

MARKOV CHAIN MODELS TO ESTIMATE THE PREMIUM FOR EXTENDED HEDGE FUND LOCKUPS

by

Emanuel Derman*, Kun Soo Park, and Ward Whitt

Department of Industrial Engineering and Operations Research
Columbia University, New York, NY 10027-6699
{ed2114, kp2143, ww2040}@columbia.edu

July 25, 2009

Abstract

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. We estimate the premium for such extended lockup, 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, if permitted, we define the annual lockup premium as the difference between the 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. The model quantifies the way the lockup premium depends on these parameters. Data from the TASS database are used to estimate the persistence, which is found to be statistically significant.

*Prisma Capital Partners

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*. A good hedge fund has had superior returns and thrives; a sick fund has had inferior returns but still continues to survive; a dead fund has had very inferior returns and consequently gone out of business. We assume that transitions among these states occur randomly according to a Markov chain. The investor in a hedge fund that ends up in a dead state has suffered a very low return and receives his share of the net asset value of the fund when the fund unwinds its investments. At the next yearly reinvestment opportunity the investor is repaid his share of the final net asset value of the dead fund and is assumed to invest in a new fund, which we take to be in an initially good state. A dead state in our model, a year later, therefore transforms into a good state.

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 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 paper, we assume that the hedge fund starts off in a good state.

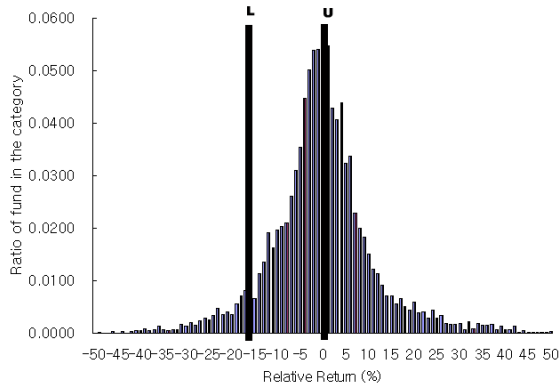


Figure 1: The distribution of hedge fund annual relative return rates based on 4788 selected observations from the TASS database from 2001 to 2005. Tentative levels L and U divide the funds into the three states G , S and D .

A fundamental principle 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 1 shows the auto-regression results from the data we selected. We did the analysis by strategy category. Included are 95% confidence intervals for

Table 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					

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 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 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 paper, 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 the

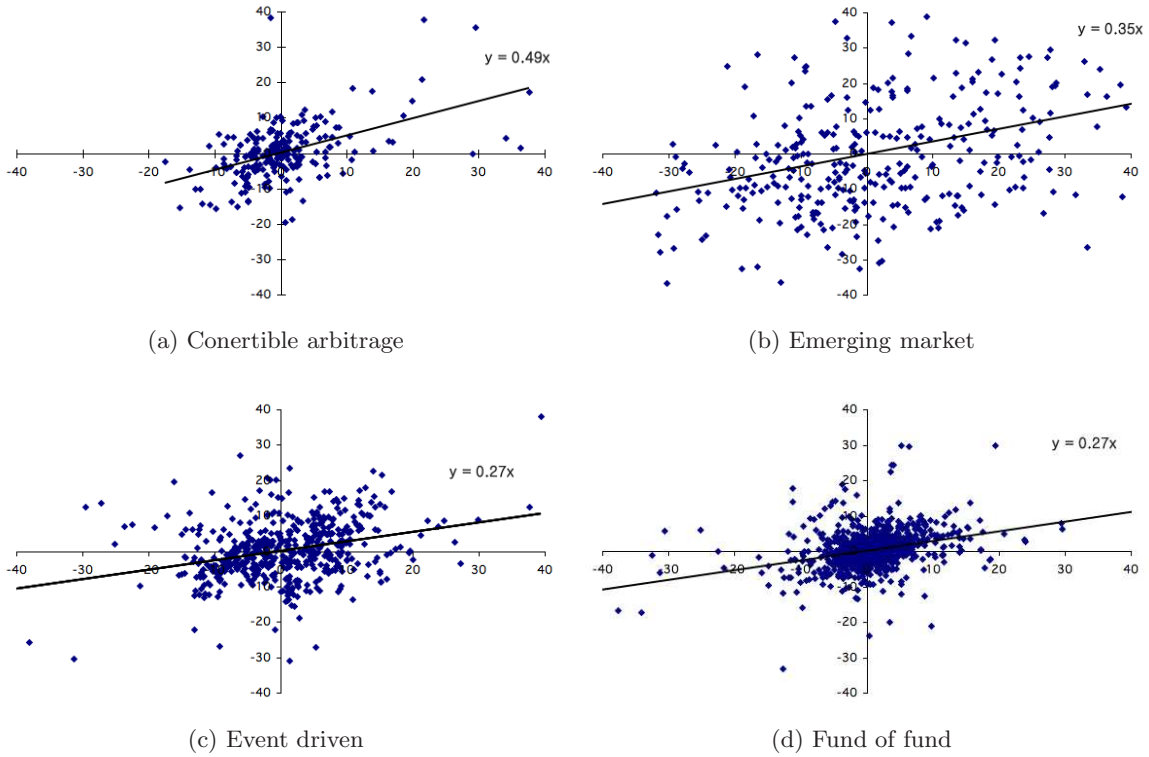


Figure 2: Scatter plots and associated least-squares lines for hedge fund annual relative return rates in successive years from 2000 to 2005 for four strategy categories from the TASS database.

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

$$\psi(\delta, \gamma, \sigma) = 0.047 \delta^{-0.11} \gamma^{0.69} \sigma^{0.64}, \quad (1.1)$$

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 paper. We start in §2 by reviewing the related literature on liquidity, including premiums for hedge fund lockup. In §3 we carefully specify what we mean by the lockup premium. In §4 we discuss persistence of hedge fund returns, reviewing the literature and analyzing data from the TASS database. In §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 §6 we introduce and analyze our three-state DTMC model, carrying out our indirect fitting procedure. In §7 we show how the model parameters and the lockup premium depend on basic hedge fund performance measures. Finally, in §8 we draw conclusions. We present additional material in an appendix.

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 et al. (2003). 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 different 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 et al. (2003) - 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 (2007), which is a preliminary account of the research reported here.

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

$$V_n \equiv e^{(\int_0^n X(t) dt)} \quad (3.1)$$

dollars. We now define a deterministic value r_n such that

$$e^{nr_n} = \mathbb{E}[V_n] \quad (3.2)$$

for V_n in (3.1); 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 (3.1) and (3.2), r_n can be expressed directly as

$$r_n = \frac{\log \mathbb{E}[V_n]}{n} = \frac{1}{n} \log \left(\mathbb{E} \left[e^{(\int_0^n X(t) dt)} \right] \right), \quad (3.3)$$

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

$$p_n \equiv r_n^1 - r_n^2, \quad (3.4)$$

where r_n^i is the rate of return of X^i , defined as in (3.3). 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.,

$$e^{B_i} = (1 + M_{i,1})(1 + M_{i,2}) \cdots (1 + M_{i,12}); \quad (3.5)$$

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

$$V_n = \prod_{i=1}^n e^{B_i} = e^{\sum_{i=1}^n B_i} \quad (3.6)$$

for A_i in (3.5). Equivalently, starting from the reported monthly returns $M_{i,j}$, we let the return-rate stochastic process X be defined by

$$X(t) \equiv 12 \log(1 + M_{i,j}) \quad \text{for } (i-1) + (j-1)/12 \leq t < (i-1) + (j/12), \quad (3.7)$$

so that

$$\int_{(i-1)+(j-1)/12}^{(i-1)+(j/12)} X(t) dt \equiv \log(1 + M_{i,j}) \quad (3.8)$$

for $1 \leq i \leq n$ and $1 \leq j \leq 12$. With definition (3.7), equation (3.6) is consistent with definition (3.1).

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 (3.5). 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

$$R_i \equiv B_i - \mathbb{E}[B_i] \equiv B_i - \mu_i. \quad (3.9)$$

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

Combining equations (3.6) and (3.9), we see that the (random) total value at the end of year n of hedge fund j is

$$V_n^j = \prod_{i=1}^n e^{(\mu_i + R_i)} = e^{(\sum_{i=1}^n \mu_i)} e^{(\sum_{i=1}^n R_i^j)} \quad (3.10)$$

and the difference between the expected total returns is

$$\mathbb{E}[V_n^1] - \mathbb{E}[V_n^2] = e^{(\sum_{i=1}^n \mu_i)} \left(\mathbb{E} \left[e^{(\sum_{i=1}^n R_i^1)} \right] - \mathbb{E} \left[e^{(\sum_{i=1}^n R_i^2)} \right] \right). \quad (3.11)$$

Hence, the premium in (3.4) becomes

$$p_n \equiv r_n^1 - r_n^2 = \frac{1}{n} \left[\log \left(\mathbb{E} \left[e^{(\sum_{i=1}^n R_i^1)} \right] \right) - \log \left(\mathbb{E} \left[e^{(\sum_{i=1}^n R_i^2)} \right] \right) \right], \quad (3.12)$$

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 (3.12), we will also set the initial state as Good, as mentioned in §1.

4. Persistence of Hedge Fund Returns

As indicated in §3, we specify how hedge funds perform by looking at the *relative rate of return* of a fund, given by the R_i in (3.9). 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, ter Horst et al. (2001) 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 §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 (3.9)). 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 $\pm 30\%$ for fixed income arbitrage, equity macro and $\pm 40\%$ for convertible, dedicated short bias, and global macro strategy categories. We also exclude pairs of relative returns exceeding $\pm 50\%$ 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 (3.5), the monthly returns are annualized to produce annual returns, from which we calculate relative rates of return R_i as indicated in §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 and Table 1 in §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 C for the details.

5. An Approximation for the Lockup Premium Based on Persistence Alone

Given the expression for the lockup premium p_n in (3.3), (3.4) and (3.12), 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 (3.3). Expression (3.3) 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 (3.12), that yields the approximation

$$p_n \approx \tilde{p}_n \equiv \frac{1}{n} \left[\log \left(e^{\left(\mathbb{E} \left[\sum_{i=1}^n R_n^1 \right] \right)} \right) - \log \left(e^{\left(\mathbb{E} \left[\sum_{i=1}^n R_n^2 \right] \right)} \right) \right] = \frac{1}{n} \sum_{i=1}^n \left(\mathbb{E}[R_i^1] - \mathbb{E}[R_i^2] \right). \quad (5.1)$$

Unlike p_n in (3.12), the approximation \tilde{p}_n in (5.1) 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 (5.1) 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 (5.1), while the two-term approximation is

$$p_n \approx \tilde{p}_n + 1/(2n) \left(\mathbb{E} \left[\left(\sum_{i=1}^n R_i^1 \right)^2 \right] - \mathbb{E} \left[\left(\sum_{i=1}^n R_i^2 \right)^2 \right] \right). \quad (5.2)$$

The second term explains most of the error for small n , e.g., for $n \leq 5$; see §E 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 (5.1). 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 (5.1), 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 §3, let R_n be the relative return rate in the n^{th} year. As mentioned in §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

$$\mathbb{E}[R_n] = \gamma \cdot \mathbb{E}[R_{n-1}] = \gamma^n \cdot Y_G, \quad n \geq 1. \quad (5.3)$$

As a consequence of (5.3), the sum of the expected relative return rates up to the n^{th} year can be expressed as a product

$$\sum_{i=1}^n \mathbb{E}[R_i] = Y_G \left(\frac{\gamma(1 - \gamma^n)}{1 - \gamma} \right), \quad 0 \leq \gamma < 1. \quad (5.4)$$

Combining this simple analysis with approximation (5.1), 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 (5.4). 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

$$\begin{aligned}\tilde{p}_n &\equiv \tilde{p}_n(\gamma, Y_G) = \left(\frac{1}{n} \sum_{i=1}^n \gamma Y_G \right) - \left(\frac{1}{n} \sum_{i=1}^n \gamma^i Y_G \right) \\ &= Y_G \gamma \left(1 - \frac{1 - \gamma^n}{n \cdot (1 - \gamma)} \right), \quad n \geq 1,\end{aligned}\tag{5.5}$$

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

$$\gamma Y_G \left(1 - \frac{1}{n(1 - \gamma)} \right) \leq \tilde{p}_n \leq \gamma Y_G \left(1 - \frac{1}{n} \right), \quad n \geq 1,\tag{5.6}$$

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

$$Y_G = \mathbb{E}[N(0, \sigma^2) | N(0, \sigma^2) > 0] = \mathbb{E}[|N(0, \sigma^2)|] = \sigma \mathbb{E}[|N(0, 1)|] = \sqrt{2/\pi} \sigma \approx 0.8\sigma.\tag{5.7}$$

We can combine (5.5) and (5.7) to obtain the following general approximate no-death lockup premium function

$$\tilde{p}_n(\gamma, \sigma) = 0.8\sigma\gamma \left(1 - \frac{1 - \gamma^n}{n \cdot (1 - \gamma)} \right), \quad n \geq 1. \quad (5.8)$$

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 (5.7), but we can generalize for arbitrary boundary point U , using the following formula for the conditional expectation of a normal random variable:

$$\mathbb{E}[N(m, \sigma^2) | a \leq N(m, \sigma^2) \leq b] = m + \sigma \frac{[\phi((a - m)/\sigma) - \phi((b - m)/\sigma)]}{[\Phi((b - m)/\sigma) - \Phi((a - m)/\sigma)]} \quad (5.9)$$

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 (5.9), 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 8 in §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 8 in §6.7.

Figure 8 shows plots of two curves for positive death rates δ , obtained using the DTMC model in §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 8, we will see

that the lockup premium is decreasing in the hedge fund death rate with our Markov chain model. Consequently, formulas (5.5) and (5.8) 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.

6. The Discrete-Time-Markov-Chain Model

We start in §6.1 by discussing two important hedge-fund performance measures: persistence and the death rate. Next in §6.2 we define the basic three-state DTMC model, which has six parameters. Then in §6.3 we introduce four equations that the six parameters must satisfy, based on standard hedge fund performance measures. In §6.4 we develop explicit formulas for the three parameters appearing in the DTMC transition probabilities. In §6.5 we show how to calculate all the parameters after specifying two of the relative returns. We present numerical examples in §6.6. Finally, we show how to calculate the lockup premium in §6.7.

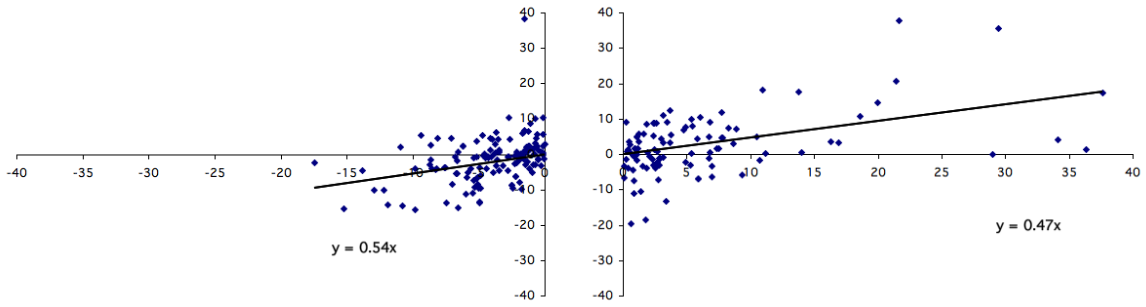
6.1. Important Hedge-Fund Performance Measures

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.

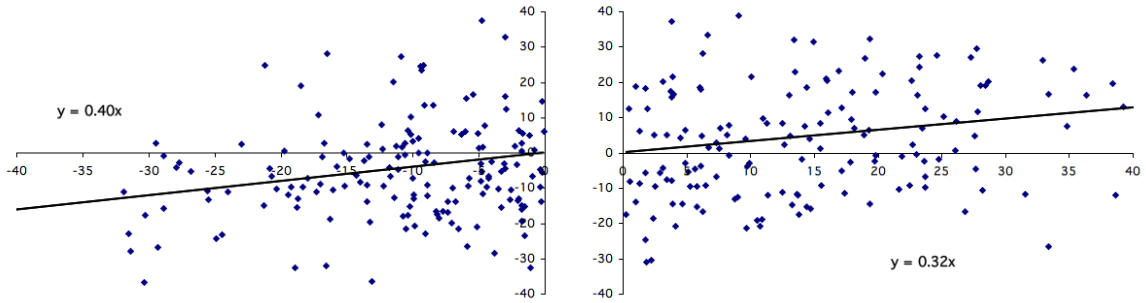
Two persistence factors: γ_G and γ_S . In equations (6.5) and (6.6) below we will introduce two state-dependent persistent factors γ_G and γ_S , instead of just the single γ . Clearly, this generalization is important if the persistence factors for the two states do in fact differ significantly. To illustrate what actually may happen, Figure 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 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.

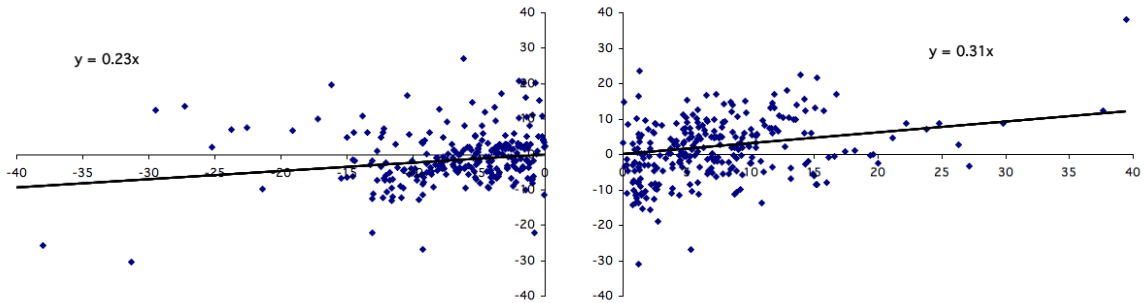
After checking the reasons for funds being terminated in the HFR data, Rouah (2006) concluded that, after removing these biases, 3 to 5% of the hedge funds leave the database each year because of failure. As noted in §4, Park (2006) estimated that the fund death rate



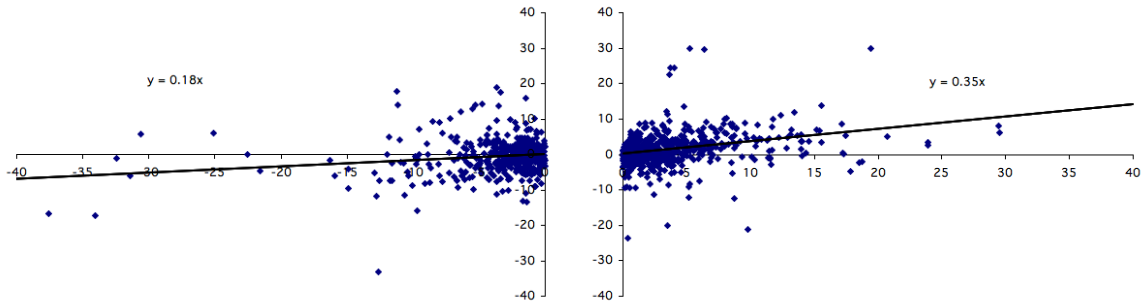
(a) Convertible arbitrage



(b) Emerging market



(c) Event driven



(d) Fund of fund

Figure 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.

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 (6.1). the *probability of surviving n years* is $S_n = P_{G,G}^n + P_{G,S}^n$ for $n \geq 1$. Figure 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 §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 $\delta = 0.06$.

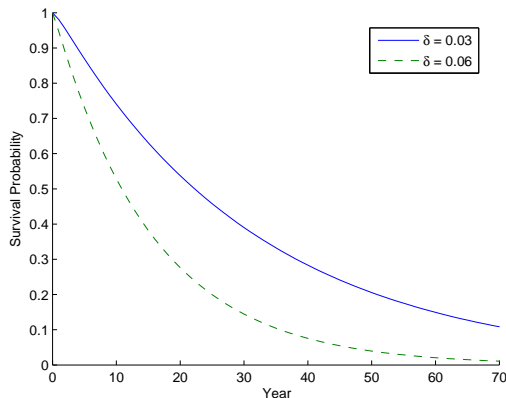


Figure 4: The survival probability for the DTMC model when $\delta = 0.03$ and 0.06 , for parameter values given in the Table 2.

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 :

$$P = \begin{matrix} G \\ S \\ D \end{matrix} \begin{pmatrix} p & 1-p & 0 \\ q & r & 1-q-r \\ 0 & 0 & 1 \end{pmatrix}. \quad (6.1)$$

which corresponds to the following diagram: We have assumed that it is impossible to transition

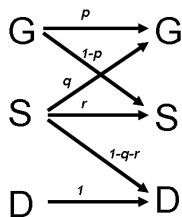


Figure 5: Transition probabilities in the absorbing Markov chain

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

$$P = \begin{matrix} G \\ S \\ D \end{matrix} \begin{pmatrix} p & 1-p & 0 \\ q & r & 1-q-r \\ p & 1-p & 0 \end{pmatrix}. \quad (6.2)$$

In (6.2), 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 (6.2) is

$$\pi_G = \frac{q + p(1 - q - r)}{2 - p - r}, \quad \pi_S = \frac{1 - p}{2 - p - r}, \quad \pi_D = \frac{(1 - p)(1 - q - r)}{2 - p - r}. \quad (6.3)$$

Our DTMC model uses both transition matrices. We use the absorbing transition matrix in (6.1) when we compute the expected return of a fund, while we use the ergodic transition matrix in (6.2) 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 .

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 (6.1), 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 (6.1) and (6.3), the *death rate* is

$$\delta = \pi_S \cdot P_{S,D} = \frac{1-p}{2-p-r}(1-q-r) = \pi_D . \quad (6.4)$$

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:

$$\gamma_G \cdot Y_G = p \cdot Y_G + (1-p) \cdot Y_S \quad \text{and} \quad (6.5)$$

$$\gamma_S \cdot Y_S = q \cdot Y_G + r \cdot Y_S + (1-q-r) \cdot Y_D . \quad (6.6)$$

We explain these DTMC-persistence equations as follows: In equation (6.5), the fund starts with state G ; in equation (6.6) 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 (6.1).

Our fourth equation is for the steady-state variance of the annual returns R_i in (3.9). Notice that its variance equals the variance of B_i , defined in (3.5). 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

$$\sigma^2 = \pi_G \cdot Y_G^2 + \pi_S \cdot Y_S^2 + \pi_D \cdot Y_D^2 . \quad (6.7)$$

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 (6.4), (6.5) and (6.6).

The three formulas. Assuming that γ_G , γ_S , δ , Y_G , Y_S and Y_D are specified, the three equations in (6.4), (6.5), and (6.6) 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 (6.5), because that is a single equation for the single unknown variable p . The solution is

$$p = \frac{\gamma_G \cdot Y_G - Y_S}{Y_G - Y_S}. \quad (6.8)$$

Having found the explicit expression for p in (6.8), 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 (6.4), we get

$$q \equiv q(r) = 1 - r - \frac{\delta(2 - p - r)}{1 - p} = 1 - \delta \left(\frac{2 - p}{1 - p} \right) - r \frac{(1 - p - \delta)}{(1 - p)}. \quad (6.9)$$

Since $\delta < 1 - p$ by (6.4), the function $q(r)$ in (6.9) is necessarily strictly decreasing in r .

Next, (6.6) can be rewritten as

$$q \equiv q(r) = \frac{\gamma_S \cdot Y_S - Y_D - r(Y_S - Y_D)}{Y_G - Y_D} = \frac{(\gamma_S - r)Y_S - (1 - r)Y_D}{Y_G - Y_D}. \quad (6.10)$$

Combining the two equations (6.9) and (6.10), we get an explicit expression for r , first in terms of p and then in terms of the basic model parameters, namely,

$$r = \frac{\left(\frac{1-p-\delta(2-p)}{1-p} \right) - \left(\frac{\gamma_S \cdot Y_S - Y_D}{Y_G - Y_D} \right)}{\left(\frac{1-p-\delta}{1-p} \right) - \left(\frac{Y_S - Y_D}{Y_G - Y_D} \right)} = \frac{\left(\frac{(1-\delta)(1-\gamma_G)Y_G - \delta(Y_G - Y_S)}{(1-\gamma_G)Y_G} \right) - \left(\frac{\gamma_S \cdot Y_S - Y_D}{Y_G - Y_D} \right)}{\left(\frac{(1-\gamma_G)Y_G - \delta(Y_G - Y_S)}{(1-\gamma_G)Y_G} \right) - \left(\frac{Y_S - Y_D}{Y_G - Y_D} \right)} \quad (6.11)$$

To be feasible, we of course need $0 \leq q \leq 1 - r$ and $0 \leq r \leq 1$. Formulas (6.9) and (6.11) simplify when $\delta = 0$; see §D in the Appendix.

By further analysis, we can determine what parameter values can occur; see §F in the Appendix for a detailed analysis. Figure 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 $\delta = 0.07$.

6.5. Determining All Model Parameters

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

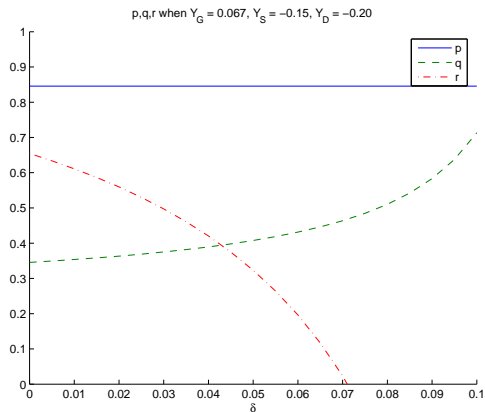


Figure 6: The DTMC parameter values p , q and r as a function of δ when $Y_G = 0.067$, $Y_S = -0.15$, $Y_D = -0.20$ and $\gamma_G = \gamma_S = 0.5$

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 §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 (6.7). 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 7 below for the special case in which $Y_S = -0.15$, $Y_D = -0.20$, $\gamma_G = \gamma_S = 0.5$ and $\delta = 0.03$.

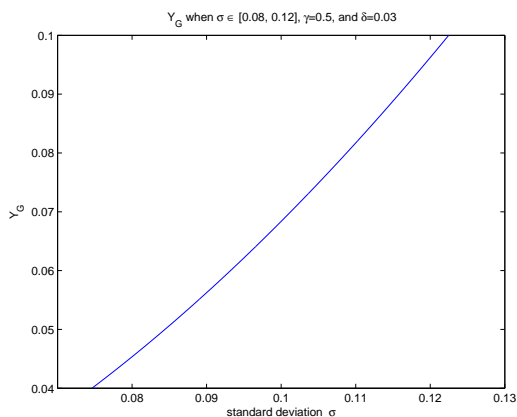


Figure 7: The standard deviation of relative return σ versus Y_G when $Y_S = -0.15$, $Y_D = -0.20$, $\gamma_G = \gamma_S = 0.5$, and $\delta = 0.03$.

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 (6.7) to obtain

$$1 = \pi_G \cdot (Y_G/\sigma)^2 + \pi_S \cdot (Y_S/\sigma)^2 + \pi_D \cdot (Y_D/\sigma)^2 . \quad (6.12)$$

Observe that the steady-state probability vector π in (6.3) and the death rate δ in (6.4) depend only on DTMC parameters p , q and r , while the equations (6.8), (6.10) and (6.11) 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.*

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 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 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: 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

6.7. The Lockup Premium Calculation

To calculate the lockup premium, we use formula (3.12). 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 §1, we assume that $S_0 = G$. Then,

$$\mathbb{E} \left[e^{\sum_{i=1}^n R_i^1} | S_0 = G \right] = (pe^{Y_G} + (1-p)e^{Y_S})^n. \quad (6.13)$$

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

$$m(t, s) \equiv \mathbb{E} \left[e^{\sum_{i=1}^t R_i^2} | S_0 = G, S_t = s \right] \cdot P_{G,s}^t, \quad (6.14)$$

where P^t , the t^{th} power of matrix P defined in (6.1), represents the probability of reaching S_t from G at t^{th} year. Then, we obtain the following recursion formulas

$$\begin{aligned} m(t, G) &= pe^{Y_G}m(t-1, G) + qe^{Y_G}m(t-1, S), \\ m(t, S) &= (1-p)e^{Y_S}m(t-1, G) + re^{Y_S}m(t-1, S) \quad \text{and} \\ m(t, D) &= (1-q-r)e^{Y_D}m(t-1, S), \end{aligned} \quad (6.15)$$

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

$$\mathbb{E} \left[e^{\sum_{i=1}^n R_i^2} | S_0 = G \right] = m(n, G) + m(n, S) + \sum_{t=2}^n m(t, D) (pe^{Y_G} + (1-p)e^{Y_S})^{n-t}. \quad (6.16)$$

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 §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 8 shows the lockup premium calculated from (6.15) and (3.12) as well as the analytical approximation for $\delta = 0$ in (5.1). It is observed that both the one-term and two-term analytical approximations in §5 constitute upper bounds for the lockup premium.

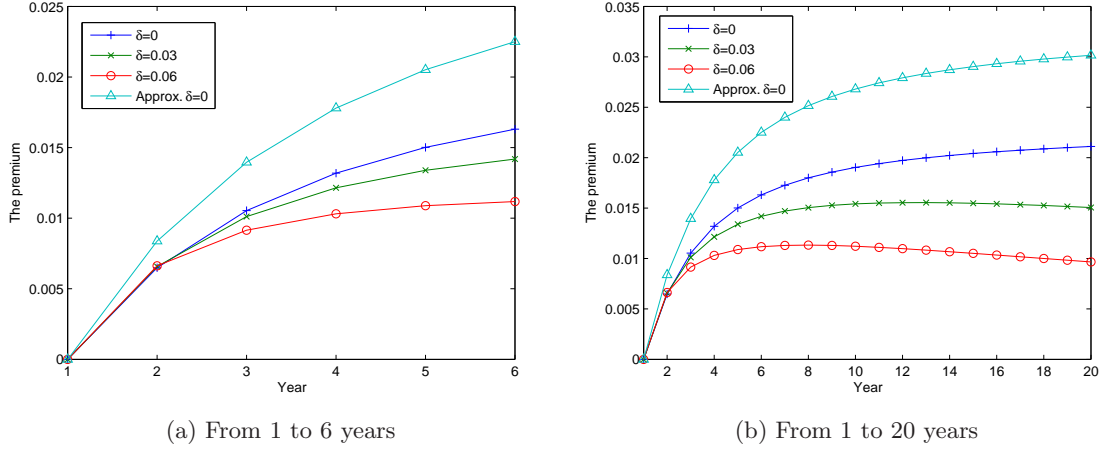


Figure 8: The lockup premium function for DTMC model for three values of the hedge-fund death rate δ and analytic approximation of lockup premium (§5) for $\delta = 0$. The remaining model parameters are $Y_G = 0.067$, $Y_S = -0.15$, and $Y_D = -0.20$ and $\gamma_G = \gamma_S = \gamma = 0.5$.

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.

Figure 9 (i) shows the lockup premium for five values of γ : 0.1, 0.2, 0.3, 0.4 and 0.5 while Figure 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 G.

We next consider how the DTMC model parameters p , q and r depend on the other driving variables. To supplement Figure 6 and the commentary in §6.4, Figure 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 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;

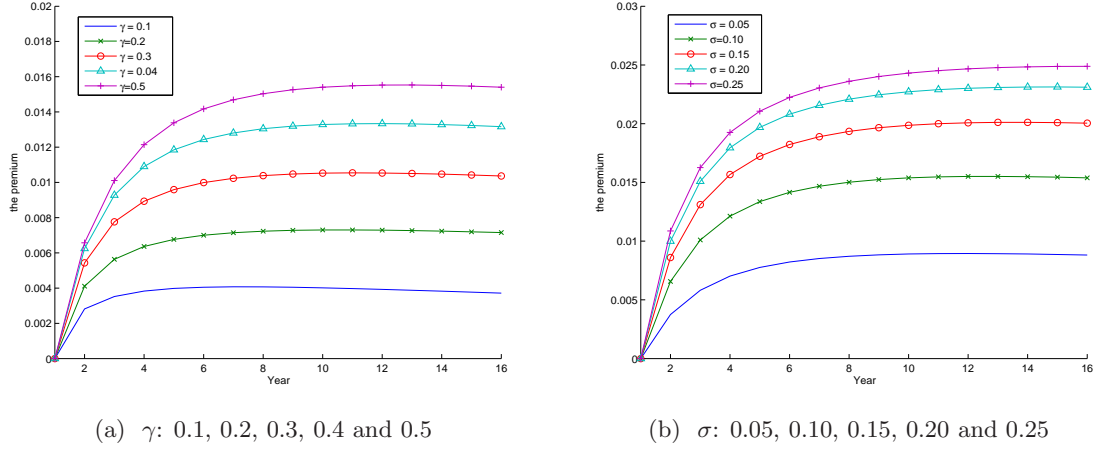


Figure 9: The lockup premium for the DTMC model in the base case with five values of γ and σ .

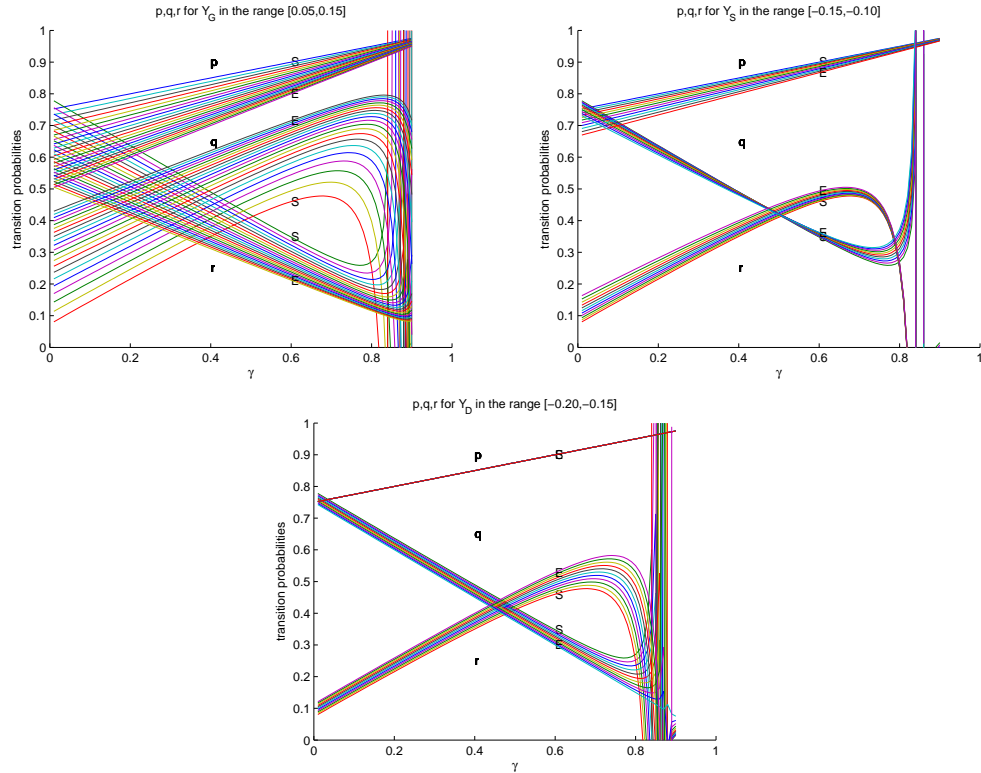
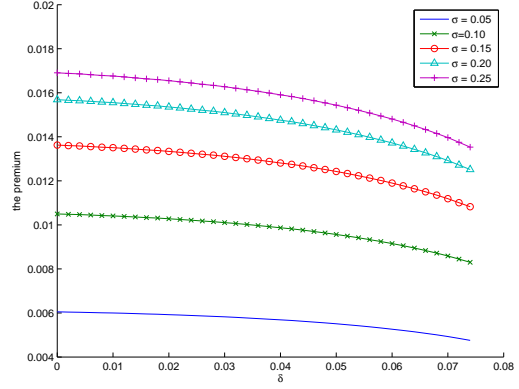
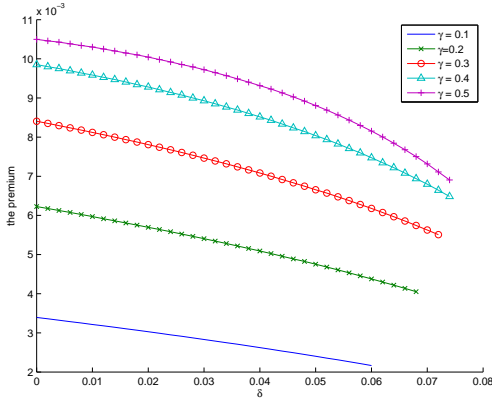


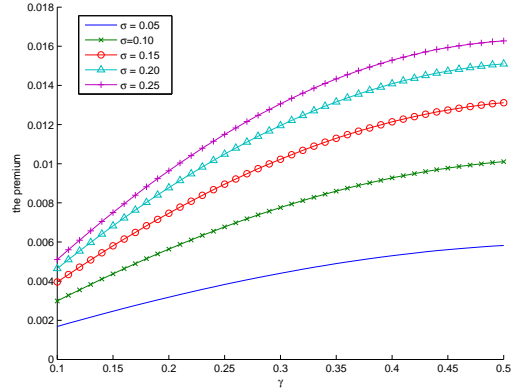
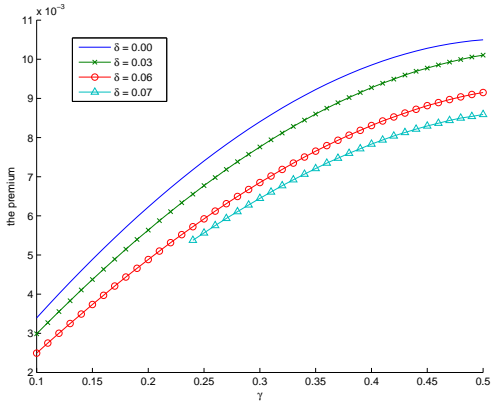
Figure 10: The parameters p , q and r as a function of γ in the base case for values of Y_G ranging from 0.05 (starting value, denoted by S) to 0.15 (ending value, denoted by E), Y_S ranging from -0.15 to -0.10 , and Y_D from -0.20 to -0.15

there is concavity in γ but convexity in σ . Also, the three-year lockup premium is relatively insensitive to δ .

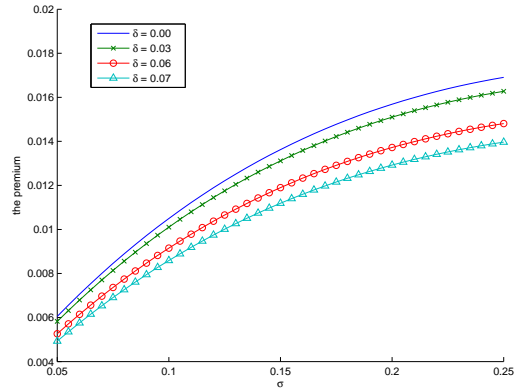
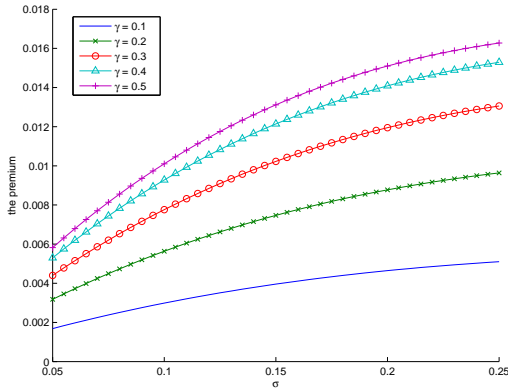
We remark that the lockup premium in the DTMC model can be approximated by a simple functional form of three variables δ , γ , and σ with the choice of $Y_S/\sigma = -1.5$ and $Y_D/\sigma = -2.0$. The approximation with a simple functional form is helpful to quickly estimate how the lockup



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 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.

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 $\psi_{(3)}^p = 0.15 \delta^{0.11} \gamma^{0.74} \sigma^{1.00}$ 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 G.4 for further discussion.

8. Conclusion

As we explained in §1 and §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 receives the final net asset value of his investments in the dead fund when it terminates its business, so the lockup restriction ends and the investor is now free to reinvest in a good fund. The deleterious returns that result from having been locked into a fund that became sick are therefore mitigated by its death.

In §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 (1.1) for our main numerical example.

As we explained in §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 and §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. 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. (2009). 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).

Acknowledgments. We thank Soonmin Ko for assistance in the early stages of this study. Ward Whitt was partially supported by NSF grant DMI-0457095.

References

- 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.
- 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. *The Journal of Finance* **46**(4) 1411–1425.
- Aragon, G. 2007. Share restrictions and asset pricing: evidence from the hedge fund industry. *Journal of Financial Economics* **83**(1) 33–58.
- 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.
- Boyson, N. M, M. J. Cooper. 2004. Do hedge funds exhibit performance persistence? a new approach. Forthcoming to *Financial Analysts Journal*.
- 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. J., M. A. Milevsky, T. S. Salisbury. 2003. Asset allocation and the liquidity premium for illiquid annuities. *The Journal of Risk and Insurance* **70**(3) 509–526.
- Chordia, T., R. Roll, A. Subrahmanyam. 2001. Market liquidity and trading activity. *Journal of Finance* **56**(2) 501–530.
- Darar, V., N. Y. Naik, R. Radcliffe. 1998. Liquidity and asset returns: An alternative test. *Journal of Financial Markets* **1** 203–219.
- Derman, E. 2007. A simple model for the expected premium for hedge fund lockups. *Journal of Investment Management* **5**(3) 5–15.
- Derman, E., K. S. Park, W. Whitt. 2009. A stochastic-difference-equation model for hedge-fund relative returns. Forthcoming to *Quantitative Finance*.

- 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.
- Fung, W., D. A. Hsieh. 2000. Performance characteristics of hedge fund 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.
- Gregoriou, G. N. 2002. Hedge fund survival lifetimes. *Journal of Asset Management* **3**(3) 237–252.
- Hasanhodzic, J., A. 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.
- 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, 12015.
- 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.
- Krishnamurthy, A. 2002. The bond / old-bond spread. *Journal of Financial Economics* **66**(2-3) 463–506.
- Liang, B. 1999. On the performance of hedge fund. *Financial Analyst Journal* **55**(4) 72–85.
- 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.
- Park, H. 2006. Risk measures for hedge funds and a survival analysis. Ph.D. thesis, University of Massachusetts.

- Pastor, L., R. F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* **111**(3) 642–685.
- Ross, S. M. 2003. *Introduction to Probability Models*. Eighth ed. Academy Press.
- Rouah, F. 2006. Competing risks in hedge fund survival. Ph.D. thesis, McGill University.
- ter Horst, J. R., T. Nijman, M. Verbeek. 2001. Eliminating look-ahead bias in evaluating persistence in mutual fund performance. *Journal of Empirical Finance* **8**(4) 345–373.
- Vayanos, D. 2004. Flight to quality, flight to liquidity, and the pricing of risk. NBER working paper No. W10327.
- Warga, A. 1992. Bond returns, liquidity, and missing data. *Journal of Financial and Quantitative Analysis* **27**(4) 605–617.

APPENDIX

A. Overview

In this appendix we elaborate on several issues investigated in the main paper. We start with statistical issues. In §B we provide additional details about our statistical analysis of the TASS database. Specifically, we describe our estimates of the standard deviation σ . In §C we describe an alternative ratio method for estimating the persistence of hedge fund returns.

We next turn to the DTMC models. In §D we consider the two-state DTMC model without dying funds, which provides a link between §5 and §6 in the main paper. In §E, we investigate if the difference between the premium p_n and the linear approximation \tilde{p}_n derived in §5 of the main paper 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 §F, we analyze what possible parameter values can occur in the DTMC model. Lastly, we supplement §7 in the main paper in §G 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 (1.1) in the main paper.

B. Statistical Results

In this section we supplement our discussion of our statistical methods provided in §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 3 and observe that they are within the range we consider in this paper: 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 3: 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

C. 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 4 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 1 and 4, these persistence estimates tend to be similar to the regression estimates.

Table 4: 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

D. 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

$$P = \begin{matrix} G \\ S \end{matrix} \begin{pmatrix} p & 1-p \\ 1-r & r \end{pmatrix}, \quad (\text{D.1})$$

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 (D.1). A convenient explicit expression for π is

$$\pi \equiv (\pi_G, \pi_S) = \left(\frac{1-r}{(1-r) + (1-p)}, \frac{1-p}{(1-r) + (1-p)} \right) = \left(\frac{1-r}{2-r-p}, \frac{1-p}{2-r-p} \right). \quad (\text{D.2})$$

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

$$\sigma^2 = \pi_G \cdot Y_G^2 + \pi_S \cdot Y_S^2. \quad (\text{D.3})$$

To satisfy, (D.3), we calibrate p and r in the transition matrix P of (D.1). We do this from two equations for persistence factor γ , expressed as a function of p and q .

$$\gamma \cdot Y_G = p \cdot Y_G + (1-p) \cdot Y_S \quad \text{and} \quad \gamma \cdot Y_S = (1-r) \cdot Y_G + r \cdot Y_S. \quad (\text{D.4})$$

From (D.4), we derive that

$$p = \frac{\gamma \cdot Y_G - Y_S}{Y_G - Y_S}, \quad r = \frac{Y_G - \gamma \cdot Y_S}{Y_G - Y_S}. \quad (\text{D.5})$$

From the above equations, it is straightforward to verify that $(Y_G - Y_S)/\sigma = 1$. Hence, if we fix $Y_S/\sigma = -1.5$, then we $Y_G/\sigma = 1/1.5 \approx 0.67$, 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 $Y_G/\sigma = \text{median} \{|N(0, 1)|\} = 0.67$.

E. Approximation of the No-death Lockup Premium

This section supplements §5 of the main paper 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

$$\log(x + 1) \approx x \quad \text{and} \quad e^x \approx 1 + x + \frac{1}{2!}x^2 \quad (\text{E.1})$$

for x close to 0. Applying these approximation formulas to (3.12), we have

$$\begin{aligned} p_n &\approx \frac{1}{n} \left[\log \left(1 + \mathbb{E} \left[\sum_{i=1}^n R_i^1 \right] + \mathbb{E} \left[\frac{1}{2} \left(\sum_{i=1}^n R_i^1 \right)^2 \right] \right) \right. \\ &\quad \left. - \log \left(1 + \mathbb{E} \left[\sum_{i=1}^n R_i^2 \right] + \mathbb{E} \left[\frac{1}{2} \left(\sum_{i=1}^n R_i^2 \right)^2 \right] \right) \right] \\ &\approx \frac{1}{n} \left[\left(\mathbb{E} \left[\sum_{i=1}^n R_i^1 \right] - \mathbb{E} \left[\sum_{i=1}^n R_i^2 \right] \right) + \frac{1}{2} \left(\mathbb{E} \left[\left(\sum_{i=1}^n R_i^1 \right)^2 \right] - \mathbb{E} \left[\left(\sum_{i=1}^n R_i^2 \right)^2 \right] \right) \right] \end{aligned} \quad (\text{E.2})$$

Thus, we have the approximation $p_n \approx \tilde{p}_n + e_n$ for \tilde{p}_n in (5.1). 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 (6.16), with negligible difference.

We compute $\tilde{p}_n + e_n, n = 1, 2, \dots, 20$ numerically for three different death rates ($\delta = 0.00, 0.03, 0.06$), $\gamma = 0.5$ and the other parameter values in Table 2 of the main paper. 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 12. However, it is also observed that as n increases, other higher-order terms omitted in the approximation formula (E.2) 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.

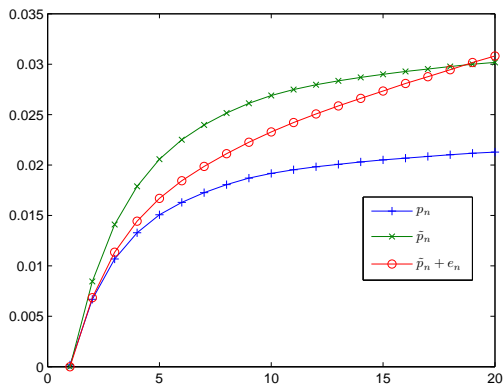


Figure 12: The DTMC lockup premium (p_n), analytic approximation (\tilde{p}_n), and two-term approximation ($\tilde{p}_n + e_n$) for $\delta = 0$, $\gamma = 0.5$, $Y_G = 0.067$, $Y_S = -0.15$ and $Y_D = -0.20$.

F. Possible parameter values in the DTMC Model

In this section, we supplement §6.4 in the main paper and determine what parameter values can occur in the DTMC model. Figure 13 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 (6.8) in the main paper, 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 (6.8), we see that p is independent of δ . Since (6.4) implies that $\delta < 1 - p$, there is

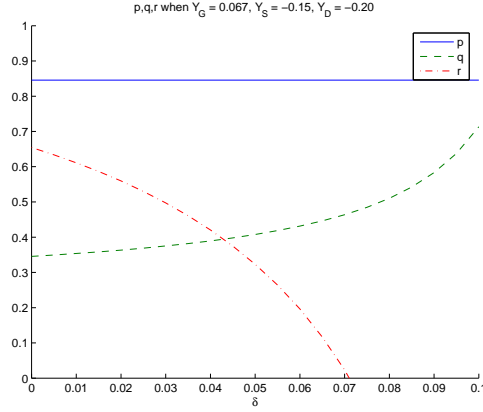


Figure 13: The DTMC parameter values p , q and r as a function of δ when $Y_G = 0.067$, $Y_S = -0.15$, $Y_D = -0.20$ and $\gamma_G = \gamma_S = 0.5$

an upper bound on the possible δ , consistent with Figure 6. Moreover, that inequality can be restated as $p < 1 - \delta$. When combined with (6.8), that yields an upper bound on 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 (6.10) 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 (6.10) 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

$$r < \frac{|Y_D/Y_S| + \gamma_S}{|Y_D/Y_S| + 1} < 1 \quad \text{for } 0 < \gamma_S < 1, \quad (\text{F.1})$$

where $|Y_D/Y_S| > 1$. On the other hand, if $Y_S \geq 0 > Y_D$, then we have

$$r < \frac{(|Y_D|/Y_S) - \gamma_S}{(|Y_D|/Y_S) + 1} < 1, \quad (\text{F.2})$$

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

$$a - B \equiv \left(\frac{(1 - \delta)(1 - \gamma_G)Y_G - \delta(Y_G - Y_S)}{(1 - \gamma_G)Y_G} \right) - \left(\frac{\gamma_S \cdot Y_S - Y_D}{Y_G - Y_D} \right) > 0 \quad \text{or} \quad (\text{F.3})$$

$$b - A \equiv \left(\frac{Y_S - Y_D}{Y_G - Y_D} \right) - \left(\frac{(1 - \gamma_G)Y_G - \delta(Y_G - Y_S)}{(1 - \gamma_G)Y_G} \right) > 0. \quad (\text{F.4})$$

Examination of (6.11) 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 $\delta = 0.07$.

G. Sensitivity Analysis for the DTMC Model

In this section we do more sensitivity analysis, expanding on the discussion in §7 in the main paper. We first carry out the calculations for the base case, using the parameter values derived in §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 .

G.1. How the Lockup Premium Depends on γ (γ_G, γ_S) and σ

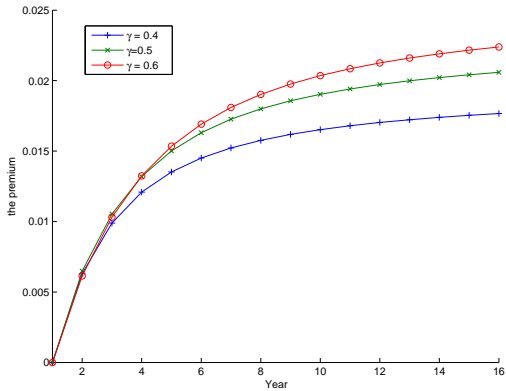
We now see how much the premium depends on the model parameters $\gamma(\gamma_G, \gamma_S)$ and σ . Table 5 shows the model parameters for $\gamma = 0.4, 0.5$, and 0.6 and Figure 14 shows the lockup premium for $\gamma = 0.4, 0.5$, and 0.6 . The figure suggests that as the persistence factor γ decreases, 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 14 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 6 and the sensitivity of the lockup premium with respect to γ_G and γ_S . Note that in the third line of Table 6, r is negative, which breaks down the DTMC model. Figure 15 shows the lockup premium for parameters in Table 6.

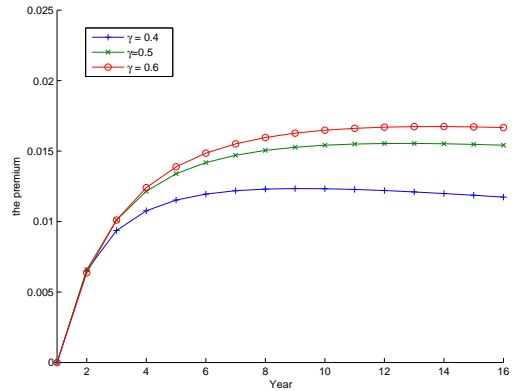
We lastly check the sensitivity of the lockup premium with respect to σ . Our TASS database analysis estimates σ of annual returns for each year is lower than 0.20 in most cases. We here highlight the sensitivity of the lockup premium for $\sigma = 0.05, 0.10$, and 0.15 with $\gamma = 0.5$. Table 7 is the parameter value sets and Figure 16 is the corresponding lockup premium. We see that the premium increases in σ .

Table 5: Parameter value sets for γ ranging from 0.2 to 0.6

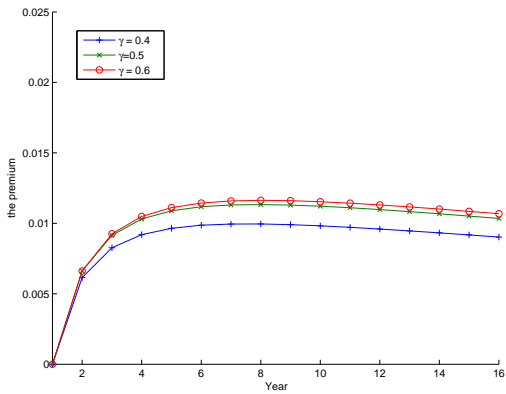
γ	δ	p	q	r	Y_G	Y_S	Y_D	Calculated σ
0.4	0.00	0.8147	0.4147	0.5853	0.067	-0.15	-0.20	0.1002
0.4	0.03	0.8147	0.4427	0.4360	0.067	-0.15	-0.20	0.1000
0.4	0.06	0.8147	0.4892	0.1877	0.067	-0.15	-0.20	0.0998
0.5	0.00	0.8456	0.3456	0.6544	0.067	-0.15	-0.20	0.1002
0.5	0.03	0.8432	0.3719	0.5030	0.0685	-0.15	-0.20	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



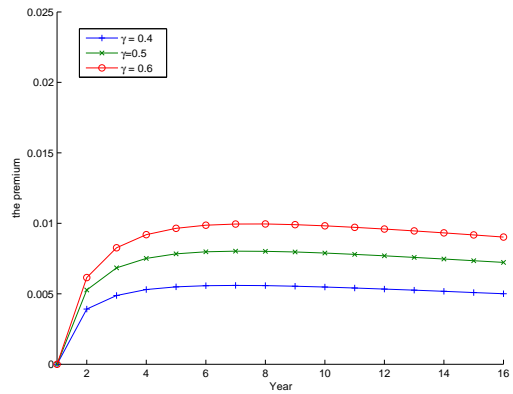
(a) With $\delta = 0$ for $\gamma = 0.4, 0.5, 0.6$



(b) With $\delta = 0.03$ for $\gamma = 0.4, 0.5, 0.6$



(c) With $\delta = 0.06$ and $\gamma = 0.4, 0.5, 0.6$

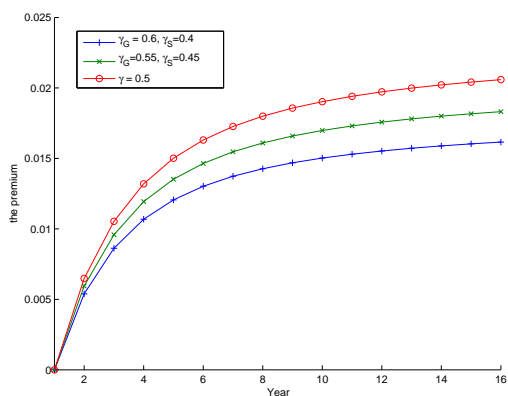


(d) With $\delta = 0.06$ and $\gamma = 0.2, 0.3, 0.4$

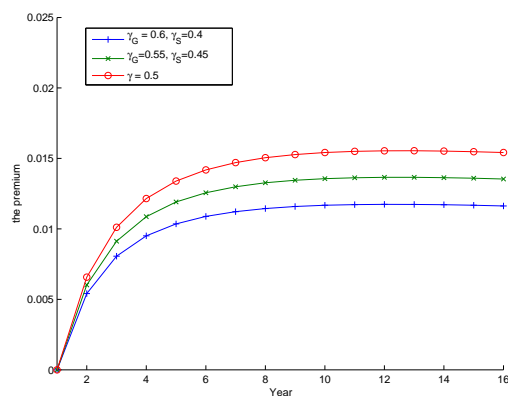
Figure 14: The lockup premium for the DTMC model for parameter values in Tables 5.

Table 6: Parameter value sets for γ_G and γ_S

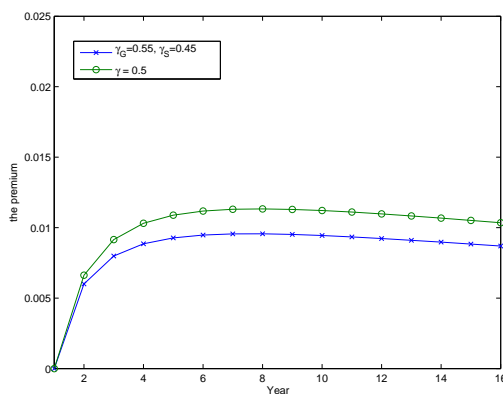
γ_G	γ_S	δ	p	q	r	Y_G	Y_S	Y_D	Calculated σ
0.6	0.4	0.00	0.8655	0.3982	0.6018	0.076	-0.15	-0.20	0.1000
0.6	0.4	0.03	0.8643	0.4320	0.4069	0.077	-0.15	-0.20	0.1002
0.6	0.4	0.06	0.8637	0.5068	-0.0127	0.775	-0.15	-0.20	0.1000
0.55	0.45	0.00	0.8547	0.3725	0.6275	0.0715	-0.15	-0.20	0.1000
0.55	0.45	0.03	0.8527	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



(a) With $\delta = 0$



(b) With $\delta = 0.03$



(c) With $\delta = 0.06$

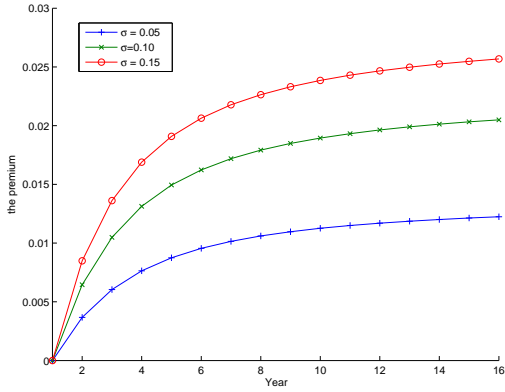
Figure 15: The lockup premium for the DTMC with $\gamma_G \neq \gamma_S$. $\gamma_G = 0.6, \gamma_S = 0.4$, $\gamma_G = 0.55, \gamma_S = 0.45$, and $\gamma_G = \gamma_S = 0.5$.

G.2. Sensitivity of p, q , and r with respect to δ, Y_G, Y_S , and Y_D

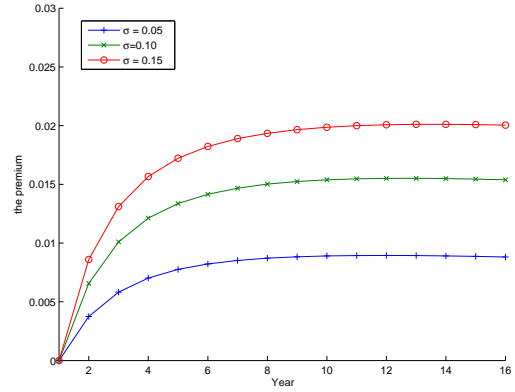
In this section, we observe the effect of δ to the implied transition probabilities p, q, r . Figure 6 is the implied transition probabilities for δ from 0 to 0.1. It is clear that the transition

Table 7: Parameter value sets for $\sigma = 0.05, 0.10, 0.15$

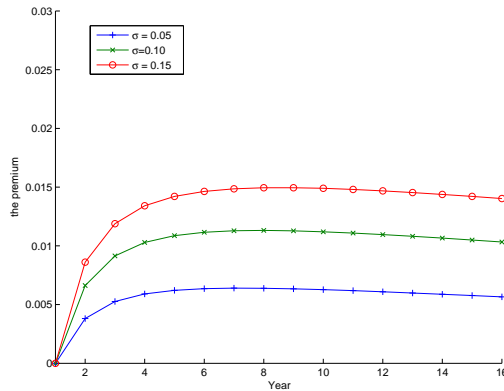
σ	δ	p	q	r	Y_G	$Y_S = -1.5\sigma$	$Y_D = -2.0\sigma$	σ (calculated)
0.05	0.00	0.8461	0.3461	0.6539	0.0333	-0.075	-0.10	0.05
0.10	0.00	0.8461	0.3461	0.6539	0.0667	-0.150	-0.20	0.10
0.15	0.00	0.8461	0.3461	0.6539	0.1000	-0.225	-0.30	0.15
0.05	0.03	0.8434	0.3721	0.5025	0.0342	-0.075	-0.10	0.05
0.10	0.03	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 $\delta = 0.00$



(b) With $\delta = 0.03$



(c) With $\delta = 0.06$

Figure 16: The lockup premium for the DTMC for $\sigma = 0.05, 0.10,$ and 0.15 .

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

$$\frac{\partial r}{\partial \delta} = -\frac{\frac{2-p-r}{1-p}(Y_G - Y_D)}{(Y_G - Y_S) - \frac{\delta}{1-p}(Y_D - Y_G)}. \quad (\text{G.1})$$

From (G.1), we observe that as δ increases, r decreases more rapidly. Note that the coefficient of δ in (G.1) 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 rate in the DTMC model is 0.07.

The implied transition probabilities are calculated for Y_G , Y_S , and Y_D in Figure 10 in §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 10. This justifies that our parameter fitting method which changes Y_G while fixing Y_S and Y_D since Y_G has greater effect to the implied transition probabilities than Y_S and Y_D . Notice that from (6.8) of the main paper, 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 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 17.

G.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 9 in the main paper, it is clear that the lockup premium increases as δ decreases, γ increases, or σ increases. 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 18 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 18 (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

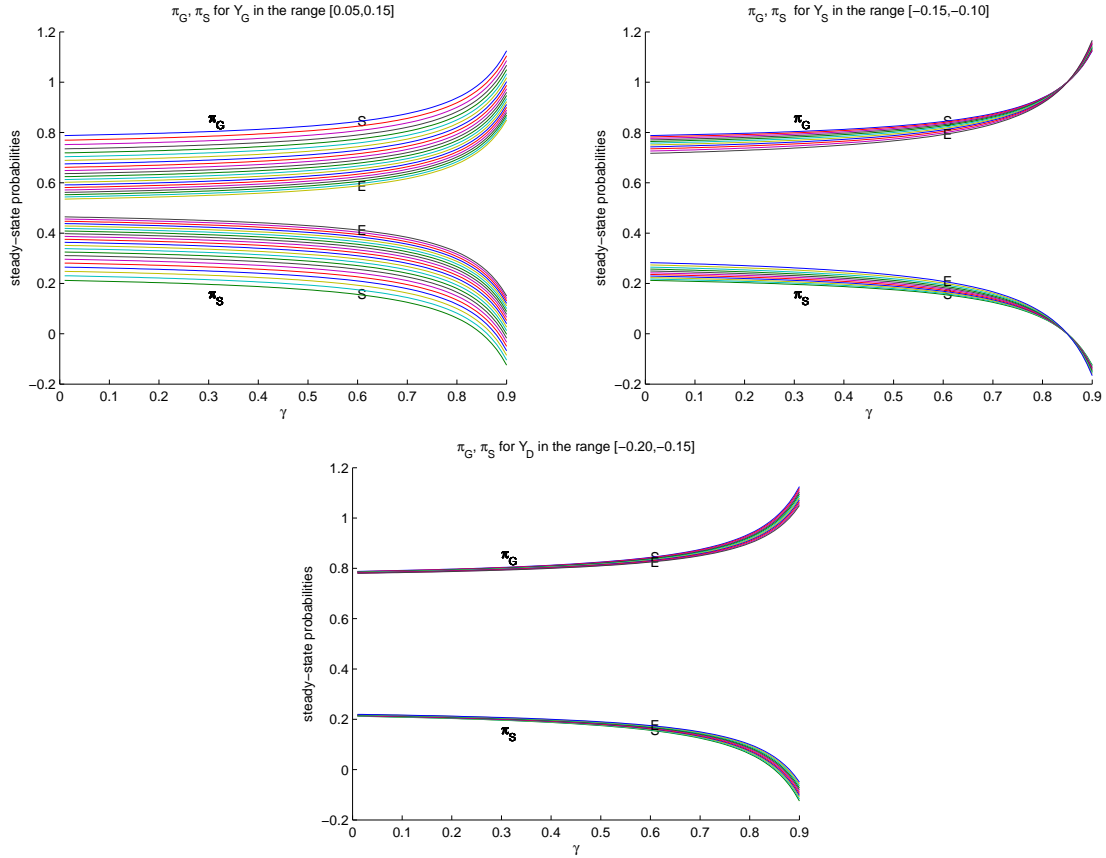
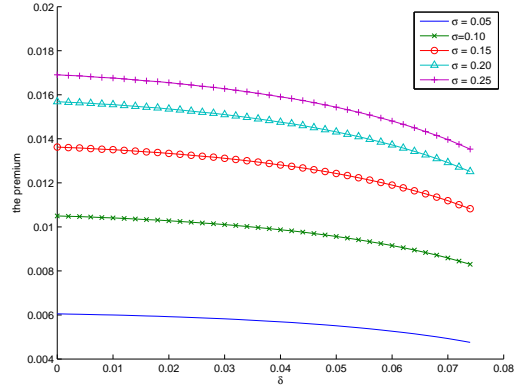
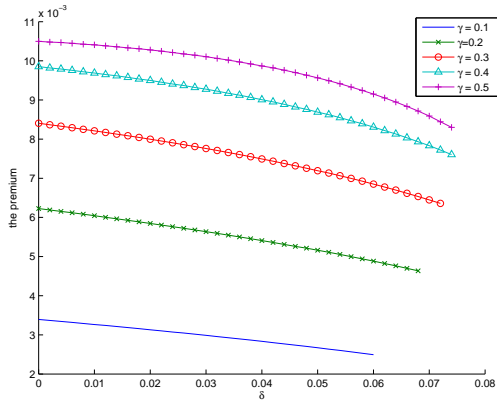


Figure 17: The steady-state probabilities π_G and π_S as a function of γ in the base case for values of Y_G ranging from 0.05 (starting value, denoted by S) to 0.15 (ending value, denoted by E), Y_S ranging from -0.15 to -0.10 , and Y_D ranging from -0.20 to -0.15 .

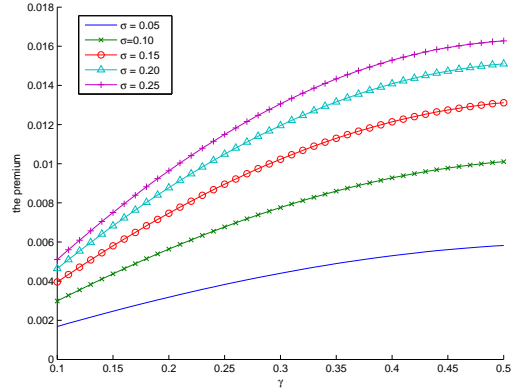
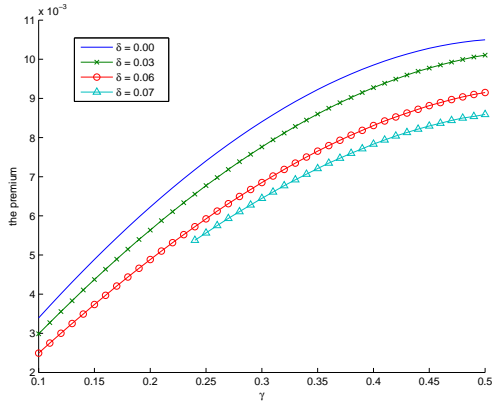
concavity of the premium for γ increases as γ increases. Figures 18 (i) and (ii) suggest that the three-year lockup premium is quite insensitive to δ , which implies that the effect of δ on the lockup premium is relatively small. Figure 18 (iii) and (iv) show the three-year lockup premium for $\delta = 0.03$. To supplement that, Figure 19 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.

G.4. Estimating the Functional Form of the Three-year Lockup Premium

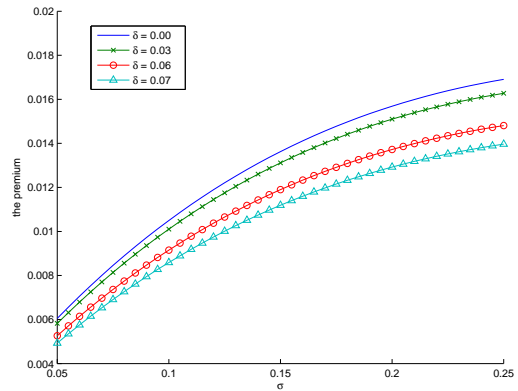
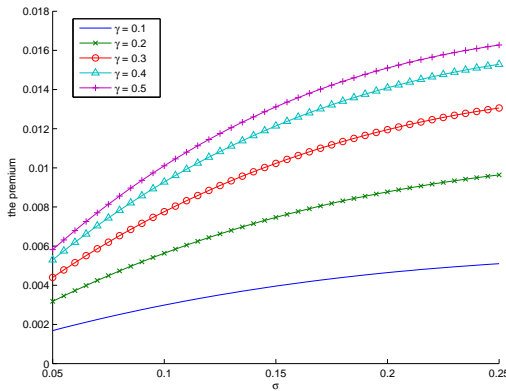
So far, we have calculated the lockup premium for variables γ , σ , and δ with the DTMC model. Since we have calculated the premium as a function of three variables, it is then natural to consider a simple functional form to describe the premium. If the estimation can be done relatively easily, it is useful to approximate the premium with a closed-form 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



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 18: 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.

may then understand the effect of these three variables more intuitively. 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

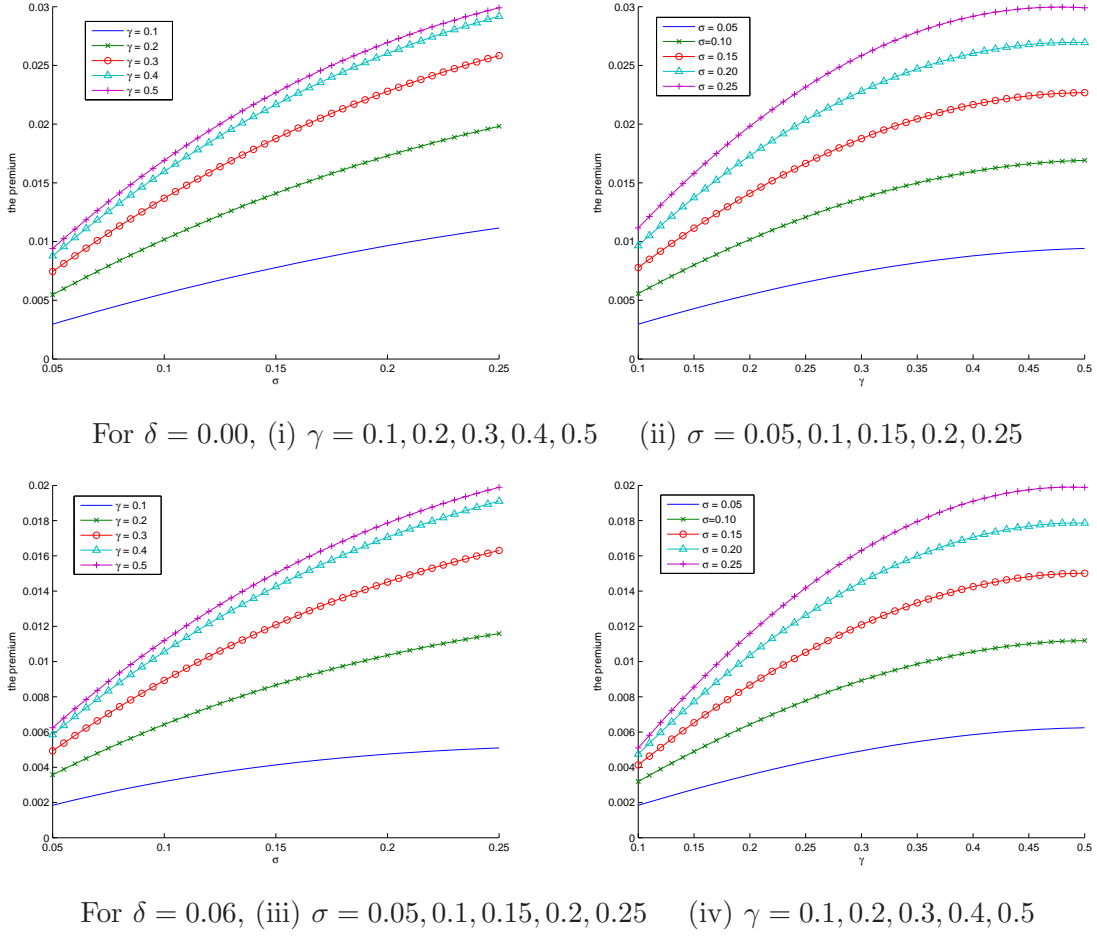


Figure 19: The three-year lockup premium for the DTMC model with $Y_S = -1.5\sigma, Y_D = -2.0\sigma$.

a good state may become 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 18 suggests that the three-year lockup premium is a weakly concave function of γ , a 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:

$$\psi_{(3)}(\delta, \gamma, \sigma) \approx \psi_{(3)}^p(\delta, \gamma, \sigma) \equiv a \delta^b \gamma^c \sigma^d. \quad (\text{G.2})$$

Taking logarithms of both sides of (G.2), it is straightforward to estimate the parameters a, b, c ,

and d from the calculated three-year lockup-premium values with linear regression, because

$$\ln \psi_{(3)}^p(\delta, \gamma, \sigma) = \ln a + b \ln \delta + c \ln \gamma + d \ln \sigma . \quad (\text{G.3})$$

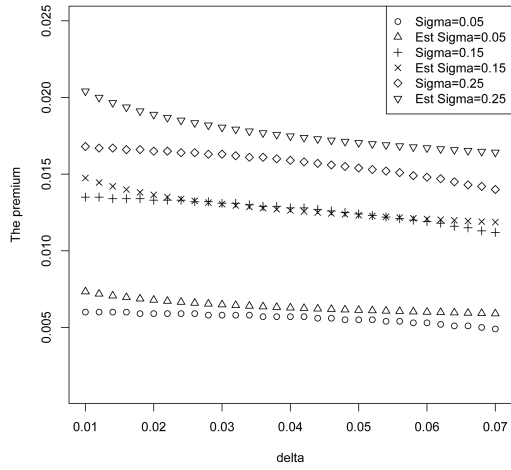
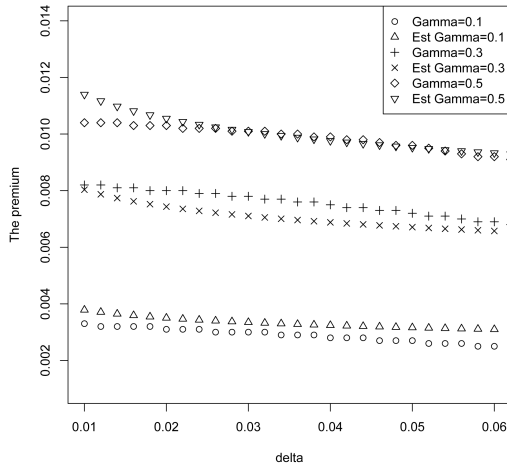
Since $\lim_{\delta \rightarrow 0} \psi_{(3)}^p = \infty$ when $b < 0$ and 0 when $b > 0$, which is not desirable for our estimation purpose, we have to restrict range of δ away from 0. Thus, we restrict the range of δ to $[0.01, 0.07]$.

It turns out that without further restricting the ranges of the variables δ, γ and σ , the candidate function $\psi_{(3)}^p(\delta, \gamma, \sigma)$ approximates the three-year lockup premium reasonably well. For example, for $\delta \in [0.01, 0.07]$, the linear regression of (G.3) approximates the three-year lockup premium by

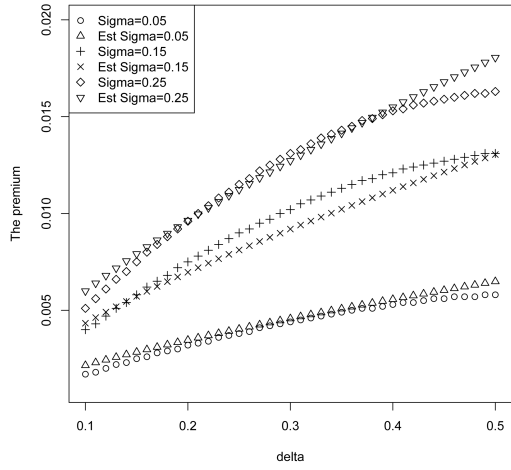
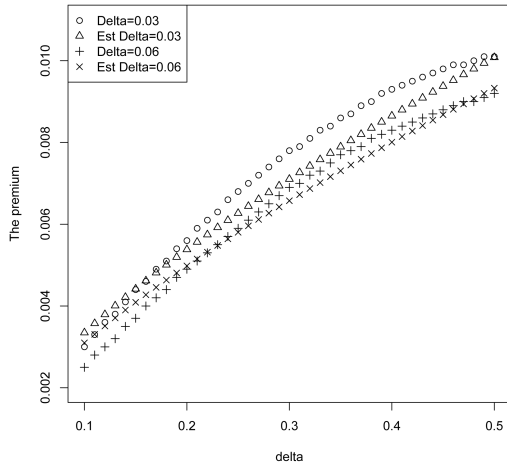
$$\psi_{(3)}^p(\delta, \gamma, \sigma) = 0.047 \delta^{-0.11} \gamma^{0.69} \sigma^{0.64},$$

with maximum error of 0.0036. Notice that the exponent to δ is -0.11 , which eventually makes $\lim_{\delta \rightarrow 0} \psi_{(3)}^p(\delta, \gamma, \sigma) = \infty$. 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 20 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.

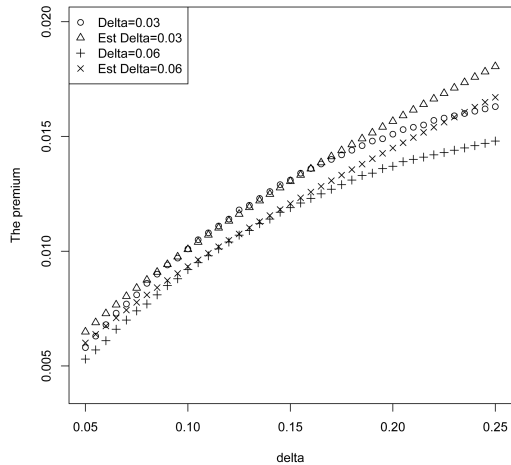
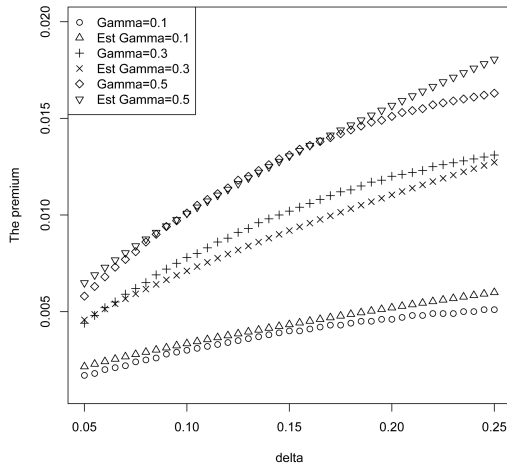
We remark that the approximation of the fixed-year lockup premium in the DTMC model by a product function $\psi^p(\delta, \gamma, \sigma) = a \delta^b \gamma^c \sigma^d$ works reasonably well for the other lockup periods (n) and other choices of Y_S/σ . For example, the four-year lockup premium in the DTMC model is approximated by $\psi_{(4)}^p(\delta, \gamma, \sigma) = 0.03 \delta^{-0.20} \gamma^{0.69} \sigma^{0.64}$ with maximum error of 0.0043. If we choose $Y_S/\sigma = -1.0$, $\psi_{(3)}^p(\delta, \gamma, \sigma) = 0.06 \delta^{-0.20} \gamma^{0.64} \sigma^{0.72}$ 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 $\psi_{(4)}^p(\delta, \gamma, \sigma) = 0.05 \delta^{-0.26} \gamma^{0.65} \sigma^{0.73}$ 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 20: Evaluating the product approximation $\psi_{(3)}^p(\delta, \gamma, \sigma) = 0.15 \delta^{-0.11} \gamma^{0.74} \sigma^{1.00}$ for the three-year lockup premium