Applications of Stochastic Modeling to Quantitative Finance and Operations Management

Kun Soo Park

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2010

ABSTRACT

Applications of Stochastic Modeling to Quantitative Finance and Operations Management

Kun Soo Park

This thesis consists of four essays on applications of stochastic modeling to problems in the areas of quantitative finance and operations management.

The first part contains two essays on quantitative finance, presented in Chapter 2 and 3. The essays develop stochastic process models designed to better understand the performance of hedge funds. Recently, the number of hedge funds and the amount of assets they manage have been increasing rapidly. However, hedge funds reveal relatively little about their performance, and, since hedge funds report their returns voluntarily, their performance data is limited and not clearly reliable. Thus, models of hedge fund performance that can be easily analyzed and fit to limited data are valuable.

In Chapter 2, we propose and develop a new stochastic process model to solve a specific problem for hedge funds: quantifying the premium from extended hedge-fund lockups. A lockup period for a hedge fund is a time period after making the investment during which the investor cannot freely redeem his investment. Recently, lockup periods have been increasing from one year to multiple years. Then, for an investor in hedge funds, it is important to calculate the premium in compensation deserved for the restricted investment opportunities imposed by an extended lockup restriction. We model returns from an investment in hedge funds with a discrete-time Markov chain (DTMC). We use this model to calculate the premium from an extended lockup period. One key modeling feature is the statistical persistence in the quality of relative returns of hedge funds, that is, the tendency for a fund that generates relatively high (or low) returns in a period to continue generating relatively high (or low) returns again in the next period. By solving systems of equations, we fit the Markov chain transition probabilities to three directly observable hedge fund performance

measures from the limited data: the persistence of return, the variance of return and the hedge-fund death rate. This so-called "calibration" of a model is a common and time-tested strategy in the practical use of contingent claim models. We also quantify how the lockup premium depends on the model parameters and the lockup period.

In Chapter 3, we extend the model just described: a stochastic-difference-equation is introduced to directly model the relative returns of a hedge fund. An important feature of the model is that for the relative returns of a hedge fund, the limiting distribution is easily analyzed. Just as in Chapter 2, we incorporate the persistence of returns in our modeling. Specifically, for the relative return X_n of a hedge fund in year n, we propose a stochastic-difference-equation of the form $X_n = A_n X_{n-1} + B_n$ where A_n represents persistence and B_n represents noise. This model is appealing because it involves relatively few parameters, can be analyzed, and can be fit to the limited and less reliable data reasonably well. We show that a simple model framework where A_n is constant and B_n is normal random variable provides a good fit for hedge funds with light return tails. We also show that the model within the same general framework can also be fit to the heavy-tail case successfully.

The second part contains two essays on operations management, presented in Chapter 4 and 5. These essays employ stochastic modeling to better understand operational decisions and behavior of firms in the business of procurement and supply chain management. In operations management, stochastic models are popular in modeling uncertainties in the demand of a customer or cost of a product. We study a procurement problem in operations management that interfaces with economics in Chapter 4 and a supply chain problem that interfaces with accounting in Chapter 5.

In Chapter 4, we consider a procurement system where a buyer wants to procure a product from sellers who have random production costs. We especially study a procurement that combines both auctions and bargaining, a combination that has become increasingly popular recently. Although both auction and bargaining in procurement have been studied extensively in the both economics and operations management literature recently, research that combines auctions and bargaining is limited. We model and analyze a combined auction and bargaining procurement system where an auction is followed by bargaining between the buyer and the winning seller in the auction. For this auction-bargaining model, we find a symmetric equilibrium bidding strategy for the sellers in a closed form. We also show that the buyer's expected profit in the combined procurement is higher than the profit in auction or bargaining only procurement.

In Chapter 5, we study the impact of a transfer pricing scheme for tax purposes for intra-firm transactions in the supply chain of a multinational firm. Although the impact of transfer pricing has been studied in the cost accounting literature, a detailed impact of the transfer pricing method on operational decisions and divisional profits in a supply chain has not yet been explicitly studied in both the cost accounting and operations management literature. In this chapter, we consider a supply chain where a retailer sub-division of a multinational firm orders a product from a manufacturing sub-division of the firm through an intra-firm transaction and sells it to customers under random demands. Our analysis shows that the problem can be analyzed as a variant of well known price-setting newsvendor framework in operations management. We also study the efficiency of a supply chain under the two popular transfer pricing schemes for tax reporting and show how transfer pricing methods affect operational decisions and profits of a firm and its sub-divisions.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deepest appreciation to my advisors and co-authors, Professors Emanuel Derman, Woonghee Tim Huh, and Ward Whitt. It was my great honor and pleasure to have an opportunity to work with them. I truly appreciate their encouragement and inspiration. This thesis would never have been possible without their kind and supportive guidance. Upon reflection, I feel so fortunate to learn and study under these distinguished scholars in operations research at Columbia University. I truly thank everyone for letting me have that exciting opportunity.

I also would like to thank Professors Paul Glasserman and Alp Muharremoglu for kindly agreeing to serve as dissertation committee members. I want to thank all the other professors in the department of industrial engineering and operations research, especially Professors Daniel Bienstock, Guillermo Gallego, Donald Goldfarb, Garud Iyengar, and Karl Sigman, from whom I took invaluable courses and with whom I had interesting opportunities to serve as a teaching assistant.

I thank my colleagues at Columbia University who always have inspired me and broadened my knowledge.

I give my special thanks to my wife, Minjeong Kang, and my family - my father, Jongwon Park and mother, Sooyoung Bae.

I finally thank God who constantly keeps me inspired to sail into the Sea of Knowledge.

To my wife and family

Chapter 1

Introduction

This thesis consists of four essays on the application of stochastic modeling to problems in the areas of quantitative finance and operations management. In the first part of the thesis, we study problems in the area of quantitative finance, focusing on hedge funds. In the second part, we study problems in the area of operations management, with special attention to the intersections of operations management with economics and accounting.

1.1 Contributions to Quantitative Finance

A stochastic process model of hedge fund return process is useful for research and investment analysis purposes. However, hedge funds are usually not required to report their returns to any authorities. Due to this, it is hard to find reliable and accurate data on hedge-fund returns. Hence, a model of hedge-fund returns that can easily be fit to the data and be analyzed and applied is valuable in hedge-fund research. In Part I, we propose stochastic process models that incorporate the persistence observed in hedge-fund returns. Persistence in returns is a tendency for a fund that generates relatively high (or low) returns in a period to continue generating relatively high (or low) returns again in the next period. We analyze hedge-fund returns data from the Tremont Advisory Shareholders Services (TASS) database over the period 2000 to 2005 and show that there exists a statistically significant degree of persistence in the annual relative returns of hedge funds. In this part of the thesis, we apply a discrete-time Markov chain (DTMC) that incorporates persistence to quantify the return premium for extended hedge-fund lockups in Chapter 2. In Chapter 3, we develop models based on a stochastic-difference-equation whose limiting distributions fit the same TASS data reasonably well.

In Chapter 2, we study how to quantify the premium from extended hedge fund lockups. A lockup period in a hedge fund is a time period after making the investment during which the investor cannot freely redeem his investment. It is routine to have a one-year lockup period, but recently hedge funds have requested longer lockup periods. Since investors lose an opportunity to redeem their money and re-invest in other funds until a lockup period ends, they deserve a premium for tying up their money, and the problem of quantifying the premium from a lockup is an important problem in hedge-fund research. We estimate the premium for such extended lockups, taking the point of view of a manager of a fund of funds, who has to choose between two investments in similar funds in the same strategy category, the first having a one-year lockup and the second having an n-year lockup. Assuming that the manager will rebalance his portfolio of hedge funds on a yearly basis, we define the annual lockup premium as the difference between the expected rates of return from these investments. We develop a Markov chain model to estimate this lockup premium. By solving systems of equations, we fit the Markov chain transition probabilities to three directly observable hedge fund performance measures: the persistence of return, the variance of return and the hedge-fund death rate. We contribute by quantifying the way the lockup premium depends on these parameters and the lockup period. Data from the TASS database are used to estimate the persistence, which is found to be statistically significant.

While Chapter 2 is focused on developing a stochastic process model to the lockup premium problem, Chapter 3 is devoted to developing a stochastic process model directly for the hedge-fund relative returns themselves. Despite the abundance of stochastic models for other securities like stocks, commodities and market indices, relatively few stochastic models have been developed for hedge funds. Since hedge-fund returns are reported only voluntarily by the funds, performance data of hedge funds are not only limited but also less reliable. In this chapter, we contribute by developing a stochastic-process model of the relative returns of a hedge fund. This model is appealing because it can involve relatively few parameters, can be analyzed, and can be fit to the limited and somewhat unreliable data reasonably well. Specifically, we propose a stochastic difference equation of the form $X_n = A_n X_{n-1} + B_n$ to model the annual returns X_n of a hedge fund relative to other funds in the same strategy group in year n. We let $\{A_n\}$ and $\{B_n\}$ be independent sequences of independent and identically distributed random variables, allowing general distributions, with A_n and B_n independent of X_{n-1}, where $E[B_n] = 0$. The key model parameters are the year-to-year persistence factor $\gamma \equiv E[A_n]$ and the noise variance $\sigma_b^2 \equiv Var(B_n)$. The model was chosen primarily to capture the observed persistence, which ranges from 0.11 to 0.49 across eleven different hedge-fund strategies, according to regression analysis. The constant-persistence normal-noise special case with $A_n = \gamma$ and B_n (and thus X_n) normal provides a good fit for some strategies, but not for others, largely because in those other cases the observed relative-return distribution has a heavy tail. We show that the heavy-tail case can also be successfully modelled within the same general framework. The model is evaluated by comparing model predictions with observed values of (i) the relative-return distribution, (ii) the lag-1 auto-correlation and (iii) the hitting probabilities of high and low thresholds within the five-year period.

1.2 Contributions to Operations Management

In the second part of the thesis, we study problems in operations management area, with special attention to interfaces with economics and accounting. In operations management, stochastic modeling is popular in modeling uncertainties on customer demand or production cost. In part II, we model random production cost in procurement and random customer demand in supply chain and apply it to analyze the problems below.

In Chapter 4, we study a profit optimization problem of a buyer in a procurement system. A procurement decision is one of the most important decision that a firm faces in managing its supply chain. It is involved with selecting suitable suppliers among many potential competing sellers and reducing the purchase cost. Both auctions and bargaining are popular in procurement in practice and they are also extensively studied in both the economics and operations management literatures. However, recently, auctions and bargaining are often combined in one procurement system, although research on this combined procurement is relatively limited. In this paper, we consider a combined auction-bargaining model in a setting where a single buyer procures an indivisible good from one of many competing sellers. The procurement model that we analyze is a sequential model consisting of the auction phase followed by the bargaining phase. In the auction phase, the sellers submit bids, and the seller with the lowest bid is selected as the winning bidder. In the bargaining phase, the buyer audits the cost of the winning seller and then negotiates with him to determine the final price. For this auction-bargaining model, we contribute by finding a symmetric equilibrium bidding strategy for the sellers in a closed form, which is simple to understand and closely related to the classical results in the auction and bargaining literature. We also prove that the auction-bargaining model generates at least as much profit to the buyer as the standard auction or sequential bargaining model.

In Chapter 5, we study a supply chain problem of a multinational firm, especially under a certain transfer pricing scheme for tax purposes for its intra-firm transaction. Transfer pricing refers to the pricing of an intra-firm transactions of an intermediate product or service. It has a significant impact on how the divisional performances are evaluated. Thus, transfer pricing is regulated by tax authorities who impose a set of pre-specified transfer pricing methods for tax purposes. Such regulations provide some flexibility to the firm's tax reporting practice, and the particular choice of the transfer pricing method can have a significant impact on the profits of the divisions and the entire firm. In this chapter, we contribute by studying how the transfer pricing for tax purposes affects operational decisions and the corresponding profits of a firm. We consider a firm consisting of two divisions (a manufacturing division and a retail division) where a retailer division faces a random customer demand in consideration. The retail division sets the retail price, and orders an intermediate product from the upstream manufacturing division. The manufacturing division accepts or rejects the retail division's order. We particularly consider two commonly used transfer pricing methods for tax purpose – the cost-plus method and the resale-price method. We also extend our analysis to a case of a multinational firm whose manufacturing and retailing divisions are under different income tax rates. Our analysis of both the retailer's problem and the central planner's problem is based on the classical price-setting newsvendor problem and its variants. We show several sensitivity results to the parameters of these methods. Numerical results indicate that the cost-plus method tends to allocate a higher percentage of profit to the retail division while the resale-price method tends to achieve a higher firm-wide profit. These results suggest that the choice of transfer pricing method has a significant impact on the profit as well as operational decisions of a firm.

Part I

Contributions to Quantitative Finance

Chapter 2

Estimating the Premium for Extended Hedge Fund Lockups

2.1 Introduction

A *lockup period* for investment in a hedge fund is a time period after making the investment during which the investor cannot freely redeem his investment. It is routine to have a one-year lockup period, but recently the requested lockup periods have grown longer. It is reasonable for an investor in a hedge fund to expect compensation for the restricted investment opportunities imposed by an extended lockup condition, with the compensation increasing as the length of the lockup period increases. We regard that compensation as a *lockup premium*, and we ask: What should that lockup premium be as a function of the length of the lockup period?

In asking this question, we take the point of view of a manager of a fund of funds, who has to choose between two investments in similar funds in the same strategy category, with the first having a one-year lockup and the second having an n-year lockup. We assume that the manager will re-balance his portfolio of hedge funds on a yearly basis, as permitted. This perspective leads us to *define* the lockup premium as the incremental deterministic return rate required to make the expected total returns of the two alternatives equal. Our definition accounts for lost gains due to the inability to re-balance the investment portfolio in hedge funds, but not for other lost investment opportunities, so we provide a conservative estimate of the lockup premium. Investors can separately consider the consequence of other lost investment opportunities, if that is desired. Indeed, recent financial history indicates that the other component may be very important. Nevertheless, for clarity, we think it is desirable to separate these issues.

With that definition specified, our goal is to develop a mathematical model to estimate the lockup premium as a function of the lockup period and key hedge-fund performance measures. There are significant challenges in deciding what modelling approach to use. We want a model that is easy to understand, properly reflects the specific lockup conditions, has predictive power, can be effectively analyzed and can be fit to available data.

These requirements lead us to propose a relatively simple *three-state Markov chain model.* This model formulation is admittedly highly stylized, but we think that actually is an advantage rather than a disadvantage, because the data and their quality are quite limited. Nevertheless, this stylized model may be viewed with skepticism, because it is unfamiliar. It is thus good to remember that many of the most frequently used models are highly stylized, having very few parameters; e.g., the geometric Brownian motion model underlying the Black-Scholes formula.

By introducing a model with relatively few parameters, we have fewer parameters to fit to data. In this context, we contribute by developing an innovative way to calibrate (fit) the model to data. We do not directly fit the natural model parameters, which are the Markov chain transition probabilities and the state-dependent returns, but instead we *indirectly fit the model* to more directly observable hedge fund performance measures, specifically, the persistence of return, the variance of return and the hedge-fund death rate. This indirect approach requires that we solve systems of equations to determine the required model parameters. We carry out this model fitting using hedge-fund return data from the Tremont Advisory Shareholders Services (TASS) database.

Even though estimating the value of the premium for hedge-fund lockup is a liquidity problem similar to determining the appropriate rate of return for a long term certificate of deposit, it has its own special character. There is a complication with hedge funds, because investors may actually have an early opportunity to redeem their investment. If the hedge fund performs very poorly, so that it ceases operating, then a significant portion of the investment is returned to investors, even if the lockup period has not expired. At first, glance, it might appear that consequently there should be no liquidity problem at all, but the two extreme alternatives are not the only possibilities: Hedge fund performance may be weak, so that returns are low and future prospects are dim, even though the fund does not cease operating. The lockup prevents the investor from moving his investment away from such "sick" funds. This special way hedge fund lockup is treated makes the liquidity premium more complicated, providing motivation for more careful analysis.

Our proposed model directly responds to this special feature of hedge fund investments: We consider *three possible states for a hedge fund: good, sick and dead*, and we assume that transitions among these states occur randomly according to a Markov chain. In a dead state, the investor suffers a low return, but at the next yearly reinvestment opportunity the state changes to a good state, because the investor gets his money back and can invest in a new fund, which we take to be in the good state. (We assume that the investor gets all his investment back, even though he suffers the low return.) There is no extra penalty from the lockup associated with a dead fund, but there is from a sick fund. With only nominal one-year hedge fund lockup, we assume that investors will reinvest in a good fund at the next yearly reinvestment opportunity whenever any fund they have invested in becomes sick. In contrast, with the extended lockup period, no reinvestment is possible until the end of the lockup period. In the meantime, the sick fund may perform poorly, and produce low returns, but there also is a chance that it may rebound and become a good fund. Clearly, some care is needed to properly account for the various good and bad possibilities, which inevitably must be regarded as random events. The Markov chain models can capture the behavior described above, so provide insight.

It remains to specify the three Markov chain states. We propose classifying the funds according to their return rates. Specifically, we focus on the relative return rates, represented as the percent-point difference from the average annual return rate for that strategy category of funds. We say that a fund is in a: *good* state when its relative return rate is higher than U percent, *sick* state when its relative return rate is between L and U percent, and *dead* state when its relative return rate is less than L percent. We leave U and L as model parameters in general. Figure 2.1 illustrates possible state definitions in a plot of the distribution of annual return rates based on 4788 selected observations from 2001 to 2005 from the TASS database. Tentative levels U and L show how states might be defined. Throughout this chapter, we assume that the hedge fund starts off in a good state.

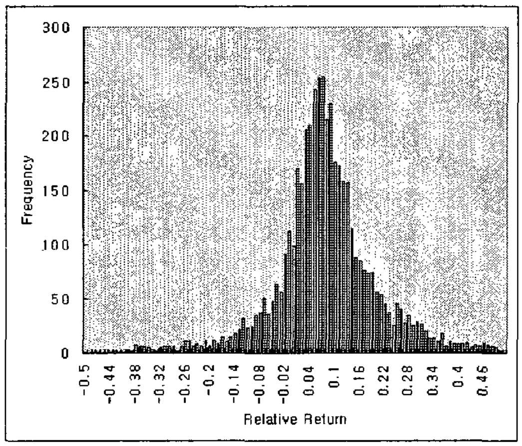


Figure 2.1: The distribution of hedge fund annual relative return rates based on 4788 selected observations from the TASS database from 2001 to 2005. Tentative levels L and U divide the funds into the three states G, S and D.

A fundamental principal guides our analysis: the *persistence hypothesis.* We postulate that there is a persistence in hedge fund performance within a particular hedge fund strategy category: We assume that above-average funds will tend to continue doing well, while below-average funds will tend to continue faltering. A persistence of γ means that *for every* 1 *percentage point you earn above the average in the current year, you expect to earn* γ *percentage points above the average in the next year.*

We estimated the persistence by doing a regression analysis on the hedge-fund-return data from TASS, and found strong statistical evidence to support the persistence hypothesis. There are eleven strategy categories of hedge funds in TASS. (See, e.g., Boyson and Cooper (2004), Hasanhodzic and Lo (2007) or visit Credit Suisse/Tremont (www.hedgeindex.com) to find out more about strategy categories.) Table 2.1 shows the auto-regression results from the data we selected. We did the analysis by strategy category. Included are 95% confidence intervals for each persistence factor (the regression coefficient). Zero persistence is contained in the 95% confidence interval for only three strategies. The P-values give the probability

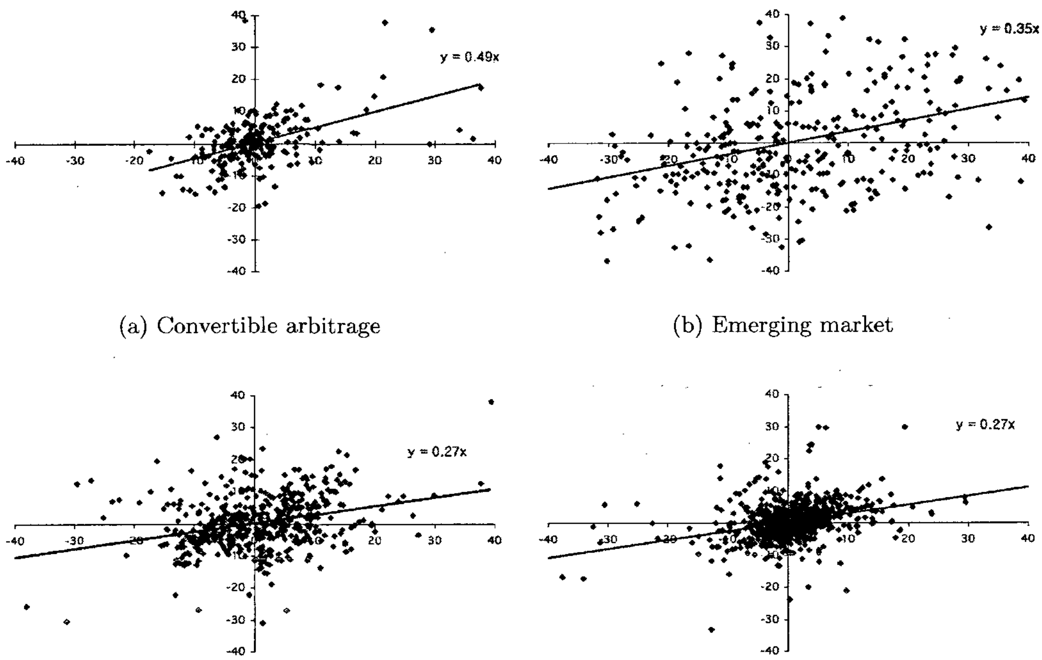
Table 2.1: Auto-regression analysis results

strategy	number of observations	persistence γ	lower 95%	upper 95%	R^2	P-value
Convertible arbitrage	244	0.49	0.38	0.60	0.24	4.24×10^{-16}
Dedicated short bias	30	0.29	-0.04	0.62	0.10	0.08
Emerging market	325	0.35	0.26	0.45	0.13	1.02×10^{-11}
Equity macro	270	0.06	-0.05	0.16	0.004	0.28
Event driven	534	0.27	0.20	0.34	0.09	9.27×10^{-13}
Fixed income arbitrage	196	0.24	0.12	0.36	0.07	1.28×10^{-4}
Fund of fund	982	0.27	0.22	0.32	0.10	4.48×10^{-24}
Global macro	176	0.10	-0.06	0.27	0.009	0.21
Long short equity	1654	0.15	0.11	0.20	0.03	8.03×10^{-12}
Managed future	238	0.22	0.09	0.35	0.04	1.12×10^{-3}
Other	167	0.42	0.27	0.57	0.15	1.34×10^{-7}
All	4816					

of seeing the observed persistence if there actually were none. The estimated persistence factors vary, but for most strategy categories, the P-values are very small. The regression analysis shows that R^2 is very low, implying that there is considerable randomness. To illustrate, Figure 2.2 is the scatter plot of two consecutive relative returns and the associated least-squares-fit with zero intercept for four of these fund categories.

In this chapter, we only consider fund strategy as a basis for persistence. Other classifications might also produce persistence; e.g., one can estimate persistence based on the fund manager's tenure, asset size, fee structure, and so on, depending on the investor's judgement. As long as persistence is found or anticipated, our Markov chain model can be applied to estimate the lockup premium.

The Markov chain model can be used to estimate how the lockup premium depends on



(a) Convertible arbitrage (b) Emerging market

(c) Event driven (d) Fund of fund

Figure 2.2: Scatter plots and associated least-squares lines for hedge fund annual relative return rates in successive years from 2000 to 2005 for four strategy categories from the TASS database.

the hedge-fund performance measures. Consistent with intuition, we show that the lockup premium increases with both the variance of the return and the persistence of the return, provided that the persistence is not too high. (There necessarily is no lockup premium with total persistence, when $\gamma = 1$.) What is less obvious, but consistent with intuition upon reflection, is that the lockup premium *decreases* with the hedge fund death rate. Of course, increased death rate is bad for the investor, but the investor experiences the low return associated with a dead fund whether or not there is a lockup. With the extended lockup, the higher death rate can only help by giving the investor an opportunity to reinvest his money.

The models do more: The models *quantify* the effect of these observable hedge fund performance measures on the lockup premium. For example, we show that the three-year lockup premium in the DTMC model is quite well approximated by the function

$$\psi(\delta, \gamma, \sigma) = 0.047\ \delta^{-0.11}\gamma^{0.69}\sigma^{0.64}, \tag{2.1}$$

where δ is the death rate, γ is the persistence and σ is the standard deviation of the yearly relative returns (under parametric assumptions to be explained later).

Organization of the chapter. We start in §2.2 by reviewing the related literature on liquidity, including premiums for hedge fund lockup. In §2.3 we carefully specify what we mean by the lockup premium. In §2.4 we discuss persistence of hedge fund returns, reviewing the literature and analyzing data from the TASS database. In §2.5 we develop a simple approximation for the lockup premium based on persistence alone, without any Markov chains, assuming no dying funds. This simple analysis provides a useful reference case, because it yields a simple formula. In §2.6 we introduce and analyze our three-state DTMC model, carrying out our indirect fitting procedure. In §2.7 we show how the model parameters and the lockup premium depend on basic hedge fund performance measures. Finally, in §2.8 we draw conclusions. We present additional material in an appendix.

2.2 Liquidity Literature Review

There is a substantial literature on liquidity, including hedge fund lockup, but it mostly has a different character.

Liquidity premiums in asset pricing. The liquidity premium is well recognized as an important factor in asset pricing, but it is commonly measured by transaction cost; e.g, see Amihud and Mendelson (1986), Pastor and Stambaugh (2003), Chordia et al. (2001), and Eleswarapu and Reinganum (1993). For example, in the stock market, bid-ask spread is one measure of the liquidity premium. Amihud and Mendelson (1986) showed that there exists an increasing and concave relationship between the asset return and the bid-ask spread. Darar et al. (1998) confirmed this result, using the reciprocal of the stock turnover rate to measure the liquidity premium. More recently, Vayanos (2004) considered liquidity in an equilibrium model. He considered the liquidity premium in asset pricing with different transaction costs. He showed that as assets become more volatile, the required excess return from a riskless asset increases with the transaction costs.

Studies of liquidity have also been performed for the bond market; e.g., Amihud and Mendelson (1991), Warga (1992), Krishnamurthy (2002) and Longstaff (2004). For bonds, it is argued that there should exist a clear premium for liquidity, separate from the credit risk premium. Most-recently-issued U.S. Treasury bonds are considered the most liquid bonds available, among all bonds with similar conditions. Since US Treasury bonds are assumed to be riskless, they provide a natural way to measure the liquidity premium, without having to consider credit risk. The papers above study the liquidity premium by comparing the price of most-recently-issued US Treasury bonds (on the run) to the price of the bonds issued three months previously (off the run).

There are a few papers that are more closely related to what we do here, namely, Longstaff (1995, 2001) and Browne and Whitt (1996). These papers also view the liquidity premium as arising from the investor's inability to rebalance his portfolio in a timely way. Specifically, they define the liquidity premium as the additional required fixed return to compensate for the loss of the investor's utility from the inability to rebalance the investor's portfolio. They calculate the required liquidity premium as a function of the degree of risk averseness in the utility function, the market growth rate, and the liquidity restriction period. They rely heavily on mathematical models and mathematical analysis for this purpose. Unlike these references, we do not use utility functions.

We conclude this section by mentioning Hayes (2006), which used a Markov chain model for a difference purpose – to develop a model for the maximum drawdown of hedge funds.

Empirical studies on hedge fund lockup. There is a growing literature on hedge funds, e.g., see Agarwal and Naik (2005), but only a few researchers have focused on hedge fund lockup. Liang (1999) found that the average hedge fund returns are related positively to the lockup periods from the analysis of Hedge Fund Research, Inc. (HFR) database. Aragon (2007) quantified the lockup premium for hedge funds empirically. He compared the hedge fund performance with and without extended lockup conditions. He estimated that the average difference in the annual returns is around 4 to 7 percentage points.

There also are empirical studies on the liquidity premium for funds other than hedge funds. For example, Ippolito (1989) conducted a similar study for mutual funds. There is a load-type mutual fund, which assesses sales charges. Ippolito (1989) found that the load-type mutual funds make approximately 3.5 percentage points higher return than no-load mutual funds.

In summary, from our investigation of the literature, we find that only a few papers - Longstaff (1995, 2001) and Browne and Whitt (1996) - have interpreted liquidity premium as quantification of the cost of a restricted rebalance opportunity. We found no previous papers employing models calibrated to data in order to estimate the liquidity premium. And none of the papers have used Markov chains, with the exception of Derman et al. (2009a), which is a preliminary account of the research reported here.

2.3 Definition of the Lockup Premium

In this section we carefully specify what we mean by the lockup premium. To do so, we first define the rate of return of an investment having a *stochastic process* $X \equiv \{X(t) : t \geq 0\}$; i.e., the (random) value at the end of n years of one dollar invested in this investment at the beginning of the first year is

$$V_n \equiv e^{\left(\int_0^n X(t)\,dt\right)} \tag{2.2}$$

dollars. We now define a deterministic value r_n such that

$$e^{nr_n} = \mathbb{E}[V_n] \tag{2.3}$$

for V_n in (2.2); i.e., r_n is the constant rate of return, with continuous compounding, that yields the same expected value $\mathbb{E}[V_n]$ over n years. Following common practice, we have "backed out" the rate of return r_n from the expected cash value $\mathbb{E}[V_n]$. By (2.2) and (2.3), r_n can be expressed directly as

$$r_n = \frac{\log \mathbb{E}[V_n]}{n} = \frac{1}{n} \log\left(\mathbb{E}\left[e^{\left(\int_0^n X(t)\,dt\right)}\right]\right), \tag{2.4}$$

where we use the natural logarithm (base e).

Now consider two different hedge funds within the same strategy, with rate of return stochastic processes X^1 and X^2, as above. Let X^1 have a 1-year lockup and let X^2 have an n-year lockup. Let the *lockup premium* p_n be

$$p_n \equiv r_n^1 - r_n^2, \tag{2.5}$$

where r_n^i is the rate of return of X^i, defined as in (2.4). To evaluate the premium, we need to determine the two return-rate stochastic processes X^1 and X^2; to do that, we will apply the Markov chain model.

However, we do not actually have the continuous-time return-rate stochastic processes directly available from the TASS database. Instead, monthly returns are reported. Consistent with the framework above, we define a continuously compounded annual returns B_i for year i from monthly returns $M_{i,j}$ for the j^{th} month within year i by geometric compounding, i.e.,

$$e^{B_i} = (1 + M_{i,1})(1 + M_{i,2}) \cdots (1 + M_{i,12}); \tag{2.6}$$

i.e., the (random) value at the end of j months of one dollar invested in this investment at the beginning of year i is $(1 + M_{i,1})(1 + M_{i,2}) \cdots (1 + M_{i,j})$ dollars. Consequently, the (random) total value V_n at the end of n years is the n-fold product

$$V_n = \prod_{i=1}^{n} e^{B_i} = e^{\sum_{i=1}^{n} B_i} \tag{2.7}$$

for A_i in (2.6). Equivalently, starting from the reported monthly returns $M_{i,j}$, we let the return-rate stochastic process X be defined by

$$X(t) \equiv 12 \log(1 + M_{i,j}) \quad \text{for} \quad (i-1) + (j-1)/12 \leq t < (i-1) + (j/12), \tag{2.8}$$

so that

$$\int_{(i-1)+(j-1)/12}^{(i-1)+(j/12)} X(t)\, dt \equiv \log(1 + M_{i,j}) \tag{2.9}$$

for $1 \leq i \leq n$ and $1 \leq j \leq 12$. With definition (2.8), equation (2.7) is consistent with definition (2.2).

As indicated above, we start with the monthly returns $M_{i,j}$ and then construct the annual rate of return B_i by geometric compounding, as in (2.6). In order to reduce the effects of systematic yearly fluctuations, and more closely approach stationarity, we focus on *relative return rates* for each fund strategy. To do so, we let $\mu_i \equiv \mathbb{E}[B_i]$, the mean return rate for a particular hedge fund strategy within year i, estimated as the average of the observed values of B_i over all funds within that strategy. Then the (random) relative return rate is

$$R_i \equiv B_i - \mathbb{E}[B_i] \equiv B_i - \mu_i. \tag{2.10}$$

We exploit persistence of hedge fund returns in the setting of these relative return rates R_i.

Combining equations (2.7) and (2.10), we see that the (random) total value at the end of year n of hedge fund j is

$$V_n^j = \prod_{i=1}^{n} e^{(\mu_i + R_i)} = e^{(\sum_{i=1}^n \mu_i)} e^{(\sum_{i=1}^n R_i^j)} \tag{2.11}$$

and the difference between the expected total returns is

$$\mathbb{E}[V_n^1] - \mathbb{E}[V_n^2] = e^{(\sum_{i=1}^n \mu_i)} \left(\mathbb{E}\left[e^{(\sum_{i=1}^n R_i^1)}\right] - \mathbb{E}\left[e^{(\sum_{i=1}^n R_i^2)}\right]\right). \tag{2.12}$$

Hence, the premium in (2.5) becomes

$$p_n \equiv r_n^1 - r_n^2 = \frac{1}{n}\left[\log\left(\mathbb{E}\left[e^{(\sum_{i=1}^n R_i^1)}\right]\right) - \log\left(\mathbb{E}\left[e^{(\sum_{i=1}^n R_i^2)}\right]\right)\right], \tag{2.13}$$

which is independent of the average rates μ_i.

To determine what this lockup premium p_n should be, we will be modelling the relative-return-rate stochastic process $\{R_k^1 : k \geq 1\}$ by a discrete-time Markov chain, and then defining the associated relative-return-rate stochastic process $\{R_k^2 : k \geq 1\}$ to account for the extended n-year lockup. With (2.13), we will also set the initial state as Good, as mentioned in §2.1.

2.4 Persistence of Hedge Fund Returns

As indicated in §2.3, we specify how hedge funds perform by looking at the *relative rate of return* of a fund, given by the R_i in (2.10). In that context, we say there is persistence if R_{i+1} tends to be similar to R_i. in particular, we measure persistence by the regression coefficient for pairs (R_i, R_{i+1}). Before discussing our regression analysis, we review the literature on persistence.

The persistence literature. Persistence has been studied quite extensively within the hedge-fund literature, but it remains a highly controversial topic. A consensus has not yet been reached on the degree of persistence in hedge-fund returns, or even whether it exists at all. Also, there are differences in the specific definition of persistence; e.g., Jagannathan et al. (2006), Fung et al. (2008) and Kosowski et al. (2007) are about alpha persistence. However, persistence always represents the ability to predict future returns from past and present returns.

There are serious questions about the quality of the data and the proper way to analyze it. Researchers have tried to take advantage of the two main hedge fund databases - TASS and HFR. In doing so, researchers have discovered that it is difficult to make unbiased estimates because reporting is voluntary, and some funds stop reporting, especially those performing poorly; see Jagannathan et al. (2006), Fung et al. (2008), Fung and Hsieh (2009) and Kosowski et al. (2007).

Despite the difficulty with biases in the hedge fund data, researchers have conducted studies. Although some researchers did not find evidence of performance persistence, others did. Brown et al. (1999) used a simple two-state categorization - win or lose - to measure performance persistence, recording a win if the fund beats the median return, but they did not find evidence of persistence. Boyson and Cooper (2004) carried out a similar analysis and still did not find evidence of persistence.

However, several papers found performance persistence for shorter periods ranging from a quarter to three years. Koh et al. (2003) used the method of Brown et al. (1999) for Asian hedge funds and found strong persistence in short horizons from monthly to quarterly. Agarwal and Naik (2000) and Jagannathan et al. (2006) used linear regression, as we do, as well as the previous two-way classifications. Agarwal and Naik (2000) did not provide regression slope explicitly but showed that depending on the strategy category of hedge fund, the percentage of funds which have statistically significant positive slope in regression ranges from 5 to 34 percent, where most of the strategy categories have around 20 percent. Using the same parametric linear regression and non-parametric two-way classifications, Agarwal and Naik (2000) claimed that the evidence of persistence is strongest for the shorter quarterly time periods. On the other hand, Edwards and Caglayan (2001) found strong persistence in over 1-2 years from the Managed Account Reports (MAR/Hedge) data. Furthermore, Jagannathan et al. (2006) found a significantly high performance persistence for a three-year period in their empirical study with HFR data. Jagannathan et al. (2006) carefully took account of the bias from voluntary reports and did regression analysis for the relative returns for three consecutive years. Using generalized method-of-moment (GMM) estimation, they found a statistically significant persistence factor of 0.56 for a three-year period.

There also exists indirect evidence of performance persistence from the study of hedge-fund liquidation or survival. Brown et al. (2001) indirectly supported performance persistence when they found that a negative aggregated return over the previous two years increases the probability that a fund will liquidate. Furthermore, Horst (1971) concluded that hedge-fund survival is strongly related to historical performance. Baquero et al. (2005) conducted probit regression analysis of hedge-fund liquidation. They found that funds with high returns are much less likely to liquidate than funds with low returns from quarterly return data, which again indirectly supports persistence.

Our regression analysis. As indicated in §2.1, we conducted linear autoregression analysis with the TASS data to find the best linear regression line between two consecutive year's relative rates of return (the R_i in (2.10)). Specifically, letting the current year's (annual) relative return rate be denoted by R_c and the next year's relative return rate be denoted by R_n, we find the slope γ for the line $R_n = \gamma \cdot R_c$, which produces the minimum sum of squared errors.

The actual data analysis procedure is somewhat complicated. A fund usually keeps reporting its monthly returns as long as it continues operating. If a fund ceases reporting its returns to TASS, then the last date of the report is marked as the *drop date* in the data. A fund may stop reporting its returns if it is liquidated due to successive losses. However, it is not always true that a hedge fund suffers huge losses when it ceases reporting. In fact, even a successful fund may cease reporting if it no longer wants to reveal its performance publicly. Thus, we cannot simply count the number of funds dropped from the data to estimate the liquidation rate of hedge funds. If the reason why a fund ceases reporting is available, TASS reports it, but the reason is often not reported.

As mentioned above, TASS differentiates between the date the fund starts reporting and the date the fund starts operating. Thus, we can exclude one possible bias: the so called *backfill bias.* When a fund starts reporting returns after operating for several months or years, the fund simultaneously reports several monthly returns at the time its first return is reported. It is then possible for the fund manager to drop or change some bad monthly returns which have been made before the reporting date, which may increase reported returns from the actual returns. Fung and Hsieh (2000) calculate that the difference from actual returns and reported returns is about 3.6% per year from this reason. Therefore, we consider monthly returns only after the fund's first reporting date. Similarly, if a fund's monthly returns are reported less than six times a year, we exclude these data due to the possibility of hiding or altering bad returns.

The other criterion we consider is the Net Asset Value (NAV) managed by a fund, which is also archived in TASS. If a fund's managed assets are too small, then the monthly return might be too volatile, since it may have relatively less ability to diversify its risks. We assume that a fund has an ability to produce relatively stable returns once its managed assets reach a certain level. Specifically, we consider monthly returns only if the fund's NAV has reached 25 million dollars at least once, at which point we assume that the fund becomes mature, so that it can produce relatively stable returns. Similar criteria were used by Boyson and Cooper (2004).

Before conducting the regression, we also exclude pairs of return rates with extreme values, depending on the distribution of the pairs of returns for each strategy category. Even one or two outliers can seriously affect the regression, especially if we do not have a large number of observations. Specifically, we exclude pairs of relative returns when one absolute relative return exceeds ± 30% for fixed income arbitrage, equity macro and ± 40% for convertible, dedicated short bias, and global macro strategy categories. We also exclude pairs of relative returns exceeding ± 50% for emerging market, event driven, fund of fund, long/short equity, managed future, and others strategy categories of funds.

After selecting the monthly returns based on the above criteria, we make pairs of two successive annual returns for each hedge fund from 2000 to 2005. Thus, there are possibly six pairs of annual returns of a fund, if it does not cease reporting during that period. As indicated in (2.6), the monthly returns are annualized to produce annual returns, from which we calculate relative rates of return R_i as indicated in §2.3. The regression analysis results in very low intercept for all strategy categories. Thus, we conducted an auto-regression without an intercept to obtain our final estimate. The results are shown in Figure 2.2 and Table 2.1 in §2.1. As can be seen there, we found eight out of eleven strategies of fund with significant persistence: (i) convertible, (ii) dedicated short bias, (iii) emerging market, (iv) event driven, (v) fixed income, (vi) fund of fund, (vii) managed future, and (viii) others. For these fund strategies, the least-squares-fit slope, γ, ranges from 0.15 to 0.49.

A different way to estimate the persistence factor is to look at the ratio of the next-year average return rates to the current-year average return rate, restricting attention to the returns that are positive in the current year. (The same estimate is produced when you repeat that procedure, but instead restrict attention to the return rates that are negative in the current year.) See Appendix A.3 for the details.

2.5 An Approximation for the Lockup Premium Based on Persistence Alone

Given the expression for the lockup premium p_n in (2.4), (2.5) and (2.13), it should be evident that no exact analysis is possible based on persistence alone. However, we now show that it is possible to obtain a useful rough approximation for the lockup premium based on persistence alone if we make an appropriate approximation in our definition of the lockup premium.

Modifying the definition of the premium. The idea is to simplify the expression for the rate of return r_n in (2.4). Expression (2.4) is complicated primarily because the expectation operator appears in between the logarithm and the exponential functions, so they cannot cancel each other out. What we do for our rough approximation, then, is act as if we can interchange the order of the expectation operator and the exponential function.

In the setting of (2.13), that yields the approximation

$$\begin{aligned} p_n \approx \tilde{p}_n &\equiv \frac{1}{n}\left[\log\left(e^{(\mathbb{E}[\sum_{i=1}^n R_n^1])}\right) - \log\left(e^{(\mathbb{E}[\sum_{i=1}^n R_n^2])}\right)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}[R_i^1] - \mathbb{E}[R_i^2]\right). \end{aligned} \tag{2.14}$$

Unlike p_n in (2.13), the approximation $\tilde{p}_n$ in (2.14) is a linear function of the expected return rates $\mathbb{E}[R_i^j]$, and so is much easier to analyze.

We can also regard approximation (2.14) as an approximation derived from asymptotic analysis, where we use the classic approximations $\log(1+x) \approx x$ and $e^x \approx 1 + x + \frac{1}{2!}x^2$ for x close to 0. A one-term approximation is (2.14), while the two-term approximation is

$$p_n \approx \tilde{p}_n + 1/(2n)\left(\mathbb{E}\left[(\sum_{i=1}^{n} R_i^1)^2\right] - \mathbb{E}\left[(\sum_{i=1}^{n} R_i^2)^2\right]\right). \tag{2.15}$$

The second term explains most of the error for small n, e.g., for $n \leq 5$; see §A.5 in the Appendix.

Assumptions based on persistence alone. We now show how persistence alone, without any Markov chains, can be used to generate an estimate of the lockup premium, provided that we use the linear approximation (2.14). This simple analysis depends on four additional assumptions:

1. There is a single persistence factor γ with $0 < \gamma < 1$.

2. We can ignore the phenomenon of hedge funds dying.

3. The return rates R_i each year are normally distributed with a fixed variance σ^2.

4. The performance of a fund is considered good if its annual return exceeds the average annual return.

Together with approximation (2.14), the first two assumptions imply that the expected relative returns over time evolve *linearly*, enabling us to derive a simple *approximate no-death lockup premium* as a function of the expected excess return rate of a good fund. The last two assumptions enable us to determine the expected excess return rate of a good fund. The third assumption can be weakened, but some analogous assumption is needed. The fourth assumption is just one possible case; it can easily be varied without altering the rest of the analysis.

The no-death lockup premium. Let Y_G denote the expected excess relative rate of return of a good fund, assumed to be strictly positive. As in §2.3, let R_n be the relative return rate in the n^{th} year. As mentioned in §2.1, we assume that the hedge fund starts off in a good state. Then, the assumed γ persistence implies that the expected relative return rate during the first year is $\mathbb{E}[R_1] = \gamma Y_G$, for $0 \leq \gamma < 1$. The notion of γ persistence, with no funds dying, implies that we can *recursively* determine the expected relative return rates in successive years by

$$\mathbb{E}[R_n] = \gamma \cdot \mathbb{E}[R_{n-1}] = \gamma^n \cdot Y_G, \quad n \geq 1 \ . \tag{2.16}$$

As a consequence of (2.16), the sum of the expected relative return rates up to the n^{th} year can be expressed as a product

$$\sum_{i=1}^{n} \mathbb{E}[R_i] = Y_G \left(\frac{\gamma(1-\gamma^n)}{1-\gamma} \right) \ , \ 0 \leq \gamma < 1 \ . \tag{2.17}$$

Combining this simple analysis with approximation (2.14), we can compute the approximate premium for an n-year lockup compared to 1-year lockup. Under a 1-year lockup, investors have a chance to replace all sick funds with good funds at the end of each year. If they do, the expected return each year is the same as in the first year: $\mathbb{E}[R_1] = \gamma Y_G$. Thus, at the end of the n^{th} year, the total expected relative return is simply $n\gamma Y_G$. On the other hand, under an n-year lockup, the fund just evolves without replacement up to the n^{th} year, as in (2.17). We assume that after the n^{th} year, the funds with 1-year and n-year lockups are both replaced by funds with the same 1-year lockup, so that there necessarily will be no difference in a fund's return after the n^{th} year.

Consequently, the *approximate no-death lockup premium* is

$$\begin{aligned} \tilde{p}_n \equiv \tilde{p}_n(\gamma, Y_G) &= \left(\frac{1}{n} \sum_{i=1}^{n} \gamma Y_G \right) - \left(\frac{1}{n} \sum_{i=1}^{n} \gamma^i Y_G \right) \\ &= Y_G \gamma \left(1 - \frac{1-\gamma^n}{n \cdot (1-\gamma)} \right), \quad n \geq 1 \ , \end{aligned} \tag{2.18}$$

which is a concave increasing function in n for each γ, $0 < \gamma < 1$, and a concave function of γ for each $n \geq 1$. The approximate lockup premium $\tilde{p}_n(\gamma)$ is *not* an increasing function of γ overall; e.g., for $n = 2$, $\tilde{p}_n(\gamma) = Y_G\gamma(1-\gamma)/2$, which is increasing for $0 < \gamma < 1/2$, but decreasing for $1/2 < \gamma < 1$. More generally, $\tilde{p}_n = 0$ for both $\gamma = 0$ and $\gamma = 1$, with $\tilde{p}_n(\gamma) > 0$ for $0 < \gamma < 1$. However, the lockup premium function $\tilde{p}_n(\gamma)$ is increasing in γ for all sufficiently small γ, for each $n \geq 1$.

From (2.18), we see that $\tilde{p}_1 = 0$, $\tilde{p}_n \to \gamma Y_G$ as $n \to \infty$, and we have the bounds

$$\gamma Y_G \left(1 - \frac{1}{n(1-\gamma)}\right) \leq \tilde{p}_n \leq \gamma Y_G \left(1 - \frac{1}{n}\right), \quad n \geq 1 , \tag{2.19}$$

which yield convenient approximations. For large n or small γ, the lower bound is an accurate approximation of $\tilde{p}_n$.

The excess rate of return from a good fund. The approximate no-death lockup premium function $\tilde{p}_n(\gamma)$ clearly shows how the approximate lockup premium depends on the three quantities: the length n of the extended lockup period, the persistence factor γ and the expected excess rate of return of a good fund, Y_G. Clearly, n is directly observable, and we have seen how we can estimate γ, but it remains to specify Y_G.

However, if we define Y_G as the expected excess rate of return of a good fund and apply the last two assumptions, then we can calculate Y_G as well. Letting $N(m, \sigma^2)$ denote a normally distributed random variable with mean m and variance σ^2, we have

$$Y_G = \mathbb{E}[N(0,\sigma^2)|N(0,\sigma^2) > 0] = \mathbb{E}[|N(0,\sigma^2)|] = \sigma\mathbb{E}[|N(0,1)|] = \sqrt{2/\pi}\sigma \approx 0.8\sigma . \tag{2.20}$$

We can combine (2.18) and (2.20) to obtain the following general approximate no-death lockup premium function

$$\tilde{p}_n(\gamma, \sigma) = 0.8\sigma\gamma \left(1 - \frac{1-\gamma^n}{n \cdot (1-\gamma)}\right), \quad n \geq 1 . \tag{2.21}$$

With assumptions 3 and 4 above, we see that the no-death lockup premium should be approximately directly proportional to the standard deviation σ. Assumption 4 plays a key role in getting the simple formula (2.20), but we can generalize for arbitrary boundary point U, using the following formula for the conditional expectation of a normal random variable:

$$\mathbb{E}[N(m,\sigma^2)|a \leq N(m,\sigma^2) \leq b] = m + \sigma \frac{[\phi((a-m)/\sigma) - \phi((b-m)/\sigma)]}{[\Phi((b-m)/\sigma) - \Phi((a-m)/\sigma)]} \tag{2.22}$$

for $-\infty \leq a < b \leq +\infty$, based on the relation $x\phi(x) = -\phi'(x)$ where ϕ is the standard normal density. From formula (2.22), we see that Y_G will *not* be proportional to σ if we change the upper boundary point U.

We emphasize that, even under assumption 4 above, having $\tilde{p}_n$ be directly proportional to σ depends critically on the third ceteris-paribus assumption made above. Since we are free to choose the monetary units, we can choose to define all returns relative to the standard deviation σ, which must be in the same units as the returns. In that sense, the lockup premium is *automatically* proportional to σ. The proportionality conclusion becomes more meaningful when we assume that the distribution of returns depends on σ as a simple scale factor, as provided by assumption 3 above. We need to impose a strong condition on the way the return distribution changes with σ in order to deduce the desired proportionality conclusion. The normality is only used to compute the precise value of the mean.

Relating to the calibration by Markov chains. We remark that the Markov chain model calibration will also produce its own estimates of the excess return Y_G, but we will find that analysis yields similar conclusions. Indeed, our main numerical example has $Y_G = 0.67\sigma$. We remark that we can obtain exactly that value if we take Y_G to be the *median* of the positive relative returns, because the median of the random variable $|N(0,1)|$ is 0.67.

Anticipating our future numerical examples with Markov chains, we refer to our estimate for the lockup premium in Figure 2.8 in §2.6.7 for the case $\gamma = 0.5$, $\sigma = 0.1$, $\delta = 0$ and $Y_G = 0.067$. Our estimate without death appears as the upper curve in Figure 2.8 in §2.6.7.

Figure 2.8 shows plots of two curves for positive death rates δ, obtained using the DTMC model in §2.6 under the same approximation. The plotted cases for $\delta = 0.03$ and $\delta = 0.06$ show the importance of going beyond the no-death model. Consistent with Figure 2.8, we will see that the lockup premium is decreasing in the hedge fund death rate with our Markov chain model. Consequently, formulas (2.18) and (2.21) in this section, derived under the assumption of zero death rate, provide upper bounds on our estimated lockup premium with positive δ, with a simple closed-form formula.

2.6 The Discrete-Time-Markov-Chain Model

We start in §2.6.1 by discussing two important hedge-fund performance measures: persistence and the death rate. Next in §2.6.2 we define the basic three-state DTMC model, which has six parameters. Then in §2.6.3 we introduce four equations that the six parameters must satisfy, based on standard hedge fund performance measures. In §2.6.4 we develop explicit formulas for the three parameters appearing in the DTMC transition probabilities. In §2.6.5 we show how to calculate all the parameters after specifying two of the relative returns. We present numerical examples in §2.6.6. Finally, we show how to calculate the lockup premium in §2.6.7.

2.6.1 Important Hedge-Fund Performance Measures

Our Markov chain model will depend critically on the persistence of returns and the hedge-fund death rate. So we discuss these performance measures further now.

Two persistence factors: γ_G and γ_S. In equations (2.27) and (2.28) below we will introduce two state-dependent persistent factors γ_G and γ_S, instead of just the single γ. Clearly, this generalization is important if the persistence factors for the two states do in fact differ significantly. To illustrate what actually may happen, Figure 2.3 shows the results of a regression analysis applied to two consecutive-year relative returns for positive and negative parts of the current relative-return data separately. Figure 2.3 shows a significant difference in the slope of regression line for several fund categories, suggesting that it may be important to use separate state-dependent persistence factors.

The stationary death rate δ. We calibrate our models by specifying the overall annual death rate, denoted by δ. Unfortunately, estimating the death rate from the TASS database is difficult, in part because poorly performing funds often stop reporting, but funds also stop reporting for other reasons, e.g., because they seek no new investors.

After checking the reasons for funds being terminated in the HFR data, Rouah (2006) concluded that, after removing these biases, 3 to 5% of the hedge funds leave the database each year because of failure. As noted in §2.4, Park (2006) estimated that the fund death rate is only 3.1 %, even though the total attrition rate from the TASS database was 8.7 % , based on her analysis from 1995 to 2004.

The death rate is closely related to the survival probability and median life of the fund. Clearly, as the death rate increases, the survival probability and the median life decrease. Since median life is more easily observable, it is convenient to verify the death rate of our model through the median life in the hedge fund data.

One way to check the validity of the model is to calculate the survival probability curve produced by the model. In terms of the transition matrix P to be introduced in (2.23). the *probability of surviving n years* is $S_n = P^n_{G,G} + P^n_{G,S}$ for $n \geq 1$. Figure 2.4 shows the survival probability curve for the DTMC model when $\delta = 0.03$ and 0.06. When $\delta = 0.03$, about 90% survive for 5 years, whereas the survival probability goes down to around 80% when $\delta = 0.06$. If we increase δ above 0.07, then we are unable to fit the DTMC model.

Studies estimating the median survival time of hedge funds were discussed in §2.4. In addition, Gregoriou (2002) estimated that median survival time of all hedge funds is 5.5 years, depending on factors such as millions managed, performance fee, leverage, minimum purchase and also on the redemption period. More recently, Rouah (2006) reported estimates of median survival time due before liquidation as ranging from 5.8 to 7.4 years based on the HFR data and from 7.2 to 17.4 years based on the TASS database. This last observation by Rouah (2006) suggests that the mean life of a fund across all strategies is approximated reasonably by the DTMC model with $\delta = 0.06$.

2.6.2 The Basic DTMC Model

As indicated in the introduction, we let our Markov chain models have three states: good, sick and dead. We model the changing fund state over time as a DTMC, as in Chapter 4 of Ross (2003). We let time be discrete, with the unit of time representing one year. The initial DTMC is an *absorbing Markov chain*, with the D state being the sole absorbing state; once a fund becomes dead, it remains dead forever. We consider a transition matrix depending on three parameters: p, q and r:

$$P = \begin{matrix} G \\ S \\ D \end{matrix} \begin{pmatrix} p & 1-p & 0 \\ q & r & 1-q-r \\ 0 & 0 & 1 \end{pmatrix} . \tag{2.23}$$

which corresponds to the following diagram: We have assumed that it is impossible to transition from good to dead in a single year, thus eliminating one parameter.

We now move on to consider an associated *ergodic Markov chain*, having a non-degenerate limiting steady-state distribution, by assuming that a new hedge fund appears in the good state to replace a dead hedge fund right after it dies. This can be done with the new three-state DTMC transition matrix

$$P = \begin{matrix} G \\ S \\ D \end{matrix} \begin{pmatrix} p & 1-p & 0 \\ q & r & 1-q-r \\ p & 1-p & 0 \end{pmatrix} . \tag{2.24}$$

In (2.24), the transition probabilities from a dead state are the same as from a good state, because a dead fund is immediately replaced by a good fund.

From the basic theory of DTMC's, as in Theorem 4.1 of Ross (2003), we obtain the steady-state probability vector $\pi \equiv (\pi_G, \pi_S, \pi_D)$ by solving $\pi = \pi P$ under the condition that $\pi_G + \pi_S + \pi_D = 1$. The stationary probability vector π for the transition matrix P in (2.24) is

$$\pi_G = \frac{q + p(1-q-r)}{2-p-r}, \quad \pi_S = \frac{1-p}{2-p-r}, \quad \pi_D = \frac{(1-p)(1-q-r)}{2-p-r}. \tag{2.25}$$

Our DTMC model uses both transition matrices. We use the absorbing transition matrix in (2.23) when we compute the expected return of a fund, while we use the ergodic transition matrix in (2.24) when we calculate the steady-state death rate and performance variance.

We will act as if the fund earns a state-dependent fixed relative rate of return in each state. We must specify these relative rates of return. Let Y_G, Y_S and Y_D denote the relative rate of return in the states G, S and D, respectively. In other words, e^{Y_G}, e^{Y_S} and e^{Y_D} are the return at the end of one year in the states G, S and D, if one dollar is invested in a fund at the beginning of the year. Overall, we have six parameters: p, q, r, Y_G, Y_S and Y_D.

2.6.3 The Four Model-Fitting Equations

We first consider the death rate, which is defined as the proportion of live funds (in a good or sick state) that die during one transition period, which we take to be one year. For the transition matrix in (2.23), only sick funds can die in one transition. Thus, the death rate equals the product of the steady-state probability that a fund is sick times the transition probability from sick to dead. By (2.23) and (2.25), the *death rate* is

$$\delta = \pi_S \cdot P_{S,D} = \frac{1-p}{2-p-r}(1-q-r) = \pi_D \ . \tag{2.26}$$

We now introduce two equations determined by the persistence. For greater model flexibility, we allow different persistence in states G and S. The two *DTMC-persistence* equations are:

$$\gamma_G \cdot Y_G = p \cdot Y_G + (1-p) \cdot Y_S \quad \text{and} \tag{2.27}$$

$$\gamma_S \cdot Y_S = q \cdot Y_G + r \cdot Y_S + (1-q-r) \cdot Y_D \ . \tag{2.28}$$

We explain these DTMC-persistence equations as follows: In equation (2.27), the fund starts with state G; in equation (2.28) the fund starts with state S. The left side describes expected return in the next period calculated using the relevant persistence factor, whereas the right side calculates expected return in the next period using the transition probabilities of the DTMC in (2.23).

Our fourth equation is for the steady-state variance of the annual returns R_i in (2.10). Notice that its variance equals the variance of B_i, defined in (2.6). Since we are working with return rates relative to the mean, the variance of the steady-state rate of return coincides with the second moment. Thus, the *variance* equation is

$$\sigma^2 = \pi_G \cdot Y_G^2 + \pi_S \cdot Y_S^2 + \pi_D \cdot Y_D^2 \ . \tag{2.29}$$

2.6.4 Explicit Formulas for the Transition Probabilities

We now derive formulas for the DTMC transition probability parameters p, q and r in terms of Y_G, Y_S, Y_D, γ_G, γ_S and δ using the three equations (2.26), (2.27) and (2.28).

The three formulas. Assuming that γ_G, γ_S, δ, Y_G, Y_S and Y_D are specified, the three equations in (2.26), (2.27), and (2.28) produce three equations in the three unknowns p, q and r. We first observe that the variable p can be solved from the single equation in (2.27), because that is a single equation for the single unknown variable p. The solution is

$$p = \frac{\gamma_G \cdot Y_G - Y_S}{Y_G - Y_S} \ . \tag{2.30}$$

Having found the explicit expression for p in (2.30), we substitute in for p to obtain two equations in the remaining two unknowns q and r. Indeed, given p, we can rewrite each of the two remaining equations to express q directly as a function of r. First, from (2.26), we get

$$q \equiv q(r) = 1 - r - \frac{\delta(2-p-r)}{1-p} = 1 - \delta\left(\frac{2-p}{1-p}\right) - r\frac{(1-p-\delta)}{(1-p)} \ . \tag{2.31}$$

Since $\delta < 1 - p$ by (2.26), the function $q(r)$ in (2.31) is necessarily strictly decreasing in r.

Next, (2.28) can be rewritten as

$$q \equiv q(r) = \frac{\gamma_S \cdot Y_S - Y_D - r(Y_S - Y_D)}{Y_G - Y_D} = \frac{(\gamma_S - r)Y_S - (1-r)Y_D}{Y_G - Y_D} \ . \tag{2.32}$$

Combining the two equations (2.31) and (2.32), we get an explicit expression for r, first in terms of p and then in terms of the basic model parameters, namely,

$$r = \frac{\left(\frac{1-p-\delta(2-p)}{1-p}\right) - \left(\frac{\gamma_S \cdot Y_S - Y_D}{Y_G - Y_D}\right)}{\left(\frac{1-p-\delta}{1-p}\right) - \left(\frac{Y_S - Y_D}{Y_G - Y_D}\right)} = \frac{\left(\frac{(1-\delta)(1-\gamma_G)Y_G - \delta(Y_G - Y_S)}{(1-\gamma_G)Y_G}\right) - \left(\frac{\gamma_S \cdot Y_S - Y_D}{Y_G - Y_D}\right)}{\left(\frac{(1-\gamma_G)Y_G - \delta(Y_G - Y_S)}{(1-\gamma_G)Y_G}\right) - \left(\frac{Y_S - Y_D}{Y_G - Y_D}\right)} \tag{2.33}$$

To be feasible, we of course need $0 \leq q \leq 1 - r$ and $0 \leq r \leq 1$. Formulas (2.31) and (2.33) simplify when $\delta = 0$; see §A.4 in the Appendix.

By further analysis, we can determine what parameter values can occur; see §A.6 in the Appendix for a detailed analysis. Figure 2.6 shows the three parameters as a function of δ with $Y_G = 0.067$, $Y_S = -0.15$, $Y_D = -0.20$ and $\gamma_G = \gamma_S = 0.5$. From the analysis, it can be shown that there is an upper limit on how high the death rate δ and the persistence γ can be. For the other parameters we consider, the maximal possible death rate is $\delta = 0.07$.

2.6.5 Determining All Model Parameters

We now put everything together to develop an algorithm for computing all the model parameters.

An iterative algorithm. There are several ways we may proceed. We choose to specify Y_S and Y_D in addition to δ, γ_G, γ_S and σ. (This decision is supported by the fact that the model parameters are less sensitive to Y_S and Y_D than to Y_G, as we will see in §2.7.) Specifying these two quantities determines all the parameters. We then calculate the model parameters iteratively. We do so by guessing Y_G, which enables us to directly calculate the DTMC parameters p, q and r, and then the steady-state probability vector π. Given π, we can then calculate σ from (2.29). We then iterate until the calculated σ agrees with the initially specified value of σ.

Although it is not entirely evident from the equations, because π depends on Y_G too, our calculations indicate that σ is an increasing function of Y_G, so it is easy to find the appropriate value of Y_G, e.g., by performing bisection search. A simple plot of σ versus Y_G verifies this property, and reveals the appropriate value of Y_G. We illustrate in Figure 2.7 below for the special case in which $Y_S = -0.15$, $Y_D = -0.20$, $\gamma_G = \gamma_S = 0.5$ and $\delta = 0.03$.

Denominating in terms of σ. For additional insight, it is helpful to express our returns in units of the standard deviation σ. We can divide through by σ^2 in (2.29) to obtain

$$1 = \pi_G \cdot (Y_G/\sigma)^2 + \pi_S \cdot (Y_S/\sigma)^2 + \pi_D \cdot (Y_D/\sigma)^2 . \tag{2.34}$$

Observe that the steady-state probability vector π in (2.25) and the death rate δ in (2.26) depend only on DTMC parameters p, q and r, while the equations (2.30), (2.32) and (2.33) for p, q and r are invariant under scale multiples of Y_G, Y_S and Y_D.

Paralleling Figure above, it is useful to see how Y_G/σ behaves as a function of σ when we fix Y_S/σ and Y_D/σ in addition to δ and γ. It turns out that, after fixing $Y_S/\sigma = -1.5$ and $Y_D/\sigma = -2.0$, the value of Y_G/σ is constant when $\delta = 0$ and almost constant (very weakly increasing) when $\delta > 0$. For the special case in which $Y_S/\sigma = -1.5, Y_D/\sigma = -2.0, \gamma_G = \gamma_S = 0.5$, and $\delta = 0.03$, $Y_G/\sigma \approx 0.685$ for σ ranging from 0.07 to 0.13. It is thus convenient and useful to set Y_S and Y_D proportional to σ. *We hereafter set $Y_S = -1.5\sigma$ and $Y_D = -2.0\sigma$ for our analysis.*

2.6.6 Numerical Examples

We now consider some numerical examples. Our base case is $\delta = 0.03, \gamma_G = \gamma_S = \gamma = 0.5, \sigma = 0.1$, $Y_S = -1.5\sigma = -0.15$, and $Y_D = -2.0\sigma = -0.20$. If we try $Y_G = 0.685\sigma = 0.0685$, then we get $p = 0.8432$, $q = 0.3719$, $r = 0.5030$, and $\sigma = 0.1001$.

Table 2.2 shows parameter values for various δ, γ_G, γ_S, with Y_S, Y_D and σ fixed as above, the return Y_G is calculated iteratively by the method above. The last line of the Table 2.2 shows that r is negative. If $\gamma_G = \gamma_S = 0.5$, our numerical analysis shows that r reaches 0 and becomes negative when δ is above 0.07.

Table 2.2: Parameter value sets

δ	γ_G	γ_S	σ	Y_G	Y_S	Y_D	Calculated σ	p	q	r
0.00	0.5	0.5	0.1	0.067	-0.15	-0.20	0.1002	0.8456	0.3456	0.6544
0.03	0.5	0.5	0.1	0.0685	-0.15	-0.20	0.1001	0.8432	0.3719	0.5030
0.06	0.5	0.5	0.1	0.070	-0.15	-0.20	0.1001	0.8409	0.4207	0.2282
0.07	0.5	0.5	0.1	0.075	-0.15	-0.20	0.1001	0.8401	0.4474	0.0796
0.00	0.6	0.4	0.1	0.076	-0.15	-0.20	0.1000	0.8655	0.3982	0.6018
0.03	0.6	0.4	0.1	0.077	-0.15	-0.20	0.1002	0.8643	0.4320	0.4069
0.06	0.6	0.4	0.1	0.0775	-0.15	-0.20	0.1000	0.8637	0.5068	-0.0127

2.6.7 The Lockup Premium Calculation

To calculate the lockup premium, we use formula (2.13). Without extended lockup, we start with a good fund, so that R_i^1 for $1 \leq i \leq n$ are i.i.d. random variables each with two possible values. Let S_t denote one of the three possible states at year t $(t \geq 1)$: G, S or D. We define S_0 as the state of a fund at the beginning of the first year. As mentioned in §2.1, we assume that $S_0 = G$. Then,

$$\mathbb{E}\left[e^{\sum_{i=1}^{n} R_i^1} | S_0 = G\right] = \left(pe^{Y_G} + (1-p)e^{Y_S}\right)^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{(kY_G + (n-k)Y_S)}. \tag{2.35}$$

The corresponding expectation for the fund with extended lockup is more complicated, but it can be calculated recursively. It is immediate to see that $R_1^2 = R_1^1$, resulting in $p_1 = 0$. We define

$$m(t,s) \equiv \mathbb{E}\left[e^{\sum_{i=1}^{t} R_i^2} | S_0 = G, S_t = s\right] \cdot P_{G,s}^t, \tag{2.36}$$

where P^t, the t^{th} power of matrix P defined in (2.23), represents the probability of reaching S_t from G at t^{th} year. Then, we obtain the following recursion formulas

$$\begin{aligned} m(t,G) &= pe^{Y_G}m(t-1,G) + qe^{Y_G}m(t-1,S), \\ m(t,S) &= (1-p)e^{Y_S}m(t-1,G) + re^{Y_S}m(t-1,S) \quad \text{and} \\ m(t,D) &= (1-q-r)e^{Y_D}m(t-1,S), \end{aligned} \tag{2.37}$$

where $m(1,G) = pe^{Y_G}$, $m(1,S) = (1-p)e^{Y_S}$ and $m(1,D) = 0$. Notice that if a fund becomes dead before year n, it starts with a good state. Furthermore, the new good fund is now under 1-year lockup instead of n-year. Because of this, care must be taken for a sample path once a fund becomes dead. We finally have

$$\mathbb{E}\left[e^{\sum_{i=1}^{n} R_i^2} | S_0 = G\right] = m(n,G) + m(n,S) + \sum_{t=2}^{n} m(t,D)\left(pe^{Y_G} + (1-p)e^{Y_S}\right)^{n-t}. \tag{2.38}$$

For example, if we set $\sigma = 0.1$, $Y_G = 0.685\sigma$, $Y_S = -1.5\sigma$, $Y_D = -2.0\sigma$, $\gamma_G = \gamma_S = \gamma = 0.5$ and $\delta = 0.03$, we get $p = 0.8432$, $q = 0.3719$ and $r = 0.5030$ from §2.6.6. The difference between a 2-year lockup and a 1-year lockup is 0.66 percentage points of return whereas the difference between a 3-year lockup and a 1-year lockup is 1.01 percentage points of return. Figure 2.8 shows the lockup premium calculated from (2.37) and (2.13) as well as the analytical approximation for $\delta = 0$ in (2.14). It is observed that both the one-term and two-term analytical approximations in §2.5 constitute upper bounds for the lockup premium.

2.7 Sensitivity Analysis for the DTMC model

The mathematical models developed here are useful to estimate how the lockup premium depends on the different variables. We describe highlights of such analyses here and present more details in the appendix. Our results here are related to the standard *base case* with $\gamma_G = \gamma_S = \gamma = 0.5$, $\sigma = 0.1$, $Y_S = -1.5\sigma$, $Y_D = -2.0\sigma$ and $\delta = 0.03$, as in the second row of Table 2.2.

Figure 2.9 (i) shows the lockup premium for five values of γ: 0.1, 0.2, 0.3, 0.4 and 0.5 while Figure 2.9 (ii) shows the lockup premium for five values of σ: 0.05, 0.10, 0.15, 0.20 and 0.25. In both cases, these changes produce minor changes in Y_G and the other model parameters; see the Appendix A.7.

We next consider how the DTMC model parameters p, q and r depend on the other driving variables. To supplement Figure 2.6 and the commentary in §2.6.4, Figure 2.10 shows how these parameters p, q and r depend on γ (assuming $\gamma_G = \gamma_S = \gamma$) and each of the return values Y_G, Y_S and Y_D, taken one at a time. We see that the model becomes unstable if γ gets very large, but there is nice near-linear behavior for values of $\gamma \leq 0.5$. We also see that the parameters p, q and r are considerably more sensitive to Y_G than the other two returns Y_S and Y_D.

Lastly, we consider how the DTMC lockup premium for a fixed lockup period depends on three variables δ, γ, and σ. Figure 2.11 shows how the three-year lockup premium depends on two of the three variables while fixing the remaining variable. We see that the three-year lockup premium is reasonably well approximated by a linear function of γ and σ, respectively; there is concavity in γ but convexity in σ. Also, the three-year lockup premium is relatively insensitive to δ.

We remark that the lockup premium in the DTMC model can be approximated by a simple functional form of three variables δ, γ, and σ with the choice of $Y_S/\sigma = -1.5$ and $Y_D/\sigma = -2.0$. The approximation with a simple functional form is helpful to quickly estimate how the lockup premium changes if the variables change.

We had success fitting the simple product form of the three variables with an exponent for each variable for the fixed year lockup premium, denoted by $\psi^p(\delta, \gamma, \sigma) = a\, \delta^b \gamma^c \sigma^d$. After taking logarithms, we can easily apply linear regression for the lockup premium values in the DTMC model to estimate parameters a, b, c and d. By that method, the three-year lockup premium is approximated by $\psi^p_{(3)} = 0.047\ \delta^{-0.11} \gamma^{0.69} \sigma^{0.64}$ with maximum error of 0.0039 in the base case $Y_S/\sigma = -1.5$ and $Y_D/\sigma = -2.0$. The product approximation can be extended to different lockup periods (n) and choice of Y_S/σ and Y_D/σ. See Appendix A.7.4 for further discussion.

2.8 Conclusion

As we explained in §2.1 and §2.3, we have defined the hedge-fund lockup premium as the average difference (per year) between the annual returns from investments in hedge funds, where one has a nominal one-year lockup and the other has an extended n-year lockup. (In doing so, we pointed out that we are not considering the lost opportunity cost of other investments, which may be very important.) We have developed DTMC models to estimate the hedge-fund lockup premium as a function of the length n of the extended lockup period and the model parameters. To account for immediate redemption of investment when a hedge fund fails, we include a death state in the model. The lockup premium represents the cost of not being able to switch from sick funds to good funds while under the lockup condition. We assume that the investor can redeem all his investment if the fund dies, so the effect of the lockup is mitigated by the death rate. That makes the lockup premium more difficult to analyze, justifying the care we give to it.

In §2.6 we showed how the Markov chain model can be fit to basic hedge-fund performance measures, notably, the persistence of relative returns, γ (also allowing different γ_G and γ_S), the standard deviation of returns, σ, and the hedge-fund death rate δ. We then have applied the models to estimate how the lockup premium depends on these important performance measures. The models quantify how the lockup premium increases as a function of the persistence factor γ and the standard deviation σ, but decreases as a function of the death rate δ; this is summarized by (??) for our main numerical example.

As we explained in §2.1, the primary basis for our analysis is the persistence hypothesis: We postulate that there is a persistence in hedge fund performance within a particular hedge fund strategy category. Specifically, a persistence of γ means that *for every* 1 *percentage point you earn above the average in the current year, you expect to earn* γ *percentage points above the average in the next year.*

As reviewed in §2.2 and §2.4, we examined the literature to see what other researchers have concluded about hedge-fund performance persistence and the other hedge-fund performance measures, but we found varying conclusions. Indeed, the literature indicates that persistence in hedge fund returns is highly controversial. We also performed our own statistical analysis using the TASS hedge fund data to estimate these hedge fund performance measures. We found strong evidence of persistence, but the specific persistence values cannot be predicted with great confidence, as is evident from the scatter plots in Figure 2.2. Moreover, the most serious challenge to our analysis is not in the statistical conclusions based on the TASS data, which strongly support persistence, but instead in possible biases in the data, stemming from voluntary reporting. Thus we think that we have been more successful showing how the lockup premium depends on persistence and other the hedge-fund performance measures than in determining the values of persistence and the other performance measures.

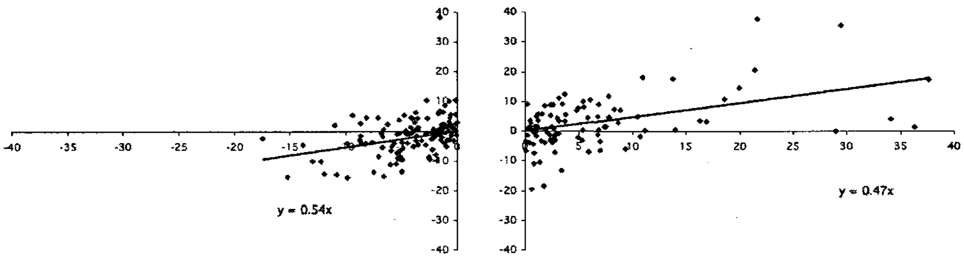
The model fitting requires solving equations. For the DTMC, we were able to give explicit formulas for the three DTMC parameters p, q and r as a function of Y_G, Y_S, Y_D, γ_G and γ_S, but in order to calibrate the standard deviation of returns, σ, we needed to use an iterative method. We developed an efficient algorithm for doing the model fitting.

We conclude that all three performance measures - δ, γ and σ - can have a significant impact on the lockup premium, but we predict that the effect will be negligible if either γ or σ is small. We estimated these key hedge-fund performance measures from the TASS database, but further work needs to be done to obtain more reliable estimates.

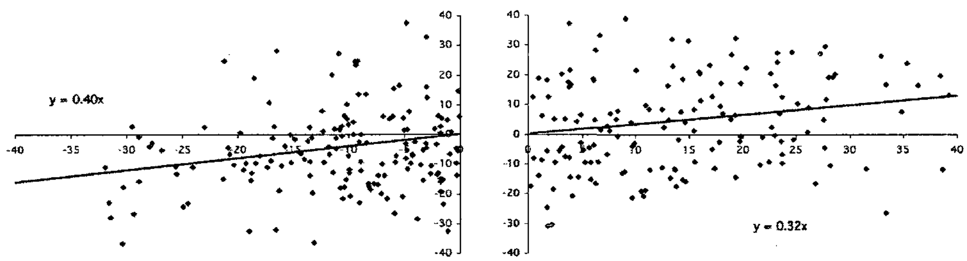
There are a number of directions for further research. One weakness of our DTMC model is that it takes two years for a fund to transition from good to dead. We have developed an analogous three-state continuous-time Markov chain (CTMC) model that does not suffer from that shortcoming. Preliminary analysis indicates that the mathematical analysis is substantially more complicated, but the numerical results are not too different; we hope to report on these results soon.

We have also shown how the persistence we have found in hedge fund relative returns can be exploited to develop a stochastic-difference-equation model for the sequence of relative returns (random variables) themselves in Derman et al. (2009b). The discrete time feature is included because of the infrequent reporting.

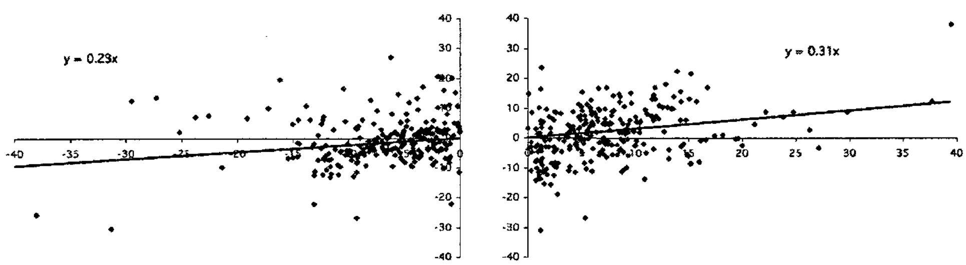
In this paper, we have considered a very specific application for our DTMC model, but it is evident that variants of the same model may be useful in other contexts. With richer data, it may be possible to include more states. Even for the specific hedge-fund liquidity premium problem we consider, one might exploit the approach here in other ways. For example, evidently a minor variation of the same procedure would work if, instead of relative returns, we focused on hedge fund alpha values, as in Jagannathan et al. (2006), Fung et al. (2008) and Kosowski et al. (2007).



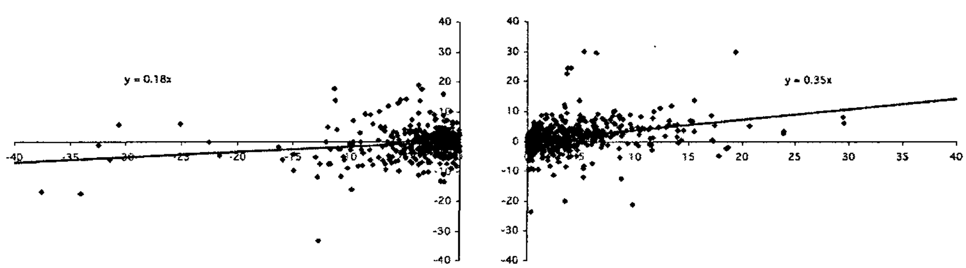
(a) Convertible arbitrage



(b) Emerging market



(c) Event driven



(d) Fund of fund

Figure 2.3: Scatter plots and least-squares lines for positive current relative returns and negative current relative returns of hedge funds from 2000 to 2005 in four categories: (i) convertible arbitrage (ii) emerging market, (iii) event driven, and (iv) fund of fund.

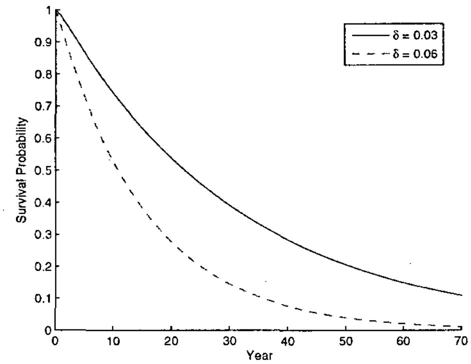


Figure 2.4: The survival probability for the DTMC model when $\delta = 0.03$ and 0.06, for parameter values given in the Table 2.2.

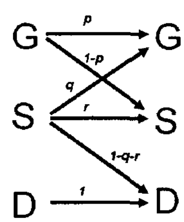


Figure 2.5: Transition probabilities in the absorbing Markov chain

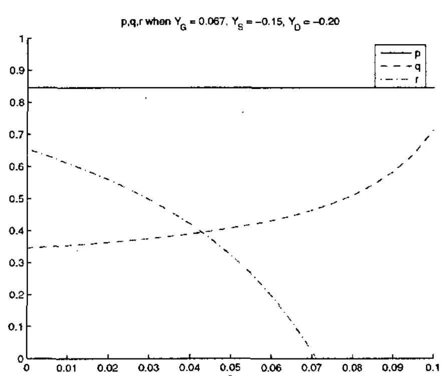


Figure 2.6: The DTMC parameter values p, q and r as a function of δ when $Y_G = 0.067$, $Y_S = -0.15$, $Y_D = -0.20$ and $\gamma_G = \gamma_S = 0.5$

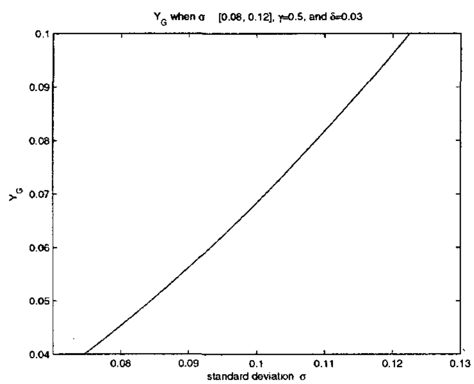
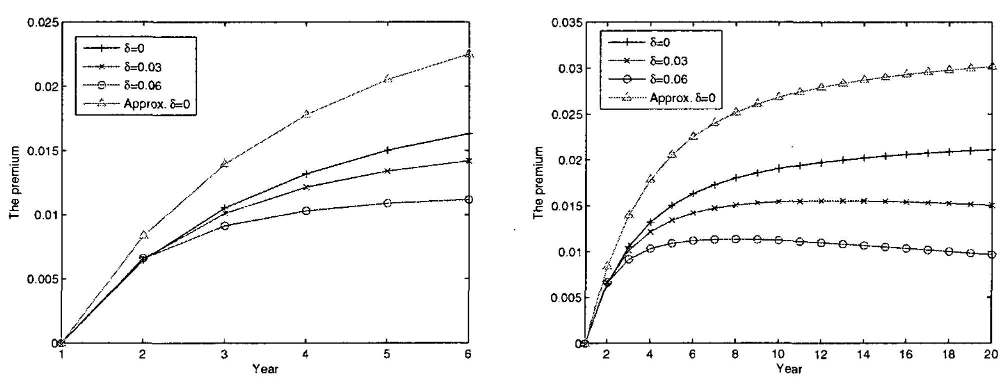


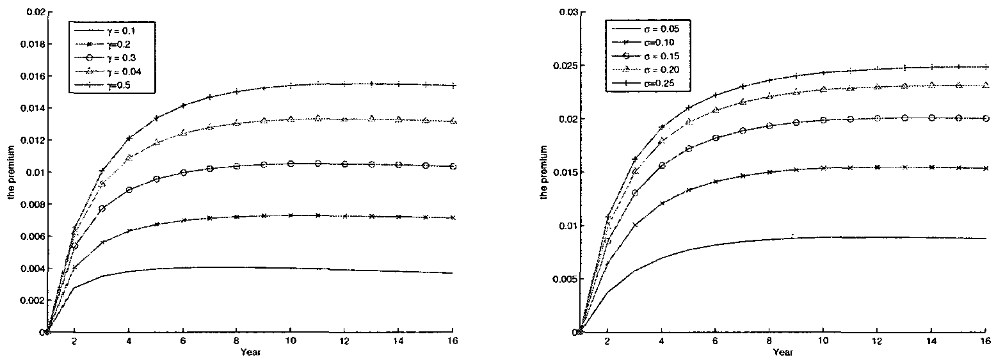
Figure 2.7: The standard deviation of relative return σ versus Y_G when $Y_S = -0.15, Y_D = -0.20$, $\gamma_G = \gamma_S = 0.5$, and $\delta = 0.03$.



(a) From 1 to 6 years

(b) From 1 to 20 years

Figure 2.8: The lockup premium function for DTMC model for three values of the hedge-fund death rate δ and analytic approximation of lockup premium (§2.5) for $\delta = 0$. The remaining model parameters are $Y_G = 0.067, Y_S = -0.15$, and $Y_D = -0.20$ and $\gamma_G = \gamma_S = \gamma = 0.5$.



(a) γ: 0.1, 0.2, 0.3, 0.4 and 0.5 (b) σ: 0.05, 0.10, 0.15, 0.20 and 0.25

Figure 2.9: The lockup premium for the DTMC model in the base case with five values of γ and σ.

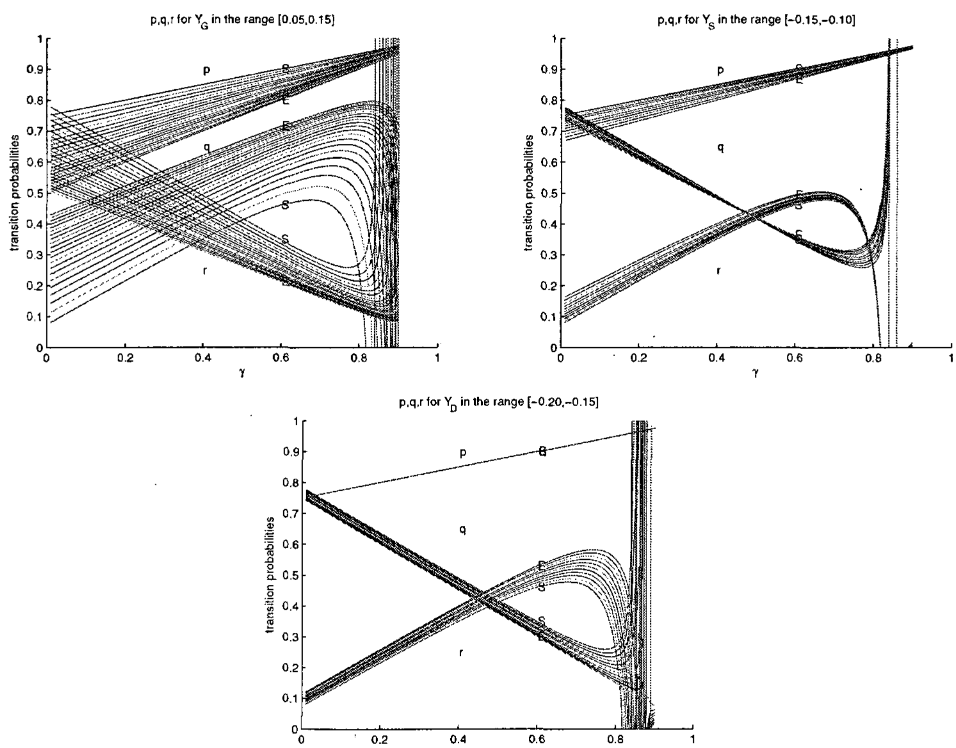
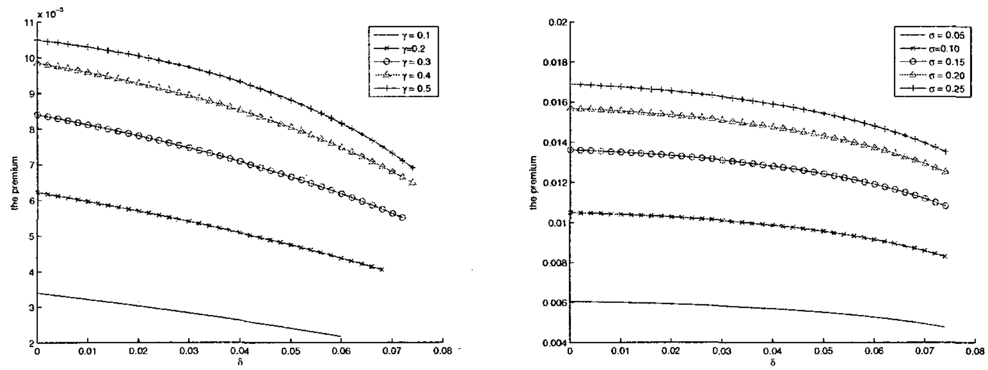
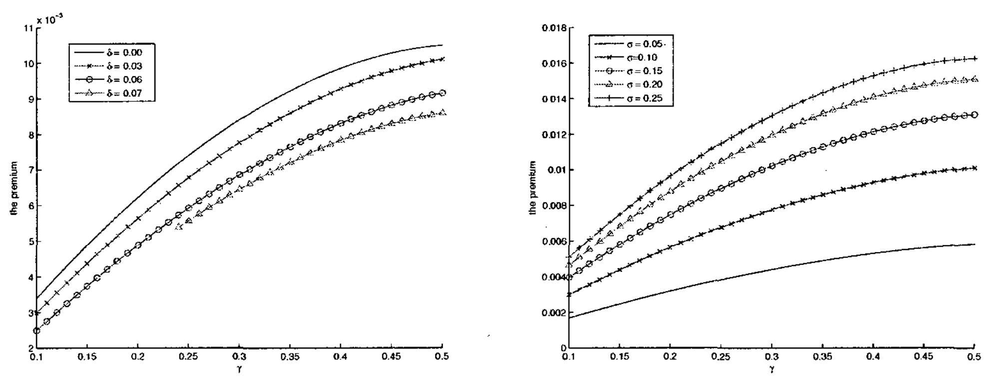


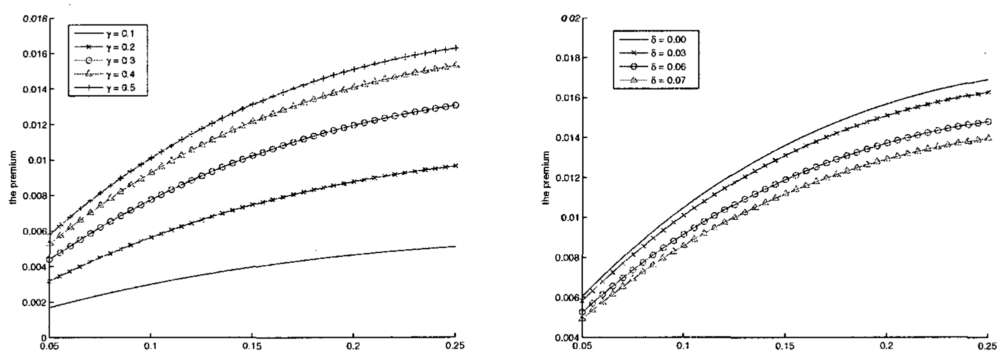
Figure 2.10: The parameters p, q and r as a function of γ in the base case for values of Y_G ranging from 0.05 (starting value, denoted by S) to 0.15 (ending value, denoted by E), Y_S ranging from -0.15 to -0.10, and Y_D from -0.20 to -0.15



For $\delta = 0.00$ to 0.08, (i) $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$ with $\sigma = 0.1$ (ii) $\sigma = 0.05, 0.1, 0.15, 0.2, 0.25$ with $\gamma = 0.5$



For $\gamma = 0.1$ to 0.5, (iii) $\delta = 0.00, 0.03, 0.06, 0.07$ with $\sigma = 0.1$ (iv) $\sigma = 0.05, 0.1, 0.15, 0.2, 0.25$ with $\delta = 0.03$



For $\sigma = 0.05$ to 0.25, (v) $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$ with $\delta = 0.03$ (vi) $\delta = 0.00, 0.03, 0.06, 0.07$ with $\gamma = 0.5$

Figure 2.11: The three-year lockup premium for the DTMC model with $Y_S = -1.5\sigma, Y_D = -2.0\sigma$. The lockup premium does not exist if q or r becomes negative.

Chapter 3

A Stochastic-Difference-Equation Model for Hedge-Fund Relative Returns

3.1 Introduction

Despite the abundance of stochastic models for stocks, commodities and market indices, relatively few stochastic models have been developed for hedge funds. That is not entirely surprising since hedge funds are not too transparent; there are only a few sources of data, with infrequent voluntary reporting. We contribute by developing a stochastic-process model of the relative annual returns of a hedge fund, exploiting data from the Tremont Advisory Shareholders Services (TASS) hedge-fund database for the period 2000-2005.

3.1.1 Relative Annual Returns Within the Fund Strategy

The TASS database archives monthly returns and the managed asset value for each hedge fund. In addition, TASS also archives various fund-specific data, such as the strategy of the fund. The eleven strategies and the sample size for each are given in the first and second columns of Table 3.1; we will explain the rest of Table 3.1 later. (The appendixes of Hasanhodzic and Lo (2007) and Chan et al. (2006) describe the hedge-fund strategies.)

Table 3.1: Estimated persistence γ and auto-correlation ρ for the eleven strategies.

Strategy	Sample Size	γ from regression[1]	γ from ratio of exp. returns[2]	ρ auto-correlation[3]
Convertible	238	0.44±0.10	0.39	0.49+0.09/-0.11
Dedicated Short	29	0.49±0.38	0.44	0.16+0.25/-0.35
Emerging Market	315	0.36±0.10	0.36	0.32+0.09/-0.10
Equity Marcro	268	0.09±0.10	0.12	0.12±0.12
Event Driven	533	0.24±0.08	0.16	0.13±0.08
Fixed Income	193	0.29±0.14	0.38	0.37+0.12/-0.14
Fund of Fund	986	0.33±0.05	0.31	0.31+0.05/-0.06
Global Macro	166	0.13±0.15	0.14	0.06±0.15
Long-short Equity	1658	0.15±0.04	0.11	0.07±0.05
Managed Future	235	0.22±0.13	0.17	0.21+0.12/-0.13
Other	167	0.41±0.15	0.38	0.39+0.12/-0.13

1. 95% confidence interval for the regression coefficient

2. Ratio of expected relative returns from the previous to current year for pairs of two successive years whose return values are both above the average.

3. confidence interval of correlation coefficient from 95% confidence interval of Fisher-Z statistic in (3.24).

In order to highlight differences in hedge fund performance within its strategy and to approach a stationary environment, we focus on the *relative* annual returns. We use geometric compounding to convert the twelve reported monthly returns into one annual return, i.e.,

$$r_{annual} = (1 + r_1) \cdot (1 + r_2) \cdots (1 + r_{12}) - 1 \ .$$

We then obtain the relative annual returns by subtracting the average for the strategy for that year.

We think of the TASS relative return data as being observations from a *stationary* discrete-time stochastic process $\{X_n : n \geq 0\}$, with X_n representing the relative annual return from year n. Assuming that the process $\{X_n\}$ is indeed approximately stationary (which is made more plausible by our focus on relative returns), we combine all the data for each category to estimate the distribution of the single-year relative return for each strategy. For each strategy, we seek a stochastic-process model that matches both the observed single-year relative-return distribution and the observed dependence structure. To have a model useful for prediction, it is desirable that the stochastic process be a Markov process, with a state that is as simple as possible.

Since we focus on relative returns, the relative-return distribution necessarily has mean 0, so a key parameter of the distribution to be matched is the *variance* $\sigma^2 \equiv Var(X_n)$, but we also want to match the entire distribution as much as possible. Indeed, in some cases we find that the return distribution has a heavy tail, consistent with an infinite variance.

For a stationary stochastic process, a key parameter describing the dependence structure is the *autocorrelation* $\rho \equiv Cor(X_n, X_{n+1}) \equiv Cov(X_n, X_{n+1})/\sigma^2$. Estimates of the auto-correlation ρ appear in the final column of Table 3.1. However, we also want to match the full time-dependent behavior of the stochastic process as much as possible. To partially test the time-dependent behavior beyond the auto-correlation ρ, we evaluate the probability that the relative returns will ever hit specified levels within a five-year period. That also illustrates how the model can be applied.

3.1.2 Persistence of Hedge-Fund Returns

Our modelling approach is motivated by our observation of persistence in the relative returns. Broadly, persistence in hedge-fund returns is a tendency for a fund which generates relatively high (or low) returns in a period to continue generating relatively high (or low) returns again in the next period.

Persistence has been studied quite extensively within the hedge-fund literature, but it remains a controversial topic. A consensus has not yet been reached on the degree of persistence in hedge-fund returns. In fact, some studies did not find significant persistence; e.g., Brown et al. (1999), Capocci and Hüber (2004), and Boyson and Cooper (2004). However, several studies have found evidence of strictly positive persistence, depending on the time period measured; Agarwal and Naik (2000) found significant persistence for quarterly returns, while Edwards and Caglayan (2001) found significant persistence over one to two years, and Jagannathan et al. (2006) found significant persistence over three years of returns. For hedge-fund indexes, Amenc et al. (2003) found statistically meaningful persistence for most of the strategies.

In this chapter, we consider persistence in the (relative) returns. (It is important to note that others have looked for persistence in different ways; e.g., Jagannathan et al. (2006) is about alpha persistence.) We say that there is a *persistence factor* of γ if for every 1 percentage point the fund makes above the average in the current year, it is expected to earn γ percentage point above the average in the next year. For the stochastic process $\{X_n : n \geq 0\}$, the persistence implies that we should have the following relation between the conditional expected relative return at the end of the current year, given the previous relative return, and the previous relative return itself:

$$E[X_n|X_{n-1}] = \gamma X_{n-1} \tag{3.1}$$

for all n and all values of X_{n-1}. We estimate the persistence factors by performing a regression analysis. In particular, we combine the relative-return data for all pairs (X_n, X_{n+1}) and perform a standard linear regression. Our estimated persistence factors for the eleven hedge-fund strategies ranged from 0.11 to 0.49; estimates by two different methods appear in the third and fourth columns of Table 3.1. The 95% confidence intervals show that positive persistence is confirmed statistically for all but two strategies; See §3.4 for more on our data-selection and analysis procedure.

In our statistical analysis we do find strong evidence for persistence, but we hasten to admit that the issue remains controversial. The voluntary reporting has led to questions about the reliability of the data. Possible biases in reported hedge-fund returns are discussed by Fung and Hsieh (2000) and Boyson and Cooper (2004). As we explain in §3.4.1, in our data selection procedure, we attempt to reduce the bias, but the TASS data should be regarded as somewhat unreliable. We emphasize that *our primary goal is not to make a case for persistence, but instead is to show how persistence can be exploited, if it is there, in order to create a flexible and tractable stochastic-process model of hedge-fund returns.* Our approach should also have other useful applications, where persistence may exist. We introduce the model in the next section. In subsequent sections, we elaborate on the appealing mathematical structure of the model, we describe our data analysis methods and results, and we show that the model provides a flexible framework for fitting.

3.2 The Proposed Stochastic Difference Equation Model

In order to capture the observed persistence in the performance of hedge-fund relative returns, we first propose the simple *stochastic difference equation* (SDE)

$$X_n = \gamma X_{n-1} + B_n, \quad n \geq 1, \tag{3.2}$$

where γ is a constant with $0 < \gamma < 1$, B_n is independent of X_{n-1} and $\{B_n : n \geq 1\}$ is a sequence of independent and identically distributed (i.i.d.) random variables, each distributed as $N(0, \sigma_b^2)$, where $N(a, b)$ denotes a normally distributed random variable with mean a and variance b.

The SDE in (3.2) is a linear, recursive Markov process; it is also a first-order autoregressive process. Moreover, the SDE in (3.2) is a natural discrete-time analog of the familiar continuous-time *stochastic differential equation*

$$dX(t) = -\nu X(t) + \sigma_c dB(t), \tag{3.3}$$

where $\{B(t) : t \geq 0\}$ is a standard Brownian motion, commonly used in finance, as can be seen by subtracting X_{n-1} from both sides in (3.2) to get

$$X_n - X_{n-1} = -(1-\gamma)X_{n-1} + B_n, \quad n \geq 1. \tag{3.4}$$

We choose the discrete-time process in (3.2) instead of the continuous-time process in (3.3) because hedge-fund returns are reported much less frequently than stock prices.

The initial SDE model in (3.2) is very appealing because, first, it clearly matches the persistence as specified in (3.1) with the same parameter γ and, second, one need to choose only one remaining model parameter σ_b^2 in order to match the steady-state variance σ^2. That is easily done, because for the model (3.2) it turns out that one variance must be a constant multiple of the other:

$$\sigma^2 = \frac{\sigma_b^2}{1-\gamma^2}, \tag{3.5}$$

Moreover, as a consequence of (3.2), the distribution of X_n (assuming stationarity) must itself be normal, distributed as $N(0, \sigma_b^2/(1-\gamma^2))$. Both these conclusions are demonstrated in §3.3.

This is a beautiful simple story when it works. Clearly, it works (from this preliminary checking) if indeed the two variances are related by (3.5) and the steady-state distribution of the relative returns is approximately normal. Fortunately, for some hedge fund strategies, we find that both conditions are satisfied reasonably well. Moreover, we can go beyond the distribution of relative annual returns to check the time-dependent behavior. In §3.3 we show that in steady-state, the SDE in (3.2) necessarily has autocorrelation equal to the persistence:

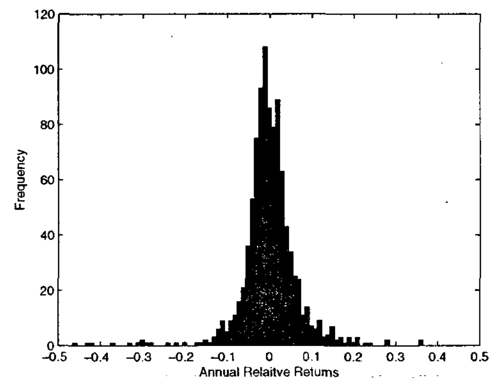
$$\text{autocorrelation} \equiv \rho = \gamma \equiv \text{persistence factor}. \tag{3.6}$$

This special relation in (3.6) turns out to match the TASS data remarkably well, given the limited data, as shown in Table 3.1, which displays estimates of both ρ and γ.

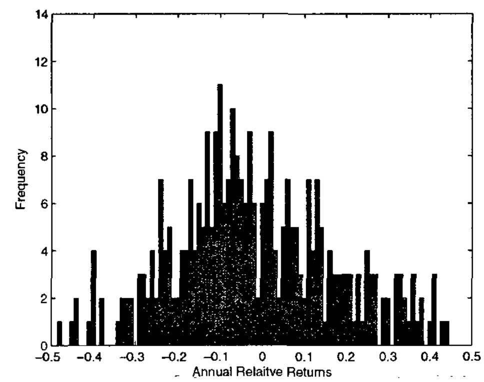
We find that the simple SDE model in (3.2) provides a remarkably good fit for some of the hedge-fund strategies, e.g., for the emerging-market strategy. However, it does not provide a good fit for all strategies; e.g., for the fund-of-fund and event-driven strategies, largely because for those other strategies the empirical distribution of the relative annual returns is quite far from normal, having a heavy tail. Figure 3.1 substantiates this claim, showing the histogram and Q-Q plots of the relative annual returns of hedge funds within the fund-of-fund and emerging-market strategies. (The units are chosen so that a relative annual return of 0.10 corresponds to 10 *percentage points* above average.)

We selected these two strategies for three reasons: (i) because these strategies have relatively large numbers of observations (ii) because they have relatively high persistence factors and (iii) because the return distributions exhibit very different tail behavior. Figure 3.1 shows that the distributions for those two strategies differ significantly. The Q-Q plots in Figure 3.1 (c) and (d) show that the distribution of the relative returns for the emerging-market strategy is close to normal, whereas for the fund-of-fund strategy it is not.

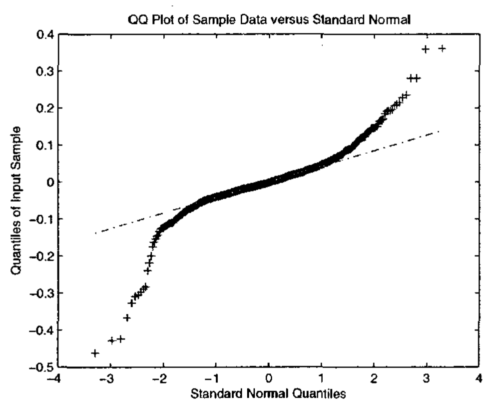
The fund-of-fund strategy is somewhat special, involving investments in other strategies. It might be considered surprising that the relative returns from the fund-of-fund strategy are less normal, since they tend to be more diversified, but correlations among the returns from different strategies may possibly explain this phenomenon. Understanding the observed tail behavior of different strategies remains a problem for future research. We do emphasize that heavy tails are also observed in other strategies, such as the event-driven strategy, as



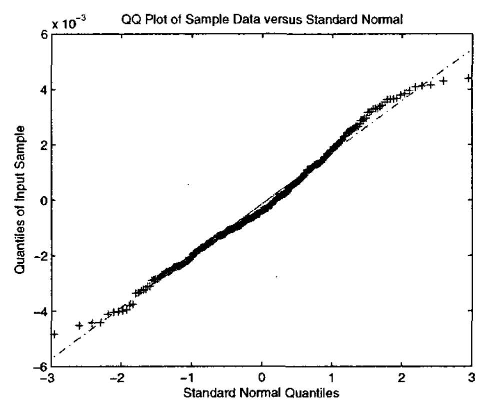
(a) Histogram from fund-of-fund strategy



(b) Histogram from emerging-market strategy



(c) Q-Q plot for fund-of-fund strategy



(d) Q-Q plot for emerging-market strategy

Figure 3.1: (a)(b) Histograms of 986 relative returns within the fund-of-fund strategy and 315 relative returns within the emerging-market strategy from the TASS database. (A relative return of 0.15 means 15 percentage points above the average.) (c)(d) Q-Q plots comparing the model to the normal distribution.

we show in Appendix §B.10. Corresponding figures for other strategies appear in Appendix §B.3.

Just as for performance persistence, the distribution and other statistical properties of hedge-fund returns are not yet well understood, despite the importance (Lhabitant, 2004; Kassberger and Kiesel, 2006; Tran, 2006). Several authors have reported that the normal distribution may not approximate hedge-funds returns well, primarily because of heavy tails

(Lo, 2001; Lhabitant, 2004; Tran, 2006; Geman and Kharoubi, 2003; Eling and Schuhmacher, 2007). It should thus not be surprising that we find that the relative returns are reasonably well approximated by the normal distribution for some strategies, but not for all strategies. Consistent with our analysis, Amo et al. (2007) pointed out that autocorrelation, high-peak, and heavy-tail may be observed from the distributions of hedge-fund returns.

Kassberger and Kiesel (2006) studied the distribution of daily hedge-fund indices within each strategy. Based on the daily indices data from March 2003 to June 2006, they show that the distributions of indices have heavy-tails by Q-Q plots. They claimed that a Normal Inverse Gaussian (NIG) distribution fits the distribution of indices well, since it may have heavy-tail and skewness depending on parameter values.

3.2.1 A More General SDE Model

The non-normal distribution shown in Figure 3.1 (c), and in other return distributions, leads us to look for other models. Fortunately, we find that a natural generalization of the simple SDE in (3.2) provides a robust and tractable model for capturing different behavior observed in the TASS data. As a generalization of the simple SDE in (3.2), we propose the SDE

$$X_n = A_n X_{n-1} + B_n \ , \ n \geq 1 \ , \qquad (3.7)$$

where A_n and B_n are independent of X_{n-1} and $\{A_n : n \geq 1\}$ and $\{B_n : n \geq 1\}$ are independent sequences of i.i.d. random variables with general distributions, satisfying

$$E[A_n] = \gamma \quad \text{for} \quad 0 < \gamma < 1, \quad \text{and} \quad E[B_n] = 0. \qquad (3.8)$$

In going from (3.2) to (3.7), we have replaced the constant persistence factor γ by the random persistence A_n, but the moment conditions in (3.8) imply that the basic persistence relation (3.1) still holds. Moreover, the autocorrelation still satisfies (3.6), as we show in §3.3. By allowing A_n and B_n to have general distributions, we have produced a much more flexible class of models. Fortunately, this class of models is also remarkably tractable, as was shown by Vervaat (1979), where many additional references can be found.

We classify the specific models we consider by the assumptions we make about the distributions of A_n and B_n. When $P(A_n = \gamma) = 1$, we have a *constant-persistence* model;

when A_n has a nondegenerate distribution, we have a *stochastic-persistence* model. When B_n is normally distributed, we have a *normal-noise* model. To capture the heavier tails we see in the data, we also consider as distributions for B_n the Student-t distribution, a mixture of two distributions, an empirical distribution and a stable distribution.

3.2.2 The Constant-Persistence Stable-Noise Model

We highlight the constant-persistence stable-noise model, because it is now common to use stable distributions to represent heavy-tailed distributions, building on early work by Mandelbrot (1963), Fama (1965) and others; see Embrechts et al. (1997), Samorodnitsky and Taqqu (1994) and §4.5 of Whitt (2002) for general background. Indeed, there is now a vast literature on heavy tails in financial data; e.g., see Lux (1996), Rachev and Mittnik (2000), Cont (2001) and Gabaix et al. (2007).

A random variable Y is said to have a (strictly) *stable law* if, for any positive numbers a_1 and a_2, there is a positive number $c \equiv c(a_1, a_2)$ such that

$$a_1 Y_1 + a_2 Y_2 \stackrel{\mathrm{d}}{=} cY, \tag{3.9}$$

where Y_1 and Y_2 are independent copies of Y and $\stackrel{\mathrm{d}}{=}$ means equality in distribution. It turns out that the constant c must be related to the constants a_1 and a_2 by

$$a_1^\alpha + a_2^\alpha = c^\alpha \tag{3.10}$$

for some constant α with $0 < \alpha \leq 2$, called the *index* of the stable law. A random variable Y_α with stable distribution having index α with $0 < \alpha < 2$ satisfies $P(Y_\alpha > x)/x^{-\alpha} \to c_+$ and $P(Y_\alpha < -x)/x^{-\alpha} \to c_-$ as $x \to \infty$ for some positive constants c_+ and c_-. Consequently, $E[|Y_\alpha|^p] < \infty$ for all $p < \alpha$, but $E[|Y_\alpha|^p] = \infty$ for all $p > \alpha$. We will be considering α with $1 < \alpha < 2$, so that our stable distributions will have infinite variance but finite mean, which we take to be zero.

Just as for the normal distribution (which can be regarded as a special stable distribution), the structure of the SDE in (3.2) implies that the stochastic structure of the distribution of B_n is inherited by the distributions of X_n for the constant-persistence models; i.e., the distribution of X_n is again stable with the same index and skewness parameter;

that is, we have

$$X_n \stackrel{\mathrm{d}}{=} \left(\frac{1}{1-\gamma^\alpha}\right)^{1/\alpha} B_n, \tag{3.11}$$

as we prove in §3.3. We use this relation (3.11) in what we think are novel ways: *We use* (3.11) *to test both the constant-persistence stable-noise model and the stable index* α (using the persistence factor γ already estimated); see §3.7.

For the constant-persistence stable-noise model, the SDE in (3.2) also has the continuous-time analog in (3.3), but where now $\{B(t) : t \geq 0\}$ is a non-Gaussian stable Lévy motion, as in Samorodnitsky and Taqqu (1994). More generally, when the random variable B_n has a non-normal distribution, (3.2) has continuous analog (3.3) where $\{B(t) : t \geq 0\}$ is a Lévy process; see Wolfe (1982). In §5 of Wolfe (1982), he shows how to construct the continuous-time analog from the discrete-time SDE if it is desired. By now, there is a substantial literature on non-standard stochastic differential equations in finance; e.g., see Barndorff-Nielsen and Shephard (2001) and Borland (2002).

We will show that the constant-persistence stable-noise model is remarkably effective for the fund-of-fund strategy. Nevertheless, other versions of the model in (3.7) are worth considering as well, in part because they have finite variance, which allows us to use the observed variance σ^2 to calibrate the model.

3.2.3 Previous Models of Hedge-Fund Returns

A conventional assumption is that a firm's net asset value evolves in continuous time as a geometric Brownian motion. Following that convention, a log-normal distribution was used to model hedge fund net asset value by Atlan et al. (2006) and the risky investment the hedge fund holds by Hodder and Jackwerth (2007). However, the log-normal assumption is not empirically tested in those papers.

Others have previously used Markov process models to model hedge-fund returns. Hayes (2006) used discrete-time birth-and-death process to calculate the maximum drawdown in hedge-fund returns, and used the autocorrelation condition to calibrate the model. In Derman et al. (2009a) we used three-state Markov chain models to estimate the premium from extended hedge-fund lockup. We used the same TASS data to calibrate that model.

Several econometric models have been proposed as well. A seminal paper is Amin and

Kat (2003), which sought a trading strategy with cash and a market portfolio such as S&P 500 to *replicate* the distribution of a hedge-fund's returns. If a replicating portfolio can be found, by considering the required initial investment in the replicating portfolio and the hedge-fund management fee, then it may be possible to evaluate whether or not an investment in the hedge fund is justifiable or not. A similar replicating approach is also found in Hasanhodzic and Lo (2007). They tried to replicate hedge-fund returns with six common risk factors such as the S&P 500, US Dollar Indexes, Bond index, etc, by means of linear regression analysis. Chan et al. (2006) is a paper closely related to Hasanhodzic and Lo (2007). However, the purpose of Chan et al. (2006) was somewhat different; they wanted to decompose the risk factors underlying the hedge fund in order to compare the systematic risks of hedge funds to that of other traditional asset classes.

3.2.4 Applications of the Stochastic Model

As usual, a stochastic-process model allows us to go far beyond a direct examination of historical data to ask various "what if" questions. There are many ways to apply the model to answer questions, which cannot easily be answered from the data directly. We might simply want to know the probability distribution of the relative return for a particular hedge fund over the following year, given all available past data. From the past data, we can observe the most recent relative return, say $X_0 = c$. We would then apply the model in (3.7) to conclude that the relative return next year should be distributed as $A_1 c + B_1$, where A_1 and B_1 are the independent stochastic persistence and noise, respectively, for that hedge-fund strategy, whose distributions can be determined by data fitting, as described in this chapter. We could go further and calculate the discounted present value of the return stream over many years; see (3.22) - (3.23).

We might want to invest in that particular hedge fund because we believe that it will be especially well managed. We could use the model to provide a "measurement-based" quantification of what we mean by good management. In particular, we may postulate that a good fund manager improves the fund performance in one or more of three possible ways: increasing the expected persistence $\gamma \equiv E[A_n]$, reducing the standard deviation of the persistence $\sigma_a \equiv \sqrt{Var(A_n)}$, or reducing the standard deviation of the additive noise

$\sigma_b \equiv \sqrt{Var(B_n)}$. With the model, we can quantify the impact of such effects. We first fit the model to the data for that hedge-fund strategy in order to obtain random variables A_n and B_n. We then produce new random variables A_n' and B_n' consistent with the postulated consequences of good management. We then calculate future relative returns, both with the original model and with the revised model. In that way, we can estimate the value added by the good management.

We illustrate with a concrete example: Suppose that the relative returns for a specific fund in the last year are $X_0 = c$. We start by quantifying what it mean for a "good" manager to be effective. Suppose that we conclude that the impact of superior management should increase its nominal estimated expected persistence from γ to 1.5γ, reduce the estimated standard deviation of the persistence from σ_a to $0.8\sigma_a$, and reduce the estimated standard deviation of the noise from from σ_b to $0.5\sigma_b$. As a numerical example, we choose the beta-persistence t-noise model developed in §3.6.2 for the fund-of-fund strategy (which has parameter values $\gamma = 0.33, \sigma_a = 0.0381, \sigma_b = 0.0642$, and $\alpha = 50, \beta = 101.52$). We then choose new random variable A_n' and B_n' with $\gamma' = 1.5\gamma, \sigma_a' = 0.8\sigma_a, \sigma_b' = 0.5\sigma_b$ and define X_n' based on the new parameter values. Then, algebraic manipulation yields $\alpha' = 84.75$ and $\beta' = 86.46$. It is then immediate that $\mathbb{E}[X_1'|X_0' = c] - \mathbb{E}[X_1|X_0 = c] = (\gamma' - \gamma)c = 0.1650c$, $Var(X_1|X_0 = c) - Var(X_1'|X_0' = c) = c^2(0.36\sigma_a^2) + 0.75\sigma_b^2 = 0.0005c^2 + 0.0031$. We have thus shown how the model can be applied to quantify the impact of good management.

3.3 Background on the General SDE

The behavior of the general SDE in (3.7) is well described in Vervaat (1979); we will be stating implications from the general results there. We will be considering the standard (good) case in which the expectation $E[\log(A_n)]$ is well defined (at least one of the positive part or the negative part has finite expectation) and the following (minimal) logarithmic-moment conditions are satisfied:

$$-\infty \leq E[\log(A_n)] < 0 \quad \text{and} \quad E[\log^+(B_n)|] < \infty \; , \tag{3.12}$$

where $\log^+(x) \equiv \max\{0, \log(x)\}$. Note that $\log(A_n) = -\infty$ occurs if $A_n = 0$, which is a possibility we want to allow. That corresponds to no persistence at all.

Under condition (3.12), Vervaat shows that we have convergence in distribution $X_n \Rightarrow X_\infty$ as $n \to \infty$, where the distribution of X_∞ is independent of the initial conditions and is characterized as the unique solution to the *stochastic fixed-point equation*

$$X_\infty \stackrel{\mathrm{d}}{=} A_n X_\infty + B_n, \tag{3.13}$$

where the random vector (A_n, B_n) is independent of X_∞ on the right. There is thus a unique stationary version of the process $\{X_n : n \geq 0\}$, obtained by letting the initial value X_0 be distributed as X_∞, while being independent of A_1 and B_1. With our notion of persistence in mind, it is natural to go beyond condition (3.12) and assume in addition that $P(0 \leq A_n < 1) = 1$. That will immediately imply extra moment conditions we make for A_n below. But that extra assumption is actually not required.

Moreover, we actually do not need to assume that A_n is independent of B_n, as we have done, but the strong results in Vervaat (1979) do require that the sequence $\{(A_n, B_n)\}$ be a sequence of i.i.d. random vectors. It is worth noting, though, that the general model in (3.7) has been further generalized beyond Vervaat (1979). First, Brandt (1986) established results for the case in which independence for the sequence $\{(A_n, B_n) : n \geq 1\}$ is dropped; he assumes only that it is a stationary sequence. Next Horst (2001) considers the time-dependent version, allowing the distribution of (A_n, B_n) to depend on n. Finally, Horst (2003) embeds the model in a game-theoretic setting, letting the values of (A_n, B_n) depend on the strategic decisions of multiple players. These extensions are significantly less tractable than (3.7) here, but they open the way to interesting new applications.

Given (3.12), we can also characterize the distribution of X_∞ via an *infinite-series representation*

$$X_\infty \stackrel{\mathrm{d}}{=} \sum_{k=1}^{\infty} A_1 A_2 \cdot A_{k-1} B_k, \tag{3.14}$$

where the series on the right converges with probability 1 (w.p.1). It is thus easy to approximately generate samples from the distribution of X_∞ by considering a truncated version of the series. If $|A_n|$ tends to be relatively small, as with our persistence estimates, then relatively few terms are required.

Moreover, it is easy to apply the stochastic fixed-point equation (3.13) in order to deduce that the steady-state value X_∞ is distributed simply as a constant multiple of B_n, as given in (3.11), when B_n has a stable law. We have the following elementary proposition:

Proposition 3.3.1. *For the simple SDE in* (3.2)*, if B_n has a stable law with index α, i.e., if* (3.9) *and* (3.10) *hold for $0 < \alpha \leq 2$ (with $\alpha = 2$ being the case of a normal distribution), then*

$$X_\infty \stackrel{\mathrm{d}}{=} \left(\frac{1}{1-\gamma^\alpha}\right)^{1/\alpha} B_n; \tag{3.15}$$

i.e., (3.11) *is valid.*

Proof. First, since we are considering the simple SDE in (3.2), we have $A_n \equiv \gamma$. Since the distribution of X_∞ is the unique solution to the stochastic fixed-point equation (3.13), it suffices to show that $X_\infty \equiv cB_n$ satisfies equation (3.13) for some constant c, i.e., it suffices to show that

$$cB \stackrel{\mathrm{d}}{=} \gamma(cB) + B_n, \tag{3.16}$$

where B and B_n are independent random variables with the common distribution of B_n. Since B_n has a stable law with index α, we can apply (3.10) to get the equation $c^\alpha = (\gamma c)^\alpha + 1^\alpha$, which has the desired value for c as its unique solution. □

Important moment properties of the SDE in (3.7) are given in §5 of Vervaat (1979), but these require extra conditions on the moments of the model elements. Prior to the moment conditions made in (3.8), in addition to the conditions above, we assume the technical regularity conditions

$$E[|A_n|] < 1, \quad E[|B_n|] < \infty \quad \text{and} \quad E[|X_0|] < \infty. \tag{3.17}$$

Under these conditions, it follows that $E[|X_\infty|] < \infty$ and $E[|X_n|] < \infty$ for all n. By 5.2.1 of Vervaat (1979), if (3.17) holds, then in general

$$E[X_\infty] = \frac{E[B_n]}{1 - E[A_n]} \quad \text{and} \quad E[X_n] \to E[X_\infty] \quad \text{as} \quad n \to \infty. \tag{3.18}$$

Since we assume condition (3.8) in addition to conditions (3.12) and (3.17), we can conclude that $E[X_\infty] = 0$ and $E[X_n|] \to 0$ as $n \to \infty$.

We will not want to go beyond these first-moment conditions for B_n in (3.17) when we consider stable noise, because then B_n will have infinite variance. However, for the finite-variance case, we also assume that $E[A_n^2] < 1$, and $E[B_n^2] < \infty$ and $E[X_0^2] < \infty$. Then

5.2.2 of Vervaat (1979) provides the following important expression for the variance of the steady-state distribution:

$$\sigma^2 \equiv Var(X_\infty) = \frac{E[B_n^2]}{1 - E[A_n^2]} = \frac{Var(B_n)}{1 - E[A_n^2]} \equiv \frac{\sigma_b^2}{1 - \sigma_a^2 - \gamma^2}, \tag{3.19}$$

where we have introduced the new notation $\sigma_a^2 \equiv Var(A_n)$ and used the assumption that $E[A_n] = \gamma$ in the final expression. Paralleling (3.18), it also implies the convergence $Var(X_n) \to Var(X_\infty)$ as $n \to \infty$. When $P(A_n = \gamma) = 1$, then (3.19) reduces to (3.5).

We now exploit the variance limit above under the the moment conditions in order to characterize the auto-correlation of the stationary version of the stochastic process $\{X_n\}$. We will characterize the asymptotic behavior, with a non-stationary initial condition. For that purpose, assume that $E[X_0] = 0$ along with the moment conditions, so that we have $E[X_n] = 0$ for all n. Then the time-dependent auto-covariance is simply

$$Cov(X_{n+1}, X_n) = E[X_{n+1}X_n] = \gamma E[X_n^2] = \gamma Var(X_n), \tag{3.20}$$

which implies that the associated auto-correlations satisfy

$$\rho_n \equiv Cor(X_{n+1}, X_n) = \frac{Cov(X_{n+1}, X_n)}{\sqrt{Var(X_{n+1})Var(X_n)}} = \gamma\sqrt{\frac{Var(X_n)}{Var(X_{n+1})}} \to \gamma \quad \text{as} \quad n \to \infty. \tag{3.21}$$

We have thus shown for the general SDE model in (3.7) that $\rho = \gamma$, where $\rho \equiv \rho_\infty$ is the auto-correlation for the stationary version of $\{X_n\}$, obtained by letting X_0 be distributed as X_∞, just as claimed in (3.6) for the simple SDE in (3.2).

In our hedge-fund context it is natural to be interested in the discounted present value of a return stream. It is thus convenient that the discounting can be incorporated into our current framework. First, if we postulate a constant rate of interest r compounded continuously, so that the annual discounting factor is e^{-r}, then the (random) present value of the entire relative-return stream and its conditional expected value are

$$V(r) = \sum_{n=1}^{\infty} e^{-nr} X_n \quad \text{and} \quad E[V(r)|X_0] = \frac{X_0}{1 - \gamma e^{-r}}. \tag{3.22}$$

More generally, we may have random annual interest rate R_n in year n, so that the present value is

$$V = \sum_{n=1}^{\infty} \left(\prod_{k=1}^{n} R_k\right) X_n \ . \tag{3.23}$$

Given our model with specified distributions for A_n and B_n, a well-defined stochastic process $\{R_n : n \geq 1\}$, which could be (but need not be) a sequence of i.i.d. random variables with specified distribution, and the initial value X_0, we can easily determine the distribution of V by simulation. We can first generate a segment of the process $\{X_n\}$ recursively, and then do the same for the sum in (3.23). Given typical discounting processes $\{R_n\}$, the series will converge quickly, so that truncated versions will yield good approximations.

3.4 Empirical Observations from the TASS Data

3.4.1 Hedge-Fund Data Selection and Analysis

We first explain how we try to remove biases in the TASS data. We then describe the regression procedure to estimate the persistence factor.

TASS differentiates between the date the fund starts reporting and the date the fund starts operating. When a fund starts reporting returns after operating for several months or years, the fund may simultaneously report several monthly returns at the time its first return is reported. It is then possible for the fund manager to report only good returns. Otherwise, if the returns are bad, the manager may choose not to report them. This phenomenon creates the so-called *backfill bias*, since the backfilled returns tend to be higher due to the missing bad returns. Fung and Hsieh (2000) calculate that the difference from actual returns and reported returns is about 3.6% per year from this reason. In order to at least partially address this problem, we consider monthly returns only after the fund's first reporting date. Similarly, if a fund's monthly returns are reported less than six times a year, we exclude these data, due to the possibility of hiding bad returns.

We also considered the asset value managed by a fund. We treat all funds equally, without regard to the asset value, so we present a "fund view" as opposed to a "dollar view." However, we did start by removing very small funds from our sample. Specifically, we consider monthly returns only if the fund's asset value managed has reached 25 million dollars at least once, at which point we assume that the fund becomes mature, so that it can produce relatively stable returns. A similar data selection strategy was used by Boyson and Cooper (2004). To better understand this issue, we computed the average asset value managed for each fund and plotted the distribution of the values; it is shown in the Appendix §B.2. As might be expected, the distribution of the sizes has a heavy tail.

After selecting the monthly returns based on the above criteria, we proceeded to estimate the persistence factor by regression. In particular, we made pairs of two successive annual returns for each hedge fund from 2000 to 2005. Thus, there are possibly five pairs of annual returns of a fund, if it does not cease reporting during that period. (Thus, our sample sizes in Table 3.1 are the number of pairs in the strategy.) The monthly returns are annualized to measure the yearly persistence of returns, using geometric compounding. We next calculate relative annual returns for each fund by subtracting the average annual returns of the funds in the same strategy. The relative returns for two successive years are then coupled as a pair to estimate yearly persistence factor. In order to make meaningful pairs of relative returns for two successive years, the averages of annual returns for the first year and each strategy of the funds are calculated first. When calculating the average annual returns and the associated relative returns for the next consecutive year, we only include returns from the funds which existed and were not dropped from the TASS database during the previous year. Thus, the average annual return for any given year depends on whether that year is treated as an initial year or a next year. They are not necessarily equal, since some funds may start reporting to TASS in the next consecutive year. In this way, we finally construct pairs of two consecutive relative returns from 2000 to 2005 for each strategy of the fund.

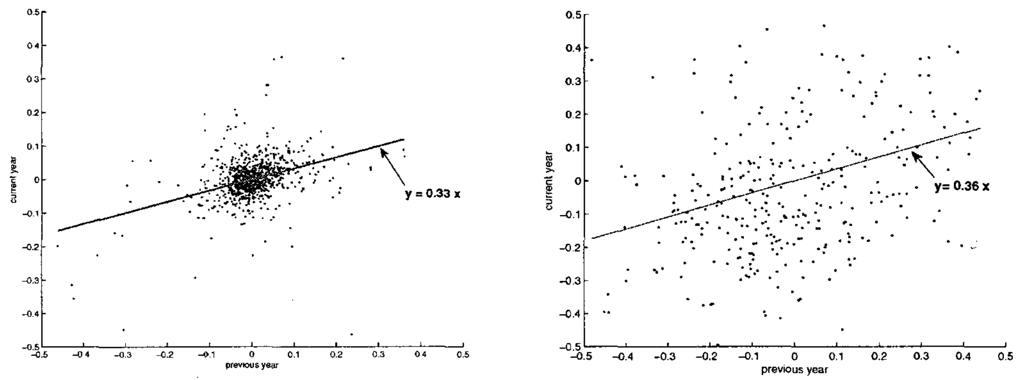
Before conducting regression, we also exclude pairs of returns with extreme values, depending on the distribution of the pairs of returns for each strategy category. Even one or two outliers can seriously affect the regression, especially if we do not have a large number of observations. Specifically, we excluded pairs of relative returns when one absolute relative return exceeds $\pm$ 30 % for the fixed-income and equity-macro and $\pm$ 40% for the convertible, dedicated-short-bias, and global-macro strategies. We also excludes pairs of relative returns exceeding $\pm$ 50% for the emerging-market, event-driven, fund-of-fund, long/short-equity, managed-future, and others strategies. (These percentages were chosen to be appropriate by visual inspection. The percentages are roughly equivalent relative to the overall standard deviation of the return distribution for the strategy.) On the positive side, this data-selection procedure helps us avoid data errors. On the other hand, this data-selection procedure might lead us to underestimate heavy tails. As a consequence, our heavy-tail findings should be even more convincing.

We conducted a linear auto-regression analysis with pairs of two successive years of annual relative returns. The coefficient from this linear regression, i.e., the least square fit is the calculated persistence. The regression analysis results in very low intercept for all strategy category. Thus, we finally conduct a auto-regression without intercept and consider only the coefficient term. The results are shown in the third column of Table 3.1.

An alternative way to estimate the persistence factor is to consider the ratio of the next-year average returns to the current-year average return, restricting attention to the returns that are positive in the current year. The fourth column of Table 3.1 shows the ratio of two successive average returns restricting attention to the returns that are positive and negative in the current year, respectively. We observe that these alternative persistence estimates tend to be similar to the regression estimates.

3.4.2 Persistence of Relative Returns

We started by constructing scatter plots of the relative returns for each hedge-fund strategy, using all pairs (X_n, X_{n+1}), and performed auto-regression analysis in that setting in order to estimate the persistence factor, which thus becomes the the regression coefficient. Figure 3.2 shows the scatter plots of the relative annual returns for the fund-of-fund and emerging-market strategies. A linear relationship is not overwhelmingly clear in Figure 3.2. Nevertheless, we do observe more pairs of returns in the lower left and higher right sides of the scatter plot, indicating the existence of persistence. We mention that the persistence factor may also be derived in another way. We can also estimate the persistence factor from the ratio of the two successive years' expected relative returns, when those values are both above the average. This directly measures the ratio of current year's expected relative returns to the previous year's expected relative returns, but we have yet to develop the statistical properties of this estimator. The estimated persistence factors by both these methods are given in Table 3.1.



(a) Fund-of-fund strategy (b) Emerging-market strategy

Figure 3.2: Scatter plots and auto-regression lines for relative returns from two successive years within (a) the fund-of-fund strategy and (b) the emerging-market strategy.

3.4.3 Distribution of Relative Returns

We now turn to the distribution of the relative annual returns. As illustrated by Figure 3.1, we constructed histograms showing the empirical distribution and constructed Q-Q plots to test for normality. As we have indicated before, the emerging-market strategy relative-return distribution seems to be approximately normal, but the fund-of-fund relative-return distribution does not. The distributions and Q-Q plots for the other strategies are given in §B.3 of the Appendix. The Q-Q plots there show that the relative-return distribution for the global-macro strategy also is well approximated by the normal distributions, but all others have significant departures from normality in the tails. We also performed the Lilliefors test in Appendix §B.3, from which we conclude, statistically, that the relative returns from most of the strategies are not consistent with the normal distribution. (See Lilliefors (1967) for the details of the test.) In order to facilitate visual comparison with the normal distribution, we also plotted histograms from a simulation of i.i.d. normal random variable with the same sample sizes; see, Appendix §B.4. Finally, we note that the fund-of-fund relative-return distribution has a relatively high peak in the center.

3.4.4 Autocorrelation of Relative Returns

In §3.3 we showed that the auto-correlation is equal to the persistence for the general SDE model in (3.7). Thus we want to see if that is true for the TASS data. To examine this issue, we estimate the auto-correlations in the data, using the sample correlation coefficient estimator, denoted by r. In order to estimate the 95% confidence intervals for the auto-correlation correlation, we use the well-known result that the Fisher Z statistic, defined by

$$Z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) , \qquad (3.24)$$

is approximately normally distributed with mean zero and standard deviation $\sigma_z = 1/\sqrt{n-3}$, where n is the sample size; e.g., see Serfling (1980) or Lin (1989).

From (3.24), we derive the confidence interval of the correlation coefficient ρ from the confidence interval of Z. The confidence interval is not symmetric around the observed sample autocorrelation coefficient r because r is a non-symmetric function of Z in (3.24). The last column in Table 3.1 summarizes the results. Table 3.1 shows that the two 95% confidence intervals – for the persistence γ and the auto-correlation ρ – overlap significantly for most strategies. Thus we conclude that γ and ρ coincide with each other and regard this as support for the validity of the SDE model in (3.7). Figure 3.3 adds by providing a graphical comparison of these confidence intervals.

3.5 Testing the Constant-Persistence Normal-Noise Model

We now describe how we evaluated the fit of the constant-persistence normal-noise model. This model has only two parameters γ and $\sigma_b \equiv SD(B_n) \equiv \sqrt{Var(B_n)}$, so the fit to the observed persistence γ and standard deviation $\sigma \equiv SD(X_n)$ is immediate. If we use only those two parameters, we obtain a perfect fit by applying (3.5) and letting $\sigma_b^2 = (1-\gamma^2)\sigma^2$. Such a fit seems to provide a reasonable rough model in all cases.

In this section we want to evaluate the quality of that fit more closely. One test is the auto-correlation; the predicted relation between the autocorrelation and persistence in (3.6) holds more generally, and was just discussed above; Table 3.1 shows that the fit is

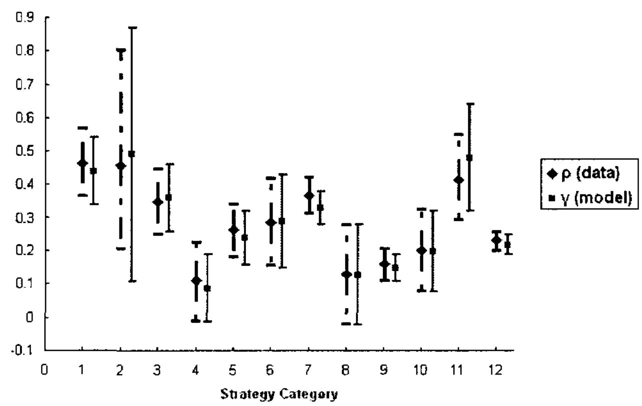


Figure 3.3: A comparison of estimates of the auto-correlation ρ and the persistence γ, showing the 95% confidence intervals for both. As before, the horizontal axis represents the strategy: 1. Convertible 2. Dedicated Short 3. Emerging Market 4. Equity Marcro 5. Event Driven 6. Fixed Income 7. Fund of Fund 8. Global Macro 9. Long-short Equity 10. Managed Future 11. Other 12. All

pretty good, given the limited data. There are two principal remaining issues: (i) Is the relative-return distribution approximately normal? and (ii) Are the standard deviations (or variances) actually related by (3.5)? We have already addressed the first question in §3.4.3, finding that the return distribution is approximately normal in some cases, but not all. Now we turn to the one remaining question.

As indicated before, assuming stationarity, we combine all the relative-return data to estimate the one-year relative-return distribution. The standard deviation of that distribution is denoted by σ; it is estimated directly by the sample standard deviation once the data have been combined.

Testing is possible because we can also directly observe the values of the noise variables B_n. We estimate $\sigma_b \equiv SD(B_n) \equiv \sqrt{Var(B_n)}$ by acting as if the model is valid, implying that $B_n \equiv X_{n+1} - \gamma X_n$ would be i.i.d random variables, using the previously estimated value of the persistence γ. We thus estimate σ_b directly by the sample standard deviation as well, but we are here assuming the model to get the i.i.d. structure and we are using our estimate of the persistence γ. From (3.5), the constant-persistence normal-noise model (and other finite-variance-noise models) predict that $\sigma/\sigma_b = \sqrt{1/(1-\gamma^2)}$. Since we have already estimated γ from the data, we can compare σ/σ_b and $\sqrt{1/(1-\gamma^2)}$ in order to test the validity of the model.

Table 3.2 shows the results. From the last two columns in the table, we observe that

Table 3.2: Estimation of the standard deviations σ and σ_b to test the constant-persistence model.

Strategy	σ	σ_b	ratio	ratio
	return[1]	noise[2]	data[3]	model[4]
1. Convertible	0.0686+0.0068/-0.0056	0.0579+0.0057/-0.0048	1.18	1.11+0.07/-0.05
2. Dedicated Short	0.1393+0.0480/-0.0284	0.1353+0.0466/-0.0275	1.03	1.15+0.88/-0.14
3. Emerging Market	0.1903+0.0161/-0.0138	0.1797+0.0152/-0.0130	1.06	1.07+0.05/-0.04
4. Equity Macro	0.0801+0.0074/-0.0062	0.0655+0.0061/-0.0051	1.22	1.00+0.01/-0.00
5. Event Driven	0.1007+0.0064/-0.0057	0.0884+0.0056/-0.0050	1.14	1.03+0.03/-0.02
6. Fixed Income	0.0693+0.0077/-0.0063	0.0661+0.0073/-0.0060	1.05	1.04+0.06/-0.03
7. Fund of Fund	0.0681+0.0031/-0.0029	0.0565+0.0026/-0.0024	1.21	1.06+0.02/-0.02
8. Global Macro	0.1070+0.0129/-0.0104	0.1027+0.0124/-0.0100	1.04	1.01+0.03/-0.01
9. Long-short Equity	0.1520+0.0054/-0.0050	0.1376+0.0048/-0.0045	1.10	1.01+0.01/-0.01
10. Managed Future	0.1265+0.0126/-0.0105	0.1214+0.0121/-0.0101	1.04	1.02+0.03/-0.02
11. Other	0.1003+0.0120/-0.0097	0.0976+0.0117/-0.0094	1.03	1.14+0.16/-0.08

1. σ: Standard deviation and 95 % confidence interval of the relative annual return
2. σ_b: Standard deviation and 95 % confidence interval of $B_n \equiv X_n - \gamma X_{n-1}$.
3. Ratio: σ/σ_b observed from the data.
4. Ratio: $\sqrt{1/(1-\gamma^2)}$, ratio σ/σ_b from the constant-persistence normal-noise model; see (3.5). 95% confidence interval of the ratio is obtained from 95% confidence interval of γ in Table 3.1.

σ/σ_b and $\sqrt{1/(1-\gamma^2)}$ are quite close for some fund strategies, but not for others. In particular, we see a good match for the emerging-market, fixed-income, global-macro, and managed-future strategies, but we see a poor match, in various degrees, for the others; the worst being the equity-macro and fund-of-fund strategies.

Where the match is good, we need to also test the normal-distribution property, which we have done, and discussed in §3.4.3. Where the match is poor, we see right away that we need to consider a different model, which is what much of the rest of this chapter is about.

3.6 Stochastic-Persistence Models

In this section, we consider the stochastic-persistence models with various stochastic noise distributions as an alternative to the constant-persistence normal-noise model. Our analysis here illustrates the great model flexibility for fitting to data. Our goal in this section is to remedy both deficiencies found in the constant-persistence normal-noise model for some strategies: With the extra flexibility, we obtain a perfect match of the variance σ^2, remedying the problems observed in the last two columns of Table 3.2, and in addition seek a good match in the overall distribution.

3.6.1 Beta Persistence

In order to achieve this new flexibility in a controlled way, we assume that A_n has a *beta distribution*, which is a probability distribution that concentrates on the open unit interval $(0, 1)$. The beta distribution has two parameters, α and β, with mean $\alpha/(\alpha+\beta)$ and variance $\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$. We can choose α and β to match the mean $E[A_n]$ and the variance $Var(A_n)$, provided that the variance is not too large. We remark that the beta distribution arises naturally in Bayesian frameworks when focusing on an unknown parameter lying in a fixed interval; e.g., see Browne and Whitt (1996). However, other persistence distributions can be used in essentially the same way.

By introducing beta persistence, we have thus increased the parameters associated with the persistence from only one (γ) in the deterministic case to two with this beta distribution. We can fit the beta parameters α and β to the mean and variance by

$$\gamma = \frac{\alpha/\beta}{1+\alpha/\beta} \quad \text{and} \quad c_a^2 \equiv \frac{\sigma_a^2}{\gamma^2} = \frac{\beta}{\alpha(\alpha+\beta+1)} = \frac{1}{\frac{\alpha}{\beta}\left(\frac{\alpha}{\beta}+1+\frac{1}{\beta}\right)} . \tag{3.25}$$

From (3.25), we see that the mean γ depends on α and β only through their ratio, while

c_a^2, the squared coefficient of variation (SCV, variance divided by the square of the mean), is strictly increasing in both α and β for any given ratio α/β.

The full beta-persistence stochastic-noise model has three basic parameters: σ_b^2, γ and σ_a^2, but we only directly observe γ and σ^2. We have used γ to specify the mean $E[A_n]$. We thus have only σ^2 to use in order to determine the two model variances σ_a^2 and σ_b^2. Hence, there is one extra degree of freedom.

We apply the variance formula (3.19) to determine a relation that all these variances must satisfy. Formula (3.19) implies that we must have

$$0 \le \sigma_b^2 \le (1-\gamma^2)\sigma^2 \quad \text{and} \quad 0 \le \sigma_a^2 \le 1-\gamma^2. \tag{3.26}$$

Given both σ^2 and σ_b^2, formula (3.19) gives a formula for σ_a^2. In summary, there is a one-parameter family of variance pairs (σ_a^2, σ_b^2) consistent with our data.

We can draw some initial conclusions. First, if $\sigma_a^2 = 0$, so that $A_n = \gamma$ w.p.1, then we can estimate σ_b^2 directly by looking at $X_n - \gamma X_{n-1}$, as we already did. By formula (3.5) or (3.19), we then should have $\sigma_b^2 = (1-\gamma^2)\sigma^2$, but that is inconsistent with the results in Table 3.2. Hence we conclude that we do need to have stochastic persistence; i.e., we should consider some non-degenerate beta distribution for A_n.

One way to proceed at this point is to exploit what we have done in the previous section, and assume that we have already fit the variance σ_b^2 by acting as if the persistence A_n were constant. In other words, we let σ_b^2 be the estimated variance of $X_n - \gamma X_{n-1}$, using our estimate of the persistence γ.

Given that we start with an estimate of σ^2 and have already estimated γ and σ_b^2 by the methods already described, we can choose the variance $\sigma_a^2 \equiv Var(A_n)$ to satisfy (3.19). For the fund-of-fund return data, we have $\gamma = 0.33$ from Table 3.1, while $\sigma = 0.0681$ and $\sigma_b = 0.0565$ from Table 3.2, so that our estimated beta parameters are, first, $\sigma_a^2 = 0.2028$ and then $\alpha = 0.03$ and $\beta = 0.06$. However, the result is not plausible, because these small values of α and β produce a strongly U-shaped density for A_n; see Appendix §B.5.

We deduce that we should consider larger values of α and β, and thus smaller values for the variance σ_a^2 and larger values for σ_b^2. For given α, β is determined to match γ. From visual inspection, we estimate that $\alpha = 50$ should be reasonable; see Appendix §B.7.

Once we have chosen α, that determines β and thus σ_a^2, which in turn determines σ_b^2 by (3.19). For $\alpha = 50$, we get $\beta = 101.51$, $\sigma_a^2 = 0.0014$ and $\sigma_b = 0.9369\sigma = 0.0642$. Having calibrated the model parameter values, we then approximate the random variable X_∞ by taking a truncated version of the infinite series in (3.14). In our context, where we always have $\gamma < 1/2$, fewer than 10 terms suffices. We use only 5 for the fund-of-fund data with $\gamma = 0.33$. That yields the approximation

$$X_\infty \approx B_1 + A_1B_2 + A_1A_2B_3 + A_1A_2A_3B_4 + A_1A_2A_3A_4B_5 \; . \tag{3.27}$$

We get one realization from X_∞ by generating four independent copies of A_n and five independent copies of B_n.

3.6.2 The Beta-Persistence Normal-Noise and t-Noise Models

So far, by this rather involved process, we have specified only the variance of the noise $\sigma_b^2 \equiv Var(B_n)$. A simple specific noise distribution with that variance is the normal distribution that we have been considering; we get it by simply assuming that $B_n \stackrel{\mathrm{d}}{=} N(0, \sigma_b^2)$. For that special noise distribution, the single parameter σ_b^2 fully specifies the noise distribution. We call this the *beta-persistence normal-noise* model. However, when we apply this procedure and apply simulation to estimate the relative-return distribution, we see that the return distribution remains too close to the normal distribution. That remains the case for a wide range of α values; See, Appendix §B.5. Thus we rule out the beta-persistence normal-noise model. Our analysis leads us to conclude that this beta-noise feature, by itself, does not address the heavy tails seen in the data for the fund-of-fund strategy.

In order to capture the heavy tails in the observed relative-return distribution, we consider non-normal noise distributions. In doing so, we build on our previous analysis. As before, we aim to match the estimated values of γ and σ. We exploit the beta persistence we have already constructed, with $\alpha = 50$, $\sigma_a^2 = 0.0133$ and $\sigma_b = 0.0638$.

As a new candidate noise distribution, we propose the (Student)-t distribution, which is known to have a heavier tail than the normal distribution. Specifically, we assume that $B_n \stackrel{\mathrm{d}}{=} \kappa T(\nu)$ where $T(\nu)$ denotes a random variable with the standard t-distribution having parameter ν, which is commonly referred to as the degrees of freedom, and κ is a constant scale factor. Since we keep the beta persistence, we call the overall model the *beta-persistence t-noise model.*

For $\nu > 2$, the variance of a t-distributed random variable T is $\nu/(\nu-2)$. Since $E[B_n] = 0$, we can match the given variance via

$$\sigma_b^2 \equiv Var(B_n) = E[B_n^2] = E[\kappa^2 T^2] = \frac{\kappa^2 \nu}{\nu - 2} \; . \tag{3.28}$$

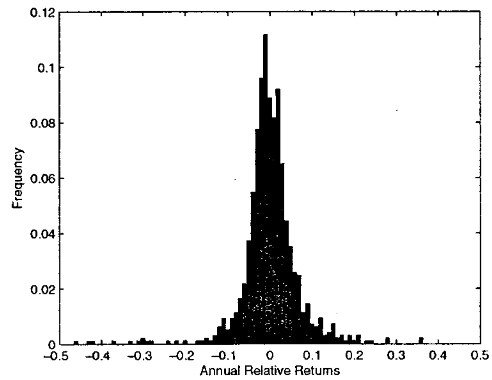
We first use ν as a parameter to choose in order to select the desired shape of the distribution of X_n, consistent with a fixed first two moments of B_n (mean 0 and variance σ_b^2). We then use κ to match the observed variance. Thus, for any given ν, κ is determined by $\kappa = \sigma_b \sqrt{(\nu-2)/\nu}$.

Figure 3.4 shows the simulated distribution of the relative return X_n from the beta-persistence t-noise model with $A_n \stackrel{\mathrm{d}}{=} Beta(50, 101.51)$ and $B_n \stackrel{\mathrm{d}}{=} 0.0278 \cdot T(2.4)$ compared to the observed relative-return distribution for the fund-of-fund strategy. Comparing Figures 3.1 and 3.4, we see that the beta-persistence t-noise model approximates the observed relative-return distribution much better than the constant-persistence normal-noise model does. The two-sample Kolmogorov-Smirnov test also statistically shows that we cannot reject the hypothesis that the simulated data and empirical data come from the same distribution, with p value of 0.3080 (The high p value indicates that we cannot reject the hypothesis that the two random variables are drawn from the same distribution; e.g., see Massey (1951).)

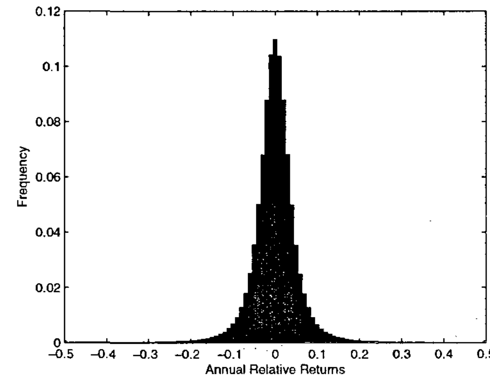
However, looking closely at Figure 3.4, we see that the observed relative-return distribution still has heavier tails than predicted by the model, especially in the left tail. That conclusion is confirmed by the Q-Q plot in Figure 3.4(c).

3.6.3 The Beta-Persistence Empirical-Noise Model

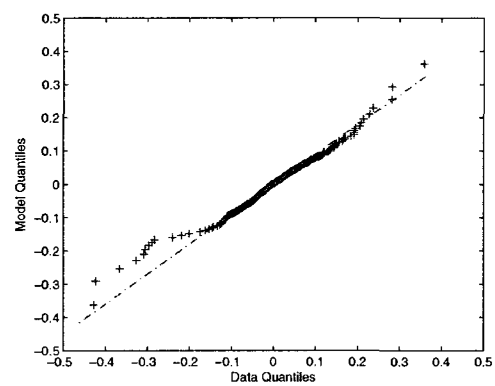
A relatively simple way to obtain a better fit to the data within the beta-persistence class of models is to let B_n have the observed empirical distribution for $X_n - \gamma X_{n-1}$, using the estimated value of γ. This automatically gives B_n and its estimated variance σ_b^2. It now goes further to directly match the shape. This procedure works quite well, as we show in Appendix §B.7. Overall, the approach works well if we are content to use the model for



(a) Histogram of relative returns from the data



(b) Beta-persistence t-noise model



(c) Q-Q plot comparing the model to the data

Figure 3.4: (a) The relative-return distribution from the data within the fund-of-fund strategy (986 observations). (b) Simulation estimate of the relative-return distribution (sample size 10^6) using the beta-persistence t-noise model, with the sample size of 10^6, with $\alpha = 50$, $\beta = 101.51$, $\nu = 2.4$, $k = 0.0278$, $\gamma = 0.33$ and $\sigma = 0.0681$. (c) Q-Q plot of the beta-persistence t-noise model to the data.

simulation. However, we might want a parametric model, with not too many parameters, so we consider further refinements.

3.6.4 The Beta-Persistence Mixed-Noise Model

Since the beta-persistence t-noise model did not adequately capture the heavy left tail of the observed relative-return distribution for the fund-of-fund strategy, we continue to search for a better parametric model. In order to better match this feature, we consider a mixture of two distributions for our noise distribution. We do this both to illustrate the flexibility of our general modelling framework and to obtain a better fit.

Again building upon our previous fitting, we let the distribution of B_n be a mixture of an exceptional normal distribution with some small probability p and the t distribution with probability $1-p$. We start with the beta stochastic persistence in order to calibrate the two variances σ^2 and σ_b^2, and then we introduce the t-noise distribution in order to capture the main shape of the return distribution. In addition, we now add a small normal component to capture the heavy left tail. We call this overall construction our *beta-persistence mixed-noise model.*

The noise random variable B_n in this model can be defined explicitly by

$$B_n = \begin{cases} Z_1 \stackrel{\mathrm{d}}{=} \mu_1 + \kappa T(\nu) \text{ with probability } 1-p \\ Z_2 \stackrel{\mathrm{d}}{=} N(\mu_2, \sigma_2^2) \text{ with probability } p\ . \end{cases} \tag{3.29}$$

Here it is understood that Z_1 represents the regular returns, while Z_2 represents the exceptional low returns. We intend to make the probability p small.

From (3.29), we have six parameters to fit: $p, \kappa, \nu, \mu_1, \mu_2$ and σ_2. We start by controlling the overall shape. That is done by choosing the t parameter ν, in the method just described. We then calibrate p by counting the number of relative returns less than -2σ. Then it remains to fit the four remaining parameters κ, μ_1, μ_2 and σ_2. But now we can write down expressions for the mean and variance of B_n:

$$\begin{aligned} \mathbb{E}[B_n] &= (1-p)\mu_1 + p\mu_2 = 0, \\ \sigma_b^2 &= \mathbb{E}[B_n^2] \ = \ (1-p)\left(\kappa^2 \frac{\nu}{\nu-2} + \mu_1^2\right) + p(\mu_2^2 + \sigma_2^2)\ . \end{aligned} \tag{3.30}$$

Since we have two equations in four parameter values, we have two degrees of freedom. Thus, we fit μ_2 and σ_2 directly from the data. We directly fit the mean and standard deviation of the relative returns counted for estimating p. In this way, we can fit p, μ_2 and

σ_2 at the same time. Then, from (3.30), we can obtain explicit representations for μ_1 and κ, namely,

$$\mu_1 = -p\mu_2/(1-p) \quad \text{and} \quad \kappa = \sqrt{[(\nu-2)/\nu(1-p)]\left(\sigma_b^2 - p(\mu_2^2+\sigma_2^2) - (1-p)\mu_1^2\right)} \ . \quad (3.31)$$

For the fund-of-fund relative returns, out of 986 data points in our sample, we find 18 relative returns below $-2\sigma = -0.1363$. Thus our estimate for p is $p = 18/986 = 0.0183$. As indicated above, in this step we also select the mean and standard deviation of this "exceptional distribution." We find that the mean and standard deviation of those 18 returns are $\mu_2 = -0.2746$ and $\sigma_2 = 0.0717$. Finally, we fit the remaining parameters, getting $\mu_1 = -0.0051$ and $\kappa = 0.0232$. Again, after calibrating parameters for X_n, we use (3.27) to generate realizations of the modelled stationary return X_∞.

Figure 3.5 (a) and (b) shows the simulated return distribution for this beta-persistence mixed-noise model. We now do see a heavier left tail in the model, just like that in the data, but unfortunately now the left tail of the return distribution generated by the model now is heavier than the left tail of the observed distribution from the data. This actually should not be surprising because our model exaggerates the probability of a return below -2σ, including the t-variable as well as the exceptional normal component.

In order to reduce the gap between the model and the data in the left tail, we consider a new parameter fitting procedure that reduces p while keeping μ_2 and σ_2 as specified. The new procedure starts from the given parameter values $p, \mu_2, \sigma_2, \mu_1, \kappa$ and the simulation obtained from the fitting procedure stated above. We first calculate the probability of relative returns falling below the threshold in the model, denoted by f. Since $\mu_2 \ll -2\sigma$ and $\sigma_2 \ll (-2\sigma + \mu_2)$, we ignore the probability of exceptional random variables exceeding the threshold. Let t be the probability that t-distribution falls below the threshold (which we do not evaluate directly). From the definition of t and the observed f, we obtain $p + (1-p)t \approx f$, which yields $t \approx (f-p)/(1-p)$. To obtain a corrected model, we replace f by p and p by p', and have $p' + (1-p')t \approx p$ for $t \approx (f-p)/(1-p)$. Combining these two equations, we get the following expression for p' (which is to replace p):

$$p' = \frac{2p - p^2 - f}{1-f}. \quad (3.32)$$

Our revised model is (3.29) with p replaced with p' in (3.32). We assume that μ_2 and σ_2

remain unchanged. We thus need to calculate new values of μ_1' and κ' via (3.31), using p' instead of p.

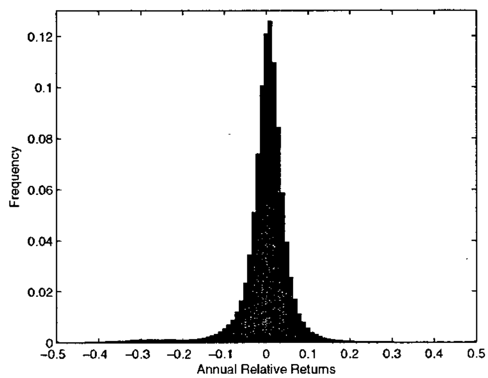
Then, we perform simulation once more with new parameters. Since the first simulation has $f = 0.0284$, we obtain $p' = 0.0081, \mu_1' = 0.0022$ and $\kappa' = 0.0236$ from the new procedure. We found that this procedure significantly improves the fitting. As shown in Figure 3.5, the left tail from the new procedure matches the data much better than before.

3.7 The Constant-Persistence Stable-Noise Model

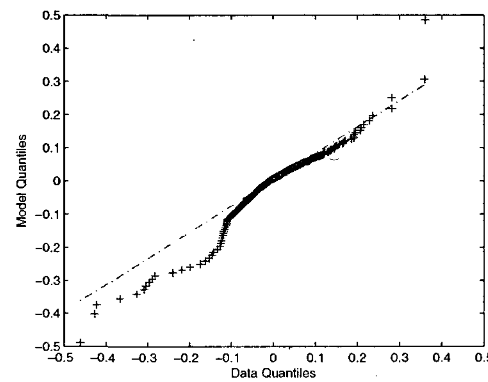
The procedures in §3.6 introduced more and more complexity in order to obtain a better and better fit. A more parsimonious alternative is to directly address the heavy-tail property at the outset by using a stable distribution. In doing so, we have to abandon the information provided by the variance σ^2 and the other variances, because the stable distribution has infinite variance. We thus lose a convenient model parameter when we take this step.

However, we gain simplicity, because we can use the constant-persistence model and avoid any representation of the distribution of A_n. Moreover, the stable distribution has the advantage of providing additional tractability. In particular, with constant persistence, stable noise provides the nice relation between the distribution of X_n and the distribution of B_n given in (3.11) and Proposition 3.3.1. That relation says that X_n will be distributed the same as a constant multiple of B_n.

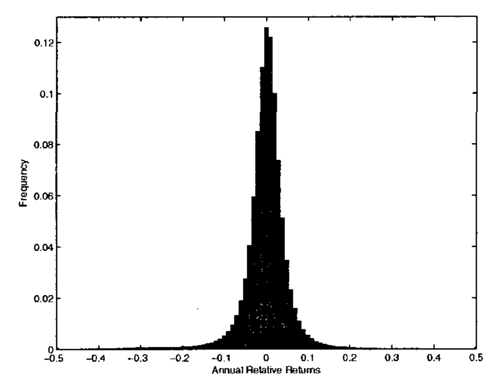
Indeed, Proposition 3.3.1 provides an ideal way to test whether the constant-persistence stable-noise might be appropriate. A simple test is to plot the distributions of X_n and B_n and see if they look similar. As noted before, we obtain B_n directly from $X_n - \gamma X_{n-1}$, using the previous estimate for the persistence γ. Figures 3.4 (a) and 3.6 (a) show the empirical distributions of X_n (stationary version) and B_n obtained from the fund-of-fund data. Clearly, these distributions look remarkably similar, although the Q-Q plot in Figure 3.6 (b) shows some discrepancy in the tails. Moreover, the relationship is further substantiated by Table 3.3, where the ratio of the quantile differences of these distributions are calculated at different levels. These quantile ratios constitute estimates of the proportionality constant c. These quantile ratios are consistently around 1.2, with some discrepancy again in the



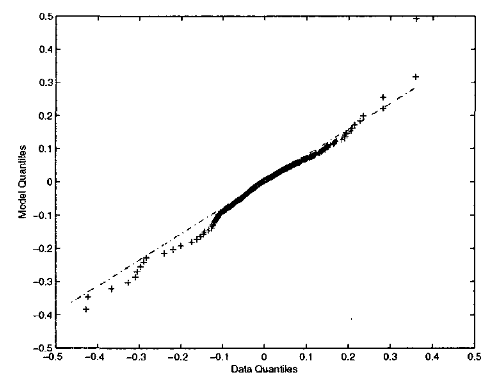
(a) Beta-persistence mixed-noise model



(b) Q-Q plot of the model to the data



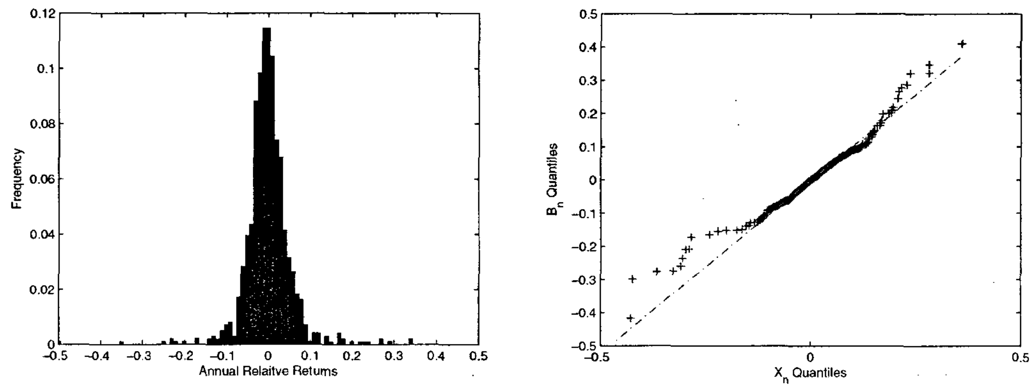
(c) Beta-persistence mixed-noise model re-calibrated



(d) Q-Q plot of the model to the data

Figure 3.5: (a) Simulation estimate of the relative-return distribution (sample size 10^6) using the beta-persistence mixed-noise model with $\gamma = 0.33, \alpha = 50, \nu = 2.4, \sigma = 0.0681, \mu_2 = -0.2746, \sigma_2 = 0.0717, p = 0.0186, \mu_1 = -0.0051$ and $\kappa = 0.0232$. (to be compared to Figure 3.4 (a)). (b) Q-Q plot comparing the model to the data. (c) (d) Simulation estimate of the relative-return distribution and Q-Q plot for the same model re-calibrated with $p' = 0.0098, \mu_1' = 0.0027, \kappa' = 0.0237$ in (3.32).

tails. Thus, Figure 3.6 and Table 3.3 suggest that $X_n \stackrel{d}{=} cB_n$ approximately, where c is a constant whose value is about 1.2. We also performed the two sample Kolmogorov-Smirnov test to compare the distributions, and obtained a p value of 0.5196, which provides further support.



(a) $B_n \equiv X_n - \gamma X_{n-1}$ from the data

(b) Q-Q plot comparing X_n to cB_n

Figure 3.6: (a) Distribution of $B_n \equiv X_n - \gamma X_{n-1}$ for the fund-of-fund relative returns, to be compared to Figure 3.4 (a), and (b) Q-Q plot comparing the distributions of X_n and cB_n with $c = 1.2$.

Table 3.3: The Quantile Differences of X_n and B_n and Their Ratios

Quantile Difference[1]	X_n	B_n	Ratio [2]
55% – 45%	0.0111	0.0085	1.3096
60% – 40%	0.0210	0.0170	1.2321
65% – 35%	0.0327	0.0265	1.2342
70% – 30%	0.0425	0.0364	1.1683
75% – 25%	0.0566	0.0492	1.1506
80% – 20%	0.0709	0.0609	1.1633
85% – 15%	0.0907	0.0778	1.1656
90% – 10%	0.1211	0.1053	1.1509
95% – 5%	0.1887	0.1430	1.3194

1. Difference between two quantile values.

2. Ratio: Quantile Difference for X /Quantile Difference for B.

Recall from our discussion in §3.1 that the index α of a stable law coincides with its tail-decay parameter (of the form $Cx^{-\alpha}$ for some constant C). The conventional elementary way to investigate power tails and estimate the index α is to directly construct a log-log plot of the tails of the distributions. Figure 3.7 shows the log-log plots of the two distribution tails for the fund-of-fund relative-return data. (Figure 3.7 also shows corresponding plots for a model, to be discussed below.) We observe that the left tail of the return distribution is approximated quite well by the linear slope of -1.6, which implies that there is approximately a power tail and that $\alpha \approx 1.6$. As we have observed before, the heavy-tail behavior is more evident in the left tail than in the right tail. The two sample Kolmogorov-Smirnov test result also shows high p value (0.1446), which statistically shows that the two samples could be drawn from the same distribution. (In Appendix §B.6 we provide log-log plots of the tails of the simulated distributions from the other models for contrast.)

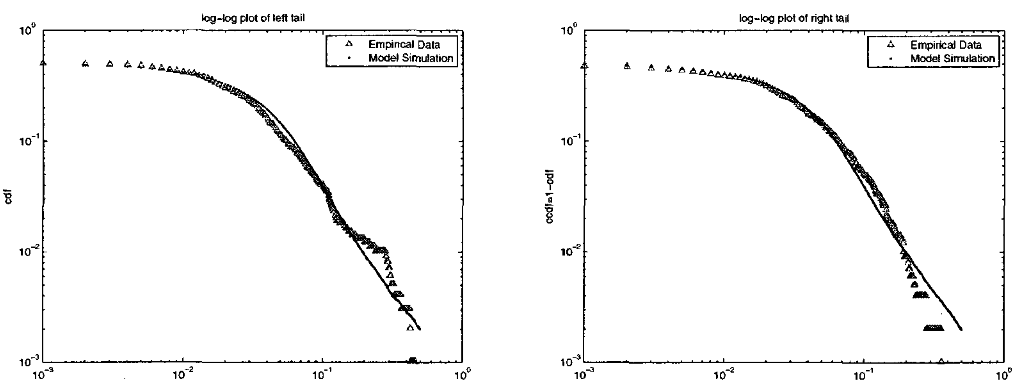


Figure 3.7: Log-log plots of the left and right tails of the fund-of-fund relative-return distribution, from the TASS data and the constant-persistence stable-noise model with parameters $\gamma = 0.33$, $\alpha = 1.6, \beta = 0, k = 0.029$.

We now combine the last two observations to develop a test for the constant-persistence stable-noise model. On the one hand, we have directly estimated the stable index α from the log-log plots of the distribution tails (getting $\alpha \approx 1.6$), but on the other hand, for the constant-persistence stable-noise model, the observed quantile ratio $c \approx 1.2$ also provides an estimate of the index α. That is true because, given the quantile ratio c and the persistence

γ, we can solve for α in the equation

$$c^{\alpha} = \frac{1}{1-\gamma^{\alpha}}, \tag{3.33}$$

obtained from (3.11). We see that the observed value $c = 1.20$ is consistent with the other parameter values: $\gamma \approx 0.33$ and $\alpha \approx 1.6$. Thus the constant-persistence stable-noise model passes this test.

Non-Gaussian stable laws actually have four parameters, and are commonly referred to by $S_{\alpha}(\kappa, \beta, \mu)$; see Samorodnitsky and Taqqu (1994). (We use κ instead of the conventional σ to avoid confusion with the standard deviation considered previously.) As before, α is the index, which ranges in $0 < \alpha < 2$. The other three parameters are: the scale κ, the skewness β and the location parameter μ. When the stable law has a finite mean, μ is that mean. Since we are considering stable laws with finite mean, where that mean is zero, we always have $\mu = 0$. For $\alpha > 1$ and $\mu = 0$, we have the scaling relation

$$cS_{\alpha}(\kappa, \beta, 0) \stackrel{\mathrm{d}}{=} S_{\alpha}(c\kappa, \beta, 0) \quad \text{for all} \quad c > 0 \tag{3.34}$$

for all model parameters. Choosing the scale parameter κ is like choosing the measuring units. In addition to the index, the shape is determined by the skewness parameter β which ranges in $-1 \leq \beta \leq 1$. From (3.34), we see that the scale has no effect on the index or the skewness.

Given the index α, we also have available the two parameters κ and β. As α increases, the shape of the distribution is more centered. As β increases, the distribution is skewed more to the left. Thus we formulate the constant-persistence stable-noise model by letting $B_n \stackrel{\mathrm{d}}{=} \kappa \cdot S_{\alpha}(1, \beta, 0)$. Using Proposition 3.3.1 and the scaling relation (3.34) for the constant-persistence stable-noise model (3.2), we have

$$X_{\infty} \stackrel{\mathrm{d}}{=} \left(\frac{1}{1-\gamma^{\alpha}}\right)^{1/\alpha} \kappa \cdot S_{\alpha}(1, \beta, 0) \, . \tag{3.35}$$

We emphasize that this characterization of the limiting distribution in the constant-persistence stable-noise model simplifies further analysis and simulation; e.g., we do not need the approximation formula in (3.27).

We are now ready to consider specific parameter values for our constant-persistence stable-noise model. We can select the index from the slope of the log-log plots, as in Figure

3.7. We then can set the scale parameter κ by looking at the quantile ratios. We have chosen the value $\kappa = 0.029$. We can choose the skewness to match the shape. We compare plots of the distribution of either B_n or X_n to plots of stable distributions as a function of the skewness parameter β. In this informal way, we picked $\beta = 0$; see, Appendix §B.8 for the details.

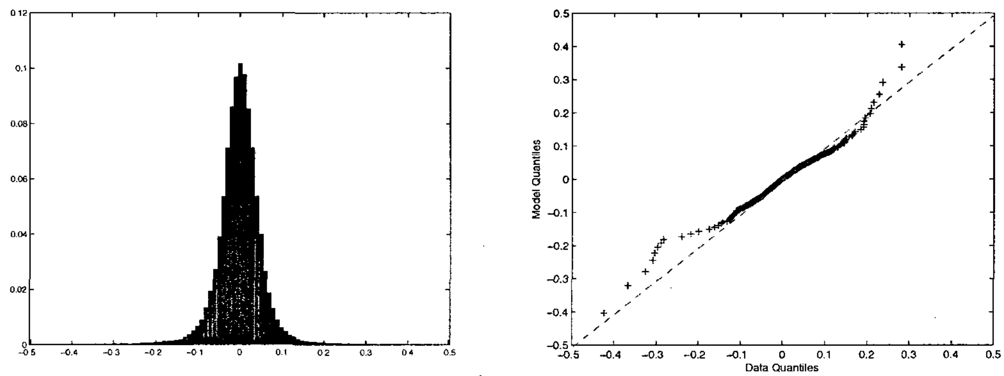
Figure 3.8 (a) shows the estimated relative-return distribution from the calibrated constant-persistence stable-noise model. Note that the chosen value of $\kappa = 0.029$ matches the peak of the distribution from the data and model reasonably well; see Figure 3.4 (a) for comparison. Figure 3.8 shows that the model approximates the empirical distribution reasonably well. However, Figure 3.8 (b) shows that the tails of the simulated distribution from the model fits the tails of the distribution from the data only roughly, not as good as Figure 3.5 (d).

Now we further test the validity of the model by comparing the quantile ratio in Table 3.3 and c in (3.35). Since the quantile ratio is estimated from the data and c is predicted by the model, if they coincide, the validity of the model is verified. It turns out that the model with calibrated $\alpha = 1.6$ and $\gamma = 0.33$, $\kappa = 0.029$ from the data generates $c = 1.1232$ which is consistent with Table 3.3. This provides solid support for the constant-persistence stable-noise model.

3.8 An Additional Model Test: Hitting Probabilities

In this section, we consider the probability that the hedge-fund relative return ever exceeds some level during the 5-year time period. Such hitting probabilities are important for risk management. We consider high or low levels of relative returns, measured in units of (sample) standard deviation σ. By simply counting the number of hedge funds whose relative returns have ever reached the level during 5-year period (2000-2004), we calculate the hitting probability from the data.

Table 3.4 shows the hitting probabilities of each level for five years from the data within the fund-of-fund strategy and the corresponding beta-persistence t-noise, beta-persistence mixed noise and constant-persistence stable-noise models. The probability estimate from



(a) Constant-persistence stable-noise model

(b) Q-Q plot comparing model and data

Figure 3.8: (a) A simulation estimate of the relative-return distribution (sample size 10^6) of the constant-persistence stable-noise model with $\alpha = 1.6, \beta = 0, \kappa = 0.029$ (to be compared to Figure 3.4 (a)). (b) Q-Q plot comparing the predicted relative-return distribution based on the constant-persistence stable-noise model to the empirical distribution from the fund-of-fund TASS data.

the data is the observed proportion of funds whose relative returns had ever hit the level during the entire five-year period, among the 92 total number of funds within fund-of-fund strategy in 2000. The initial relative return in the model simulation is set to have the stationary limiting distribution of each model, i.e., X_∞.

We perform two different simulation estimates. First, in order to estimate the true hitting probabilities, we generate 10,000 independent values of X_∞ for initial relative returns, using (3.27) and (3.35) and then use the recursion $X_n = A_n X_{n-1} + B_n$ to calculate 95% confidence interval of hitting probability throughout five years. Second, in order to assess whether the model is consistent with the data, given the small sample size, we simulate 92 independent values of the X_∞ random variables as the initial relative returns in 2000 and then use the recursion formula of $X_n = A_n X_{n-1} + B_n$ to determine the hitting probability within 5 years. We repeat 20 of these simulations and record the maximum and minimum hitting probability observed and investigate if the range of hitting probabilities includes the probability from the data. It is observed that the hitting probabilities for the high level fit the probability from the data relatively well. However, all the first estimates predict higher hitting probabilities for the low levels than are predicted from the data estimates. Nevertheless, the range of probabilities from the 20 simulations includes the hitting-probability estimates from data in most cases. (See, also Appendix B.7 for corresponding results for the Beta-persistence empirical-noise model.)

Table 3.4: Hitting probabilities of thresholds over a five-year period (2000-2004)

Level[1]	data[2]	t-noise		Mixed noise		Stable noise	
		$N = 92$[3]	$N = 10,000$[4]	$N = 92$[3]	$N = 10,000$[4]	$N = 92$[3]	$N = 10,000$[4]
3 σ	0.0326	[0,0.0435]	0.0280±0.0032	[0,0.0543]	0.0174±0.0026	[0,0.0870]	0.0326±0.0035
2 σ	0.0761	[0.0326,0.1087]	0.0696±0.0050	[0.0217,0.0761]	0.0464±0.0041	[0.0217,0.1630]	0.0712±0.0050
1 σ	0.2363	[0.1739,0.3478]	0.2569±0.0086	[0.1304,0.2717]	0.2012± 0.0079	[0.1304,0.3696]	0.2593±0.0086
-1 σ	0.2391	[0.1848,0.3043]	0.2603±0.0086	[0.1196,0.3152]	0.2028± 0.0079	[0.1739,0.3587]	0.2590±0.0086
-2 σ	0.0542	[0.0326,0.1413]	0.0718±0.0051	[0.0217,0.1522]	0.0797± 0.0053	[0.0217,0.1087]	0.0670±0.0049
-3 σ	0.0326	[0,0.0543]	0.0273±0.0032	[0.0109,0.0978]	0.0516±0.0043	[0,0.0652]	0.0328±0.0035

1. $\sigma = 0.0681$, the observed standard deviation of the fund-of-fund relative returns.
2. Number of funds that have ever hit the level for 2000-2004 divided by 92, the total number in 2000.
3. Minimum and maximum of the probabilities from 20 simulations with sample size of 92 initially.
4. 95 % confidence interval of hitting probability from simulation with sample size of 10,000 initially

3.9 Conclusion

In this chapter, we proposed a stochastic difference equation (SDE) of the form $X_n = A_n X_{n-1} + B_n$ to model the relative returns of hedge funds. In §3.2-§3.3 we showed that the model is remarkably tractable, with many convenient analytical properties. Afterwards, we showed that the model is remarkably flexible for model fitting by showing how it can be calibrated to the data from the TASS database from 2000 to 2005. The foundation of our approach is persistence. It is quantified in the model via $\gamma \equiv \mathbb{E}[A_n]$. We presented a strong case for basing the model on persistence by showing that the observed persistence estimated from the data by regression is statistically significant for all but two strategies (see Table 3.1). The persistence was found to range from 0.11 to 0.49 across the eleven fund strategies.

For the emerging-market strategy, the parsimonious (two parameter) constant-persistence normal-noise model with $A_n = \gamma$ and $B_n \stackrel{d}{=} N(0, \sigma_b^2)$ provides an excellent fit, with σ_b^2 fit to the estimated relative-returns variance σ^2 directly by (3.5). However, the constant-persistence normal-noise model is not suitable for the fund-of-fund strategy, and most other strategies, largely because the relative-return distribution has heavy tails. However, we find that some strategies are well approximated by the beta-persistence normal-noise model. In particular, that is the case for the long-short equity strategy, as we show in the Appendix §B.9. We do a complete fitting for that strategy there.

For the heavy-tailed distributions, we demonstrated the SDE model flexibility by showing that a good fit can be obtained for the fund-of-fund relative-return process by choosing variables A_n and B_n in different ways. The beta-persistence mixed-noise model in §3.6.4, the constant-persistence stable-noise model in §3.7 and the beta-persistence empirical-noise model in Appendix §B.7 all produced remarkably good fits, given the limited and unreliable data. Each of these models has advantages and disadvantages: The empirical-noise model is evidently most accurate, but it is a complicated non-parametric model, which may only be useful in simulation studies. The stable-noise model has the most appealing mathematical form, but it is not as accurate and it cannot exploit the variance for fitting (since it implies infinite variance). The mixed-noise model falls in between: it has good accuracy and it is a parametric model that can use the variance for fitting, but the parametric structure is complicated, making it harder to use in mathematical analysis. But these three models are just a sample of what could be considered. They illustrate that our SDE model offers a flexible model for fitting.

We paid special attention to matching the (assumed stationary) single-year relative-return distribution, but we also evaluated the fit of the stochastic-process model over time. As shown in (3.21), the SDE model predicts that the autocorrelation coefficient should coincide with the persistence factor γ. Table 3.1 shows that is consistent with the data. In §3.8 we also showed that the model predicted 5-year hitting probabilities of high (or low) thresholds reasonably well too. The fit here was especially good for the beta-persistence empirical-noise model, as shown in Appendix §B.7. In this test, our conclusions were not as strong as we would like because of the relatively small sample sizes and the somewhat unreliable data. We think that there is the potential for even better fitting with better data.

Overall, we contend that the value of our proposed modelling approach has been demonstrated. It should be useful in other financial contexts as well, wherever persistence may exist. As we explained in §3.2, our SDE is a discrete-time analog of the common stochastic differential equation, which should be regarded as an attractive alternative when time is naturally regarded as discrete. §3.2.4 contains a numerical example illustrating how our model can be applied to go beyond data description to answer various "what if" questions. There we briefly considered how the model might be applied to quantify the value of good fund management.

Part II

Contributions to Operations Management

Chapter 4

A Sequential Auction-Bargaining Model in Procurement

4.1 Introduction

Since procurement costs often constitute a large portion of a firm's total operating costs (Bonser and Wu (2001)), selecting suppliers with attractive prices is an important decision that a firm faces in managing its supply chain. The type of transaction method for conducting procurement varies greatly depending on the context, ranging from (i) pricing (including both static and dynamic pricing) to (ii) auctions and (iii) one-to-one bargaining (Elmaghraby et al. (2005)). While the posted price mechanism is popular in business-to-consumer transactions, both auctions and bargaining are now widespread in business-to-business procurement (Bajari et al. (2008)). Each of these three methods has been studied in the operations management literature, e.g., Federgruen and Heching (1999), Chen (2007), Chen et al. (2005), and Nagarajan and Bassok (2008), but we note that many transactions in business practice cannot simply be categorized as pricing, auctions or bargaining, as they may have the characteristics of multiple methods. There have been a number of papers that combine features of pricing and auctions, e.g., Caldentey and Vulcano (2007) and Gallien and Gupta (2007). In this chapter, we consider a procurement method that sequentially combines an auction and bargaining, where the outcome of the auction is not final but is subject to further negotiation.

The motivation for this chapter stems from our observation that the use of auctions in the procurement process often does not completely determine the final outcome of the procurement decision. Auctions perform poorly when the projects are complex and the contracts cannot be fully specified (Bajari et al. (2008)). Suppliers feel that the auction format of procurement erodes their control over the final price and "commoditizes" their products (Jap (2002)). However, it is quite common that a purchasing manager solicits bids from the pool of potential suppliers (either by telephone, mail, or Internet), and based on the bids that she has received, she decides with which buyers to examine closely and to possibly negotiate. The round of indicative bidding is valuable to the buyer in the case that the cost associated with studying a supplier as part of due diligence is high. For example, in the sale of Daewoo Motors in 1999, Ford bid in the indicative round between \$5.4 and \$6.3 billion, higher than DaimlerChrysler-Hyundai's joint bid of \$4.5 billion, and subsequently Ford was chosen as the sole bidder in the final negotiation phase (Ye (2007)). In a supply chain setting, the post-auction negotiation provides the buyer and the supplier with an opportunity to learn more about the costs, knowledge, and performance expectation (Dalya and Nath (2005)).

Both auctions and bargaining are commonly used in practice, and have been thoroughly studied in the literature. Auctions provide efficiency and simplicity in connecting the buyer to the seller with the lowest cost (Manelli and Vincent (1995)), and the buyer prefers auctions when she has relatively less information about the sellers' costs (Waehrer (1999)). There exists a vast amount of papers on the theoretical analysis of auctions in both the economics and the operations management literature; we refer the reader to Krishna (2002) and Menezes and Monteiro (2005) for a review of the auction theory and Myerson (1981) for the optimal auction design. The types of auctions include: first price or second price auctions, sealed-bid or open-bid auctions, and auctions with or without a reserve price. Another common way of determining the terms of trade (e.g., supplier selection and price decision) is bargaining, which occurs between the buyer and one or many sellers (Nagarajan and Bassok (2008)). According to Bajari et al. (2008), bargaining was used in 45 percent of procurement decisions for the public-sector non-residential construction projects in Northern California between 1995 and 2000. For more references in the bargaining literature, we refer to Rubinstein (1982) or Osborne and Rubinstein (1990) for a general review and Wu (2004) for an emphasis on procurement. Examples of commonly used bargaining models include the ultimatum take-it-or-leave-it offers and the sequential alternating offers.

While there are several papers that compare auctions and bargaining in terms of minimizing the procurement cost, empirical and experimental studies yield no clear verdict. For example, Kjerstad (2005) argues from an empirical study of procurement contracts of medical and surgical products that auctions do not provide significantly lower prices compared to bargaining. Bajari et al. (2008) empirically analyze contracts awarded in the construction industry, and they find that auctions do not perform well in some cases because of the insufficient number of bidders. In an empirical study of timber sales, Leffler et al. (2003) note that conducting an auction might incur a significantly high cost to the auctioneer if she does not use professional assistance from a forestry consultant; in this case, auctions would be less preferable to bargaining. In an experimental study, Thomas and Wilson (2002) compare the first price auction to multilateral bargaining between a single buyer and multiple sellers where the buyer solicits price offers by showing each seller his rivals' price offers while restricting communication between sellers. They report that, with four sellers, the buyer's acquisition costs through auction and bargaining are almost the same. While the above-mentioned comparisons are with respect to cost, many papers make comparisons along other dimensions such as the quality of the product; see, for example, Manelli and Vincent (1995), Bonaccorsi et al. (1999), and Tunca and Zenios (2006).

In this chapter, we consider a setting where a risk-neutral profit maximizing buyer procures an indivisible product from one of many competing suppliers. We propose a model that combines an auction and bargaining sequentially in two phases. The first phase is the standard auction, such as the first or second price auctions where one seller is selected among multiple competing sellers. At this time, while the buyer's value is public information, each seller's cost is private information unknown to other sellers and the buyer. In the second phase, the buyer bargains with the chosen seller over the final price of the product. We assume that, at the beginning of the bilateral bargaining, information regarding the seller's true cost becomes available to the buyer. This assumption is justifiable *if* the value and cost information can be accurately estimated (through the on-going supplier-customer relationship, the maturity of the market, or an additional investigation as a part of due-diligence), or *if* such information is equally uncertain to both parties of bargaining. (The buyer's true value has already become public in the first stage – thus, the buyer does not have any information advantage over the chosen seller.) The outcome of bargaining results in a price that is between the seller's true cost and the buyer's valuation. During the bargaining process, the buyer has the option to purchase the product at the price that is equal to the seller's winning bid in the first phase. For this combined auction-bargaining system, we study the sellers' equilibrium bidding strategy in the first phase, and also the buyer's choice of the reserve price, if allowed.

4.1.1 Literature Review

While procurement systems that combine auctions and bargaining are not uncommon in practice, the literature on the analysis of such models is rather limited (Engelbrecht-Wiggans and Katok (2006)). Single-unit sequential auction-bargaining models with the first-phase auction and the second-phase bargaining have been studied by Bulow and Klemperer (1996), Elyakime et al. (1997), and Wang (2000).

The seminal paper in this literature is Bulow and Klemperer (1996). They propose a sequential auction-bargaining model, where the winning bidder is determined by an open-bid second price auction, followed by a take-it-or-leave-it offer by the auctioneer (which may be accepted or rejected by the only remaining bidder). The major result of this paper is that while the auctioneer's ability to make an ultimatum bargaining offer increases her expected profit, the amount of this increase is bounded above by the expected gain of having one additional bidder in the standard auction. (Thus, the buyer prefers having an additional seller in the auction as opposed to conducting the second phase of ultimatum bargaining.) While Bulow and Klemperer (1996) use the ultimatum bargaining model with incomplete information, we are able to incorporate the relative bargaining power of the buyer and the seller in a bargaining model with complete information. We also show that their major result mentioned above may not hold if a bargaining model other than the ultimatum bargaining is used.

Elyakime et al. (1997) study a single-unit sequential model, where the first phase is the first price sealed-bid auction where the auctioneer also submits a secret sealed-bid reserve price. If no bid meets the reserve price, then the second-phase bargaining takes place, and the trade occurs between the auctioneer and the most attractive bidder where the gains from the trade is split equally between them. Thus, the cost and value information becomes public at the beginning of the second phase (as in our model). They present an equilibrium bidding strategy as a solution to a first-order differential equation. Numerical results indicate that both the auctioneer and the bidders prefer this model to the auction model that does not have the possibility of second- phase bargaining.

In the model proposed by Wang (2000), the buyer has a private valuation (unknown to sellers) and does not submit any bids in the auction phase. The seller with the most attractive bid becomes the winning seller. The buyer has the choice of either accepting the winning bid of the auction as the price of the product or entering into second-phase bargaining with the winning seller. In the latter case, the winning seller's cost becomes known to the buyer while the buyer's valuation remains private, and the second-phase bargaining is modeled as a Rubinstein-style dynamic game with one-sided uncertainty. For this model, the sellers' symmetric equilibrium strategy is given as a solution to a first-order differential equation. As his model is the closest to our model, we highlight the difference between the two models. (1) While the buyer's valuation is private in his model, we model her valuation as public information. (2) Whereas the buyer in his model decides whether to continue to the second phase where the winning seller's bid no longer has any effect, the buyer in our model always continues to the bargaining phase where the winning bid remains consequential and acts as an outside option. (3) While he models equal bargaining power between the buyer and the winning seller, we allow the possibility that the buyer may have stronger bargaining power than the seller. (4) While he considers only the first price auction in the first phase, we consider both the first price and the second price auctions. (Our analysis later shows that the second price auction generates more profit to the buyer than the first price auction.) (5) More importantly, while there is no closed-form bidding strategy in Wang (2000), we find an equilibrium bidding strategy in a closed form that is simple to analyze and intuitive to understand.

While the above papers involve only a single unit, Salmon and Wilson (2008) consider a sequential auction-bargaining model with two identical units. The buyer procures one unit through an auction in the first phase, followed by a second phase where the second unit may be procured through a take-it-or-leave-it offer from the seller with the second highest bid in the auction phase. They show the nonexistence of a pure strategy for the sellers. Experimental results indicate that the buyer's profit in this model is higher than in the sequential auction where the second phase is also conducted as an auction.

In all of the auction-bargaining models mentioned thus far, an auction is followed by bargaining; we mention that Engelbrecht-Wiggans and Katok (2006) consider a model where bargaining precedes an auction. The buyer wants to procure multiple products from sellers, each capable of producing one unit. In the first phase, the sellers' costs have not been realized, and the buyer offers some sellers an opportunity in which they may commit to sell one unit at a price to be established later in the second-phase auction. These sellers are excluded from the second-phase auction, regardless of whether or not they have accepted the buyer's offer, and the second-phase auction is conducted as a generalization of the second price auction. In addition, the growing literature on the auctions with the buyout option ("buy it now" option) could be considered as a combined bargaining-auction model, where the buyout price acts as a take-it-or-leave-it offer before the bidders participate in the auction. It has been noted that when the bidders exhibit impatience over time, this option can increase the auctioneer's profit (Mathews (2004), Budish and Takeyama (2001), Hidvégi et al. (2006), Caldentey and Vulcano (2007), and Gallien and Gupta (2007)).

4.1.2 Contribution

There are a limited number of papers that combine auctions and bargaining, especially compared to the vast literature on auctions and on bargaining, and this chapter studies a sequential auction-bargaining model that complements existing models in the literature. Our model admits a symmetric equilibrium bidding strategy in a closed form. This is our main contribution. (In all existing auction-bargaining models, the bidders' equilibrium strategies are either simple truthtelling or unable to be expressed in a closed form.) As a result, we are able to compute the buyer's expected profit, and we show that she generates higher profit in the auction-bargaining model than in the standard auction or bargaining stand-alone models. In the special case of uniform cost distributions, we show that the equilibrium strategy that we find is unique. Interestingly, the equilibrium bidding strategy in our model is closely tied to the standard results in the classical auction and bargaining literature.

As our second contribution, we study the auction-bargaining model where the buyer announces a reserve price at the beginning of the auction phase. The reserve price increases the buyer's expected profit in this model, consistent with the auction model, and we show how to find the optimal reserve price. In the auction-bargaining model, we find that the buyer sets a reserve price that is less aggressive than in the standard auction-only model. We also characterize the sellers' equilibrium bidding strategies when the buyer's reserve price is present.

Our third contribution is to compare the use of the first price auction and the second price auction during the auction phase of the model. While the expected profit of the buyer in a standard auction remains the same regardless of the auction format, we find that, in the sequential auction-bargaining model where the Revenue Equivalence Theorem does not apply, the expected profit of the buyer is higher when the second price auction is used in the auction phase than when the first price auction is used. Thus, a risk-neutral profit-maximizing buyer prefers the second price auction-bargaining model to the first price auction-bargaining model.

The organization of this chapter is as follows. In Section 4.2, we describe the auction-bargaining model in detail. In Section 4.3, we summarize some preliminary results in the auction and bargaining theory for further analysis. In Section 4.4, we study the auction-bargaining model and in particular develop the sellers' symmetric equilibrium bidding strategies. Our analysis includes the use of the second price auction, the first price auction, and the buyer's reserve price. We consider extensions and variants of the auction-bargaining model in Section 4.5, and we conclude in Section 4.6. All proofs are in Chapter B of Part III, which is the appendix of this chapter.

4.2 Model Description

Suppose that a buyer (e.g., a manufacturer) wants to procure an indivisible product from one of $n+1$ potential sellers (e.g., suppliers), where $n \geq 1$. The products offered by these sellers are identical, and the sellers are indexed by $i = 1, 2, \ldots, n+1$. Each seller i's opportunity cost of the product is drawn from a random variable C_i, and we assume that the distributions of C_i's are independent and identical, having the support of $[\underline{c}, \overline{c}]$, where $\underline{c} \geq 0$ and $\overline{c} < \infty$. Let $F(\cdot)$ and $f(\cdot)$ denote the cumulative distribution function and the probability density function of each C_i, respectively. We assume that the *ex post* cost c_i of seller i is initially private information, but the distribution function $F(\cdot)$ is public information. It is convenient to define a random variable Y, which denotes the minimum of n independent and identically distributed random variables having the common distribution F, i.e., $Y \sim \min\{C_1, \ldots, C_n\}$. It follows that Y has the cumulative distribution function $G(c) = 1 - (1 - F(c))^n$ and the probability density function $g(c) = n \cdot (1 - F(c))^{n-1} f(c)$. Also, let $\overline{G}(c) = 1 - G(c)$. For simplicity of the exposition, we assume that $f(c) > 0$ for all $c \in [\underline{c}, \overline{c}]$. Lastly, we denote an indicator function by $I\{\cdot\}$.

The value of the product to the buyer is denoted by v. We assume that v is public information, thus known to every seller, and lies above the support of F, i.e., $\overline{c} \leq v < \infty$. (We can easily extend our model to the scenario where the value of the product is uncertain to both the buyer and sellers, in which case we use v to denote the expected value of the product for the risk-neutral buyer.) The assumption of publicly available v is applicable if the suppliers have long-term relationships with the buyer, if suppliers can infer the value of the product from market conditions, or if the value is equally uncertain to both the buyer and the suppliers.

We take the viewpoint of the buyer, who wants to maximize her expected profit. Maximizing the buyer's expected profit is an objective commonly employed in the optimal mechanism-design literature. In this literature, a procurement *mechanism* refers to the determination of allocation and payment, where the outcome depends only on the bids submitted by the sellers. In this chapter, however, we do not restrict our attention to the class of procurement mechanisms; instead, we allow that the outcome may also depend on the sellers' costs, possibly by auditing private information. We use the term *procurement* *system* to refer to this broader class of procurement methods. Examples of procurement systems include auctions and bargaining (discussed in Section 4.3).

In this chapter, we study a procurement system that combines an auction and bargaining sequentially. In our model, the first-phase auction is used to select the winning seller, and the second-phase bargaining determines the price (payment), which depends on both the bids submitted during the auction phase and the winning seller's cost. We call this an *auction-bargaining* (A-B) procurement system. In the first phase, each seller i observes his *ex post* cost $C_i = c_i$, which is private information known to him only, and then submits a sealed bid to the buyer. This bid represents a price at which the seller is willing to sell the product, should the buyer find it agreeable. The buyer selects the seller with the most competitive bid (i.e., the lowest bid price). Once the winning seller is chosen, the buyer can purchase the product at that price, or enter into the second-phase bilateral bargaining with the winning seller for a possibly better price. As soon as the winning seller is determined at the end of the first phase, and prior to entering the bargaining phase, we assume that the buyer discovers the true cost to the winning bidder either through additional investigation or auditing the cost structure of the seller. (The notion of a post-auction audit has been used in multi-attribute auctions; see the scoring rule auction by Che [11 and 12] and its application in Beil and Wein (2003), Cachon and Zhang (2006), Benjaafar et al. (2007), and Wan and Beil (2008).) We also assume that the buyer obtains this information without additional effort; we can easily extend our model to the case of positive cost associated with this additional effort by comparing it to the benefit of the bargaining phase. (Without the post-auction audit, the analysis of the model becomes quite involved as in Wang (2000), and it does not yield results that are as simple as those presented in this chapter.) At any time during the bargaining process, the buyer can purchase the product from the seller at the winning bid, which acts as an "outside option" that the buyer can exercise. Thus, the final price is bounded above by the winning seller's bid and below by the winning seller's cost. Since the auction phase of the A-B model is conducted using a first price sealed-bid format, we refer to this model as the *first price sealed-bid* A-B model, or simply the first price A-B model.

If the first-phase auction is instead conducted using an open descending-price auction, then we refer to this model as the *second price open-bid* A-B model, or the second price A-B model. Here, the buyer starts the bidding process at a high price, and continuously lowers the price. Each bidder is initially active in the bidding process, and continues to remain active as long as he is willing to sell the product to the seller at the current bid price, and drops out when the current bid is no longer attractive. The bidding stops when there remains exactly one seller, who becomes the winning bidder. The buyer enters a bargaining process with the winning bidder, and the winning bid acts as an outside option for the buyer, as in the first price A-B model.

We remark that our intention is *not* to design the optimal procurement system, departing from the main focus of the mechanism design literature of economics. (For the buyer's profit-maximizing mechanism in the presence of the post-auction audit, it is possible to design a mechanism where the audit eliminates information asymmetry and the buyer extracts full rent; see Eső and Szentes (2007). However, such a mechanism is relatively complicated and may not be suited for business practice.) Rather, we restrict our attention to the sequential auction-bargaining system, which is simple and easy to implement in practice.

4.3 Preliminaries

Both auctions and bargaining have been extensively studied in the literature. In this section, we summarize some of the results in the auction theory (Section 4.3.1) and the bargaining theory (Section 4.3.2), and we establish elementary properties. These results will be useful later in analyzing the auction-bargaining (A-B) system and also in comparing this system to the auction-only or bargaining-only systems.

4.3.1 Auction

In a single-unit procurement auction, also known as a reverse auction, many potential sellers bid for the right to sell to a single buyer. The seller with the lowest bid wins the auction, and the payment to the seller is set by the lowest bid (in the first price auction) or by the second lowest bid (in the second price auction). Under the assumption that the sellers' opportunity costs are independent and identically distributed, the auction has been well analyzed in the literature (see, for example, a textbook by Krishna (2002)). We review some classical results here.

We consider the symmetric bidding strategy, denoted by $\beta(c)$, where c is the *ex post* opportunity cost of a seller. Let π_A and Π_A be the expected payment and expected profit made by the buyer. Recall that C_i's are independent and identically distributed, and $\overline{G}$ denotes the complementary cumulative distribution function of $Y \sim \min\{C_1, \ldots, C_n\}$.

Lemma 4.3.1. *In the first price or the second price procurement auction without the buyer's reserve price,*

$$\beta(c) = \begin{cases} \mathbb{E}[Y|Y > c] & \textit{in the first price auction} \\ c\ , & \textit{in the second price auction} \end{cases}$$

$$\pi_A = (n+1) \cdot \mathbb{E}\Big[\mathbb{E}\big[Y \cdot I\{Y > C_{n+1}\}|C_{n+1}\big]\Big]\ ,$$

$$\Pi_A = v - \pi_A\ .$$

See, for example, Section 2.3 of Krishna (2002) for the proof of Lemma 4.3.1. Also, notice that as long as $f(c) > 0$ for $c \in [\underline{c}, \overline{c}]$, $\beta(c)$ is a strictly increasing function of c. In the above lemma, the expected payment π_A made by the buyer is the same for both the first price and the second price auctions, and this result is a consequence of the celebrated Revenue Equivalence Theorem.

In the auction theory literature, the analysis and optimization of the reserve price has been well-studied. Suppose that at the beginning of the auction, the buyer announces a reserve price r, above which she commits herself not to pay. Let β^r be the symmetric equilibrium bidding strategy under the reserve price r in the auction only model. If $r < \overline{c}$, the analysis uses $\beta^r(r) = r$ as a boundary condition instead of $\beta(\overline{c}) = \overline{c}$. Let r_A^* denote the optimal reserve price that maximizes the buyer's expected profit in the auction-only model. Let π_A^r and Π_A^r denote the expected payment and profit of the buyer, respectively. Note that the buyer can procure the product only if there exists at least one seller who bids below the reserve price r, which occurs with probability $1 - (1 - F(r))^{n+1}$.

Lemma 4.3.2. *In the first price or the second price procurement auction with the buyer's* *reserve price r,*

$$\beta^r(c) = \begin{cases} \mathbb{E}[\min\{Y,r\}|Y > c] & \text{in the first price auction with } \underline{c} \le c \le r \\ c\ , & \text{in the second price auction with } \underline{c} \le c \le r \\ \infty\ , & \text{if } c > r \end{cases}$$

$$\pi_A^r = (n+1)\cdot\mathbb{E}\Big[\mathbb{E}\big[\min\{Y,r\}\cdot I\{Y > C_{n+1}\}|C_{n+1}\big]\cdot I\{C_{n+1}\le r\}\Big],$$

$$\Pi_A^r = \Big[1-(1-F(r))^{n+1}\Big]\cdot v - \pi_A^r\ .$$

Furthermore, the optimal reserve price $r = r_A^*$ satisfies

$$v - r = \frac{F(r)}{f(r)}\ .$$

See, for example, Section 2.5 of Krishna (2002) for the proof of Lemma 4.3.2. Also, notice that if $F(r)/f(r)$ is weakly increasing in r, the solution to the above equation is unique.

4.3.2 Bargaining

We review Rubinstein (1982)'s bilateral bargaining model of the alternating offers under complete information. Consider a bargaining game between the buyer with the valuation v and the seller with the opportunity cost c, where $v \ge c$. Let $1-\lambda$ be the *bargaining power* of the seller, where $\lambda \in [0,1]$. (The bargaining power depends on the relative discount rates of the buyer and the seller, and also on which player first proposes a take-it-or-leave-it offer. If the seller proposes first, and $\delta_s, \delta_b \in (0,1)$ denote the discount factors of the seller and the buyer, then $1-\lambda = (1-\delta_b)/(1-\delta_s\delta_b)$. Note that the seller's bargaining power $1-\lambda$ is increasing in δ_s and decreasing in δ_b.) We suppose that v, c, and λ are public information.

Lemma 4.3.3. *In the unique subgame perfect equilibrium of the alternating-offers bargaining game, the pricing outcome function is given by*

$$\gamma(c) = \lambda c + (1-\lambda)v.$$

Thus, if the seller has as much bargaining power as possible, i.e., $1-\lambda = 1$, the price outcome of bargaining is v, capturing all the gains from the trade and leaving zero profit to the buyer. The proof of Lemma 4.3.3 is standard and can be found, for example, in

Rubinstein (1982) and Osborne and Rubinstein (1990). Note that the seller makes profit of $(1-\lambda)(v-c)$, which is bargaining power times the gains from the trade, whereas the buyer earns $\lambda(v-c)$. Lemma 4.3.3 implies that a strong bargaining power of the seller (with a low value of λ) results in a high negotiated price $\gamma(c)$. While the price depends on both c and v, we hereafter suppress its dependency on v for the simplicity of notation by treating the buyer's valuation v as fixed.

We now consider the case with multiple sellers, in which the buyer bargains with the sellers in a sequential manner. As before, the buyer's valuation v is public information. We suppose that the opportunity costs and bargaining powers of the sellers (c_i and $1-\lambda_i$) are also public. (This is an additional assumption to those assumptions given in Section 4.3.1.)

At any point, the buyer may continue bargaining with the current seller or discontinue the current bargaining and start bargaining with a new seller. Once the buyer aborts bargaining with a seller, she cannot re-enter another round of bargaining with the same seller at a later point in time. For our discussion in this section, suppose that the sequence of sellers for bargaining is exogenously fixed and that the objective is to minimize the buyer's payment (the cost of bargaining is negligible).

To analyze this sequential bargaining game, we first consider a simpler case with two sellers. Let $\gamma_i(c) = \lambda_i \cdot c + (1-\lambda_i) \cdot v$, where $1-\lambda_i$ represents the bargaining power of seller i. If the buyer bargains with the second seller (seller 2) only, then the outcome of the price is $\gamma_2(c_2)$. Thus, in bargaining with the first seller (seller 1), the buyer has an option of aborting bargaining with seller 1 and then starting bargaining with seller 2. This option is analytically equivalent to exercising an option to purchase from buyer 2 at price $\gamma(c_2)$. During the bargaining with the first seller, this option involves procuring the product from someone not involved in the current bargaining process, and it is referred to as an *outside option* in the bargaining theory literature. (By contrast, the inside option refers to an option to purchase from the seller currently being negotiated. See Binmore (1985), Shaked (1994), and Muthoo (1999) for details.) If $\gamma_1(c_1) < \gamma_2(c_2)$, then the buyer will bargain with seller 1 to reach the negotiated price of $\gamma_1(c_1)$; if $\gamma_1(c_1) > \gamma_2(c_2)$, then either the first seller will propose the price of $\gamma_2(c_2)$, or the buyer will immediately move to seller 2 for bargaining. Thus, the price that the buyer pays is $\min\{\gamma_1(c_1), \gamma_2(c_2)\}$. Extending the analysis to $n+1$ sellers (where $n \geq 1$), we can easily observe that the pricing outcome function of the sequential bargaining game is $\min_i \gamma_i(c_i)$. Note that the expression for the price is independent of the sequence of sellers. It is shown in a similar framework that the buyer is indifferent to the sequence of bargaining while the sellers prefer to bargain earlier (Nagarajan and Bassok (2008) and Nagarajan and Sosic (2008)).

Let π_B be the expected payment of the buyer by taking the expected value with respect to all possible realizations of $(C_1, \ldots, C_n, C_{n+1})$, and let Π_B be the expected profit of the buyer.

Lemma 4.3.4. *In the sequential bargaining model,*

$$\begin{aligned} \pi_B &= \sum_{i=1}^{n+1} \mathbb{E}\Big[\mathbb{E}\big[\gamma_i(C_i) \cdot \mathbb{P}[\gamma_i(C_i) \leq \gamma_j(C_j),\ \forall j \neq i]|C_i\big]\Big], \\ \Pi_B &= v - \pi_B \ . \end{aligned}$$

Furthermore, if $\lambda_i = \lambda$ *for each* i*, then* $\min_i\{\gamma_i(c_i)\} = \gamma(\min_i c_i)$*, and it follows that*

$$\pi_B = (n+1) \cdot \mathbb{E}[\gamma(C_{n+1}) \cdot \overline{G}(C_{n+1})] \ .$$

Both in the auction model and in the sequential bargaining model, the buyer procures the product from one of the competing sellers, but the expected payments are not identical. The celebrated revenue equivalence result is not applicable here since, in the bargaining model, the payment by the buyer to the winning seller depends on the bargaining powers of all the sellers. Thus, the bargaining model is not a procurement *mechanism* in the classical sense. In fact, we caution the reader that these two models should not be compared directly since the auction-only model assumes that the buyer does *not* know the sellers' costs, contrary to the assumption made in the sequential bargaining model.

4.4 Analysis of the Auction-Bargaining (A-B) Model

In this section, we analyze the auction-bargaining (A-B) model described in Section 4.2. We first consider the first price A-B model where the sellers have identical bargaining power $1-\lambda$ and the buyer does not set any reserve price (Section 4.4.1). We find a symmetric bidding strategy equilibrium, and we derive an expression for the expected payment by the buyer, which we compare to the auction-only and sequential bargaining-only counterparts. We also consider the model where a reserve price is set by the buyer in the auction phase (Section 4.4.2) and the second price A-B model (Section 4.4.3).

4.4.1 First Price A-B Model

The A-B model consists of two phases: the auction phase followed by the bargaining phase. The bids submitted by sellers during the auction phase not only are used to determine the winning seller (with the lowest bid) but also act as a price for the outside option which the buyer can exercise subsequently in the bargaining phase. Thus, when a seller submits his bid, he strikes a balance between increasing the probability of winning in the auction phase and decreasing the price of the buyer's option in the event that he becomes the winning bidder. In this section, we study the symmetric bidding strategy of sellers in the auction phase of the first price A-B model.

We suppose that the sellers have the identical bargaining power, i.e., $1 - \lambda_i = 1 - \lambda$ for all i. We are interested in the symmetric bidding strategy of the first phase, which we denote by ψ. It is straightforward to show that any reasonable bidding strategy satisfies $\psi(c) \geq c$; otherwise, the bidder's profit would be negative. We assume that ψ is a strictly increasing function of the bidder's opportunity cost c. Recall that β and γ represent the symmetric bidding strategy in the first price auction-only model and the pricing outcome function of the bilateral bargaining game, respectively. We assume a technical condition: these two functions β and γ intersect finitely many times in $[\underline{c}, \overline{c}]$.

We start with the following lemma stating the relationship of the ψ function and γ function. The proof of this result is based on the following observation that if a bidder with cost c places an auction phase bid larger than $\gamma(c)$, then the price outcome in the subsequent bargaining phase cannot exceed $\gamma(c)$. Thus, by decreasing his bid price to $\gamma(c)$, the bidder increases the probability of his winning without affecting his profit in the case that he wins the auction phase. The proof of Lemma 4.4.1 and all other proofs are located in Chapter B of this thesis.

Lemma 4.4.1. *In the first price A-B model, any strictly increasing equilibrium bidding* *strategy ψ in the auction phase satisfies, for all $c \in [\underline{c}, \overline{c}]$,*

$$\psi(c) \leq \gamma(c) .$$

If the bidder with *ex post* cost c wins the auction phase, then the price outcome of the subsequent bargaining phase is the minimum of his winning bid in the first phase and $\gamma(c)$. Thus, as a corollary of the above lemma, the winning bidder receives $\psi(c)$ from the buyer.

In Theorem 4.4.2 below, we present a symmetric equilibrium strategy ψ in the auction phase. It turns out that this equilibrium strategy is given by a simple expression involving β and γ. However, we introduce a technical condition that is required by Theorem 4.4.2. Indeed, we show later that if this condition fails to be satisfied, then ψ may not exist. Recall that $1 - \lambda$ is the bargaining power of the seller and λ is the slope in the definition of the bargaining outcome price function γ. Also, recall that g and G represent the probability density function and cumulative distribution function of $Y \sim \min\{C_1, \ldots, C_n\}$, respectively. For any given bidding strategy $\psi : [\underline{c}, \overline{c}] \rightarrow \Re^+$, define Γ^ψ to be a subset of $[\underline{c}, \overline{c}]$ where the inequality in Lemma 4.4.1 is tight, i.e., $\Gamma^\psi = \{c \in [\underline{c}, \overline{c}] \mid \psi(c) = \gamma(c)\}$.

Condition 1. *A first phase bidding strategy $\psi : [\underline{c}, \overline{c}] \rightarrow \Re^+$ satisfies the following condition: for any interior point c of Γ^ψ,*

$$\lambda \geq \frac{g(c)}{\overline{G}(c)}(\gamma(c) - c) . \tag{4.1}$$

Condition 1 is a technical assumption required to prove Theorem 4.4.2, which provides a simple expression for the first phase equilibrium bidding strategy. A sufficient condition for this bidding strategy to satisfy Condition 1 will be given later in Lemma 4.4.3. In the uniform $[0, 1]$ distribution case, this sufficient condition is equivalent to the fact that the seller's bargaining power $(1 - \lambda)$ does not exceed $1/(n + 1)$. This is not unreasonable since there are a total of $n + 2$ agents in the system ($n + 1$ sellers and one buyer).

Theorem 4.4.2. *Let ψ be a strictly increasing function defined on $[\underline{c}, \overline{c}]$ by*

$$\psi(c) = \min\{\beta(c), \gamma(c)\} .$$

If ψ satisfies Condition 1, then it is a symmetric equilibrium bidding strategy in the first phase of the first price A-B model. Furthermore, if ψ does not satisfy Condition 1, then it is not a symmetric bidding strategy.

Theorem 4.4.2 implies that a seller's bid is more aggressive in the A-B model than in the auction-only model. The upper bound of γ on ψ (Lemma 4.4.1) partly explains the aggressiveness in the sellers' first-stage bidding strategy. Each bidder does not submit any bid higher than the corresponding γ value, and furthermore he knows that the other bidders will be more aggressive than they were in the auction-only model. It is also interesting to note that as the sellers' bargaining power $1-\lambda$ decreases, $\gamma(c)$ decreases and sellers bid more aggressively. Thus, the degree of the sellers' aggressiveness in the first phase represents the relative weakness in their bargaining position against the buyer. In the extreme case where $1-\lambda$ is 0, $\gamma(c)$ becomes c, i.e., each seller bids his true cost and gains zero profit. In the other extreme case of $1-\lambda=1$, it follows $\gamma(c)=v$ and $\psi(c)$ becomes $\beta(c)$, a symmetric equilibrium bidding strategy of the auction-only model.

The following result provides a sufficient condition for Condition 1, which is easy to verify when $\beta'(c)$ is readily available.

Lemma 4.4.3. *A first-phase bidding strategy* $\psi = \min\{\beta(c), \gamma(c)\}$ *satisfies Condition 1 if* $\lambda \geq \beta'(c)$ *for all* $c \in [\underline{c}, \overline{c}]$.

Lemma 4.4.3 is useful when the bidding strategy β of the auction-only model is absolutely continuous. For example, if F is a uniform distribution, then β is linear and one can easily verify whether $\lambda \geq \beta'(c)$ is satisfied. (We remark that Lemma 4.4.3 also holds if the condition $\gamma'(c) = \lambda \geq \beta'(c)$ holds only for interior points of Γ^{ψ}; however this condition is stated such that it is independent of ψ.)

The following corollary compares the expected payment of the buyer in the first price A-B model under the symmetric equilibrium given in Theorem 4.4.2 to the expected payment in the auction-only model or in the sequential bargaining model. It states that the A-B model generates higher profit to the buyer than the auction-only or sequential bargaining model. This result can be explained by the fact that each seller's bid in the A-B model (see Theorem 4.4.2) is less than or equal to both his corresponding bid in the auction-only model or the price outcome in the bargaining model. For a given symmetric equilibrium bidding strategy ψ for the auction phase, let π^{ψ}_{AB} and Π^{ψ}_{AB} denote the expected payment and profit of the buyer, respectively. (Recall that the pair of π_A and Π_A and the pair of π_B and Π_B have been similarly defined for the auction-only model and the sequential bargaining-only model, respectively. Also, recall that C_{n+1} has a cumulative density function F.)

Corollary 4.4.4. *In the first price A-B model, suppose that* $\psi(c) = \min\{\beta(c), \gamma(c)\}$ *is a strictly increasing symmetric equilibrium satisfying Condition 1. Then, the buyer's expected payment and profit are given by*

$$\begin{aligned} \pi^{\psi}_{AB} &= (n+1) \cdot \mathbb{E}\left[\min\{\beta(C_{n+1}),\ \gamma(C_{n+1})\} \cdot \overline{G}(C_{n+1})\right], \\ \Pi^{\psi}_{AB} &= v - \pi^{\psi}_{AB}. \end{aligned}$$

Furthermore, $\Pi^{\psi}_{AB} \geq \max\{\Pi_A,\ \Pi_B\}$.

The proof of Corollary 4.4.4 is straightforward and therefore omitted. To establish the inequality of $\Pi^{\psi}_{AB} \geq \max\{\Pi_A,\ \Pi_B\}$, we observe from Lemmas 4.3.1 and 4.3.4 that $\pi_A = (n+1) \cdot \mathbb{E}\left[\beta(C_{n+1}) \cdot \overline{G}(C_{n+1})\right]$ and $\pi_B = (n+1) \cdot \mathbb{E}\left[\gamma(C_{n+1}) \cdot \overline{G}(C_{n+1})\right]$.

We recall from Section 4.3 that the ranking between the auction-only model and the sequential bargaining model depends on the model parameters. Corollary 4.4.4 above establishes that, for the risk-neutral buyer, the equilibrium of the A-B model given in Theorem 4.4.2 is preferable to both the auction-only model and the sequential bargaining model. Thus, if this equilibrium is the unique equilibrium, then it is not necessary for the buyer to study model parameters when she is faced with the question of which procurement system would result in the largest expected profit. The A-B model results in an expected profit for the buyer that is higher than the expected profit of the other two models studied in Section 4.3. We caution the reader that this result is based on our modeling assumption that running an auction or a bargaining round is costless. When the cost of running a *bargaining round* is not negligible, we determine whether the first price A-B model is still preferable to the auction-only model by comparing this cost to the difference in expected profits, $\Pi^{\psi}_{AB} - \Pi_A$. Similarly, when the cost of running an *auction* is not negligible, we compare the first price A-B model to the sequential bargaining model by considering $\Pi^{\psi}_{AB} - \Pi_B$. We also note that while the first price A-B model runs only one bargaining round, the sequential bargaining model can possibly run multiple rounds of bargaining if the buyer does not discover a seller's cost until she enters bargaining with the seller. Thus, the first price A-B model has an advantage of identifying the most competitive seller with whom the buyer can bargain. (See Section 4.5 for details.)

While Theorem 4.4.2 shows one equilibrium bidding strategy for the A-B model, we now consider the possibility of other symmetric increasing bidding strategies for ψ. Before we continue with our analysis, we revisit the symmetric equilibrium bidding strategy in the first price auction-only model, discussed in Section 4.3.1. For this model, $\beta(c) = \mathbb{E}[Y|Y > c]$ is an equilibrium bidding strategy, and it is a solution to the following differential equation:

$$0 = \frac{\partial}{\partial z}\left[\, \overline{G}(z) \cdot (\beta(z) - c)\right]\Big|_{z=c} = -g(c)(\beta(c) - c) + \overline{G}(c)\frac{\partial}{\partial c}\beta(c) .$$

Solving this differential equation, we obtain a family of non-intersecting solutions characterized by the constant of integration $K \in (-\infty, \infty)$:

$$\beta_K(c) = \mathbb{E}[Y|Y > c] + \frac{K}{\overline{G}(c)} .$$

If $K = 0$, then β_K corresponds to β given above in Lemma 4.3.1. If $K > 0$, then it can be shown that β_K forms a symmetric equilibrium for the first price auction-only model, but it has the undesirable property that it approaches ∞ as c approaches $\overline{c}$ from the left. If $K < 0$, then β_K is not monotone in $[\underline{c}, \overline{c}]$; however, it is used to define the first price bidding strategy when a buyer specifies a reserve price. (When the reserve price is $r \in [\underline{c}, \overline{c}]$, then the equilibrium bidding strategy is given by $\beta^r = \beta_K$, where $\beta_K(r) = r$.) The above definition of β_K proves to be convenient in defining bidding strategies for the first price A-B model.

Theorem 4.4.5. *In the first price A-B model, suppose that a continuous and increasing function ψ satisfies Condition 1. Then, ψ is an equilibrium bidding strategy for the auction phase of the first price A-B model if and only if*

(i) $\psi(c) \le \gamma(c)$ *for* $c \in [\underline{c}, \overline{c}]$, *and*

(ii) there exists $\{a_0, a_1, \ldots, a_m\}$ *satisfying* $\underline{c} = a_0 \le a_1 \le a_2 \le \cdots \le a_m = \overline{c}$ *such that for each interval* $I_i = [a_{i-1}, a_i]$ *where* $i = 1, \ldots, m$, *either*

$$\psi(c) = \beta_{K_i}(c) \text{ for some } K_i \in (-\infty, \infty), \qquad \text{or} \qquad \psi(c) = \gamma(c) .$$

Furthermore, $\psi(\overline{c}) = \overline{c}$.

Since the family of curves $\{\beta_K \mid K \in (-\infty, \infty)\}$ parameterized by K are non-intersecting and ψ is continuous, we can assume without loss of generality that the sequence of intervals I_i's alternatively satisfies $\psi(c) = \beta_{K_i}(c)$ or $\psi(c) = \gamma(c)$. Thus, Theorem 4.4.5 shows that an equilibrium bidding strategy ψ in the first price A-B model consists of alternating γ and β_{K_i} functions, provided that Condition 1 is satisfied whenever ψ coincides with γ. This result is useful in constructing ψ. The next result states that Condition 1 must be satisfied for every symmetric equilibrium bidding strategy.

Lemma 4.4.6. *If a continuous increasing function ψ does not satisfy Condition 1, then ψ is not a symmetric equilibrium bidding strategy in the first phase of the A-B model.*

We now remark on statement (ii) in Theorem 4.4.5, focusing on the right-most interval $I_m = [a_{m-1}, a_m]$. For $c \in I_m$, either $\psi(c) = \beta_{K_m}(c)$ or $\psi(c) = \gamma(c)$. In the former case, we must have $K_m = 0$. (Otherwise, strictly positive K_m implies $\lim_{c \uparrow \bar{c}} \beta_{K_m}(c) \to \infty$, contradicting the definition of ψ in Theorem 4.4.2, and strictly negative K_m implies decreasing β_{K_m}, contradicting the monotonicity of ψ.) In the latter case, we have $\psi(\bar{c}) = \gamma(\bar{c})$, which is possible only if $\psi(\bar{c}) = \gamma(\bar{c}) = \beta_0(\bar{c}) = \bar{c}$. Thus, we obtain $\psi(c) \leq \bar{c}$, which states that the first-phase bid would not be higher than the highest possible cost, which is a reasonable outcome.

The following result is useful in constructing another symmetric equilibrium bidding strategy from a given equilibrium. In the first proof of Corollary 4.4.7, we replace a sequence of a symmetric bidding strategy ψ with γ; in the second part, we replace it with β_K for some K. The proof of Corollary 4.4.7 follows from Theorem 4.4.5 and Lemma 4.4.6 and is omitted.

Corollary 4.4.7. *Let ψ be an equilibrium symmetric bidding strategy for the auction phase of the first price A-B model. Let s and t satisfy $\underline{c} \leq s < t \leq \bar{c}$, and let ψ° be an increasing function such that $\psi^\circ(c) = \psi(c)$ for $c \notin (s, t)$.*

(a) Suppose both $\psi(s) = \gamma(s)$ and $\psi(t) = \gamma(t)$ hold, and inequality (4.1) holds for $c \in (s, t)$. If $\psi^\circ(c) = \gamma(c)$ for $c \in (s, t)$, then ψ° is an equilibrium bidding strategy with $\Pi_{AB}^{\psi^\circ} \leq \Pi_{AB}^{\psi}$.

(b) *Suppose both* $\psi(s) = \beta_K(s)$ *and* $\psi(t) = \beta_K(t)$ *hold for some* K, *and* $\psi(c) \leq \beta_K(c)$ *holds for* $c \in (s,t)$. *If* $\psi^\circ(c) = \beta_K(c)$ *for* $c \in (s,t)$, *then* ψ° *is an equilibrium bidding strategy with* $\Pi_{AB}^{\psi^\circ} \leq \Pi_{AB}^{\psi}$.

Corollary 4.4.7 shows that there exists a partially ordered ranking among symmetric equilibria with respect to the expected profit of the buyer. Corollary 4.4.8 below shows that there is a biggest and smallest element in this partially ordered set, through yet another method of constructing an equilibrium.

Corollary 4.4.8. *Let* ψ^1 *and* ψ^2 *be equilibrium symmetric bidding strategies for the auction phase of the first price A-B model. Then, both* $\psi^m(c) = \min\{\psi^1(c), \psi^2(c)\}$ *and* $\psi^M(c) = \max\{\psi^1(c), \psi^2(c)\}$ *are also equilibrium symmetric bidding strategies. Furthermore,*

$$\Pi^{\psi^M} \leq \min\{\Pi^{\psi^1}, \Pi^{\psi^2}\} \leq \max\{\Pi^{\psi^1}, \Pi^{\psi^2}\} \leq \Pi^{\psi^m} .$$

The proof of Corollary 4.4.8 is straightforward and therefore omitted.

Notice that Theorem 4.4.5 implies the possibility of multiple equilibrium bidding strategies, each characterized by a collection of intervals and K_i values. The profit-comparison result of Corollary 4.4.4 (i.e., $\Pi_{AB}^{\psi} \geq \max\{\Pi_A, \ \Pi_B\}$) corresponds to a particular equilibrium (given by $\psi(c) = \min\{\beta(c), \gamma(c)\}$), and this result may not hold with other equilibrium bidding strategies. The following theorem shows that the above comparison of profits holds for any symmetric equilibrium bidding strategy.

Theorem 4.4.9. *For any increasing and continuous symmetric equilibrium bidding strategy* ψ *in the A-B model that satisfies Condition 1, we have* $\Pi_{AB}^{\psi} \geq \max\{\Pi_A, \ \Pi_B\}$.

EXAMPLE: UNIFORM [0,1] COST IN THE FIRST PRICE A-B MODEL.

We consider an example of uniform opportunity costs, where each C_i has a uniform distribution on $[0,1]$, with $v \geq 1$. In this uniform $[0,1]$ cost case, it turns out that Condition 1 is satisfied if and only if $\lambda \geq n/(n+1)$, and in this case, we obtain the following uniqueness result.

Theorem 4.4.10. *In the first price A-B model, suppose that sellers' opportunity costs have uniform* $[0,1]$ *distribution and* $\lambda > n/(n+1)$. *Then, there exists a unique continuous strictly* *increasing symmetric equilibrium bidding strategy for the auction phase, given by*

$$\psi(c) \;=\; \min\{\beta_0(c), \gamma(c)\}\ .$$

The proof of Theorem 4.4.10 is based on the observation that whenever $\gamma(c) < \beta_0(c)$, any β_K intersecting with γ at c has the property $\beta'_K(c) < \lambda = \gamma'(c)$.

We now describe in detail the symmetric equilibrium bidding strategy given in Theorem 4.4.10. We compute the buyer's expected profit in the A-B model and compare it to the auction-only model and the sequential bargaining model. Since each seller's cost is uniformly distributed, we obtain $\overline{G}(c) = (1-c)^n$. Thus, the auction phase bidding strategy is given by $\psi(c) = \min\{\beta_0(c),\ \gamma(c)\}$, where

$$\beta_0(c) \;=\; \frac{1}{n+1} + \frac{n}{n+1}\cdot c \qquad \text{and} \qquad \gamma(c) \;=\; (1-\lambda)\cdot v + \lambda\cdot c.$$

Since both β_0 and γ are linear functions, they intersect at most once depending on the buyer's value v and the sellers' bargaining power $1-\lambda$. Let $s = \inf\{c \in [0,1] \mid \beta(c) \le \gamma(c)\}$. (Since $n/(n+1)$ and λ correspond to the slopes of β_0 and γ, s represents the intersection of these two lines if $s \in (0,1)$.) Thus,

$$\psi(c) \;=\; \min\{\beta(c),\gamma(c)\} \;=\; \begin{cases} \gamma(c)\ , & \text{if } c \in [0,s) \text{ or } c = s \in (0,1] \\ \beta(c)\ , & \text{if } c \in (s,1] \text{ or } c = s \in [0,1). \end{cases}$$

Also, from Lemma 4.4.3 and $\lambda > n/(n+1) = \beta'(c)$, it is easy to verify that ψ satisfies Condition 1.

We comment on the assumption $\lambda > n/(n+1)$. It is reasonable to expect that if there is only one potential seller, the bargaining powers of the seller and the buyer are similar, i.e., $\lambda \approx 1/2$. The seller's bargaining power typically decreases as the number of sellers increases. When there are $n+1$ potential sellers, there are a total of $n+2$ players in the market (including the buyer), and it is reasonable to expect that each seller's bargaining power satisfies $1-\lambda = 1/(n+2)$. In this case, $\lambda = (n+1)/(n+2) > n/(n+1) = \beta'(c)$. Furthermore, if $\lambda > n/(n+1)$ does not hold, then by Lemma 4.4.6 there exists no strictly increasing equilibrium bidding strategy in the A-B model with the exception of $\psi(c) = \beta(c)$.

With the Uniform $[0,1]$ cost distribution, the unique equilibrium bidding strategy in the A-B model is given by Theorem 4.4.10 of Section 4.4.1. Here, we give a closed-form expression for the buyer's expected profit in the A-B model and compare it to the other models.

Recall that the buyer's expected profits in the auction-only model and in the sequential bargaining-only model are given by the following expressions. These results are straightforward to verify.

$$
\begin{aligned}
\Pi_A &= v - (n+1)\ \mathbb{E}\left[\overline{G}(c)\beta(c)\right] \\
&= v - (n+1)\int_0^1 \beta(c)\overline{G}(c)f(c)dc \;=\; v - \frac{2}{n+2}\,, \\
\Pi_B &= v - (n+1)\ \mathbb{E}\left[\overline{G}(c)\gamma(c)\right] \\
&= v - (n+1)\int_0^1 \gamma(c)\overline{G}(c)f(c)dc \;=\; \lambda v - \frac{\lambda}{n+2}\,.
\end{aligned}
$$

Note that the ranking between Π_A and Π_B depends on the values of λ and v.

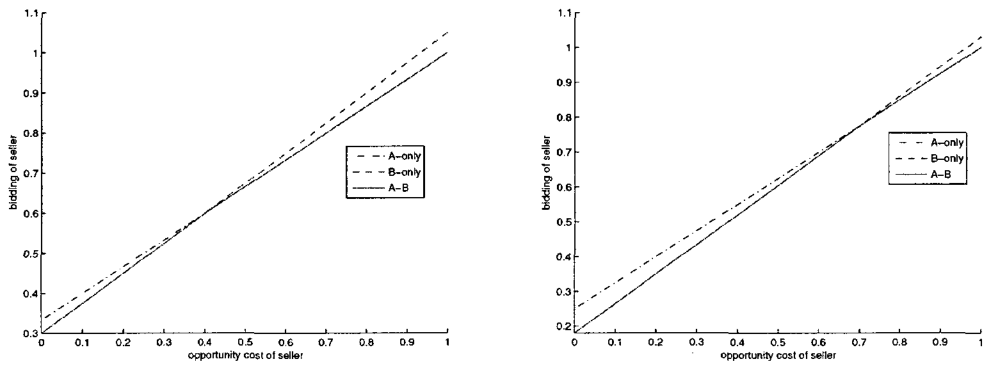
We now compute the expected profit Π_{AB} of the buyer in the A-B model. Suppose that the two linear functions β and γ intersect in $[0,1]$. The intersection point is given by

$$
s = \frac{1-(n+1)(1-\lambda)v}{(n+1)\lambda - n}\,.
$$

An algebraic manipulation shows that

$$
\begin{aligned}
\Pi_{AB} &= v - (n+1)\mathbb{E}[\psi(c)\overline{G}(c)] \\
&= v - (n+1)\left(\int_0^s \gamma(c)\overline{G}(c)f(c)dc + \int_s^1 \beta(c)\overline{G}(c)f(c)dc\right) \\
&= v - \frac{2}{n+2} + [1-(n+1)(1-\lambda)v]\ \frac{1-(1-s)^{n+1}}{n+1} \\
&\quad + [(n+1)\lambda - n]\left[\frac{s(1-s)^{n+1}}{n+1} - \frac{1-(1-s)^{n+2}}{(n+1)(n+2)}\right]\,. \qquad (4.2)
\end{aligned}
$$

Figure 4.1 shows the symmetric equilibrium bidding strategy of the A-B model in comparison to the symmetric equilibrium bidding strategy of the auction-only model as well as the pricing outcome function of the bargaining model. The expected profit Π^{ψ}_{AB} of the buyer in the A-B model can be computed in a closed form. (See equation (4.2).) Figure 4.2 shows that the expected profit of the buyer in the first price A-B model is higher than in the auction-only model or the sequential bargaining-only model. In addition, while it can be shown that the buyer's expected profit in the auction-only model or the sequential

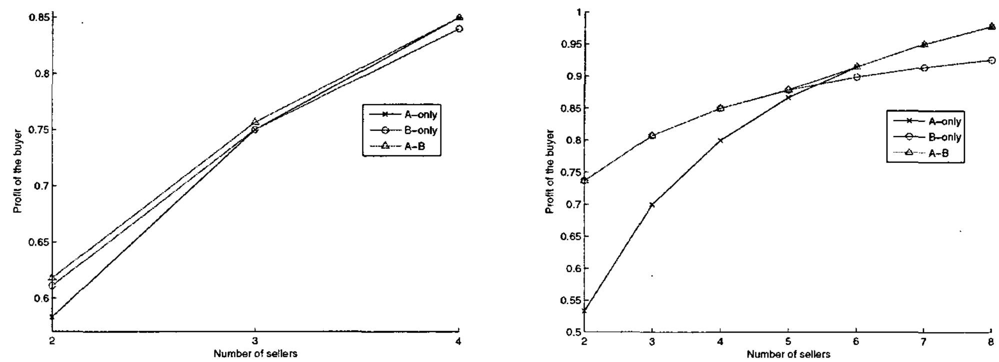


(a) $\lambda = 0.75$, and $n = 2$ (i.e., 3 sellers). (b) $\lambda = 0.85$, and $n = 3$ (i.e., 4 sellers).

Figure 4.1: The symmetric equilibrium bidding strategy of the A-B model (A-B) compared to the bidding strategy of the auction-only (A-only) model and the pricing outcome function of the sequential bargaining model (B-only). $v = 1.2$.

bargaining model can be shown to be concave with respect to the number of the sellers, it may fail to be concave in the A-B model.

It is interesting to observe from Figure 4.2 that it is not always more profitable for the buyer to add one more seller to the auction-only model than to have the second-phase of bargaining as in the A-B model. Consider, for example, the case where the number of sellers is 2 in Figure 4.2(b). This is in contrast to the conclusion of Bulow and Klemperer (1996), in which the benefit of the second-phase ultimatum bargaining is always outweighed by the benefit of having an extra bidder. Our observation is consistent with a similar finding of Wang (2000) for his auction-bargaining model, where he assumes a one-sided uncertainty of the private buyer's valuation in the bargaining phase. Whereas Bulow and Klemperer (1996) do not consider the presence of an audit, both Wang (2000) and this chapter allow an audit prior to the bargaining phase, thereby equipping the buyer with more information to benefit from the auction phase. Thus, in these two papers, it is quite plausible that bargaining may become preferable to attracting one more bidder.



(a) $\lambda = (n+1)/(n+2)$ under 2, 3, 4 sellers. (b) $\lambda = 0.85$ under $2, \ldots, 8$ sellers.

Figure 4.2: The expected profit of the buyer in the A-B model (A-B) compared to the auction-only model (A-only) and the sequential bargaining model (B-only), as a function of the total number of sellers, $n+1$. $v = 1.2$.

4.4.2 A-B Model with a Reserve Price

While we have assumed in Section 4.4.1 that the buyer does not set any reserve price, we now consider the case where the buyer announces a reserve price, over which she commits not to purchase the product from any of the sellers. This announcement is made before the bids are submitted in the first phase. By setting a reserve price, the buyer faces the risk of not being able to procure, but she may increase her expected profit by encouraging sellers to bid more aggressively.

We use r to denote the reserve price set by the buyer. Since the buyer does not want to procure the product above its value v, we proceed by assuming $r \leq v$. Recall that β^r represents the symmetric bidding strategy in the first price auction-only model with a reserve price r. Also recall that γ denotes the pricing outcome function of the bilateral bargaining game. We assume that these two functions β^r and γ intersect finitely many times in $[\underline{c}, r]$. Similar to the first price A-B model without a reserve price, we introduce a technical condition that is required by Theorem 4.4.11 below. Recall that from the definition of Γ, $\Gamma^{\psi^r} = \{c | \ \psi^r(c) = \gamma(c)\}$.

Condition 2. *A bidding strategy ψ^r satisfies the following condition: for any c satisfying*

$\underline{c} < c < r$ *in the interior point of* Γ^{ψ^r},

$$\lambda \geq \frac{g(c)}{\overline{G}(c)}(\gamma(c) - c) .$$

Note that Condition 2 is identical to Condition 1 except that the above inequality is required for $c \in (\underline{c}, r)$ only. The following theorem summarizes the main results for the first price A-B model with a reserve price.

Theorem 4.4.11. *Let* ψ^r *be a strictly increasing function defined on* $[\underline{c}, r]$ *by*

$$\psi^r(c) \quad = \quad \min\{\beta^r(c), \gamma(c)\} .$$

If ψ^r *satisfies Condition 2, then it is a symmetric equilibrium bidding strategy in the first phase of the first price A-B model with a reserve price* r.

The proof of Theorem 4.4.11 is similar to the proof of Theorem 4.4.2 and is omitted. We observe from Theorem 4.4.11 that a seller's bid in the first phase of the A-B model is more aggressive than in the auction-only model with the same reserve price. This observation is analogous to the case without any reserve prices (Section 4.4.1). Furthermore, note that the seller's bid $\psi^r(c)$ becomes more aggressive as r decreases, just as in the auction-only model. We remark that Condition 2 is required to guarantee that ψ^r forms an equilibrium.

The following corollary shows the expected profit of the buyer in the first price A-B model with a reserve price. The result is immediate, and thus we omit the proof of the corollary.

Corollary 4.4.12. *Under the conditions of Theorem 4.4.11, the buyer's expected payment and profit are given by*

$$\begin{aligned} \pi^r_{AB} &= (n+1) \cdot \mathbb{E}\left[\overline{G}(C_{n+1})\ \psi^r(C_{n+1}) \cdot I\{C_{n+1} \leq r\}\right] , \\ \Pi^r_{AB} &= \left[1 - (1 - F(r))^{n+1}\right] v - \pi^r_{AB} . \end{aligned}$$

From Lemma 4.3.2 and Corollary 4.4.12, it is clear that, for the risk-neutral buyer, the A-B model with a reserve price is preferable to the auction-only model with the same reserve price. We now consider the optimal reserve price which maximizes the buyer's expected profit. Let r^*_{AB} denote the optimal reserve price of the buyer in the first price A-B model.

The first order condition for the optimal reserve price in the first price A-B model is derived in the following lemma.

Lemma 4.4.13. *Under the condition of Corollary 4.4.12, the optimal reserve price* $r = r^*_{AB}$ *satisfies the following equation:*

$$v - r = \frac{\int_{\underline{c}}^{r} I\{\beta^r(c) \leq \gamma(c)\} \cdot f(c)\, dc}{f(r)} .$$

Note that the integral part of the above equation, $\int_{\underline{c}}^{r} I\{\beta^r(c) \leq \gamma(c)\} \cdot f(c)\, dc$, represents the probability that a seller's *ex post* cost c satisfies both $\beta^r(c) \leq \gamma(c)$ and $c \leq r$. Lemma 4.4.13 states that the optimal reserve price in the auction-only model, r^*_A, is not necessarily the optimal reserve price in the A-B model. In fact, if $F(r)/f(r)$ is weakly increasing, then the unique optimal reserve price r^*_A given in Lemma 4.3.2 can be shown to be a lower bound for r^*_{AB}. We obtain this result by comparing the equation in Lemma 4.3.2 to the equation in Lemma 4.4.13. In other words, in the A-B model, the buyer sets a less-aggressive reserve price than in the auction-only model. We attribute this observation to the fact that the seller's bid is less important in the A-B model, because the buyer learns the seller's cost before the second-phase bargaining process.

While Theorem 4.4.11 shows one equilibrium bidding strategy for the A-B model with a reserve price, we now consider the possibility of other symmetric bidding strategies for ψ^r as in the previous section. The following theorem is analogous to Theorem 4.4.5 and Lemma 4.4.6, and we omit this proof.

Theorem 4.4.14. *In the first price A-B model with a reserve price, suppose that a continuous increasing function* ψ^r *satisfies Condition 2. Then,* ψ^r *is a symmetric equilibrium bidding strategy for the first phase if and only if*

(i) $\psi^r(r) = r$

(ii) $\psi^r(c) \leq \gamma(c)$ *for* $c \in [\underline{c}, r]$

(iii) There exists $\{a_0, a_1, \ldots, a_m\}$ *and* $K_i \in (-\infty, \infty)$ *for each* K_i *such that* $\underline{c} = a_0 \leq a_1 \leq a_2 \leq \cdots \leq a_m = r$ *and for each* $I_i \in [a_{i-1}, a_i]$*, either*

$$\psi^r(c) = \beta_{K_i}(c) \text{ for some } K_i \in (-\infty, \infty), \qquad \text{or} \qquad \psi^r(c) = \gamma(c) .$$

Furthermore, $\psi^r(c) = \min\{\beta^r(c), \gamma(c)\}$ *for* $c \in I_m$.

Also, if a continuous increasing function ψ^r *does not satisfy Condition 2, then* ψ^r *is not a symmetric equilibrium bidding strategy in the first phase of the A-B model with a reserve price* r.

EXAMPLE: UNIFORM $[0,1]$ COST IN THE A-B MODEL WITH A RESERVE PRICE.

We return to the example of the uniform $[0,1]$ opportunity cost. As before, we assume $\lambda > n/(n+1)$. We also suppose $\gamma(0) \leq \beta^r(0)$; otherwise, it can be shown that $\psi^r(c) = \beta^r(c)$, implying that the A-B model becomes the same as the auction-only model. The following lemma establishes that ψ^r given in Theorem 4.4.11, $\psi^r(c) = \min\{\beta^r(c), \gamma(c)\}$, satisfies Condition 2, and is a unique first-phase bidding strategy for the first price A-B model with a reserve price r.

Lemma 4.4.15. *In the first price A-B model with a reserve price, suppose that the sellers' opportunity costs are drawn from uniform* $[0,1]$ *distribution. Let* $r \in [0,1]$ *be the reserve price, and suppose that both* $\gamma(0) \leq \beta^r(0)$ *and* $\lambda > n/(n+1)$ *hold. Then, the symmetric equilibrium bidding strategy* $\psi^r(c)$ *for the first phase is unique, given by*

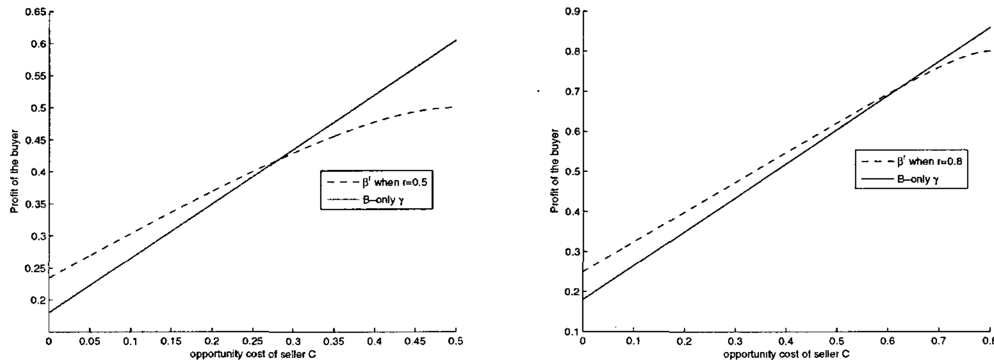
$$\psi^r(c) = \min\{\beta^r(c),\ \gamma(c)\}\ .$$

The proof of Lemma 4.4.15 is similar to Theorem 4.4.10 and is based on the observation that whenever $\gamma(c) < \beta^r(c)$, any β_K intersecting with γ at c has the property $\beta_K'(c) < \lambda = \gamma'(c)$.

In the uniform $[0,1]$ cost case, it follows from Lemma 4.3.2 that

$$\beta^r(c) = \frac{nc+1}{n+1} - \frac{(1-r)^{n+1}}{(n+1)(1-c)^n}\ ,$$

and β^r is concave with respect to c in $[0,r]$. Since $\gamma(0) \leq \beta^r(0)$ and $\gamma(r) \geq r = \beta^r(r)$, there exists a unique intersection of β^r and γ in $[0,r]$. Let $s^r \in [0,r]$ such that $\beta^r(s^r) = \gamma(s^r)$. Figure 4.3 shows the optimal bidding strategy in the auction-only model with a reserve price $r \in \{0.5, 0.8\}$ and the pricing outcome function of bargaining. We observe that the intersection of these two functions is unique for each value of r. From Lemma 4.4.15, $\psi^r(c)$ is given by the minimum of these two functions.



(a) $r = 0.5$ (b) $r = 0.8$

Figure 4.3: The bidding strategy of a seller in the auction-only model with a reserve price (β^r) and the pricing outcome function (B-only γ) with $\lambda = 0.85$, $v = 1.2$, and $n = 3$, i.e., 4 sellers.

From Corollary 4.4.12, it is straightforward to verify that the expected payment from the buyer, $\pi_{AB}^{\psi^r}$, and the expected profit of the buyer, $\Pi_{AB}^{\psi^r}$, are

$$\pi_{AB}^{\psi^r} = (n+1)\left(\int_0^r (1-c)^n \cdot \min\{\beta^r(c), \gamma(c)\}\, dc\right),$$
$$\Pi_{AB}^{\psi^r} = \left(1-(1-r)^{n+1}\right)v - \pi_{AB}^r .$$

From the first order condition and Lemma 4.4.13, the optimal reserve price r_{AB}^* maximizing the expected profit of the buyer satisfies

$$v - r \;=\; r - s^r ,$$

where s^r is the unique intersection between β^r and γ in $[0, r]$. Since $v - r$ is nonnegative, it follows that $r \geq s^r$. Thus,

$$r_{AB}^* \;=\; \frac{v + s^{r_{AB}^*}}{2} .$$

In the auction-only model, the optimal reserve price for the uniform $[0, 1]$ example is given by $r_A^* = v/2$ (Lemma 4.3.2), and it follows that $r_{AB}^* \geq r_A^*$ from the above equation. Thus, the optimal reserve price in the A-B model is less aggressive than in the auction-only model. This observation is consistent with the remark following Lemma 4.4.13.

Figure 4.4 is a numerical example of the expected profit of the buyer in the A-B model as a function of the reserve price. Clearly, the buyer's maximum expected profit in the first price A-B model with a reserve price is higher than in the auction-only model with the same reserve price. For the auction-only model, the optimal reserve price is given by $r_A^* = v/2 = 0.6$; for the A-B model, the optimal reserve price is given by $r_{AB}^* = (v+s)/2 \approx 0.95$, where s is numerically calculated to be approximately 0.7 in this case. Thus, we observe that r_{AB}^* is bigger than r_A^*. We also observe that for low values of r, the expected profit to the buyer is the same in both models. This occurs since a sufficiently small reserve price r violates the condition $\gamma(0) \leq \beta^r(0)$, implying that both ψ^r and β^r functions coincide; thus, the A-B model essentially becomes the auction-only model.

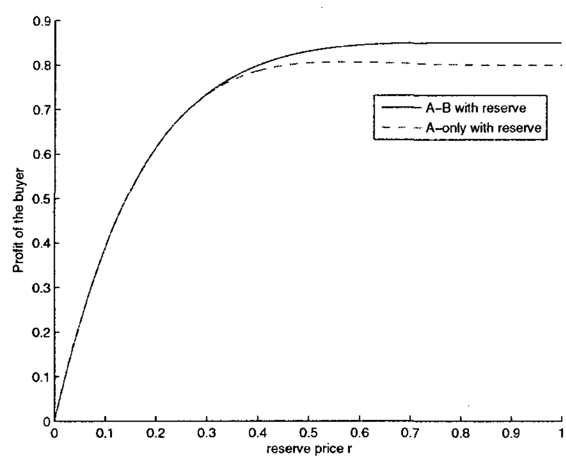


Figure 4.4: The expected profit of the buyer in the auction-only model (A-only) and the first price A-B model (A-B), both with a reserve price r, $\lambda = 0.85$, $v = 1.2$, and $n = 3$ (i.e., 4 sellers).

4.4.3 Second Price A-B Model

In the second price A-B model, the seller with the lowest bid in the first phase (auction phase) is selected for the second phase of bargaining, just as in the first price A-B model. However, the price at which the buyer can purchase the product from the winning bidder is not his bid (the lowest bid), but the second lowest bid. The second price A-B model can easily be implemented if the auction phase is conducted as an open descending-price auction. In this section, we investigate an equilibrium bidding strategy for the first phase in the second price A-B model. We also compare the expected profit of the buyer in the first and second price A-B models, and we study whether a preference ranking between the first and second price A-B models can be established.

In this section, let $\psi(c)$ be a symmetric equilibrium bidding strategy of the second price A-B model. A symmetric equilibrium bidding strategy in the second price A-B model is given in Theorem 4.4.16.

Theorem 4.4.16. *In the second price A-B model, a dominant bidding strategy for each bidder in the first phase is*

$$\psi(c) = c .$$

The proof of Theorem 4.4.16 is similar to the case of the second price auction-only model and is based on the observation that the amount of payment a bidder expects to receive conditioned on his winning does not depend on his first phase bid. Note that the above strategy is the same as the second price auction-only model and does not depend on the number of competing sellers in the system.

The celebrated Revenue Equivalence Theorem implies that the expected profit of the buyer in the first price auction-only model is the same as the corresponding quantity in the second price auction-only model. This theorem is not applicable in the A-B model which is not a *mechanism.* Now, we compare the expected profits of the buyer, and it turns out that they are generally not identical.

Theorem 4.4.17. *The expected profit of the buyer in the second price A-B model is at least that of the first price A-B model with* $\psi(c) = \min\{\beta(c), \gamma(c)\}$.

The proof of Theorem 4.4.17 is based on comparing the *ex post* expected payment received by each seller in both of the two models. Let us first review an analogous result in the auction-only model. It is well known that while the mean of the final price is the same in both the first price auction-only model and the second price auction-only model, the distribution of the final price in the latter model is a *mean-preserving spread* of the final price in the former model (e.g. Krishna (2002)). In the A-B model, a similar analysis shows that the *ex post* final price in the second price A-B model is a "spread" of the *ex post* final price in the first-price A-B model; however, the minimum operator in the definition of ψ introduces concavity in taking the expectation over the "spread". Then, by Jensen's Inequality, the expected final price is lower in the second price A-B model, as stated in Theorem 4.4.17.

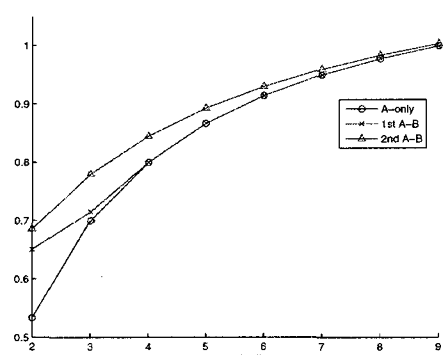
EXAMPLE: UNIFORM [0, 1] COST IN THE SECOND PRICE A-B MODEL.

In the second price A-B model, if the opportunity cost is drawn from uniform $[0,1]$ distribution, the expected profit of the buyer, denoted by Π^2_{AB}, is

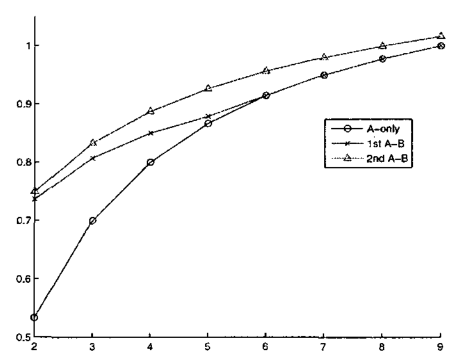
$$\begin{aligned}\Pi^2_{AB} &= v - \sum_{i=1}^{n+1} \mathbb{E}\Big[\mathbb{E}\left[\ \overline{G}(c) \cdot \mathbb{E}[\min\{Y, \gamma(c)\}|Y > c]|C_i = c\right]\Big] \\ &= v - \frac{2}{n+2} + \frac{[1-(1-\lambda)v]^{n+2} - [(1-\lambda-(1-\lambda)v]^{n+2}}{(n+2)\lambda},\end{aligned}$$

where the last equality results from algebraic manipulation. It can be shown from Lemma 4.3.1 that $\Pi_A = v - 2/(n+2)$. Since the last term above is nonnegative, it follows that $\Pi^2_{AB} \geq \Pi_A$.

Figure 4.5 shows the expected profit of the buyer in the auction-only, first price A-B, and second price A-B models. We observe that the buyer's expected profit is the highest in the second price A-B model.



(a) $\lambda = 0.75$



(b) $\lambda = 0.85$

Figure 4.5: The expected profit of the buyer as a function of the number of sellers in the auction-only (A-only), first price A-B (1st A-B), and second price A-B (2nd A-B) models. $v = 1.2$.

4.5 Extensions and Further Discussions

In this section, we consider extensions of the models discussed in Sections 4.3 and 4.4. In Section 4.5.1, we first revisit the auction-bargaining model of Section 4.4. While only one seller was selected in the auction phase for bargaining in the second phase, in Section 4.5.1, we consider the case where the buyer can select more than one seller from the first phase. In Section 4.5.2, we discuss the impact of positive fixed costs for conducting an auction or bargaining. In Section 4.5.3, we consider the sequential bargaining model of Section 4.3.2 by assuming that the buyer discovers the sellers' costs one at a time and is not allowed to resume bargaining with a seller that she had previously aborted.

4.5.1 Multiple Sellers in the Bargaining Phase of the A-B Model

In the A-B models studied in Section 4.4, the buyer selects only one seller in the auction phase and enters the second phase bargaining with this seller. In this section, suppose that the buyer selects multiple sellers in the first phase and conducts sequential bilateral bargaining with these sellers. At the end of the first phase, the buyer observes all the bids b_i submitted by sellers i, and selects the bidders with the m lowest bids, where $m \in \{2, 3, \ldots, n+1\}$. (For notational convenience, suppose that bidders $i = 1$ through m are selected for the second phase.) For the second phase bargaining process, the buyer decides the order in which she will sequentially bargain with the m sellers. At the beginning of the second phase, the buyer knows the opportunity costs c_i for all sellers $i = 1, \ldots, m$. (The second phase bargaining with m sellers follows the sequential bargaining model described in Section 4.3.2.) We refer to this model as Model M.

We find an equilibrium bidding strategy of sellers in the first phase and the bargaining strategy of the buyer in the second phase. Consider the buyer's problem. By an argument similar to the proof of Lemma 4.3.4, the buyer's payment is independent of the bargaining order and is given by $\min\{\min\{b_i, \gamma(c_i)\} \mid i = 1, \ldots, m\}$. Thus, one of the buyer's weakly dominant strategies is to bargain first with the seller with the least $\min\{b_i, \gamma(c_i)\}$ (with the least bid) until an agreement is reached. For the sellers' problem, the first-phase bidding strategy can be shown to remain unchanged from Theorem 4.4.2 of Section 4.4.

Theorem 4.5.1. *Let ψ be a strictly increasing function defined on $[\underline{c}, \overline{c}]$ by*

$$\psi(c) \;=\; \min\{\beta(c),\ \gamma(c)\}\ .$$

If ψ satisfies Condition 1, an equilibrium bidding strategy in Model M is as follows: (i) in the first phase, seller i submits bid $\psi(c_i)$, (ii) in the second phase, the buyer bargains with the seller with the least $\min\{b_i, \gamma(c_i)\}$ and reaches an agreement with him.

We consider a variation of Model M. During the second phase, the seller does not know the *ex post* cost c_i of seller i until she enters a round of bilateral bargaining with the seller. Thus, the buyer cannot use $\{c_i \mid i = 1, \ldots, m\}$ in deciding the order of bargaining. This is similar to the sequential bargaining model of Section 4.5.3. For this model, we can establish a result similar to Theorem 4.5.1, except that the buyer's strategy is to bargain with the seller with the least b_i until an agreement is reached.

4.5.2 Fixed Costs of Auctions and Bargaining

In our discussion in Section 4.4, we have assumed that the cost of conducting bilateral bargaining or an auction is negligible. Such an assumption is valid, for example, when there is a good infrastructure for conducting auctions or bargaining transactions or when the buyer's relationships with the sellers have been well maintained such that the auction or bargaining outcomes can be reached without much effort.

In this section, we consider the case where there are positive fixed costs for auction and bargaining, denoted by K_A and K_B, respectively. Then, in the auction-only model, the buyer incurs the cost of K_A, and in the sequential bargaining model, the buyer incurs the cost of at least one K_B (see the discussion in Section 4.5.3). In the A-B model, the buyer incurs the fixed costs associated with the auction and one round of bargaining, i.e., $K_A + K_B$. By comparing these fixed costs with the expected profits of the buyer in each of these models, the buyer can select the procurement system that maximizes her expected profit.

4.5.3 Sequential Bargaining and Optimal Stopping

We consider the sequential bargaining model where the buyer enters a series of bilateral bargaining with each of $n+1$ potential sellers in an exogenously determined order. Without loss of generality, we assume that this order is given by $i = 1, 2, \ldots, n+1$. In Section 4.3.2, we have assumed that the buyer knows the *ex post* cost c_i of each seller i, and it has been shown that the buyer's optimal choice is to purchase from the most efficient seller by paying $\gamma(\min_i c_i)$. In this section, we assume that the buyer does not know the realized cost c_i of seller i until she starts a bargaining process with this seller. Furthermore, the buyer cannot resume bargaining with a seller once she starts bargaining with another seller, i.e., the buyer cannot go back to a previous seller.

We note that this model is an alternative to the A-B model discussed in Sections 4.3 and 4.4. In both models, the buyer does not know a seller's cost, unless she initiates bargaining with that seller. In the A-B model, the buyer uses an auction to select one seller with whom she bargains; in the sequential bargaining model discussed here, the buyer does not use any auction, but uses a series of bargaining rounds. In this subsection, we outline a basic analysis for the sequential bargaining model.

We assume that a seller's opportunity cost is independently drawn from a common distribution F. In this case, we show that the buyer's profit-maximizing decision strategy of whether to accept a bargaining outcome with each seller is an optimal stopping problem. The following lemma for the optimal stopping criterion follows from Ferguson (2000).

Lemma 4.5.2. *In the sequential bargaining model without the buyer's* a priori *knowledge of c_i's, the optimal stopping criterion for the buyer is to accept the bargaining outcome with the j'th seller if $c_j \leq A_j$ where*

$$A_j = \begin{cases} \infty, & j = n+1 \\ \mathbb{E}[\min\{C_{j+1}, A_{j+1}\}], & j \in \{n, n-1, \ldots, 1\}. \end{cases}$$

Furthermore, if C_i is uniformly distributed on $[0,1]$, then $A_n = 1/2$ and $A_j = A_{j+1} - A_{j+1}^2/2$ for $j = n-1, \ldots, 1$.

Given the above decision rule of the buyer, seller j will choose to sell to the buyer at the price of $\gamma(c_j)$ if $c_j \leq A_j$. If $A_j < c_j \leq \gamma(A_j)$, then he will sell and receive $\gamma(A_j)$ instead since it is still profitable to do so. Otherwise, he will not reach an agreement with the buyer and the buyer will start bargaining with seller $j+1$.

We can easily extend the above analysis to the case where the buyer incurs a fixed cost K_B for each round of bargaining. This cost is incurred, for example, when the buyer researches a seller's *ex post* cost at the beginning of bargaining with the seller. Then, it can be shown that the optimal stopping criterion of Lemma 4.5.2 remains valid except for the following modification: for $j \in \{n, n-1, \dots, 1\}$,

$$A_j = \mathbb{E}[\min\{C_{j+1}, A_{j+1} - K_B/\lambda\}].$$

(Note that $1-\lambda$ is the bargaining power of the seller.)

In the special case where the buyer knows that the first seller is the most efficient seller, i.e., $c_1 = \min_i c_i$, then it is optimal for the buyer to reach an agreement with the first seller. In this case, the buyer pays $\gamma(c_1) = \min_i \gamma(c_i)$.

4.6 Conclusion

In this chapter, we have examined a combined auction-bargaining model in a procurement setting, where the buyer procures an indivisible item from one of many competing sellers. The model consists of two phases: in the auction phase, the buyer selects the seller, and in the bargaining phase, the final price is determined. The winning bid in the auction phase serves as an outside option for the buyer in the bargaining phase. As a result, each seller's bidding strategy in the auction phase strikes a balance between increasing his probability of winning and increasing the final price in the case that he wins. In this chapter, we take the perspective of the expected profit maximizing buyer, and we allow the buyer to set a reserve price in the first phase.

For our model, we find a symmetric first-phase bidding strategy for the sellers which is simple and intuitive to understand. We show that the combined auction-bargaining model produces a higher expected profit to the buyer than the standard auction or sequential bargaining models. The buyer's expected profit can be improved by setting an appropriate reserve price in the auction phase, and in many cases, the optimal reserve price can easily be computed from the first order condition. We also show that the buyer prefers conducting the auction phase using a second price rather than a first price auction. Our results are illustrated using an example with uniformly distributed costs.

We believe that there are several interesting extensions that can be addressed in the framework of the auction-bargaining model proposed above. For example, when the sellers are asymmetric, both in terms of the distribution of cost and the bargaining power, it would be interesting to investigate what kind of sellers would benefit from the auction-bargaining model as opposed to the standard auction-only or bargaining models.

Chapter 5

Impact of Transfer Pricing Methods for Tax Purposes on Operational Decisions of a Multinational Firm

5.1 Introduction

Transfer pricing refers to the pricing of an intra-firm transaction of an intermediate product or service between two divisions[1] of a firm (Feinschreiber, 2004). Because it has a significant impact on how the division performances are evaluated, the choice of the transfer pricing method influences the decisions of the divisions, and consequently affects the magnitude of the well-known "double marginalization" effect (Spengler, 1950). The loss of optimality due to double marginalization may in theory be mitigated by carefully designing an appropriate payment scheme between the divisions within the firm, and as a result, incentive alignment and supply chain coordination have been a subject of interest in both the management accounting and the supply chain coordination literatures.

In practice, for a multinational firm, transfer pricing is closely related not only to perfor-

[1]Or affiliates, subsidiaries, or departments.

mance evaluations but also to tax reporting. If the firm is subject to several tax jurisdictions, it is profitable for the firm to shift the majority, if not all, of its profits to the jurisdiction with the lowest tax rate. To prevent this abuse, tax authorities use a set of pre-specified transfer pricing methods that are intended to ensure reasonable price allocation among the divisions (e.g. Section 482 of Internal Revenue Service (IRS) Tax Code). Such regulations provide some flexibility to the firm's tax reporting practice, and the particular choice of the transfer pricing method can have a significant impact on the before and after-tax profits of the divisions and of the entire firm (Horst, 1971; Copithorne, 1971).

The set of transfer pricing methods approved by tax authorities typically does not eliminate the double marginalization effect, and as a result, the maximum firm-wide after-tax profit may not be achieved (See, e.g. Halperin and Srinidhi (1987)). To induce the best firm-wide decisions from its divisions, one possible solution for the firm is to maintain two separate accounting systems – one for performance evaluation and the other for tax reporting. Such practice, however, is not always used; in fact, it has been reported that over 80 percent of multinational firms based in the U.S. choose the same accounting system for both performance evaluation and tax reporting (Czechowicz et al., 1982; Nielsen et al., 2008). This phenomenon occurs not only because of the cost and effort associated with maintaining two accounting systems, but also due to the possible tax disputes with the regulatory authorities (Halperin and Srinidhi, 1991; Granfield, 1995; Baldenius et al., 2004). Tax disputes are not uncommon as an Ernst & Young survey reports that five out of six companies have experienced transfer pricing issues with tax authorities, 49 percent of which are on-going (Johnston, 1995). Based on this, we examine the case where a common transfer pricing method serves both performance evaluation and tax reporting purposes.

In this chapter, we study operational decisions of a multinational firm with special attention to transfer pricing methods. We are motivated by the continuing globalization of the world economy where production and consumption increasingly occur in different countries; in fact, one-third of the total exports from the U.S. are attributed to multinational firms (Bernard et al., 2006). We develop a stylized model of a multinational firm consisting of the manufacturing division and the retail division, where the retail division faces price-sensitive stochastic demand and orders an intermediate product from the manufacturing division. We allow the retailer to set the price of the final product, which is a common practice in many industries where the retailer's power has increased (Ailawadi et al., 1999). The manufacturer either accepts or rejects the order. The objective of each division is to maximize its expected profit. We also consider the central planner's problem of maximizing the expected firm-wide after-tax profit as a benchmark. Our modeling of the firm's supply chain is similar to price-setting newsvendor models used in the supply chain management literature, such as Lariviere and Porteus (2001) and Song et al. (2008). The transfer pricing methods that we consider are the cost-plus method and the resale-price method, which are the most commonly-used non-market pricing methods for multinational firms in the U.S. (Benvignati, 1985; Halperin and Srinidhi, 1987), which we describe in Section 5.1.1 below.

5.1.1 The U.S. Tax Regulations for Transfer Pricing

The U.S. transfer pricing methods for tax purposes are based on the *arm's length principle* (defined in the OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations) that the transfer price should be the best estimate of the price as if the two divisions involved were indeed two independent entities, not part of the same firm structure. The details on how to determine the transfer price are specified in detail in "the transfer pricing regulations" (or "the Section 482 Regulations"), which were published by the U. S. Treasury Department on July 1994. Determination of the transfer price is less complicated when an intermediate product has its own market outside of the firm, in which case, the arm's length price is the market price. This method is referred to as the *comparable uncontrolled price* (CUP) method (Treasury Regulation Section 1.482-3(b)).

An intermediate product, however, is often specific to the firm and not sold outside of the firm, in which case, the market price does not exist and the CUP method is inapplicable. The price is then determined by considering the cost and profit structure of the "most similar product" available in the market. Benvignati (1985) reports that only approximately one-quarter of the multinational firms in the U.S. use the CUP method. More recently, Tang (2002) and Ernst&Young (2008) report that the non-market transfer pricing methods are commonly used by multinational firms that conduct international transfers of tangible products. Two of the most common non-market transfer pricing methods are the *cost-plus* method (Treasury Regulation Section 1.482-3(d)) and the *resale-price* method (Treasury Regulation Section 1.482-3(c)), which are described below.

- COST-PLUS (CP) METHOD. Under the cost-plus method, the transfer price of the intermediate product is determined by multiplying manufacturing cost by a fixed constant. This markup is based on the gross profit percentage of the sales made by the manufacturing division for the *most similar product* to the market outside the firm. This method is appropriate if there exists a similar product that the manufacturing division produces and that is sold outside the firm such that reliable information about its profit margin can be obtained.

- RESALE-PRICE (RP) METHOD. Under the resale-price method, the transfer price is calculated based on the price of the final product sold to the customers, by marking down the percentage earned by the retail division for the *most similar product* purchased from the market. Thus, the resale-price method is appropriate if the retail division procures a similar product from the market such that its profit margin can be obtained.

Whereas the decision regarding the use of the CUP method is dictated by the existence of the market for an intermediate product outside of the firm, the choice between the cost-plus and resale-price methods is not always clear. A firm typically searches for the most similar product that is either sold by the manufacturing division or purchased by the retail division, in order to apply an appropriate non-market transfer pricing method. In practice, however, not only is it difficult to identify the most similar product in the market, but the reliability and accuracy of its transaction data can be questionable. When the most similar product is not very close to the original product (for example, in terms of its function, type and market geography), the profit percentage needs to be adjusted accordingly. See Feinschreiber (2004) and Eden (1998) for details. As a result of these ambiguities, the choice of the transfer pricing method and the markup or markdown parameter is often at the discretion of the firm (although its decisions need to be justified and are subject to investigation by the IRS). Therefore, understanding the impact of the transfer pricing method on operational decisions as well as on the profits of the firm and its division can provide valuable insight to the firm.

5.1.2 Literature Review

TRANSFER PRICING. The performance of a firm under a decentralized decision-making environment has been well addressed in the accounting and economics literature starting from the classical paper of Hirshleifer (1956). Each division makes decisions to optimize its own objective, which often leads to suboptimal outcomes for the firm. The decision of a division depends on how the intra-firm transactions are priced, and several transfer pricing schemes have been proposed and analyzed. See, for example, Kanodia (1979), Ronen and McKinney (1970), Ronen and Balachandran (1988), Yeom et al. (2000) and Baldenius and Reichelstein (2006). The focus of these papers is the coordination of divisions in order to achieve the firm-wide optimal decision, and it is not the compliance of the transfer pricing schemes to tax regulations.

There are several papers that consider the tax rate differential and the transfer pricing constraints imposed by tax authorities. Among these papers, the exact modeling of the transfer price and the minutiae of detail vary significantly, reflecting the presence of intrinsic ambiguity and flexibility in the transfer pricing regulations and also in practice. In Horst (1971) and Copithorne (1971), the transfer price is optimized over an exogenously specified range, and this decision is decoupled with other decisions of the firm; consequently, the firm's optimal decision shifts as much profit as possible to the tax jurisdiction with the lower tax rate, subject to satisfying given constraints. In subsequent papers, the selection of the transfer price is coupled with other decisions of the firm. In Samuelson (1982), an upper bound imposed on the transfer price is given by the resale price, an endogenous decision of the firm (here, a lower bound is given by the production cost, an exogenous parameter). The transfer price in Eden (1983) depends on the firm's order quantity decision and its dependence is given by a deterministic function, called the customs valuation method. As in this chapter, Halperin and Srinidhi (1987) study the cost-plus and resale-price methods for transfer pricing and demonstrate how the firm's pricing decision on the final product as well as its "most similar product" is distorted from the optimal decision when there is no regulation on transfer pricing. All the papers mentioned thus far in this paragraph share two commonalities: deterministic demand modeling and a single decision maker (centralized problem). Halperin and Srinidhi (1991) extend their earlier model to a decentralized, where coordination can possibly be achieved through profit sharing.

In this chapter, we study the impact of transfer pricing methods for tax purposes in a firm's pricing and production decisions, both in centralized and decentralized settings. Since Halperin and Srinidhi (1987) are the closest to this chapter, we highlight the differences between the two papers. First, Halperin and Srinidhi (1987) assume that the gross profit margin is set to that of the most similar product the firm produces. They are concerned about distortion of the production of the most similar product under different transfer pricing rules. We, however, assume that such profit margin comes from industry average and are more focused on the production and pricing decision of the product directly. Second, we incorporate into our model a newsvendor-type supply chain framework, which is well established in the operations management literature for studying quantity decisions under the random demand, whereas Halperin and Srinidhi (1987) only consider production decision under the deterministic demand. Lastly, while the main interest of Halperin and Srinidhi (1987) is to show the fact that the current transfer pricing regulation causes distortion in the pricing decision, our interest is in comparing the effectiveness and impact of the two commonly-used transfer pricing methods, in order to guide managers who can exercise discretion in selecting one of these two methods.

DECENTRALIZED SUPPLY CHAIN: WITHOUT RETAIL SALES PRICING. In the absence of tax issues, the performance of decentralized supply chains consisting of two divisions (manufacturer and retailer) has been well studied both in the accounting literature (e.g., Hirshleifer (1956)) and the operations management literature (e.g. Cachon (2003)). In the majority of the papers, each division maximizes the expected profit of its division, and the retailer makes an ordering decision but not a resale price decision. The most common form of the transfer payment by the retailer to the manufacturer is the *price-only contract*, where the payment is proportional to the order quantity. It is well known that the central optimal solution (referred to as coordination) cannot generally be achieved by using the price-only contract unless the manufacturer earns zero profit. Several papers further analyze the price-only contract. In Lariviere and Porteus (2001) and Li and Atkins (2002, 2005), the manufacturer is a Stackelberg leader who sets the transfer price, and the retailer is a follower who determines the order quantity. These models contrast with our models, where the transfer price is either set exogenously (cost-plus method) or implicitly determined by the retailer's decisions (resale-plus method). Recently, Perakis and Roels (2007) characterize the loss of optimality due to decentralization under the price-only contract.

While the loss of optimality is unavoidable in the price-only contract that has a single parameter, coordination becomes possible through one of several multi-parameter contracts, which include the *returns* or *buyback* (Pasternack, 1985; Emmons and Gilbert, 1998), *revenue sharing* (Cachon and Lariviere, 2001; Koulamas, 2006), *rebates* (Taylor, 2002; Chen et al., 2007), and *quantity discount* (Weng, 1995; Khouja, 1996). In all of these contracts, the total transfer payment by the retailer depends not only on the order quantity but also on the sales quantity, except for the quantity discount contract.

PRICE-SETTING NEWSVENDOR FRAMEWORK. There has been a growing interest to model simultaneous decisions in resale pricing and ordering. By incorporating external demand that is price-sensitive, the price-setting extensions of the classical newsvendor problem have permeated the operations management literature, and they are reviewed in Petruzzi and Dada (1999) and Yano and Gilbert (2003). The model of price-sensitive demand has also been included in the supply chain literature with multiple decision-makers, consisting of the manufacturer and the retailer. For example, Li and Atkins (2002, 2005) study the price-only contract, which is shown to be incapable of achieving coordination as in the previous exogenous demand case. Our model with the cost-plus method is an example of the price-only contract since the transfer price, though it depends on the production cost of the manufacturing division, is exogenous to the decisions of the divisions and the realization of stochastic demand. Also, we mention that the buyback contract has been studied by Granot and Yin (2005) and Song et al. (2008) in this framework.

The transfer payment under the contracts mentioned above, including our cost-plus method, depends on the retailer's quantity decision but not on her resale price decision. In contrast, the transfer payment under the resale-price method depends on both quantity and pricing decisions – it is a product of the order quantity and a transfer price that is a fraction of the retailer's resale price. The resale-price method can be considered a variant of the *price-discount scheme* (Bernstein and Federgruen, 2005) or *trade promotion* (Ailawadi et al., 1999), in which the transfer price depends on the resale price charged by the retailer. Although the resale-price method and revenue-sharing contract share some similarities, the transfer payment does not depend on the realized demand in the resale-price method while it does depend on it in the revenue sharing contract.

All of the aforementioned papers in the operations management literature do not consider the tax impact of the transfer pricing method. Shunko and Gavirneni (2007) and Shunko et al. (2008) are the only papers, to our knowledge, that incorporate the tax rate differential in their modeling of a two-stage supply chain similar to our own. These models allow some leeway in determining the transfer price as opposed to following a specific transfer pricing method, and this decision is made by the firm before the retailer's decisions are made. In Shunko and Gavirneni (2007), computational results show the importance of considering the randomness in demand in setting the transfer pricing decision. Motivated by this, we consider a supply chain model with stochastic demand. In Shunko et al. (2008), they consider outsourcing decisions in a model where demand is deterministic while the cost is stochastic. In contrast, we explicitly account for the specific transfer pricing methods (the cost-plus and the resale-price methods) in the decentralized supply chain within a price-setting newsvendor framework, allowing for the possibility that the transfer price may depend on the retailer's decision. We study the flexibility of the firm in determining the transfer prices in a more controlled manner – through both the choice of the transfer pricing methods and the sensitivity analysis of the markup or markdown parameter for each of these methods. Our interest focuses on how the structural form of the transfer pricing methods for tax purposes impact the firm's performance.

OPERATIONS MANAGEMENT AND FINANCE INTERFACE. This chapter rests on the interface between operations management and marketing (through the price-setting newsvendor model), and also belongs to the broad area of the operations management and finance interface, through the modeling of the interplay between supply chain management and tax planning. This interface between operations management and finance has drawn much attention recently, and while it is not our intention to review papers in this area, we refer interested readers to Birge (2000), Babich (2007), Caldentey and Haugh (2009), Federgruen and Yang (2008), and the references therein. Most papers in this area study financing and risk management, and the application of operations management to the area of accounting and tax has been limited.

5.1.3 Contribution and Organization

In this chapter, we address the effect of transfer pricing on supply chain performance under demand uncertainty. We specifically examine two transfer pricing methods for tax purposes that are sanctioned by tax authorities and widely used in practice. While these methods have previously been studied by Halperin and Srinidhi (1987, 1991) who focus on the existence of the distortion in pricing the most similar product, our main interest is in the direct comparison of the two transfer pricing methods. Also, while their papers are based on the deterministic demand, we model the uncertainty in demand, and thus the ordering decision explicitly incorporates the inventory risk of overage and underage.

We model the transfer pricing methods with the variants of the price-setting newsvendor framework that is well established in the operations management literature, such that our results can easily be understood in the context of previous research streams in supply chain management. Our analysis shows that the pricing and inventory decision under the cost-plus method reduces to the price-setting newsvendor with the suitably modified cost parameter. Also, analysis of the resale-price method is a modified version of the price-setting newsvendor problem, where the modification affects the revenue component of the model. While the analysis of the model under the cost-plus follows relatively easily from the previous price-setting newsvendor results, the analysis of the model under the resale-pricing method needs further work. We show that the profit optimization problem under the resale-pricing method is quasi-concave problem where the optimal price and quantity is obtained uniquely. This approach enables us to find, for each of the two transfer pricing methods, several structural properties for the firm's decisions and their profit outcomes. We show several sensitivity results; for example, as the parameter of each transfer pricing method changes, the total expected firm-wide profit (before-tax) is monotone in the cost-plus method; however, it is shown numerically that this profit is concave in the resale-price method.

We perform a comparison of the two transfer pricing methods to address the question of which method would be preferable to each division and to the firm's central planner. We present some evidence supporting that the resale-price method tends to generates a higher firm-wide before-tax profit than the cost-plus method. Numerical results show that the cost-plus method tends to allocate a higher percentage of profit to the retail division. This result is consistent with the result of Eden (1998) based on the notion of transfer pricing continuum, which is applicable to the case where the resale price is however fixed. We, furthermore, consider the impact of the difference in the tax rates between the divisions. We show how tax differential may change the structural results in our model. Such findings will provide insights to tax planners of multinational firms interested in effectively designing and managing global supply chains.

The remainder of this chapter is organized as follows. In Section 5.2, we present the model by introducing appropriate notations. In Section 5.3, we analyze the cost-plus method and the resale-price method individually and then in comparison. We conclude in Section 5.4.

5.2 Model

We consider a firm consisting of two divisions subject to separate tax jurisdictions. The upstream division produces the intermediate good and ships it to the downstream division, which faces price-sensitive stochastic customer demand. The focus of this chapter is to investigate the impact of the transfer payment between these two divisions when the form of payment follows one of several transfer pricing methods commonly used in the international tax context, and study the behavior and performance of the decentralized global supply chain.

In our model, the downstream division (hereafter, *retailer*) determines the retailer's selling price p and the order quantity q of the product. Then, the upstream division (hereafter, *manufacturer*) decides whether or not to accept the retailer's order, based on the order quantity and the pre-specified transfer payment scheme. We emphasize that the details of the payment scheme are exogenously given, in compliance with transfer pricing regulations.

If the manufacturer accepts the order, he produces the order quantity and ships to the retailer. Then, the stochastic demand $D(p)$ arrives at the retailer, and the sales quantity is $\min\{D(p), q\}$. For simplicity, we assume that any demand in excess of q units is lost without any additional penalty, and that any excess inventory is scrapped at no additional cost or salvage value.

We assume that the retail demand function is given by $D(p) = D(p, \epsilon) = y(p) \cdot \epsilon$, which is composed of the deterministic function $y(p)$ and the nonnegative random variable ϵ. The multiplicative demand function is one of the most widely accepted models for the price-sensitive demand; see, e.g., Petruzzi and Dada (1999), Monahan et al. (2004), Granot and Yin (2005), Wang et al. (2004), Wang (2006) and Song et al. (2008). (The other commonly used model is the additive from of demand given by $D(p) = y(p) + \epsilon$; however, we use the multiplicative form since it exhibits a desirable property that the magnitude of uncertainty increases as the expected demand increases.) It has been documented that the multiplicative form appears to be more tractable than the additive counterpart. We suppose that the stochastic variable ϵ is nonnegative, and has the support of $[\underline{d}, \overline{d}]$, where $0 \leq \underline{d} \leq \overline{d} \leq \infty$. Let $\mu = \mathbb{E}[\epsilon] > 0$. We assume that ϵ has the increasing generalized failure rate (IGFR) property where a function $r(z)$ defined by

$$r(z) = \frac{f(z)z}{1 - F(z)} \tag{5.1}$$

is increasing in z. It is a mild assumption that is satisfied by many distributions such as uniform, exponential, normal and some Weibull distributions. (See Lariviere (2006) for the detail.) For the deterministic component, we assume $y(p) = ap^{-b}$ where $a > 0$ and $b > 1$. Such a negative power demand function has been commonly used both in the price-setting newsvendor literature and in the economics literature, including the references mentioned above. Note that b indicates the price elasticity of demand. If $b > 1$, then the demand is price elastic, whereas $b < 1$ implies that it is price inelastic. Just as Wang et al. (2004) and Wang (2006), we focus on the case where $b > 1$ (otherwise, the optimal price could go to infinity).

For the cost model, let c^M be the per-unit production cost of the manufacturer. We assume that the payment from the retailer to the manufacturer is proportional to the order quantity q, and we denote the per-unit transfer price by w. We note that this price is charged to the retailer on the order quantity q, irrespective of the actual quantity she has sold. In determining the transfer price w, we consider two transfer pricing methods that are commonly used for U.S. tax reporting purposes. (i) Under the cost-plus (CP) method, the transfer price is determined by a percentage markup of the manufacturer's cost, i.e., $w = \gamma \cdot c^M$, where $\gamma \geq 1$ denotes the *markup rate* of the manufacturer. The value of an appropriate markup γ is determined based on the profits of other comparable companies in the industry. (ii) Under the resale-price (RP) method, the transfer price is a percentage markdown of the retailer's selling price, i.e., $w = \beta p$, where $\beta \in [0,1]$ is the *markdown rate* of the retailer, and is once again determined based on similar comparable transactions. Observe that while w is independent of the decision vector (p, q) in the cost-plus method, it depends on p in the resale-price method. In this chapter, we treat both γ and β as exogenous constants. We let τ^M denote one minus the effective tax rate of the manufacturer, such that the manufacturer's after-tax profit is τ^M times his pre-tax profit. Similarly, we define τ^R for the retailer.

The manufacturer's pre-tax profit is given by

$$\Pi^M(q) \quad = \quad (w - c^M) \cdot q \ .$$

For the retailer, the ordering cost includes not only the payment to the manufacturer, but also additional costs due to transportation, inspection and possibly the purchase of raw materials. We denote this additional per-unit ordering cost by c^R, and thus the retailer incurs $w + c^R$ per unit of the product ordered from the manufacturer. The revenue to the retailer is given by $p \cdot \min\{D(p), q\}$. Therefore, the retailer's expected pre-tax profit satisfies

$$\Pi^R(q,p) \quad = \quad -(c^R + w) \cdot q + p \cdot \mathbb{E}[\min\{D(p), q\}] \ .$$

In our model, the retailer bears the inventory risk of unsold units, and her decision balances the tradeoff between the possibility of losing sales opportunities and incurring ordering costs for unsold products. The after-tax profits of the manufacturer and the retailer are $\tau^M \Pi^M(q)$ and $\tau^R \Pi^R(q,p)$, respectively.

We compare the decentralized model to the centralized system, in which the central planner's problem is to maximize the firm's total after-tax profit:

$$\Pi^C(q,p) \quad = \quad \tau^M \Pi^M(q) + \tau^R \Pi^R(q,p) \ .$$

(The above objective function assumes that the firm does not repatriate profit from one tax jurisdiction to another, and rather use it for reinvestment and expansion within the division where the profit is earned. Such a practice is consistent with the evidence provided by Shunko and Gavirneni (2007), and this assumption has also been adopted by Halperin and Srinidhi (1987, 1991)). We impose a constraint that each of the two divisions must make nonnegative profit (to avoid manipulating the transfer price to shift an unlimited amount of profit from the high-tax regime to the low-tax regime), i.e., $\Pi^R(q,p) \geq 0$ and $\Pi^M(q,p) \geq 0$. Let $c = c^M + c^R$. If the effective tax rates for the manufacturer and for the retailer are the same (i.e., $\tau^M = \tau^R$), then it can be shown that Π^C equals τ^M $(= \tau^R)$ multiplied by the following benchmark quantity:

$$\Pi^B(q,p) \quad = \quad -c \cdot q \; + \; p \cdot \mathbb{E}[\min\{D(p), q\}] \; , \tag{5.2}$$

which is an objective function of the well-known price-setting newsvendor problem due to Petruzzi and Dada (1999). (Here, the superscript B stands for the benchmark system without considering taxes.) Otherwise, if the tax rates differ and the transfer price w is *endogenous* (i.e., the central planner can arbitrarily choose a price for its intra-firm transaction), then the optimal value of w will be chosen such that all the profits of the firm will be shifted to the division with the lower tax rate (see Samuelson (1982) for a similar argument). Furthermore, in this case, it can be shown that the optimal value of (p,q) will again be the solution to the problem of maximizing (5.2), and that the optimal value of Π^C is $\max\{\tau^M, \tau^R\}$ times the optimal value of $\Pi^B(q,p)$ in (5.2). We can also view the benchmark quantity as the total before-tax profit of a firm. For the remainder of this chapter, we assume that the choice of the transfer price is *not* arbitrarily endogenous and must follow either the cost-plus method or the resale-price method.

We summarize the relationship among the optimal expected profits in the following proposition. Let (p^R, q^R), (p^C, q^C) and (p^B, q^B) be the optimal solutions maximizing Π^R, Π^C and Π^B, respectively.

Proposition 5.2.1.

(a) $\max\{\tau^M, \tau^R\} \cdot \Pi^B(q,p) \;\geq\; \Pi^C(q,p) \;=\; \tau^M \Pi^M(q) + \tau^R \Pi^R(q,p)$ *for any* (q,p).

(b) $\max\{\tau^M, \tau^R\} \cdot \Pi^B(q^B,p^B) \;\geq\; \Pi^C(q^C,p^C) \;\geq\; \tau^M \Pi^M(q^R) + \tau^R \Pi^R(q^R,p^R)$.

The proof of this proposition is straightforward and is omitted. Part (b) of Proposition 5.2.1 shows that the gap in the performance between the decentralized system and the benchmark system consists of the two components. The first inequality in the left hand side occurs since the intra-firm payment cannot be made arbitrarily and must comply to transfer pricing regulations for tax purposes. We refer to this as the *tax regulation* effect, which is given by $\max\{\tau^M, \tau^R\} \cdot \Pi^B(q,p) - \Pi^C(q^C, p^C)$. The second inequality, given by $\Pi^C(q^C, p^C) - \tau^M \Pi^M(q^R) + \tau^R \Pi^R(q^R, p^R)$, comes from the *double marginalization* effect which is due to the fact that the retailer maximizes her own profit as opposed to the overall firm profit.

5.3 Analysis

In this section, we analyze our model. Since demand is a random function of price, the retailer makes both pricing and inventory decisions by striking a balance not only between higher price and more sales, but also between excess inventory and lost sales. When the retailer sets price p and order quantity q, the retailer pays the manufacturer $w \cdot q$ where w is the transfer price, set by one of the transfer pricing methods. The retailer's revenue is $p \cdot \min\{D(p), q\}$.

It is convenient to define the *stocking factor*

$$z = q/y(p) ,$$

and consider the retailer's decision with respect to (z, p) instead of (q, p). Such a substitution has been used in the operations management literature by Petruzzi and Dada (1999), Li and Atkins (2002) and Wang et al. (2004). Also define the overage and underage functions by

$$\Lambda(z) = \mathbb{E}[(z - \epsilon)^+] \qquad \text{and} \qquad \Theta(z) = \mathbb{E}[(\epsilon - z)^+] .$$

Then, it follows that

$$\mu - \Theta(z) = z - \Lambda(z) . \tag{5.3}$$

In Section 5.3.1, we first study the benchmark system without considering tax rate differential. Then, we analyze the cost-plus method and the resale-price method in Sections

5.3.2 and 5.3.3, respectively, and we compare the cost-plus and resale-price methods in Section 5.3.4. We examine the effect of demand variability in Section 5.3.5.

5.3.1 Benchmark System

We consider the *benchmark system*, where the objective is to maximize the total expected profit of the supply chain without considering taxes. This problem is formulated as the price-setting newsvendor problem, well-known in the operations management literature. We review some known results for this problem, and establish new properties. The benchmark system is useful in analyzing the optimal decisions of the decentralized system as well as in comparing its performance to the decentralized system.

The expected profit of the benchmark system with price p and stocking factor z with cost c satisfies

$$\begin{aligned} \Pi^B(z,p;c) &= p \cdot \mathbb{E}[\min\{q, D(p)\}] - c \cdot q \\ &= y(p)\left[(\mu - \Theta(z))p - cz\right] . \end{aligned} \tag{5.4}$$

Recall $c = c^M + c^R$. We denote by z^B the optimal stocking factor, and let $q^B = q^B(z^B)$ and $p^B = p^B(z^B)$ be the optimal price and quantity, respectively. The corresponding optimal expected profit is denoted by $\Pi^{B*} = \Pi^B(z^B, p^B(z^B); c)$. The problem of maximizing (5.4) is a price-setting newsvendor problem, for which we establish structural properties in Lemma 5.3.1 below. Let $\perp\!\!\!\perp$ denote independence, and let $\propto$ denote proportionality.

Lemma 5.3.1. *In the benchmark system,*

(a) The optimal price for a given z is given by

$$p^B(z) = \frac{bc}{b-1} \cdot \frac{z}{\mu - \Theta(z)} .$$

(b) $\Pi^B(z, p^B(z); c)$ is equal to $\frac{c}{b-1} \cdot q^B(z)$, which is quasi-concave in z. Also,

$$z^B \perp\!\!\!\perp c , \qquad p^B \propto c , \qquad q^B \propto c^{-b} \qquad \text{and} \qquad \Pi^{B*} \propto c^{-(b-1)} .$$

Part (a) is a standard result from Petruzzi and Dada (1999), which is a consequence of the particular choice of the price-dependent demand function. In part (b), the result that $z^B \perp\!\!\!\perp c$ is due to Wang et al. (2004). The stocking factor z^B is independent of the cost c because the optimal price in (a) is also linear in c, thus canceling its effect. They have also shown an intuitive result that p^B is increasing in c, but have not established this relationship is linear. The results regarding the quantity q^B and the profit Π^{B*}, that both of these quantities are convex decreasing in the cost c, are not surprising, but they have not yet appeared in the literature, to the best of our knowledge. We believe that these results shed additional light to the classical price-setting newsvendor problem.

5.3.2 Cost-Plus Method

In the cost-plus method, the transfer price is set to a markup rate times the manufacturer's cost, i.e., $w = \gamma \cdot c^M$ where $\gamma \geq 1$. In the decentralized system, the decisions of both price and order quantity are determined by the retailer, and the manufacturer decides whether or not to accept the retailer's order. Any transfer price $w \geq c^M$ results in a non-negative profit for the manufacturer, and as a result, he accepts the order provided that the parameter γ is at least 1. In the centralized system, where the central planner makes both pricing and quantity decisions, the nonnegativity requirement for the manufacturer's profit is automatically satisfied provided that $\gamma \geq 1$. We first consider the decentralized system, and then compare it to the centralized system. The analysis of this section is based on the reformulation of the given problems as price-setting newsvendor problems, and applying the properties that have now been established for such problems (Lemma 5.3.1).

DECENTRALIZED SYSTEM.

Consider the retailer's optimization problem in the decentralized system. Define

$$\hat{c}(\gamma) \quad = \quad c^R + \gamma c^M \; . \tag{5.5}$$

This represents the effective per-unit cost to the retailer. Under the cost-plus method with the transfer price set at $w = \gamma c^M$, the retailer's expected profit is

$$\begin{aligned} \Pi^R(z,p) &= p \cdot \mathbb{E}[\min\{D(p), q\}] - \hat{c}(\gamma) \cdot q \\ &= \Pi^B(z, p; \hat{c}(\gamma)) \end{aligned} \tag{5.6}$$

Note that this expression is the same as $\Pi^B(z, p; c)$ in (5.4) except that c is replaced by $\hat{c}(\gamma)$. Thus, the retailer's problem is also a price-setting newsvendor problem. As discussed in Section 5.3.1, the retailer's optimal price p^R for a given stocking factor z is given by

$$p^R(z) \quad = \quad \frac{b\hat{c}(\gamma)}{b-1} \cdot \frac{z}{\mu - \Theta(z)} \ ,$$

and $\Pi^R(z, p^R(z))$ has the unique maximizer z^R which satisfies the first-order condition $d\Pi^R(z, p^R(z))/dz = 0$.

The following lemma summarizes the sensitivity of the optimal price, quantity and the expected profit on the markup rate γ. Increasing γ results in an increase in the transfer price from the retailer to the manufacturer. Let $p^R = p^R(z^R)$ and $q^R = q^R(z^R)$ denote the retailer's optimal price and order quantity. We denote the expected profit of the retailer and the corresponding expected profit of the manufacturer when the retailer chooses her optimal price and quantity with $z = z^R$, by $\Pi^{R*} \equiv \Pi^R(z^R, p^R)$ and $\Pi^M(z^R, p^R)$, respectively.

Lemma 5.3.2. *In the decentralized system under the cost-plus method, $\Pi^R(z, p^R(z))$ is quasi-concave in z. Also,*

$$z^R \perp\!\!\!\perp \hat{c}(\gamma) \ , \qquad p^R \propto \hat{c}(\gamma) \ , \qquad q^R \propto \hat{c}(\gamma)^{-b} \qquad \textit{and} \qquad \Pi^{R*} \propto \hat{c}(\gamma)^{-(b-1)} \ .$$

Furthermore, $z^R = z^B$.

The above lemma shows that the retailer's profit $\Pi^R(z, p^R(z))$ initially increases in the stocking factor because of the increased sales and revenue, but eventually decreases because of the overage cost. Thus, the optimal choice of the stocking factor z^R can be found using the first-order condition. Since $\hat{c}(\gamma) = c^R + \gamma c^M$, the lemma implies that z^R is also independent of the markup rate γ. This result can again be explained by the fact that the retailer's price p^R is linearly increasing in the retailer's effective cost $\hat{c}(\gamma)$ (and also in γ). We also obtain that as the markup rate γ increases, the retailer's cost for each unit increases, resulting in lower quantity q^R. It is expected that the retailer's profit Π^{R*} decreases in the markup rate γ.

The manufacturer's expected profit is expressed as

$$\Pi^M(z^R, p^R) \quad = \quad q(z^R)(w - c^M) \quad = \quad y(p^R) \cdot z^R \cdot (\gamma - 1) \cdot c^M \ . \tag{5.7}$$

While Lemma 5.3.2 has shown that the retailer's optimal expected profit Π^{R*} is decreasing in γ, the following lemma shows that the corresponding manufacturer's expected profit

$\Pi^M(z^R, p^R)$ is unimodal in γ. Increasing in γ initially increases the manufacturer's profit as it increases the margin of each unit, but it eventually decreases his profit because the retailer is too squeezed to order sufficient quantities. However, Lemma 5.3.3 (b) shows that the ratio $\Pi^{R*}/\Pi^M(z^R, p^R)$ is decreasing in γ, which means that the retailer's share of the overall profit also decreases as the manufacturer increases its markup rate to the retailer.

Lemma 5.3.3. *In the decentralized system under the cost-plus method,*

(a) $\Pi^M(z^R, p^R)$ *is quasi-concave with respect to* γ *achieving its maximum at* γ^M *where*

$$\gamma^M = \frac{c^R/c^M + b}{(b-1)}, \tag{5.8}$$

and $\gamma^M > 1$.

(b) The ratio of the retailer's optimal expected profit to the manufacturer's corresponding expected profit satisfies

$$\frac{\Pi^{R*}}{\Pi^M(z^R, p^R)} = \frac{\gamma + c^R/c^M}{\gamma - 1} \cdot \frac{1}{b-1}.$$

which is decreasing in γ.

(c) With respect to γ $(\gamma \geq 1)$, *the total after-tax profit,* $\tau^R\Pi^{R*} + \tau^M\Pi^M(z^R, p^R)$, *is either decreasing (if* $\tau^R \geq \tau^M$*) or quasi-concave (if* $\tau^R < \tau^M$*) that is maximized at* γ^C *where*

$$\gamma^C = \frac{b + (1 - \tau^R/\tau^M)c^R/c^M}{(b-1) + \tau^R/\tau^M}.$$

Although γ is not a decision variable in our model, the analysis in Lemma 5.3.3 provides insights for the retailer or central planner's choice between the two methods. Observe that the manufacturer's optimal markup rate γ^M given in part (a) is robust such that it is independent of the randomness in demand ϵ, and depends only on the ratio of costs and the demand elasticity b. In addition, for fixed γ, note that the ratio of retailer's optimal expected profit to the manufacturer's corresponding expected profit is independent of the randomness in demand ϵ. These result comes from the fact that the stocking factor z^R is independent of ϵ in the multiplicative demand model. Lemma 5.3.3 (b) also implies that the retailer's share of the total before-tax profit is bounded above by $1/b$. Thus the retailer are assured to make more than $1/b$ of the total before-tax profit in the cost-plus method. (See, section 5.3.4 for its implication in the comparison of cost-plus and resale-price methods.)

It is intuitive to see that γ^M is decreasing with respect to c^R/c^M, i.e., the manufacturer makes relatively less margin when its profit is maximized as the cost of the manufacturer becomes relatively higher than the retailer. Furthermore, as for the total after-tax profit, the ratio of the tax rates to the manufacturer and retailer is another important factor that affects the γ^C. It is intuitive to observe that as τ^R/τ^M decreases, γ^C decreases to γ^M, which implies that the manufacturer's profit becomes more important in the total after-tax profit

CENTRALIZED SYSTEM.

We now consider the central planner's problem of maximizing the after-tax total expected profit of the firm's supply chain. Let $\Pi^C(z,p)$ denote the expected profit of the centralized system. Then,

$$\begin{aligned}\Pi^C(z,p) &= \tau^R \cdot \Pi^R(z,p) + \tau^M \cdot \Pi^M(z,p) \\ &= \tau^R\{y(p)\,[(\mu - \Theta(z))p - \hat{c}(\gamma)z]\} \;+\; \tau^M\left\{y(p)z(\gamma-1)c^M\right\} \\ &= \tau^R \Pi^B(z,p;\tilde{c})\ ,\end{aligned}$$

where $\hat{c}(\gamma) = c^R + \gamma c^M$ as defined in (5.5), and

$$\tilde{c}(\gamma) \;=\; c + ((\tau^R - \tau^M)/\tau^R)\cdot(\gamma-1)\cdot c^M\ . \tag{5.9}$$

Note that this is another price-setting newsvendor problem of type (5.4) where the cost parameter is given by $\tilde{c}(\gamma)$. Also, notice that the second term in (5.9) represents the adjustment to the cost parameter accounting for tax differences. If the retailer's effective tax rate is lower than the manufacturer's tax rate (i.e., $\tau^R > \tau^M$), then it is better to accumulate profit at the retailer than at the manufacturer. It is achieved in our model by setting the cost higher than the actual cost (i.e., $\tilde{c}(\gamma) > c$), which induces that the retailer orders those units only if he is more certain to sell them. The cost parameter $\tilde{c}(\gamma)$ is also expressed as

$$\tilde{c}(\gamma) \;=\; \left(1 - \frac{\tau^R}{\tau^M}\right)\hat{c}(\gamma) + \left(\frac{\tau^R}{\tau^M}\right)c\ .$$

It is observed from the above equation that the new cost parameter is weighted average of c and $\hat{c}(\gamma)$; as the effective tax rate of the retailer becomes more greater than the rate of the manufacturer, the new cost parameter becomes closer to $\hat{c}(\gamma)$, implying that the problem is closer to the retailer's profit optimization problem. The central planner's problem is to maximize $\Pi^C(z,p)$ subject to the condition that both the manufacturer and the retailer earn nonnegative profit, i.e., $\Pi^M(z,p) \geq 0$ and $\Pi^R(z,p) \geq 0$. The first constraint is always satisfied from the definition of $\Pi^M(z,p)$ in (5.7) and $\gamma \geq 1$. The second constraint can be shown to be a lower bound constraint on the retail sales price p. The price-setting newsvendor problem with a lower bound constraint is equivalent to a concave-function maximization with a linear inequality constraint, which is not difficult to solve analytically.

Let $p^C(z)$ be the value of p that maximizes $\Pi^C(z,p)$ subject to the constraints. Similarly, let z^C and $p^C = p^C(z^C)$ denote the optimal decisions in the centralized system.

Lemma 5.3.4. *In the centralized system under the cost-plus method,*

$$p^C(z) \;=\; \max\left\{\hat{c}(\gamma),\, \frac{b\tilde{c}(\gamma)}{b-1}\right\} \cdot \frac{z}{\mu - \Theta(z)} \;=\; \begin{cases} \frac{b}{b-1} \cdot \tilde{c}(\gamma) \cdot \frac{z}{\mu-\Theta(z)} & \text{if } \gamma - 1 \leq \frac{c^R + c^M}{(b\tau^M - \tau^R)c^M} \\ \hat{c}(\gamma) \cdot \frac{z}{\mu - \Theta(z)} & \text{if } \gamma - 1 \geq \frac{c^R + c^M}{(b\tau^M - \tau^R)c^M}. \end{cases}$$

Furthermore, $z^C = z^B$.

In Lemma 5.3.4, the first case corresponds to an unconstrained price-setting newsvendor solution of Lemma 5.3.1, and the second case corresponds to a boundary solution. Interestingly, the condition separating these two cases is independent of the distribution of ϵ. Furthermore, if ϵ is deterministic, then the multiplicative factor $z^C/(\mu - \Theta(z^C))$ is 1, simplifying the expression for p^C.

EXAMPLE: UNIFORM $[0,2]$ RANDOM DEMAND.

We consider the case where $\epsilon \sim U[0,2]$, one of the most commonly used examples in the literature – see, for example, Granot and Yin (2005) and Emmons and Gilbert (1998). Thus, $\mu = 1$. Consistent with Chen et al. (2007), Granot and Yin (2005), we set $a = 1$ and $b = 2$; thus, $y(p) = p^{-2}$. Also set $c^R = c^M = 1$. For the tax rates, we consider three cases where (τ^R, τ^M) is either $(1.0, 1.0)$, $(1.0, 0.8)$ or $(0.8, 1.0)$. (The corporate tax rate for the U.S. is from 15% to 35%. Assuming 35% tax rate for the U.S., the tax rate for Thailand (30%) or Indonesia (28%) is around 80% of that for the U.S.) Figure 5.1 corresponds to the

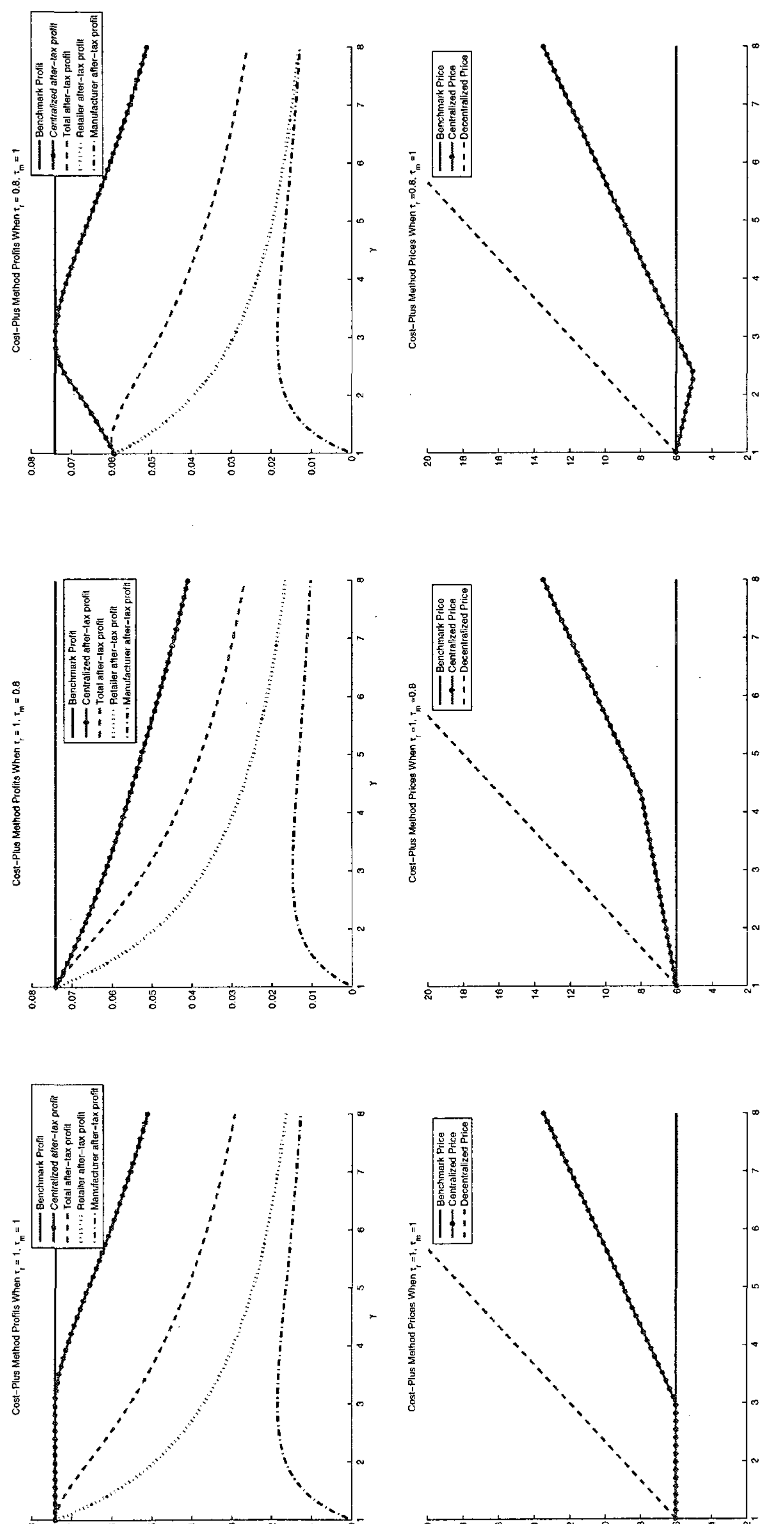


Figure 5.1: Multiplicative random demand model in cost-plus method when $a = 1, b = 2, c^R = c^M = 1$

cost-plus method, where the first row displays the profits as a function of the markup rate γ and the second row displays the price decisions as a function of γ.

In all three cases, the retailer faces the same problem, and the profits of the retailer and the manufacturer also remain the same after being adjusted for tax rates. In the decentralized system, the retailer's profit is decreasing in γ (which affects her purchase cost from the manufacturer), and the manufacturer's profit is unimodal achieving its maximum around $\gamma = 3$. However, the total after-tax profit in the decentralized system, $\tau^R\Pi^{R*} + \tau^M\Pi^M(p^R)$, depends on the tax rates. When the manufacturer's tax rate is at least the retailer's tax rate, i.e., $\tau^M \leq \tau^R$ (the first two columns), we observe that it is decreasing in γ, and is maximized when the manufacturer makes no profit (i.e., $\gamma = 1$). This is due to the fact that either only the *double marginalization* effect exists (in the case of $\tau^R = \tau^M$) here, or the *tax regulation* effect favors increased profit at the retailer's side (in the case of $\tau^R > \tau^M$). In addition, the retailer in the decentralized system selects price p^R within $[p^B, p^C]$. However, in contrast, when the retailer's tax rate is higher, i.e., $\tau^R < \tau^M$ (the third column), the total after-tax profit is quasi-concave and is maximized when γ is slightly yet strictly greater than 1, as consistent with Lemma 5.3.3. In this case, the quasi-concavity of the profit is due to the trade-off between the *double marginalization* effect and *tax regulation* effect which here favors increased profit at the manufacturer's side. In addition, in this case, p^R may be smaller than the benchmark price p^B, which occurs to satisfy the manufacturer's profit feasibility constraint $\Pi^M(p^R) \geq 0$.

Note that the *centralized system* can achieve the profit of the benchmark system for a specific value of γ. This value of γ is exactly 1 if the manufacturer's tax rate is at least the retailer's tax rate (i.e.,$\tau^R \geq \tau^M$); otherwise, the best value of parameter γ is greater than 1, and is obtained by balancing the opposing forces of the tax regulation effect (preferring higher γ) and the double marginalization effect (preferring lower γ).

We discuss the impact of demand uncertainty in Section 5.3.5.

5.3.3 Resale-Price Method

We now study the resale-price method with the random demand model. Under the resale-price method, the transfer price is set to βp, where $\beta \in [0, 1]$ is the exogenously determined markdown rate, and p is the selling price of the retailer. We first analyze the decentralized system where the retailer decides both the pricing and quantity decisions and the manufacturer either accepts or rejects the order, and then consider the problem of the centralized system.

DECENTRALIZED SYSTEM.

In the decentralized system with the transfer price $w = \beta p$, the retailer's revenue is p per unit sold, and her cost is $c^R + \beta p$ per unit ordered. Thus, her expected profit for given (z, p) is

$$\begin{aligned} \Pi^R(z,p) &= p \cdot \mathbb{E}[\min\{D(p), q\}] - (c^R + \beta p) \cdot q \\ &= y(p)\left[(\mu - \Theta(z) - \beta z)p - c^R z\right] = y(p)\left[X(z)p - c^R z\right] , \end{aligned} \tag{5.10}$$

where for notational convenience, $X(z)$ is defined by

$$X(z) = -\beta\Lambda(z) - (1-\beta)\Theta(z) + (1-\beta)\mu = \mu - \Theta(z) - \beta z . \tag{5.11}$$

Notice that (5.10) is similar to (5.4) except that $\mu - \Theta(z)$ term adds $-\beta z$ and c is replaced with c^R.

We note that while the retailer's problem with the cost-plus method (in Section 5.3.2) was transformed to a price-setting newsvendor problem (benchmark system), the above problem with the resale-price method cannot be transformed to a price-setting newsvendor problem. This is due to the fact that the retailer's effective cost $c^R + \beta p$ depends on the price decision p. The manufacturer's profit is given by

$$\Pi^M(z,p) = (\beta p - c^M) \cdot q . \tag{5.12}$$

The retailer's decision must ensure the nonnegativity of $\Pi^M(z,p)$, which is equivalent to

$$p \geq c^M/\beta . \tag{5.13}$$

(Otherwise, the manufacturer would not accept the order.)

We first consider the retailer's pricing decision for fixed z, and show that her profit is quasi-concave in her price. We differentiate Π^R in (5.10) with respect to p. Since $y(p) = ap^{-b}$, we have $y'(p) = -by(p)/p$, it follows

$$\frac{\partial \Pi^R(z,p)}{\partial p} = \frac{y(p)}{p}\Big[-(b-1)X(z)p + c^R bz\Big] . \tag{5.14}$$

From (5.10), we observe that $X(z) \leq 0$ implies that the retailer's profit is always non-positive. Thus, we restrict our attention to the case $X(z) > 0$ for the remainder of this section.

Since the expression in (5.14) is a product of a positive factor and a linear factor, $\partial \Pi^R(z,p)/\partial p$ changes sign at most once from positive to negative as p increases, and thus $\Pi^R(z,p)$ is quasi-concave in p. The *unconstrained* maximizer of $\Pi^R(z,p)$ for fixed z is obtained by setting (5.14) to zero, which results in $\frac{bc^R}{b-1} \cdot \frac{z}{X(z)}$. If this solution is not feasible in (5.13) then the boundary solution $p = c^M/\beta$ is the maximizer. Therefore, we obtain the following result.

Lemma 5.3.5. *In the decentralized system under the resale-price method, we fix z such that $X(z) > 0$. Then, $\Pi^R(z,p)$ is quasi-concave in p, and the optimal solution for $\max_p\{\Pi^R(z,p) \mid p \geq c^M/\beta\}$ is given by*

$$p^R(z) \quad = \quad \max\left\{\frac{bc^R}{b-1} \cdot \frac{z}{X(z)}, \frac{c^M}{\beta}\right\} \quad = \quad \begin{cases} \frac{c^M}{\beta} & \text{if } \beta \leq \frac{b-1}{b} \cdot \frac{c^M}{c^R} \cdot \frac{X(z)}{z} \\ \frac{bc^R}{b-1} \cdot \frac{z}{X(z)} & \text{otherwise} \end{cases}.$$

We comment on the boundary solution $p^R(z) = c^M/\beta$ which occurs when $\beta \leq \frac{b-1}{b} \cdot \frac{c^M}{c^R} \cdot \frac{X(z)}{z}$. This boundary solution results in zero profit for the manufacturer. For the retailer's problem, since $p^R(z)$ is independent of z, the maximization of $\Pi^R(z, p^R(z))$ is a regular (not price-setting) newsvendor problem. Thus, the problem of finding the optimal value of z, denoted by z^R, is easy to solve. Therefore, we proceed by assuming otherwise, i.e., $\beta > \frac{b-1}{b} \cdot \frac{c^M}{c^R} \cdot \frac{X(z)}{z}$.

Substituting $p^R(z) = \frac{bc^R}{b-1} \cdot \frac{z}{X(z)}$ to Π^R and simplifying the expression lead to

$$\Pi^R(z, p^R(z)) \quad = \quad y(p^R(z)) \cdot \frac{c^R z}{b-1} \quad = \quad \frac{c^R}{b-1} \cdot q^R(z) \ , \tag{5.15}$$

where $q^R(z) = y(p^R(z)) \cdot z$. The maximization of $\Pi^R(z, p^R(z))$ is similar to the maximization of the benchmark profit $\Pi^B(z, p^B(z))$ in Lemma 5.3.1, but additional technical difficulty arises here because the linearity of $p^R(z)$ in z no longer holds in this case. However, we can establish that $\Pi^R(z, p^R(z))$ is unimodal in z and has a unique maximizer z^R.

Define $G(z) = [X'(z) \cdot z]/X(z)$. Let z^R denote a maximizer of $\Pi^R(z, p^R(z))$. The next lemma states that the optimal price and stocking factor under the resale-price method can be obtained uniquely. (Recall that $\overline{d}$ is an upper bound on the range of ϵ.)

Theorem 5.3.6. *In the decentralized system under the resale-price method, suppose that* $\beta \geq \frac{b-1}{b} \cdot \frac{c^M}{c^R} \frac{X(z)}{z}$. *Then,* $p^R(z)$ *is increasing in* z, *and* $\Pi^R(z, p^R(z))$ *in* (5.15) *is unimodal in* z. *Furthermore, if there exists* z *such that* $G(z) = (b-1)/b$, *then* $z^R = z$; *otherwise,* $z^R = \overline{d}$.

Now, we address the sensitivity results with respect to the markdown rate β. We recall from Section 5.3.2 that under the cost-plus method, as the markup rate γ increases, the retailer's stocking factor z^R does not change, price p^R increases, quantity q^R decreases, retailer's expected profit Π^{R*} decreases, and finally the retailer's share of the overall profit also decreases (Lemma 5.3.3). Below, we show analogous sensitivity results with respect to the markdown rate β under the resale-plus model, with the difference being that the stocking factor z^R now decreases in β. (Loosely speaking, z^R is related to the newsvendor ratio, which is $(p-w)/p = 1-\beta$. As β increases, $1-\beta$ decreases.) Denote the optimal z^R for given β by z^R_β. While most sensitivity results in the cost-plus method are independent of the randomness of the demand, some of the following results regarding the sensitivity of the optimal solutions with respect to β need a condition that depends on the distribution of ϵ. Recall that $r(z) = (f(z)z)/(1-F(z))$ is the generalized failure rate given in (5.1).

Theorem 5.3.7. *In the decentralized system under the resale-price method, suppose that* $\beta \geq \frac{b-1}{b} \cdot \frac{c^M}{c^R} \frac{X(z)}{z}$ *holds. Also, suppose that the maximizer* z^R *of* $\Pi^R(z, p^R(z))$ *exists and satisfies* $G(z^R) = (b-1)/b$. *Then, (a)* z^R *decreases in* β, *(b)* $\Pi^R(z^R, p^R)$ *decreases in* β, *(c)* q^R *decreases in* β, *and (d) if* $(b-1)r(z^R_\beta) > 1-\beta$, p^R *increases in* β.

Now, we turn our attention to the ratio of expected profits between the retailer and manufacturer. The retailer's profit $\Pi^{R*} = \Pi^R(z^R, p^R(z^R))$ is given by (5.15), and the manufacturer's expected profit from (5.12) is

$$\Pi^M(z^R, p^R) = (\beta \cdot p^R(z^R) - c^M) \cdot y(p^R(z^R)) \cdot z^R .$$

Thus, we have

$$\frac{\Pi^{R*}}{\Pi^M(z^R, p^R)} = \frac{c^R}{(b-1)(\beta p(z^R) - c^M)} ,$$

and this ratio is decreasing in β since $p^R = p(z^R)$ is increasing in β (Theorem 5.3.7), under the condition in the theorem. This result is consistent with the similar monotonicity of the analogous quantity with the cost-plus method (Lemma 5.3.3).

CENTRALIZED SYSTEM.

We now consider the after-tax optimal profit from the perspective of the central planner, and show that this problem can be cast in a form that is similar to the retailer's problem in the decentralized system. The firm's total after-tax profit under the resale-price method, $\Pi^C = \tau^R \Pi^R + \tau^M \Pi^M$, can be expressed as

$$\Pi^C \quad = \quad \tau^R y(p) \left[(\mu - \Theta(z) - \breve{\beta})p - \breve{\beta}\breve{c}z \right] \tag{5.16}$$

where $\breve{\beta} = (1 - \tau^M/\tau^R) \cdot \beta$ and $\breve{c} = c^R + (\tau^M/\tau^R) \cdot c^M$. We remark that if $\tau^R = \tau^M$, then the right-hand-side expression in (5.16) corresponds to the standard price-setting newsvendor problem, whose optimal price and stocking factor are independent of τ^R and τ^M.

We require the constraint that both the manufacturer and the retailer earn nonnegative expected profit, i.e., $\Pi^M \geq 0$ and $\Pi^R \geq 0$. The first constraint $\Pi^M \geq 0$ is equivalent to $p \geq c^M/\beta$. The second constraint $\Pi^R \geq 0$ is not guaranteed by a simple inequality such as $p \geq c^R/(1-\beta)$ because of the excess inventory risk assumed by the retailer, and this constraint is equivalent to $X(z)p - c^R z \geq 0$ from (5.10).

Note that this expression (5.16) is the same form as the retailer's objective function (5.10) where β and c are replaced by $\breve{\beta}$ and $\breve{c}$, respectively. Therefore, we can obtain the optimal pair of z and p in a similar manner as the retailer's problem in the decentralized system. Similar to the definition of X in (5.11), we define $\breve{X}(z) = \mu - \Theta(z) - \breve{\beta}z$. Then, similar to (5.10), we obtain that $\breve{\Pi}^C(z,p)/\tau^R = y(p) \cdot [\breve{X}(z)p - \breve{c}z]$. This expression is quasi-concave in p for fixed z. Furthermore, the two constraints $\Pi^M \geq 0$ and $\Pi^R \geq 0$ impose lower bounds on p, i.e., $p \geq c^M/\beta$ and $p \geq c^R z/X(z)$. Thus, finding the optimal p for given z is an easy problem, and the optimal value of $p^C(z)$ is either the unconstrained minimizer $\frac{b}{b-1}\frac{\breve{c}\cdot z}{\breve{X}(z)}$ or the maximum of the lower bounds. (The result follows the argument of Lemma 5.3.5.) Now, the problem of finding the optimal z^C maximizing $\breve{\Pi}^C(z, p^C(z))$ can be conducted through a simple single-dimensional search in the interval $[\underline{d}, \overline{d}]$.

EXAMPLE: UNIFORM $[0,2]$ RANDOM DEMAND.

We revisit the earlier example of $\epsilon \sim U[0,2]$ with $a = 1$, $b = 2$ and $c^R = c^M = 1$. Figure 5.2 shows the profits and prices under the resale-price method when $(\tau^R, \tau^M) \in \{(1,1), (1,0.8), (0.8,1)\}$. In the decentralized system, the retailer faces the same problem in all three cases, and her optimal price satisfies $p^R = \frac{bc^R}{b-1} \cdot \frac{z}{X(z)}$. This expression is valid only

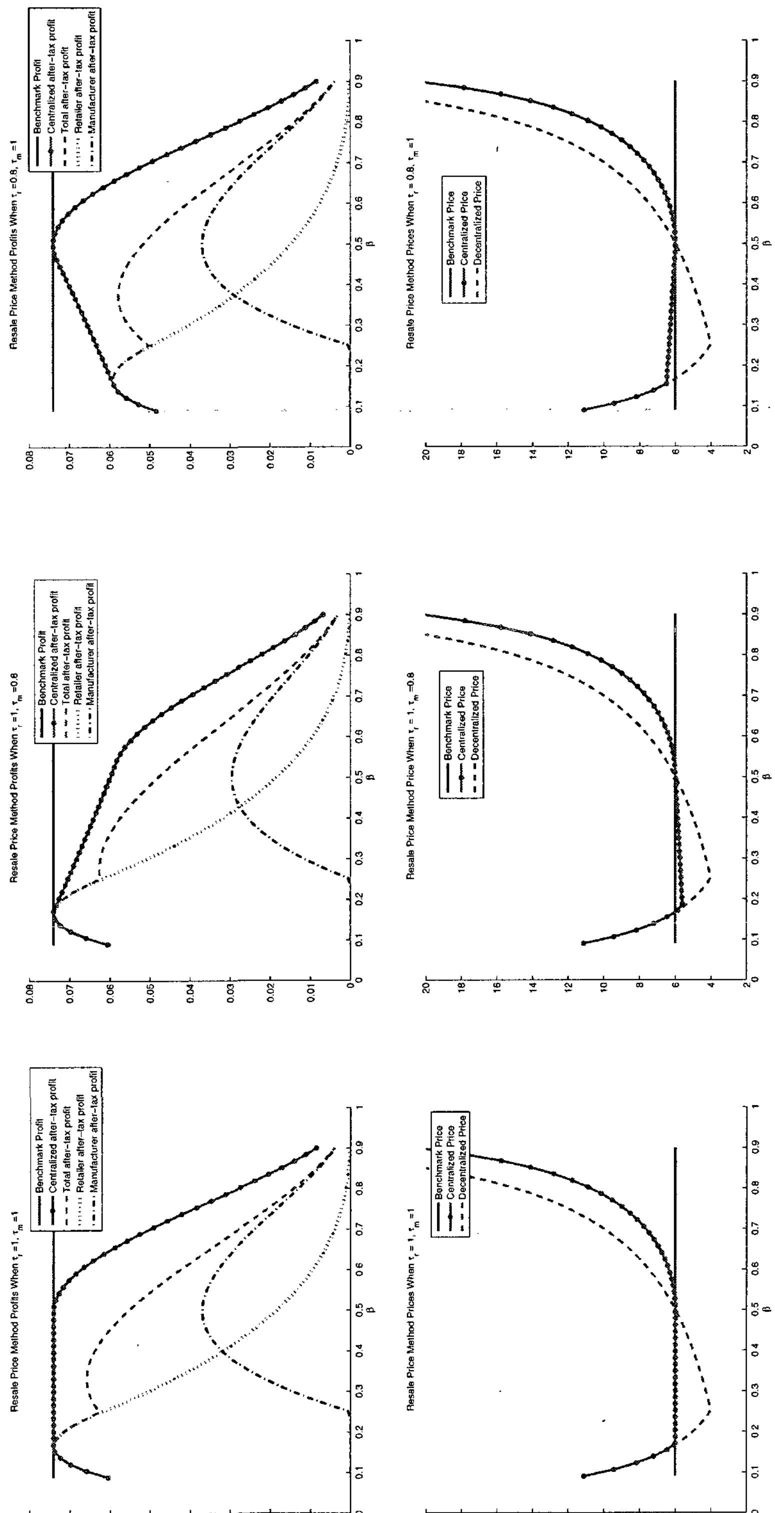


Figure 5.2: Multiplicative random demand model in resale-price method when $a = 1, b = 2, c^R = c^M = 1$

for β exceeding a certain threshold value. If this condition does not hold, then the manufacturer's profit $\tau^M \Pi^M(p^R)$ is zero. We observe that the retailer's after-tax profit $\tau^R \Pi^{R*}$ achieves the highest when the markdown rate β is around 0.2, whereas the manufacturer's profit $\Pi^M(z^R, p^R)$ is highest at a large value of around $\beta = 0.5$. We note that both $\tau^R \Pi^{R*}$ and $\tau^M \Pi^M(z^R, p^R)$ are quasi-concave with respect to β.

The total after-tax profit in the decentralized system, $\tau^R \Pi^{R*} + \tau^M \Pi^M(z^R, p^R)$ is quasi-concave in each of the two intervals based on the above-mentioned threshold value, and the local maximizers in each of these intervals are interior solutions. (This is in contrast to the cost-plus case where, if $\tau^R \geq \tau^M$, the total after-tax profit is maximized by the extreme markup rate value of $\gamma = 1$, in which case, the manufacturer's profit is always zero.) Thus, as Ailawadi et al. (1999) have pointed out, the resale-price method has a desirable property which is that both the retailer and manufacturer can make positive profits when the total before-tax profit is maximized. Then, it may be easier for the two divisions to coordinate under the resale-price method than under the cost-plus method.

5.3.4 Comparison of the Cost-Plus Method and the Resale-Price Method

In this section, we compare the properties of the cost-plus method and the resale-price method, focusing on the decentralized system.

We first directly compare the profits under each parameter values γ and β. For each combination of parameters values γ and β in the model, Figure 5.3 shows which of the two transfer pricing methods generates higher expected profit for the retailer, for the manufacturer, or for the total system. The gray area in the figure indicates that the profit under the resale-price is higher than that of the cost-plus method. Demand is uniformly distributed between 0 and 2. We note that the retailer makes more profit under the resale-price method when the markdown parameter β of the resale-price method is close to 1 (thus keeping most of the revenue), and the markup parameter γ of the cost-plus method is small (thus not paying much margin to the manufacturer). The figure also shows that the manufacturer generally prefers to resale-price method to the cost-price method. The firm-wide profit of the firm is higher under the resale-price method generally when β is reasonably high (such that the retailers profit becomes similar to the firm-wide profit), but the threshold level of β is dependent of γ. This is helpful to the central and local management who face the decision of selecting which transfer pricing method to use.

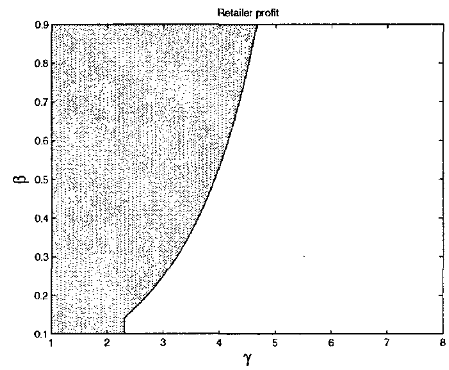
Since the parameters γ and β of these methods do not naturally correspond to each other, comparing these methods directly is not straightforward. We thus take two indirect approaches. In the first approach, we compare two methods by choosing the parameters such that the retailer's profits remain the same in both methods, and then we compare the manufacturer's profits. By restricting our attention to the *deterministic* demand case where $\epsilon = 1$, we are able to prove that the resale-price yields higher profit to the manufacturer than the cost-plus method in this comparison. In the second approach, we choose parameters such that the *share* of the retailer's profits from the total profit remain the same in both methods and compare their profits. In our notation, we use the subscript CP and RP to denote the cost-plus method or the resale-price method, respectively.

Fixing Retailer's Profits. For the fist approach, we assume that the demand is deterministic. We now fix the retailers' profits to remain the same in both methods. In particular, we choose the markdown rate of the resale-price method, such that the revenue is shared between the retailer and the manufacturer according to their relative contribution to the production, i.e., $\beta = c^M/c$. Since the retailer decides the price and quantity of the product, we then calculate the markup rate of the cost-plus method such that the retailer's profits are the same under both methods. In the following theorem, we establish the relationship between the manufacturers' profits under the two transfer pricing methods.

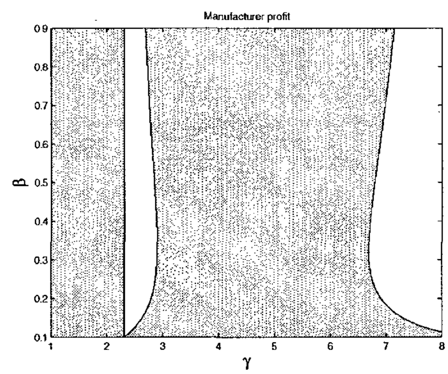
Theorem 5.3.8. *Assume that the demand is deterministic, i.e., $\epsilon = 1$. Let $\beta = c^M/c$, and let γ such that $\Pi_{CP}^{R*} = \Pi_{RP}^{R*}$. Let p_{CP}^R and p_{RP}^R be the retailer's optimal selling pricing decision under the cost-plus method and the resale price method, respectively. Then, $\Pi_{CP}^M(p_{CP}^R) \leq \Pi_{RP}^M(p_{RP}^R)$.*

The theorem implies that the manufacturer makes relatively less profit in the cost-plus method than in the resale-price method, if the parameters are chosen such that the retailer makes the same amount of profit under both methods. It is surprising that such inequality can be established. The proof of the above theorem is found in Appendix D.7.

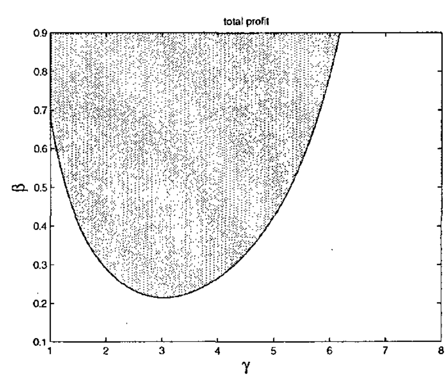
Fixing Ratio of Profits. We compare the profits under the cost-plus method and the resale-price method computationally, by fixing the retailer's share of the total after-tax



(a) Comparison of the retailer's profits under CP and RP



(b) Comparison of the manufacturer's profits under CP and RP



(c) Comparison of the total profits under CP and RP

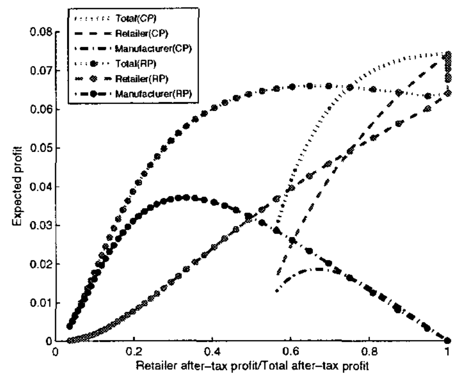
Figure 5.3: Comparison of the profits under the cost-plus and the resale-price methods. The gray area indicates that the profit under the resale-price method is higher than that under the cost-plus method.

profit in the decentralized system (shown on the horizontal axes). We continue with the example of $a = 1$, $b = 2$, $c^M = c^R = 1$ and $\epsilon \sim U[0, 2]$. As shown in Figure 5.4 (a), while the resale-price method allows an arbitrary division of profits between the retailer and the manufacturer, the retailer's share of the profit under the cost-plus method is always higher than half ($1/b = 0.5$) of the total profit if tax rates are the same (from Lemma 5.3.3 (b)). Thus, the resale-price method is not appropriate when the retailer wants to avoid a situation where the retailer's profit is less than half of the total firm profit. We also observe that for both the retailer's profit and the manufacturer's profit, the cost-plus method is preferable when the retailer's profit is much higher than the manufacturer's (i.e., the ratio is close to 1); otherwise, in general, the resale-price method tends to outperform the cost-plus method. The comparison of after-tax profits, shown in Figure 5.4 (b) and (c), is not straightforward as that of before-tax profits; for example, the retailer's share of the total after-tax profit under the cost-plus method is now bounded above by $1/[(b-1)\tau^M/\tau^R + 1]$ and this ratio depends on the ratio of τ^M and τ^R.

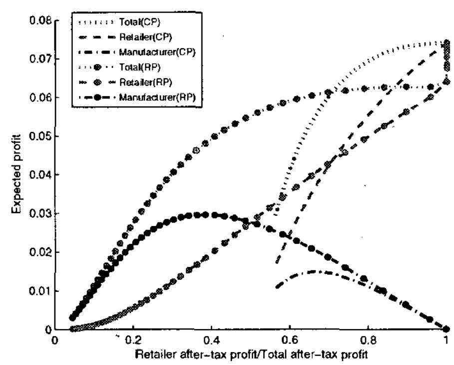
5.3.5 The Effect of Variability in Demand

In Sections 5.3.2 and 5.3.3, our analysis has shown the robustness of our results with respect to the demand variability, as the structural properties are not sensitive to the distribution of the stochastic component ϵ. In this section, we focus on the decentralized system, and examine the impact of demand variability on the optimal pricing and quantity decisions. We parameterize the random variable ϵ by δ, which represents the magnitude of demand variability. Let ϵ be distributed uniformly in the interval $[1-\delta, 1+\delta]$, where $0 \leq \delta \leq 1$. Then, $E[\epsilon] = 1$ and $\text{Var}[\epsilon] = \delta^2/3$. Note that $\delta = 0$ corresponds to the deterministic demand model. We use $a = 1$, $b = 2$, and $c^R = c^M = 1$.

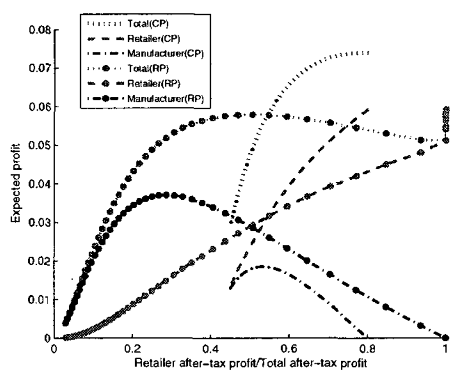
Figure 5.5 shows the expected profits of the firm, the retailer and the manufacturer. The shapes of the plots are familiar from Figures 5.1 and 5.2. While it is not surprising that demand variability causes a decline in expected profits, we learn that the cost of variability is shared by both the manufacturer and the retailer, and thus any measure to reduce demand variability would benefit both divisions. Figure 5.6 shows the optimal pricing and quantity decisions of the retailer. Under the cost-plus method, we observe that the optimal prices are



(a) $\tau^R = 1, \tau^M = 1$

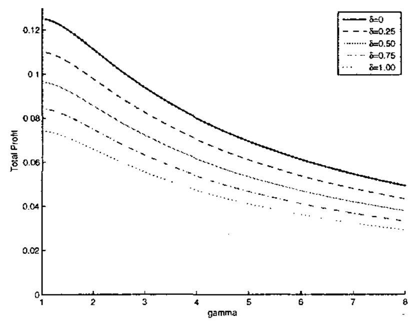


(b) $\tau^R = 1, \tau^M = 0.8$

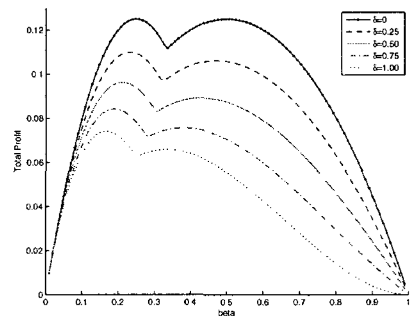


(c) $\tau^R = 0.8, \tau^M = 1$

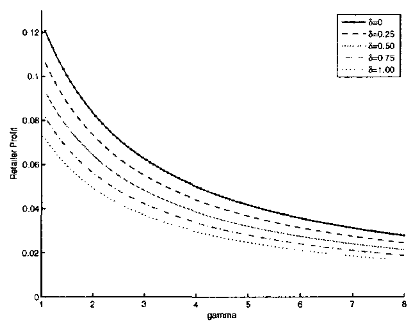
Figure 5.4: Expected profits of the central system, retailer, manufacturer, and total supply chain against the retailer's expected profit to the total supply chain expected profit. CP represents cost-plus method and RP represents resale-price method.



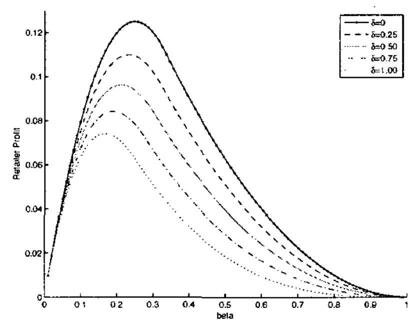
(a) Total before-tax profit (cost-plus)



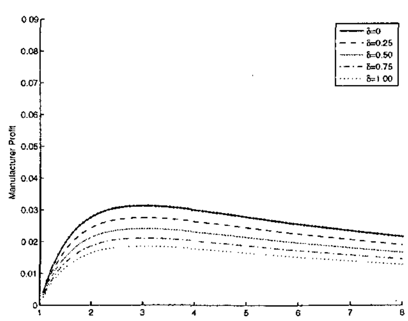
(b) Total before-tax profit (resale price)



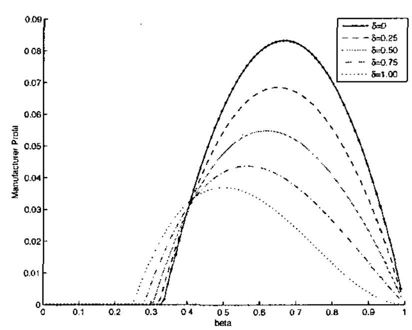
(c) Retailer's expected profit (cost-plus)



(d) Retailer's expected profit (resale price)

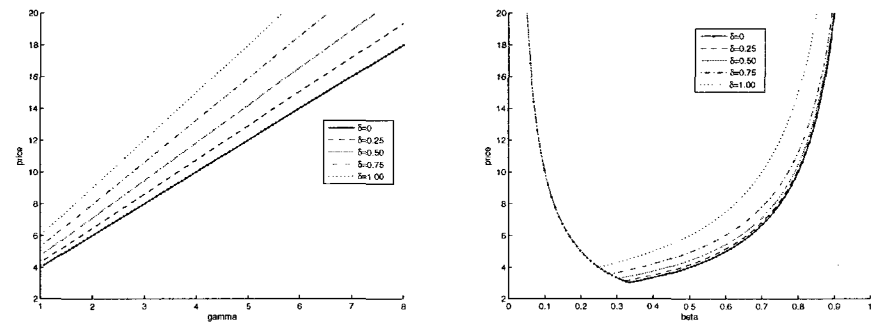


(e) Manufacturer's expected profit (cost-plus)

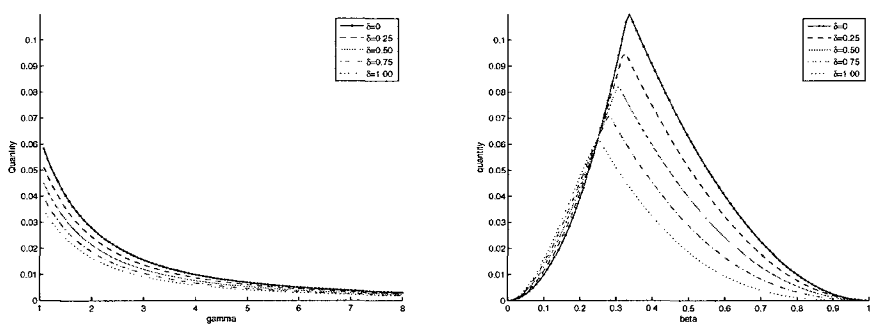


(f) Manufacturer's expected profit (resale price)

Figure 5.5: Expected profits in Decentralized Systems: $a = 1, b = 2, c^R = c^M = 1, \epsilon \sim U[1-\delta, 1+\delta]$ where $0 < \delta \leq 1$.



(a) Retailer's Price Decision p^R (cost-plus) (b) Retailer's Price Decision p^R (resale price)



(c) Retailer's Quantity Decision (cost-plus) (d) Retailer's Quantity Decision (resale price)

Figure 5.6: Retailer's Decisions in Decentralized Systems: $a = 1, b = 2, c^R = c^M = 1, \epsilon \sim U[1-\delta, 1+\delta]$ where $0 < \delta \leq 1$.

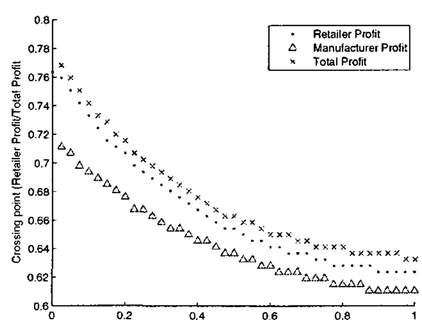


Figure 5.7: Resale-Price vs. Cost-Plus Preference Threshold as a Function of Demand Variability: $\epsilon \sim U[1-\delta, 1+\delta]$ where $0 < \delta \leq 1$.

always higher in the random demand model compared to the deterministic demand model. This property is linked to the multiplicative modeling of demand uncertainty – in their study of the single-stage price-setting newsvendor model, Petruzzi and Dada (1999) note and explain why the multiplicative form of demand uncertainty increases the optimal price (whereas the additive form decreases it). Correspondingly, demand uncertainty decreases optimal quantities. Now, under the resale-price method, if β is low, then the prices are independent of demand variability δ since they are at the boundary of the nonnegativity constraint of the manufacturer's profit, i.e., $p = c^M/\beta$; if β is high, this constraint is no longer active and the price in the random demand model is higher, as in the cost-plus method.

Recall that, in Figure 5.4, we have compared the cost-plus and resale-price methods by fixing the retailer's percentage share of the firm profit, and have shown that when this percentage share is low or intermediate, the resale-price method is preferable and when it is high, the cost-plus method is preferable. In Figure 5.7, we plot the threshold at which the cost-plus method becomes preferable. We see that the thresholds decrease as demand variability δ increases, which indicates that the cost-plus method becomes more attractive as δ increases. While this phenomenon cannot be explained easily, the simplicity of the cost-plus method seems to have positive effects when demand is highly uncertain. Furthermore, the presence and magnitude of demand variability affects optimal decisions and outcomes.

5.4 Conclusion

In this chapter, we have considered the impact of the transfer pricing methods for tax purposes on the optimal pricing and ordering decisions of a multinational firm. Since transfer pricing is regulated by tax authorities, the choice of transfer pricing methods is restricted, and two of the commonly-used methods (cost-plus and resale price) are considered in this chapter. We have incorporated demand uncertainty into our model, and analyzed the problem of quantity and pricing decisions in the framework of the price-setting newsvendor model and its variant. Our analysis includes details of how to solve the optimization problems as well as the properties of the optimal solution.

In the cost-plus method, the retailer's proportion of the firm's expected profit can be bounded below by parameters of the demand model, and the performance of the decentralized firm is maximized when the retailer takes all the profit if the tax rates are the same. However, the resale-price method allows a wider range of profit sharing between the two divisions, and the optimal firm-wide solution yields positive profits to both divisions. The above comparison of after-tax profit under two methods is not straightforward if tax rates differ; a similar conclusion can be reached when the difference in tax rates is not significant or the tax rate to the manufacturer is higher than the rate to the retailer. However, when the retailer's tax rate is higher, the result depends on the balancing of the double marginalization effect and tax regulation effect. We compare the two transfer pricing methods in terms of profit for a range of parameter values used in these methods. Furthermore, we have also compared these two methods by fixing the percentage of the total firm's profit allocated to each division. When the retailer's share is small or moderate (i.e., relatively weak bargaining power of the retailer), the resale-price method is preferred by both divisions, and when it is high, the cost-plus method is preferred. Interestingly, the performance of the cost-plus method, relative to the resale-price method, improves as demand becomes more uncertain.

The current literature of operations management have largely ignored important issues arising from the accounting and taxation arena. From our interaction with practitioners in public accounting companies, many supply chain decisions are made based on tax considerations. This chapter demonstrates an example of modeling how the accounting and tax considerations can make an impact on operational decisions of that firm, and that many important research questions are yet to be explored in the interface of operations management and accounting.

Part III

Bibliography

Bibliography

Agarwal, V., N. Y. Naik. 2000. Multi-period performance persistence analysis of hedge funds. *Journal of Financial and Quantitative Analysis* **35**(3) 327–342.

Agarwal, V., N. Y. Naik. 2005. Hedge funds. *Foundations and Trends in Finance* **1**(2) 103–169.

Ailawadi, K., P. Farris, E. Shames. 1999. Trade promotion: Essential to selling through resellers. *Sloan Management Review,* Fall **41**(1) 83–92.

Amenc, N., S. E. Bied, L. Martellini. 2003. Predictability in hedge fund returns. *Financial Analysts Journal* **59**(5) 32–46.

Amihud, Y., H. Mendelson. 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* **17**(2) 223–249.

Amihud, Y., H. Mendelson. 1991. Liquidity, maturity, and the yields on U.S. treasury securities. *Journal of Finance* **46**(4) 1411–1425.

Amin, G. S., H. M. Kat. 2003. Hedge fund performance 1990-2000: Do the "money machines" really add value? *Journal of Financial and Quantitative Analysis* **38**(2) 251–274.

Amo, A., H. Harasty, P. Hillion. 2007. Diversification benefits of funds of hedge funds: Identifying the optimal number of hedge funds. *The Journal of Alternative Investments* **10**(2) 10–22.

Aragon, G. 2007. Share restrictions and asset pricing: evidence from the hedge fund industry. *Journal of Financial Economics* **83**(1) 33–58.

Atlan, M., H. Geman, M. Yor. 2006. Options on hedge funds under the high-water mark rule. *Working Paper, Université Pierre et Marie Curie and Université Paris Dauphine* .

Babich, V. 2007. Competition and diversification effects in supply chains with supplier default risk? *Manufacturing & Service Operations Management* **9**(2) 123–146.

Bajari, P. L., R. S. McMillan, S. Tadelis. 2008. Auctions versus negotiations in procurement: An empirical analysis. *forthcoming to Journal of Law, Economics, and Organization* .

Baldenius, T., N. D. Melumad, S. Reichelstein. 2004. Integrating managerial and tax objects in transfer pricing. *The Accounting Review* **79**(3) 591–615.

Baldenius, T., S. Reichelstein. 2006. External and internal pricing in multidivisional firms. *Journal of Accounting Research* **44**(1) 1–28.

Baquero, G., J. ter Horst, M. Verbeek. 2005. Survival, look-ahead bias, and persistence in hedge fund performance. *Journal of Financial and Quantitative Analysis* **40**(3) 493–517.

Barndorff-Nielsen, O. E., N. Shephard. 2001. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B* **63**(2) 167–241.

Beil, D. R., L. M. Wein. 2003. An inverse-optimization-based auction mechanism to support a multiattribute rfq process. *Management Science* **49**(11) 1529–1545.

Benjaafar, S., E. Elahi, K. L. Donohue. 2007. Outsourcing via service competition. *Management Science* **53**(2) 241–259.

Benvignati, A.M. 1985. An empirical investigation of international transfer pricing by u.s. manufacturing firms. A.M. Rugman, L. Eden, eds., *Multinationals and Transfer Pricing*. St. Martin's Press, 308–314.

Bernard, A.B., J.B. Jensen, P.K. Schott. 2006. Transfer pricing by U.S.-based multinational firms. *NBER Working Paper 12493* .

Bernstein, F., A. Federgruen. 2005. Decentralized supply chains with competing retailers under demand uncertainty. *Management Science* **51**(1) 18–29.

Binmore, K. 1985. Bargaining and coalitions. A. E. Roth, ed., *Game Theoretic Models of Bargaining*. Cambridge University Press, 269–304.

Birge, J. 2000. Option methods for incorporating risk into linear capacity planning models. *Manufacturing & Service Operations Management* **2**(1) 19–31.

Bonaccorsi, A., P. L. Thomas, F. Pammolli, G. Turchetti. 1999. Auctions vs. bargaining: An empirical analysis of medical device procurement. *Laboratory of Economics and Management Papers Series 1999/20, Sant'Anna School of Advanced Studies, Pisa, Italy* .

Bonser, J. S., S. D. Wu. 2001. Procurement planning to maintain both short-term adaptiveness and long-term perspective. *Management Science* **47** 769–786.

Borland, L. 2002. A theory of non-Gaussian option pricing. *Quantitative Finance* **2**(6) 415–431.

Boyson, N. M, M. J. Cooper. 2004. Do hedge funds exhibit performance persistence? A new approach. *Working Paper, Northeastern and Purdue University* .

Brandt, A. 1986. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* **18**(1) 211–220.

Brown, S. J., W. N. Goetzmann, R. G. Ibbotson. 1999. Offshore hedge funds survival and performance 1989-1995. *Journal of Business* **72**(1) 91–117.

Brown, S. J., W.N. Goetzmann, J. Park. 2001. Careers and survival: competition and risk in the hedge fund and cta industry. *Journal of Finance* **56**(5) 1869–1886.

Browne, S., W. Whitt. 1996. Portfolio choice and the Bayesian Kelly criterion. *Advances in Applied Probability* **28**(4) 1145–1176.

Budish, E.B., L.N. Takeyama. 2001. Buy prices in online auctions: Irrationality on the internet? *Economic Letters* **72**(3) 325–333.

Bulow, J., P. Klemperer. 1996. Auctions versus negotiations. *The American Economic Review* **86**(1) 180–194.

Cachon, G. 2003. *Supply Chain Coordination with Contracts*, chap. 6. Handbooks in Operations Research and Management Science: Supply Chain Management, North Holland.

Cachon, G. P., M. A. Lariviere. 2001. Turning the supply chain into a revenue chain. *Harvard Business Review* **79**(3) 20–21.

Cachon, G. P., F. Zhang. 2006. Procuring fast delivery: Sole sourcing with information asymmetry. *Management Science* **52**(6) 881–896.

Caldentey, R., M. Haugh. 2009. Supply contracts with financial hedging. *Operations Research* **57**(1) 47–65.

Caldentey, R., G. Vulcano. 2007. Online auctions and list price revenue management. *Management Science* **53**(5) 795–813.

Capocci, D., G. Hüber. 2004. Analysis of hedge fund performance. *Journal of Empirical Finance* **11**(1) 55–89.

Chan, N., M. Getmansky, S. M. Hass, A. W. Lo. 2006. Do hedge fund increase systematic risk? *Economic Review - Federal Reserve Bank of Atlanta* **91**(4) 49–80.

Chen, F. 2007. Auctioning supply contracts. *Management Science* **53**(10) 1562–1576.

Chen, R. R., R. O. Roundy, R. Q. Zhang, G. Janakiraman. 2005. Efficient auction mechanisms for supply chain procurement. *Management Science* **51** 467–482.

Chen, X., C. Li, B. Rhee, D. Simchi-Levi. 2007. The impact of manufacturer rebates on supply chain profits. *Naval Research Logistics* **54**(6) 667–680.

Chordia, T., R. Roll, A. Subrahmanyam. 2001. Market liquidity and trading activity. *Journal of Finance* **56**(2) 501–530.

Cont, R. 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **1**(2) 223–236.

Copithorne, L. W. 1971. International corporate transfer prices and government policy. *Canadian Journal of Economics* **4**(3) 324–341.

Czechowicz, I.J., F.D.S. Choi, V.B. Bavishi. 1982. *Assessing Foreign Subsidiary Performance Systems and Practices of Leading Multinational Companies.* Business International Corporation.

Dalya, S. P., P. Nath. 2005. Online reverse auctions: Issues, themes, and prospects for the future. *Industrial Marketing Management* **34**(2) 157–166.

Darar, V., N. Y. Naik, R. Radcliffe. 1998. Liquidity and asset returns: An alternative test. *Journal of Financial Markets* **1** 203–219.

Derman, E., K. S. Park, W. Whitt. 2009a. Markov chain models to estimate the premium for extended hedge fund lockups. *forthcoming to Wilmott Journal* .

Derman, E., K. S. Park, W. Whitt. 2009b. A stochastic-difference-equation model for hedge-fund relative returns. Forthcoming to *Quantitative Finance.*

Eden, L. 1983. Transfer pricing policies under tariff barriers. *Canadian Journal of Economics* **16**(4) 669–685.

Eden, L. 1998. *Taxing Multinationals: Transfer Pricing and Corporate Income Taxation in North America.* University of Toronto Press.

Edwards, F. R., M. O. Caglayan. 2001. Hedge fund performance and manager skill. *Journal of Futures Markets* **21**(11) 1003–1028.

Eleswarapu, V. R., M. R. Reinganum. 1993. The seasonal behavior of liquidity premium in asset pricing. *Journal of Financial Economics* **34**(3) 373–386.

Eling, M., F. Schuhmacher. 2007. Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking and Finance* **31**(9) 2632–2647.

Elmaghraby, W., P. Keskinocak, M. Rothkopf. 2005. Introduction. *Electronic Commerce Research* **5** 175–180.

Elyakime, B., J. Laffont, P. Loisel, Q. Vuong. 1997. Auctioning and bargaining: An econometric study of timber auctions with secret reservation prices. *Journal of Business and Economic Statistics* **15**(2) 209–220.

Embrechts, P., C. Klüppelberg, T. Mikosch. 1997. *Modeling Extremal Events*. Springer.

Emmons, H., S. Gilbert. 1998. The role of returns policies in pricing and inventory decisions for catalogue goods. *Management Science* **44**(2) 276–283.

Engelbrecht-Wiggans, R., E. Katok. 2006. E-sourcing in procurement: Theory and behavior in reverse auctions with noncompetitive contracts. *Management Science* **52**(4) 581–596.

Ernst&Young. 2008. *Pricing under Pressure: Global transfer pricing survey 2007-2008*. Ernst and Young International.

Eső, P., B. Szentes. 2007. Optimal information disclosure in auctions and the handicap auction. *Review of Economic Studies* **74**(3) 705–731.

Fama, E. 1965. The behavior of stock-market prices. *Journal of Business* **38**(1) 34–105.

Federgruen, A., A. Heching. 1999. Combined pricing and inventory control under uncertainty. *Operations Research* **47**(3) 454–475.

Federgruen, A., N. Yang. 2008. Selecting a portfolio of suppliers under demand and supply risks. *Operations Research* **56**(4) 916–936.

Feinschreiber, R. 2004. *Transfer Pricing Methods: An Applications Guide*. John Wiley & Sons, Inc.

Ferguson, T. S. 2000. *Optimal Stopping and Applications*. Mathematics Department, UCLA. URL `http://www.math.ucla.edu/~tom/Stopping/Contents.html`.

Fung, W., D. A. Hsieh. 2000. Performance characteristics of hedge funds and commodity funds: Natural versus spurious biases. *Journal of Financial and Quantitative Analysis* **35**(3) 291–307.

Fung, W., D. A. Hsieh. 2009. Measurement biases in hedge fund performance data: an update. *Financial Analysts Journal* **65**(3) 36–38.

Fung, W., D. A. Hsieh, N. Y. Naik, T. Ramadorai. 2008. Hedge fund performance, risk and capital formation. *Journal of Finance* **63**(4) 1777–1803.

Gabaix, X., P. Gopikrishnan, V. Plerou, E. Stanley. 2007. A unified econophysics explanation for the power-law exponents of stock market activity. *Physica A: Statistical Mechanics and its Applications* **382**(1) 81–88.

Gallien, J., S. Gupta. 2007. Temporary and permanent buyout prices in online auctions. *Management Science* **53**(5) 814–833.

Geman, H., C. Kharoubi. 2003. Hedge funds revisited: Distributional characteristics, dependence structure and diversification. *Journal of Risk* **5**(4) 55–73.

Granfield, M. 1995. *Economic Analysis and the Defense of Transfer Pricing Structures and Results*. Washington D.C.:Tax Management Inc., Bureau of National Affairs.

Granot, D., S. Yin. 2005. On the effectiveness of returns policies in the price-dependent newsvendor model. *Naval Research Logistics* **52**(8) 765–779.

Gregoriou, G. N. 2002. Hedge fund survival lifetimes. *Journal of Asset Management* **3**(3) 237–252.

Halperin, R., B. Srinidhi. 1987. The effect of the U.S. income tax regulations' transfer pricing rules on allocative efficiency. *The Accounting Review* **62**(4) 686–706.

Halperin, R. M., B. Srinidhi. 1991. U.S. income tax transfer-pricing rules and resource allocation: The case of decentralized multinational firms. *The Accounting Review* **66**(1) 141–157.

Hasanhodzic, J., A. W. Lo. 2007. Can hedge-fund returns be replicated?: The linear case. *Journal of Investment Management* **5**(2) 5–45.

Hayes, B. T. 2006. Maximum drawdowns of hedge funds with serial correlation. *Journal of Alternative Investments* **8**(4) 26–38.

Hidvégi, Z., W. Wang, A.B. Whinston. 2006. Buy-price english auction. *Journal of Economic Theory* **129**(1) 31–56.

Hirshleifer, J. 1956. On the economics of transfer pricing. *Journal of Business* **29**(3) 172–184.

Hodder, J. E., J. C. Jackwerth. 2007. Incentive contracts and hedge fund management. *Journal of Financial and Quantitative Analysis* **42**(4) 811–826.

Horst, T. 1971. The theory of the multinational firm: Optimal behavior under different tariff and tax rates. *Journal of Political Economy* **79**(5) 1059–1072.

Horst, U. 2001. The stochastic equation $Y_{t+1} = A_t Y_t + B_t$ with non-stationary coefficients. *Journal of Applied Probability* **38**(1) 80–94.

Horst, U. 2003. Stability of linear stochastic difference equations in strategically controlled random environments. *Advances in Applied Probability* **35**(4) 961–981.

Ippolito, R. 1989. Efficiency with costly information: A study of mutual fund performance. *Quarterly Journal of Economics* **104**(1) 1–23.

Jagannathan, R., A. Malakhov, D. Nonikov. 2006. Do hot hands persist among hedge fund managers. *NBER Working Paper* .

Jap, S. D. 2002. Online reverse auctions: Issues, themes, and prospects for the future. *Journal of the Academy of Marketing Science* **30**(4) 506–525.

Johnston, D. C. 1995. International business; cross-border taxes complicate trade. *New York Times* .

Kanodia, C. 1979. Risk sharing and transfer price systems under uncertainty. *Journal of Accounting Research* **17**(1) 74–98.

Kassberger, S., R. Kiesel. 2006. A fully parametric approach to return modelling and risk management of hedge funds. *Financial Markets and Portfolio Management* **20**(4) 472–491.

Khouja, M. 1996. The newsboy problem with multiple discount offered by suppliers and retailers. *Decision Science* **27**(3) 589–599.

Kjerstad, E. 2005. Auctions vs negotiations: A study of price differences. *Health Economics* **14**(12) 1239–1251.

Koh, F., W. T. H. Koh, M. Teo. 2003. Asian hedge funds: Return persistence style and fund characteristics. Working Paper, Singapore Management University.

Kosowski, R., N. Y. Naik, M. Teo. 2007. Do hedgefunds deliver alpha? a bayesian and bootstrap analysis. *Journal of Financial Economics* **84**(1) 229–264.

Koulamas, C. 2006. A newsvendor problem with revenue sharing and channel coordination. *Decision Science* **37**(1) 91–100.

Krishna, V. 2002. *Auction Theory*. Academy Press.

Krishnamurthy, A. 2002. The bond / old-bond spread. *Journal of Financial Economics* **66**(2-3) 463–506.

Lariviere, M. 2006. A note on probability distributions with increasing generalized failure rates. *Operations Research* **54**(3) 602–604.

Lariviere, M., E. Porteus. 2001. Selling to the newsvendor: An analysis of price-only contracts. *Manufacturing & Service Operations Management* **3**(4) 293–305.

Leffler, K. B., T. R. Rucker, I. A. Munn. 2003. The choice among sales procedures: Auction vs. negotiated sales of private timber. *Working Paper, Department of Economics, University of Washington* .

Lhabitant, F. 2004. *Hedge Funds: Quantitative Insights*. Wiley.

Li, Q., D. Atkins. 2002. Coordinating replenishment and pricing in a firm. *Manufacturing & Service Operations Management* **4**(4) 241–257.

Li, Q., D. Atkins. 2005. On the effect of demand randomness on a price/quantity setting firm. *IIE Transactions* **37**(12) 1143–1153.

Liang, B. 1999. On the performance of hedge fund. *Financial Analysts Journal* **55**(4) 72–85.

Lilliefors, H. W. 1967. On the Kolmogorov-Smirnov test for normality with mean and variances unknown. *Journal of the American Statistical Association* **62**(318) 399–402.

Lin, L. I. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**(1) 255–268.

Lo, A. 2001. Risk management for hedge funds: Introduction and overview. *Financial Analysts Journal* **57**(6) 16–33.

Longstaff, F. A. 1995. How much can marketability affect security values ? *Journal of Finance* **50**(5) 1767–1774.

Longstaff, F. A. 2001. Optimal portfolio choice and the valuation of illiquid securities. *Review of Financial Studies* **14**(2) 407–431.

Longstaff, F. A. 2004. The flight-to-liquidity premium in US treasury bond price. *Journal of Business* **77**(3) 511–526.

Lux, T. 1996. The stable Paretian hypothesis and the frequence of large returns: an examination of major German stocks. *Applied Financial Economics* **6**(6) 463–475.

Mandelbrot, B. 1963. The variation of certain speculative prices. *Journal of Business* **36**(4) 394–419.

Manelli, A. M., D. R. Vincent. 1995. Optimal procurement mechanisms. *Econometrica* **63**(3) 591–620.

Massey, F. J. 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of American Statistical Association* **46**(253) 68–78.

Mathews, T. 2004. The impact of discounting on an auction with a buyout option: a theoretical analysis motivated by ebay's buy-it-now feature. *Journal of Economics* **81**(1) 24–52.

Menezes, F. M., P. K. Monteiro. 2005. *Introduction to Auction Theory*. Oxford University Press.

Monahan, G. E., N. C. Petruzzi, W. Zhao. 2004. The dynamic pricing problem from a newsvendor's perspective. *Manufacturing & Service Operations Management* **6**(1) 73–91.

Muthoo, A. 1999. *Bargaining Theory with Applications*. Cambridge University Press.

Myerson, R. 1981. Optimal auction design. *Mathematics of Operations Research* **6**(1) 58–73.

Nagarajan, M., Y. Bassok. 2008. A bargaining framework in supply chains. *Management Science* **54**(8) 1482–1496.

Nagarajan, M., G. Sosic. 2008. Game-theoretic analysis of cooperation among supply chain agents: Review and extensions. *European Journal of Operational Research* **187**(3) 719–745.

Nielsen, S. B., P. Raimondos-Møller, G. Schjelderup. 2008. Taxes and decision rights in multinationals. *Journal of Public Economic Theory* **10**(2) 245–258.

Osborne, M., A. Rubinstein. 1990. *Bargaining and Markets*. Academy Press.

Park, H. 2006. Risk measures for hedge funds and a survival analysis. PhD Dissertion, University of Massachusetts.

Pasternack, B.A. 1985. Optimal pricing and returns policies for perishable commodities. *Marketing Science* **4**(2) 166–176.

Pastor, L., R. F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* **111**(3) 642–685.

Perakis, G., G. Roels. 2007. The price of anarchy in supply chains: Quantifying the efficiency of price-only contracts. *Management Science* **53**(8) 1249–1268.

Petruzzi, N.C., M. Dada. 1999. Pricing and the newsvendor problem: A review with extensions. *Operations Research* **47**(2) 183–194.

Rachev, S., S. Mittnik. 2000. *Stable Paretian Models in Finance*. Wiley.

Ronen, J., K. R. Balachandran. 1988. An approach to transfer pricing under uncertainty. *Journal of Accounting Research* **26**(2) 300–312.

Ronen, J., G. McKinney, III. 1970. Transfer pricing for divisional autonomy. *Journal of Accounting Research* **8**(1) 99–112.

Ross, S. M. 2003. *Introduction to Probability Models*. Eighth ed. Academy Press.

Rouah, F. 2006. Competing risks in hedge fund survival. PhD Dissertion, McGill University.

Rubinstein, A. 1982. Perfect equilibrium in a bargaining model. *Econometrica* **50**(1) 97–109.

Salmon, T. C., B. J. Wilson. 2008. Second chance offers versus sequential auctions: Theory and behavior. *Economic Theory* **34**(1) 47–67.

Samorodnitsky, G., M. S. Taqqu. 1994. *Stable Non-Gaussian Random Processes*. Chapman & Hall.

Samuelson, L. 1982. The multinational firm with arm's length transfer price limits. *Journal of International Economics* **13**(3-4) 365–374.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley.

Shaked, A. 1994. Opting-out: Bazaars versus "high-tech" markets. *Investigations Economics* **18**(3) 421–432.

Shunko, M., L. Debo, S. Gavirneni. 2008. Optimal transfer pricing and offshoring strategies of global supply chains. *Working Paper, Carnegie Mellon University* .

Shunko, M., S. Gavirneni. 2007. Role of transfer prices in global supply chains with random demands. *Journal of Industrial and Management Optimization* **3**(1) 99–117.

Song, Y., S. Ray, S. Li. 2008. Structural properties of buyback contracts for price-setting newsvendors. *Manufacturing & Service Operations Management* **10**(1) 1–18.

Spengler, J. 1950. Vertical integration and antitrust policy. *Journal of Political Economy* **58**(4) 347–352.

Tang, R. Y. W. 2002. *Current Trends and Corporate Cases in Transfer Pricing*. Quorum Books.

Taylor, T.A. 2002. Supply chain coordination under channel rebates with sales effort effects. *Management Science* **48**(8) 992–1007.

Thomas, C. J., B. J. Wilson. 2002. A comparison of auctions and multilateral negotiations. *RAND Journal of Economics* **33**(1) 140–155.

Tran, V. 2006. *Evaluating Hedge Fund Performance.* Wiley Finance.

Tunca, T. I., S. A. Zenios. 2006. Supply auctions and relational contracts for procurement. *Manufacturing and Service Operations Management* **8**(1) 43–67.

Vayanos, D. 2004. Flight to quality, flight to liquidity, and the pricing of risk. NBER Working Paper No. W10327.

Vervaat, W. 1979. On a stochastic difference equation and a representation of non-negative infinitely divisible random variable. *Advanced Applied Probability* **11**(4) 750–783.

Waehrer, K. 1999. The ratchet effect and bargaining power in a two-state model of competive bidding. *Economic Theory* **13**(1) 171–181.

Wan, Z., D. R. Beil. 2008. RFQ auctions with supplier qualification screening. *forthcoming to Operations Research* .

Wang, R. 2000. Bidding and renegotiation in procurement auctions. *European Economic Review* **15**(8) 1577–1597.

Wang, Y. 2006. Joint pricing-production decisions in supply chains of complementary products with uncertain demand. *Operations Research* **54**(6) 1110–1127.

Wang, Y., L. Jiang, Z. Shen. 2004. Channel performance under consignment contract with revenue sharing. *Management Science* **50**(1) 34–47.

Warga, A. 1992. Bond returns, liquidity, and missing data. *Journal of Financial and Quantitative Analysis* **27**(4) 605–617.

Weng, Z. F. 1995. Channel coordination and quantity discount. *Management Science* **41**(9) 1509–1522.

Whitt, W. 2002. *Stochastic-Process Limits.* Springer.

Wolfe, S. J. 1982. On a continuous analogue of the stochastic difference equation $X_{n+1} = \rho X_n + B_n$. *Stochastic Processes and Their Applications* **12**(3) 301–312.

Wu, S. D. 2004. Supply chain intermediation: A bargaining theoretic framework. D. Simchi-Levi, S.D. Wu, M. Shen, eds., *Handbook of Quantitative Supply Chain Analysis in the eBusiness Era*, chap. 3. Kluwer Academic Publishers, 67–115.

Yano, C.A., S.M. Gilbert. 2003. Coordinated pricing and production/procurement decisions: A review. A. Chakravarty, J. Eliashberg, eds., *Managing Business Interfaces: Marketing, Engineering and Manufacturing Perspectives*. Kluwer Academic Publishers.

Ye, L. 2007. Indicative bidding and a theory of two-stage auctions. *Games and Economic Behavior* **58**(1) 181–207.

Yeom, S., K. R. Balachandran, J. Ronen. 2000. The role of transfer price for coordination and control within a firm. *Review of Quantitative Finance and Accounting* **14**(2) 161–192.

Part IV

Appendices

Appendix A

Appendix to Chapter 2

A.1 Overview

In this appendix we elaborate on several issues investigated in Chapter 2. We start with statistical issues. In §A.2 we provide additional details about our statistical analysis of the TASS database. Specifically, we describe our estimates of the standard deviation σ. In §A.3 we describe an alternative ratio method for estimating the persistence of hedge fund returns.

We next turn to the DTMC models. In §A.4 we consider the two-state DTMC model without dying funds, which provides a link between §2.5 and §2.6 in Chapter 2. In §A.5, we investigate if the difference between the premium p_n and the linear approximation $\tilde{p}_n$ derived in §2.5 of Chapter 2 can be understood by the second term in the two-term asymptotic expansion developed there. We show that the second term explains the difference well for relatively short-term lockup premiums. In §A.6, we analyze what possible parameter values can occur in the DTMC model. Lastly, we supplement §2.7 in Chapter 2 in §A.7 by providing additional descriptions of the way the three-state DTMC model parameters and the lockup premium depend on basic hedge fund performance measures. We also show how the three or four-year lockup premium can be approximated with a simple multiplicative form of the three parameters δ, γ and σ, as shown in (2.1) in Chapter 2.

A.2 Statistical Results

In this section we supplement our discussion of sour statistical methods provided in §2.4. In particular, we describe our estimates of the standard deviations of the annual returns selected above. We display sample standard deviation for selected annual returns in Table A.1 and observe that they are within the range we consider in this chapter: from 0.05 to 0.25. For several strategy categories, the number of selected returns is too small to obtain meaningful estimates of the standard deviation. Thus, merging all returns from 2000 to 2004 for each strategy category may give better insight about the real variability of the annual returns.

Table A.1: Estimated standard deviation of annual returns (%)

strategy	Number of observation	2000	2001	2002	2003	2004	All
Convertible arbitrage	244	0.07	0.07	0.08	0.10	0.06	0.08
Dedicated short bias	30	0.02	0.15	0.18	0.22	0.14	0.16
Emerging market	325	0.20	0.22	0.15	0.19	0.12	0.17
Equity macro	270	0.12	0.06	0.09	0.06	0.07	0.08
Event driven	534	0.13	0.09	0.12	0.10	0.08	0.10
Fixed income arbitrage	196	0.07	0.03	0.08	0.09	0.06	0.07
Fund of fund	982	0.12	0.08	0.05	0.06	0.03	0.06
Global macro	176	0.07	0.11	0.12	0.15	0.08	0.11
Long short equity	1654	0.19	0.17	0.15	0.14	0.09	0.14
Managed future	238	0.13	0.13	0.12	0.14	0.10	0.12
Other	167	0.15	0.06	0.07	0.10	0.07	0.09

A.3 Persistence from Ratios of Average Relative Returns

An alternative way to estimate the persistence factor is to consider the ratio of the next-year average returns to the current-year average return, restricting attention to the returns that are positive in the current year. Table A.2 is the ratio of two successive average returns restricting attention to the returns that are positive and negative in the current year, respectively.

Since the average of relative returns is zero by definition, the ratio of averages for positive and negative returns should be identical. However, it does not hold always since we excluded outliers of relative returns that would not make the average of relative returns zero. As can be seen from Table 2.1 and A.2, these persistence estimates tend to be similar to the regression estimates.

Table A.2: Ratio of average relative returns for good states

(i) For positive current relative returns

Strategy	current year average relative return	next year average relative return	ratio (γ)
Convertible	6.04	2.32	0.38
Emerging market	15.31	4.70	0.31
Event driven	7.29	1.29	0.18
Fund of fund	4.26	1.56	0.37

(ii) For negative current relative returns

Strategy	current year average relative return	next year average relative return	ratio (γ)
Convertible	-4.12	-1.55	0.38
Emerging market	-12.21	-6.25	0.51
Event driven	-7.33	-1.28	0.18
Fund of Fund	-3.73	-0.69	0.18

A.4 The DTMC Model Without Death

We now return to the DTMC model and elaborate upon the analysis of the case $\delta = 0$. If we consider the DTMC without hedge funds dying, then we can work with a two-state DTMC, which has the transition matrix

$$P = \begin{matrix} G \\ S \end{matrix} \begin{pmatrix} p & 1-p \\ 1-r & r \end{pmatrix} , \tag{A.1}$$

which has only the two parameters p and r.

Let $\pi \equiv (\pi_G, \pi_S)$ be the steady-state probability vector of the two-state DTMC with transition matrix in (A.1). A convenient explicit expression for π is

$$\pi \equiv (\pi_G, \pi_S) = \left(\frac{1-r}{(1-r)+(1-p)}, \frac{1-p}{(1-r)+(1-p)} \right) = \left(\frac{1-r}{2-r-p}, \frac{1-p}{2-r-p} \right) . \tag{A.2}$$

Since Y_G and Y_S are the assumed relative returns (deviations from the mean return), we can express variance of the fund's relative performance in steady state as

$$\sigma^2 = \pi_G \cdot Y_G^2 + \pi_S \cdot Y_S^2 . \tag{A.3}$$

To satisfy, (A.3), we calibrate p and r in the transition matrix P of (A.1). We do this from two equations for persistence factor γ, expressed as a function of p and q.

$$\gamma \cdot Y_G = p \cdot Y_G + (1-p) \cdot Y_S \tag{A.4}$$

and

$$\gamma \cdot Y_S = (1-r) \cdot Y_G + r \cdot Y_S . \tag{A.5}$$

From (A.4) and (A.5), we derive that

$$p = \frac{\gamma \cdot Y_G - Y_S}{Y_G - Y_S}, \quad r = \frac{Y_G - \gamma \cdot Y_S}{Y_G - Y_S} . \tag{A.6}$$

From the above equations, it is straightforward to verify that

$$\frac{Y_G}{\sigma} \frac{-Y_S}{\sigma} = 1 . \tag{A.7}$$

If we fix $Y_S/\sigma = -1.5$, we have

$$\frac{Y_G}{\sigma} = \frac{1}{1.5} \approx 0.67 , \tag{A.8}$$

which exactly matches the analysis without Markov chain when $\delta = 0$. By assuming normally distributed annual relative returns, and letting Y_G be the median of the positive returns, we found that $Y_G/\sigma =$ median $\{|N(0,1)|\} = 0.67$.

A.5 Approximation of the No-death Lockup Premium

This section supplements §2.5 of Chapter 2 by examining the approximation of p_n by the two-term Taylor series expansion, in order to estimate p_n better than $\tilde{p}_n$ with a simple form. From the Taylor expansion, we obtain that

$$\log(x+1) \approx x \quad \text{and} \tag{A.9}$$

$$e^x \approx 1 + x + \frac{1}{2!}x^2, \tag{A.10}$$

for x close to 0. Applying these approximation formulas to (2.13), we have

$$\begin{aligned} p_n &\approx \frac{1}{n}\left[\log\left(1 + \mathbb{E}\left[\sum_{i=1}^{n} R_i^1\right] + \mathbb{E}\left[\frac{1}{2}\left(\sum_{i=1}^{n} R_i^1\right)^2\right]\right)\right. \\ &\quad \left. - \log\left(1 + \mathbb{E}\left[\sum_{i=1}^{n} R_i^2\right] + \mathbb{E}\left[\frac{1}{2}\left(\sum_{i=1}^{n} R_i^2\right)^2\right]\right)\right] \\ &\approx \frac{1}{n}\left[\left(\mathbb{E}\left[\sum_{i=1}^{n} R_i^1\right] - \mathbb{E}\left[\sum_{i=1}^{n} R_i^2\right]\right) + \frac{1}{2}\left(\mathbb{E}\left[\left(\sum_{i=1}^{n} R_i^1\right)^2\right] - \mathbb{E}\left[\left(\sum_{i=1}^{n} R_i^2\right)^2\right]\right)\right] \\ &\approx \tilde{p}_n + e_n \end{aligned} \tag{A.11}$$

Thus, we observe that p_n can be approximated by the simple formula $\tilde{p}_n + e_n$. However, it is hard to evaluate e_n analytically due to the dependence between R_i^j and R_{i+1}^j in general. We thus use Monte Carlo simulation with a large number (10^5) of relative returns for a fund under 1-year and n-year lockup in order to evaluate the lockup premium from 1 to 20 years. We find that 10^5 samples are sufficient to produce the same lockup premium to the premium obtained from (2.38), with negligible difference.

We compute $\tilde{p}_n + e_n, n = 1, 2, ..., 20$ numerically for three different death rates ($\delta = 0.00, 0.03, 0.06$), $\gamma = 0.5$ and the other parameter values in Table 2.2 of Chapter 2. As usual, we assume that a fund starts with a good state at the beginning. Numerical analysis shows that this e_n explains the difference between the exact lockup premium (p_n) and the analytical approximation ($\tilde{p}_n$) reasonably well for relatively small values of n, specifically, for $n \leq 5$. For example, for $\delta = 0$, we find that e_n explains more than 70% of the difference between p_n and $\tilde{p}_n$ for less than five years; see Figure A.1. However, it is also observed that as n increases, other higher-order terms omitted in the approximation formula (A.11) become increasingly important in the lockup premium. As n increases, it is more likely to see sample paths of a fund that become dead and start as a 1-year lockup fund. Notice that a 1-year lockup fund produces higher expected relative returns than n-year lockup fund. Thus, it is not surprising to see that $\tilde{p}_n + e_n$ becomes less accurate as n increases.

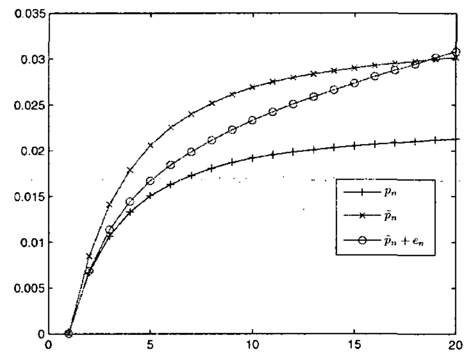


Figure A.1: The DTMC lockup premium (p_n), analytic approximation ($\tilde{p}_n$), and two-term approximation ($\tilde{p}_n + e_n$) for $\delta = 0$, $\gamma = 0.5, Y_G = 0.067, Y_S = -0.15$ and $Y_D = -0.20$.

A.6 Possible parameter values in the DTMC Model

In this section, we supplement §2.6.4 in Chapter 2 and determine what parameter values can occur in the DTMC model. Figure A.2 shows the three parameters as a function of δ with $Y_G = 0.067$, $Y_S = -0.15$, $Y_D = -0.20$ and $\gamma_G = \gamma_S = 0.5$.

From (2.30) in Chapter 2, we see that p is a linear function of γ_G with positive slope $Y_G/(Y_G - Y_S)$. If $Y_S \leq 0$, then we necessarily have $\gamma_G < p < 1$. The minimum possible value of p, attained when $\gamma_G = 0$, is $|Y_S|/(Y_G + |Y_S|)$. For example, if $Y_G = 0.05 > 0 > Y_S = -0.15$, then the minimum value of p is 0.75 (at $\gamma_G = 0$) and the slope is 0.25. On the other hand, if $Y_G > Y_S > 0$, then we must have $p \leq \gamma_G$. If, instead, $Y_G > Y_S > 0$, then we require that $\gamma_G \cdot Y_G > Y_S$.

From (2.30), we see that p is independent of δ. Since (2.26) implies that $\delta < 1 - p$, there is an upper bound on the possible δ, consistent with Figure 2.6. Moreover, that inequality can be restated as $p < 1 - \delta$. When combined with (2.30), that yields an upper bound on

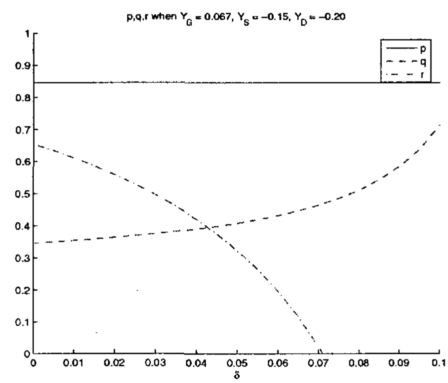


Figure A.2: The DTMC parameter values p, q and r as a function of δ when $Y_G = 0.067$, $Y_S = -0.15$, $Y_D = -0.20$ and $\gamma_G = \gamma_S = 0.5$

γ_G, which is strictly less than 1: $\gamma_G \leq (1-\delta)+\delta(Y_S/Y_G)$. For $Y_G = 0.067$ and $Y_S = -0.150$, $\gamma_G \leq 1 - 3.23\delta$.

Under the general condition that $Y_G > Y_S > Y_D$, we see that $q \equiv q(r)$ via (2.32) is a strictly decreasing function of r. The largest possible value of q occurs for $r = 0$, which is $(\gamma_S \cdot Y_S - Y_D)/(Y_G - Y_D)$. In order for q to be feasible (nonnegative), we must have that largest possible value be nonnegative. Hence to have a feasible nonnegative value of q, we must have $\gamma_S \cdot Y_S \geq Y_D$. That is always satisfied provided that $Y_D \leq 0$ (given that $Y_G > Y_S > Y_D$).

From (2.32) alone, we can find an upper bound on r in terms of γ_S, Y_S and Y_D. If $0 > Y_S > Y_D$, then we must have $(1-r)|Y_D| \geq (r - \gamma_S)|Y_S|$, so that

$$r < \frac{|Y_D/Y_S| + \gamma_S}{|Y_D/Y_S| + 1} < 1 \quad \text{for} \quad 0 < \gamma_S < 1 \;, \tag{A.12}$$

where $|Y_D/Y_S| > 1$. On the other hand, if $Y_S \geq 0 > Y_D$, then we have

$$r < \frac{(|Y_D|/Y_S) - \gamma_S}{(|Y_D|/Y_S) + 1} < 1 \;, \tag{A.13}$$

where now $|Y_D|/Y_S$ can assume a wide range of values.

When $Y_G > 0 \geq Y_S > Y_D$, r has the form $r = (a - B)/(A - b)$, where $a < A$ and $b < B$, so that we always have $r < 1$. We then have $r > 0$ if and only if either $a > B$ or $A < b$; r is negative otherwise. To have $r > 0$, we must have

$$a - B \equiv \left(\frac{(1-\delta)(1-\gamma_G)Y_G - \delta(Y_G - Y_S)}{(1-\gamma_G)Y_G}\right) - \left(\frac{\gamma_S \cdot Y_S - Y_D}{Y_G - Y_D}\right) > 0 \quad \text{or} \tag{A.14}$$

$$b - A \equiv \left(\frac{Y_S - Y_D}{Y_G - Y_D}\right) - \left(\frac{(1-\gamma_G)Y_G - \delta(Y_G - Y_S)}{(1-\gamma_G)Y_G}\right) > 0\ . \tag{A.15}$$

Examination of (2.33) shows that there can be difficulties in r as $\gamma_G \uparrow 1$, because the term $\delta(Y_G - Y_S)/(1-\gamma_G)Y_G$ appearing in the terms a and A diverges as $\gamma_G \uparrow 1$.

In summary, from this analysis, we see that there is an upper limit on how high the death rate δ and the persistence γ can be. For the other parameters we consider, the maximal possible death rate is $\delta = 0.07$.

A.7 Sensitivity Analysis for the DTMC Model

In this section we do more sensitivity analysis, expanding on the discussion in §2.7 in Chapter 2. We first carry out the calculations for the base case, using the parameter values derived in §2.6.4. Our model depends on three exogenous variables, δ, γ, σ. We at first emphasized how the lockup premium depends on the death rate δ. It is also important to investigate how the lockup premium depends upon γ and σ.

A.7.1 How the Lockup Premium Depends on γ (γ_G, γ_S) and σ

We now see how much the premium depends on the model parameters $\gamma(\gamma_G, \gamma_S)$ and σ. Table A.3 shows the model parameters for $\gamma = 0.4, 0.5$, and 0.6 and Figure A.3 shows the lockup premium for $\gamma = 0.4, 0.5$, and 0.6. The figure suggests that as the persistence factor γ decreases, the n-year lockup premium decreases. The DTMC model works for persistence factor as low as 0.1. However, the n-year lockup premium decreases to the amount lower than 0.5 percentage points for any n. Figure A.3 suggests that the estimation of γ is important, especially for small δ and large n.

We next consider two separate persistence factors, γ_G and γ_S in Table A.4 and the sensitivity of the lockup premium with respect to γ_G and γ_S. Note that in the third line of Table A.4, r is negative, which breaks down the DTMC model. Figure A.4 shows the lockup premium for parameters in Table A.4.

We lastly check the sensitivity of the lockup premium with respect to σ. Our TASS database analysis estimates σ of annual returns for each year is lower than 0.20 in most cases. We here highlight the sensitivity of the lockup premium for $\sigma = 0.05, 0.10$, and 0.15

Table A.3: Parameter value sets for γ ranging from 0.2 to 0.6

γ	δ	p	q	r	Y_G	Y_S	Y_D	Calculated σ
0.4	0.00	0.8147	0.4147	0.5853	0.067	-0.15	-0.20	0.1002
0.4	0.03	0.8147	0.4427	0.4360	0.067	-0.15	-0.20	0.1000
0.4	0.06	0.8147	0.4892	0.1877	0.067	-0.15	-0.20	0.0998
0.5	0.00	0.8456	0.3456	0.6544	0.067	-0.15	-0.20	0.1002
0.5	0.03	0.8432	0.3719	0.5030	0.0685	-0.15	-0.20	0.1001
0.5	0.06	0.8409	0.4207	0.2282	0.070	-0.15	-0.20	0.1001
0.6	0.00	0.8765	0.2765	0.7235	0.067	-0.15	-0.20	0.1002
0.6	0.03	0.8727	0.3029	0.5645	0.070	-0.15	-0.20	0.0997
0.6	0.06	0.8679	0.3590	0.2324	0.074	-0.15	-0.20	0.1002
0.2	0.06	0.7615	0.6298	0.0782	0.0637	-0.15	-0.20	0.1002
0.3	0.06	0.7879	0.5586	0.1374	0.0652	-0.15	-0.20	0.1000
0.4	0.06	0.8147	0.4892	0.1877	0.067	-0.15	-0.20	0.0998

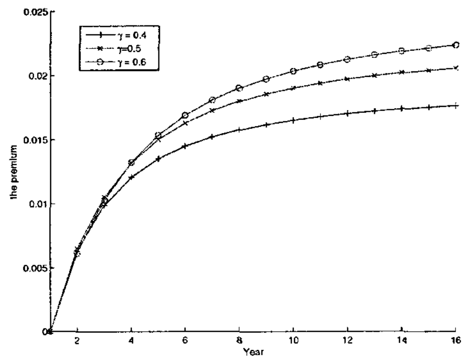
with $\gamma = 0.5$. Table A.5 is the parameter value sets and Figure A.5 is the corresponding lockup premium. We see that the premium increases in σ.

A.7.2 Sensitivity of p, q, and r with respect to δ, Y_G, Y_S, and Y_D

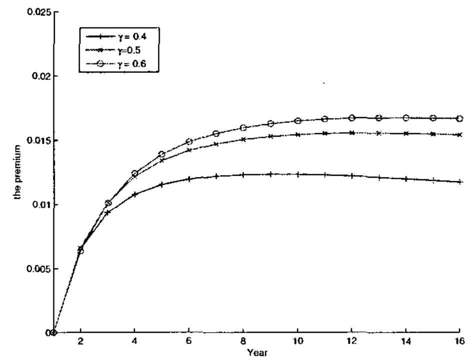
In this section, we observe the effect of δ to the implied transition probabilities p, q, r. Figure 2.6 is the implied transition probabilities for δ from 0 to 0.1. It is clear that the transition probability from sick to sick state, r, is the most sensitive to δ. Simple calculation of the partial derivative of r with respect to δ shows that

$$\frac{\partial r}{\partial \delta} = -\frac{\frac{2-p-r}{1-p}(Y_G - Y_D)}{(Y_G - Y_S) - \frac{\delta}{1-p}(Y_D - Y_D)} . \tag{A.16}$$

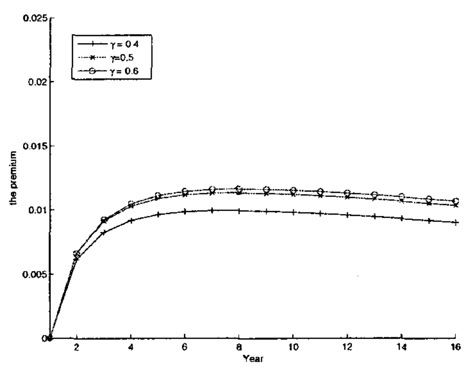
From (A.16), we observe that as δ increases, r decreases more rapidly. Note that the coefficient of δ in (A.16) is $(Y_G - Y_D)/(1-p) \approx 1$ thus its impact is big. We also observe that r becomes negative as δ increases above 0.07. Thus, the maximum allowable death



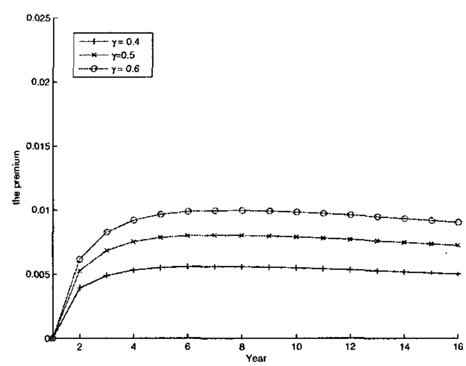
(a) With $\delta = 0$ for $\gamma = 0.4, 0.5, 0.6$



(b) With $\delta = 0.03$ for $\gamma = 0.4, 0.5, 0.6$



(c) With $\delta = 0.06$ and $\gamma = 0.4, 0.5, 0.6$



(d) With $\delta = 0.06$ and $\gamma = 0.2, 0.3, 0.4$

Figure A.3: The lockup premium for the DTMC model for parameter values in Tables A.3.

rate in the DTMC model is 0.07.

The implied transition probabilities are calculated for Y_G, Y_S, and Y_D in Figure 2.10 in §2.7. The plots show that p, q, and r are sensitive to Y_G, but that there is even more dependence upon γ, especially when $\gamma > 0.75$. The sensitivity of p, q, and r to Y_S and Y_D is much less, as is shown in Figure 2.10. This justifies that our parameter fitting method which changes Y_G while fixing Y_G and Y_D since Y_G has greater effect to the implied transition probabilities than Y_S and Y_D. Notice that from (2.30) of Chapter 2, p is independent of Y_D and a linear function of γ.

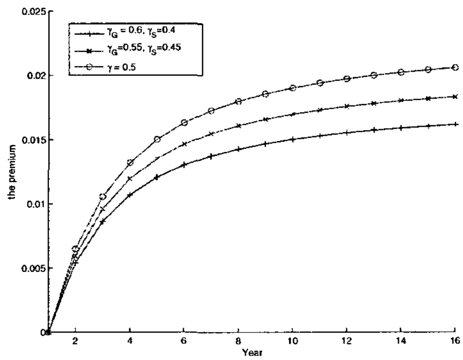
We next consider the sensitivity of the steady-state probabilities π_G and π_S to the model parameters. Up until the critical point in γ, the steady-state probabilities π_G and π_S are

Table A.4: Parameter value sets for γ_G and γ_S

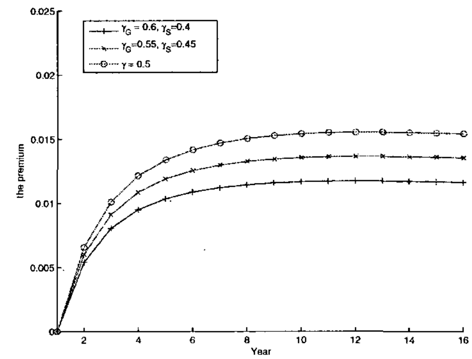
γ_G	γ_S	δ	p	q	r	Y_G	Y_S	Y_D	Calculated σ
0.6	0.4	0.00	0.8655	0.3982	0.6018	0.076	-0.15	-0.20	0.1000
0.6	0.4	0.03	0.8643	0.4320	0.4069	0.077	-0.15	-0.20	0.1002
0.6	0.4	0.06	0.8637	0.5068	-0.0127	0.775	-0.15	-0.20	0.1000
0.55	0.45	0.00	0.8547	0.3725	0.6275	0.0715	-0.15	-0.20	0.1000
0.55	0.45	0.03	0.8527	0.4014	0.4583	0.0730	-0.15	-0.20	0.1002
0.55	0.45	0.06	0.8513	0.4603	0.1276	0.074	-0.15	-0.20	0.1001
0.5	0.5	0.00	0.8456	0.3456	0.6544	0.067	-0.15	-0.20	0.1002
0.5	0.5	0.03	0.8432	0.3719	0.5030	0.0685	-0.15	-0.20	0.1001
0.5	0.5	0.06	0.8409	0.4207	0.2282	0.070	-0.15	-0.20	0.1001

Table A.5: Parameter value sets for $\sigma = 0.05, 0.10, 0.15$

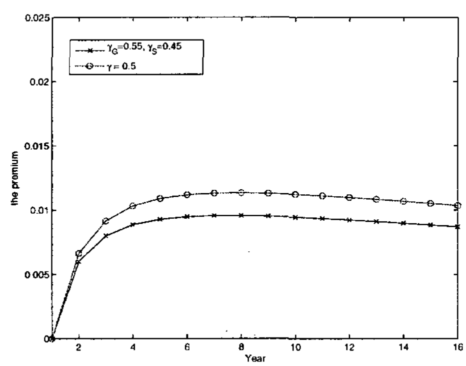
σ	δ	p	q	r	Y_G	$Y_S = -1.5\sigma$	$Y_D = -2.0\sigma$	σ (calculated)
0.05	0.00	0.8461	0.3461	0.6539	0.0333	-0.075	-0.10	0.05
0.10	0.00	0.8461	0.3461	0.6539	0.0667	-0.150	-0.20	0.10
0.15	0.00	0.8461	0.3461	0.6539	0.1000	-0.225	-0.30	0.15
0.05	0.03	0.8434	0.3721	0.5025	0.0342	-0.075	-0.10	0.05
0.10	0.03	0.8434	0.3721	0.5025	0.0684	-0.150	-0.20	0.10
0.15	0.03	0.8434	0.3721	0.5025	0.1026	-0.225	-0.30	0.15
0.05	0.06	0.8410	0.4209	0.2275	0.0350	-0.075	-0.10	0.05
0.10	0.06	0.8410	0.4209	0.2275	0.0699	-0.150	-0.20	0.10
0.15	0.06	0.8410	0.4209	0.2275	0.1049	-0.225	-0.30	0.15



(a) With $\delta = 0$



(b) With $\delta = 0.03$



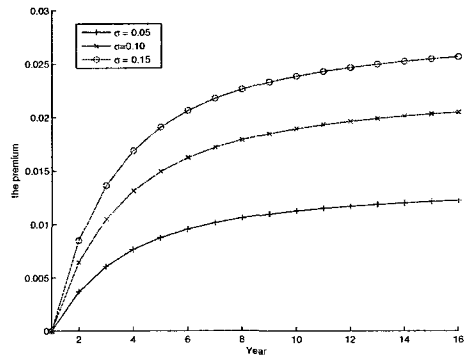
(c) With $\delta = 0.06$

Figure A.4: The lockup premium for the DTMC with $\gamma_G \neq \gamma_S$. $\gamma_G = 0.6, \gamma_S = 0.4$, $\gamma_G = 0.55 \gamma_S = 0.45$, and $\gamma_G = \gamma_S = 0.5$.

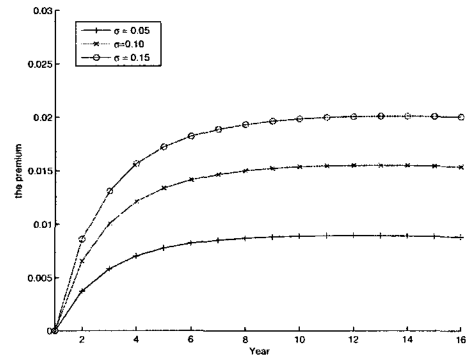
less sensitive to γ, Y_S, and Y_D, but is sensitive to Y_G, which can be regarded as a function of σ, as illustrated in Figure A.6.

A.7.3 Sensitivity of the Premium for a Fixed Lockup Period

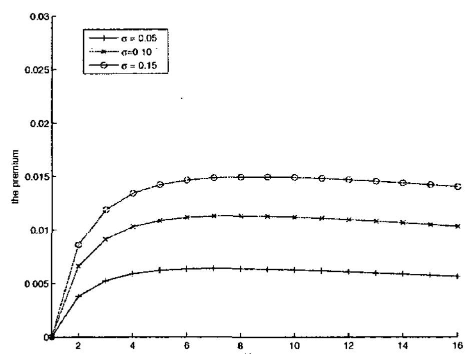
In this section, we investigate how the lockup premium for a fixed lockup period depends on the three variables δ, γ, and σ. We consider $n = 3$ with the choice of parameters $Y_S/\sigma = -1.5$ and $Y_D/\sigma = -2.0$. The result is helpful to estimate the lockup premium for a fixed lockup period when some of the three variables change. From Figure 2.9 in Chapter 2, it is clear that the lockup premium increases as δ decreases, γ increases, or σ increases.



(a) With $\delta = 0.00$



(b) With $\delta = 0.03$



(c) With $\delta = 0.06$

Figure A.5: The lockup premium for the DTMC for $\sigma = 0.05, 0.10$, and 0.15.

It turns out that the three-year lockup premium can be approximated reasonably well by a simple linear function of each variable separately, at least over a narrow range.

Figure A.7 shows the three-year lockup premium as a function of σ, γ, and δ for the DTMC models. In this figure, we see how the three-year lockup premium depends on two of the three variables σ, γ, and δ while fixing the remaining one variable. For example, Figure A.7 (i) shows the change of the three-year lockup premium for δ and γ while fixing σ as 0.1. We observe the near-perfect linearity of the three-year lockup premium for σ. We also observe that the concavity of the premium for γ increases as γ increases. Figures A.7 (i) and (ii) suggest that the three-year lockup premium is quite insensitive to δ, which implies that the effect of δ on the lockup premium is relatively small. Figure A.7 (iii) and (iv) show

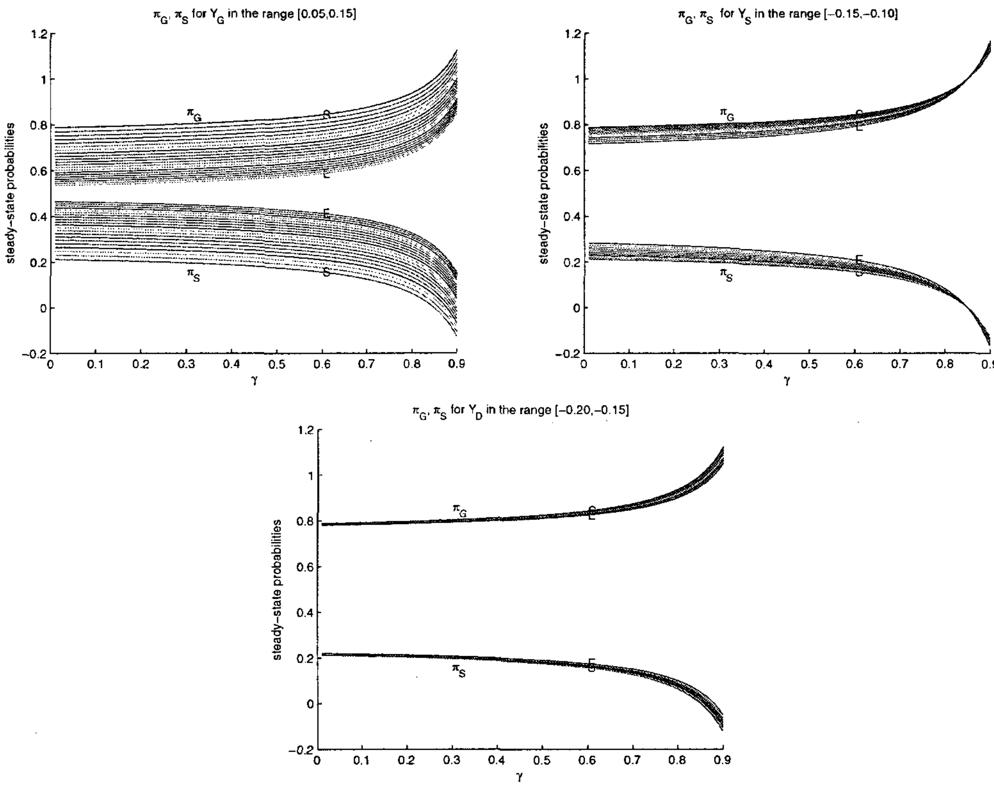
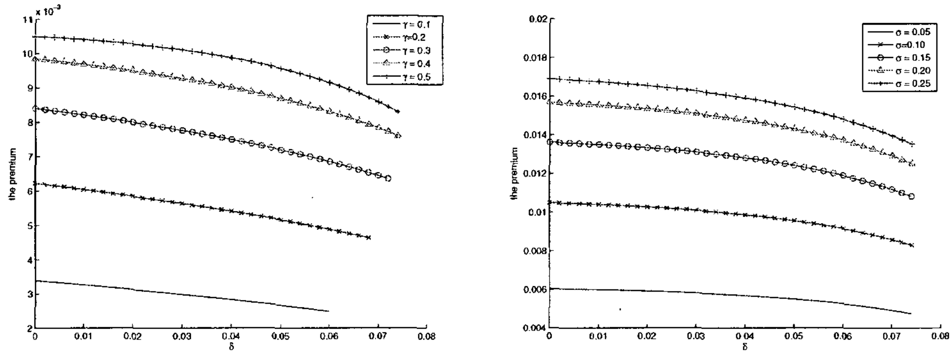


Figure A.6: The steady-state probabilities π_G and π_S as a function of γ in the base case for values of Y_G ranging from 0.05 (starting value, denoted by S) to 0.15 (ending value, denoted by E), Y_S ranging from -0.15 to -0.10, and Y_D ranging from -0.20 to -0.15.

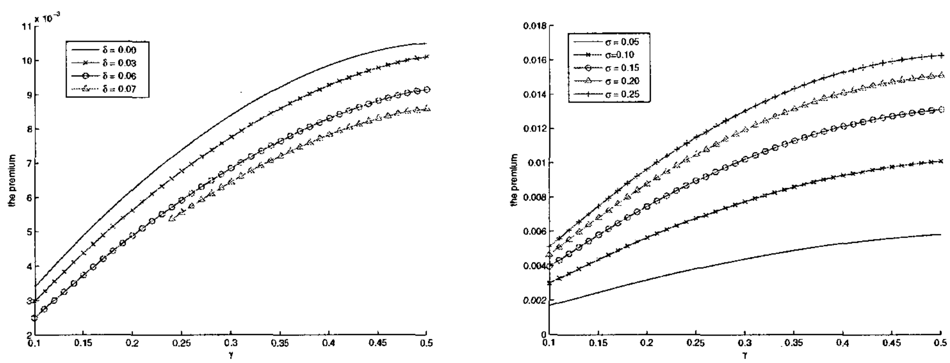
the three-year lockup premium for $\delta = 0.03$. To supplement that, Figure A.8 illustrates how the three-year lockup premium for different γ and σ changes with $\delta = 0.00$ and 0.06. We observe that the shape of the three-year lockup premium function does not change as δ changes.

A.7.4 Estimating the Functional Form of the Three-year Lockup Premium

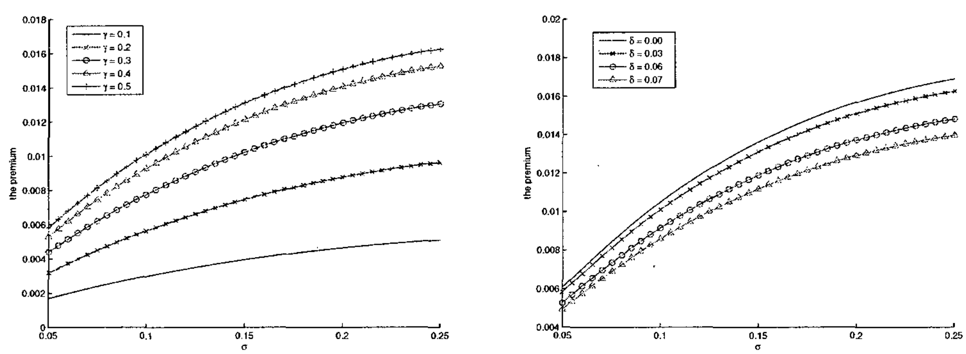
So far, we have calculated the lockup premium for variables γ, σ, and δ with the DTMC model. Since we have calculated the premium as a function of three variables, it is then natural to consider a simple functional form to describe the premium. If the estimation can be done relatively easily, it is useful to approximate the premium with a closed-form



For $\delta = 0.00$ to 0.08, (i) $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$ with $\sigma = 0.1$ (ii) $\sigma = 0.05, 0.1, 0.15, 0.2, 0.25$ with $\gamma = 0.5$



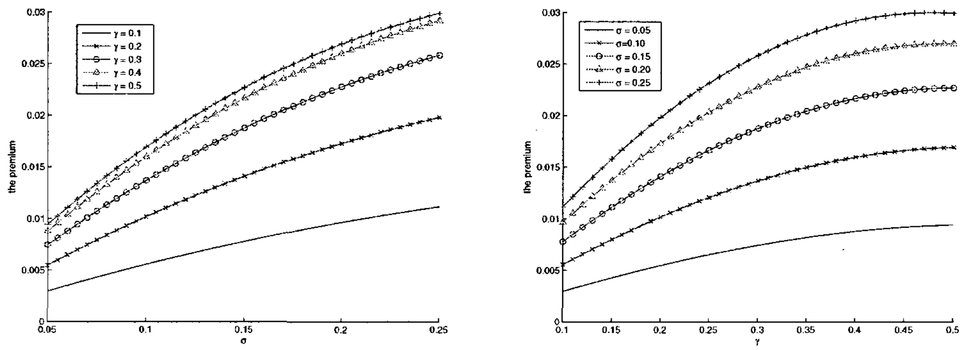
For $\gamma = 0.1$ to 0.5, (iii) $\delta = 0.00, 0.03, 0.06, 0.07$ with $\sigma = 0.1$ (iv) $\sigma = 0.05, 0.1, 0.15, 0.2, 0.25$ with $\delta = 0.03$



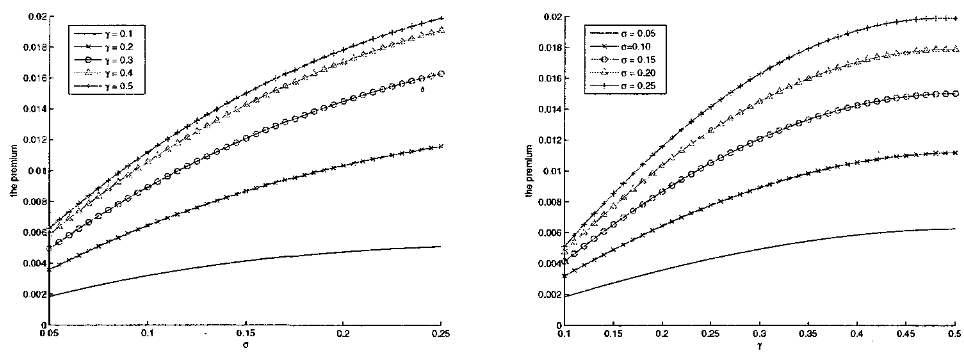
For $\sigma = 0.05$ to 0.25, (v) $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$ with $\delta = 0.03$ (vi) $\delta = 0.00, 0.03, 0.06, 0.07$ with $\gamma = 0.5$

Figure A.7: The three-year lockup premium for the DTMC model with $Y_S = -1.5\sigma, Y_D = -2.0\sigma$. The lockup premium does not exist if q or r becomes negative.

expression of three variables, denoted by $\psi(\delta, \gamma, \sigma)$. (Again, we fix $Y_S/\sigma = -1.5$ and $Y_D/\sigma = -2.0$.) We may then understand the effect of these three variables more intuitively.



For $\delta = 0.00$, (i) $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$ (ii) $\sigma = 0.05, 0.1, 0.15, 0.2, 0.25$



For $\delta = 0.06$, (iii) $\sigma = 0.05, 0.1, 0.15, 0.2, 0.25$ (iv) $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$

Figure A.8: The three-year lockup premium for the DTMC model with $Y_S = -1.5\sigma, Y_D = -2.0\sigma$.

It is also easy to quickly estimate how the premium changes if the variables change.

The three-year lockup period is interesting because this is the first year a fund starting in a good state may becomes dead in the DTMC model. Thus, we can see the effect of the death of a fund on the lockup premium. Furthermore, three years is a practical case to consider. Thus, we consider an estimation of three-year lockup premium as a closed form expression of γ, σ, and δ. However, the approximation also works for different lockup period and the choice of Y_S/σ. (See the remark below.)

Figure A.7 suggests that the three-year lockup premium is weakly concave function of γ, linear function of σ and relatively insensitive to δ. We thus try a simple product form of three variables with an exponent for each variable. Specifically, denoting the three-year lockup premium as a function of δ, γ, and σ by $\psi_{(3)}(\delta, \gamma, \sigma)$, we consider the following simple candidate approximation:

$$\psi_{(3)}(\delta, \gamma, \sigma) \approx \psi^p_{(3)}(\delta, \gamma, \sigma) \equiv a\ \delta^b \gamma^c \sigma^d\ . \tag{A.17}$$

Taking logarithms of both sides of (A.17), it is straightforward to estimate the parameters a, b, c, and d from the calculated three-year lockup-premium values with linear regression, because

$$\ln \psi^p_{(3)}(\delta, \gamma, \sigma) = \ln a + b \ln \delta + c \ln \gamma + d \ln \sigma\ . \tag{A.18}$$

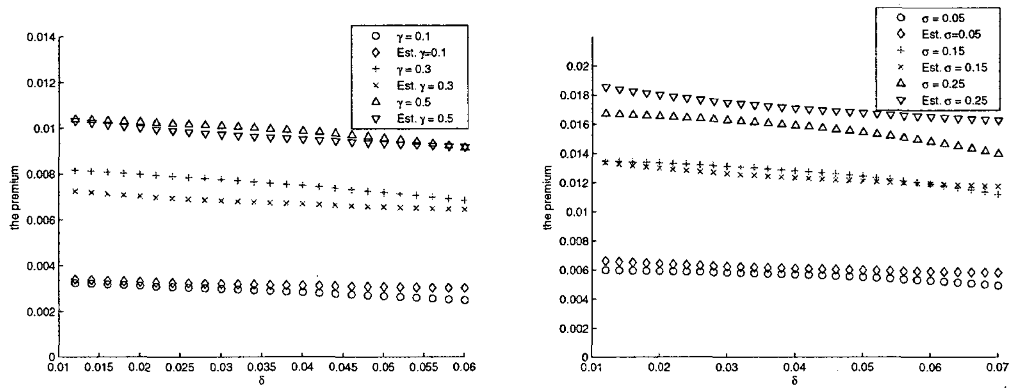
Since $\lim_{\delta \to 0} \psi^p_{(3)} = \infty$ when $b < 0$ and 0 when $b > 0$, which is not desirable for our estimation purpose, we have to restrict range of δ away from 0. Thus, we restrict the range of δ to [0.01.0.07].

It turns out that without further restricting the ranges of the variables δ, γ and σ, the candidate function $\psi^p_{(3)}(\delta, \gamma, \sigma)$ approximates the three-year lockup premium reasonably well. For example, for $\delta \in [0.01, 0.07]$, the linear regression of (A.18) approximates the three-year lockup premium by

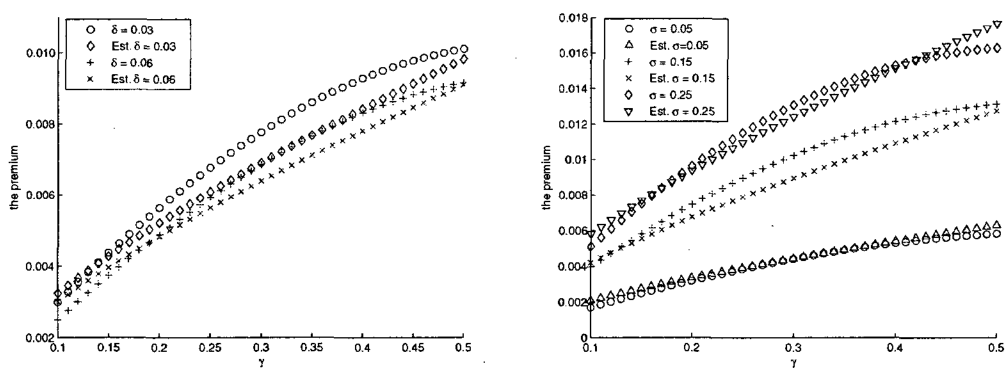
$$\psi^p_{(3)}(\delta, \gamma, \sigma) = 0.047\ \delta^{-0.11} \gamma^{0.69} \sigma^{0.64},$$

with maximum error of 0.0036. Notice that the exponent to δ is -0.11, which eventually makes $\lim_{\delta \to 0} \psi^p_{(3)}(\delta, \gamma, \sigma) = \infty$. Thus, we expect that as δ approaches 0, e.g., for $\delta \ll 0.01$, the estimation function will not approximate the three-year lockup premium in DTMC model well. Figure A.9 shows the estimation of the three-year lockup premium with the function obtained from the regression above for the selected ranges of variables. We observe that the estimation function approximates the three-year lockup premium reasonably well. If we further restrict the ranges of the variables such that $\gamma \in [0.2, 0.4]$, the maximum error reduces to 0.0004, which is only less than 11% of the three-year lockup-premium values in the DTMC model.

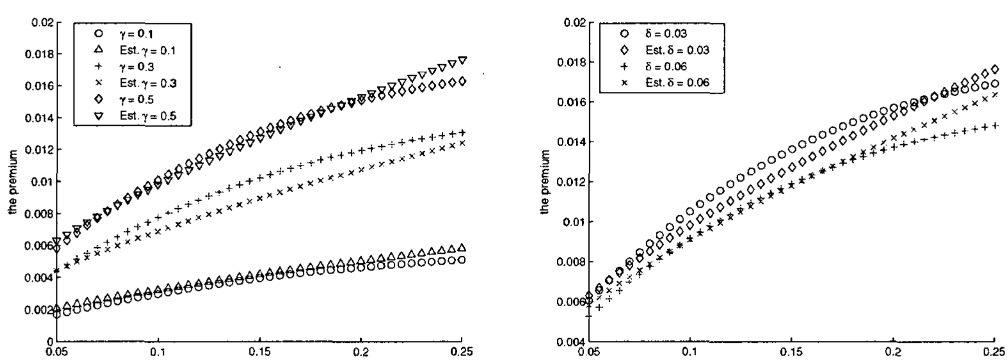
We remark that the approximation of the fixed-year lockup premium in the DTMC model by a product function $\psi^p(\delta, \gamma, \sigma) = a\ \delta^b \gamma^c \sigma^d$ works reasonably well for the other lockup periods (n) and other choices of Y_S/σ. For example, the four-year lockup premium in the DTMC model is approximated by $\psi^p_{(4)}(\delta, \gamma, \sigma) = 0.03\ \delta^{-0.20} \gamma^{0.69} \sigma^{0.64}$ with maximum error of 0.0043. If we choose $Y_S/\sigma = -1.0$, $\psi^p_{(3)}(\delta, \gamma, \sigma) = 0.06\ \delta^{-0.20}\gamma^{0.64}\sigma^{0.72}$ approximates the three-year lockup premium with maximum error of 0.0072, which can be reduced to 0.0024 if we restrict the range of γ, requiring that it be between 0.2 and 0.4. The approximation also holds reasonably well if we change n and Y_S/σ at the same time, although the maximum error increases slightly. As before, for $\delta \in [0.01, 0.07]$, if we choose $Y_S/\sigma = -1.0$, the four-year lockup premium in the DTMC model is approximated by $\psi^p_{(4)}(\delta, \gamma, \sigma) = 0.05\ \delta^{-0.26}\gamma^{0.65}\sigma^{0.73}$ with maximum error of 0.0084. Again, the error reduces to 0.0024 if we further restrict the range of γ, requiring that it be between 0.2 and 0.4.



For $\delta = 0.01$ to 0.07, (i) $\gamma = 0.1, 0.3, 0.5$ with $\sigma = 0.1$ (ii) $\sigma = 0.05, 0.15, 0.25$ with $\gamma = 0.5$



For $\gamma = 0.1$ to 0.5, (iii) $\delta = 0.03, 0.06$ with $\sigma = 0.1$ (iv) $\sigma = 0.05, 0.15, 0.25$ with $\delta = 0.03$



For $\sigma = 0.05$ to 0.25, (v) $\gamma = 0.1, 0.3, 0.5$ with $\delta = 0.03$ (vi) $\delta = 0.03, 0.06$ with $\gamma = 0.5$

Figure A.9: Evaluating the product approximation $\psi^{p}_{(3)}(\delta, \gamma, \sigma) = 0.047\ \delta^{-0.11}\gamma^{0.69}\sigma^{0.64}$ for the three-year lockup premium

Appendix B

Appendix to Chapter 3

B.1 Introduction and Summary

This appendix has nine more sections. In §B.2 we display plots of the sizes of the managed assets of the funds in our sample. In §B.3, we provide the relative-return distributions of hedge funds across 10 strategies in the TASS database from 2000-2005. It is observed that the relative-return distributions for some strategies are approximately normal, while others have high peaks or heavy tails, which is not fit by the normal distribution. In §B.4, we show how the relative-return distribution in the constant-persistence normal-noise model depends on the sample size of the simulation. We compare simulations with the sample size of the data to larger simulations with sample size of 10^6.

We supplement the analysis of the other models in the remaining sections. In §B.5, we show how the beta-persistence model depends on the beta-distribution parameters α and β. It is shown that the shape of the estimated relative-return distribution is insensitive to α and β. In §B.6, we show that the heavy-tail and light-tail distributions behave differently in log-log scale. In §B.7, we show that the beta-persistence empirical-noise model provides a good fit the the data and reasonable estimates to the hitting probabilities. In §B.8, we show how the tails of the relative-return distribution in the constant-persistence stable-noise model behave, depending on the parameter β in the stable distribution. It is observed that the estimated relative-return distribution fits the data reasonably well for fund-of-fund strategy when $\beta = 0$. In §B.9, we provide a fitting for long-short equity strategy, which has the largest sample size in the data. We conclude that the beta-persistence normal-noise model fits the data well. Finally, in §B.10, we provide a fitting for the event-driven strategy whose relative-return distribution has heavy tails. It is observed that the beta-persistence t-noise model and constant-persistence stable-noise model provides a good fit to the data.

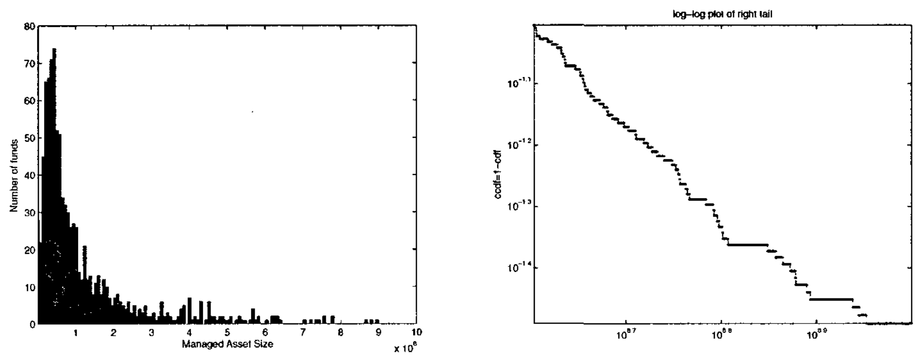
B.2 The Values of Managed Assets

As described in §3.4 of Chapter 3, we started by examining the TASS data. We followed the previous researchers, such as Boyson and Cooper (2004), in our data selection procedure. For each strategy, in order to avoid very small funds, which might have different characteristics, we first removed all funds from the data for which the managed asset value never reaches our 25 million dollar threshold. For the fund-of-fund strategy, we first removed 407 fund pairs from the data; that left the 986 fund pairs in our sample. (A pair is the relative annual returns for two successive years.)

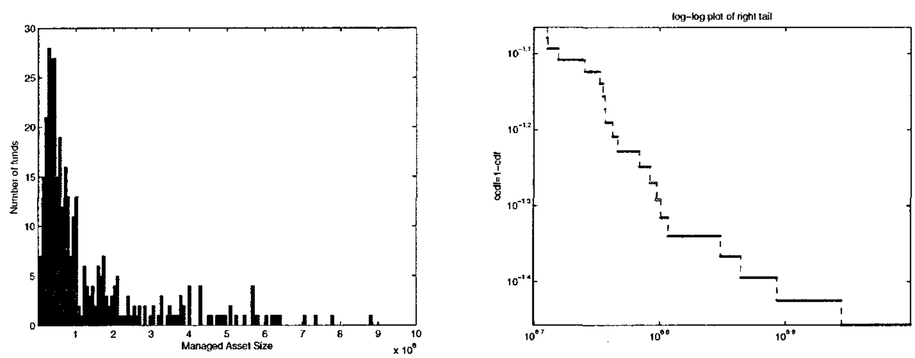
To further explore the data, we considered the distribution of the average asset values managed by the fund. In Figure B.1 (a) below, we plot the histogram of the average managed asset value among the the 986 funds in the fund-of-fund strategy. These 986 observations are taken only from the funds exceeding the 25 million dollar threshold. We see that the largest managed asset values are of order $\$10^8$. We also show a corresponding log-log plot in Figure B.1 (b), which shows that the size distribution has a heavy tail.

We also measure the total value of asset managed by the larger and smaller funds (in terms of managed asset values) in Table B.1. We first study the total value of asset managed for all 986 returns observed for fund-of-fund strategy. Since the relative returns from 2000 to 2004 are included at the same time for all 986 observations, asset values of some funds are counted multiple times for their life during the period. Thus, we also choose one specific year, namely, 2004, and take a snapshot of that year in terms of asset size such that we can see how the asset size of each fund, not the returns over the years, is distributed in one year.

The table shows that top 10% funds constitute large portion of total asset values, up to 65%. It also shows that the percentage of total asset values in two methods are not



(a) Histogram and log-log plot of the asset value managed by funds (all)



(b) Histogram and log-log plot of the asset value managed by funds (2004)

Figure B.1: Histogram and log-log plot of the value of managed assets for funds under the fund-of-fund strategy.

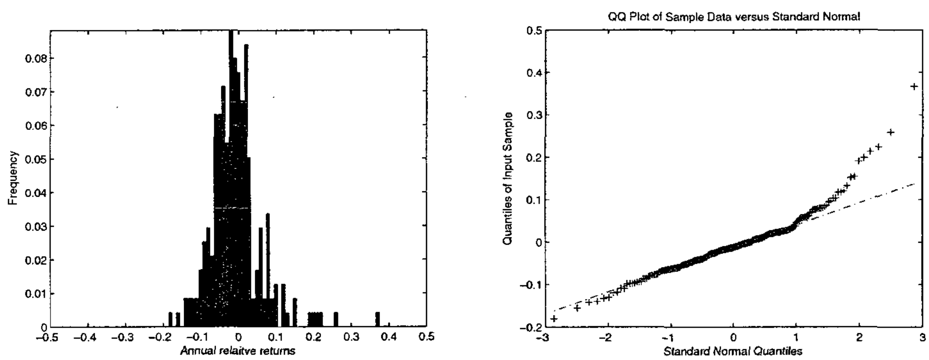
significantly different. Although the 65% is not small, we believe that this is not an extreme value such that we need some other measure to analyze the relative returns under the same strategy.

Table B.1: Managed asset values for fund-of-fund strategy

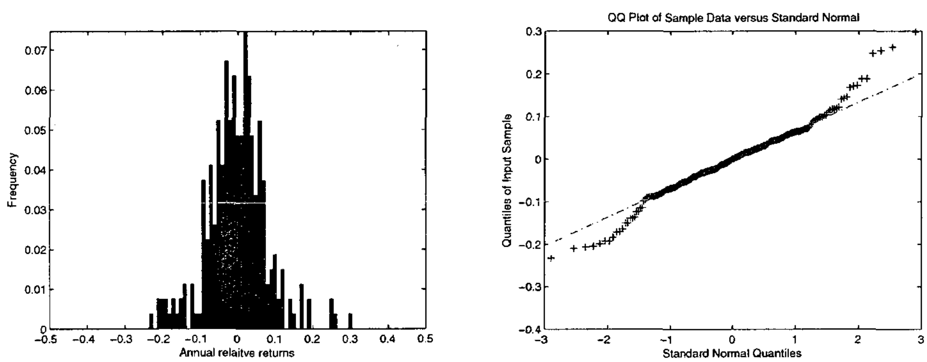
Ranking	Managed asset	Manages asset for 2004
Top 1%	33 %	38 %
Top 5%	55 %	58 %
Top 10%	65 %	67 %
Bottom 10%	0.5 %	1 %
Bottom 5%	0.1 %	0.2 %
Bottom 1%	0 %	0 %
Total Managed Asset	$\$1 \times 10^{11}$	$\$2 \times 10^{11}$

B.3 Distribution of Relative Returns from the Data

In this section, we carry out the analysis of Figure 3.1 in Chapter 3 for the other hedge-fund strategies. In particular, we display histograms of the relative returns within each of these strategies and provide Q-Q plots comparing the empirical distribution to the normal distribution. It is pointed out by Lhabitant (2004), Tran (2006), Geman and Kharoubi (2003), Eling and Schuhmacher (2007), Kassberger and Kiesel (2006) that hedge fund returns or indexes have heavy-tails, which are not fitted by normal distribution. In contrast, although most returns do indeed show heavy tails, we find that relative returns within the global-macro and emerging-market strategies can be fit to the normal distribution; see Figure 3.1 in Chapter 3 and Figure B.2 below. (We omit dedicated-short-biased strategy since we only have 29 observations.)

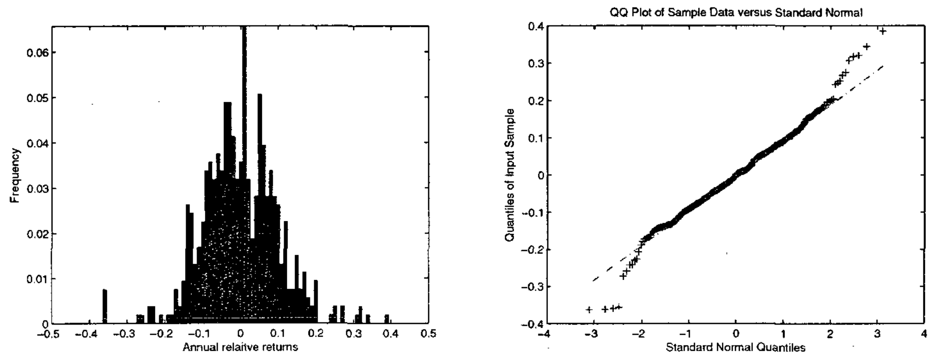


(a) Convertible ($\gamma = 0.44$)

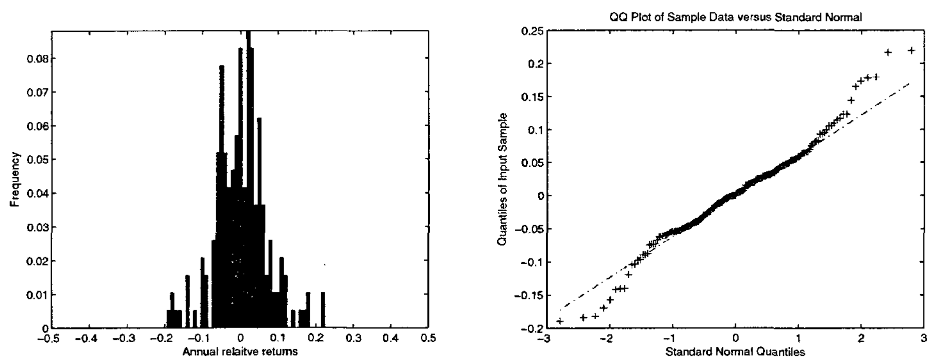


(b) Equity macro ($\gamma = 0.09$)

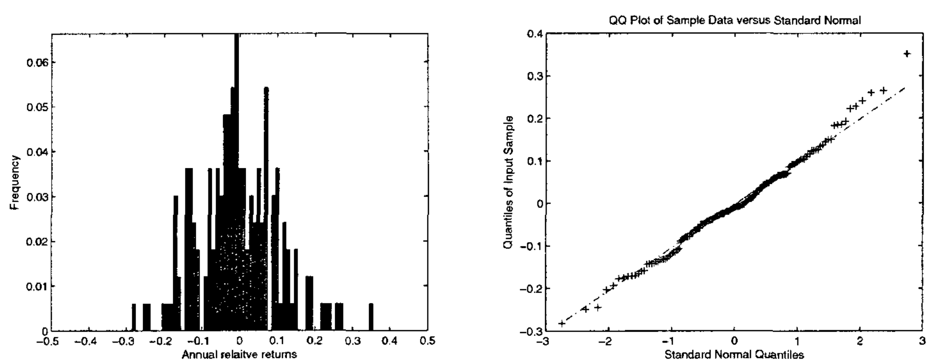
Figure B.2: Relative-return distributions and Q-Q plots comparing the empirical distribution to the normal distribution for 10 strategies in TASS database from 2000-2004.



(c) Event driven ($\gamma = 0.24$)

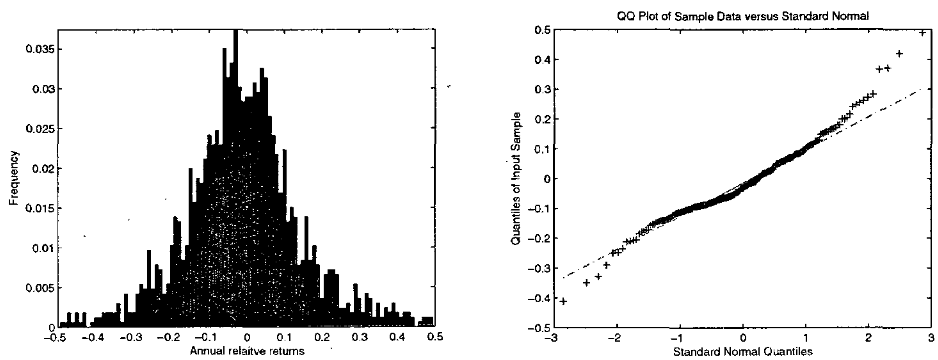


(d) Fixed income ($\gamma = 0.29$)

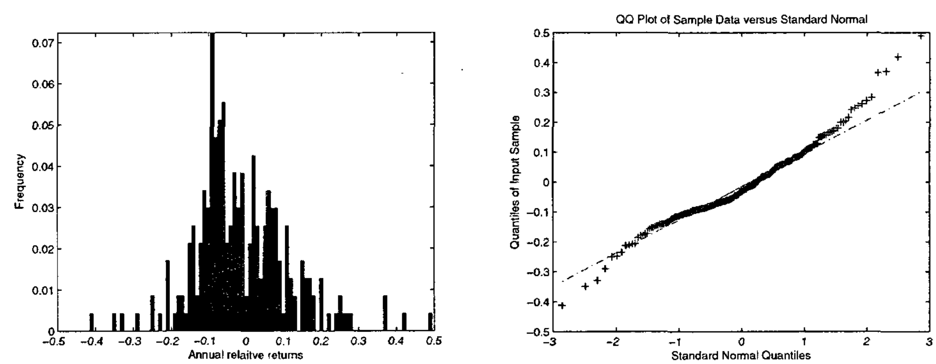


(e) Global macro ($\gamma = 0.13$)

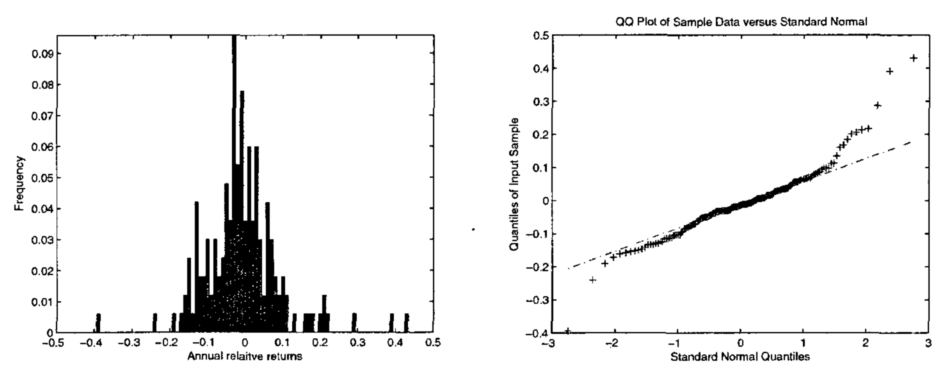
Figure B.2: (Continue) Relative-return distributions and Q-Q plots comparing the empirical distribution to the normal distribution for 10 strategies in TASS database from 2000-2004.



(f) Long-short equity ($\gamma = 0.15$)



(g) managed future ($\gamma = 0.20$)



(h) Others ($\gamma = 0.48$)

Figure B.2: (Continue) Relative-return distributions and Q-Q plots comparing the empirical distribution to the normal distribution for 10 strategies in TASS database from 2000-2004.

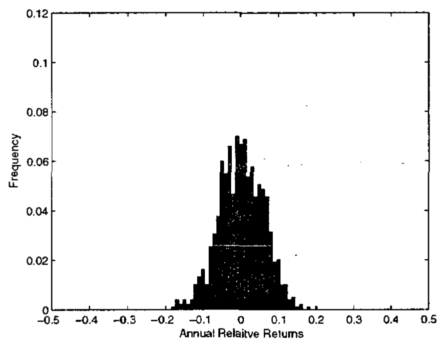
The table below shows results for the Lilliefors test. It tests the hypothesis that the sample comes from a normal distribution. The two distributions with relatively high p-values (greater that 0.05) from the Lilliefors test have distributions the look like the normal distribution in Figure B.2, both directly and in the Q-Q plot .

Table B.2: Lilliefors test results with 95 % significance level

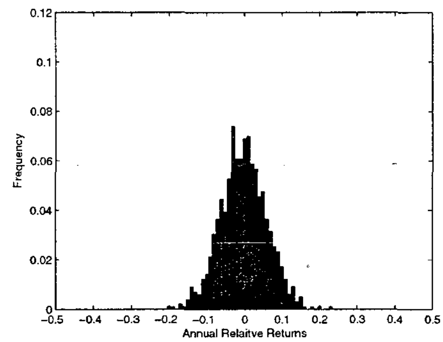
Strategies	Result	p-value
Convertible	Reject	0.0001
Equity Macro	Reject	0.0071
Event Driven	Accept	0.1204
Fixed Income	Reject	0.0424
Global Macro	Accept	0.3002
Long-short Equity	Reject	0.0001
Managed Future	Reject	0.0021
Other	Reject	0.0001

B.4 Constant-Persistence Normal-Noise Model Simulation

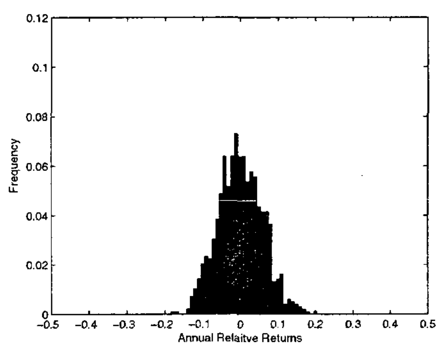
In this section, we show how the relative-return distribution in the constant-persistence normal-noise model depends on the sample size of the simulation. Since the relative returns we have from the data is limited, when fitting the relative-return distribution, it might be helpful to compare the empirical distribution to the estimated distribution with the sample size of the data. Figure B.3 (a)-(c) illustrate estimated distributions, each with the same size of the data, 986, for the fund-of-fund strategy. We then do the same for the emerging-market strategy in Figure B.3 (e)-(g) with sample size of 315. We also provide the estimated relative-return distribution with sample size of 10^6 in Figure B.3 (d) and (h) in order to see how the shape of the estimated relative-return distribution changes as the sample size increases.



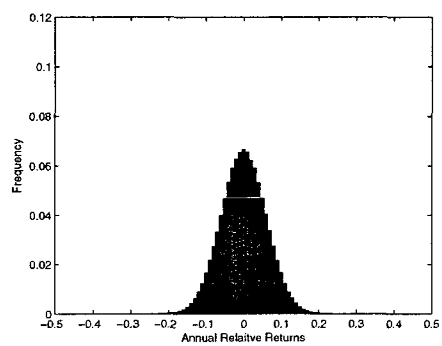
(a) 986 simulation of the model



(b) Relative-return distribution from 986 samples

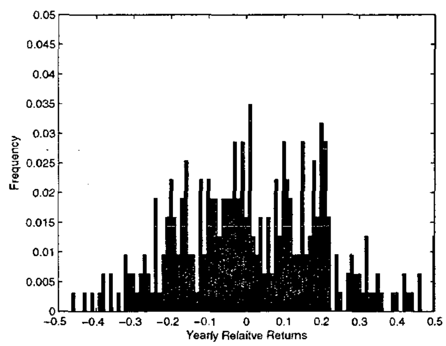


(c) Relative-return distribution from 986 samples

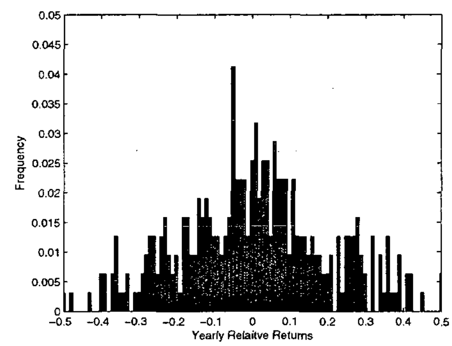


(d) Relative-return distribution from 10^6 samples

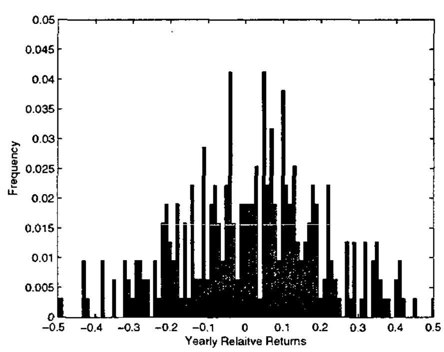
Figure B.3: (a)(b)(c) The estimated relative-return distribution with the sample size of 986 in the constant-persistence normal-noise model with $\gamma = 0.33, \sigma_b = 0.0565$ for fund-of-fund strategy. (d) The estimated relative-return distribution with the sample size of 10^6.



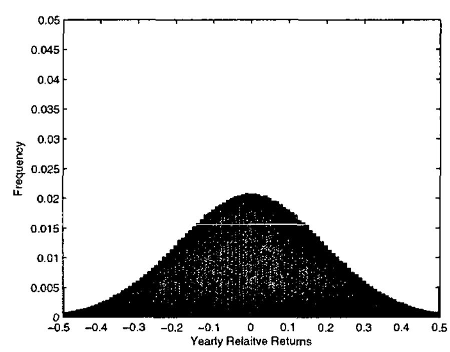
(e) 315 simulation of the model



(f) 315 simulation of the model



(g) 315 simulation of the model

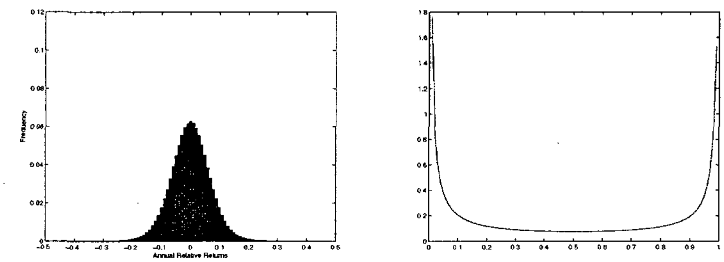


(h) 1,000,000 simulation of the model

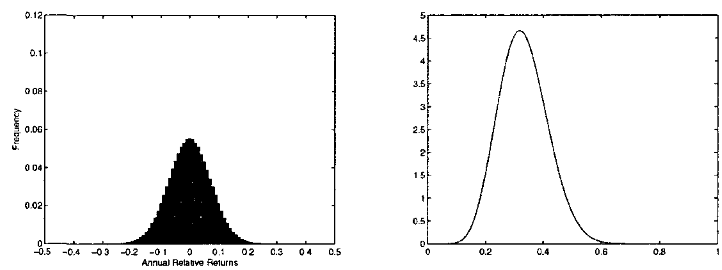
Figure B.3: (Continue)(e)(f)(g) The estimated relative-return distribution with the sample size of 315 in the constant-persistence normal-noise model with $\gamma = 0.36, \sigma_b = 0.1797$ for emerging-market strategy (h) The estimated relative-return distribution with the sample size of 10^6.

B.5 Beta-Persistence Model Simulations

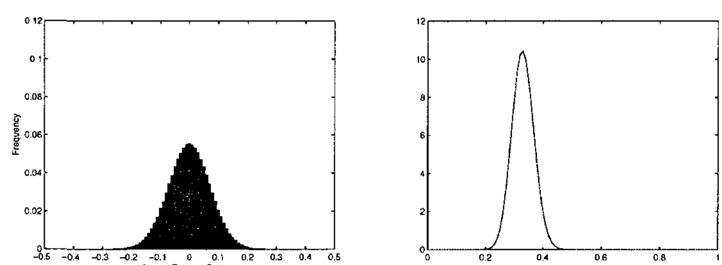
In this section, we illustrate how the beta-persistence model depends on the beta-distribution parameters α and β. It is observed that the overall relative-return distribution predicted by the model does not depend much on beta-distribution parameters. See, Figure B.4 for the beta-persistence normal-noise model. The observation also holds for the other beta-persistence models with t and mixture noise.



(a) The beta-persistence normal-noise model with $\alpha = 0.03$ and corresponding beta PDF



(b) The beta-persistence normal-noise model with $\alpha = 10$ and corresponding beta PDF

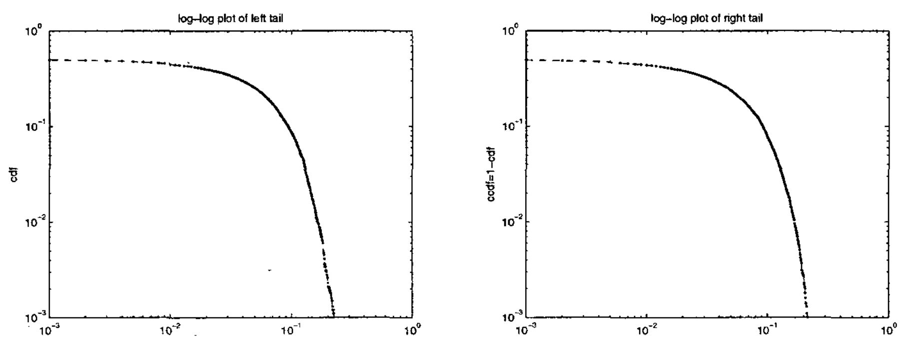


(c) The beta-persistence normal-noise model with $\alpha = 50$ and corresponding beta PDF

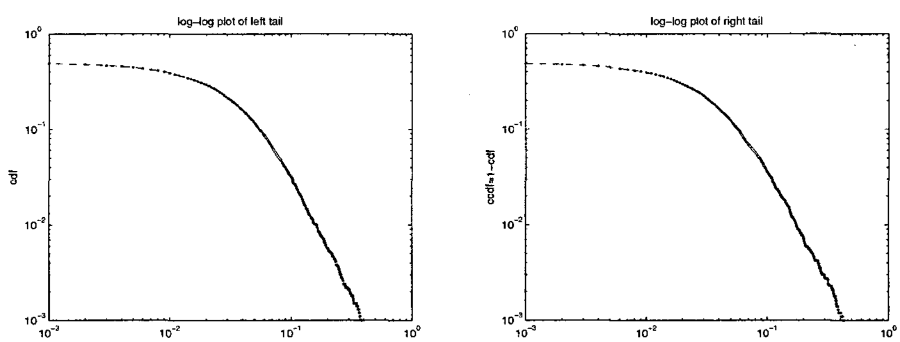
Figure B.4: Simulation estimate of the relative-return distribution and the associated beta pdf from the beta-persistence normal-noise model for the fund-of-fund strategy with $\gamma = 0.33$, $\sigma = 0.0681$ and (a) $\alpha = 0.03$ and $\beta = 0.06$, (b) $\alpha = 10$ and $\beta = 20.30$, (c) $\alpha = 50$ and $\beta = 101.51$.

B.6 Log-Log Plots of Distribution Tails in Different Models

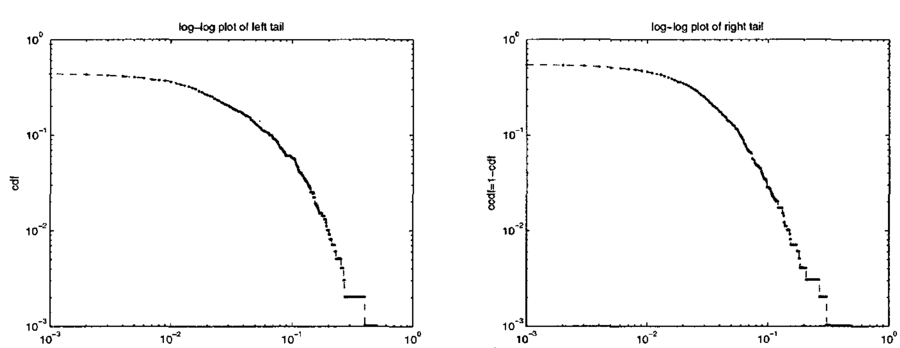
In this section, we plots the distribution tails for the normal, t, and mixture noise model in order to show the differences in their tail behavior. All except the normal have heavy tails, which is shown as linear behavior for larger values (at the right in each plot) in Figure B.5.



(a) 10,000 simulation of constant-persistence normal-noise model



(b) 10,000 simulation of constant-persistence t-noise model

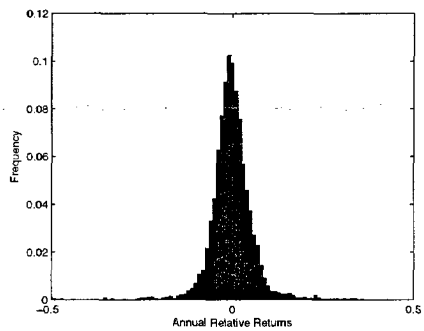


(c) 10,000 simulation of beta-persistence mixed-noise model

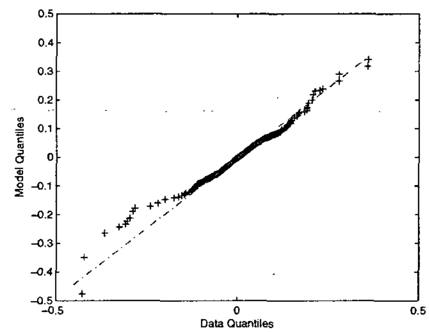
Figure B.5: Log-log plots of the estimated relative-return distributions with sample size of 10^4 in the (a) constant-persistence normal-noise model, (b) constant-persistence t-noise model, (c) constant-persistence mixed-noise model.

B.7 The Beta-Persistence Empirical-Noise Model

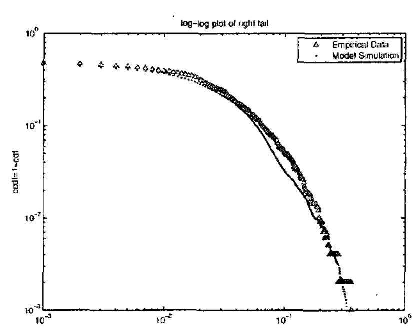
To seek a still better fit to the data within the beta-persistence class of models, we can let B_n have the observed empirical distribution for $X_n - \gamma X_{n-1}$, using the estimated value of γ. This automatically gives B_n and its estimated variance σ_b^2. It now goes further to directly match the shape, but sacrifices the explicit parametric form. In order to simulate B following the same distribution of B_n obtained from the data, we construct distribution function of B_n numerically. This is done by splitting the support of relative returns, $[-0.5, 0.5]$ equally and cumulatively count the number of returns falling each interval, from left to the right. As a numerical example, we construct distribution function of B_n from the relative returns within fund-of-fund strategy. Given the distribution function, we can generate B using inverse transform method; we generate uniform random variable and find the inverse value of given distribution function numerically. Figure B.6 shows the distribution function of X based on the simulation of B constructed from empirically obtained B_n. As we see from the figure, the beta-persistence empirical-noise model also provides a good fit to the data.



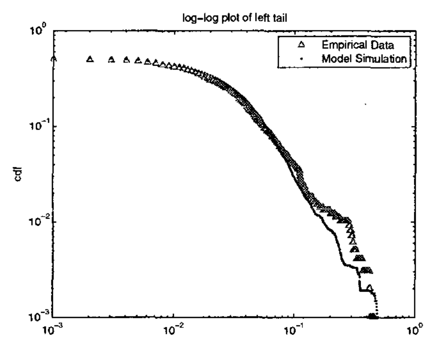
(a) Relative-return distribution



(b) Q-Q plot comparing the model to the data



(c) Left-tail log-log plot



(d) Right-tail log-log plot

Figure B.6: Simulated samples from the beta-persistence empirical-noise model with $\gamma = 0.33$, $\alpha = 50$, $\sigma = 0.0681$ and the empirical relative-return distribution for the fund-of-fund strategy from the data.

Table B.3 below shows the hitting probabilities from the beta-persistence empirical-noise model. It is observed that the maximum and minimum of 20 simulations of hitting probabilities cover the empirically observed hitting probabilities from the data. The large number (10^4) of simulation results in the fourth column of Table B.3 also suggests that the beta-persistence empirical-noise model provides reasonable estimates of the hitting probabilities.

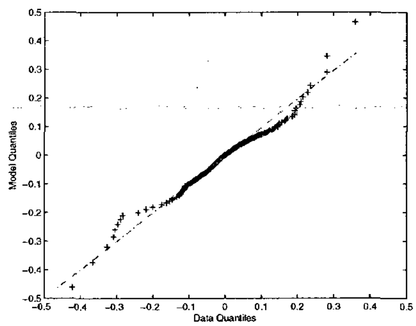
Table B.3: Hitting probabilities of thresholds over a five-year period (2000-2004)

Level[1]	data[2]	empirical-noise	
		$N = 92$[3]	$N = 10,000$[4]
3 σ	0.0326	[0,0.0652]	0.0313±0.0034
2 σ	0.0761	[0.0217,0.1196]	0.0659±0.0049
1 σ	0.2363	[0.1630,0.3261]	0.2226±0.0082
-1 σ	0.2391	[0.1413,0.2826]	0.2021±0.0079
-2 σ	0.0542	[0.0109,0.0870]	0.0477±0.0042
-3 σ	0.0326	[0,0.0543]	0.0271±0.0032

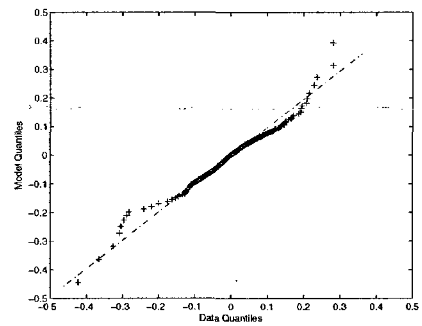
1. $\sigma = 0.0681$, the observed standard deviation of the fund-of-fund relative returns.
2. Number of funds that have ever hit the level for 2000-2004 divided by total 92 funds in 2000.
3. Minimum and maximum of the probabilities from 20 simulations with sample size of 92 initially.
4. 95 % confidence interval of hitting probability from simulation with sample size of 10,000 initially

B.8 Constant-Persistence Stable-Noise Model Simulations

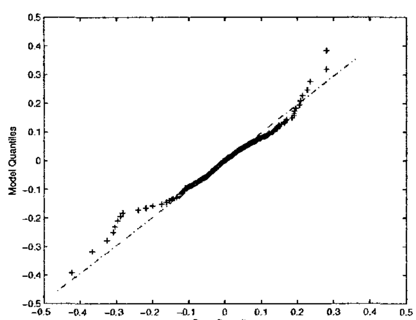
In this section, we show how the relative-return distribution in the constant-persistence stable-noise model depends on parameter β in the stable distribution. Figure B.7 shows Q-Q plots and log-log plots of the left and right tails of the estimated distributions for $\beta = -0.2, -0.1, 0$, and 0.1. It is observed that the constant-persistence stable-noise model with $\beta = -0.1$ fits the Q-Q plot well whereas the left and right tails of the distribution are approximated well with $\beta = 0.1$. Overall, $\beta = 0$ fits both the Q-Q plot and the left and right tails relatively well at the same time. It is hard to find stable random variable parameters that can fit both Q-Q plot and log-log figures at the same time. It is because the shape of the stable distribution cannot match the observed distribution exactly. However, the constant-persistence stable-noise model still provides a reasonably good fit to the data with fewer parameters than the other models, such as the beta-persistence mixed-noise model.



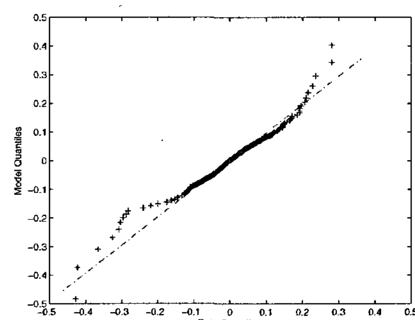
(a) Q-Q plot with $\beta = -0.2$



(b) Q-Q plot with $\beta = -0.1$

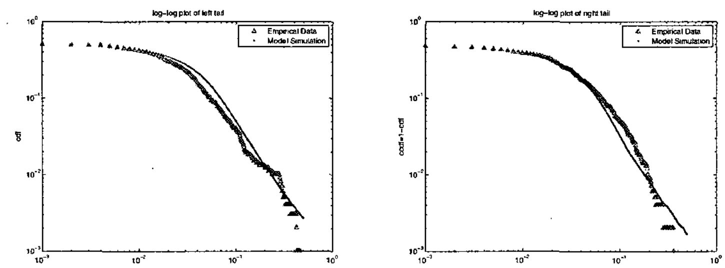


(c) Q-Q plot with $\beta = 0$

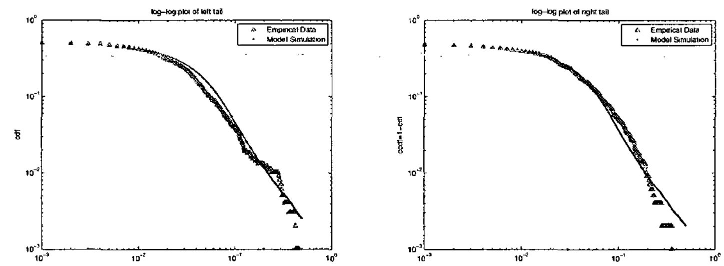


(d) Q-Q plot with $\beta = 0.1$

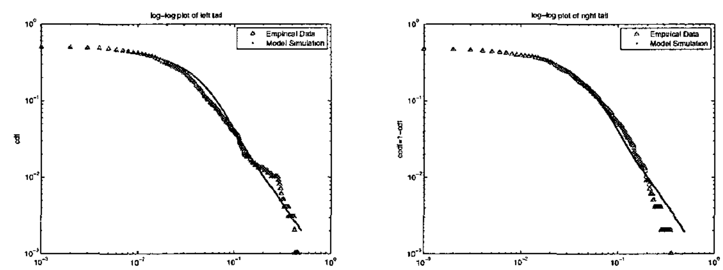
Figure B.7: Q-Q plots and Log-log plots of left and right tails of the relative-return distributions from the constant-persistence stable-noise model with $\alpha = 1.6, k = 0.0029$ for $\beta = -0.2, -0.1, 0$, and 0.1.



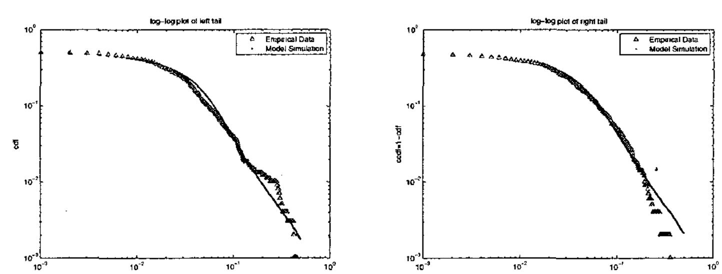
(e) Constant-persistence normal-noise model with $\beta = -0.2$



(f) Constant-persistence normal-noise model with $\beta = -0.1$



(g) Constant-persistence normal-noise model with $\beta = 0$



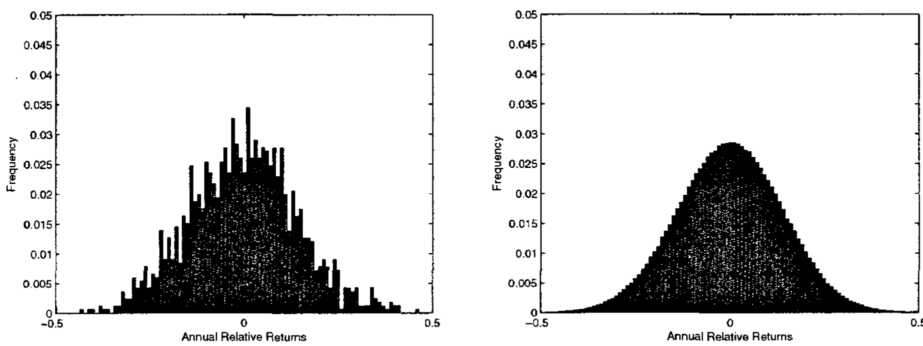
(h) Constant-persistence normal-noise model with $\beta = 0.1$

Figure B.7: (Continued) Q-Q plots Log-log plots of left and right tails of the relative-return distributions from the constant-persistence stable-noise model with $\alpha = 1.6, k = 0.0029$ for $\beta = -0.2, -0.1, 0$, and 0.1.

B.9 Analysis of Relative Returns within the Long-Short Equity Strategy

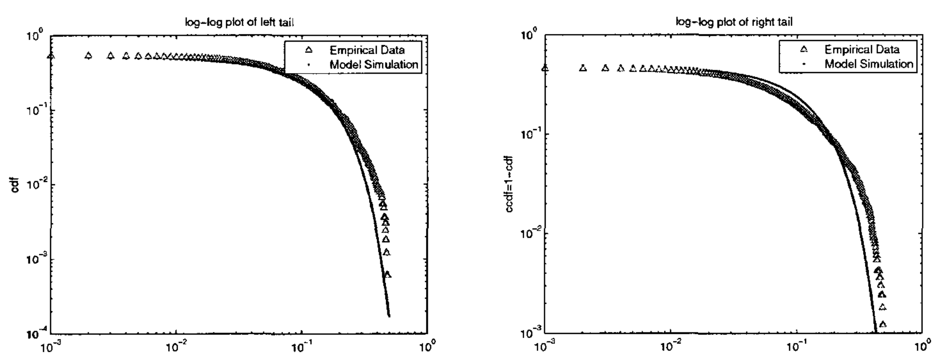
In this section, we fit the relative returns within the long-short equity strategy. Table 3.1 in Chapter 3 shows that this strategy has the largest sample size. Thus it is natural to fit our SDE model to the data in this case. Although we observe relative large number of observations from the data for this strategy, we see that the relative returns does not have high performance persistence.

The Q-Q plot of the relative returns in Figure B.2 (f) suggests that the distribution does not have heavy tails. That is also supported in the log-log plots of the distribution tails in Figure B.8 since both the left and right tails do not end with a linear line and instead decrease quickly in the right side of the Figure B.8 (c). Thus, we start from normal-noise model to fit the data. As observed in Table 3.2 in Chapter 3, the ratio σ/σ_b from the data and model do not match. Thus, we use the beta-persistence normal-noise model first with $\alpha = 50$. For given $\sigma = 0.1520$ and $\gamma = 0.15$ from the data, we calibrate other parameters β, σ_a and σ_b, following §3.6 of Chapter 3. Figure B.8 (a) and (b) show the estimated relative-return distribution. It is observed from Figure B.8 (d) that the Q-Q plot of the model to the data is close to a linear line with slope 1. Thus, we conclude that the relative-return distribution is approximated well by the beta-persistence normal-noise model for the long-short equity strategy.

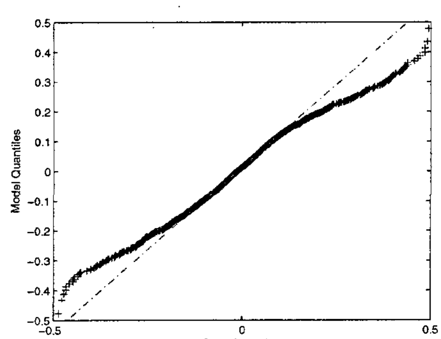


(a) Relative-return distribution with 1658 samples

(b) Relative-return distribution with 10^6 samples



(c) Log-log plot of the left and right tails from the beta-persistence normal-noise model

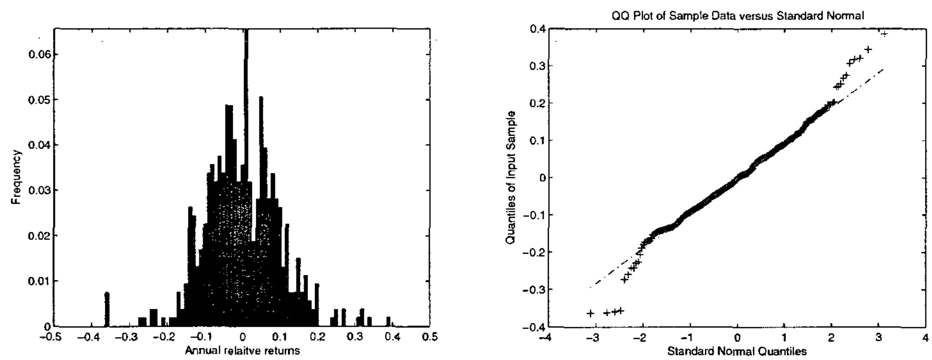


(d) Q-Q plot (simulated distribution to empirical one)

Figure B.8: Relative returns simulated from the beta-persistence normal-noise model with $\alpha = 50$, $\sigma = 0.1520$, $\gamma = 0.15$ for the long-short equity strategy.

B.10 Analysis of Relative Returns within the Event-Driven Strategy

In this section, we analyze another single strategy whose relative-return distribution has heavy tails. In particular, we analyze the event-driven strategy since it has relative big sample size (533) and high persistence factor ($\gamma = 0.24$). The Q-Q plot in Figure B.9 shows that the relative-return distribution has heavier tails than a normal distribution. We thus proceed using our beta-persistence t-noise and constant-persistence stable-noise models to fit the data.



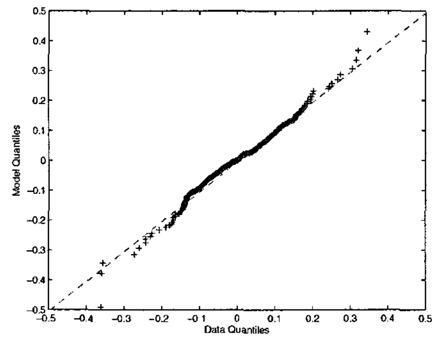
(a) Event-driven strategy ($\gamma = 0.24$)

Figure B.9: Distribution of relative returns from event-driven strategy and Q-Q plot comparing the distribution to the normal distribution.

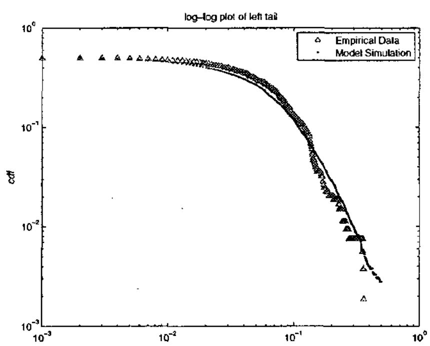
B.10.1 Beta-Persistence t-Noise Model

In this section, we test whether the beta-persistence t-noise model can fit the data for the event-driven strategy. Recall that in the beta-persistence t-noise model, once α is set, then the other parameter β in the beta random variable is determined to fit the mean ($\gamma = 0.24$). Just as we did for the fund-of-fund strategy, we set $\alpha = 50$, so that the persistence random variable is relatively narrowly distributed around $\gamma = 0.24$. We then set the degrees of freedom in the t random variable to fit the distribution of relative returns from the data. Another parameter k in the model is determined to fit the standard deviation of X_n ($\sigma = 0.1007$). We find that $v = 3.5$ fits the distribution well.

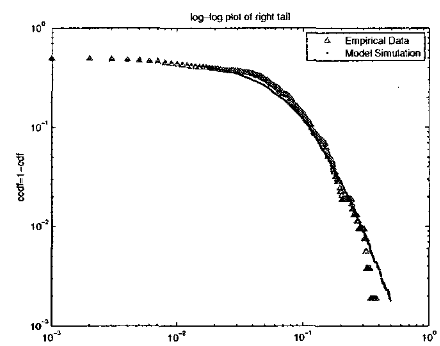
From Figure B.10, we observe that the quantiles in the Q-Q plot comparing the samples from the model to the data coincide reasonably well. We obtain p value of 0.1349 from Kolmogorov-Smirnov two sample test. Thus, we cannot reject the hypothesis that the simulated returns and empirical returns come from the same distribution.



(a) Q-Q plot comparing the model the the data



(b) Log-log plot of left tails



(c) Log-log plot of right tails

Figure B.10: The beta-persistence t-noise model 10^4 number of simulation with $\alpha = 50, v = 3.5, \gamma = 0.24$ comparing to the data for event-driven strategy.

B.10.2 Constant-Persistence Stable-Noise Model

In this section, we test whether the constant-persistence stable-noise model provides a good fit the data. In order to test that, we measure the quantiles of X_n and B_n that directly come from $X_n - \gamma X_{n-1}$, using previous estimate for the persistence factor γ. Table B.4 shows that the ratios of quantiles from X and B are roughly equal to 1.3. We thus proceed the model fitting by assuming that $c = 1.3$.

Given $c = 1.3$, we now compare X_n and cB_n from the data for the event-driven strategy. Figure B.11 shows the histograms of X_n and cB_n from the data, which look similar. We also conducted Kolmogorov-Smirnov two-sample test and obtained a p-value of 0.2834. Thus we cannot reject the hypothesis that these two sets of samples come from the same distribution. The Q-Q plot also shows that the quantiles from the distributions of the samples from the model and the data coincide with each other remarkably well.

Figure B.12 shows that the constant-persistence stable-noise model fits the relative returns within the event-driven strategy reasonably well with stable-distribution parameters

Table B.4: The Quantile Differences of X_n and B_n and Their Ratios

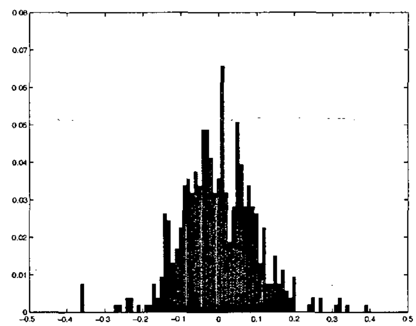
Quantile Difference[1]	X_n	B_n	Ratio [2]
55% − 45%	0.0259	0.0207	1.2533
60% − 40%	0.0460	0.0372	1.2378
65% − 35%	0.0783	0.0578	1.3552
70% − 30%	0.1012	0.0703	1.4396
75% − 25%	0.1270	0.0921	1.3789
80% − 20%	0.1580	0.1204	1.3132
85% − 15%	0.1878	0.1587	1.1832
90% − 10%	0.2935	0.2067	1.1587
95% − 5%	0.3051	0.2876	1.0610

1. Difference between two quantile values.

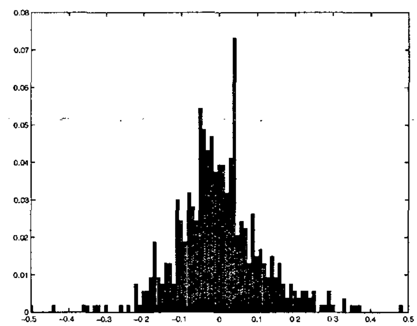
2. Ratio: Quantile Difference for X /Quantile Difference for B.

$\alpha = 1.75$, $\beta = -0.2$ and $\kappa = 0.055$. The Q-Q plots in the figure show that the quantiles of the distributions of the samples from the model and the data coincide well. Also, log-log plots of the left and right tails show that the tail behaviors of the distribution of the samples from the model approximate the distribution of the samples from the data reasonably well.

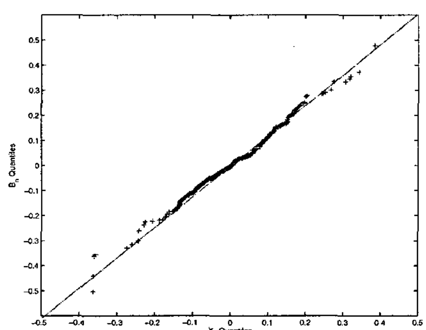
We test if the c and α in Figure B.12 and γ reasonably fit (3.33) in the main papr. We observe that $c^\alpha = 1.58$ and $1/(1-\gamma^\alpha) = 1.08$ coincide only roughly. Nevertheless, in summary, we conclude that the fitting to a heavy-tailed distribution works reasonably well, given the limited data.



(a) Event-driven strategy X_n

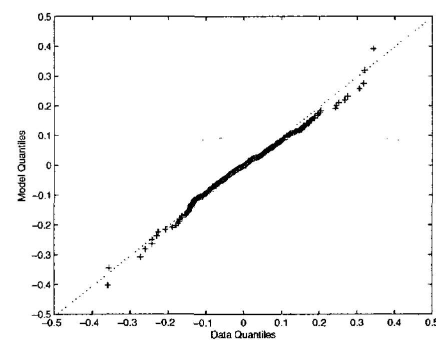


(b) Event-driven strategy $cB_n = c(X_n - \gamma X_{n-1})$

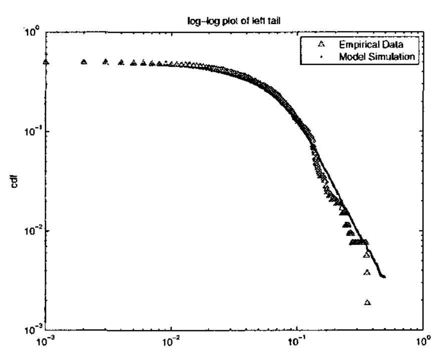


(c) Q-Q plot comparing X_n and cB_n

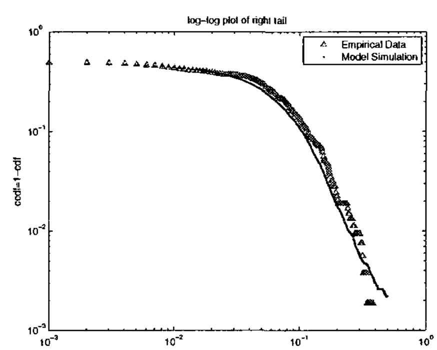
Figure B.11: X_n and cB_n from event-driven strategy and Q-Q plot comparing the distribution of X_n and cB_n with $c = 1.3$ for event-driven strategy.



(a) Q-Q plot comparing the model to the data



(b) Log-log plot of left tails



(c) Log-log plot of right tails

Figure B.12: Event-driven strategy Q-Q plot comparing the distribution of 533 samples from the data and 10^4 samples from the constant-persistence stable-noise model with $\alpha = 1.75, \beta = -0.2, \gamma = 0.24, k = 0.055$ for event-driven strategy.

Appendix C

Appendix to Chapter 4

C.1 Proof of Lemma 4.4.1

Proof. We prove $\psi(c) \leq \gamma(c)$. Suppose, by a way of contradiction, $\psi(c) > \gamma(c)$. Suppose a seller has the *ex post* cost c and places a bid of $\psi(c)$ in the first-phase auction of the A-B model. If he becomes the winning bidder, then the final price outcome in the second-phase bargaining process will be $\gamma(c)$ since $\gamma(c) < \psi(c)$. Since the probability density function f is strictly positive at $c \in (\underline{c}, \overline{c})$, by perturbing his bid lower, the seller strictly increases the probability of his winning without affecting the outcome of the final price in the scenario in which he wins the auction phase. Thus, bidding more than $\gamma(c)$ in the auction phase is not optimal for the seller. □

C.2 Proof of Theorem 4.4.2

Proof. To show that $\psi(c) = \min\{\beta(c), \gamma(c)\}$ is a symmetric equilibrium, we suppose that all the other sellers except seller $i \in \{1, 2, \ldots, n+1\}$ follow this strategy and show that seller i achieves the maximum expected profit by also following ψ. Suppose $c \in [\underline{c}, \overline{c}]$ is the *ex post* cost of bidder i, and let b be his first-phase bid. By Lemma 4.4.1, $b \leq \gamma(c)$. Also, it is clearly not optimal for seller i to place a bid lower than $\psi(\underline{c})$. Thus, $\psi(\underline{c}) \leq b \leq \gamma(c)$.

Let z be defined by $\psi(z) = b$. Since ψ is a strictly increasing and continuous function, z is well-defined. We refer to z as the implied cost associated with his bid b. From the above argument, $\psi(\underline{c}) \leq \psi(z) \leq \gamma(c)$. Then, seller i's expected profit is given by the probability of winning multiplied by his expected profit, conditioned on winning the auction. The probability of winning the auction phase is $\overline{G}(z)$. If seller i wins the auction phase, then the price outcome of the bargaining phase is the minimum of his bid $\psi(z)$ and $\gamma(c)$, i.e., $\min\{\psi(z), \gamma(c)\}$ which is $\psi(z)$. Thus, seller i's expected profit is

$$\Pi_i^{\psi}(z, c) \;=\; \overline{G}(z)\,(\psi(z) - c) \;=\; \overline{G}(z)\,(\min\{\beta(z), \gamma(z)\} - c)\ . \qquad \text{(C.1)}$$

We prove that, if ψ satisfies Condition 1, $\Pi_i^{\psi}(z,c)$ is maximized at $z=c$, by showing that $\Pi_i(z,c)$ increases in z if $z<c$ and decreases in z if $z>c$. We consider the following three cases.

Case $\beta(z)<\gamma(z)$: Since β and γ intersect finitely many times, there exists $\epsilon>0$ such that $\beta(z')<\gamma(z')$ for any $z'\in(z-\epsilon,z+\epsilon)$. Thus, it follows $\psi(z')=\beta(z')$ and $\Pi_i(z',c)=\overline{G}(z')(\beta(z')-c)$. Differentiating $\Pi_i^{\psi}(z,c)$ with respect to z in $(z-\epsilon,z+\epsilon)$ when $\psi(z)=\beta(z)$, we obtain

$$\frac{\partial}{\partial z}\Pi_i^{\psi}(z,c) = -g(z)(\beta(z)-c)+\overline{G}(z)\frac{\partial}{\partial z}\beta(z) .$$

Since $\beta(z)=\mathbb{E}[Y|Y>z]$ follows from Lemma 4.3.1, it is straightforward to verify

$$\frac{\partial}{\partial z}\beta(z) = \frac{g(z)}{\overline{G}(z)}(\beta(z)-z) . \tag{C.2}$$

Combining these two equations,

$$\frac{\partial}{\partial z}\Pi_i^{\psi}(z,c) = g(z)(c-z) .$$

If $z<c$, then the above expression is positive, and perturbing z higher increases the expected profit. Otherwise, perturbing z lower increases the expected profit.

Case $\gamma(z)<\beta(z)$: As before, there exists $\epsilon>0$ such that $\gamma(z')<\beta(z')$ for any $z'\in(z-\epsilon,z+\epsilon)$. Differentiating the expected profit with respect to z when $\psi(z)=\gamma(z)$, we obtain

$$\frac{\partial}{\partial z}\Pi_i^{\psi}(z,c) = -g(z)(\gamma(z)-c)+\overline{G}(z)\lambda . \tag{C.3}$$

Note that by Lemma 4.4.1, we obtain $\gamma(z)=\psi(z)\le\gamma(c)$, which implies $z\le c$. If $z<c$, then $\psi(z)=\gamma(z)$ and Condition 1 together imply that (C.3) is nonnegative since

$$\lambda \ge \frac{g(z)}{\overline{G}(z)}(\gamma(z)-z) \ge \frac{g(z)}{\overline{G}(z)}(\gamma(z)-c) ,$$

where the first inequality follows from Condition 1 and the second inequality follows from $z<c$. Thus, perturbing z higher weakly increases the expected profit.

Case $\gamma(z)=\beta(z)$: There exists $\epsilon>0$ such that either $\beta(z')<\gamma(z')$ or $\beta(z')>\gamma(z')$ for either $z'\in(z,z+\epsilon)$ or $z'\in(z-\epsilon,z)$. Then, apply the argument used in one of the two cases discussed above accordingly.

If Condition 1 does not hold, then there exists c such that $\psi(c)=\gamma(c)<\beta(c)$ and $\frac{\partial}{\partial z}\Pi_i^{\psi}(z,c)\big|_{z=c}<0$. Thus, bidding lower than $\gamma(c)$ improves the profit of seller i and it follows that ψ is not an equilibrium bidding strategy.

□

C.3 Proof of Lemma 4.4.3

Proof. To prove the lemma, suppose $\gamma'(c) = \lambda \geq \beta'(c)$ for any $c \in [\underline{c}, \overline{c}]$. Recall $\psi(c) = \min\{\gamma(c), \beta(c)\}$. Since $\psi(c) = \gamma(c)$ implies $\gamma(c) \leq \beta(c)$,

$$\frac{g(c)}{\overline{G}(c)}(\gamma(c) - c) \leq \frac{g(c)}{\overline{G}(c)}(\beta(c) - c) = \frac{\partial}{\partial c}\beta(c) ,$$

where the equality follows from the definition of β in Lemma 4.3.1 (see (C.2)). Thus, from $\beta'(c) \leq \lambda$, we obtain (4.1), completing the proof. □

C.4 Proof of Theorem 4.4.5

Proof. Let ψ be a continuous increasing function defined on $[\underline{c}, \overline{c}]$. First, if ψ satisfies both (i) and (ii), then we can show that ψ is an equilibrium by using the argument in the proof of Theorem 4.4.2. (Notice that equation (C.2) holds for any β_K function, and the only change is to replace β with β_{K_i} in the proof of Theorem 4.4.2.)

Suppose now ψ is a symmetric equilibrium function. We show that ψ satisfies (i) and (ii). By Lemma 4.4.1, we obtain $\psi(c) \leq \gamma(c)$ for each $c \in [\underline{c}, \overline{c}]$, proving (i). For (ii), let $c^\circ \in (\underline{c}, \overline{c})$ such that $\psi(c^\circ) < \gamma(c^\circ)$. Let $s = \sup\{c < c^\circ | \psi(c) = \gamma(c) \text{ or } c = \underline{c}\}$ and $t = \inf\{c > c^\circ | \psi(c) = \gamma(c) \text{ or } c = \overline{c}\}$. By the continuity of both ψ and γ, it follows that $c^\circ \in (s, t)$. Let $\Pi_i(z, c)$ be seller i's expected profit, where seller i submits the bid of $\psi(z)$ when his cost is c, and all other bidders follow the bidding strategy ψ. Since ψ is an equilibrium function, $\Pi_i(z, c)$ for fixed c is maximized when $z = c$. Differentiating $\Pi_i(z, c) = \overline{G}(z)(\psi(z) - c)$ with respect to z in (s, t), we obtain

$$\frac{\partial}{\partial z}\Pi_i(z, c) = -g(z)(\psi(z) - c) + \overline{G}(z)\frac{\partial}{\partial z}\psi(z) .$$

It is straightforward to verify that the family of solutions for ψ satisfying the above differential equation to be 0 at $z = c$ is

$$\beta_K(c) = \mathbb{E}[Y | Y > c] + \frac{K}{\overline{G}(c)} ,$$

parameterized by K. The choice of K is unique by the boundary condition at c°, namely $\beta_K(c^\circ) = \psi(c^\circ)$. Thus, $\psi(c) = \beta_K(c)$ for $c \in [s, t]$, and we verify (ii).

Furthermore, we prove the property of ψ in the neighborhood of $\overline{c}$. Recall from Section 4.2 that we assumed $v \geq \overline{c}$, which implies $\gamma(\overline{c}) \geq \beta_0(\overline{c}) = \overline{c}$. By the first part of Theorem 4.4.5, in the interval $[a_{m-1}, a_m]$ where $a_m = \overline{c}$, we have either $\psi(c) = \beta_K(c)$ for some K, or $\psi(c) = \gamma(c)$. In the former case, K cannot be positive since $\beta_K(c)$ should be well-defined and finite for all c in the closed interval $[a_{m-1}, a_m]$. Also, K cannot be negative since $\beta_K(c)$ should be increasing in the neighborhood of $\overline{c}$. Thus, we only need to consider β_0 and γ.

If $\gamma(\overline{c}) = \beta_0(\overline{c})$, then the above result shows that $\psi(\overline{c}) = \gamma(\overline{c}) = \beta_0(\overline{c}) = \overline{c}$. We proceed with the case $\gamma(\overline{c}) > \beta_0(\overline{c})$ and will prove $\psi(c) = \beta_0(c)$ in $[a_{m-1}, a_m]$. Suppose by way of contradiction that $\psi(c) = \gamma(c)$ in this interval. Then, since $g(c) = n \cdot (1 - F(c))^{n-1} f(c)$ and $\overline{G}(c) = (1 - F(c))^n$, we have

$$\frac{g(c)}{\overline{G}(c)}(\gamma(c) - c) = n \cdot \frac{f(c)}{1 - F(c)}(\gamma(c) - c) .$$

By the assumption that $f(c) > 0$ for all $c \in [\underline{c}, \overline{c}]$ where $\overline{c} < \infty$ (from Section 4.2), it follows that the above expression can be arbitrarily large as $c \to \overline{c}$. In particular, the above expression exceeds λ, which violates Condition 1. Thus, we conclude $\psi(c) = \beta_0(c)$ in $[a_{m-1}, a_m]$.

□

C.5 Proof of Lemma 4.4.6

Proof. We prove the lemma by contradiction. Suppose there exists c in the interior of Γ^ψ such that inequality (4.1) does not hold, i.e., $\psi(c) = \gamma(c)$ and

$$\lambda < \frac{g(c)}{\overline{G}(c)}(\gamma(c) - c) . \tag{C.4}$$

It follows that there exists $\epsilon > 0$ such that $(c - \epsilon, c) \subseteq \Gamma^\psi$, and the above inequality still holds in this interval. Suppose that all the other bidders follow the bidding strategy ψ. We proceed to show that seller i's expected profit is higher when he bids $\psi(c - \epsilon)$ compared to $\psi(c)$, which violates the definition of an equilibrium.

Differentiating the expected profit of seller i with respect to $z \in (c - \epsilon, c)$, we obtain from (C.3) that

$$\frac{\partial}{\partial z}\Pi_i^\psi(z, c) = \overline{G}(z) \cdot \left(-\frac{g(z)}{\overline{G}(z)}(\gamma(z) - c) + \lambda\right) ,$$

which is negative by the choice of ϵ above. Thus, $\Pi_i(c - \epsilon, c) < \Pi_i(c, c)$, and it follows that ψ is not an equilibrium strategy. □

C.6 Proof of Theorem 4.4.9

Proof. From Lemma 4.4.1, recall $\psi(c) \leq \gamma(c)$ for any $c \in (\underline{c}, \overline{c})$. Thus, if we can prove $\psi(c) \leq \beta(c)$ for $c \in (\underline{c}, \overline{c})$, then we obtain $\psi(c) \leq \min\{\beta(c), \gamma(c)\}$, which implies the required result. Therefore, for the remainder of the proof, we prove $\psi(c) \leq \beta(c)$ for $c \in (\underline{c}, \overline{c})$.

Consider the characterization of $\psi(c)$ given in the statement of Theorem 4.4.5. From the discussion following Lemma 4.4.6, we have $\psi(\overline{c}) = \beta_0(\overline{c}) \leq \gamma(\overline{c})$. Let

$$s_1 = \min\{\tilde{c} \in [\underline{c}, \overline{c}] \mid \beta_0(c) \leq \gamma(c) \quad \text{for any} \quad c \in [\tilde{c}, \overline{c}]\} .$$

If $s_1 = \underline{c}$, then we must have $\psi(c) = \beta(c)$, which proves the required result. If $s_1 > \underline{c}$, define

$$s_2 = \min\{\tilde{c} \in [\underline{c}, s_1] \mid \beta_0(c) \geq \gamma(c) \quad \text{for any} \quad c \in [\tilde{c}, s_1]\} .$$

Then, from the continuity of ψ and the monotonicity of β_K in K, it follows that, for any $c \in [s_2, s_1]$, we have $\psi(c) = \gamma(c)$ or $\psi(c) = \beta_K(c)$ for some $K \leq 0$. If $s_2 = \underline{c}$, then the required result holds. Note that $\gamma(c) = \beta_0(c)$ in the interval $[s_1, \overline{c}]$. Thus, if $s_1 = \underline{c}$, then we obtain the required result. Otherwise, we proceed by assuming that $s_1 > \underline{c}$. Define

$$s_2 \quad = \quad \min\{\tilde{c} \in [\underline{c}, s_1] \mid \psi(c) \leq \beta_0(c) \quad \text{for any} \quad c \in [\tilde{c}, s_1]\} \ .$$

If $s_2 = \underline{c}$ holds, we obtain the required result; thus, we proceed by assuming otherwise, i.e., $s_2 > \underline{c}$. Then, there exists $s^\circ < s_2$ sufficiently close to s_2 such that $\psi(c) = \gamma(c) > \beta_0(c)$ and $\beta_0'(c) > \gamma'(c) = \lambda$ in the interval $[s^\circ, s_2)$. We obtain

$$\frac{g(s^\circ)}{G(s^\circ)}[\gamma(s^\circ) - s^\circ] \ > \ \frac{g(s^\circ)}{G(s^\circ)}[\beta_0(s^\circ) - s^\circ] \ = \ \beta_0'(s^\circ) \ > \ \lambda \ ,$$

which is contrary to Condition 1. Thus, we conclude that $s_2 = \underline{c}$. □

C.7 Proof of Theorem 4.4.10

Proof. From the remarks following Theorem 4.4.5, we obtain $\psi(1) = 1$. Thus, by applying Theorem 4.4.5, the right-most segment is the β_0 segment; more precisely, let $s = \inf\{c \in [0,1] \mid \beta_0(c) \leq \gamma(c) \text{ or } c = 0\ \}$. Then, $\psi(c) = \beta_0(c)$ for all $c \in [s, 1]$. Since $s = 0$ implies that the statement of the theorem holds, we proceed by assuming $s > 0$. Then, it is easy to verify that if $K > 0$, $\beta_K(c) > \beta_0(c) > \gamma(c)$ in $[0, s)$, implying that β_K and γ do not intersect in $[0, s)$; thus, β_K does not specify any of the segments. If $K < 0$, then β_K crosses γ at most once at $[0, s]$, in which case, it crosses from above to below (not from below to above). Therefore, because of the constraint $\psi(c) \leq \gamma(c)$, we cannot construct a *continuous* increasing function ψ satisfying $\psi(\tilde{c}) = \beta_K(\tilde{c}) < \gamma(\tilde{c})$ and $\psi(\hat{c}) = \beta_K(\hat{c}) < \gamma(\hat{c})$ where $0 \leq \tilde{c} < \hat{c} \leq 1$. Thus, β_K does not specify any of the segments. □

C.8 Proof of Lemma 4.4.13

Proof. To derive the optimal reserve price r^*_{AB} in the first price A-B model, we calculate the expected profit of the buyer with the reserve price r and maximize her expected profit with respect to r. The expected profit of the buyer in the first price A-B model with a reserve price r is

$$\Pi^{\psi^r}_{AB} \quad = \quad \left(1 - (1 - F(r))^{n+1}\right) \cdot v - (n+1) \cdot \int_{\underline{c}}^{r} \psi^r(c)\overline{G}(c)f(c)\ dc \ .$$

Differentiating the expected profit with respect to r yields,

$$\frac{\partial}{\partial r}\Pi^{\psi^r}_{AB} \quad = \quad (n+1) \cdot \overline{G}(r) \cdot f(r) \cdot (v - r) - (n+1) \cdot \int_{\underline{c}}^{r} \overline{G}(c) \cdot f(c) \cdot \frac{\partial}{\partial r}\psi^r(c)dc \ .$$

Since $\psi^r(c) = \min\{\beta^r(c), \gamma(c)\}$, differentiating ψ^r with respect to r gives

$$\frac{\partial}{\partial r}\psi^r(c) = \begin{cases} \frac{\partial}{\partial r}\beta^r(c) = \frac{\overline{G}(r)}{\overline{G}(c)} & \text{if } \psi^r(c) = \beta^r(c) \\ \frac{\partial}{\partial r}\gamma(c) = 0 & \text{if } \psi^r(c) = \gamma(c) . \end{cases}$$

Substituting this into the above differentiation of the expected profit, we have

$$\frac{\partial}{\partial r}\Pi_{AB}^{\psi^r} = (n+1)\cdot\overline{G}(r)\cdot\left[f(r)\cdot(v-r) - \int_{\underline{c}}^{r} I\{\beta^r(c) \leq \gamma(c)\}\cdot f(c)\,dc\right] .$$

Note that the first order condition implies $\partial\Pi_{AB}^{\psi^r}/\partial r = 0$. Since $r < \overline{c}$ and $\overline{G}(r)$ is strictly positive for any $r \in [\underline{c}, \overline{c})$, we obtain

$$f(r)\cdot(v-r) - \int_{\underline{c}}^{r} I\{\beta^r(c) \leq \gamma(c)\}\cdot f(c)\,dc = 0 ,$$

as required. □

C.9 Proof of Theorem 4.4.17

Proof. Suppose that seller i's *ex post* cost is c. We compare the payment received by seller i in the first price and the second price A-B models. Since the first-phase bidding functions in Theorems 4.4.10 and 4.4.16 are symmetric and increasing, seller i wins the auction phase if $c < C_j$ for each $j \neq i$. Otherwise, seller i does not receive any payment from the buyer.

We compare the payments received by seller i given that he wins the auction phase of the A-B model, i.e., $Y > c$. Let $P_i^1(c, Y)$ and $P_i^2(c, Y)$ denote these quantities in the first price and the second price A-B models, respectively. (The superscript indicates the first price or the second price auction.) Let ψ^1 and ψ^2 denote the equilibrium bidding strategy given in Theorems 4.4.10 and 4.4.16, respectively. Then,

$$P_i^1(c,Y) = \min\{\gamma(c), \beta(c)\} \quad \text{and} \quad P_i^2(c,Y) = \min\{\gamma(c), Y\} .$$

Let $m_i^1(c)$ denote the *ex post* conditional expected revenue received by seller i when his realized cost is c, i.e., the expected value of $P_i^1(c, Y)$ where the expectation is taken for all values of Y satisfying $Y > c$. Similarly, define $m_i^2(c)$. Then,

$$\begin{aligned} m_i^1(c) &= \min\{\gamma(c), \beta(c)\} = \min\{\gamma(c), \mathbb{E}[Y \mid Y > c]\} \text{ and} \\ m_i^2(c) &= \mathbb{E}[\min\{\gamma(c), Y\} \mid Y > c]. \end{aligned}$$

Observe that $\min\{\gamma(c), y\}$ is a concave function with respect to y. We apply Jensen's Inequality to this function for the conditional distribution $[Y|Y > c]$, and obtain $m_i^1(c) \geq m_i^2(c)$. Since the *ex post* expected revenue received by seller i is higher in the first price A-B model, the buyer's expected profit is lower in the first price A-B model. □

C.10 Proof of Theorem 4.5.1

Proof. From the discussion preceding the statement of this theorem, it is easy to verify that the buyer's strategy is the best response to the sellers' bidding strategy, which is symmetric and increasing.

We now consider the seller i's best response given that the buyer and all the other sellers follow the strategy given in the statement. Suppose that seller i has the cost of c_i and bids b_i in the first phase. Since ψ is an increasing function, there exists z such that $\psi(z) = b_i$. Also, from Lemma 4.4.1, we have $\psi(\underline{c}) \leq \psi(z) \leq \gamma(c_i)$. Without loss of generality, suppose that $i = n+1$, and that the other sellers $j \in \{1, \ldots, n\}$ are indexed in an increasing order of b_j. Since $b_j = \psi(c_j) = \min\{\beta(c_j), \gamma(c_j)\}$ for $j \in \{1, \ldots, n\}$, sellers other than $n+1$ are indexed in an increasing order of $\psi(c_j)$, and also in an increasing order of c_j.

Thus, seller $i = n+1$ is one of m sellers selected in the first phase if and only if $\min\{b_i, \gamma(c_i)\} < \psi(c_m)$ (assuming no ties). Since $b_i = \psi(z) \leq \gamma(c_i)$, this condition is equivalent to $\psi(z) < \psi(c_m)$. Furthermore, seller i wins in the second phase if and only if $b_i = \psi(z) < \min\{\psi(c_j), \gamma(c_j)\} = \psi(c_j)$ for each $j = 1, \ldots, m$. In summary, seller i is selected if and only if $z < c_1$. In the case that seller i wins, the final price is $\psi(z)$ and his profit is $\psi(z) - c_i$. Then we obtain (C.1) where c is replaced with c_i and the remainder of the proof is similar to the proof of Theorem 4.4.2 subsequent to (C.1). (The only change is to replace c with c_i throughout.) □

Appendix D

Appendix for Chapter 5

D.1 Proof of Lemma 5.3.1

Proof. Part (a) is obtained by differentiating $\Pi^B(z,p)$ in (5.4) with respect to p and setting it to zero.

Now, for (b), by substituting the expression of $p^B(z)$ into (5.4) and using the identity $\mu - z = \Theta(z) - \Lambda(z)$, we obtain

$$\begin{aligned}\Pi^B(z,p^B(z)) &= y(p^B(z))[(\mu-\Theta(z))(p^B(z)-c)-c\Lambda(z)] \\ &= \frac{c}{b-1}\cdot y(p^B(z))z \qquad\qquad (D.1) \\ &= \frac{c}{b-1}\cdot q^B(z)\ , \qquad\qquad (D.2)\end{aligned}$$

where the last equality follows from the definition of z, i.e., $q^B(z) = y(p^B(z))\cdot z$. Since ϵ has the IGFR property and $b > 0$, Wang et al. (2004) have shown that $\Psi^B(z,p(z))$ is quasi-concave in z, and that the value of z satisfying $d\Pi^B(z,p(z))/dz = 0$ is unique, which is denoted by z^B.

From the expression (D.1), it is easy to see that the optimal solution z^B is independent of c, i.e., $z^B \perp\!\!\!\perp c$. Now, the first part (a) implies $p^B \propto c$. Then, from $q^B(z) = y(p^B(z))\cdot z$ and (D.2), we obtain $q^B \propto c^{-b}$ and $\Pi^{B*} \propto c^{-(b-1)}$, respectively. □

D.2 Proof of Lemma 5.3.2

Proof. The first two parts follow directly from the definition of $\hat{c}(\gamma)$ and Lemma 5.3.1. The last part follows from $z^B \perp\!\!\!\perp c$ in Lemma 5.3.1 since both $\Pi^R(z,p)$ and $\Pi^B(z,p)$ are essentially the same price-setting newsvendor problem where the only difference lies in the cost parameter. □

D.3 Proof of Lemma 5.3.3

Proof. (a) From Lemma 5.3.2, there exists K independent of γ such that

$$\Pi^M(z^R,p^R) = y(p^R)\cdot z^R\cdot(\gamma-1)\cdot c^M = K\cdot\hat{c}(\gamma)^{-b}\cdot(\gamma-1)\cdot c^M .$$

Differentiating it with respect to γ, we obtain

$$\frac{\partial\Pi^M(z^R,p^R)}{\partial\gamma} = K\cdot c^M\cdot\tilde{c}(\gamma)^{-(b+1)}[-(b-1)\cdot c^M\cdot\gamma+(c^R+b\cdot c^M)] .$$

Since $\tilde{c}(\gamma)^{-(b+1)} > 0$, it is clear that $d\Pi^M(z^R,p^R)/d\gamma$ changes the sign only once from positive to negative at $\gamma=\gamma^M$.

(b) From (D.2) and (5.7), we have

$$\Pi^{R*} = \frac{\hat{c}(\gamma)}{b-1}\cdot q(z^R) , \qquad \text{and}$$
$$\Pi^M(z^R,p^R) = q(z^R)\cdot(\gamma-1)\cdot c^M .$$

By dividing the first equation by the second equation, we obtain the required result.

(c) Rewriting Π^{R*} and $\Pi^M(p^R,z^R)$, we have

$$\begin{aligned}
\Pi^{R*} &= \frac{y(p^R)}{b-1}\hat{c}(\gamma)z^R,\\
\Pi^M(p^R,z^R) &= y(p^R)z^R(\hat{c}(\gamma)-c),\\
\tau^R\Pi^{R*}+\tau^M\Pi^M(p^R,z^R) &= \frac{y(p^R)z^R}{b-1}\left[(\tau^R+(b-1)\tau^M)\hat{c}(\gamma)-(b-1)\tau^M c\right].
\end{aligned}$$

Using the fact that z^R is independent of $\hat{c}(\gamma)$ and $p^R\propto\hat{c}(\gamma)$, we have

$$\tau^R\Pi^{R*}+\tau^M\Pi^M(p^R,z^R) = K\hat{c}(\gamma)^{-b}\left[(\tau^R+(b-1)\tau^M)\hat{c}(\gamma)-(b-1)\tau^M c\right],$$

where K is all the terms that are independent of $\hat{c}(\gamma)$. Differentiating the above equation yields

$$\frac{d\left(\tau^R\Pi^{R*}+\tau^M\Pi^M(p^R,z^R)\right)}{d\gamma} = Kc^M\hat{c}(\gamma)^{-(b+1)}\left[-(\tau^R+(b-1)\tau^M)\hat{c}(\gamma)+b\tau^M c\right].$$

Thus the total after-tax profit is maximized when

$$\gamma = \frac{b\tau^M+(\tau^M-\tau^R)c^R/c^M}{\tau^R+(b-1)\tau^M}.$$

It is then immediate to verify that the total-after tax profit is either quasi-concave with respect to γ ($\gamma\geq 1$) if $\tau^R\geq\tau^M$ or decreasing if $\tau^R<\tau^M$. □

D.4 Proof of Lemma 5.3.4

Proof. We have already observed that the constraint $\Pi^M(z,p) \geq 0$ is always satisfied. We consider the constraint $\Pi^R(z,p) \geq 0$. From the definition of $\Pi^R(z,p)$ in (5.6) and the identify $\mu - z = \Theta(z) - \Lambda(z)$, algebraic simplification shows that $\Pi^R(z,p) \geq 0$ is equivalent to

$$p \;\geq\; \hat{c}(\gamma)\,\frac{z}{\mu - \Theta(z)}\;.$$

First, we consider the optimal price $p^R(z)$ for any fixed z. Without the above constraint, the optimal choice of price p that maximizes $\Pi^C(z,p)$ for given z would have been

$$\rho \;=\; \frac{b}{b-1}\cdot\tilde{c}(\gamma)\cdot\frac{z}{\mu-\Theta(z)}\;,$$

by following the analysis of Lemma 5.3.1. This price would be optimal provided that $p = \rho$ satisfies the above constraint, i.e., $[b/(b-1)]\cdot\tilde{c}(\gamma) \geq \hat{c}(\gamma)$. This condition holds if and only if, by algebraic manipulation involving (5.5) and (5.9),

$$\gamma - 1 \;\geq\; \frac{c^R + c^M}{(b\tau^M - \tau^R)c^M}\;.$$

Note that this condition does not depend on the value of z. The proof of $z^C = z^B$ is similar to the proof of Lemma 5.3.2.

Otherwise, we have $\rho < [b/(b-1)]\cdot\hat{c}(\gamma)\cdot z/(\mu-\Theta(z))$, and also $[b/(b-1)]\cdot\tilde{c}(\gamma) < \hat{c}(\gamma)$. In this case, $\Pi^C(z,p)$ is decreasing in p within the interval of $p \geq \rho$ since

$$\frac{\partial \Pi^C}{\partial p} \;=\; \frac{y(p)}{p}[-(b-1)(\mu-\Theta(z))p + b\cdot\hat{c}(\gamma)\cdot z] \;\leq\; 0.$$

Thus, the optimal feasible price for fixed z is $p^C(z) = \hat{c}(\gamma)\cdot z/(\mu-\Theta(z))$. Then, similar to (D.1), we derive

$$\begin{aligned}\Pi^C(z,p^C(z)) \;&=\; y(p^C(z))[(p^C(z)-\hat{c}(\gamma))(\mu-\Theta(z)) - \hat{c}(\gamma)\Lambda(z)]\\ &=\; [\hat{c}(\gamma)-\tilde{c}(\gamma)]\cdot y(p^C(z))\cdot z\;.\end{aligned}$$

Note that this expression is a constant multiple of $\Pi^B(z,p^B(z))$ in (D.1), and we conclude $z^C = z^B$. □

D.5 Proof of Theorem 5.3.6

Recall that $[\underline{d},\overline{d}]$ is the support of ϵ. Let $l = \inf\{z \geq \underline{d} : X(z) > 0\}$.

Lemma D.5.1. *There exists $\hat{z} \in [\underline{d},\overline{d}]$ such that $X(z) > 0$ for $z < \hat{z}$ and $X(z) \leq 0$ for $z > \hat{z}$. Also, $G(z)$ is strictly decreasing in $[\underline{d},\hat{z}]$, and $G(\underline{d}) = 1$.*

Proof. Recall $G(z) = [X'(z) \cdot z]/X(z)$, and $X(z) = \mu - \Theta(z) - \beta z$ where $\Theta(z) = \mathbb{E}[(\epsilon - z)^+]$. Let F and f denote the cumulative density function and the probability density function of ϵ, respectively. Then,

$$X'(z) = 1 - \beta - F(z) \qquad \text{and} \qquad X''(z) = -f(z)\,. \tag{D.3}$$

Thus, X is concave in z. Since $X(\underline{d}) = \mu - (\mu - \underline{d}) - \beta \cdot \underline{d} = (1-\beta) \cdot \underline{d} \geq 0$, it follows that $\hat{z}$ satisfying the condition in the statement of the lemma exists.

We first claim that $G(z) \leq 1$ for any $z \in [\underline{d}, \hat{z}]$. To see this claim, observe

$$1 - G(z) = 1 - \frac{X'(z) \cdot z}{X(z)} = 1 - \frac{z - F(z)z - \beta z}{\mu - \Theta(z) - \beta z} = \frac{X(z) - X'(z) \cdot z}{X(z)}\,.$$

Clearly, $X(z) = \mu - \Theta(z) - \beta z > 0$. Also,

$$\begin{aligned} X(z) - X'(z) \cdot z &= [\mu - \Theta(z) - \beta z] - [z - F(z)z - \beta z] \\ &= \mu - \Theta(z) - (1 - F(z)) \cdot z \\ &= \mathbb{E}[\epsilon] - \mathbb{E}[(\epsilon - z) \cdot \mathbf{1}[\epsilon > z]] - \mathbb{E}[z \cdot \mathbf{1}[\epsilon > z]] \\ &= \mathbb{E}[\epsilon \cdot \mathbf{1}[\epsilon \leq z]] \;\geq\; 0. \end{aligned} \tag{D.4}$$

Thus, we obtain $1 - G(z) \geq 0$, proving the claim. Since the above expression is 0 at $z = \underline{d}$, we also obtain that $G(\underline{d}) = 1$.

Differentiating $G(z)$ with respect to z, we have

$$\begin{aligned} G'(z) &= \frac{[-f(z)z + X'(z)]X(z) - X'(z)z[(1-\beta) - F(z)]}{X(z)^2} \\ &= \frac{1 - F(z)}{X(z)} \left((1 - G(z)) \left(1 - \frac{\beta}{1 - F(z)} \right) - \frac{zf(z)}{1 - F(z)} \right)\,. \end{aligned} \tag{D.5}$$

Then, since $z \downarrow \underline{d}$, we obtain $\lim_{z \downarrow \underline{d}} G'(z) \leq 0$.

We now show $G(z)$ is strictly decreasing by contradiction. If G is not decreasing, then by the continuity of G and the above claim, there exist $z_1, z_2 \in (\underline{d}, \hat{z}]$ such that $z_1 < z_2$, $G(z_1) \geq G(z_2)$, and $G'(z_1) \leq 0 < G'(z_2)$. The derivative condition implies

$$(1 - G(z_1)) \left(1 - \frac{\beta}{1 - F(z_1)} \right) - \frac{z_1 f(z_1)}{1 - F(z_1)} \leq 0 < (1 - G(z_2)) \left(1 - \frac{\beta}{1 - F(z_2)} \right) - \frac{z_2 f(z_2)}{1 - F(z_2)}.$$

Here, since the rightmost expression is positive, we have $\beta/(1 - F(z_2)) < 1$. Then, by the increasing property of F, it follows that $1 - \beta/(1 - F(z_1)) \geq 1 - \beta/(1 - F(z_2)) > 0$. By the increasing generalized failure rate (IGFR) property of ϵ, the above inequality leads to contradiction with $G(z_1) \geq G(z_2)$. □

Proof of Theorem 5.3.6. For $X(z) \geq 0$, for a given z, from (5.14), we find that the optimal price p for a given z satisfies

$$p^R(z) = \frac{bc^R}{b-1} \cdot \frac{z}{X(z)}\,. \tag{D.6}$$

Plugging $p^R(z)$ to Π^R, we have

$$\Pi^R(z,p^R(z)) \;=\; y(p^R(z))\cdot\frac{c^R z}{b-1} \;=\; \frac{c^R}{b-1}\cdot q^R(z)\;, \tag{D.7}$$

where $q^R(z) = y(p^R(z))\cdot z$. From the definition of $y(p) = ap^{-b}$ and (D.6), we get $\frac{d}{dp}y(p) = -by(p)/p$, and

$$\frac{d}{dz}p^R(z) \;=\; \frac{bc^R}{b-1}\cdot\frac{X(z)-z\cdot X'(z)}{X^2(z)} \;=\; \frac{p^R(z)}{z}\cdot\left(1-\frac{zX'(z)}{X(z)}\right) \;=\; \frac{p^R(z)}{z}\cdot(1-G(z))\;.$$

We note that, by Lemma D.5.1, the above expression is nonnegative, and $p^R(z)$ is increasing in z.

Differentiating (D.7) with respect to z, we obtain

$$\begin{aligned}\frac{d\Pi^R(z,p^R(z))}{dz} &= \frac{c^R}{b-1}\left[y(p^R(z)) + z\cdot\frac{d}{dp}y(p)\cdot\frac{d}{dz}p^R(z)\right]\\ &= \frac{c^R}{b-1}y(p^R(z))\,[1-b+b\cdot G(z)]\\ &= y(p^R(z))\cdot\frac{bc^R}{b-1}\cdot\left[G(z)-\frac{b-1}{b}\right]\;.\end{aligned}$$

By Lemma D.5.1, this expression is strictly increasing in z, changes sign from positive to negative only once. Thus, $\Pi^R(z,p^R(z))$ is quasi-concave in z. If there exists z satisfying $G(z) = (b-1)/b$, then such z is the unique maximizer. Otherwise, it can be shown that $\Pi^R(z,p^R(z))$ is maximized when $z = \overline{d}$. □

D.6 Proof of Theorem 5.3.7

(a) We denote the dependence of various functions on β explicit in our notation. Recall $G(z|\beta) = [\frac{d}{dz}X(z|\beta)\cdot z]/X(z|\beta)$, and $X(z|\beta) = \mu - \Theta(z) - \beta z$ where $\Theta(z) = \mathbb{E}[(\epsilon - z)^+]$. Since

$$X(z|\beta) \;=\; \mu-\Theta(z)-\beta z \qquad\text{and}\qquad \frac{d}{dz}\left[X(z|\beta)\cdot z\right] \;=\; z-\beta z - zF(z)\;,$$

we obtain

$$\frac{d}{d\beta}X(z|\beta) \;=\; -\,z \qquad\text{and}\qquad \frac{d}{d\beta}\left[\frac{d}{dz}\left[X(z|\beta)\cdot z\right]\right] \;=\; -\,z\;.$$

Then,

$$\begin{aligned}\frac{d}{d\beta}G(z|\beta) &= \frac{d}{d\beta}\left[\frac{\frac{d}{dz}X(z|\beta)\cdot z}{X(z|\beta)}\right] = \frac{X(z|\beta)\cdot(-z) \;-\; \left[\frac{d}{dz}X(z|\beta)\cdot z\right]\cdot(-z)}{X(z|\beta)^2}\\ &= \frac{\left[X(z|\beta)-\frac{d}{dz}X(z|\beta)\cdot z\right]\cdot(-z)}{X(z|\beta)^2}\le\;0\end{aligned}$$

where the last inequality follows from (D.4). Thus $G(z|\beta)$ is decreasing in β. Since G is also decreasing in z, the value of z^R satisfying $G(z^R) = (b-1)/b$ decreases in β.

(b) By differentiating (5.10) with respect to β to obtain

$$\frac{\partial \Pi^R(z,p|\beta)}{\partial \beta} = -p \cdot y(p) \cdot z \leq 0 ,$$

for any z and p. Thus, it follows that $\max_{z,p} \Pi^R(z,p|\beta)$ is decreasing in β.

(c) From (5.15),

$$\Pi^R(z^R, p^R(z^R)) = \frac{c^R}{b-1} \cdot q^R(z^R) .$$

Since the left-hand-side expression decreases in β by part (c), $q^R(z^R)$ also decreases in z^R.

(d) For any z, $p^R(z)$ satisfies $\partial \Pi^R(z, p^R(z))/\partial p = 0$ in (5.14). Thus,

$$p^R(z) = \frac{c^R bz}{(b-1)X(z)} .$$

Now, since z^R satisfies $G(z^R) = (b-1)/b$ where $G(z) = X'(z) \cdot z/X(z)$, we have

$$\frac{z^R}{X(z^R)} = \frac{b-1}{b} \cdot \frac{1}{X'(z^R)} .$$

Thus, from (D.3), it follows

$$p^R(z^R) = c^R \cdot \frac{1}{X'(z^R)} = \frac{c^R}{1-\beta-F(z^R)} .$$

Now, in order to consider the dependence of z^R on β, we use the notation z^R_β for given β. We obtain

$$\begin{aligned} \frac{dp^R(z^R_\beta)}{d\beta} &= \frac{d}{d\beta}\frac{c^R}{X'(z^R_\beta)} = -c^R \frac{dX'(z^R_\beta)}{d\beta}\frac{1}{X'(z^R_\beta)^2} \\ &= -\frac{c^R}{X'(z^R_\beta)^2}\left(-1 - f(z^R_\beta)\frac{dz^R_\beta}{d\beta}\right) = -\frac{c^R}{X'(z^R_\beta)^2}\frac{(b-1)f(z^R_\beta)z^R - X'(z^R_\beta)}{-bf(z^R_\beta)z^R + X'(z^R_\beta)} . \end{aligned}$$

where $dz^R_\beta/d\beta$ is derived from the fact that $dG(z^R_\beta|\beta)/d\beta = 0$. Similar to (D.8), we have

$$\begin{aligned} \left.\frac{\partial G(z|\beta)}{\partial z}\right|_{z=z^R_\beta} &= \frac{1}{X(z^R_\beta)}\left(-f(z^R_\beta)z^R_\beta + X'(z^R_\beta)(1-G(z^R_\beta))\right) \\ &= \frac{1}{X(z^R_\beta)}\left(-f(z^R_\beta)z^R_\beta + X'(z^R_\beta)\frac{1}{b}\right) , \end{aligned}$$

and

$$\frac{\partial G(z^R_\beta|\beta)}{\partial \beta} = \frac{z^R_\beta}{X(z^R_\beta)}\left(-1 + G(z^R_\beta)\right) = \frac{z^R_\beta}{bX(z^R_\beta)} .$$

Therefore,

$$\begin{aligned}
\frac{dG(z^R_\beta|\beta)}{d\beta} &= \frac{\partial G(z^R_\beta|\beta)}{\partial\beta} + \frac{\partial G(z|\beta)}{\partial z}\Big|_{z=z^R_\beta}\cdot\frac{dz^R\beta}{d\beta} \\
&= -\frac{z^R_\beta}{bX(z^R_\beta)} + \frac{1}{X(z^R_\beta)}\left(-f(z^R_\beta)z^R_\beta + X'(z^R_\beta)\frac{1}{b}\right)\frac{dz^R_\beta}{d\beta} = 0 , \\
\frac{dz^R_\beta}{d\beta} &= \frac{z^R_\beta}{-bf(z^R_\beta)z^R_\beta + X'(z^R_\beta)} ,
\end{aligned}$$

Since we know that $dz^R_\beta/d\beta < 0$, we know that $-bf(z^R_\beta)z^R_\beta + X'(z^R_\beta) < 0$. Thus, $dp^R(z^R_\beta)/d\beta > 0$ if $(b-1)f(z^R_\beta)z^R_\beta - X'(z^R_\beta) > 0$.

D.7 Proof of Theorem 5.3.8

Proposition D.7.1. *For any positive numbers k, c^R and c^M,*

$$\frac{c^R}{c^M}\left[1 - \left(\frac{c^R}{c^R+c^M}\right)^k\right] \leq k .$$

Proof. We first observe that both sides of the desired inequality approaches 0 as $k \downarrow 0$. Thus, it suffices to show that their derivatives satisfy the inequality, i.e.,

$$-\frac{c^R}{c^M}\left(\frac{c^R}{c^R+c^M}\right)^k \ln\left(\frac{c^R}{c^R+c^M}\right) \leq 1 .$$

In fact, we will show a stronger result

$$-\frac{c^R}{c^M}\ln\left(\frac{c^R}{c^R+c^M}\right) \leq 1 ,$$

for any positive c^R and c^M. To show the above inequality, we fix c^M and show that (i) the limit of the left-side expression as $c^R \to \infty$ is 1, and (ii) the left-side expression is increasing in c^R.

To prove (i), we use l'Hôpital's rule to obtain

$$\lim_{c^R\to\infty}\frac{\ln\left(\frac{c^R}{c^R+c^M}\right)}{c^M/c^R} = \lim_{c^R\to\infty}\frac{\frac{c^R+c^M}{c^R}\cdot\frac{c^M}{(c^R+c^M)^2}}{c^M/c^{R^2}} = \lim_{c^R\to\infty}\frac{c^R}{c^R+c^M} = 1.$$

To prove (ii), we take the derivative with respect to c^R,

$$\frac{d}{dc^R}\left[-\frac{c^R}{c^M}\ln\left(\frac{c^R}{c^R+c^M}\right)\right] = -\frac{1}{c^M}\ln\left(\frac{c^R}{c^R+c^M}\right) - \frac{1}{c^R+c^M} .$$

The nonnegativity of this expression follows from the following Taylor expression:

$$-\ln\left(\frac{c^R}{c^R+c^M}\right) = -\ln\left(1-\frac{c^M}{c^R+c^M}\right) = \sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{c^M}{c^R+c^M}\right)^n \geq \frac{c^M}{c^R+c^M} .$$

Thus, we complete the required proof. □

Proof of Theorem 5.3.8. Set $\beta = c^M/c$. In the deterministic model, the optimal price under the resale-price method is $p = bc^R/[(1-\beta)(b-1)]$. Then,

$$\Pi_{RP}^{R*} = \left(\frac{c^R}{c}\right)^b \cdot \frac{a}{b} \cdot \left(\frac{b-1}{bc^R}\right)^{b-1} .$$

Similarly, the optimal price under the cost-plus method is $p = b(c^R + \gamma c^M)/(b-1)$ and

$$\Pi_{CP}^{R*} = \frac{a}{b}\left(\frac{b-1}{b(c^R+\gamma c^M)}\right)^{b-1} .$$

To find the value of the appropriate markup parameter γ, we equate the above expression to Π_{CP}^{R*}:

$$\left(\frac{c^R}{c}\right)^b \cdot \frac{a}{b} \cdot \left(\frac{b-1}{bc^R}\right)^{b-1} = \frac{a}{b} \cdot \left(\frac{b-1}{b\hat{c}(\gamma)}\right)^{b-1}$$

where $\hat{c}(\gamma) = c^R + \gamma c^M$. Then, γ satisfies

$$\hat{c}(\gamma) = \left[\frac{c^b}{c^R}\right]^{\frac{1}{b-1}} \qquad \text{and} \qquad \gamma - 1 = \frac{1}{c^M}\left[\left(\frac{c^b}{c^R}\right)^{\frac{1}{b-1}} - c\right] .$$

Now, we compare $\Pi_{CP}^M(p_{CP}^R)$ and $\Pi_{RP}^M(p_{RP}^R)$. From the above analysis, the condition $\Pi_{CP}^M(p_{CP}^R) \leq \Pi_{RP}^M(p_{RP}^R)$ is equivalent to each of the following inequalities:

$$(\gamma-1)\cdot c^M \cdot a \cdot \left(\frac{b-1}{b\hat{c}(\gamma)}\right)^b \leq \left(\frac{c^R}{c}\right)^{b-1} \cdot a \cdot \frac{(b-1)^{b-1}}{(b\cdot c^R)^b} \cdot \left[\frac{c^R c^M}{c}\right]$$

$$(\gamma-1)(b-1)\left(\frac{1}{\hat{c}(\gamma)}\right)^b \leq \frac{1}{c^b}$$

$$\frac{1}{c^M}\left[\left(\frac{c^b}{c^R}\right)^{\frac{1}{b-1}} - c\right](b-1)\left[\frac{c^R}{c^b}\right]^{\frac{b}{b-1}} \leq \frac{1}{c^b} .$$

Simplifying the above inequality, it is equivalent to

$$\frac{c^R}{c^M}\left[1 - \left(\frac{c^R}{c}\right)^{\frac{1}{b-1}}\right] \leq \frac{1}{b-1} ,$$

which holds by an application of Proposition D.7.1. □