tau_i^*, P_{\tau_i^*}) P_{\tau_i^*}^2 \sigma^2(t_i^*, P_{t_i^*}) P_{t_i^*}^2 \right] \frac{\Delta t}{2} \\
&= \frac{1}{2} \int_0^1 \mathbf{E}_{t=0} \left[\left(\frac{\partial^2 H(t, P_t)}{\partial P_t^2} \sigma^2(t, P_t) P_t^2 \right)^2 \right] dt,
\end{aligned}$$

which establishes the result of part (a). ■

2.8.5 Proof of Theorem 2(b)

Using the result of part 1 and Markov inequality, we conclude that the sequence of probability measures induced on the real line by $\sqrt{N}(H(1, P_1) - V_1^{(N)})$ is tight and therefore relatively compact (see Billingsley (1986)). This implies that every subsequence of the original sequence of probability measures contains further weakly converging subsequence. To prove the statement of part 2, we need to show that every weakly converging subsequence of the original sequence converges weakly to the stochastic integral in (2.17). Consider a weakly converging subsequence $\sqrt{N_k}(H(1, P_1) - V_1^{(N_k)})$. To show that

$$\sqrt{N_k}(H(1, P_1) - V_1^{(N_k)}) \Rightarrow \frac{1}{\sqrt{2}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW_t',$$

it suffices to check that for any function $\phi(x) = \exp(\sqrt{-1} \lambda x)$,

$$\mathbf{E}_{t=0} \left[\phi \left(\sqrt{N_k} (H(1, P_1) - V_1^{(N_k)}) \right) \right] \rightarrow \mathbf{E}_{t=0} \left[\phi \left(\frac{1}{\sqrt{2}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW_t' \right) \right].$$

We would like to show that for any $\epsilon > 0$, there exists an integer K , such that $k > K$ implies that

$$\left| \mathbf{E}_{t=0} \left[\phi \left(\sqrt{N_k} (H(1, P_1) - V_1^{(N_k)}) \right) - \phi \left(\frac{1}{\sqrt{2}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW_t' \right) \right] \right| < \epsilon. \quad (2.50)$$

As before, let $V_r^{(N_k)}$ denote the value of the replicating portfolio at time r . Then

$$\begin{aligned}
& \left| \mathbf{E}_{t=0} \left[\phi \left(\sqrt{N_k} (H(1, P_1) - V_1^{(N_k)}) \right) \right] - \right. \\
& \quad \left. \mathbf{E}_{t=0} \left[\phi \left(\frac{1}{\sqrt{2}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW'_t \right) \right] \right| \\
& \leq \left| \mathbf{E}_{t=0} \left[\phi \left(\sqrt{N_k} (H(r, P_r) - V_r^{(N_k)}) \right) \right] - \right. \\
& \quad \left. \mathbf{E}_{t=0} \left[\phi \left(\frac{1}{\sqrt{2}} \int_0^r \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW'_t \right) \right] \right| + \\
& \quad \left| \mathbf{E}_{t=0} \left[\phi \left(\sqrt{N_k} (H(1, P_1) - V_1^{(N_k)}) \right) \right] - \mathbf{E}_{t=0} \left[\phi \left(\sqrt{N_k} (H(r, P_r) - V_r^{(N_k)}) \right) \right] \right| + \\
& \quad \left| \mathbf{E}_{t=0} \left[\phi \left(\frac{1}{\sqrt{2}} \int_0^1 \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW'_t \right) \right] - \right. \\
& \quad \left. \mathbf{E}_{t=0} \left[\phi \left(\frac{1}{\sqrt{2}} \int_0^r \sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} dW'_t \right) \right] \right| \\
& \equiv J_1 + J_2 + J_3.
\end{aligned}$$

Fix an arbitrary $\epsilon' > 0$. Given (2.49), we can always pick r such that

$$\mathbf{E}_{t=0} \left[\frac{1}{2} \int_r^1 \left(\sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} \right)^2 dt \right] < \epsilon'.$$

According to Theorem 2,¹⁹ there exists K_1 , such that $k > K_1$ implies that $|J_1| < \epsilon'$. Since $|\phi'(x)| \leq |\lambda|$,

$$\mathbf{E}_{t=0} [J_2^2] \leq \lambda^2 N_k \mathbf{E}_{t=0} \left[\left(H(1, P_1) - V_1^{(N_k)} - H(r, P_r) + V_r^{(N_k)} \right)^2 \right]. \quad (2.51)$$

Using the arguments of part 1, we conclude that there exists K_2 , such that $k > K_2$ implies that

$$\begin{aligned} & \mathbf{E}_{t=0} \left[\left(H(1, P_1) - V_1^{(N_k)} - H(r, P_r) + V_r^{(N_k)} \right)^2 \right] \\ & < \mathbf{E}_{t=0} \left[\frac{1}{2} \int_r^1 \left(\sigma^2(t, P_t) P_t^2 \frac{\partial^2 H(t, P_t)}{\partial P_t^2} \right)^2 dt \right] + \epsilon' < 2\epsilon'. \end{aligned}$$

Therefore, for $k > K_2$, $\mathbf{E}_{t=0} [J_2^2] < 2\lambda^2\epsilon$. It is also clear from the previous argument that $\mathbf{E}_{t=0} [J_3^2] < \lambda^2\epsilon$. Thus, if we set $K = \max(K_1, K_2)$ and $\epsilon' = (1 + 3\lambda^2)^{-1}\epsilon$, inequality (2.50) will hold for any $k > K$. This completes the proof of part (b). ■

2.8.6 Proof of Theorem 3

(2.20) follows from (2.18). The closed-form expression for the option price as a function of time and stock price is given by the Black-Scholes option-pricing formula (see Black and Scholes (1973)). The diffusion coefficient of a geometric Brownian motion is constant and the price of the stock P_t has lognormal distribution. We use Fubini's theorem to change the order of integration in (2.20) and calculate expected values in closed form. ■

¹⁹Necessary regularity conditions can be established using (2.16) Theorem 10, p. 72, Friedman (1964) and Theorems 4.5, 4.6, pp. 141-142, Friedman (1975).

2.8.7 Proof of Theorem 4

(2.23) follows from (2.18). Regularity conditions of the Theorem 2 are not satisfied here. This is not surprising, since in our derivation we assumed that the stock price process can be characterized as a strong solution of the corresponding stochastic differential equation. Therefore, regularity conditions that we impose on the coefficients of such equation are at least as strong as those required by the existence and uniqueness theorem for stochastic differential equations. In case of the mean-reverting process these regularity conditions are not satisfied (the growth rate of the drift coefficient is faster than linear), however the stock price process is still well defined: there exists a unique solution of the diffusion equation (2.21) and the stock price process is obtained from it by exponentiation. Now it is straightforward to verify that our derivation of (2.18) is still valid.

The closed-form expression for the option price as a function of time and stock price is given by the Black-Scholes option-pricing formula (see Black and Scholes (1973)). The diffusion coefficient of a mean-reverting process is constant and the price of the stock P_t has lognormal distribution. We use Fubini's theorem to change the order of integration in (2.20) and calculate expected values in closed form. ■

References

- Avellaneda, M. and A. Paras, 1994, "Dynamic Hedging Portfolios for Derivative Securities in the Presence of Large Transactions Costs", *Applied Mathematical Finance* 1, 165–193.
- Bachelier, L., 1900, "Theory of Speculation," reprinted in P. Cootner, ed., 1964, *The Random Character of Stock Market Prices*. Cambridge, MA: MIT Press.
- Bertsimas, D., Lo, A. and L. Kogan, 1997, "Pricing and Hedging Derivative Securities in Incomplete Markets: An ϵ -Arbitrage Approach", Working Paper No. LFE-1027-97, MIT Laboratory for Financial Engineering.
- Bertsimas, D., Lo, A. and L. Kogan, 1998, "When Is Time Continuous?", Working Paper No. LFE-1033-98, MIT Laboratory for Financial Engineering.
- Bensaid, B., Lesne, J., Pages, H. and J. Scheinkman, 1992, "Derivative Asset Pricing with Transaction Costs", *Mathematical Finance* 2, 63–86.
- Billingsley, P., 1986, *Probability and Measure*, Second Edition. New York: John Wiley & Sons.
- Black, F., M. Scholes, 1973, "Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81, 637–654.
- Boyle, P., D. Emanuel, 1980, "Discretely Adjusted Option Hedges", *Journal of Financial Economics*, 8, 259–282.
- Boyle, P. and T. Vorst, 1992, "Option Replication in Discrete Time with Transaction Costs", *Journal of Finance* 47, 271–294.
- Breeden, D., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities", *Journal of Financial Economics* 7, 265–196.
- Cox, J., Ingersoll, J. and S. Ross, 1985, "An Intertemporal General Equilibrium Model of Asset Prices", *Econometrica*, 36, 363–384.
- Cox, J., Ross, S. and M. Rubinstein, 1979, "Option Pricing: A Simplified Approach", *Journal of Financial Economics* 7, 229–263.
- Duffie, D., 1987, "Stochastic Equilibria with Incomplete Financial Markets", *Journal of Economic Theory* 41, 405–416.
- Duffie, D., 1996, *Dynamic Asset Pricing Theory*, Second Edition. Princeton, NJ: Princeton University Press.
- Duffie, D. and C. Huang, 1985, "Implementing Arrow-Debreu Equilibria by Continuous Trading of Few Long-Lived Securities", *Econometrica* 53, 1337–1356.

- Duffie, D. and M. Jackson, 1990, "Optimal Hedging and Equilibrium in a Dynamic Futures Market", *Journal of Economic Dynamics and Control* 14, 21–33.
- Duffie, D. and P. Protter, 1992, "From Discrete- to Continuous-Time Finance: Weak Convergence of the Financial Gain Process", *Mathematical Finance*, 2, 1–15.
- Duffie, D. and W. Shafer, 1986, "Equilibrium in Incomplete Markets II: Generic Existence in Stochastic Economies", *Journal of Mathematical Economics* 15, 199–216.
- Eberlein, E., 1992, "On Modeling Questions in Security Valuation", *Mathematical Finance*, 2, 17–32.
- Edirisinghe, C., Naik, V. and R. Uppal, 1993, "Optimal Replication of Options with Transaction Costs and Trading Restrictions", *Journal of Financial and Quantitative Analysis* 28, 117–138.
- Ethier, S., T. Kurtz, 1986, *Markov Processes: Characterization and Convergence*. New York: J. Wiley & Sons.
- Figlewski, S., 1989, "Options Arbitrage in Imperfect Markets", *Journal of Finance* 44, 1289–1311.
- Foley, D., 1975, "On Two Specifications of Asset Equilibrium in Macroeconomic Models", *Journal of Political Economy* 83, 303–324.
- Föllmer, H. and D. Sonderman, 1986, "Hedging of Non-Redundant Contingent-Claims", in W. Hildebrand and A. Mas-Colell, eds., *Contributions to Mathematical Economics, in Honor of Gérard Debreu*. Amsterdam: North-Holland.
- Friedman, A., 1964, *Partial Differential Equations of Parabolic Type*. Englewood Cliffs, NJ: Prentice-Hall.
- Friedman, A., 1975, *Stochastic Differential Equations and Applications*, Volume 1. New York: Academic Press.
- Garman, M., 1976, "A General Theory of Asset Valuation under Diffusion State Processes", Working Paper No. 50, University of California Berkeley.
- Grannan, E. and G. Swindle, 1996, "Minimizing Transaction Costs of Option Hedging Strategies", *Mathematical Finance* 6, 341–364.
- Harrison, J.M., D.M. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory*, 20, 381–408.
- He, H., 1989, "Essays in Dynamic Portfolio Optimization and Diffusion Estimations", Ph.D. Thesis, MIT.
- He, H., 1990, "Convergence from Discrete- to Continuous- Time Contingent Claim Prices", *Review of Financial Studies*, 3, 523–546.
- He, H., 1991, "Optimal Consumption-Portfolio Policies: A Convergence from Discrete to Continuous Time Models", *Journal of Economic Theory*, 55,340–363.

- He, H. and N. Pearson, 1991, "Consumption and Portfolio Policies with Incomplete Markets and Short-Sale Constraints: The Infinite-Dimensional Case", *Journal of Economic Theory* 54, 259–304.
- Henrotte, P., 1993, "Transactions Costs and Duplication Strategies", unpublished working paper, Stanford University.
- Hodges, S. and A. Neuberger, 1989, "Optimal Replication of Contingent Claims Under Transactions Costs", unpublished paper, Financial Options Research Center, University of Warwick.
- Huang, C., 1987, "An Intertemporal General Equilibrium Asset Pricing Model: The Case of Diffusion Information", *Econometrica* 55, 117–142.
- Hull, J., A. White, 1987, "The Pricing of Options on Assets with Stochastic Volatilities", *Journal of Finance*, 42, 281–300.
- Hutchinson, J., Lo, A., T. Poggio, 1994, "A Nonparametric Approach to the Pricing and Hedging of Derivative Securities Via Learning Networks", *Journal of Finance*, 49, 851–889.
- Ingersoll, J., 1987, *Theory of Financial Decision Making*. Totowa, NJ: Rowman & Littlefield.
- Itô, K., 1951, "On Stochastic Differential Equations", *Memoirs of the American Mathematical Society* 4, 1–51.
- Jerison, D., Singer, I. and D. Stroock, eds., 1997, *The Legacy of Norbert Wiener: A Centennial Symposium*. Providence, RI: American Mathematical Society.
- Karatzas, I. and S. Shreve, 1991, *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.
- Kushner, H. and P. Dupuis, 1992, *Numerical Methods for Stochastic Control Problems in Continuous Time*. New York: Springer Verlag.
- Leland, H., 1985, "Option Pricing and Replication with Transaction Costs", *Journal of Finance*, 40, 1283–1301.
- Lo, A. and J. Wang, 1995, "Implementing Option Pricing Models When Asset Returns Are Predictable", *Journal of Finance* 50, 87–129.
- Merton, R., 1969, "Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case", *Review of Economics and Statistics* 51, 247–257.
- Merton, R., 1973, "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science* 4, 141–183.
- Merton, R., 1975, "Theory of Finance from the Perspective of Continuous Time", *Journal of Financial and Quantitative Analysis* 10, 659–674.
- Merton, R., 1982a, "On the Microeconomic Theory of Investment Under Uncertainty", in K. Arrow and M. Intriligator, eds., *Handbook of Mathematical Economics*, Volume II. Amsterdam: North-Holland Publishers.

- Merton, R., 1982b, "On the Mathematics and Economic Assumptions of Continuous-Time Models", in W. Sharpe and C. Cootner, eds., *Financial Economics: Essays in Honor of Paul Cootner*. Englewood Cliffs, NJ: Prentice-Hall.
- Milstein, G.N., 1995, *Numerical Integration of Stochastic Differential Equations*. Boston, MA: Kluwer Academic Publishers.
- Milstein, G.N., 1974, "Approximate Integration of Stochastic Differential Equations", *Theory Prob. Appl.*, 19, 557–563.
- Milstein, G.N., 1987, "A Theorem on the Order of Convergence of Mean-Square Approximations of Solutions of Systems of Stochastic Differential Equations", *Theory Prob. Appl.*, 32, 738–741.
- Neuberger, A., 1994, "Option Replication with Transaction Costs: An Exact Solution for the Pure Jump Process", in D. Chance and R. Trippi, eds., *Advances in Futures and Options Research*, Volume 7. Greenwich, CT: JAI Press.
- Platen, E., 1981, "An Approximation Method for a Class of Itô Processes", *Lit. Mat. Sb.*, 21, 121–133 (In Russian).
- Rubinstein, M., 1994, "Implied Binomial Trees", *Journal of Finance* 49, 771–818.
- Stroock, D., 1993, *Probability Theory: an Analytic View*. Cambridge, UK: Cambridge University Press.
- Toft, K., 1996, "On the Mean-Variance Tradeoff in Option Replication with Transactions Costs", *Journal of Financial and Quantitative Analysis*, 31, 233–263.
- Wiener, N., 1923, "Differential Space", *Journal of Mathematical Physics* 2, 131–174.
- Wiggins, J., 1987, "Option Values Under Stochastic Volatility: Theory and Empirical Estimates", *Journal of Financial Economics*, 5, 351–372.
- Willinger, W., Taqqu M. S., 1991, "Towards a Convergence Theory for Continuous Stochastic Securities Market Models", *Mathematical Finance*, 1, 55–99.

Chapter 3

Asset Prices and Irreversible

Investment

3.1 Abstract

This paper presents a general equilibrium model of financial asset prices with irreversible real investment. The focus is on the effects of the irreversibility of real investment on financial asset prices. The model shows how this irreversibility leads to time variation in volatility and systematic risk of stock returns. Changes in these variables are driven by real economic activity, in particular, by firms' investment decisions. Thus, systematic risk of stock returns and their volatility are affected by economy-wide and industry-specific shocks. Firm-specific variables, particularly market-to-book ratios, are linked to real activity and contain information about the dynamic behavior of stock returns. The model of this paper also provides a framework for analyzing futures prices. A comparison between the economy with irreversible investment and an identical economy without the irreversibility shows that all of these results should be attributed to the irreversibility of real investment.

3.2 Introduction

Most asset pricing models focus on the demand side of the economy, making extremely simple assumptions about the supply side. For example, consider two of the most influential papers in this literature, Lucas (1978), and Cox, Ingersoll and Ross (1985). Lucas (1978) assumes that the supply of risky assets in the economy is completely exogenous. Thus, the elasticity of supply is equal to zero and demand shocks are absorbed entirely by changes in asset prices. On the other hand, Cox, Ingersoll and Ross (1985) assume the opposite extreme: in their model the supply of basic risky assets is perfectly elastic. As a result, demand shocks have no effect on the prices of these assets. In both cases the elasticity of supply is fixed, either at infinity or at zero.

The focus on the demand side of the economy proves to be fruitful by delivering tractable models. The obvious drawback is that such models do not lead to a realistic description of supply dynamics, limiting one's understanding of the interaction between real economic activity and prices of financial assets. To learn more about such interaction, the traditional paradigm must be augmented by incorporating real economic activity, such as production and investment decisions by firms.

In this paper I develop a general equilibrium model with a nontrivial production sector. The most prominent feature of the production sector is the irreversibility of real investment. This irreversibility restricts real activity, affecting firms' investment decisions, which in turn determine properties of asset prices, such as stock-return volatility and systematic risk. Since market-to-book ratios and other firms' parameters are closely connected to real economic activity, my model leads to a structural relation between these variables and stock returns.

This paper makes several contributions to the literature on asset pricing and investment.

I formulate and explicitly solve a two-sector continuous-time general equilibrium model of a production economy with irreversible investment. This differs from the bulk of the literature, which analyzes partial equilibrium models. With a few exceptions, models of irreversible investment cannot be solved in closed form and require complex numerical computations.¹ One of the methodological contributions of this paper is its use of singular perturbation techniques to obtain accurate closed-form approximations for the exact solution of the model.

The equilibrium interaction between the supply side and the demand side of the economy leads to several asset-pricing results. One equilibrium effect of irreversibility is that stock-return volatility is time-varying and is a function of real variables, such as investment. This points to a mechanism via which the volatility of returns can increase as the stock price falls. Another effect is that systematic risk of stock returns can change over the business cycle, leading to time-variation in expected returns. Irreversibility also gives rise to a structural relation between the properties of stock returns and firm-specific variables, which proxy for real economic activity. The market-to-book ratio stands out as a particularly natural proxy for real investment and is strongly related to the volatility and systematic risk of stock returns.

The irreversibility of real investment has a number of testable empirical implications. Industries with a higher degree of irreversibility should exhibit more time-variation in their betas and expected returns. The volatility and systematic risk of stock returns are expected to change over time as functions of real investment and market-to-book ratio. These relations have been relatively unexplored in the literature on real investment, which is concerned with the behavior of real variables, and in the finance literature, which focuses primarily

¹Examples of models with closed-form solutions include Pindyck (1988), Dixit (1989b, 1991), He and Pindyck (1992), Abel and Eberly (1994, 1995, 1996, 1997a,b).

on the demand side of the economy. Several papers that do examine both asset prices and real investment, address primarily the properties of short-term interest rates and the aggregate risk premium, but not the behavior of individual stocks. I discuss a few important exceptions in Section 3.3.

My model can also be used to study futures prices. This particular application is more limited in scope, since there is only a small number of commodities for which futures contracts are traded. Moreover, in many cases, futures prices are heavily influenced by the dynamics of inventories, suggesting a price-formation mechanism different from the one considered here. Nevertheless, when storage is relatively expensive, it is reasonable to expect that futures prices are determined to a large extent by the output dynamics (this could apply, for example, to the rapidly growing market for electricity contracts). Therefore, I believe that my model provides a potentially useful method of examining the formation of futures prices, complementing currently accepted modelling paradigms. I include the discussion of futures prices separately in Appendix 3.9.1.

The paper is organized as follows. In Section 3.3, I briefly review the literature. In Section 3.4, I formulate and analyze the general equilibrium model of the economy with irreversible investment. Section 3.5 develops a benchmark model of the economy with perfectly reversible investment. In Section 3.6, I present the results on the behavior of stock prices and futures prices in the economy. Section 3.7 extends the basic model, testing the robustness of the main results. Section 3.8 is the conclusion.

3.3 Literature Review

Extensive literature analyzes the effects of irreversibility and adjustment costs on investment activity. In one of the earliest papers on the subject, Arrow (1968), shows that because of

irreversibility, the optimal investment path of a firm can consist of a sequence of alternating periods of positive and zero investment. This result was obtained under perfect certainty. Implications of capital immobility under certainty have also been analyzed in the growth literature. Examples of this line of research include Johansen (1967), Dasgupta (1969), Ryder (1969), Bose (1970), Floyd and Hynes (1979), Smith and Starnes (1979), LeRoy (1983) and others.² More recent literature has focused on the interaction between irreversibility and uncertainty. I briefly review several aspects of this literature.

One avenue of the literature is concerned with optimal timing of irreversible investment projects, emphasizing the value of the option to delay investment that arises due to irreversibility. This avenue is explored by Henry (1974), Baldwin and Meyer (1979), Baldwin (1982), Brennan and Schwartz (1985), McDonald and Siegel (1986), Ingersoll and Ross (1987), Dixit (1989, 1992) and others. Some of the models incorporate the process of firms' learning about the parameters of the model, e.g., Cukierman (1980), Bernanke (1983) and Caplin and Leahy (1993). Clearly, the very possibility of firms' learning over time creates an additional incentive to delay investment. In this paper, however, I completely ignore such informational problems, assuming that firms possess complete knowledge of their economic environment.

Another strand of the literature includes models of incremental capital accumulation by a single firm facing an uncertain economic environment. In particular, some researchers focus on a competitive firm, taking the stochastic process of the output price and factor prices as a given. These include Majd and Pindyck (1987), Pindyck (1988), Bertola and Caballero (1994), Caballero (1991), Leahy (1993), Abel and Eberly (1994, 1997a) and Abel et al. (1996). Yet others focus on a monopolist facing an exogenously given stochastic

²As LeRoy (1983) argues, Keynes's theory of investment (Keynes (1930, 1936)) can be interpreted as an early analysis of investment under irreversibility.

sequence of demand curves. Models of this type are developed by Pindyck (1988), Bertola (1989), He and Pindyck (1992), Abel and Eberly (1994, 1995, 1996, 1997b) and Dixit and Pindyck (1994). Methodologically, two solution approaches can be identified. One is based on dynamic programming techniques, such as singular stochastic control, e.g., Abel and Eberly (1994, 1995, 1996, 1997a,b) (for rigorous treatment of mathematical techniques, see Harrison (1990), Fleming and Soner (1993)). Another popular approach utilizes methods of contingent claims pricing, e.g., Majd and Pindyck (1987), Pindyck (1988). Both approaches are discussed and compared in great detail in Dixit and Pindyck (1994) and Abel et al. (1996).

In order to characterize the impact of uncertainty and irreversibility on the investment behavior of a competitive firm, it is important to recognize that the output price is determined endogenously in equilibrium, as highlighted by Pindyck (1993). This leads to the third branch of the literature, encompassing equilibrium models of a competitive industry. Examples of such models are Lucas and Prescott (1971), Lippman and Rumelt (1985), Dixit (1989c, 1991, 1992), Caballero and Pindyck (1992), Leahy (1993) and Dixit and Pindyck (1994). Following Lucas and Prescott (1971), most researchers determine the equilibrium allocations by maximizing the total social surplus. Another important methodological insight is based on the observation that firms in a competitive industry can ignore the competition when formulating their investment plans (see Leahy (1993)). This observation not only significantly simplifies the solution process, but also provides an additional justification for popular partial-equilibrium models. Mathematically, the result is based on a deep connection between the singular control problem, faced by the social planner (see, for example, Lucas and Prescott (1971)) and the optimal stopping problem, faced by a *myopic* firm. Leahy (1993) presents a detailed discussion and references to the literature. Sargent (1979)

and Olson (1989) present representative agent, one-sector general equilibrium growth models with irreversible investment. These models are designed to provide some insight into equilibrium dynamics of aggregate investment. However, as pointed out by Bertola and Caballero (1991), the empirical relevance of such models is questionable, since aggregate investment is not volatile enough to make aggregate irreversibility constraints binding.

All papers mentioned above focus on the real side of the economy. There is a small amount of literature that examines the effects of convex adjustment costs on the behavior of prices of financial assets. This literature includes Basu and Chib (1985), Huffman (1985), Basu (1987), Balvers, Cosimano and McDonald (1990), Dow and Olson (1992), Basu and Vinod (1994), Naik (1994), Benavie, Grinols and Turnovsky (1996) and others. All of these are single-sector general-equilibrium models. Out of these, Naik (1994) is the model best suited to incorporate irreversibility of investment. His focus is on the effects of exogenous changes in output uncertainty on the price of aggregate capital and on the evolution of the aggregate risk premium. Coleman (1997) works out a discrete-time, general equilibrium model with two sectors and irreversible investment. He concentrates on the dynamic behavior of the short-term interest rate and its relation to sectoral shocks. The structure of his model and the method of analysis are very different from those in this paper. Several researchers used partial equilibrium models to analyze the behavior of asset prices. Cochrane (1991, 1996) provides a theoretical model and empirical analysis based on a producer's first-order conditions. In the first paper, he uses arbitrage arguments to impose restrictions on investment returns and he tests the model empirically. His assumption that producers are facing smooth adjustment costs leads to unrealistic investment dynamics, but significantly simplifies the analysis. His second paper provides some supporting empirical evidence on the extent to which investment returns can explain the variation in expected returns on

financial assets in a conditional dynamic asset-pricing model.

My results are most closely related to the recent work by Berk, Green and Naik (1998), who target a very similar set of issues. In particular, they develop a partial equilibrium model, in which a firm's investment affects the risk of its stock returns and consequently links expected stock returns to such variables as market-to-book ratio and size. The logic of their model is quite different from the logic of this paper. As the authors readily admit, their model completely ignores the equilibrium interaction between firms' activities and market prices. This equilibrium effect is the main driving force behind my results.

3.4 Irreversible Investment in General Equilibrium

My focus on irreversibility as the main property of real investment is motivated by empirical evidence. In many (if not most) industries physical investment is to a large extent irreversible. Little value can be salvaged by selling off a firm's capital, since many production factors are industry-specific. This prompts companies to adjust their investment rules, taking into account the option value of waiting created by irreversibility (see, for example, Dixit and Pindyck (1994)). As a result, the capital accumulation process is drastically different from what it would have been under perfectly reversible investment.

In my model, the economy consists of two sectors, each using its own, sector-specific production factor (a *capital good*). Thus, there are two capital goods in the economy. One sector represents the industry under consideration, while the other sector represents the rest of the economy (if the industry is sufficiently small relative to the rest of the economy, the other sector can be thought of as a proxy for the market). While it is possible to transfer capital from the rest of the economy into the industry, the reverse process is assumed to be technologically infeasible. As a result, the capital stock of the industry cannot be maintained

at the level that would be optimal if investment were perfectly reversible. Investment is infrequent: it takes place only when the demand for the industry's assets (installed capital) is sufficiently high. Since firms are more likely to respond to a demand shock when the level of demand is already relatively high, the elasticity of supply is a function of the state of the industry and of general economic conditions.

Alternatively, one can characterize the investment process using q . This concept was pioneered by Tobin (1969) and later refined by Abel (1979) and Hayashi (1982). Tobin defined q as the ratio of the market value of a firm to the replacement cost of its capital, which is now known as *average q*. However, in my model, the appropriate concept for the task is *marginal q*, defined as the marginal value of installed capital.³ Marginal q coincides with Tobin's q only under certain conditions.⁴ In my model, investment is triggered when q reaches an endogenously determined threshold. In particular, because of the absence of adjustment costs, firms find it profitable to invest whenever the market value of capital exceeds its replacement cost.

The following informal construction can be used to develop some understanding of how irreversibility affects the prices of financial assets. Consider a single firm in the industry. By definition, its market value can be computed as a product of its average q and the replacement cost of installed capital: $V = qK$. Over a short time interval, given that the

³To be precise, marginal q is defined as the market value of a marginal unit of capital installed in the firm relative to its replacement cost. The value of a marginal unit of installed capital can be computed as the present value of the flow of profits from the sequence of *future* marginal units of capital. This should not be confused with the present value of future profits from a unit of capital that is *presently* marginal. The difference is exactly the opportunity cost arising due to irreversibility. See Pindyck (1988) and Abel et al. (1996) for an extensive discussion of this and related issues.

⁴Average q equals marginal q in my model, because of perfect competition, constant returns to scale and zero adjustment costs. See Hayashi (1982) for the original result in a deterministic setting and Abel and Eberly (1994) for an extension to stochastic models.

firm does not install new capital, return to its owners can be represented as

$$\frac{\pi}{qK} + \frac{\Delta K}{K} + \left(1 + \frac{\Delta K}{K}\right) \frac{\Delta q}{q}, \quad (3.1)$$

where π stands for the cumulative profit generated over the time interval under consideration and ΔK and Δq denote changes in the capital stock and the firm's q respectively (equation (3.1) follows from Itô's formula). Assume that the firm's technology is characterized by constant returns to scale, that there are no adjustment costs and that all firms in the industry behave competitively. If investment was perfectly reversible, the last term in the equation would be absent, since the market value of capital would be identical to its replacement cost. Because investment is irreversible, q can deviate from one, directly affecting the returns.

One effect of irreversibility can be seen in the time-variation of the first term in (3.1). Firms' profits depend on general business conditions, in particular, on the market price of their output. Given the downward-sloping demand curve, the market price of the output is a decreasing function of the aggregate output, which in turn depends on the size of the capital stock in the industry. Since the process of capital accumulation is constrained by firms' inability to disinvest, so is the aggregate industry output and, ultimately, firms' profits.

Further insight into the effects of irreversibility on asset prices can be gained by relating the elasticity of supply to the current state of the economy and the prevailing value of q . Since the supply of risky assets (installed capital) is more elastic when q is relatively high and investment is likely to take place, prices of these assets are not very sensitive to demand shocks and the relative importance of the last term in (3.1) can be expected to be

low. This is precisely the mechanism by which the properties of stock returns are related to real activity and firm-specific variables.

The rest of this section sets the stage for the formal analysis of asset prices, which I undertake in Section 3.6. Here, I develop a general equilibrium model with irreversible investment. The equilibrium is constructed in two steps. First, I find the Pareto optimal allocation and study its properties. In particular, I characterize the optimal consumption/investment policy and establish technical conditions under which the problem is well-posed and under which endogenous variables in the economy follow a stationary process. Second, I prove that the Pareto optimum can be implemented as an outcome of a competitive equilibrium.

Section 3.4.1 states the assumptions about the technology and preferences and formulates the central planner's problem. In Section 3.4.2, I analyze of the central planner's problem. Section 3.4.3 demonstrates that the Pareto optimal allocation can be supported as an outcome of a competitive equilibrium and presents a general characterization of asset prices. Section 3.4.4 develops accurate closed-form approximations to the solution of the central planner's problem.

3.4.1 The Central Planner's Problem

The economy. The economy consists of two sectors. Sector 1 represents the bulk of the economy, excluding *the industry* being analyzed, which is modelled as sector 2. When the industry is relatively small, the first sector is a proxy for *the market*. There are two capital goods and two consumption goods in the economy. The capital good of type 1 can be used for production within sector 1, investment into the capital stock of the industry (sector 2) and can be converted into the consumption good 1. The capital good of type 2 is industry-specific and cannot be used for anything other than production of the consumption good

2.⁵ The consumption good 1 serves as a numeraire.

Capital accumulation. Evolution of physical capital is described by

$$dK_{1t} = (\alpha K_{1t} - c_{1t})dt + \sigma_1 K_{1t} dW_{1t} - dI_t,$$

$$dK_{2t} = -\delta K_{2t}dt + \sigma_{21} K_{2t} dW_{1t} + \sigma_2 K_{2t} dW_{2t} + dI_t,$$

where $cov(dW_{1t}, dW_{2t}) = 0$, K_1 and K_2 are the capital stocks, c_{1t} is the aggregate rate of consumption of good 1 and I_t is the total amount of investment into the capital stock of the industry till time t . The specification of the investment/production technology implies that the industry responds differently to positive and negative demand shocks. Since the capital stock can be adjusted upwards freely, a positive shock can trigger an instantaneous increase in output, while the negative shock has no immediate effect. This asymmetry is an important determinant of equilibrium dynamics.

Furthermore, I assume that the consumption good 2 is perishable and cannot be used for investment.⁶ Thus, the entire industry output is used for consumption. Under the additional assumption of the production technology being characterized by constant returns to scale, this can be formalized as $c_{2t} = XK_{2t}$, where the parameter X controls the productivity of capital in the industry.

Agents. The economy is populated by identical agents with separable homothetic prefer-

⁵In the model, capital goods serve as proxies for production factors in the economy. I assume that there exists sufficient degree of specificity of production factors to prevent factor prices from being equalized across the sectors.

⁶The absence of storage is an important assumption. One would expect an inexpensive storage technology to mitigate the effects of irreversibility to some extent. Thus, the results of this paper should be directly applicable to industries with high cost of storage.

ences:

$$E_0 \left[\int_0^{\infty} e^{-\rho t} U(c_{1t}, c_{2t}) dt \right], \quad (3.2)$$

where

$$U(c_1, c_2) = \begin{cases} \frac{1}{1-\gamma} c_1^{1-\gamma} + \frac{b}{1-\gamma} c_2^{1-\gamma}, & \gamma > 0, \quad \gamma \neq 1, \\ \ln(c_1) + b \ln(c_2), & \gamma = 1. \end{cases}$$

Given the functional form of the utility function, one can set $X = 1$ without further loss of generality. I assume preferences to be separable across goods. While somewhat restrictive, this assumption simplifies the analysis and provides a benchmark for more general studies.

Information structure. I make standard technical assumptions about the information structure of the economy. I assume that there exists a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$, supporting two independent Brownian motions: W_{1t} and W_{2t} . \mathcal{P} is the corresponding Wiener measure. The flow of information is described by a right-continuous increasing filtration $\mathcal{F}_t, t \in [0, \infty)$, $\mathcal{F}_t \subset \mathcal{F}$. Each $\mathcal{F}_t, t \in [0, \infty)$ is an augmentation of the sigma-field generated by Brownian motions $\{(W_{1s}, W_{2s}) : s \in [0, t]\}$ ($\mathcal{F}_\infty \equiv \bigvee_{t \geq 0} \sigma \{(W_{1s}, W_{2s}) : s \in [0, t]\}$).⁷

The central planner enforces the Pareto optimal consumption/investment policy, subject to technological constraints. Formally, feasible investment policies are restricted to be right-continuous with left limits, nonnegative and nondecreasing; feasible consumption policies are restricted to be nonnegative and integrable on any finite time interval. Both are further constrained by the requirement that the stocks of capital goods must remain nonnegative

⁷See Karatzas and Shreve, (1991, Sec. 2.7) for definitions related to Brownian filtrations.

at all times.

Thus, the central planner's problem takes the form

$$\max_{\{c_{1t}, I_t\}} E_0 \left[\int_0^\infty e^{-\rho t} U(c_{1t}, c_{2t}) dt \right], \quad (3.3)$$

subject to

$$dK_{1t} = (\alpha K_{1t} - c_{1t})dt + \sigma_1 K_{1t} dW_{1t} - dI_t, \quad (3.4)$$

$$dK_{2t} = -\delta K_{2t}dt + \sigma_{21} K_{2t} dW_{1t} + \sigma_2 K_{2t} dW_{2t} + dI_t, \quad (3.5)$$

$$c_{2t} = X K_{2t}, \quad (3.6)$$

$$c_{1t} \geq 0, \quad \int_0^T |c_{1t} + c_{2t}| dt < \infty, \quad I_{0-} = 0, \quad (3.7)$$

$$dI_t \geq 0, \quad K_{1t} \geq 0, \quad K_{2t} \geq 0, \quad \forall t, T \geq 0. \quad (3.8)$$

3.4.2 The Solution of the Central Planner's Problem

For the central planner's problem to be well defined, certain restrictions must be imposed on the parameters of the model. I state them in the following proposition.

Proposition 3.3 *The value function of the original problem is finite if and only if*

$$\alpha(\gamma - 1) - \frac{\sigma_1^2}{2} \gamma(\gamma - 1) + \rho > 0, \quad (3.9)$$

$$-\delta(\gamma - 1) + \frac{\sigma_{21}^2 + \sigma_2^2}{2} \gamma(1 - \gamma) + \rho > 0. \quad (3.10)$$

Proof. The idea of the proof is to generate tight bounds on the value function and then to formulate the conditions under which these bounds are finite. For details, see Appendix

3.9.2. ■

Characterization of the Optimal Consumption/Investment Policy. Let $J(K_1, K_2)$

denote the value function of (3.3). Due to the homogeneity of the problem,

$$J(\beta K_1, \beta K_2) = \begin{cases} \beta^{1-\gamma} J(K_1, K_2), & \gamma \neq 1, \\ J(K_1, K_2) + \frac{1+b}{\rho} \ln(\beta), & \gamma = 1. \end{cases}$$

This implies that the value function has a particularly simple functional form:

$$J(K_1, K_2) = \begin{cases} \frac{1}{1-\gamma} K_1^{1-\gamma} j_\gamma \left(\frac{K_2}{K_1} \right), & \gamma \neq 1, \\ \frac{1}{\rho} \ln(K_1) + \frac{b}{\rho} \ln(K_2) + j_1 \left(\frac{K_2}{K_1} \right), & \gamma = 1 \end{cases}$$

(functions $j_\gamma(\cdot)$ depend on the coefficient of relative risk aversion). Thus, the state of the economy can be characterized by a single state variable, defined as the ratio of the capital stocks: $\Omega = K_2/K_1$.

The problem faced by the central planner is of singular control type (for background on singular control of diffusion processes, see Fleming and Soner (1992, Ch. 8)). Similar problems arise in analysis of portfolio decisions under transactions costs. Rigorous analysis of such problems is usually highly technical. Here I summarize only the main properties of the solution. Formal details are not included for the sake of brevity.⁸

⁸Shreve and Soner (1994) present an extensive analysis of a mathematically similar problem of portfolio optimization and consumption with transaction costs. In particular, they prove that the value function of their problem is a classical solution of the corresponding system of differential inequalities (which provides an infinitesimal representation of the dynamic programming principle for problems of this type) and use this fact to establish existence of the optimal consumption/investment policy. Their results can be applied with only minor modifications to the problem at hand. In particular, the special structure of my model leads to an important simplification of the analysis. The fact that the stock of physical capital is restricted to be nonnegative, combined with an infinite marginal utility of consumption at zero, implies that the second capital stock will remain positive at all times. The same result follows for the first stock, given the irreversibility of investment. Thus, the second-order differential operator in the dynamic programming principle becomes nondegenerate after the problem is reduced to one state variable, which insures that the

As in other problems of this type, the value function can be characterized by the differential inequality

$$\min(\rho J - \mathcal{L}J, J_{K_1} - J_{K_2}) = 0, \quad (3.11)$$

where

$$\mathcal{L}J = \sup_{c \geq 0} \left\{ \begin{array}{l} \frac{1}{1-\gamma} c^{1-\gamma} + \frac{b}{1-\gamma} K_2^{1-\gamma} + (\alpha K_1 - c) J_{K_1} + \\ (-\delta K_2) J_{K_2} + \frac{1}{2} \sigma_1^2 K_1^2 J_{K_1 K_1} + \frac{1}{2} (\sigma_2^2 + \sigma_{21}^2) K_2^2 J_{K_2 K_2} + \\ \sigma_1 \sigma_{21} K_1 K_2 J_{K_1 K_2} \end{array} \right\}.$$

J is increasing in K_1 and K_2 and concave: it inherits these properties from the utility function. Concavity of the value function implies that the zero-investment region $\{J_{K_1} - J_{K_2} > 0\}$ has the form $\{K_2 > K_2^*(K_1)\}$, or equivalently $\{\Omega > \Omega^*\}$. The value function can be shown to be twice continuously differentiable everywhere. This translates into the “value matching” and “smooth pasting” (or “super contact”) conditions at the boundary of the zero-investment region:⁹

$$(J_{K_1} - J_{K_2}) \Big|_{\frac{K_2}{K_1} = \Omega^*} = 0, \quad (3.12)$$

$$(J_{K_1 K_1} - J_{K_2 K_1}) \Big|_{\frac{K_2}{K_1} = \Omega^*} = 0, \quad (3.13)$$

$$(J_{K_2 K_2} - J_{K_2 K_1}) \Big|_{\frac{K_2}{K_1} = \Omega^*} = 0. \quad (3.14)$$

Only two of these conditions are linearly independent.

value function is twice continuously differentiable everywhere. In contrast, the value function in Shreve and Soner (1994) can lose smoothness over the set of points at which the differential operator becomes degenerate. This forces them to provide additional arguments to handle such cases.

⁹See Dumas (1991) for an extensive discussion of these optimality conditions.

The problem can be simplified by a change of variables. Let the new independent variable be $\omega = \ln(\Omega)$, and define the new unknown function

$$f_\gamma(\omega) \equiv j_\gamma \left(\frac{K_2}{K_1} \right). \quad (3.15)$$

For the case $\gamma \neq 1$, the new unknown function $f(\omega)$ (I have dropped the subscript γ to simplify the notation) satisfies

$$p_2 f'' + p_1 f' + p_0 f + \gamma \left(f - \frac{1}{1-\gamma} f' \right)^{1-1/\gamma} = -b e^{(1-\gamma)\omega}, \quad (3.16)$$

where

$$\begin{aligned} p_2 &= \frac{\sigma_1^2 + \sigma_{21}^2 + \sigma_2^2 - 2\sigma_{21}\sigma_1}{2}, \\ p_1 &= -\alpha - \delta + \frac{2\gamma - 1}{2}\sigma_1^2 - \frac{\sigma_{21}^2 + \sigma_2^2}{2} + (1-\gamma)\sigma_1\sigma_{21}, \\ p_0 &= (1-\gamma)\alpha - \gamma(1-\gamma)\frac{\sigma_1^2}{2} - \rho, \end{aligned}$$

inside the zero-investment region and the boundary conditions

$$f'(\omega^*)(1 + \Omega^*) = f(\omega^*)\Omega^*(1 - \gamma), \quad (3.17)$$

$$f''(\omega^*)(1 + \Omega^*) = f'(\omega^*)(1 + (1 - \gamma)\Omega^*). \quad (3.18)$$

The optimal consumption policy is given by

$$\bar{c}^* \equiv \frac{c_1^*}{K_1} = \left(f - \frac{1}{1-\gamma} f' \right)^{-\frac{1}{\gamma}} \quad (3.19)$$

For the case $\gamma = 1$, let

$$g(\omega) = f(\omega) - \frac{1}{\rho^2} \left((\alpha - b\delta) - \frac{\sigma_1^2 + \sigma_{21}^2 + \sigma_2^2}{2} + \rho(\ln(\rho) - 1) \right).$$

The unknown function $g(\omega)$ satisfies

$$p_2 g'' + p_1 g' - \rho g - \ln(1 - \rho g') = 0, \quad (3.20)$$

where

$$p_2 = \frac{\sigma_1^2 + \sigma_{21}^2 + \sigma_2^2 - 2\sigma_{21}\sigma_1}{2} > 0,$$

$$p_1 = -\alpha - \delta + \frac{\sigma_1^2 - \sigma_{21}^2 - \sigma_2^2}{2},$$

inside the zero-investment region and

$$b + \rho g'(\omega^*) = \Omega^*(1 - \rho g'(\omega^*)), \quad (3.21)$$

$$g''(\omega^*)(1 + \Omega^*) = \frac{b}{\rho} + g'(\omega^*), \quad (3.22)$$

at the boundary.

$$\tilde{c}^* \equiv \frac{c_1^*}{K_1} = \frac{\rho}{1 - \rho f'} \quad (3.23)$$

To characterize the value function completely, one must specify its asymptotic behavior as $\omega \nearrow \infty$. My analysis here is heuristic in nature – formal justification is provided in Appendix 3.9.4, Lemma 3.9.4.

As ω increases, the possibility of using good 1 for investment becomes less and less important and the value function is asymptotically the same as it would be under the additional constraint $I_t \equiv 0$ (i.e., if the two sectors were completely isolated from each other). The value function under the additional constraint is denoted by $J^{LB}(K_1, K_2)$:

$$\begin{aligned}
J^{LB}(K_1, K_2) &= \frac{\lambda_1}{1-\gamma} K_1^{1-\gamma} + b \frac{\lambda_2}{1-\gamma} K_2^{1-\gamma}, \quad \gamma \neq 1, \\
J^{LB}(K_1, K_2) &= \frac{1}{\rho} \ln(K_1) + \frac{b}{\rho} \ln(K_2) + \\
&\quad \frac{1}{\rho^2} \left((\alpha - b\delta) - \frac{1}{2} (\sigma_1^2 + \sigma_{21}^2 + \sigma_2^2) + \rho (\ln(\rho) - 1) \right), \quad \gamma = 1, \\
\lambda_1 &= \left(\alpha \frac{\gamma-1}{\gamma} - \frac{\sigma_1^2}{2} (\gamma-1) + \frac{\rho}{\gamma} \right)^{-\gamma}, \\
\lambda_2 &= \left(-\delta(\gamma-1) - \frac{\sigma_{21}^2 + \sigma_2^2}{2} \gamma(\gamma-1) + \rho \right)^{-1}.
\end{aligned}$$

Thus, as $\omega \nearrow \infty$,

$$f(\omega) \approx \lambda_1 + b\lambda_2 \exp((1-\gamma)\omega), \quad \gamma < 1, \quad (3.24)$$

$$f(\omega) \approx \lambda_1, \quad \gamma > 1. \quad (3.25)$$

$$g(\omega) \approx 0, \quad \gamma = 1, \quad (3.26)$$

The value function and the optimal consumption/investment policy can now be computed numerically by solving systems (3.16–3.18 and 3.24, 3.25) and (3.20–3.22 and 3.26). The solution procedure is outlined in Appendix 3.9.3.

Next, I characterize the optimal investment process I_t^* and the resulting dynamics of the ratio of capital stocks. Under the optimal choice of the consumption/investment policy, ω_t is a reflected diffusion process, restricted to the half-line $[\omega^*, \infty)$. According to the Itô's formula for semimartingales (e.g., Chung and Williams (1990, Th. 5.1) or Karatzas and

Shreve (1991, Th. 3.3)),

$$d\omega_t = \mu_\omega(\omega_t) dt + (\sigma_{21} - \sigma_1)dW_{1t} + \sigma_2 dW_{2t} + dL_t, \quad \omega_t \geq \omega^*, \quad (3.27)$$

$$\mu_\omega(\omega) = -\alpha - \delta + \tilde{c}_{1t}^*(\omega) + \frac{\sigma_1^2}{2} - \frac{\sigma_{21}^2 + \sigma_2^2}{2}, \quad (3.28)$$

where $\tilde{c}_{1t}^* \equiv c_{1t}^*/K_{1t}$ is a function of ω only (see (3.16, 3.20)) and L_t is the "reflection" process, preventing ω_t from falling below ω^* .¹⁰ Moreover, by the same formula,

$$dL_t = (1 + \Omega^*) \frac{dI_t^*}{K_{2t}^*},$$

where K_{2t}^* is the stock of capital good 2 under the optimal consumption/investment policy.

Thus, one can reconstruct the optimal investment process I_t^* as

$$I_t^* = I_0^* + (1 + \Omega^*)^{-1} \int_0^t K_{2u}^* dL_u, \quad (3.29)$$

where I_0^* is the initial investment necessary to bring ω into the region $[\omega^*, \infty)$: if $\omega_0 < \omega^*$,

$$I_0^* = K_{10}(\Omega_0 - \Omega^*)(1 + \Omega^*)^{-1},$$

otherwise $I_0^* = 0$. The investment process I_t is singular: investment takes place only when $\Omega_t = \Omega^*$ ($\omega_t = \omega^*$). Formally,

$$I_t^* = I_0^* + \int_0^t 1_{\{\omega_s = \omega^*\}} dI_s^*,$$

¹⁰The process L_t can be characterized as the local time of ω_t at ω^* . See Gihman and Skorohod (1972, Ch. 5) or Chung and Williams (1990, Ch.8) for background on reflected diffusion processes.

where $1_{\{\cdot\}}$ denotes the indicator function.

The Long-Run Dynamics. I now state a simple condition under which the ratio of capital stocks possesses a long-run stationary distribution. Existence of the limiting stationary distribution is a prerequisite for empirical analysis of “average” (unconditional) behavior of economic variables. The condition that I am about to present is “almost” necessary: except for a knife-edge case, any violation insures that the ratio of capital stocks converges to infinity.

Proposition 3.4 *If parameters of the model satisfy*

$$\alpha + \delta - \left(\frac{\gamma - 1}{\gamma} \alpha - \frac{\sigma_1^2}{2} (\gamma - 1) + \frac{\rho}{\gamma} \right) - \frac{\sigma_1^2}{2} + \frac{\sigma_{21}^2 + \sigma_2^2}{2} > 0, \quad (3.30)$$

the ratio of the capital stocks Ω_t possesses the long-run stationary distribution. If the inequality opposite to (3.30) holds, Ω_t almost surely tends to infinity and there does not exist a long-run stationary distribution.

Proof. See Appendix 3.9.4. ■

3.4.3 The Competitive Equilibrium

In this section I demonstrate that the solution of the central planner’s problem can be supported as an outcome of a competitive equilibrium in a decentralized production economy. Similar results on equilibrium implementation of Pareto optimal allocations have been developed in several papers (see, for example, Lucas and Prescott (1971), Prescott and Mehra (1980), Brock (1982)).

The decentralized economy is populated by homogeneous households, introduced in Section 3.4, and a large number of competitive firms. Each firm owns a stock of capital

good 2 and sells its output at the spot market.

At any point in time, there exists a spot market, where the consumption good 2 is traded against the numeraire good at the prevailing spot price S_t . Agents purchase good 2 for consumption at the spot market. They also have access to four long-lived financial assets.¹¹ The first asset generates the cumulative return process identical to the constant-returns-to-scale production technology of the first sector:

$$\frac{dv_{1t}}{v_{1t}} = \alpha dt + \sigma_1 dW_{1t}, \quad (3.31)$$

where v_{1t} is the amount invested in this asset at time t . The second asset, the *stock*, is a claim on the stream of cash flows generated by firms in the industry. I assume that at any point in time there is exactly one share of equity outstanding. Hence, each share generates a stream of dividends at rate $S_t K_{2t}^*$ and investment expenses, which total I_t^* by time t . I let P_t denote the ex-dividend stock price. The third asset, the *bond*, earns an instantaneously riskless rate of return r_t .¹² The fourth asset, priced at F_t , is available in zero net supply and produces the cumulative return process

$$\frac{dF_t}{F_t} = \mu_F(\omega_t) dt + dW_{2t}, \quad (3.32)$$

$$\mu_F(\omega_t) = r_t - \frac{U_{c_1 c_1}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{1t}^*, c_{2t}^*)} \begin{pmatrix} 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} d \langle c_{1t}^*, W_{1t} \rangle / dt \\ d \langle c_{1t}^*, W_{2t} \rangle / dt \end{pmatrix}, \quad (3.33)$$

where $\langle \cdot, \cdot \rangle$ denotes the cross-variation process (see Karatzas and Shreve (1991, p. 36)). I

¹¹I introduce the fourth long-lived asset because the instantaneous variance-covariance matrix of returns on the first two risky assets (see definitions below) might be degenerate and thus the market might not be dynamically complete.

¹²Given that the optimal investment process is singular, one could expect returns of financial assets to have a singular component. This turns out not to be the case in equilibrium. I will verify this directly for stock returns in Section 3.6. Absence of arbitrage implies that *financial* asset prices have no singular components in equilibrium (see Karatzas, Lehoczky and Schreve (1991, Section 4)).

define μ_F using the Consumption CAPM formula (e.g., Duffie (1996, p. 229)) to make the return process (3.32) consistent with the optimal consumption policy.

In equilibrium, each household maximizes the expected utility of consumption (3.2), subject to the nonnegativity constraint

$$c_{1t} \geq 0, \quad c_{2t} \geq 0,$$

and the budget constraint

$$\begin{aligned} dV_t = & -(c_{1t} + S_t c_{2t}) dt + v_{bt} r_t dt + \pi_{P_t} (S_t dt + dP_t) + \\ & \pi_{F_t} dF_t + \alpha v_{1t} dt + \sigma_1 v_{1t} dW_{1t}, \end{aligned} \tag{3.34}$$

$$V_t \geq v_{bt} + \pi_{P_t} P_t + \pi_{F_t} F_t + v_{1t}, \tag{3.35}$$

$$V_0 = P_0 + K_{10},$$

$$V_t \geq 0,$$

where V_t is the individual wealth process, v_{bt} is the amount of wealth invested in the bond, π_{P_t} is the number of shares of the stock and π_{F_t} is the number of shares of security F held by the household at time t . The nonnegative-wealth constraint (3.35) rules out arbitrage opportunities (see Dybvig and Huang (1989)). To make sure that the wealth process is well defined by (3.34), I assume that both the consumption policy (c_{1t}, c_{2t}) and the portfolio policy $(v_{bt}, \pi_{P_t}, \pi_{F_t}, v_{1t})$ are progressively measurable processes, satisfying standard

integrability conditions:

$$\int_0^{\tau_n} c_{1t} + S_t c_{2t} + |v_{bt} r_t + \pi_{P_t} (S_t + \mu_P(\omega_t) P_t) + \pi_{F_t} \mu_F(\omega_t) F_t + \alpha v_{1t}| dt < \infty, \quad (3.36)$$

$$\int_0^{\tau_n} |\pi_{P_t}|^2 d\langle P \rangle_t + \int_0^{\tau_n} |\pi_{F_t} F_t|^2 dt + \int_0^{\tau_n} |\sigma_1 v_{1t}|^2 dt < \infty \quad (3.37)$$

for a sequence of stopping times $\tau_n \nearrow \infty$, where $\mu_P(\omega_t) P_t$ and $\langle P \rangle_t$ are respectively the drift coefficient and the quadratic variation processes of P_t .

Since capital markets are dynamically complete, the exact form of financing is not important and I assume for simplicity that firms are financed entirely by equity. Firms make investment decisions to maximize their stock price, determined by the value of their output and investment expenses. Specifically, they solve the following problem:

$$\max_{\{I_t\}} \mathbb{E}_0 \left[\int_0^\infty \eta_{0,t} S_t K_{2t} dt - \int_0^\infty \eta_{0,t} dI_t \right], \quad (3.38)$$

subject to

$$dK_{2t} = -\delta K_{2t} dt + \sigma_{21} K_{2t} dW_{1t} + \sigma_2 K_{2t} dW_{2t} + dI_t$$

and $I_{0-} = 0$, $dI_t \geq 0$. Firms value future cash flows using the stochastic discount factor $\eta_{t,s}$, $t \leq s$. To ensure that firms hold rational expectations, $\eta_{t,s}$ has to be consistent with

observed market prices:

$$-\lim_{\Delta t \searrow 0} \frac{E_t[\eta_{t,t+\Delta t}]}{\Delta t} = r_t, \quad (3.39)$$

$$E_t \left[\int_t^T \eta_{t,s} S_s K_{2s}^* ds - \int_t^T \eta_{t,s} dI_s^* + \eta_{t,T} P_T \right] = P_t, \quad (3.40)$$

$$E_t[\eta_{t,T} F_T] = F_t, \quad (3.41)$$

$$E_t \left[\eta_{t,T} \exp \left(\left(\alpha - \frac{\sigma_1^2}{2} \right) (T-t) + \sigma_1 (W_{1T} - W_{1t}) \right) \right] = 1 \quad (3.42)$$

for arbitrary t and T , such that $T > t$. These conditions, however, are not sufficient for $\eta_{t,s}$ to be a valid discount factor. In general, there might exist more than one process satisfying (3.39– 3.42), leading to a variety of possible values for non-tradable assets. I am looking for an equilibrium with dynamically complete markets, in which case there exists only one candidate stochastic discount factor (see Harrison and Pliska (1983)).

Definition 3.5 *A competitive equilibrium with dynamically complete markets is a collection of stochastic processes K_{1t}^* , K_{2t}^* , c_{1t}^* , c_{2t}^* , I_t^* , v_{bt}^* , v_{1t}^* , π_{Pt}^* , π_{Ft}^* , S_t , r_t , P_t and $\eta_{t,s}$, such that (i) c_{1t}^* , c_{2t}^* , v_{bt}^* , v_{1t}^* , π_{Pt}^* and π_{Ft}^* solve the households' optimization problem, given S_t , r_t , P_t and K_{2t}^* ; (ii) I_t^* is the aggregate investment process solving (3.38), given S_t and $\eta_{t,s}$; (iii) K_{1t}^* and K_{2t}^* solve (3.4,3.5), given c_{1t}^* and I_t^* and the initial stocks of capital in the economy; (iv) markets clear: $c_{2t}^* = K_{2t}^*$, $v_{bt}^* = 0$, $v_{1t}^* = K_{1t}^*$, $\pi_{Pt}^* = 1$, $\pi_{Ft}^* = 0$; (v) S_t , r_t , P_t , I_t^* and K_{2t}^* are such that $\eta_{t,s}$ is the unique stochastic discount factor satisfying (3.39–3.42).*

I construct the competitive equilibrium using the solution of the central planner's problem.

Proposition 3.6 *There exists a competitive equilibrium with dynamically complete markets, satisfying the Definition 3.5. Processes K_{1t}^* , K_{2t}^* , c_{1t}^* , c_{2t}^* and I_t^* are given by the*

solution of the central planner's problem in Section 3.4.2. The optimal portfolio policy is given by $(v_{bt}^*, v_{1t}^*, \pi_{Pt}^*, \pi_{Ft}^*) = (0, K_{1t}^*, 1, 0)$. The stochastic discount factor is defined by

$$\eta_{t,s} = e^{-\rho(s-t)} \frac{U_{c_1}(c_{1s}^*, c_{2s}^*)}{U_{c_1}(c_{1t}^*, c_{2t}^*)}. \quad (3.43)$$

Prices of financial assets satisfy

$$S_t = \frac{U_{c_2}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{1t}^*, c_{2t}^*)}, \quad (3.44)$$

$$r_t = -\frac{E_t [d(e^{-\rho t} U_{c_1}(c_{1t}^*, c_{2t}^*))]}{e^{-\rho t} U_{c_1}(c_{1t}^*, c_{2t}^*) dt}, \quad (3.45)$$

$$P_t = \frac{J_{K_2}(K_{1t}^*, K_{2t}^*)}{J_{K_1}(K_{1t}^*, K_{2t}^*)} K_{2t}^*. \quad (3.46)$$

and (3.32, 3.33).

Proof. See Appendix 3.9.5. ■

3.4.4 Closed-Form Approximations

In this section, I use techniques of singular perturbation analysis to obtain an increasingly accurate sequence of closed-form approximations to the solution of the central planner's problem. These approximations are particularly close to the true solution when b is sufficiently small, in which case the industry remains small relative to the market most of the time.

While popular in physical sciences, such mathematical techniques have seen only a handful of applications in finance, e.g., Hull and White (1987), Atkinson and Wilmott (1995), Bertsimas, Kogan and Lo (1997), Whalley and Wilmott (1997). Thus, the second objective of this section is to present the methodology that might prove helpful in analyzing

other finance models with similar mathematical structure.

Assume that $\gamma \neq 1$. The case of $\gamma = 1$ is presented in Appendix 3.9.6. The unknown function $f(\omega)$ satisfies (3.16), subject to the boundary conditions (3.17, 3.18, 3.24, 3.25). I rescale the independent variable. Define Ξ to be equal to $b^{-1/\gamma}\Omega$. The optimal investment threshold is determined by Ξ^* . Let ξ be the natural logarithm of Ξ : $\xi = \omega - \ln(b)/\gamma$. ξ is the new independent variable. As a function of ξ , the unknown function $f(\xi)$ satisfies

$$p_2 f'' + p_1 f' + p_0 f + \gamma \left(f - \frac{1}{1-\gamma} f' \right)^{1-1/\gamma} = -b^{1/\gamma} e^{(1-\gamma)\xi},$$

subject to the same boundary conditions. I look for $f(\xi)$ and Ξ^* in the form

$$f(\xi) = \lambda_1 + b^{1/\gamma} \left(\lambda_2 e^{(1-\gamma)\xi} + f_{(0)}(\xi) + \dots + b^{n/\gamma} f_{(n)}(\xi) + \dots \right),$$

$$\Xi^* = \Xi_{(0)} + b^{1/\gamma} \Xi_{(1)} + \dots + b^{n/\gamma} \Xi_{(n)} + \dots .$$

Elements of the expansion $(f_{(0)}(\xi), \Xi_{(0)})$, $(f_{(1)}(\xi), \Xi_{(1)})$, etc., can be computed sequentially.

Without going into the details (see Appendix 3.9.6), I present only the resulting expressions for the first two terms in the asymptotic expansion.

The first-order terms:

$$f_{(0)}(\xi) = A_{(0)} \exp(\kappa(\xi - \xi^*)),$$

$$A_{(0)} = \frac{\lambda_2 \gamma (1-\gamma)}{\kappa(\kappa-1)} \left(\frac{\lambda_1}{\lambda_2} \frac{\kappa-1}{\kappa-(1-\gamma)} \right)^{1-1/\gamma},$$

$$\Xi_{(0)} = \left(\frac{\lambda_1}{\lambda_2} \frac{\kappa-1}{\kappa-(1-\gamma)} \right)^{-1/\gamma},$$

$$\kappa = \frac{-q_1 - \sqrt{q_1^2 - 4q_2 q_0}}{2q_2}.$$

The second-order terms:

$$f_{(1)}(\xi) = A_{(1)} \exp(\kappa(\xi - \xi^*)) + C_{(1)} \exp(2\kappa(\xi - \xi^*)),$$

$$C_{(1)} = \frac{1}{2\gamma(\gamma - 1)} \lambda_1^{-1-1/\gamma} \frac{A_{(0)}^2 (1 - \gamma - \kappa)^2}{4\kappa^2 q_2 + 2\kappa q_1 + q_0},$$

where $A_{(1)}$ and $\Xi_{(1)}$ are characterized by the following system of linear equations

$$\kappa A_{(1)} + \left(\lambda_2 (1 - \gamma)^2 \Xi_{(0)}^{-\gamma} - \lambda_1 (1 - \gamma) \right) \Xi_{(1)} = - \left((\kappa + \gamma - 1) A_{(0)} \Xi_{(0)} + 2\kappa C_{(1)} \right),$$

$$(\kappa^2 - \kappa) A_{(1)} - \lambda_2 \gamma (1 - \gamma)^2 \Xi_{(0)}^{-\gamma} \Xi_{(1)} = - \left(\kappa (\kappa + \gamma - 1) A_{(0)} \Xi_{(0)} + (4\kappa^2 - 2\kappa) C_{(1)} \right).$$

Higher-order terms in the expansion (3.106) can be also computed explicitly, providing a sequence of approximations to the optimal solution of increasing level of accuracy.

One can use the asymptotic results to derive a sequence of approximations to the optimal consumption policy. For instance, the first-order approximation is

$$\bar{c}_1^*(\xi) \approx \lambda_1^{-1/\gamma} - b^{1/\gamma} \frac{\lambda_1^{-1/\gamma-1}}{\gamma} A_{(0)} \left(1 - \frac{\kappa}{1 - \gamma} \right) e^{\kappa(\xi - \xi^*)}.$$

It is an increasing function of ξ and it approaches $\lambda_1^{-1/\gamma}$ as ξ approaches infinity. When b is small, the optimal consumption policy is approximately constant and equal to the one in an identical economy without the second sector.

To evaluate the accuracy of the asymptotic expansion, I compare the first- and the second-order approximations to the optimal consumption/investment policy with the numerical results. I plot the optimal consumption policy, \bar{c}^* as a function of Ω and the approximations in Figures 3-1, 3-2. I also tabulate the optimal investment threshold Ξ^* and the approximations in Table 3.1. The subjective discount rate is set to $\rho = 0.05$ and

the estimates of real stock-market returns are taken from Obstfeld (1994). I consider three different values of the relative risk aversion parameter: 0.5, 1.0 and 2.0. The rest of the parameters are chosen somewhat arbitrarily. Resulting moments of returns on the industry portfolio

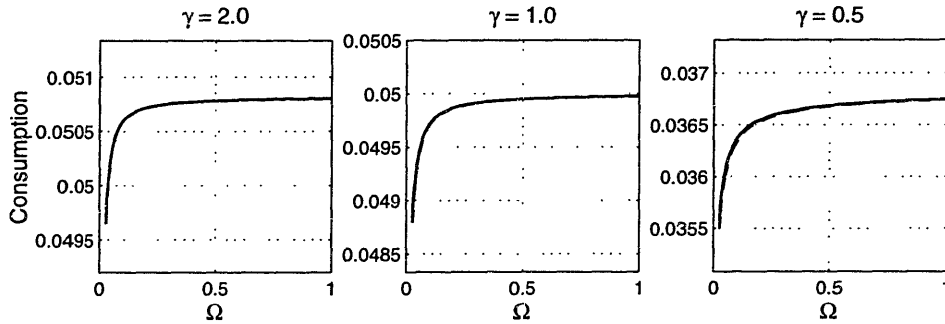


Figure 3-1: The optimal consumption policy $\bar{c}_1^*(\Omega)$ (solid), the first-order approximation (dash) and the second-order approximation (dash-dot) are plotted against the ratio of the capital stocks $\Omega = K_2/K_1$. The following set of parameter values is used: $b = 0.05$, $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$.

Figure 3-1 shows that both the first- and the second-order approximations are highly accurate. Given the choice of parameter values, the industry accounts for at least 2.5% of physical capital in the economy. Thus, the industry is not negligible relative to the market. Figure 3-2 goes even further. While the first-order approximation is not particularly accurate for $\gamma = 0.5$, the second-order approximation is practically indistinguishable from the numerical solution, even though the capital stock of the industry is at least 9% of that of the first sector. Table 3.1 shows that the second-order approximation to the optimal investment threshold is extremely close to the numerical solution. Thus, even a small number of terms in the asymptotic expansion can provide an accurate approximation to the solution of the problem at hand. The first-order approximation is within 3% for $b = 0.05$,

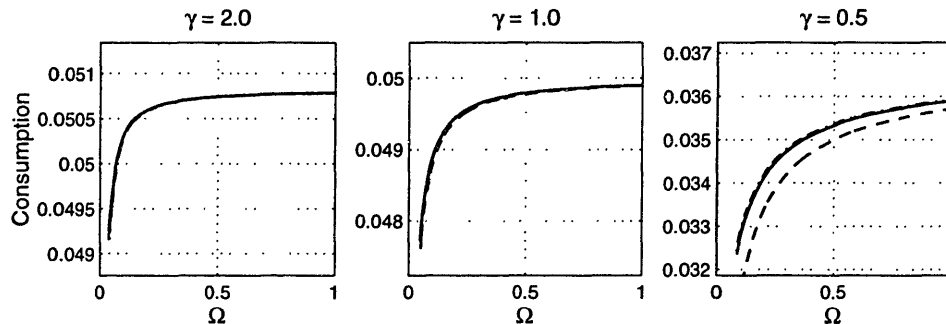


Figure 3-2: The optimal consumption policy $\tilde{c}_1^*(\Omega)$ (solid), the first-order approximation (dash) and the second-order approximation (dash-dot) are plotted against the ratio of the capital stocks $\Omega = K_2/K_1$. The following set of parameter values is used: $b = 0.1$, $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$.

which suggests that it can be highly accurate for reasonable values of the parameters.¹³

3.5 Reversible Investment in General Equilibrium

In this section I develop the model of fully reversible investment. This provides a benchmark for evaluating the effect of irreversibility on the properties of financial assets.

3.5.1 The Central Planner's Problem

Since there are no constraints on the capital transfer between the two sectors of the economy, there is no need to differentiate between the two capital stocks. One only has to keep track

¹³As the development of this section suggests, certain singular control problems arising in financial economics can be efficiently solved using computer algebra software. By automating symbolic computations, one can carry out perturbation analysis of arbitrarily high order. Such a combination of perturbation analysis with computer algebra algorithms has proved to be an effective tool for tackling challenging problems in physical sciences, e.g. in nonlinear dynamics

This approach has a number of advantages over traditional numerical methods, such as finite-difference schemes. The solution can be computed for all possible combinations of model parameters at once, which significantly simplifies estimation of the model. Numerical stability is not an issue and no accuracy is lost in computing derivatives of the solution. Pursuing this direction further would take me too far away from the main topic of this paper, therefore I leave it for future research.

γ	$b^{1/\gamma}\Xi_{(0)}$	$b^{1/\gamma}(\Xi_{(0)} + b^{1/\gamma}\Xi_{(1)})$	Numerical Solution
<u>$b = 0.05$</u>			
2.0	0.0277	0.0273	0.0274
1.0	0.0258	0.0254	0.0254
0.5	0.0250	0.0243	0.0243
<u>$b = 0.10$</u>			
2.0	0.0392	0.0384	0.0385
1.0	0.0516	0.0499	0.0500
0.5	0.1001	0.0886	0.0900

Table 3.1: The optimal investment threshold Ω^* , the first-order and the second-order approximations are tabulated for different values of the risk-aversion parameter γ and b . The following set of parameter values is used: $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$.

of the evolution of the aggregate capital stock K_t . The investment policy is specified by the choice of θ_t , defined as the proportion of the aggregate capital stock invested in sector 1 at time t . Thus, the central planner's problem takes the form

$$\max_{c_{1t}, \theta_t} \mathbb{E}_0 \left[\int_0^{\infty} e^{-\rho t} \left(\frac{1}{1-\gamma} c_{1t}^{1-\gamma} + \frac{b}{1-\gamma} (\theta_t K_t)^{1-\gamma} \right) dt \right], \quad (3.47)$$

subject to

$$dK_t = ((\alpha\theta_t - \delta(1 - \theta_t))K_t - c_{1t})dt + (\sigma_1\theta_t + \sigma_{21}(1 - \theta_t)) K_t dW_{1t} + \sigma_2(1 - \theta_t)K_t dW_{2t}, \quad (3.48)$$

$$\theta_t \geq 0, \quad (3.49)$$

$$\theta_t \leq 1, \quad (3.50)$$

$$K_t \geq 0, \quad (3.51)$$

Constraints (3.49, 3.50) restrict physical capital to be nonnegative (see (3.8)).

3.5.2 The Solution of the Central Planner's Problem

The solution of the central planner's problem (3.47–3.49) is finite as long as parameters of the model satisfy (3.9, 3.10) and the additional condition

$$\bar{\alpha}(\gamma - 1) - \frac{\bar{\sigma}^2}{2}\gamma(\gamma - 1) + \rho > 0, \quad (3.52)$$

where

$$\begin{aligned} \bar{\alpha} &= \alpha\bar{\theta} - \delta(1 - \bar{\theta}), \\ \bar{\sigma} &= \sqrt{(\sigma_1\bar{\theta} + \sigma_{21}(1 - \bar{\theta}))^2 + \sigma_2(1 - \bar{\theta})^2}, \\ \bar{\theta} &= \max \left(0, \min \left(1, \frac{\alpha + \delta - \gamma \left(\sigma_{21} \frac{\sigma_1 - \sigma_{21}}{2} - \sigma_2^2 \right)}{\gamma \left(\frac{(\sigma_1 - \sigma_{21})^2}{2} + 2\sigma_2^2 \right)} \right) \right). \end{aligned}$$

(see Appendix 3.9.7). The solution is given by the following proposition.

Under the assumption (3.52), the indirect utility function and the optimal consump-

tion/investment policy of (3.47–3.49) are characterized by

$$J(K) = \frac{\lambda}{1-\gamma} K^{1-\gamma}, \quad (3.53)$$

$$c_{1t}^* = \lambda^{-1/\gamma} K_t, \quad (3.54)$$

$$\theta_t^* \equiv \max(\theta^*, 0), \quad (3.55)$$

$$b(1 - \theta^*)^{-\gamma} + \lambda\gamma((\sigma_1 - \sigma_{21})^2 + \sigma_2^2)\theta^* + \lambda\gamma((\sigma_1 - \sigma_{21})\sigma_{21} - \sigma_2^2) - \lambda(\alpha + \delta) = 0, \quad (3.56)$$

$$b(1 - \theta^*)^{1-\gamma} + \lambda^{1-1/\gamma} + \lambda(1 - \gamma) \left((\alpha + \delta)\theta^* - \delta - \lambda^{-1/\gamma} \right) - \frac{1}{2}\lambda\gamma(1 - \gamma) \left((\theta^*(\sigma_1 - \sigma_{21}) + \sigma_{21})^2 + (1 - \theta^*)^2 \sigma_2^2 \right) - \rho\lambda = 0. \quad (3.57)$$

Proof. See Appendix 3.9.7. ■

3.5.3 The Competitive Equilibrium

As in Section 3.4.3, one can show that the Pareto optimal consumption/investment policy can be implemented as a competitive equilibrium. I will not repeat the details of the argument here.

It is straightforward to characterize the properties of asset prices in this economy. The average q is identically equal to one. Values of q above one cause entry into the industry and cannot persist in equilibrium; values of q below one cause exit from the industry and are also inconsistent with equilibrium. Thus the stock price equals $(1 - \theta_t^*) K_t$, which is simply the amount of capital in the industry.

The fact that q is identically equal to one implies a particularly simple behavior of stock prices. In particular, the beta of stock returns with respect to the first sector is constant and equals σ_{21}/σ_1 . Similarly, the volatility of stock returns is constant and is given by $\sigma_{21}^2 + \sigma_2^2$.

3.6 Irreversibility and Stock Returns

As it has been established in Section 3.4.3, the solution of the central planner's problem can be supported as an outcome of a competitive equilibrium. In this section I study properties of stock returns: their volatility and systematic risk and their relation to real economy.

Stocks in the model are claims on the entire output of the industry. Empirically, one can think of the stock as a portfolio of stocks of companies in a particular industry. In the model, there is only one share outstanding at any point in time. As it has been established in Proposition 3.6, the stock price can be computed as the shadow price of capital (i.e., the ratio of the indirect marginal utility of the second capital good and the indirect marginal utility of the numeraire good) times the total amount of capital installed:

$$P = \frac{J_{K_2}}{J_{K_1}} K_2. \quad (3.58)$$

The ratio J_{K_2}/J_{K_1} defines average q for firms in the industry, since in my model the replacement cost of capital is identically equal to one. Being equal to the marginal q , it equals one when investment takes place and is less than one otherwise.

In order to describe the behavior of financial assets in the economy, one has to characterize the dynamics of q . Using Itô's rule for semimartingales,

$$dq_t = \mu_q(\omega_t)q_t dt + \sigma_{q1}(\omega_t)q_t dW_{1t} + \sigma_{q2}(\omega_t)q_t dW_{2t} - q_t \left(\frac{q'(\omega_t)}{q_t(\omega_t)} \right) dL_t,$$

$$\mu_q(\omega) = \mu_\omega(\omega) \frac{q'(\omega)}{q(\omega)} + \frac{\sigma_2^2 + (\sigma_{21} - \sigma_1)^2}{2} \frac{q''(\omega)}{q(\omega)}, \quad (3.59)$$

$$\sigma_{q1}(\omega) = (\sigma_{21} - \sigma_1) \frac{q'(\omega)}{q(\omega)}, \quad (3.60)$$

$$\sigma_{q2}(\omega) = \sigma_2 \frac{q'(\omega)}{q(\omega)}, \quad (3.61)$$

where $\mu_\omega(\omega)$ is defined by (3.28). Due to the fact that the state variable ω_t is reflected at ω^* , $dL_t = 1_{\{\omega_t = \omega^*\}} dL_t$ and the singular component of the process q_t equals

$$q_t \left(\frac{q'(\omega^*)}{q(\omega^*)} \right) dL_t.$$

It is easy to see that this is in fact equal to zero, since $q'(\omega^*) = 0$.¹⁴ Thus, q_t has no singular component and follows a regular diffusion process.

The fact that $q'(\omega^*) = 0$ is a direct consequence of two assumptions: irreversibility of investment and instantaneous upward adjustment of the capital stock. Since this result is crucial for understanding the behavior of financial asset prices, it is important to discuss the economic intuition behind it. If $q'(\omega^*)$ was different from zero (negative), the *capital gain* due to the change in q (at $\omega = \omega^*$) over a time-period Δt would be negative and of order $\sqrt{\Delta t}$, since instantaneous investment prevents q from rising above one. Since the flow of profits over the time-period is of order Δt , it could not be optimal for firms to invest when $\omega = \omega^*$. This argument suggests that the main results of the paper are robust in the following sense: if it was not possible to invest at an infinite rate, the downward pressure on q would still imply that the upside due to capital gains is limited, hence the downside and the slope of $q(\omega)$ at ω^* would have to be limited accordingly.

Note that the above argument does not rely on the one-factor structure of the model. Even if the state of the economy was driven by several state variables, q (the market-to-book ratio) would affect stock returns in a similar fashion: variation in q would have more

¹⁴The optimality conditions (3.12-3.14) imply that

$$(J_{K_2}/J_{K_1})_{K_2} = 0$$

at $\omega = \omega^*$. The result follows from

$$q'(\omega) = K_2(J_{K_2}/J_{K_1})_{K_2}.$$

impact on returns for relatively low values of q . This singles out the market-to-book ratio as a firm-specific variable that is particularly informative from the perspective of asset pricing. This should not be surprising in view of the following observation. In the model, behavior of stock returns is driven by the elasticity of supply, which is determined by the real investment policy. The informativeness of q (market-to-book) stems from the fact that it is the sufficient statistic for real investment in my model.

Stock returns can be decomposed into a sum of three components: the change in the ex-dividend stock price, $dq_t/q_t + dK_{2t}^*/K_{2t}^* + \langle dq_t/q_t, dK_{2t}^*/K_{2t}^* \rangle$, the negative of the investment cost relative to the share price, $-dI_t^*/(q_t K_{2t}^*)$, and the dividend flow per dollar invested, S_t/q_t . Since $dI_t^* - dI_t^*/q_t = 0$, stock returns have no singular component. Thus, the instantaneous rate of return is given by

$$\begin{aligned} & \left(\mu_q(\omega) - \delta + \sigma_{q1}(\omega)\sigma_{21} + \sigma_{q2}(\omega)\sigma_2 + \frac{S(\omega)}{q(\omega)} \right) dt + \\ & \left((\sigma_{21} - \sigma_1) \frac{q'(\omega)}{q(\omega)} + \sigma_{21} \right) dW_{1t} + \left(\sigma_2 \frac{q'(\omega)}{q(\omega)} + \sigma_2 \right) dW_{2t}. \end{aligned} \quad (3.62)$$

3.6.1 Volatility

Consider the instantaneous standard deviation of the cumulative return process:

$$\sigma = \left[\left((\sigma_{21} - \sigma_1) \frac{q'(\omega)}{q(\omega)} + \sigma_{21} \right)^2 + \left(\sigma_2 \frac{q'(\omega)}{q(\omega)} + \sigma_2 \right)^2 \right]^{1/2}. \quad (3.63)$$

Since the above expression depends on the state variable, stock returns are heteroscedastic and their volatility is persistent. The relation between the volatility and the state variable can be quite complicated and is determined by model parameters. We can, however, gain some insight by considering several special cases. According to (3.63), when $\omega = \omega^*$ ($q = 1$),

σ equals $(\sigma_2^2 + \sigma_{21}^2)^{1/2}$. Compare this with the limiting behavior as $\omega \rightarrow \infty$, ($q \rightarrow 0$). In the limit, the ratio $q'(\omega)/q(\omega)$ approaches $-\gamma$ and σ tends to

$$\left(\sigma_2^2(1-\gamma)^2 + (\sigma_{21}(1-\gamma) + \gamma\sigma_1)^2\right)^{1/2}. \quad (3.64)$$

When direct shocks to industry's capital are relatively small, i.e., when $\sigma_2^2 + \sigma_{21}^2 < \sigma_1^2$, volatility is higher for low values of q .

To obtain more specific results, I assume that $b \ll 1$ and use the asymptotic approximation developed in Section 3.4.4. This results in the approximation to the industry's q :

$$q = \frac{f'\Omega^{-1}}{(1-\gamma)f - f'} = \frac{\lambda_2}{\lambda_1} e^{-\gamma\xi} + \frac{\kappa}{\lambda_1} \frac{A(0)}{1-\gamma} e^{(\kappa-1)\xi - \kappa\xi^*} + O(b^{1/\gamma}). \quad (3.65)$$

In the limit of $b \rightarrow 0$, one can check using (3.111, 3.112) and (3.113) that q is a monotonically decreasing function of its argument, which in turn implies that there is a one-to-one correspondence between q and the state variable (the ratio of capital stocks). The ratio $q'(\omega)/q(\omega)$ is negative and increasing in ω . Clearly, $q'(\omega^*)/q(\omega^*) = 0$ and (3.113) implies $\lim_{\omega \rightarrow \infty} q'(\omega)/q(\omega) = -\gamma$.

Consider the special case in which there are no direct shocks to the industry's capital stock. Then

$$\sigma = \left| \sigma_1 \frac{q'(\omega)}{q(\omega)} \right|.$$

This is an increasing function of the state variable and is negatively related to q . This implies that stock returns become more volatile as the stock price (the market-to-book

ratio) falls. According to (3.63), one would expect a similar type of behavior for low values of q when σ_1 dominates σ_{21} and σ_2 or when the risk aversion coefficient γ is sufficiently large.

I illustrate these points in Figures 3-3 and 3-4 by plotting σ against q . As one can see from these figures, the volatility as a function of q can take on various shapes, depending on the model parameters: it can be increasing, decreasing, inverted hump-shaped, convex, concave, etc. Figure 3-3 presents the case of σ_{21} and σ_2 being relatively small. Volatility is generally negatively related to q , particularly for larger values of γ . The first panel of Figure 3-4 corresponds to a relatively large value of γ (compared to the values used in the other two panels). One can see that as q gets smaller (0.75 and less), stock-return volatility increases rapidly.

Leahy and Whited (1996) present some empirical evidence on the relation between return volatility, q and investment based on a sample of six hundred U.S. manufacturing firms. While one should not expect to find an exact linear relation between volatility of stock returns and such variables as q and investment across different stocks, their results do provide a useful summary of the data. Their main findings are consistent with the predictions of this model. The amount of real investment is negatively related to the volatility of stock returns (adjusted for leverage). When Tobin's q is included in the regressions, the coefficient on the volatility becomes insignificant. They also report a negative relation between the return volatility and q .

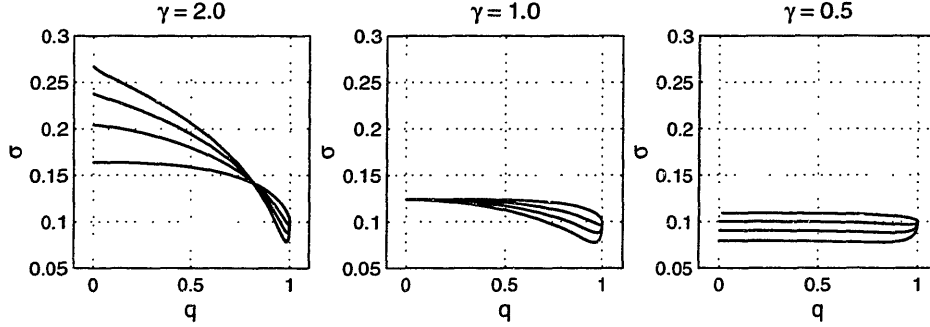


Figure 3-3: The instantaneous standard deviation of the industry portfolio is plotted as a function of the industry's q . The following set of parameter values is used: $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.01$, $\sigma_1 = 0.124$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.08$. Four different values of σ_{21} are used: 0.0, 0.02, 0.04, 0.06.

3.6.2 Systematic Risk

The beta of the stock with respect to the first sector equals

$$\beta_{P,K_1} = \frac{\sigma_{21}}{\sigma_1} + \left(\frac{\sigma_{21}}{\sigma_1} - 1 \right) \frac{q'(\omega)}{q(\omega)}. \quad (3.66)$$

It is a function of the state variable ω_t , and therefore varies over time in response to changes in economic conditions. One can see that the beta can be naturally decomposed into a sum of two terms. The first term, σ_{21}/σ_1 , can be interpreted as the technology beta: it equals the beta of K_{2t} with respect to K_{1t} ,

$$\beta_{K_2,K_1} = \frac{d\langle K_{2t}, K_{1t} \rangle}{d\langle K_{1t}, K_{1t} \rangle} = \frac{\sigma_{21}}{\sigma_1}.$$

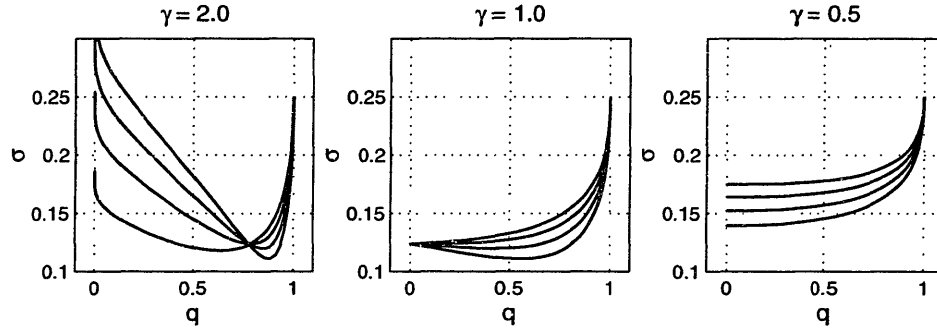


Figure 3-4: The instantaneous standard deviation of the industry portfolio is plotted as a function of the industry's q . The following set of parameter values is used: $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$. Four different values of σ_{21} are used: 0.0, 0.06, 0.12, 0.18.

The second term, $(\sigma_{21}/\sigma_1 - 1) q'(\omega)/q(\omega)$, measures the elasticity of q with respect to the state variable. It equals zero at ω^* , when real investment takes place and q equals one. As ω increases to infinity, q approaches zero and $q'(\omega)/q(\omega)$ converges to $-\gamma$.

Thus, the technology beta is the key parameter in the relation between the stock beta and the state of the economy (or the market-to-book ratio of the firm). It's effect is easy to see from the asymptotic behavior of the beta:

$$\lim_{\omega \searrow \omega^*} \beta_{P,K_1}(\omega) = \beta_{K_2,K_1}$$

$$\lim_{\omega \nearrow \infty} \beta_{P,K_1}(\omega) = \beta_{K_2,K_1} - \gamma(\beta_{K_2,K_1} - 1).$$

Thus, $\beta_{P,K_1}(\infty)$ exceeds $\beta_{P,K_1}(\omega^*)$ if and only if β_{K_2,K_1} is less than one.

Next, consider the limit of b approaching zero. The consumption of the numeraire good

is given by

$$c_{1t} = \left(\lambda_1^{-1/\gamma} + O(b^{1/\gamma}) \right) K_{1t}.$$

Thus, the instantaneous correlation between innovations in consumption and industry-specific shocks, $\text{corr}(dc_{1t}, dW_{2t})$, is of order $b^{1/\gamma}$. Therefore, the market risk premium associated with industry-specific shocks is of order $b^{1/\gamma}$ and the consumption beta of the stock is approximately the same as its beta with respect to the first sector: $\beta_{P,K_1} = \beta_{P,c_1} + O(b^{1/\gamma})$. This justifies the focus on β_{P,K_1} as a measure of the systematic risk of the stock.

In the limit, q is a monotone function of the state variable, given explicitly by (3.65). Thus, it is easy to characterize the dependence of the market beta on ω and q . It is a monotone function: if β_{K_2,K_1} is less than one, it is decreasing in q , otherwise it is increasing. Specifically,

$$\beta_{P,K_1} = \frac{\sigma_{21}}{\sigma_1} - \gamma \left(\frac{\sigma_{21}}{\sigma_1} - 1 \right) \frac{e^{(1-\gamma)(\omega-\omega^*)} - e^{\kappa(\omega-\omega^*)}}{e^{(1-\gamma)(\omega-\omega^*)} + \frac{\gamma}{\kappa-1} e^{\kappa(\omega-\omega^*)}} + O(b^{1/\gamma}). \quad (3.67)$$

The beta of stock returns is a monotone function of q . In particular, it is in inverse relation with q if and only if the technology beta, σ_{21}/σ_1 , is less than unity. Thus, one cannot expect to find the same qualitative relation between the covariance of stock returns with the market returns and q or investment across different industries or firms. This would explain why Leahy and Whited (1996) find no such relation in their panel regression, while their results do show a significant relation between investment, q and the total volatility of stock returns.

As an illustration, I compute the market beta for several sets of parameter values and plot it as a function of q in Figure 3-5. The beta is more variable as a function of q for

larger values of the risk aversion parameter.

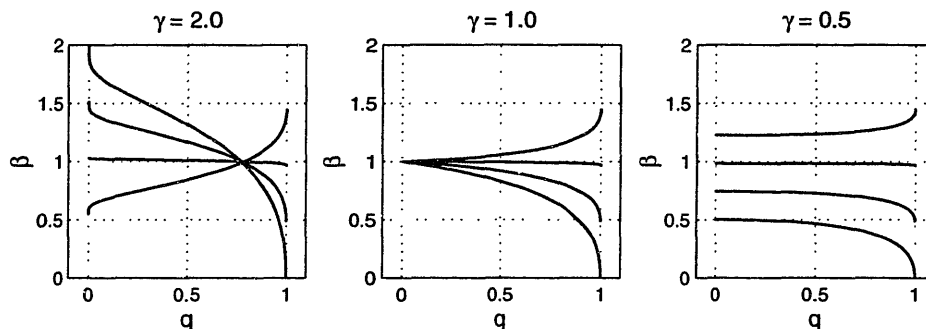


Figure 3-5: The beta of the industry portfolio is plotted as a function of the industry's q . The following set of parameter values is used: $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$. Four different values of σ_{21} are used: 0.0, 0.06, 0.12, 0.18.

Dynamically, the systematic risk of the stock follows a nonlinear mean-reverting process, determined by evolution of the state of the economy. Market-to-book ratio provides a good proxy for the underlying real variables. In fact, it contains complete information about the state of the economy in this model, being in one-to-one correspondence with the state variable. As I have pointed out in the beginning of this section, market-to-book would remain an informative proxy even if the economy was described by more than one state variable. When the market-to-book ratio (q) is below its long-run mean, it drifts down, while it is restricted from rising above one (due to new investment). In general, the beta is affected both by market-wide and industry-specific shocks via their impact on q . In Section 3.5, I have shown that the beta is not variable when investment is perfectly reversible. This suggests the following testable implication: industries with higher degree of irreversibility should exhibit higher degree of time-variation in their beta.

The dependence of the systematic risk on the state of the economy has implications for

asset allocation across sectors (industries). In particular, the systematic risk of the industry portfolio changes in response to market-wide shocks in a predictable manner. Consider the case $\beta_{K_2, K_1} < 1$ first. The industry portfolio becomes relatively less risky when the market is doing well, because in that case positive shocks to the market raise q , reducing the beta of the stocks accordingly. The opposite is true when the market experiences a sequence of negative shocks – the beta of the industry increases. In case of $\beta_{K_2, K_1} > 1$ the argument changes. Now positive shocks to the first sector reduce the industry's q and its beta, while negative shocks have the opposite effect. The bottom line is that a sequence of negative (positive) shocks to the market leads to a rise (decline) in the industry's beta. This is another testable implication of the model.

The above argument suggests that popular statements about certain industries being good “defensive” investments due to their low correlation with the market should be taken with caution: an industry can be practically uncorrelated with the market when times are good, but can become highly correlated during market declines. As a result, certain industries can be more sensitive to significant market declines than to extended periods of market growth. Such an asymmetry in the stocks' response to market-wide shocks is intimately related to the asymmetric response of investment to changes in business conditions: both are driven by irreversibility of real investment.

Since the expected excess rate of return on the stock is proportional to its systematic risk, it exhibits similar dependence on the state of the economy and firm-specific variables, such as the market-to-book ratio. In particular, expected returns are time-varying and persistent. This provides a potential theoretical underpinning for empirical findings on time-variation in expected returns (see, for example, Conrad and Kaul (1988)). It also implies that returns are predictable, more so at long than at short horizons (see Fama and

French (1988)).

The Long-Run Dynamics of the Beta. Ignoring terms of order $b^{1/\gamma}$ and higher, the dynamics of the beta can be completely characterized in terms of the dynamics of ω_t (or ξ_t , as defined in Section 3.4.4). When the condition (3.30) is satisfied, ξ_t is a reflected Brownian Motion. The stationary long-run distribution of $\xi_t - \xi^*$ is exponential with mean (scale parameter) $-\sigma_\xi^2 / (2\mu_\xi)$, where μ_ξ and σ_ξ are the drift and the diffusion coefficients of ξ_t . This makes it easy to study the long-run behavior of the beta. To provide some insight, I compute the first two unconditional moments of the beta for several sets of parameter values. These results are summarized in Table 3.2. The unconditional variance appears to increase with γ . Otherwise, both the variance and the mean of the beta exhibit complicated dependence on model parameters. Most importantly, the table shows that the systematic risk of stock returns can vary significantly over time.

3.7 Extensions and Generalizations

Gradual Adjustment. In the previous sections I assumed that the capital stock of the industry can be adjusted upwards instantaneously. This prevented q from rising above one and thus restricted the equilibrium dynamics. It is of interest to check how the results change when one restricts the rate of investment. This would slow down the adjustment of the capital stock, generating a lag in the industry's response to demand shocks. This modification of the model also serves as a check of robustness of the main results.

Specifically, assume that the capital can be transferred at most at rate $i_{\max}K_2$. Thus,

Parameters		Moments	
δ	σ_{21}	$E[\beta_{P,K_1}]$	$STD[\beta_{P,K_1}]$
<u>$\gamma = 2.0$</u>			
-0.03	0.00	1.62	0.43
-0.02	0.00	1.36	0.47
-0.03	0.08	1.20	0.16
-0.02	0.08	1.09	0.17
-0.03	0.16	0.88	0.14
-0.02	0.16	0.97	0.14
<u>$\gamma = 0.5$</u>			
-0.03	0.00	0.35	0.14
-0.02	0.00	0.31	0.14
-0.03	0.08	0.76	0.05
-0.02	0.08	0.74	0.05
-0.03	0.16	1.21	0.04
-0.02	0.16	1.22	0.04

Table 3.2: The first two unconditional moments of the beta of the industry portfolio are tabulated for several sets of parameter values. The following parameter values are used throughout the table: $\rho = 0.05$, $\alpha = 0.067$, $\sigma_1 = 0.124$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$.

the capital dynamics is described by

$$dK_{1t} = (\alpha K_{1t} - c_{1t})dt + \sigma_1 K_{1t} dW_{1t} - i_t dt, \quad (3.68)$$

$$dK_{2t} = -\delta K_{2t} dt + \sigma_{21} K_{2t} dW_{1t} + \sigma_2 K_{2t} dW_{2t} + i_t dt, \quad (3.69)$$

where $i_t \equiv dI_t/dt \in [0, i_{\max} K_{2t}]$. Without going into the details of the analysis, I only outline the optimal policy and some of the asset-pricing implications.

The optimal investment policy is of the “bang-bang” type: $i_t = i_{\max}$ for $\Omega \leq \Omega^*$ and $i_t = 0$ otherwise. This policy closely resembles the singular investment policy of the basic model. The optimal consumption policy is characterized by the envelope condition

$$U_{c_1}(c_1^*, c_2^*) = J_{K_1}(K_1, K_2).$$

Average q in the model still equals marginal q . To see this, note that the constraint on the rate of investment can be equivalently reformulated as an adjustment cost $c(i, K)$, which equals zero for i/K less than i_{\max} and is infinite otherwise. This adjustment cost function is linearly homogeneous. Combined with constant returns to scale, this equalizes average and marginal q (see Abel and Eberly (1994, Lem. 2)). Unlike in the basic case, however, q can exceed one. The logic behind this is straightforward. Before, q could not rise above one because firms would immediately invest (enter the industry), drive down the output price and, ultimately, reduce q to one. Given that instantaneous adjustment is impossible, when q exceeds one, firms invest at the maximum possible rate, but they cannot reduce q to one instantaneously. As a result of industry-specific and market-wide shocks, q can even increase temporarily. However, with sufficiently high rate of adjustment, profitable investment opportunities (high q) quickly get arbitrated away. At the same time, since

investment is irreversible, low q 's can persist, since there is no arbitrage pressure in this case. Thus, as in the basic model, dynamics of q is asymmetric, albeit less extreme.

To illustrate the behavior of asset prices, I plot the beta of stock returns as a function of q (in the limit of b approaching zero, to obtain a closed-form expression) in Figure 3-6. Note that the behavior of the beta for $q \leq 1$ closely resembles the one in the basic case, as illustrated in Figure 3-5. The main difference is that in this case q can exceed one and the beta is an increasing function of q for $q > 1$. This suggests that it is unlikely that the dependence of the beta on q can be adequately captured by a linear econometric model, since, at the very least, it can be expected to be non-monotone.

One can understand the time-variation of the beta in terms of the time-variation of q . In particular, one the state space can be partitioned into the following three regions.

First region. Low values of q : $q \ll 1$. Firms do not invest and irreversibility prevents them from disinvesting. Thus, the elasticity of supply is relatively low and q is relatively volatile.

Second region. Intermediate values of q : $q \simeq 1$. Firms are either about to invest, following an increase in q , or are already investing at the maximum possible rate and are about to stop, following a decline in q . The elasticity of supply is relatively high and, as a result, q is not sensitive to shocks and does not contribute much to the stock returns (see the discussion in the beginning of this section).

Third region. High values of q : $q \gg 1$. The industry is expanding. Firms are investing at the maximum possible rate and are likely continue investing during an extended period of time. Market-wide shocks (demand shocks) do not immediately change the rate of entry into the industry, the elasticity of supply is low and demand shocks are offset mostly by changes in the output price. Thus, q is relatively volatile.

This argument is illustrated by plotting q against its argument $\xi - \xi^*$ in Figure 3-7.

When the maximum possible rate of investment is very high, q rarely exceeds one. Thus, the third regime can be observed only infrequently, during periods of active growth of the industry. In the extreme case of instantaneous adjustment, as in the basic model, this regime is completely absent. The first two regimes, however, can still be identified.

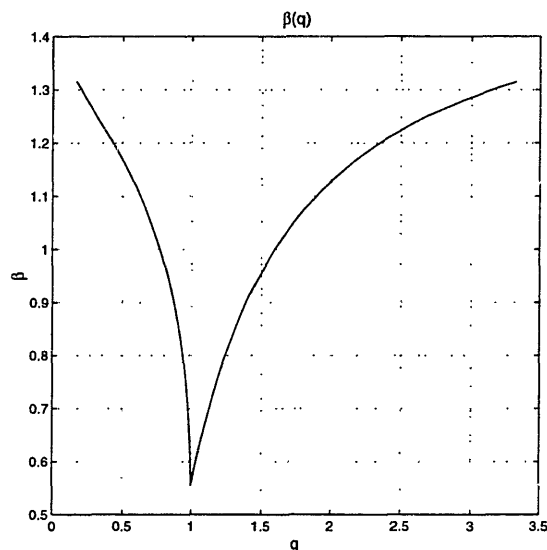


Figure 3-6: The beta of the industry portfolio is plotted as a function of the industry's q . The following set of parameter values is used: $\gamma = 2$, $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.06$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$, $i_{\max} = 1$.

Partial irreversibility and variable costs. The assumption of complete irreversibility is relatively easy to relax. For example, one can assume that capital can be transferred back from sector 2 to sector 1, but a fixed proportion would be lost during the transfer. This would impose a lower bound on q in the model, since firms would find it optimal to disinvest when q becomes sufficiently low. As a result, supply would become relatively elastic for values of q close to the lower bound, reducing the volatility of q in this region of the state space. A similar behavior would result if one introduced variable production costs into the model, allowing firms to leave their capital unused when the price of their output

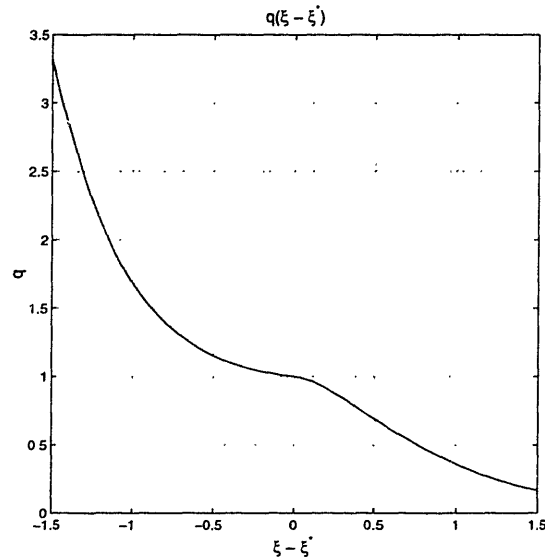


Figure 3-7: q is plotted against its argument, $\xi - \xi^*$. The following set of parameter values is used: $\gamma = 2$, $\rho = 0.05$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.06$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$, $i_{\max} = 1$.

is too low to cover their production costs.

Preferences. Separability of preferences across the consumption goods can be relaxed as well. The resulting analysis would be particularly tractable when goods are either perfect substitutes, or perfect complements. Such modifications, however, would not alter the main qualitative results of the paper.

Cross-section of industries and heterogeneous firms. An equilibrium model of irreversible investment with multiple sectors or with firm heterogeneity within the same industry would provide a significant extension of the model in this paper. Such a model would allow one to study cross-sectional properties of stock returns, as well as the joint dynamics of interest rates, real investment and stock returns. However, a straightforward modification of the model of this paper leads to serious technical difficulties, associated with singular control problems in higher dimensions. Thus, this particular extension would most likely require an alternative modelling strategy.

Strategic behavior. Another modification of the basic model would be to abandon the assumption of perfect competition within the industry, introducing elements of strategic behavior. This would change firms' investment policies, affecting the dynamics of capital accumulation. The objective would be to identify the effect of firms' strategic behavior in the product market on the behavior of associated financial assets. I leave this extension for future research.

3.8 Conclusion

In this paper, I have developed a general equilibrium model, emphasizing the effect of irreversibility of real investment on the behavior of financial asset prices. The interaction between the demand side and the supply side of the economy leads to a structural relation between real economic variables and properties of stock returns. As a result, the volatility of stock returns is stochastic and is a function of the state of the economy. In particular, irreversibility explains why stock-return volatility might rise as the stock price falls and how the systematic risk of stock returns changes over the business cycle, as well as in response to industry-specific shocks. My model also links stock returns to real investment. Because of irreversibility, market-to-book ratio (Tobin's q) plays an important role as a proxy for real variables, such as investment. It explains the time-variation in stock-return volatility, systematic risk and expected returns.

At this point it remains to be seen whether the implications of my model are supported by the data. I plan to pursue the empirical analysis and certain theoretical extensions in future research.

3.9 Appendix

3.9.1 Futures Prices

In this section I study properties of futures prices (futures are contracts on the industry's output). As I have pointed out in the introduction, the number of commodities for which futures contracts are traded is relatively small and in many cases futures prices appear to be driven by dynamics of inventories. My model would apply to contracts on commodities for which storage is relatively expensive, such as contracts on electricity, natural gas, etc. Thus, my model complements the models based on storage (inventory) models (e.g., Williams and Wright (1991)) and those with exogenous specification of convenience yield (e.g., Gibson and Schwartz (1990)).

Futures prices in my model are often in backwardation and the level of backwardation is positively related to q . The volatility of futures prices decreases with the contract horizon, which is consistent with the Samuelson proposition (see Samuelson (1965)). "Conditional violations" of the Samuelson proposition can be observed for sufficiently low values of q : the difference in volatility between contracts of different horizons is positively correlated with the degree of backwardation.

The Dynamics of the Spot Price

As it has been established in Section 3.4.3, the spot price of a unit of output produced by the industry (*the commodity*) equals the marginal rate of substitution of good 2 for good 1, i.e.,

$$S = \frac{U_{c_2}(c_1^*, c_2^*)}{U_{c_1}(c_1^*, c_2^*)} = b(\tilde{c}_1^*)^\gamma \Omega^{-\gamma}.$$

Unlike the stock price, this process generally has a singular component in equilibrium. The spot price follows a nonlinear mean-reverting process, due to the fact that there exists an upper bound on the spot price. In case of $b \ll 1$,

$$S = \lambda_1^{-1} \Xi_{(0)}^{-\gamma} \exp(-\gamma(\xi - \xi^*)) + O(b^{1/\gamma}), \quad (3.70)$$

and thus the spot price cannot exceed $\lambda_1^{-1} \Xi_{(0)}^{-\gamma}$. When below the upper bound, the spot price behaves as a Geometric Brownian Motion. Also, periods of higher q (higher investment) are associated with higher levels of the spot price.

If investment was perfectly reversible, the commodity spot price would be constant:

$$S_t = \frac{U_{c_2}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{1t}^*, c_{2t}^*)} = \frac{1}{\lambda} (1 - \max(\theta^*, 0))^{-\gamma}.$$

This suggests that the nonlinear dynamics of the spot price in the model is due entirely to irreversibility.

Ignoring terms of order $b^{1/\gamma}$ and higher, one can characterize the long-run stationary distribution of the spot price explicitly. The stationary distribution of $\xi_t - \xi^*$ is exponential with mean $1/m_\xi \equiv -\sigma_\xi^2 / (2\mu_\xi)$, where μ_ξ and σ_ξ are the drift and the diffusion coefficients of ξ_t . Therefore, the stationary distribution of the spot price is given by

$$\frac{m_\xi}{\gamma} \left(\lambda_1^{-1} \Xi_{(0)}^{-\gamma} \right)^{-\frac{m_\xi}{\gamma}} S^{\frac{m_\xi}{\gamma} - 1}, \quad S \in (0, \lambda_1^{-1} \Xi_{(0)}^{-\gamma}].$$

Depending on the model parameters, this distribution function can be either increasing or decreasing in its argument. This is reflected in the time-series properties of the spot price.

When $m_\xi/\gamma < 1$, the spot price tends to stay relatively low most of the time, exhibiting

occasional “spikes”. Such behavior is consistent with the data, as reported in Williams and Wright (1991, Ch. 6). Another important empirical property of the spot price is persistence: annual time series exhibit high first-order autocorrelation. This property is clearly implied by my model, since the spot price is a function of a persistent state variable. To illustrate these observations, I plot a single path of annual spot prices for a specific set of model parameters in Figure 3-8.

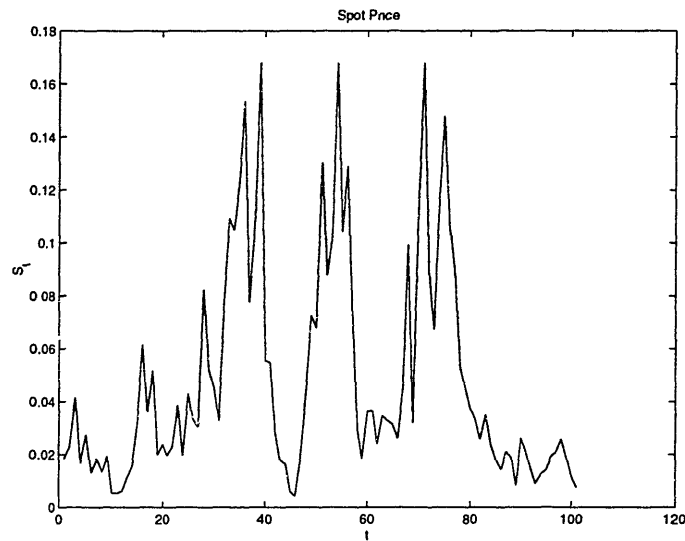


Figure 3-8: A time series of the spot price. The following set of parameter values is used: $\rho = 0.05$, $\gamma = 2$, $\alpha = 0.067$, $\delta = -0.03$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$.

The Term Structure and Volatility of Futures Prices

Let $\Phi_{t,s}$ denote the futures price at time t .¹⁵ Using the results in Cox, Ingersoll and Ross (1981), Richard and Sundaresan (1981) and Duffie and Stanton (1988), one can represent $\Phi_{t,s}$ as the time- t market value of $\int_t^s r_u du$ units of time- s commodity (good 2). The following proposition provides an explicit expression for the futures price in the limit of b approaching zero.

¹⁵The definition of futures contracts is given, for example, in Cox, Ingersoll and Ross (1981)

Proposition 3.7 *In the limit of b approaching zero, the futures price is given by*

$$\begin{aligned}\Phi_{t,s}(x) &= \lambda_1^{-1} \Xi_{(0)}^{-\gamma} (I_1 + I_2 + I_3), \\ I_1 &= \exp \left(-\gamma \left(x + \left(\mu_{\hat{\xi}} - \frac{\gamma (\sigma_{\hat{\xi}})^2}{2} \right) (s-t) \right) \right) G \left(\frac{x + \left(\mu_{\hat{\xi}} - \gamma (\sigma_{\hat{\xi}})^2 \right) (s-t)}{\sigma_{\hat{\xi}} \sqrt{s-t}} \right), \\ I_2 &= \frac{\gamma (\sigma_{\hat{\xi}})^2}{\gamma (\sigma_{\hat{\xi}})^2 - 2\mu_{\hat{\xi}}} \exp \left(\frac{\gamma (\sigma_{\hat{\xi}})^2 - 2\mu_{\hat{\xi}}}{(\sigma_{\hat{\xi}})^2} \left(x + \frac{\gamma (\sigma_{\hat{\xi}})^2}{2} (s-t) \right) \right) \\ &\quad G \left(\frac{-x + \left(\mu_{\hat{\xi}} - \gamma (\sigma_{\hat{\xi}})^2 \right) (s-t)}{\sigma_{\hat{\xi}} \sqrt{s-t}} \right), \\ I_3 &= \frac{-2\mu_{\hat{\xi}}}{\gamma (\sigma_{\hat{\xi}})^2 - 2\mu_{\hat{\xi}}} G \left(\frac{-x - \mu_{\hat{\xi}} (s-t)}{\sigma_{\hat{\xi}} \sqrt{s-t}} \right).\end{aligned}$$

If parameters of the model are such that

$$-\alpha - \delta + \lambda_1^{-1/\gamma} + \frac{\sigma_1^2}{2} - \frac{\sigma_{21}^2 + \sigma_2^2}{2} - \gamma \sigma_1 (\sigma_{21} - \sigma_1) < 0,$$

the price of contracts of long (“infinite”) maturity satisfies

$$0 < \lim_{s \rightarrow \infty} \Phi_{t,s}(x) = \frac{-2\mu_{\hat{\xi}} \lambda_1^{-1} \Xi_{(0)}^{-\gamma}}{\gamma (\sigma_{\hat{\xi}})^2 - 2\mu_{\hat{\xi}}} < \lambda_1^{-1} \Xi_{(0)}^{-\gamma}. \quad (3.71)$$

Otherwise,

$$\lim_{s \rightarrow \infty} \Phi_{t,s}(x) = 0.$$

Proof. Given the representation of the futures price as the time- t market value of $\int_t^s r_u du$ units of time- s commodity, $\Phi_{t,s} = \widehat{E}_t [S_s]$, where \widehat{E} denotes expectation under the

risk-neutral probability measure, \mathcal{Q} .¹⁶ In order to compute the futures price, one needs to specify the behavior of the state variable ξ_s under \mathcal{Q} . As it has been established in Section 3.6.2, in the limit of b approaching zero, there is no market risk premium associated with the industry-specific source of risk W_{2t} . The risk premium on W_{1t} can be extracted from returns on the market portfolio: since the short-term risk-free rate equals $\alpha - \gamma\sigma_1^2$, the risk premium is equal to $\gamma\sigma_1$. Thus, under \mathcal{Q} , the state variable ξ_t is equal in distribution to the solution of the following stochastic differential equation with instantaneous reflection at ξ^* :

$$\begin{aligned} d\hat{\xi}_s &= \mu_{\hat{\xi}} ds + \sigma_{\hat{\xi}} d\widehat{W}_s, \quad \hat{\xi}_s \geq \xi^*, \quad \hat{\xi}_t = \xi_t, \\ \mu_{\hat{\xi}} &= -\alpha - \delta + \lambda_1^{-1/\gamma} + \frac{\sigma_1^2}{2} - \frac{\sigma_{21}^2 + \sigma_2^2}{2} - \gamma\sigma_1(\sigma_{21} - \sigma_1), \\ \sigma_{\hat{\xi}} &= \sqrt{(\sigma_{21} - \sigma_1)^2 + \sigma_2^2}, \end{aligned}$$

where \widehat{W}_t is a Brownian motion under \mathcal{Q} .¹⁷ One can now represent the futures price as

$$\Phi_{t,s}(\xi_t - \xi^*) = \widehat{\mathbb{E}}_t \left[\lambda_1^{-1} \Xi_{(0)}^{-\gamma} \exp \left(-\gamma \left(\hat{\xi}_s - \xi^* \right) \right) \right]. \quad (3.72)$$

Observe that the futures price is proportional to the Moment Generating Function of the conditional distribution of $\hat{\zeta}_s \equiv \hat{\xi}_s - \xi^*$ (specifically, conditional on $\hat{\zeta}_t$), which can be computed in closed form using the expression for the corresponding cumulative distribution

¹⁶One cannot define the Equivalent Martingale Measure on the original space $(\Omega, \mathcal{F}, \mathcal{P})$ (e.g., Karatzas and Sreve (1991, p. 192) or Chung and Williams (1990, pp. 210-212)). Instead, following the construction in these references, I define the martingale measures \mathcal{Q} on the unaugmented filtration \mathcal{F}_t^o , generated by (W_{1t}, W_{2t}) , as the probability measure on (Ω, \mathcal{F}^o) , such that its restriction on \mathcal{F}_t^o is equivalent to the corresponding restriction of \mathcal{P} for all finite t . An alternative would be to define measures \mathcal{Q}_s on (Ω, \mathcal{F}_s) for every finite s . This would allow one to compute futures prices, since futures contracts are finitely lived.

¹⁷See (Karatzas, Lehoczky and Shreve, 1991) for a formal discussion of an equivalent change of measure for Itô processes with a singular component.

function (see Harrison, (1990, §3.6)):

$$\text{Prob}_{\hat{\zeta}_t} \left\{ \hat{\zeta}_s < y \right\} = 1 - G \left(\frac{-y + \hat{\zeta}_t + \mu_{\hat{\zeta}}(s-t)}{\sigma_{\hat{\zeta}}\sqrt{s-t}} \right) - e^{2\mu y/\sigma^2} G \left(\frac{-y - \hat{\zeta}_t + \mu_{\hat{\zeta}}(s-t)}{\sigma_{\hat{\zeta}}\sqrt{s-t}} \right), \quad (3.73)$$

where $G(\cdot)$ is the standard normal cumulative distribution function. The first result of the proposition is then obtained by direct calculation.

If parameters of the model are such that $\mu_{\hat{\zeta}} < 0$,¹⁸ $\hat{\zeta}_s$ has the long-run stationary distribution characterized by the exponential distribution function with scale parameter $-(\sigma_{\hat{\zeta}})^2 / (2\mu_{\hat{\zeta}})$. This allows one to compute the limit of the futures price directly as the Moment Generating Function of the exponential distribution. One can also obtain the result as a mechanical limit of the general expression. If $\mu_{\hat{\zeta}} \geq 0$,

$$\lim_{s \rightarrow \infty} \text{Prob}_{\hat{\zeta}_t} \left\{ \hat{\zeta}_s < y \right\} = 0, \quad \forall y$$

and the result follows. ■

This proposition provides a number of testable implications for the behavior of futures contracts.

1. Futures contracts can exhibit strong backwardation, i.e., the futures price can be less than the prevailing spot price. The existence of the long-run stationary distribution of the spot price implies that the futures are in backwardation whenever the prevailing spot price falls above its long-run mean. Backwardation is not restricted to

¹⁸If $\mu_{\hat{\zeta}} \geq 0$, there does not exist a long-run distribution: $\lim_{s \rightarrow \infty} \text{Prob}_t[\hat{\zeta} < \varepsilon] = 0$ for any positive ε , and thus $\lim_{s \rightarrow \infty} \Phi_{t,s}(x) = 0$. This is a manifestation of the fact that \mathcal{Q} and \mathcal{P} are mutually singular on (Ω, \mathcal{F}^o) (e.g., Karatzas and Sreve (1991, p. 192) or Chung and Williams (1990, pp. 210–212)). Thus, it is possible for a process to be ergodic under one measure, but not under the other.

long-horizon contracts. For instance, when $q = 1$, the prevailing spot price is at its maximum and, since $\Phi_{t,s} = \widehat{E}_t[S_s]$, the futures price is guaranteed to fall below the spot price regardless of the maturity of the contract. By continuity, futures will be in backwardation whenever q is sufficiently high.

2. For the long-horizon futures contracts the level of backwardation is positively correlated with the prevailing spot price. This is immediate, since the futures price is constant for long-horizon contracts.
3. Volatility of the futures prices is higher for short-horizon contracts than for very long-horizon contracts. Again, this follows from the futures price being constant for long-maturity contracts: the volatility of long-horizon contracts is zero, while the volatility of the spot price equals $\gamma\sigma_{\hat{\xi}}$.¹⁹
4. The level of backwardation is positively correlated with q . This follows directly from (3.71).

The first three implications are largely consistent with known empirical facts (e.g., Milonas (1986), Litzenberger and Rabinowitz (1995))²⁰ and have been generated in a number of theoretical models. The difference between my approach and most of the existing results is the absence of storage technology in my model. Thus, one cannot appeal to traditional arguments based on either exogenous convenience yield (e.g., Gibson and Schwartz (1990), Brennan (1991), Schwartz (1997)) or nonnegativity of inventories in competitive storage models (e.g., Williams and Wright (1991), Deaton and Laroque (1996), Routledge,

¹⁹This is a special case of the Samuelson proposition (Samuelson (1965)), which states that volatility of futures prices is decreasing with contract maturity.

²⁰See, however, Fama and French (1988) for examples of conditional violations of the Samuelson proposition.

Seppi and Spatt (1998)). The main conclusion of this analysis is that many of known properties of futures prices can arise due to irreversibility of physical investment alone. Whether or not irreversibility is an empirically important factor depends, among other things, on the actual cost of storage for a particular commodity.

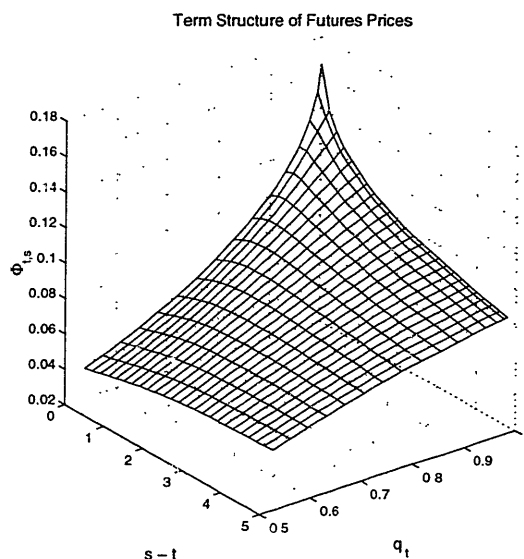


Figure 3-9: The term structure of futures prices is plotted as a function of q . The following set of parameter values is used: $\rho = 0.05$, $\gamma = 2$, $\alpha = 0.067$, $\delta = -0.02$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$.

The last implication is new. It can be understood as follows. When q is sufficiently high, firms invest and increase their output. As a result, the *future* spot price tends to be lower than the prevailing spot price, since the industry cannot reduce the rate of its output in response to a negative demand shock. This is reflected in lower futures prices and higher level of backwardation.

Term Structure of Volatility of Futures Prices

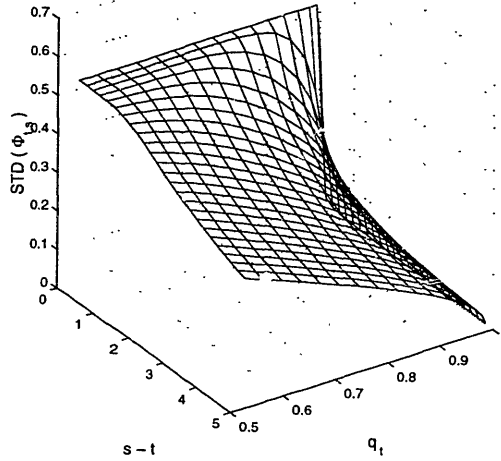


Figure 3-10: The term structure of volatility of futures prices is plotted as a function of q . The following set of parameter values is used: $\rho = 0.05$, $\gamma = 2$, $\alpha = 0.067$, $\delta = -0.02$, $\sigma_1 = 0.124$, $\sigma_{21} = 0.04$, $\sqrt{\sigma_{21}^2 + \sigma_2^2} = 0.25$.

I plot the term structure of futures prices, $\Phi_{t,s}$, and volatility of futures prices,²¹

$$\text{STD}(\Phi_{t,s}) \equiv \sqrt{E_t \left[(d_t \Phi_{t,s} / \Phi_{t,s})^2 \right] / dt},$$

for a specific set of parameter values in Figures 3-9, 3-10.

As one can see from Figure 3-9, the term structure of futures prices can have several shapes. It can be entirely in backwardation (monotonically decreasing) for large q , entirely in contango (monotonically increasing) for small q and it can be hump-shaped for intermediate values of q . Note also that the futures price increases in q , which is quite intuitive, since the spot price is an increasing function of q . To see that this is a general property of futures prices in my model (and not just a peculiarity of the numerical example), note that, according to (3.73), the distribution of ζ_s under \mathcal{Q} “increases” with ζ_t in the sense of

²¹Before applying the Itô formula to the futures price, one must complete the probability space \mathcal{F}^Q and augment the filtration \mathcal{F}_t with the \mathcal{Q} -null sets (see Chung and Williams (1990, p. 212)).

the First-Order Stochastic Dominance. Since the spot price is a monotonically decreasing function of the state variable, this implies that the futures price decreases with ζ_t (increases with q).

Figure 3-10 suggests that the volatility of futures prices decreases monotonically with the contract horizon. However, the relative difference in volatility between short- and long-horizon contracts depends on the current state of the economy. In particular, consistently with known empirical facts (see Fama and French (1988)), this difference is positively correlated with the degree of backwardation, leading to “conditional violations” of the Samuelson proposition. This behavior is explained by elasticity of supply being a function of q .

Figure 3-10 also points to another property of futures prices: their volatility decreases with q .²² Both of these features have been observed for a wide range of parameter values. The latter observation is a testable prediction that has not been previously considered in the literature. The relation between q and the volatility of futures prices suggests that the latter should change in response to market-wide and industry-specific shocks. In particular, futures prices are predicted to become more volatile in response to negative shocks to the market, as long as $\beta_{K_2, K_1} < 1$. This follows from the fact that, if $\beta_{K_2, K_1} < 1$, q decreases in response to negative shocks to the market. Independently of the parameter values, the volatility of futures prices increases in response to positive idiosyncratic (industry-specific) shocks. The reason for this is that the industry’s q must drop in response to positive idiosyncratic shocks. The bottom line is that both the market-wide and industry-specific shocks affect the term structure of futures prices and the volatility of futures prices through

²²In general, the volatility of the futures price equals zero when q equals one. This follows from the representation $\Phi_{t,s} = \hat{E}_t[S_s]$ and the fact that the state variable, ξ_s , follows a reflected diffusion process. According to Gihman and Skorohod (1972, p. 184), the first derivative of $\Phi_{t,s}$ with respect to ξ equals zero at ξ^* . Thus, the instantaneous volatility of $\Phi_{t,s}$ equals zero when $\xi = \xi^*$ (when q equals one). At the same time, as q approaches zero, the volatility of the futures price approaches the volatility of the spot price, given by $\gamma\sigma_{\xi}$.

their impact on the industry's q .

3.9.2 Proof of Proposition 3.3

To demonstrate that the conditions of the proposition are sufficient, I construct a lower and an upper bound on the value function and show that the conditions of the proposition imply that both bounds are finite.

Consider the lower bound first. It is provided by the solution of the problem in which the two sectors are completely isolated from each other and the capital stock is initially split evenly among the sectors. Consider the problem for an isolated sector 1:

$$\max_{\{c_t\}} E_0 \left[\int_0^{\infty} e^{-\rho t} \frac{1}{1-\gamma} c_t^{1-\gamma} dt \right], \quad \gamma \neq 1, \quad (3.74)$$

subject to

$$dK_t = (\mu K_t - c_t)dt + \sigma K_t dW_t,$$

$$K_t \geq 0, \quad c_t \geq 0.$$

The value function, $J_1(K)$, satisfies the dynamic programming equation²³

$$\rho J_1 = \max_{c \geq 0} \left\{ \frac{1}{1-\gamma} c^{1-\gamma} + J_{1K}(\mu K - c) + \frac{1}{2} J_{1KK} \sigma^2 K^2 \right\}. \quad (3.75)$$

²³Here and in Appendix 3.9.7, I solve the dynamic programming equations in closed form. To verify that the resulting solution is indeed the value function of the original optimization problem, one can use the standard verification theorem, e.g., Fleming and Soner (1992, Th. 9.1).

The solution of (3.75) is given by

$$\begin{aligned}
J_1(K) &= \frac{\lambda_1(\mu, \sigma, \rho)}{1-\gamma} K^{1-\gamma}, \\
c_t^* &= (\lambda_1)^{-1/\gamma} K_t, \\
\lambda_1(\mu, \sigma, \rho) &= \left(\mu \frac{\gamma-1}{\gamma} - \frac{\sigma^2}{2}(\gamma-1) + \frac{\rho}{\gamma} \right)^{-\gamma}.
\end{aligned} \tag{3.76}$$

Constraints $K_t \geq 0$, $c_t \geq 0$ are clearly satisfied. This verifies (3.9). For the case $\gamma = 1$, it is easy to check that the value function has the form $J_1(K) = \rho^{-1} \ln(K) + \text{const.}$

For an isolated sector 2, one must compute

$$\mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{b}{1-\gamma} K_t^{1-\gamma} dt \right], \quad \gamma \neq 1, \tag{3.77}$$

subject to

$$dK_t = \mu K_t dt + \sigma K_t dW_t.$$

Since $K_t = K_0 \exp((\alpha - \sigma^2/2)t + \sigma(W_t - W_0))$, the solution is given by

$$\begin{aligned}
J_2(K) &= b \frac{\lambda_2(\mu, \sigma, \rho)}{1-\gamma} K^{1-\gamma}, \\
\lambda_2(\mu, \sigma, \rho) &= \left(\mu(\gamma-1) - \frac{\sigma^2}{2}\gamma(\gamma-1) + \rho \right)^{-1}.
\end{aligned} \tag{3.78}$$

Similarly, for $\gamma = 1$, $J_1(K) = b\rho^{-1} \ln(K) + \text{const.}$

Thus, the lower bound on the value function is given by

$$J^{LB}(K_1, K_2) = \frac{\lambda_1}{1-\gamma} K_1^{1-\gamma} + b \frac{\lambda_2}{1-\gamma} K_2^{1-\gamma}, \quad \gamma \neq 1, \tag{3.79}$$

$$J^{LB}(K_1, K_2) = \frac{1}{\rho} \ln(K_1) + \frac{b}{\rho} \ln(K_2) + \frac{1}{\rho^2} \left((\alpha - b\delta) - \frac{1}{2} (\sigma_1^2 + \sigma_{21}^2 + \sigma_2^2) + \rho(\ln(\rho) - 1) \right), \quad \gamma = 1. \quad (3.80)$$

The lower bound is finite if and only if λ_1 and λ_2 are finite.

An upper bound on the value function can be obtained by adding the solutions of the following two problems:

$$\max_{\{c_t, \theta_t\}} E_0 \left[\int_0^{\infty} e^{-\rho t} \frac{1}{1-\gamma} c_t^{1-\gamma} dt \right], \quad (3.81)$$

and

$$\max_{\{c_t, \theta_t\}} E_0 \left[\int_0^{\infty} e^{-\rho t} \frac{b}{1-\gamma} K_t^{1-\gamma} dt \right], \quad (3.82)$$

both subject to the same constraints as the original problem. The idea behind these two problems is clear: one computes an upper bound by maximizing each of the components of the original objective function separately. Economically, each of these auxiliary problems corresponds to maximizing the utility of consumption of only one of the goods. The upper bound is then given by

$$J^{UB}(K_1, K_2) = J^{LB}(K_1, K_2 + K_1). \quad (3.83)$$

To see this, note that the problem (3.81) is equivalent to (3.74), since it is impossible to transfer capital from the second sector to the first. The problem (3.82) is equivalent to

(3.77) with the initial condition $K_2 + K_1$, since it is clearly optimal to transfer all the capital from the first sector to the second. Thus,

$$J^{UB}(K_1, K_2) = J_1(K_1, K_2) + J_2(K_1, K_2 + K_1).$$

To see that the conditions of the proposition are necessary, note the following: for $\gamma < 1$, both the value function and the lower bound are positive, thus the lower bound must be finite; similarly, for $\gamma > 1$, both the value function and the upper bound are negative, thus the upper bound must be finite; for $\gamma = 1$, the value function is always finite and conditions (3.9, 3.10) are trivially satisfied. ■

3.9.3 Numerical Solution of (3.16–3.18, 3.24, 3.25)

I design a sequence of finite difference approximations to the Hamilton-Jacobi-Bellman equation and the boundary conditions. First, introduce the new independent variable $\Xi = b^{-1/\gamma}\Omega$, $\xi = \ln(\Xi) = \omega - \ln(b)/\gamma$ and the new unknown function

$$g(\xi) \equiv b^{-1/\gamma} \left(f(\xi) - \lambda_1 - b^{1/\gamma} \lambda_2 \exp((1 - \gamma) \xi) \right). \quad (3.84)$$

According to (3.105), $g(\xi)$ satisfies

$$p_2 g'' + p_1 g' + p_0 g = -\gamma b^{-1/\gamma} \left(\lambda_1 + b^{1/\gamma} \left(g - \frac{1}{1 - \gamma} g' \right) \right)^{1-1/\gamma} - \lambda_1 p_0 b^{-1/\gamma}. \quad (3.85)$$

The boundary conditions (3.17, 3.18) translate into

$$g'(\xi^*) \left(1 + b^{1/\gamma} e^{\xi^*}\right) - g(\xi^*) (1 - \gamma) b^{1/\gamma} e^{\xi^*} = \lambda_1 (1 - \gamma) e^{\xi^*} - \lambda_2 (1 - \gamma) e^{(1-\gamma)\xi^*}, \quad (3.86)$$

$$g''(\xi^*) \left(1 + b^{1/\gamma} e^{\xi^*}\right) - g'(\xi^*) \left(1 + (1 - \gamma) b^{1/\gamma} e^{\xi^*}\right) = \lambda_2 (1 - \gamma) \gamma e^{(1-\gamma)\xi^*}. \quad (3.87)$$

For a given value of ξ^* , introduce a discrete grid $\{\xi_1, \xi_2, \dots, \xi_I\}$, such that $\xi_{i+1} - \xi_i = h$ and $\xi_1 = \xi^* - h$. A finite difference approximation to (3.85) is

$$\begin{aligned} & p_2 \frac{g_{i+1} - 2g_i + g_{i-1}}{h^2} + p_1 \frac{g_{i+1} - g_{i-1}}{2h} + p_0 g_i = \\ & -\gamma b^{-1/\gamma} \left(g_i - \frac{1}{1-\gamma} \frac{g_{i+1} - g_{i-1}}{2h} \right)^{1-1/\gamma} - \lambda_1 p_0 b^{-1/\gamma}, \\ & i = 2, \dots, I-1, \end{aligned} \quad (3.88)$$

while the boundary conditions (3.86, 3.87) are approximated as

$$\begin{aligned} & \frac{g_3 - g_1}{2h} \left(1 + b^{1/\gamma} e^{\xi^*}\right) - g_1 (1 - \gamma) b^{1/\gamma} e^{\xi^*} - \\ & \lambda_1 (1 - \gamma) e^{\xi^*} + \lambda_2 (1 - \gamma) e^{(1-\gamma)\xi^*} = 0, \end{aligned} \quad (3.89)$$

$$\begin{aligned} & \frac{g_3 - 2g_2 + g_1}{h^2} \left(1 + b^{1/\gamma} e^{\xi^*}\right) - \frac{g_3 - g_1}{2h} \left(1 + (1 - \gamma) b^{1/\gamma} e^{\xi^*}\right) - \\ & \lambda_2 (1 - \gamma) \gamma e^{(1-\gamma)\xi^*} = 0. \end{aligned} \quad (3.90)$$

I solve the system (3.88–3.90) iteratively. Let g^{n-1} denote the solution of the n -th step of the iterative procedure. Then g^n is computed as a solution of

$$p_2 \frac{g_{i+1}^n - 2g_i^n + g_{i-1}^n}{h^2} + p_1 \frac{g_{i+1}^n - g_{i-1}^n}{2h} + p_0 g_i^n =$$

$$-\gamma b^{-1/\gamma} \max \left(0, g_i^{n-1} - \frac{1}{1-\gamma} \frac{g_{i+1}^{n-1} - g_{i-1}^{n-1}}{2h} \right)^{1-1/\gamma} - \lambda_1 p_0 b^{-1/\gamma},$$

$$i = 2, \dots, I-1,$$

$$\frac{g_3^n - g_1^n}{2h} \left(1 + b^{1/\gamma} e^{\xi^*} \right) - g_1^n (1-\gamma) b^{1/\gamma} e^{\xi^*} = \lambda_1 (1-\gamma) e^{\xi^*} - \lambda_2 (1-\gamma) e^{(1-\gamma)\xi^*},$$

$$g_{I-1}^n = g_I^n,$$

where the last equation accounts for the asymptotic behavior of $g(\omega)$ (see Appendix 3.9.5, (3.104)). Iteration steps are repeated until

$$\max_{i=1,2,\dots,I} \left\{ |g_i^n - g_i^{n-1}| + \max \left(0, -g_i^{n-1} + \frac{1}{1-\gamma} \frac{g_{i+1}^{n-1} - g_{i-1}^{n-1}}{2h} \right) \right\} < \varepsilon,$$

where ε controls the desired accuracy level. The initial iteration can be started with $g_i^0 \equiv 0$, or a better approximation, e.g. the one developed in Section 3.4.4.

After the convergence criterion has been satisfied, compute the error in (3.90) using g_i and repeat the above procedure as a part of a line search to solve (3.90). I is chosen large enough so that $g_I^n < \varepsilon$ (see 3.104)) and refine the grid by reducing the discretization parameter h until the solution of the finite difference scheme is no longer sensitive to h . Finally, I use Richardson's deferred approach to the limit (see Press, Flannery, Teukolsky and Vetterling (1992)) to obtain an accurate approximation to the solution of the original

problem, corresponding to “ $h = 0$ ”.

3.9.4 Proof of Proposition 3.4

The proof relies on the following two lemmas.

Lemma 3.8 *The optimal solution of the central planner’s problem satisfies*

$$E_t \left[\int_t^\infty e^{-\rho(s-t)} \frac{U_{c_2}(c_{1s}^*, c_{2s}^*)}{U_{c_1}(c_{1t}^*, c_{2t}^*)} \frac{R_{2s}}{R_{2t}} ds \right] = \frac{J_{K_2}(K_{1t}^*, K_{2t}^*)}{J_{K_1}(K_{1t}^*, K_{2t}^*)}, \quad (3.91)$$

where

$$R_{2t} = \exp \left(- (\delta + \sigma_{21}^2/2 + \sigma_1^2/2) t + \sigma_{21} W_{1t} + \sigma_2 W_{2t} \right). \quad (3.92)$$

Equivalently,

$$E_t \left[\int_t^\infty e^{-\rho(s-t)} U_{c_2}(c_{1s}^*, c_{2s}^*) \frac{R_{2s}}{R_{2t}} ds \right] = J_{K_2}(K_{1t}^*, K_{2t}^*). \quad (3.93)$$

Relation (3.91) has an intuitive interpretation. It states that the relative value of a marginal unit of capital 2, expressed in terms of the numeraire good (the shadow price of capital), equals the present value of the entire future output produced by the marginal unit, taking into account the equilibrium capital accumulation dynamics (the marginal q).

Proof. Since $U_{c_1}(c_{1s}^*, c_{2s}^*) = J_{K_1}(K_{1s}^*, K_{2s}^*)$, it is sufficient to establish (3.93).

Without loss of generality, assume that in (3.93) $t = 0$. Consider an economy with initial capital stocks (K_{10}, K_{20}) , $K_{20}/K_{10} \geq \Omega^*$. \tilde{c}_{1t}^* and I_t^* are stochastic processes arising as a result of applying the optimal consumption/investment policy. Let (K_{1t}^*, K_{2t}^*) be the resulting process for the capital stocks. Consider now another economy with the initial

condition $(K_{10}, K_{20} + \Delta)$. Implement the consumption/investment policy defined by the pair of stochastic processes $(\tilde{c}_{1t}^*, I_t^*)$ (this amounts to "setting aside" Δ units of capital good 2 and treating the remaining capital stocks as if Δ did not exist). Let the corresponding indirect utility function be \tilde{J} . By construction,

$$\tilde{J} = E_0 \left[\int_0^\infty e^{-\rho t} \left(\frac{1}{1-\gamma} (\tilde{c}_{1t}^* K_{1t}^*)^{1-\gamma} + \frac{b}{1-\gamma} (K_{2t}^* + \Delta R_{2t})^{1-\gamma} \right) dt \right].$$

Since such a consumption/investment policy is suboptimal, $\tilde{J} \leq J(K_{10}, K_{20} + \Delta)$. Thus,

$$\frac{J(K_{10}, K_{20} + \Delta) - J(K_{10}, K_{20})}{\Delta} \geq \frac{\tilde{J} - J(K_{10}, K_{20})}{\Delta}.$$

One concludes, by taking the limit of $\Delta \searrow 0$ and using the dominated convergence theorem to pass the limit under the integral sign, that

$$J_{K_2}(K_{10}, K_{20}) \geq E_0 \left[\int_0^\infty e^{-\rho t} b (K_{2t}^*)^{-\gamma} R_{2t} dt \right]. \quad (3.94)$$

Next, repeat the above argument for the economy with initial conditions $(K_{10}, K_{20} - \Delta)$, $\Delta < K_{20}$. Implementing $(\tilde{c}_{1t}^*, A_t^*)$ for this economy leads to the indirect utility function

$$\tilde{J} = E_0 \left[\int_0^\infty e^{-\rho t} \left(\frac{1}{1-\gamma} (\tilde{c}_{1t}^* K_{1t}^*)^{1-\gamma} + \frac{b}{1-\gamma} (K_{2t}^* - \Delta R_{2t})^{1-\gamma} \right) dt \right].$$

Note that $K_{2t}^* > \Delta K_{2s}^*/K_{2t}^*$, due to $\Delta < K_{20}$. Suboptimality of the constructed strategy implies $\tilde{J} \leq J(K_{10}, K_{20} - \Delta)$. Therefore,

$$-\frac{J(K_{10}, K_{20} - \Delta) - J(K_{10}, K_{20})}{\Delta} \leq -\frac{\tilde{J} - J(K_{10}, K_{20})}{\Delta}$$

and the limit of $\Delta \searrow 0$ yields

$$J_{K_2}(K_{10}, K_{20}) \leq \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} b(K_{2t}^*)^{-\gamma} R_{2t} dt \right]. \quad (3.95)$$

Inequalities (3.94) and (3.95) imply

$$J_{K_2}(K_{10}, K_{20}) = \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} b(K_{2t}^*)^{-\gamma} R_{2t} dt \right],$$

which completes the proof. ■

Lemma 3.9 *The value function $J(K_1, K_2)$ satisfies*

$$J_{K_1 K_2} \leq 0, \quad \forall K_1 > 0, K_2 > 0. \quad (3.96)$$

The relation (3.96) can be justified intuitively. Note that the marginal unit of the numeraire good can be used in two ways: either for consumption, or for investment. Accordingly, its contribution to the indirect utility function can be decomposed into a sum of the discounted present value of the marginal utility of consumption and the value of the imbedded “investment option”. While the utility derived from consumption does not depend on the size of the capital stock K_2 , the value of the investment option does. Specifically, it decreases with K_2 , since so does the marginal indirect utility of K_2 . This explains why the marginal value of the numeraire good decreases with K_2 .

Proof. According to Lemma 3.8,

$$J_{K_2}(K_{1t}, K_{2t}) = \mathbb{E}_t \left[\int_t^\infty e^{-\rho(s-t)} b(K_{2s}^*)^{-\gamma} \frac{R_{2s}}{R_{2t}} dt \right].$$

I start by characterizing the dynamics of the capital stock K_2^* , resulting from the optimal investment policy I_t^* . According to (3.5) and (3.29),

$$K_{2s}^* = K_{2t}^* \frac{R_{2s}}{R_{2t}} \exp\left(\frac{1}{1 + \Omega^*} (L_s - L_t)\right), \quad t < s,$$

where R_{2t} is defined by (3.92). Then

$$J_{K_2}(K_{1t}, K_{2t}) = \mathbb{E}_t \left[\int_t^\infty e^{-\rho(s-t)} b(K_{2t})^{-\gamma} \left(\frac{R_{2s}}{R_{2t}}\right)^{1-\gamma} \exp\left(\frac{-\gamma}{1 + \Omega^*} (L_s - L_t)\right) dt \right]. \quad (3.97)$$

It is therefore sufficient to demonstrate that $L_s - L_t$ is a nondecreasing function of the initial condition K_{1t} . Consider a pair of initial conditions (K'_{1t}, K_{2t}) and (K''_{1t}, K_{2t}) , $K'_{1t} < K''_{1t}$. Denote the corresponding reflection processes L'_s and L''_s . I will show that $L'_s \leq L''_s$ almost surely.

To this end, define ω'_t and ω''_t as two solutions of (3.27) with initial conditions $\ln(K_{20}/K'_{10})$ and $\ln(K_{20}/K''_{10})$ respectively. Suppose there exists a finite moment of time s at which $\omega'_s > \omega''_s$. Since $\omega'_0 < \omega''_0$ and both solutions have almost surely continuous paths, there must exist $\tau \in (t, s)$, such that $\omega'_\tau = \omega''_\tau$. However, due to uniqueness of the solution of (3.27) (see Gihman and Skorohod (1972, §23, Th. 1)), this implies that $\omega'_u = \omega''_u$, for any $u > \tau$. I conclude that $\omega'_s \leq \omega''_s$ almost surely. Given the characterization of the reflection processes L'_t and L''_t as local times, this last inequality implies that $L'_t \leq L''_t$ almost surely.

■

Lemma 3.9 provides an important insight into the behavior of the optimal consumption policy. Since $\tilde{c}_1^* = (K_1)^{-1} (J_{K_1})^{-1/\gamma}$, it is clear that $(\tilde{c}_1^*)_{K_2} \geq 0$. This should not be surprising by now: when deciding how much to allocate for consumption, the central planner trades off the direct benefit of consumption (the marginal utility), which is independent of

K_2 , against the indirect benefit of adding a marginal unit of the numeraire good to K_1 , which has been shown to be a decreasing function of K_2 . This results in the following lemma.

Lemma 3.10 *The optimal consumption policy $\tilde{c}_1^*(\omega)$ is a nondecreasing function of its argument: $\tilde{c}_1^{*\prime}(\omega) \geq 0$.*

Being a nondecreasing function of its argument, $\tilde{c}_1^*(\omega)$ possesses a limit as ω approaches infinity. As suggested by (3.24, 3.25, 3.26), this limit is finite and equals $\lambda_1^{-1/\gamma}$ – the optimal consumption policy in the economy with completely isolated sectors. This result is formalized in the following lemma.

As the ratio of capital stocks increases, the optimal consumption policy $\tilde{c}_1^*(\omega)$ approaches a finite limit:

$$\lim_{\omega \rightarrow -\infty} \tilde{c}_1^*(\omega) = \lambda_1^{-1/\gamma} = \frac{\gamma-1}{\gamma} \alpha - \sigma_1^2 \frac{\gamma-1}{2} + \frac{\rho}{\gamma}. \quad (3.98)$$

Proof. To show that $\lim_{\omega \rightarrow \infty} \tilde{c}_1^*(\omega) = \lambda_1^{-1/\gamma}$, it is sufficient to prove that

$$\lim_{K_2 \rightarrow \infty} J_{K_1}(K_1, K_2) = \lambda_1 K_1^{-\gamma}.$$

Since $J^{LB}(K_1, K_2) \leq J(K_1, K_2) \leq J^{UB}(K_1, K_2)$, where $J^{LB}(\cdot, \cdot)$ and $J^{UB}(\cdot, \cdot)$ are given by (3.79, 3.80) and (3.83) respectively, one concludes that

$$\lim_{K_2 \rightarrow \infty} J(K_1, K_2) - \frac{b\lambda_2}{1-\gamma} K_2^{1-\gamma} = \frac{\lambda_1}{1-\gamma} K_1^{1-\gamma}, \quad (3.99)$$

convergence being uniform of compact subsets of $(0, \infty)$. This verifies (3.24–3.26). Combined with the fact that J_{K_1} is a nonincreasing function of K_2 (see Lemma 3.9), this allows

one to prove the statement of the lemma.

Consider an arbitrary value of K_1 and $h \in (0, K_1)$. Monotonicity of J_{K_1} and the mean value theorem imply

$$\frac{J(K_1 + h, K_2) - J(K_1, K_2)}{h} \leq J_{K_1}(K_1, K_2) \leq \frac{J(K_1, K_2) - J(K_1 - h, K_2)}{h}, \quad \forall K_2.$$

Combined with (3.99), this means that

$$\begin{aligned} \underline{\lim}_{K_2 \rightarrow \infty} J_{K_1}(K_1, K_2) &\geq \frac{\lambda_1}{1 - \gamma} \frac{(K_1 + h)^{1-\gamma} - K_1^{1-\gamma}}{h}, \\ \overline{\lim}_{K_2 \rightarrow \infty} J_{K_1}(K_1, K_2) &\leq \frac{\lambda_1}{1 - \gamma} \frac{K_1^{1-\gamma} - (K_1 - h)^{1-\gamma}}{h}. \end{aligned}$$

The fact that this holds regardless of h guarantees that, as K_2 approaches infinity, $J_{K_1}(\cdot, K_2)$ converges and the limit equals $\lambda_1 K_1^{-\gamma}$. ■

A brief discussion will clarify the intuition behind the main result. If parameters of the model satisfy (3.30), the drift of ω_t is strictly negative everywhere. By Bellman-Gronwall inequality (see Gihman and Skorohod (1972, §16, Lemma 4)), this implies that ω_t is bounded from above by the process $\omega^* + (\alpha + \delta - \lambda_1^{-1/\gamma} - \sigma_1^2/2 + (\sigma_{21}^2 + \sigma_2^2)/2)t + (\sigma_1 + \sigma_{21})W_{1t} + \sigma_2 W_{2t}$, reflected at ω^* , which is equivalent in law to a reflected Brownian motion with the drift coefficient $\alpha + \delta - \lambda_1^{-1/\gamma} - \sigma_1^2/2 + (\sigma_{21}^2 + \sigma_2^2)/2$, the diffusion coefficient $\left((\sigma_{21} - \sigma_1)^2 + \sigma_2^2\right)^{1/2}$ and the initial condition ω_0 . This process possesses the long-run stationary distribution, characterized by the exponential density (see Karlin and Taylor (1981)). Thus, ω_t has zero probability of reaching infinity. Similarly, ω_t is bounded from below by a reflected Brownian motion with the drift $\alpha + \delta - \tilde{c}_1^*(\omega^*) - \sigma_1^2/2 + (\sigma_{21}^2 + \sigma_2^2)/2$, therefore ω^* is not an absorbing boundary. Both the upper and the lower bounds on ω_t have long-run stationary distributions and so does ω_t .

If the inequality opposite to (3.30) holds, the drift of the process ω_t is uniformly positive for all values of ω exceeding a certain fixed value, therefore the ratio of capital stocks has a positive probability of reaching infinity (see Karlin and Taylor (1981)). In the case under consideration an even stronger result holds, as stated in the proposition.

To prove the first statement of the proposition, define a new process $\zeta = \omega - \omega^*$. This process is a diffusion on $[0, \infty)$ reflected at zero and is equivalent in law to the solution of

$$\begin{aligned} d\zeta_t &= \mu_\zeta dt + \sigma_\zeta dW_t, \quad \zeta_t \geq 0, \\ \mu_\zeta &= \left(-\alpha - \delta + \tilde{c}_{1t}^*(\omega_t) + \frac{\sigma_1^2}{2} - \frac{\sigma_{21}^2 + \sigma_2^2}{2} \right), \\ \sigma_\zeta &= \sqrt{(\sigma_{21} - \sigma_1)^2 + \sigma_2^2} dW_t. \end{aligned}$$

Following Gihman and Skorohod (1972, §23), I introduce a function $u(\cdot)$ defined on $[0, \infty)$, decreasing, possessing bounded continuous derivatives $u'(\cdot)$, $u''(\cdot)$ and satisfying

$$u(0) = 0, \quad \mu_\zeta u'(0_+) + \frac{1}{2} \sigma_\zeta^2 u''(0_+) = 0, \quad \lim_{x \rightarrow \infty} u'(x) = 1, \quad \lim_{x \rightarrow \infty} u''(x) = 0.$$

Such a function can be constructed by "pasting" together a linear and an exponential function and smoothing the resulting function around the pasting point. Then $\eta_t \equiv u(\zeta_t)$ is itself a diffusion process reflected at zero. η_t satisfies

$$d\eta_t = \mu_\eta(\eta_t) dt + \sigma_\eta(\eta_t) dW_t$$

on the interval (\mathfrak{D}, ∞) , where

$$\begin{aligned}\mu_\eta(\eta) &= \left(-\alpha - \delta + \tilde{c}_{1t}^*(\omega^* + u^{-1}(\eta)) + \frac{\sigma_1^2}{2} - \frac{\sigma_{21}^2 + \sigma_2^2}{2} \right) u'(u^{-1}(\eta)) + \\ &\quad \frac{1}{2} \sigma_\zeta^2 u''(u^{-1}(\eta)), \\ \sigma_\eta(\eta) &= \sigma_\zeta u'(u^{-1}(\eta)).\end{aligned}$$

Thus, $\mu_\eta(\cdot)$ is bounded and it approaches $-\alpha - \delta + \lambda_1^{-1/\gamma} + \sigma_1^2/2 - (\sigma_{21}^2 + \sigma_2^2)/2$ as η approaches infinity. $\sigma_\eta(\cdot)$ is also bounded and approaches σ_ζ as η approaches infinity. Next, extend the functions $\mu_\eta(\cdot)$ and $\sigma_\eta(\cdot)$ to the whole real line by means of

$$\begin{aligned}\mu_\eta^*(x) &= \mu_\eta(x), & x > 0, \\ \mu_\eta^*(x) &= -\mu_\eta(x), & x \leq 0, \\ \sigma_\eta^*(x) &= \sigma_\eta(-x), & x > 0, \\ \sigma_\eta^*(x) &= \sigma_\eta(x), & x \leq 0.\end{aligned}$$

Define η_t^* as a solution of

$$d\eta_t^* = \mu_\eta^*(\eta_t^*) dt + \sigma_\eta^*(\eta_t^*) dW_{1t}^*.$$

According to Gihman and Skorohod (1972, §23), the process $|\eta_t^*|$ is equivalent in law to η_t . Thus, given the one-to-one correspondence between η_t and ω_t , it suffices to establish the existence of the long-run stationary distribution of η_t^* . To do this, define a new function

$$\bar{\sigma}(x) \equiv \sigma_\eta^*(x) \exp \left(- \int_0^x \frac{2\mu_\eta^*(y)}{\sigma_\eta^{*2}(y)} dy \right).$$

According to Gihman and Skorohod (1972, §18, Th. 3), the process η_t^* possesses the long-run distribution if

$$\int_{-\infty}^{\infty} \frac{1}{\bar{\sigma}^2(y)} dy < \infty.$$

This last condition is clearly satisfied in view of the properties of the functions $\mu_\eta(\cdot)$ and $\sigma_\eta(\cdot)$ established above.

To prove the second statement, assume that $\text{Prob}\{\lim_{t \rightarrow \infty} \zeta_t = \infty\} < 1$. Then, there exists $M > 0$, such that for any $M_1 > M$,

$$\text{Prob} \left\{ \exists \{t_k, k = 1, 2, \dots\} : \lim_{k \rightarrow \infty} t_k = \infty, \zeta_{t_k} \leq M_1 \right\} > 0.$$

To uncover a contradiction, define $M_1 > M$, such that $\mu_\zeta > 0$ for $\zeta \geq M_1$. Fix $M_2 > M_1$. Since the process ζ_t is continuous and bounded below by a reflected Brownian motion,

$$\text{Prob} \{ \exists t \in (t_k, \infty) : \zeta_t = M_2 | \zeta_{t_k} \leq M_1 \} = 1.$$

According to a well-known property of the Brownian motion with drift (e.g., Harrison (1990, §8), Karatzas and Shreve (1991, pp.196–197)),

$$\text{Prob} \{ \exists s \in (t, \infty) : \zeta_s \leq M_1 | \zeta_t = M_2 \} = \varepsilon < 1,$$

where ε is a function of M_2 , M_1 and model parameters and does not depend on t . Thus, one concludes that

$$\text{Prob} \{ \exists \{(t, s)_l, l = 1, 2, \dots\} : t_l < s_l < t_{l+1}, \zeta_{t_l} = M_2, \zeta_{s_l} \leq M_1 \} = \lim_{l \rightarrow \infty} \varepsilon^l > 0,$$

which is a clear contradiction. ■

3.9.5 Proof of Proposition 3.6

One has to verify that the processes defined in Proposition 3.6 satisfy the Definition 3.5. I check each of the statements of the definition separately.

(i) Suppose that the agent can attain a higher level of expected utility with another feasible consumption plan $(\tilde{c}_{1t}, \tilde{c}_{2t})$:

$$E_0 \left[\int_0^{\infty} e^{-\rho t} (U(c_{1t}^*, c_{2t}^*) - U(\tilde{c}_{1t}, \tilde{c}_{2t})) dt \right] < 0.$$

By concavity of the utility function,

$$U(c_1^*, c_2^*) - U(\tilde{c}_1, \tilde{c}_2) \geq U_{c_1}(c_1^*, c_2^*) (c_1^* - \tilde{c}_1) + U_{c_2}(c_1^*, c_2^*) (c_2^* - \tilde{c}_2),$$

and therefore

$$E_0 \left[\int_0^{\infty} \eta_{0,t} ((c_{1t}^* - \tilde{c}_{1t}) + S_t (c_{2t}^* - \tilde{c}_{2t})) dt \right] < 0.$$

As I will show in (ii), $E_0 \left[\int_0^{\infty} \eta_{0,t} S_t c_{2t}^* dt - \int_0^{\infty} \eta_{0,t} dI_t^* \right] = P_0 K_{20}$. An entirely similar argument can be used to demonstrate that $E_0 \left[\int_0^{\infty} \eta_{0,t} c_{1t}^* dt + \int_0^{\infty} \eta_{0,t} dI_t^* \right] = K_{10}$. It follows that

$$E_0 \left[\int_0^{\infty} \eta_{0,t} (c_{1t}^* + S_t c_{2t}^*) dt \right] = P_0 K_{20} + K_{10} = V_0$$

and therefore

$$\mathbb{E}_0 \left[\int_0^\infty \eta_{0,t} (\bar{c}_{1t} + S_t \bar{c}_{2t}) dt \right] > V_0. \quad (3.100)$$

Given (v), the dynamic budget constraint implies that

$$\mathbb{E}_0 \left[\int_0^T \eta_{0,t} (c_{1t} + S_t c_{2t}) dt \right] \leq V_0, \quad \forall T > 0.$$

for every feasible consumption plan (c_{1t}, c_{2t}) (e.g., Karatzas and Shreve (1991, p. 374)).

Using the monotone convergence theorem,

$$\mathbb{E}_0 \left[\int_0^\infty \eta_{0,t} (c_{1t} + S_t c_{2t}) dt \right] \leq V_0,$$

which is a clear violation of (3.100). Thus, $(\bar{c}_{1t}, \bar{c}_{2t})$ cannot be budget-feasible, which proves optimality of (c_{1t}^*, c_{2t}^*) .

(ii) Without loss of generality, assume that there is only one firm in the industry, facing the problem

$$\max_{\{I_t\}} \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{U_{c_1}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{10}^*, c_{20}^*)} \frac{U_{c_2}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{1t}^*, c_{2t}^*)} K_{2t} dt - \int_0^\infty e^{-\rho t} \frac{U_{c_1}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{10}^*, c_{20}^*)} dI_t \right], \quad (3.101)$$

subject to (3.5) and $I_t \geq 0, dI_t \geq 0$. First, according to (3.91),

$$\mathbb{E}_s \left[\int_s^\infty e^{-\rho t} \frac{U_{c_2}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{1s}^*, c_{2s}^*)} \frac{R_{2t}}{R_{2s}} dt \right] = e^{-\rho s} \frac{J_{K_2}(K_{1s}^*, K_{2s}^*)}{J_{K_1}(K_{1s}^*, K_{2s}^*)}.$$

Second, the solution of (3.5) is given by

$$K_{2t} = K_{20}R_{2t} + \int_0^t \frac{R_{2t}}{R_{2s}} dI_s, \quad (3.102)$$

where R_{2t} is defined by (3.92). Finally,

$$\begin{aligned} & \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{U_{c_1}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{10}^*, c_{20}^*)} \frac{U_{c_2}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{1t}^*, c_{2t}^*)} K_{2t} dt \right] = \\ & \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{U_{c_2}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{10}^*, c_{20}^*)} \left(K_{20}R_{2t} + \int_0^t \frac{R_{2t}}{R_{2s}} dI_s \right) dt \right] = \\ & \frac{J_{K_2}(K_{10}, K_{20})}{J_{K_1}(K_{10}, K_{20})} K_{20} + \mathbb{E}_0 \left[\int_0^\infty \int_s^\infty e^{-\rho t} \frac{U_{c_2}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{10}^*, c_{20}^*)} \frac{R_{2t}}{R_{2s}} dt dI_s \right], \end{aligned}$$

where the first equality follows from (3.102) and the second is obtained using integration by parts and Lemma 3.8. According to the law of iterated expectations, the second term in the latter expression equals

$$\begin{aligned} & \mathbb{E}_0 \left[\int_0^\infty \frac{U_{c_1}(c_{1s}^*, c_{2s}^*)}{U_{c_1}(c_{10}^*, c_{20}^*)} \mathbb{E}_s \left[\int_s^\infty e^{-\rho t} \frac{U_{c_2}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{1s}^*, c_{2s}^*)} \frac{R_{2t}}{R_{2s}} dt \right] dI_s \right] = \\ & \mathbb{E}_0 \left[\int_0^\infty e^{-\rho s} \frac{U_{c_1}(c_{1s}^*, c_{2s}^*)}{U_{c_1}(c_{10}^*, c_{20}^*)} \frac{J_{K_2}(K_{1s}^*, K_{2s}^*)}{J_{K_1}(K_{1s}^*, K_{2s}^*)} dI_s \right], \end{aligned}$$

the equality being an application of (3.91). Thus, the firm's problem is equivalent to

$$\max_{\{I_t\}} \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{U_{c_1}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{10}^*, c_{20}^*)} \left(\frac{J_{K_2}(K_{1t}^*, K_{2t}^*)}{J_{K_1}(K_{1t}^*, K_{2t}^*)} - 1 \right) dI_t \right].$$

Note that, because of (3.11) and the constraint $dI_t \geq 0$, the last expression is nonpositive for all feasible investment policies. It equals zero if $1_{\{J_{K_2} \neq J_{K_1}\}} dI_t = 0$. Since the candidate solution I_t^* satisfies this condition, its optimality has been verified. Finally, the value of

the firm under the optimal investment policy equals

$$\frac{J_{K_2}(K_{10}, K_{20})}{J_{K_1}(K_{10}, K_{20})} K_{20}.$$

(iii, iv) These conditions are satisfied by construction.

(v) The short-rate process r_t is well defined by (3.45) (this can also be verified by direct calculation or indirectly, as discussed in footnote 12) and satisfies (3.39) by construction. As I have demonstrated in (ii), $P_t = E_t [\int_t^\infty \eta_{t,s} S_s K_{2s}^* ds - \int_t^\infty \eta_{t,s} dI_s^*]$, thus (3.40) is satisfied. Both (3.41) and (3.42) hold, because corresponding cumulative return processes satisfy the Consumption CAPM (see Duffie (1996, p. 229)):

$$\begin{pmatrix} \alpha \\ \mu_F \end{pmatrix} = r_t - \frac{U_{c_1 c_1}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{1t}^*, c_{2t}^*)} \begin{pmatrix} \sigma_1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} d \langle c_{1t}^*, W_{1t} \rangle / dt \\ d \langle c_{1t}^*, W_{2t} \rangle / dt \end{pmatrix}.$$

This proves that $\eta_{t,s}$ is a valid stochastic discount factor.

I will now show that $\eta_{t,s}$ is the unique stochastic discount factor consistent with (3.41) and (3.42). First, note that the instantaneous variance-covariance matrix of the cumulative return processes is constant and nonsingular. Second, according to Harrison and Kreps (1979), it is sufficient to demonstrate that the process for the market price of risk,

$$\theta_t \equiv - \frac{U_{c_1 c_1}(c_{1t}^*, c_{2t}^*)}{U_{c_1}(c_{1t}^*, c_{2t}^*)} \begin{pmatrix} d \langle c_{1t}^*, W_{1t} \rangle / dt \\ d \langle c_{1t}^*, W_{2t} \rangle / dt \end{pmatrix}, \quad (3.103)$$

is uniformly bounded. Using the relations $U_{c_1} = J_{K_1}$ and $c_1^* = (J_{K_1})^{-1/\gamma}$, rewrite (3.103)

as

$$\theta_t = \begin{pmatrix} \sigma_1 \frac{K_1 J_{K_1 K_1}}{J_{K_1}} + \sigma_{21} \frac{K_2 J_{K_1 K_2}}{J_{K_1}} \\ \sigma_{22} \frac{K_2 J_{K_1 K_2}}{J_{K_1}} \end{pmatrix}.$$

Since $J(K_1, K_2) = \frac{1}{1-\gamma} K_1^{1-\gamma} f(\omega)$, both $K_2 J_{K_1 K_2} / J_{K_1}$ and $K_1 J_{K_1 K_1} / J_{K_1}$ are functions of ω only:

$$\begin{aligned} \frac{K_2 J_{K_1 K_2}}{J_{K_1}} &= \frac{f' - \frac{1}{1-\gamma} f''}{f - \frac{1}{1-\gamma} f'}, \\ \frac{K_1 J_{K_1 K_1}}{J_{K_1}} &= -\gamma - \frac{f' - \frac{1}{1-\gamma} f''}{f - \frac{1}{1-\gamma} f'}. \end{aligned}$$

Thus, it suffices to establish that the first ratio is finite. The denominator equals $\tilde{c}_1^*(\omega)^{-\gamma}$, which is a bounded function (see Lemmas 3.9, 3.9.4). The following lemma implies that the numerator is also bounded. ■

Lemma 3.11 *As the ratio of capital stocks approaches infinity ($\omega \rightarrow \infty$), $f'(\omega) - \frac{1}{1-\gamma} f''(\omega)$ approaches zero.*

Proof. Define $g(\omega)$ using (3.84). Then $f' - \frac{1}{1-\gamma} f''$ equals $b^{-1/\gamma} \left(g' - \frac{1}{1-\gamma} g'' \right)$. According to Lemma 3.9.4, $\lim_{\omega \rightarrow \infty} \left(f(\omega) - \frac{1}{1-\gamma} f'(\omega) \right) = \lambda_1$. This implies that

$$\lim_{\omega \rightarrow \infty} \left(g(\omega) - \frac{1}{1-\gamma} g'(\omega) \right) = 0.$$

Next, (3.99) implies that $\lim_{\omega \rightarrow \infty} g(\omega) = 0$. Thus,

$$\lim_{\omega \rightarrow \infty} g(\omega) = \lim_{\omega \rightarrow \infty} g'(\omega) = 0. \quad (3.104)$$

Next, using (3.85),

$$g' - \frac{1}{1-\gamma}g'' = g' + \frac{p_1g' + p_0g + \gamma b^{-1/\gamma} \left(\lambda_1 + b^{1/\gamma} \left(g - \frac{1}{1-\gamma}g' \right) \right)^{1-1/\gamma} + \lambda_1 p_0 b^{-1/\gamma}}{(1-\gamma)p_2},$$

which converges to zero as ω approaches infinity (I used the fact that $p_0 = -\gamma\lambda_1^{-1/\gamma}$). ■

3.9.6 Perturbation Analysis, Section 3.4.4

The case $\gamma \neq 1$. As I have stated in Section 3.4.4, the first step in the analysis is to rescale the independent variable, introducing $\xi = \omega - \ln(b)/\gamma$. As a function of the new independent variable, $f(\xi)$ satisfies

$$p_2f'' + p_1f' + p_0f + \gamma \left(f - \frac{1}{1-\gamma}f' \right)^{1-1/\gamma} = -b^{1/\gamma}e^{(1-\gamma)\xi}, \quad (3.105)$$

subject to the boundary conditions (3.17, 3.18, 3.24, 3.25). I look for $f(\xi)$ and the optimal investment threshold Ξ^* in the form

$$f(\xi) = \lambda_1 + b^{1/\gamma} \left(\lambda_2 e^{(1-\gamma)\xi} + f_{(0)}(\xi) + \dots + b^{n/\gamma} f_{(n)}(\xi) + \dots \right), \quad (3.106)$$

$$\Xi^* = \Xi_{(0)} + b^{1/\gamma} \Xi_{(1)} + \dots + b^{n/\gamma} \Xi_{(n)} + \dots \quad (3.107)$$

The next step in perturbation analysis is to substitute the expansion (3.106) into (3.105) and to collect terms of the same order in $b^{1/\gamma}$. This way one obtains a sequence of equations

on $f_{(0)}(\xi)$, $f_{(1)}(\xi)$, etc.:

$$q_2 f''_{(0)} + q_1 f'_{(0)} + q_0 f_{(0)} = 0, \quad (3.108)$$

$$q_2 f''_{(1)} + q_1 f'_{(1)} + q_0 f_{(1)} = \frac{\gamma - 1}{2\gamma} \lambda_1^{-1-1/\gamma} \left(f_{(0)} - \frac{1}{1-\gamma} f'_{(0)} \right)^2, \quad (3.109)$$

⋮

where

$$q_2 = p_2 > 0,$$

$$q_1 = p_1 + \lambda_1^{-1/\gamma},$$

$$q_0 = -\lambda_1^{-1/\gamma} < 0.$$

Each of these equations is a linear differential equation with a known general solution. The inhomogeneous term in equation on $f_{(k+1)}$ depends only on $f'_{(k)}$, therefore one can solve these equations sequentially. To specify the solution completely, one has to impose boundary conditions, obtained by substituting (3.106) and (3.107) into (3.17, 3.18) and matching the terms according to their order in b . I solve for the first two terms in the asymptotic expansion (3.106).

The general solution of (3.108) is given by

$$f_{(0)}(\xi) = A_{(0)} \exp(\kappa_- (\xi - \xi^*)) + B_{(0)} \exp(\kappa_+ (\xi - \xi^*)), \quad (3.110)$$

where

$$\kappa_{\pm} = \frac{-q_1 \pm \sqrt{q_1^2 - 4q_2 q_0}}{2q_2}, \quad \kappa_+ > 0, \quad \kappa_- < 0.$$

Since $f_{(0)}(\xi)$ inherits the condition $\lim_{\xi \rightarrow \infty} f_{(0)}(\xi) = \lambda_1$ from (3.24, 3.25), one immediately concludes that $B_{(1)} = 0$. Next, substitute (3.110) and (3.107) into (3.17, 3.18). This results in a system of two equations on $A_{(0)}$ and $\Xi_{(0)}$:

$$\begin{aligned}\lambda_2(1-\gamma)\Xi_{(0)}^{1-\gamma} - \lambda_1(1-\gamma)\Xi_{(0)} + \kappa A_{(0)} &= 0, \\ -\lambda_2\gamma(1-\gamma)\Xi_{(0)}^{1-\gamma} + (\kappa^2 - \kappa)A_{(0)} &= 0.\end{aligned}$$

where $\kappa = \kappa_-$. These equations can be solved explicitly:

$$A_{(0)} = \frac{\lambda_2\gamma(1-\gamma)}{\kappa(\kappa-1)} \left(\frac{\lambda_1}{\lambda_2} \frac{\kappa-1}{\kappa-(1-\gamma)} \right)^{1-1/\gamma}, \quad (3.111)$$

$$\Xi_{(0)} = \left(\frac{\lambda_1}{\lambda_2} \frac{\kappa-1}{\kappa-(1-\gamma)} \right)^{-1/\gamma}. \quad (3.112)$$

$\Xi_{(0)}$ is well defined, since²⁴

$$\kappa - (1-\gamma) < 0. \quad (3.113)$$

Note that $\text{sign}(A_{(0)}) = \text{sign}(1-\gamma)$, which one would expect to hold, because

$$\left(b^{1/\gamma} A_{(0)} / (1-\gamma) \right) \exp(\kappa(\xi - \xi^*))$$

²⁴Since

$$\kappa - (1-\gamma) = \frac{-q_1 - 2(1-\gamma)q_2 - \sqrt{q_1^2 - 4q_2q_0}}{2q_2}$$

and $q_2 > 0$, it suffices to check the sign of

$$(q_1^2 - 4q_2q_0) - (q_1 + 2(1-\gamma)q_2)^2.$$

The last expression equals

$$-4q_2(q_0 + (1-\gamma)q_1 + (1-\gamma)^2q_2) = \frac{4q_2}{\lambda_2} > 0.$$

This implies that (3.113) holds.

approximates the difference between the optimal and a feasible solution of the maximization problem. Also, note that (3.113) implies that the first-order approximation to the value function is of the same sign as $1 - \gamma$, which must be the case for the optimal solution. It is straightforward to verify that in the limit of γ approaching 1, (3.111, 3.112) converge to their counterparts for the log-utility case (3.119, 3.120), and thus the first-order terms in (3.106, 3.107) converge to the first-order terms in (3.114, 3.115).

Next, I determine $f_{(1)}(\xi)$ and $\Xi_{(1)}$. $f_{(1)}(\xi)$ equals

$$A_{(1)} \exp(\kappa(\xi - \xi^*)) + C_{(1)} \exp(2\kappa(\xi - \xi^*)),$$

where $C_{(1)}$ is found from (3.109):

$$C_{(1)} = \frac{1}{2\gamma(\gamma - 1)} \lambda_1^{-1-1/\gamma} \frac{A_{(0)}^2 (1 - \gamma - \kappa)^2}{4\kappa^2 q_2 + 2\kappa q_1 + q_0}.$$

$A_{(1)}$ and $\Xi_{(1)}$ satisfy a system of two linear equations, resulting from (3.17, 3.18):

$$\begin{aligned} \kappa A_{(1)} + \left(\lambda_2 (1 - \gamma)^2 \Xi_{(0)}^{-\gamma} - \lambda_1 (1 - \gamma) \right) \Xi_{(1)} &= - \left((\kappa + \gamma - 1) A_{(0)} \Xi_{(0)} + 2\kappa C_{(1)} \right), \\ (\kappa^2 - \kappa) A_{(1)} - \lambda_2 \gamma (1 - \gamma)^2 \Xi_{(0)}^{-\gamma} \Xi_{(1)} &= - \left(\kappa (\kappa + \gamma - 1) A_{(0)} \Xi_{(0)} + (4\kappa^2 - 2\kappa) C_{(1)} \right). \end{aligned}$$

While it is clearly possible, I will not solve for $A_{(1)}$ and $\Xi_{(1)}$ explicitly here.

The case $\gamma = 1$. The new unknown function $g(\omega)$ satisfies the differential equation (3.20), subject to the boundary conditions (3.17, 3.18, 3.26). I expand $g(\omega)$ in powers of b as

$$g(\omega) = bg_{(1)}(\omega) + b^2 g_{(2)}(\omega) + \cdots b^n g_{(n)}(\omega) + \cdots \quad (3.114)$$

Also, I expand Ω^* as

$$\Omega^* = b\Omega_{(1)} + b^2\Omega_{(2)} + \cdots b^n\Omega_{(n)} + \cdots . \quad (3.115)$$

After substituting (3.114) into (3.20) and collecting terms of the same order in b , one obtains a sequence of equations on $g_{(1)}(\omega)$, $g_{(2)}(\omega)$, etc.:

$$\begin{aligned} p_2 g_{(1)}'' + (p_1 + \rho) g_{(1)}' - \rho g_{(1)} &= 0, \\ p_2 g_{(2)}'' + (p_1 + \rho) g_{(2)}' - \rho g_{(2)} &= -\rho^2 \left(g_{(1)}' \right)^2, \\ &\vdots \end{aligned}$$

The first equation has a general solution

$$g_{(1)}(\omega) = A_{(1)} \exp(\kappa_-(\omega - \omega^*)) + B_{(1)} \exp(\kappa_+(\omega - \omega^*)),$$

where

$$\kappa_{\pm} = \frac{-(p_1 + \rho) \pm \sqrt{(p_1 + \rho)^2 + 2\rho p_2}}{2p_2}.$$

Observe that $\kappa_+ > 0$ and $\kappa_- < 0$. According to (3.17), $\lim_{\omega \rightarrow \infty} g_{(1)}(\omega) = 0$ therefore

$B_{(1)} = 0$. Let $\kappa \equiv \kappa_-$. Then

$$g_{(1)}(\omega) = A_{(1)} \exp(\kappa(\omega - \omega^*)). \quad (3.116)$$

Similarly, the second equations yields

$$g_{(2)}(\omega) = A_{(2)} \exp(\kappa(\omega - \omega^*)) + C_{(2)} \exp(2\kappa(\omega - \omega^*)), \quad (3.117)$$

$$C_{(2)} = \frac{-\rho^2 \kappa^2 A_{(1)}^2}{4p_2 \kappa^2 - 2\kappa(p_1 + \rho) - \rho}. \quad (3.118)$$

It is clear that all other equations can be solved in closed form as well.

To find coefficients $A_{(1)}$, $\Omega_{(1)}$, $A_{(2)}$, $\Omega_{(2)}$, etc., that satisfy both of the boundary conditions on $g(\omega)$, substitute (3.116-3.118) into the boundary conditions (3.17, 3.18) and collect terms of the same order in b . After simple algebraic manipulations, the unknown coefficients are shown to satisfy the following system of equations:

$$\begin{aligned} \rho \kappa A_{(1)} - \Omega_{(1)} &= -1, \\ (\kappa^2 - \kappa) A_{(1)} &= \frac{1}{\rho}, \\ \rho \kappa A_{(2)} - \Omega_{(2)} &= -2\rho \kappa C_{(2)} - \rho \kappa \Omega_{(1)} A_{(1)}, \\ (\kappa^2 - \kappa) A_{(2)} &= -\kappa^2 \Omega_{(1)} A_{(1)} - (4\kappa^2 + 2\kappa) C_{(2)}, \\ &\vdots \end{aligned}$$

This infinite system of equations has a block-diagonal form. The first pair of equations leads to

$$\Omega_{(1)} = \frac{\kappa}{\kappa - 1} > 0, \quad (3.119)$$

$$A_{(1)} = \frac{1}{\rho \kappa (\kappa - 1)} > 0. \quad (3.120)$$

The second pair yields

$$A_{(2)} = \frac{-\kappa\Omega_{(1)}A_{(1)} - (4\kappa + 2)C_{(2)}}{\kappa - 1},$$

$$\Omega_{(2)} = \rho\kappa \frac{-\Omega_{(1)}A_{(1)} - (4 + 2\kappa)C_{(2)}}{\kappa - 1}.$$

In general, coefficients $A_{(k)+}$, $\Omega_{(k-2)}$ can be computed sequentially by solving systems of two linear equations.

To obtain an approximation to the optimal consumption policy, expand it as

$$\tilde{c}_1^*(\omega) = \tilde{c}_{1(0)}^*(\omega) + b\tilde{c}_{1(1)}^*(\omega) + b^2\tilde{c}_{1(2)}^*(\omega) + \dots + b^n\tilde{c}_{1(2)}^*(\omega) + \dots.$$

Using (3.114), one finds that

$$\begin{aligned}\tilde{c}_{1(0)}^*(\omega) &= \rho, \\ \tilde{c}_{1(1)}^*(\omega) &= \rho^2 g'_{(1)}(\omega), \\ \tilde{c}_{1(2)}^*(\omega) &= \rho^2 g'_{(2)}(\omega) + \rho^3 \left(g'_{(1)}(\omega)\right)^2, \\ &\vdots\end{aligned}$$

3.9.7 Proof of Proposition 3.5.2

First, I prove the sufficient condition for the value function to be finite. Restrictions on the model parameters (3.9, 3.10) are sufficient when $\gamma < 1$. To obtain a sufficient condition for the case $\gamma > 1$, consider the following two auxiliary optimization problems: (3.81) and (3.82), both subject to (3.48, 3.49, 3.50, 3.51, 3.6). An upper bound on the original objective function results from dropping the irreversibility constraint and maximizing each

of its components separately.

Consider the problem

$$\max_{\{c_{1t}, \theta_t\}} \mathbb{E}_0 \left[\int_0^{\infty} e^{-\rho t} \frac{1}{1-\gamma} c_{1t}^{1-\gamma} dt \right],$$

subject to

$$dK_t = ((\alpha\theta_t - \delta(1 - \theta_t))K_t - c_{1t})dt + (\sigma_1\theta_t + \sigma_{21}(1 - \theta_t)) K_t dW_{1t} +$$

$$\sigma_2(1 - \theta_t)K_t dW_{2t},$$

$$\theta_t \geq 0, \quad \theta_t \leq 1.$$

The value function, $J(K)$, satisfies

$$\rho J = \max_{c \geq 0, \theta \in [0,1]} \left\{ \begin{array}{l} \frac{1}{1-\gamma} c^{1-\gamma} + J_K((\alpha\theta - \delta(1 - \theta))K - c) + \\ \frac{1}{2} J_{KK}^2 \left((\sigma_1\theta + \sigma_{21}(1 - \theta))^2 + \sigma_2^2(1 - \theta)^2 \right) K^2 \end{array} \right\}.$$

The solution has the functional form $J(K) = \lambda(1 - \gamma)^{-1} K^{1-\gamma}$. The maximization over θ yields $\theta = \bar{\theta}$,

$$\bar{\theta} = \max \left(0, \min \left(1, \frac{\alpha + \delta - \gamma \left(\sigma_{21} \frac{\sigma_1 - \sigma_{21}}{2} - \sigma_2^2 \right)}{\gamma \left(\frac{(\sigma_1 - \sigma_{21})^2}{2} + 2\sigma_2^2 \right)} \right) \right). \quad (3.121)$$

After that the problem is reduce to (3.75) and I find that $\lambda = \lambda_1(\bar{\alpha}, \bar{\sigma}, \rho)$ (see 3.76), where

$$\bar{\alpha} = \alpha\bar{\theta} - \delta(1 - \bar{\theta}),$$

$$\bar{\sigma} = \sqrt{(\sigma_1\bar{\theta} + \sigma_{21}(1 - \bar{\theta}))^2 + \sigma_2(1 - \bar{\theta})^2}.$$

To verify that the resulting solution is indeed the value function of the original optimization problem, one can use a standard verification theorem, (see, for example, Fleming and Soner (1992, Th. 9.1)). Similarly, the second optimization problem is reduced to computing (3.77) with properly defined values of parameters. The resulting value function is finite if and only if $\lambda_2(\bar{\alpha}, \bar{\sigma}, \rho)$ is (see 3.78).

The value function of (3.47–3.51) satisfies

$$\rho J = \max_{c \geq 0, \theta \in [0,1]} \left\{ \begin{array}{l} \frac{1}{1-\gamma} c^{1-\gamma} + \frac{b}{1-\gamma} (\theta K)^{1-\gamma} + J_K((\alpha\theta - \delta(1-\theta))K - c) + \\ \frac{1}{2} J_{KK}^2 \left((\sigma_1\theta + \sigma_{21}(1-\theta))^2 + \sigma_2^2(1-\theta)^2 \right) K^2 \end{array} \right\}.$$

Due to the homogeneity of the problem, the solution has the functional form $J(K) = \lambda(1-\gamma)^{-1} K^{1-\gamma}$. The optimal choice of θ must satisfy (3.55, 3.56), while λ must satisfy (3.57). The optimal choice of c is given by (3.54). ■

References

- Abel, A., 1979. *Investment and the Value of Capital*. New York: Garland.
- Abel, A., 1983, "Optimal Investment under Uncertainty," *American Economic Review* 73, 228–233.
- Abel, A., 1988, "Stock Prices under Time-Varying Dividend Risk: An Exact Solution in an Infinite Horizon Equilibrium Model," *Journal of Monetary Economics* 22, 375–393.
- Abel, A., A. Dixit, J. Eberly and R. Pindyck, 1996, "Options, The Value of Capital, and Investment," *Quarterly Journal of Economics* 111, 753–777.
- Abel, A., and J. Eberly, 1994, "A Unified Model of Investment under Uncertainty," *American Economic Review* 84, 1369–1384.
- Abel, A., and J. Eberly, 1995, "The Effects of Irreversibility and Uncertainty on Capital Accumulation," *NBER Working Paper* 5363, 1995.
- Abel, A., and J. Eberly, 1996, "Optimal Investment with Costly Reversibility," *Review of Economic Studies* 63, 581–593.
- Abel, A., and J. Eberly, 1997a, "An exact Solution for the Investment and Value of a Firm Facing Uncertainty, Adjustment Costs, and Irreversibility," *Journal of Economic Dynamics and Control* 21, 831–852.
- Abel, A., and J. Eberly, 1997b, "The Mix and the Scale of Factors with Irreversible and Fixed Costs of Investment," *NBER Working Paper* 6148, 1997.
- Altug, S., 1984, "Time-to-Build, Delivery Lags, and the Equilibrium Pricing of Capital Goods," *Journal of Money, Credit, and Banking* 25, 301–319.
- Arrow, K., 1968, "Optimal Capital Policy and Irreversible Investment," in J.N. Wolfe, ed., *Value, Capital and Growth* (Chicago:Aldine, 1968), 1–20.
- Atkinson, C. and P. Wilmott, 1995, "Portfolio Management with Transaction Costs: An Asymptotic Analysis," *Mathematical Finance* 5, 357–367.
- Baldwin, C., 1982, "Optimal Sequential Investment When Capital is not Readily Reversible," *Journal of Finance* 37, 763–782.
- Baldwin, C. and R. Meyer, 1979, "Liquidity Preference under Uncertainty: A Model of Dynamic Investment in Illiquid Opportunities," *Journal of Financial Economics* 7, 347–374.
- Balvers, R., T. Cosimano, B. McDonald, 1990, "Predicting Stock Returns in an Efficient Market," *Journal of Finance* 45, 1109–1128.
- Barsky, R., 1989, "Why Don't the Prices of Stocks and Bonds Move Together?," *American Economic Review* 79, 1132–1145.

- Basu, P., 1987, "An Adjustment Cost Model of Asset Pricing," *International Economic Review* 28, 609–621.
- Basu, P. and S. Chib, 1985, "Equity Premium in a Production Economy," *Economics Letters* 18, 53–58.
- Basu, P. and H. Vinod, 1994, "Mean Reversion in Stock Prices: Implications from a Production Based Asset Pricing Model," *Scandinavian Journal of Economics* 96, 51–65.
- Benavie, A., E. Grinols and S. Turnovsky, 1996, "Adjustment Costs and Investment in a Stochastic Endogenous Growth Model," *Journal of Monetary Economics* 38, 77–100.
- Bernanke, B., 1983, "Irreversibility, Uncertainty and Cyclical Investment," *Quarterly Journal of Economics* 98, 85–106.
- Bentonila, S. and G. Bertola, 1990, "Firing Costs and Labor Demand: How Bad is Euroclerosis?," *Review of Economic Studies* 57, 381–402.
- Berk, J., R. Green and V. Naik, 1998, "Optimal Investment, Growth Options and Security Returns", *Working Paper*.
- Berk, J., 1998, "A Simple Approach for Deciding when to Invest," *American Economic Review*, forthcoming.
- Bertola, G., 1989, "Irreversible Investment". *Unpublished paper, Princeton University*, 1989.
- Bertola, G. and R. Caballero, 1990, "Kinked Adjustment Costs and Aggregate Dynamics," in *NBER Macroeconomics Annual 1990*, Cambridge: The MIT Press.
- Bertola, G. and R. Caballero, 1994, "Irreversibility and Aggregate Investment," *Review of Economic Studies* 61, 223–246.
- Bertsimas, D., L. Kogan, A. Lo, 1997, "Pricing and Hedging Derivative Securities in Incomplete Markets: An ε -Arbitrage Approach," *NBER Working Paper 6250, 1997*.
- Bose, S., 1970, "Optimal Growth in a Nonshiftable Capital Model," *Econometrica* 38, 128–152.
- Brennan, M., 1991, "The Price of Convenience and the Valuation of Commodity Contingent Claims" in D. Lund and B. Oksendal (eds.), *Stochastic Models and Option Models*, Elsevier Science Publishers, 1991.
- Brennan, M. and E. Schwartz, 1985, "Evaluating Natural Resource Investments," *Journal of Business* 58, 135–157.
- Brock, W., 1982, "Asset Prices in a Production Economy," in J.J. McCall (ed.), *The Economics of Information and Uncertainty*. Chicago: University of Chicago Press 1982.
- Caballero, R., 1991, "On the sign of the Investment-Uncertainty Relationship," *The American Economic Review* 81, 279–288.

- Caballero, R., 1997, "Aggregate Investment," *MIT Department of Economics Working Paper 97-20*, 1997.
- Caballero, R., E. Engel and J. Haltiwanger, 1995, "Plant-Level Adjustment and Aggregate Investment Dynamics," *Brookings Papers on Economic Activity*, 1995:2, 1-54.
- Caballero, R. and J. Leahy, 1996, "Fixed Costs: The Demise of Marginal q ," *NBER Working Paper 5508*, 1996.
- Caballero, R. and R. Pindyck, 1996, "Uncertainty, Investment and Industry Evolution," *International Economic Review* 37, 641-662.
- Chung, K. and R. Williams, 1990, *Introduction to Stochastic Integration*. Birkhäuser Boston.
- Cochrane, J., 1991, "Production-Based Asset Pricing and the Link between Stock Returns and Economic Fluctuations," *Journal of Finance* 46, 209-237.
- Cochrane, J., 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy* 104, 572-621.
- Coleman II, W., 1997, "Behavior of Interest Rates in a General Equilibrium Multisector Model with Irreversible Investment," *Macroeconomic Dynamics* 1, 206-227.
- Conrad, J. and G. Kaul, 1988, "Time-Variation in Expected Returns," *Journal of Business* 61, 409-425.
- Cox, C., J. Ingersoll and S. Ross, 1981, "The Relation Between Forward Prices and Futures Prices," *Journal of Financial Economics* 9, 321-346.
- Cukierman, A., 1980, "The Effects of Uncertainty of Investment under Risk Neutrality with Endogenous Information," *Journal of Political Economy* 88, 482-475.
- Dasgupta, P., 1969, "Optimum Growth when Capital is Non-transferable," *Review of Economic Studies* 36, 77-88.
- Davis, M. and A. Norman, 1990, "Portfolio Selection with Transactions Costs," *Mathematics of Operations Research* 15, 676-713.
- Deaton, A. and G. Laroque, 1996, "Competitive Storage and Commodity Price Dynamics," *Journal of Political Economy*, 896-923.
- Dixit, A., 1989a, "Intersectoral Capital Reallocation under Price Uncertainty," *Journal of International Economics* 26, 309-325.
- Dixit, A., 1989b, "Entry and Exit Decisions under Uncertainty," *Journal of Political Economy* 97, 620-638.
- Dixit, A., 1989c, "Hysteresis, Import Penetration, and Exchange Rate Pass-Through," *Quarterly Journal of Economics* 104, 205-228.
- Dixit, A., 1991, "Irreversible Investment with Price Ceilings," *Journal of Political Economy* 99, 541-557.

- Dixit, A., 1992, "Investment and Hysteresis", *Journal of Economic Perspectives* 1, 107–132.
- Donaldson, J., 1984, "Comparative Dynamics of an Equilibrium Intertemporal Asset Pricing Model," *Review of Economic Studies* 51, 491–508.
- Donaldson, J. and R. Mehra, 1983, "Stochastic Growth with Correlated Production Shocks," *Journal of Economic Theory* 29, 282–312.
- Dow, J. and L. Olson, 1992, "Irreversibility and the Behavior of Aggregate Stochastic Growth Models", *Journal of Economic Dynamics and Control* 16, 207–224.
- Duffie, D., 1996, *Dynamic Asset Pricing Theory*. Princeton, New Jersey: Princeton University Press.
- Duffie, D. and R. Stanton, 1988, "Pricing Continuously Resettled Contingent Claims," *Journal of Economic Dynamics and Control* 16, 561–574.
- Duffie, D. and C. Huang, 1985, "Implementing Arrow-Debreu Equilibria by Continuous Trading of Few Long-Lived Securities," *Econometrica* 53, 1337–1356.
- Dumas, B., 1991, "Super Contact and Related Optimality Conditions," *Journal of Economic Dynamics and Control* 15, 675–685.
- Dumas, B., 1992, "Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World," *Review of Financial Studies* 5, 153–180.
- Dybvig, P. and C. Huang, 1989, "Nonnegative Wealth, Absence of Arbitrage, and Feasible Consumption Plans," *Review of Financial Studies* 4, 377–401.
- Fama, E. and K. French, 1988, "Business Cycles and the Behavior of Metals Prices," *Journal of Finance* 43, 1075–1093.
- Fama, E. and K. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics* 22, 3–25.
- Fama, E. and K. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance* 47, 427–465.
- Floyd, J. and J. Hynes, 1979, "Capital immobility, adjustment costs, and the theoretical foundations of income-expenditure models," *Journal of Political Economy* 87, 267–291.
- Gibson, R. and E. Schwartz, 1990, "Stochastic Convenience Yield and the Pricing of Oil Contingent Claims," *Journal of Finance*, 959–976.
- Gihman, I. and A. Skorohod, 1972, *Stochastic Differential Equations*. Berlin: Springer-Verlag.
- Harrison, M., 1990, *Brownian Motion and Stochastic Flow Systems*. Malabar, Florida: Robert E. Krieger Publishing Company.
- Harrison, M. and D. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Security Markets," *Journal of Economic Theory* 20, 381–408.

- Harrison, M., S. Pliska, 1983, "A Stochastic Calculus Model of Continuous Trading: Complete Markets," *Stochastic Processes and Applications* 11, 215–260.
- Hayashi, F., 1982, Tobin's Marginal q and average q : a Neoclassical Interpretation," *Econometrica* 50, 213–224.
- He, H. and R. Pindyck, 1992, "Investments in Flexible Production Capacity," *Journal of Economic Dynamics and Control* 16, 575–599.
- Henry, C., 1974, "Investment Decisions under Uncertainty: The 'Irreversibility Effect'," *American Economic Review* 64, 1006–1012.
- Huang, C., 1987, "An Intertemporal General Equilibrium Asset Pricing Model: The Case of Diffusion Information," *Econometrica* 55, 117–142.
- Huffman, G., 1985, "Adjustment Costs and Capital Asset Pricing," *Journal of Finance* 40, 691–709.
- Hull, J. and A. White, 1987, "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance* 42, 281–300.
- Ingersoll, J. and S. Ross, 1987, "Waiting to Invest: Investment and Uncertainty," *Journal of Business* 65, 1–29.
- Johansen, L., 1967, "Some Theoretical Properties of a Two-Sector Model of Optimal Growth," *Review of Economic Studies* 34, 125–141.
- Karatzas, I., J. Lehoczky and S. Shreve, 1991, "Equilibrium Models with Singular Asset Prices," *Mathematical Finance* 1, 11–29.
- Karlin, S. and H. Taylor, 1981, *A Second Course in Stochastic Processes*. Academic Press, New York.
- Kothari, S. and J. Shanken, 1997, "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis," *Journal of Financial Economics* 44, 169–203.
- Kydland, F. and E. Prescott, 1982, "Time to Build and Aggregate Fluctuations," *Econometrica* 50, 1345–1370.
- Leahy, J., 1993, "Investment in Competitive Equilibrium: The Optimality of Myopic Behavior," *Quarterly Journal of Economics* 108, 1105–1133.
- Leahy, J., T. Whited, 1996, "The Effect of Uncertainty on Investment: Some Stylized Facts," *Journal of Money, Credit and Banking* 28, 64–83.
- LeRoy, S., 1983, "Keynes's Theory of Investment," *History of Political Economy* 15, 397–421.
- Lippman, S. and R. Rumelt, 1985, "Demand Uncertainty and Investment in Industry-Specific Capital," working paper, UCLA Graduate School of Management.
- Litzenberger, R. and N. Rabinowitz, 1995, "Backwardation in Oil Futures Markets: Theory and Empirical Evidence," *Journal of Finance* 50, 1517–1545.

- Long, J. and C. Plosser, 1983, "Real Business Cycles," *Journal of Political Economy* 91, 39–69.
- Lucas, R., 1978, "Asset Prices in an Exchange Economy", *Econometrica* 46, 1429–1445.
- Lucas, R., and E. Prescott, 1971, "Investment under Uncertainty," *Econometrica* 39, 659–681.
- Majd, S. and R. Pindyck, 1987, "Time to Build, Option Value, and Investment Decisions," *Journal of Financial Economics* 18, 7–27.
- McDonald, R. and D. Siegel, 1986, "The Value of Waiting to Invest," *Quarterly Journal of Economics* 101, 707–728.
- Milonas, N., 1986, "Price Variability and the Maturity Effect in Futures Markets," *Journal of Futures Markets* 6, 443–460.
- Naik, V., 1994, "Asset Prices in Dynamic Production Economies with Time-Varying Risk," *Review of Financial Studies* 7, 781–801.
- Obstfeld, M., 1994, "Risk, Diversification and Growth", *American Economic Review* 84, 1310–1329.
- Olson, L., 1989, "Stochastic Growth with Irreversible Investment," *Journal of Economic Theory* 47, 101–129.
- Pindyck, R., 1988, "Irreversible Investment, Capacity Choice, and the Value of the Firm," *American Economic Review* 83, 273–277.
- Pindyck, R., 1993, "A Note on Competitive Investment under Uncertainty," *American Economic Review* 78, 969–985.
- Prescott, E. and R. Mehra, 1980, "Recursive Competitive Equilibrium: The Case of Homogeneous Households," *Econometrica* 48, 1365–1379.
- Press, W., B. Flannery, S. Teukolsky and W. Vetterling, 1992, *Numerical Recipes in C*. Cambridge University Press.
- Richard, S. and M. Sundaresan, 1981, "A Continuous Time Equilibrium Model of Forward Prices and Futures Prices in a Multigood Economy," *Journal of Financial Economics* 9, 347–371.
- Rogers, L.C.G. and D. Williams, 1987, *Diffusions, Markov Processes, and Martingales*. John Wiley & Sons Ltd.
- Routledge, B., D. Seppi and C. Spatt, 1998, "Equilibrium Forward Curves for Commodities," Working paper, GSIA, Carnegie Mellon University.
- Ryder, H., 1969, "Optimal Accumulation in a Two-Sector Neoclassical Economy with Nonshiftable Capital," *Journal of Political Economy* 77, 665–683.
- Samuelson, P., 1965, "Proof that Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review* 6, 41–49.

- Sargent, T., 1979, "Tobin's Q and the Rate of Investment in General Equilibrium," in *On the State of Macroeconomics*, edited by K.Brunner and A.H.Metzler, North Holland 1979.
- Schwartz, E., 1997, "The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging," *Journal of Finance* 52, 923–973.
- Shreve, S. and H. Soner, 1994, "Optimal Investment and Consumption with Transaction Costs," *Annals of Applied Probability* 4, 609–692.
- Smith, G. and W. Starnes, 1979, "A Short-Run Two-Sector Model with Immobile Capital," *Journal of Money, Credit and Banking* 11, 44–67.
- Whalley, A. and P. Wilmott, 1997, "An Asymptotic Analysis of an Optimal Hedging Model for Option Pricing," *Mathematical Finance* 7, 307–324.
- Williams, J. and B. Wright, 1991, *Storage and Commodity Markets*. Cambridge: Cambridge University Press.
- Zhu, H., 1991, *Variational Inequalities and Dynamic Programming for Singular Stochastic Control*. Doctoral Dissertation, Division of Applied Mathematics, Brown University, Providence, RI.

THESIS PROCESSING SLIP

FIXED FIELD: ill. _____ name _____

index _____ biblio _____

COPIES: Archives Aero Dewey ~~1111~~ Hum
Lindgren Music Rotch Science

TITLE VARIES: _____

NAME VARIES: _____

IMPRINT: (COPYRIGHT) _____

COLLATION: 237 p

ADD: DEGREE: _____ ▶ DEPT.: _____

SUPERVISORS: _____

NOTES:

cat'r:	date:
DEPT: <u>Mgt</u>	page: <u>J215</u>
YEAR: <u>1999</u>	DEGREE: <u>Ph.D.</u>
NAME: <u>KOGAN, Leonid</u>	