Essays on the Empirical Properties of Stock and Mutual Fund Returns

by

Jonathan David Taylor

Submitted to the Operations Research Center
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Decision Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2000

© Massachusetts Institute of Technology 2000. All rights reserved.

Author .................................................. Operations Research Center
                                      May 9, 2000

Certified by ........................................... Andrew W. Lo
                                  Harris & Harris Group Professor of Finance
                                      Thesis Supervisor

Accepted by ........................................... Cynthia Barnhart
                                          Co-Director, Operations Research Center
1984-1996 for 660 mutual funds, we test for the other managerial gaming effects predicted by the model. Our results suggest a role for the theory of tournaments in studies of mutual fund behavior.

In "Are Stock Returns Stable" (Co-authored with Blake LeBaron and Andrew Lo) we revisit earlier studies of the marginal distribution of stock returns and the properties of stable laws. We apply recent advances in bootstrapping techniques to test whether stock returns exhibit stability under addition. Statistically, we are unable to reject stability for the indexes and large stocks in our sample. However, we find significant economic rejections of the stable hypothesis for all assets studied.

Thesis Supervisor: Andrew W. Lo
Title: Harris & Harris Group Professor of Finance
Acknowledgments

Whenever I pick up a thesis, I go straight to the Acknowledgements section. I guess I like to hear heartfelt words, even if I don’t really know all of the details. My story begins in the Fall of 1994 and ends in the Spring of 2000. The people who helped me along the way know all the details. I know these words are heartfelt.

James Mcleod, Massoud Amin, and William H. Danforth wrote recommendations for me in my application to MIT. I received a Hugh Hampton Young Memorial Fellowship because of Chancellor Danforth’s recommendation (this is the only plausible explanation). Alvin Drake gave me my first academic job - T.A. for 6.041. Rob Freund helped me work out what Decision Sciences means when the name is scratched off the door. Paulette Mosley made sure that people more powerful than I would respect that Decision Sciences is not Operations Research, even though I’m in the Operations Research Center. David Bell gave me a vision of what could be “out there” for me if I kept working hard. Andrew Lo showed me that I must believe if it is to be achieved. Jiang Wang overlooked my faults and helped me develop my ideas. Katharine Fennelly showed me that the grass is greener at the Harvard Business School. Mary Marshal found me TA positions in Finance, even though I’m not in Finance.

Wally’s Jazz Club - and more specifically John Lamkin on drums, Aaron Goldberg on piano, Darren Barrett on trumpet, Reuben Rogers on bass, and Jimmy Greene on tenor sax - soothed my soul on Friday and Saturday nights. St. John - St. Hugh’s Catholic Chruch in Dorchester took over from there on Sunday mornings. Dr. Howard Ramseur was the first to help me identify that my soul needed soothing.

Ramazan Demir showed me what it means to be a truly talented, hard working, and generous human being. Joelle Simpson rolled the stone away and has even agreed to marry me! Kimani Stancil (and the AFAM Chess Club) gave me a gift of chess and fellowship. Patrick Blattstone taught me how to create my own Jazz. The Samaritans of Boston enabled me to be a better one. Meyer Chambers (and the Office for Black Catholics) awakened in me a desire to serve.

My Mom and my Dad never let me forget that they love me. My Grandma always made me forget that MIT is hard. I’ll always remember that my sister Allyson was 100% in my
corner during the darkest hours. Scoot, Ron, and Kathy were never fair-weather friends.

That’s about it. I have a PhD in Decision Sciences and will start work as an Assistant Professor of Finance. There are more logical, more direct ways of doing what I’ve done. I don’t pretend that this thesis is all that great. Read it if you must, but please be understanding. God is not finished with me yet!
Contents

1 Does Survivorship Bias Matter? ........................................... 8
  1.1 Introduction .......................................................... 8
  1.2 Problem Formulation and Model Description ......................... 11
    1.2.1 Formal Statement of the Survivorship Bias Problem ............... 11
    1.2.2 Description of the Simulation Approach ......................... 12
  1.3 Estimation Procedure and Results ................................... 13
    1.3.1 Means, Standard Deviations, and Betas .......................... 13
    1.3.2 Explanatory Power of Beta ...................................... 20
    1.3.3 Short-Term Positive Autocorrelation .......................... 22
  1.4 Conclusion .................................................................. 27
    1.4.1 Means, Standard Deviations, and Betas ......................... 27
    1.4.2 Risk/Return Tradeoff .............................................. 28
    1.4.3 Return Predictability ............................................ 29

2 A Role for the Theory of Tournaments in Studies of Mutual Fund Behavior 30
  2.1 Introduction .......................................................... 30
  2.2 Model .................................................................. 33
    2.2.1 Structure .......................................................... 34
    2.2.2 Analysis of Type I Games ...................................... 37
    2.2.3 Analysis of Type II Games ..................................... 41
  2.3 Empirical Analysis ..................................................... 45
    2.3.1 Data Description ................................................. 45
    2.3.2 Methodology ..................................................... 47
2.3.3 Results .................................................. 49
2.4 Conclusion ............................................... 52

3 Are Stock Returns Stable? .......................... 54
  3.1 Introduction ......................................... 54
  3.2 A Review of Stable Distributions ............... 56
  3.3 Test of Stability – Stability Under Addition ... 58
    3.3.1 Implementation of the Test Statistic ....... 58
    3.3.2 Bootstrap Methods and Stable Distributions 60
  3.4 Empirical Results ..................................... 61
    3.4.1 Data and Implementation ...................... 61
    3.4.2 Results of the Stability Under Addition Test 65
  3.5 Stable Laws and Risk Management ............... 68
  3.6 Conclusion ............................................. 71

A Code and Derivations for Survivorship Analysis .... 72
  A.1 Splus Code for Means, Variances, Beta & Momentum Strategies 72
  A.2 Splus Code for Cross-Sectional Regressions ............ 75
  A.3 Derivation of the Total Variance of Stock and Mutual Fund Returns 76
  A.4 Conditional Expectation of a Truncated $N(0,\sigma^2)$ Random Variable 77

B Code & Fund Information for Tournament Analysis ... 78
  B.1 Splus Code for Tournament Analysis ............... 78
  B.2 Lipper-Classified Growth Funds ..................... 79

C Splus and C Code for Estimation and Bootstrap ....... 85
    C.2.1 C Code to Resample Random Blocks of 50 ("blocker") 87
    C.2.2 C Code to Form 50 Non-Overlapping Series’ ("datamaker") 90

Bibliography ............................................... 95
Chapter 1

Does Survivorship Bias Matter?

Abstract

Survivorship bias influences statistical inference in Finance. Through a series of Monte Carlo simulations in the style of Brown, Goetzmann, Ibbotson, and Ross (1992), we study the sampling distribution of the mean return, standard deviation, beta, Fama & MacBeth (1973) t-statistic, and Jegadeesh & Titman (1993) momentum strategy return in progressively truncated datasets. Survivor-biased datasets have higher mean returns, lower return standard deviations and lower betas than the full sample. Beta has no explanatory power even when the CAPM is true, a finding virtually unaffected by survivorship bias. Returns to a momentum strategy are positive even when stock idiosyncratic returns are serially and cross-sectionally uncorrelated, but survivorship bias overestimates the returns and underestimates the beta of the strategy.

1.1 Introduction

Past performance does not predict future results.

There is a tradeoff between risk and expected return.

These statements of market efficiency and capital market equilibrium form the cornerstones of modern financial economics. Yet, within the last few years, researchers have uncovered empirical regularities that appear to provide real counterexamples. Among these are:

1. the zero explanatory power of $\beta$ in cross-sectional regressions of returns on beta documented by Fama & French (1992), and

These anomalies (and a host of others not mentioned here) have encouraged some researchers to develop behavioral models of price formation based on psychological experiments. For instance, Daniel, Hirshleifer & Subrahmanyam (1998) explain the positive short-term autocorrelation of stock returns with a model of overconfident investors who overestimate the precision of their private signals. In the model, subsequent public signals confirm initial private signals on average because investor overconfidence causes them to overweight confirming information and underweight disconfirming information. The result is subsequent trading in the direction of the initial private signal, i.e. momentum.

The anomalies have also encouraged researchers to generalize rational models of price formation to include additional factors, time-varying parameters, and allowances for model misspecification. For instance, Fama & French (1996) and Fama (1998) demonstrate that many of the anomalies are not robust to a multifactor specification of the capital asset pricing model. They propose a model that, though unable to explain the positive short-term autocorrelation of stock returns documented by Jegadeesh & Titman (1993), captures the returns to portfolios formed on Earnings/Price, Cash/Price, Sales Growth and long-term past returns (DeBondt & Thaler - 1985).

Jagannathan and Wang (1996) generalize the capital asset pricing model to include time-varying betas and market risk premia. The specification they test includes the yield spread between BAA- an AAA-rated bonds as a proxy for the market risk premium. It also expands the market portfolio to include a measure of human capital. They find that the addition of these time-varying variables increases the explanatory power of the model over the static CAPM and leaves little for the Fama & French (1996) size and book-to-market factors to explain.

Roll & Ross (1994) and Kandel & Stambaugh (1995) demonstrate the effects of model misspecification. They warn that a (slightly) inefficient proxy for the true market portfolio can yield a flat relationship between beta and expected returns. These inefficient betas, if used in Fama-MacBeth (1973) regressions, can fail to drive out other "explanatory" variables.

Ultimately, researchers who adopt psychology-based models of price behavior and re-
searchers whose models are based on the theory of rational choice must demonstrate their ability to explain and predict the behavior of financial markets. Unfortunately, both groups are limited in this pursuit by the quality of the available data.

In this paper, we study stocks in a perfect world where we know for certain that the past does not predict the future and that risk is related to return through the traditional Sharpe (1964), Lintner (1965), and Black (1972) CAPM. The one hitch is that we carry out our empirical analysis only on those stocks that survive a series of market tests. We compare the sampling distributions of various financial statistics in survivor-biased datasets with those obtained from unbiased samples. The results allow us to extend the work of Brown, Goetzmann & Ross (1995) by describing further "what researchers should expect to find due to survival alone" in the hopes of "disentangling survival effects from meaningful economic phenomenon."

We focus on three sets of statistics:

- descriptive statistics, including the average mean, standard deviation and beta,
- the explanatory power of $\beta$ in cross-sectional regressions of return on beta, and
- the returns to momentum strategies.

We find that

1. Survivor-biased datasets have higher mean returns, lower return standard deviations and lower betas than the full sample.

2. Beta has no explanatory power even when the CAPM is true, a finding virtually unaffected by survivorship bias.

3. Returns to a momentum strategy are positive even when stock idiosyncratic returns are serially and cross-sectionally uncorrelated, but survivorship bias overestimates the returns and underestimates the beta of the strategy.

The paper proceeds as follows. In Section II, we formalize the problem of survivorship bias and describe our simulation approach to estimating its effects. In Section III, we describe the estimation procedure for each set of statistics and present the simulation results. Section IV discusses the results and concludes.
1.2 Problem Formulation and Model Description

1.2.1 Formal Statement of the Survivorship Bias Problem

In this section, we provide a formal statement of the problem in the style of Heckman (1976). The resulting discussion and example illustrates how a survivorship criteria introduces bias in the coefficients of a standard OLS regression.

To clarify ideas, consider a market with $N$ stocks, where the relationship between the past return of stock $i$, $X_i$, and its future return $Y_i$ is given by

$$Y_i = \gamma X_i + U_i,$$  \hspace{1cm} (1.1)

with the usual assumptions that the error $U_i$ is mean zero and uncorrelated across stocks.

For any stock $i$, the conditional expected future return given the past return is given by:

$$E(Y_i|X_i) = \gamma X_i.$$  \hspace{1cm} (1.2)

However, if the investor observes only the subset of stocks that have met a survivorship criteria, equation (2) becomes

$$E(Y_i|X_i, \Pi) = \gamma X_i + E(U_i|\Pi),$$  \hspace{1cm} (1.3)

where $\Pi$ denotes the criteria.

If the conditional expectation of $U_i$ is zero i.e. selection into the subset is random, regressions fit on the subset yield unbiased estimates of $\gamma$. However, this is not the case in general.

In particular, if stocks drop out of the sample because their realized average returns are low, we observe a stock’s return only if

$$\Pi : X_i, Y_i \geq C$$  \hspace{1cm} (1.4)

Equation (3) and equation (4) suggest that even if the true relationship between past
and future returns is zero (as in a weak-form efficient market with $\gamma = 0$), the conditional expectation of the future return $E(Y_i|X_i, \Pi)$ can be non-zero due to a non-zero conditional expected residual $E(U_i|\Pi)$.\footnote{Heckman (1976) demonstrates that the estimate of $\gamma$ is biased unless we include $E(U_i|\Pi)$ as a regressor in (1).} In other words, the investor who sees $\hat{\gamma} > 0$ in his survivorship-biased sample can not conclude that past outperformers are likely to have high future returns in general.

The effect of survivorship bias in OLS regressions is captured by $E(U_i|\Pi)$ in equation (3). In the following section, we describe our simulation approach to estimating the effects of survivorship bias in more general models.

1.2.2 Description of the Simulation Approach

In this section, we describe the structure of a simulation we use throughout the paper to study the effects of survivorship bias on statistical inference in Finance.

The main object of study is a 72-month return series for 1000 stocks that we generate according to the market model

$$x_{it} = r_f + \beta_i(r_{mt} - r_f) + \epsilon_{it}. \quad (1.5)$$

We assume that $r_f$, the monthly return on a 30-day U.S. Treasury Bill, is equal to .31%. We model $r_{mt} - r_f$, the monthly market risk premium, as a normally distributed random variable with mean .71% per month and standard deviation 5.7% per month. These parameter choices are based on historical monthly total returns of S&P 500 stocks and U.S. Treasury Bills, as reported by Ibbotson Associates over the 70-year period 1926-1996.

We assume that the monthly non-systematic component of the stock return is normally distributed with zero mean and standard deviation $\sigma_i$. From the estimates of $\beta_i$ and $\sigma_i$ in the September 1998 "Beta Book," we obtain some sense of the cross-sectional dispersion in beta among currently listed stocks and of the relationship between beta and non-systematic risk. The relationship between $\sigma_i$ (in percent per month) and $\beta_i$ suggested by the data is
given by

$$\sigma_i = 2.55\beta_i^2 - 1.25\beta_i + 12.71,$$  \hspace{1cm} (1.6)

with $R^2 = .22$ and $N = 9281$. The cross-sectional dispersion in beta is modeled by a normal random variable with mean 0.82 and standard deviation 0.79.\footnote{All of these relationships are basically the same in the August 1998 Beta Book.} Note that individual stock returns are serially uncorrelated and the non-systematic portion of these returns are cross-sectionally uncorrelated.

The test methodology is as follows: First, we choose 1000 $\beta$’s from the “Beta Book” distribution. A choice of $\beta_i$ fixes a value of $\sigma_i$ through equation (6). Next, we generate a 72-month return series for 1000 stocks according to equation (5). We then form three additional return series’ by removing the worst performing $\Pi\%$ of the stocks on a total return basis from the sample each year ($\Pi = 5, 10, 20$).\footnote{Chan, Jegadeesh & Lakonishok (1995) find that as few as 3\% and as many as 15\% of the companies in the CRSP database have missing data on COMPUSTAT.} For instance, the return series corresponding to a 10\% performance cut includes those stocks whose annual returns in each of the six years ranked in the top 90\% of their surviving peers.

Throughout the next section, we study the sampling distributions of a number of important statistics in Finance as a function of $\Pi$. Each sampling distribution is based on 20,000 replications of the methodology above. The Splus code for the procedure is located in Appendix A.

1.3 Estimation Procedure and Results

1.3.1 Means, Standard Deviations, and Betas

We begin our analysis by examining the average mean return, standard deviation, and beta of stocks in datasets suffering from survivorship bias. Our main findings generally verify the Brown, Goetzmann, Ibbotson & Ross (1992) suggestion that “survivorship bias will lead to obvious biases in first and second moments and cross moments of return, including beta.”
However, the results also suggest that survivorship bias effects stock and mutual fund studies differently.

As demonstrated for mutual funds in Brown & Goetzmann (1995) and Malkiel (1995), Table I shows that stocks that meet a survival criteria have higher average returns than the universe of stocks from which they are selected. We find that the median average monthly return for stocks meeting a 5% performance cut is .4% higher than entire sample. This translates into a 4.8% annual return difference.

By contrast, Brown & Goetzmann (1995) find a .8% annual return difference between the returns of mutual funds that survived from 1977 through 1988 and the entire universe of funds that existed during the period. A similar analysis by Malkiel (1995) finds a 1.4% difference for mutual funds over the 1982-1991 period. The simulation study of Brown, Goetzmann, Ibbotson & Ross (1992) finds a .4% difference for mutual funds over a set of simulated 4-year periods.

The analysis of Malkiel (1995) suggests that 5% of the existing mutual funds at the end of a given year disappear by the end of the following year.\(^4\) Why then does the same mortality rate in our model generate a significantly higher annual return difference for stocks? The answer is stocks have higher return variance than mutual funds. Hendricks, Patel Zeckhauser (1997) and Brown, Goetzmann, Ibbotson & Ross (1997) demonstrate that the expected return of surviving high variance assets is larger than the expected return of surviving low variance assets when all assets have the same unconditional mean return.

As a concrete example of the principle, consider the variance of mutual fund returns in the BGIR (1992) study. They also model fund returns according to the market model in equation (5), but with a relationship between nonsystematic risk and beta based on actual mutual fund returns. Their relationship is given by \(0.05349(1 - \beta)^2\). With these parameters and with their model of beta as normally distributed with mean .95 and standard deviation .25, we obtain an unconditional variance in mutual fund returns of .38% per month. A similar calculation with the parameters of our model yields a variance in stock returns of 2.79% per month.\(^5\)

---

\(^4\)The 1998 NYSE Fact Book suggests a similar fraction of existing stocks leave that exchange each year.

\(^5\)For details, see Appendix A.
Now consider a mutual fund and a stock with identical expected returns subject to a 5% performance cut, i.e. the returns of each show up in our datasets only if they are in the top 95% of their respective distributions. We can estimate the difference in conditional expected returns between the stock and the mutual fund by making use of the following fact derived in Appendix A:

\[
E(x|x \geq k) = \frac{\sigma}{\sqrt{2\pi(1 - \Phi(\frac{k}{\sigma}))}} e^{-\frac{k^2}{2\xi^2}}
\]  

(1.7)

for a \(N(0,\sigma^2)\) random variable.

When \(k\) is chosen so that \(\frac{k}{\sigma} = -1.645\) (a 5% performance cut) and the stock variance and mutual fund variance are .0279 and .0038 respectively, the difference in conditional expected returns is 13.7% per year - more than enough to account for the annual return difference we observe in our simulation.

Ultimately, a precise explanation of the return difference depends on the assumptions we make about the statistical properties of stock and mutual fund returns and the exact way in which stocks and funds are removed from the sample. But, to the extent that stock market returns have higher variance than actual mutual fund returns, our results suggest that survivorship bias has a much larger influence on estimates of average return in stock studies than in mutual fund studies, even if the attrition rates in both markets are similar. Or put another way, even small attrition rates in stock markets can have large effects on estimates of the mean return in biased samples.

Survivorship bias has the same effect on the average variance among surviving stocks as it does on surviving mutual funds, however. BGIR (1997) suggests that conditioning upon low returns can “pick out” the more volatile funds. We find that surviving stocks indeed tend to be less variable on average than the universe of stocks from which they are selected. On average, the median average monthly standard deviation for surviving stocks is about .7% lower than for the sample as a whole.

However, where the effect of survivorship bias on the mean and variance is simple to understand, its effect on beta is much more complex. BGIR (1992) finds that the average beta among surviving mutual funds is higher than among the entire sample. They also
find that the average beta among surviving funds increases as a function of the degree of survivorship bias.

Like BGIR (1992), we find that the average beta among surviving funds increases as a function of survivorship. However, we also find that the average beta tends to be lower among surviving stocks, except for extremely high rates of survivorship bias. What accounts for the difference? Again, the higher variance in stock returns plays a role, but the total effect results from a complex interaction between the cross-sectional dispersion in stock variances, the distribution of beta in the sample, and our particular model of survivorship bias.

To see this, consider the situation in Figures I and II which plot the total variance of a BGIR (1992) mutual fund and a stock, respectively, versus their beta, for a range of beta values. We also include the corresponding density functions of beta.

First, we show, by way of a graphical argument, that the average mutual fund beta is larger among surviving funds. Figure I shows that most of the fund betas are concentrated on a small range from 0 to 2. We also see that there is very little dispersion in total fund return variance in the BGIR (1992) model. Since the return variances are pretty much the same across funds, beta acts more like an index of expected returns than of risk. As market returns are positive on average, poor performers tend to be low beta funds. Conditional on survival, therefore, the mean beta is higher.

A similar argument leads to the opposite result for our stocks. Figure II shows that stocks have a wider range of possible beta values. We also see that there is more dispersion in total stock return variance in our model. Since large beta stocks tend to have the highest total return variances, beta acts more like an index of risk than of expected returns. As survivorship bias “picks out” high variance stocks, high beta stocks are more likely to drop from the sample. Conditional on survival, therefore, the mean beta is lower.

As the performance cut deepens however, surviving stocks (and surviving funds) tend to become more homogenous in their return variances. Since market returns are positive on average, poor performers tend to be low beta stocks, all other things equal. Conditional on survival, therefore, the mean beta increases in the size of the performance cut for stocks and mutual funds.

The results have immediate implications for the size effect of Banz (1981). Fama &
French (1992) find that beta and size are almost perfectly negatively correlated ($\rho = -.98$). If high beta stocks are likely to be small stocks as well, then our results suggest that the small stocks in survivor biased datasets are likely to have lower betas and higher returns. Size effects, therefore, can be artifacts of survivorship bias as well as proxies for additional factors in the true capital asset pricing model.
Table I  
Descriptive Statistics for Surviving and Non-Surviving Stocks

72-months of stock return data for 1000 stocks is simulated according to the market model in equation (5). For each market replication, we calculate the average 72-month mean return, standard deviation, and true beta of surviving and non-surviving (NS) stocks in the sample. A Π% performance cut is modeled by removing the worst performing Π% of the stocks (on a total return basis) in the sample each year. The 2.5, 50, and 97.5%iles of the sampling distribution of the average mean return, standard deviation, and beta are presented for surviving and non-surviving stocks (NS) at various performance cuts based on 20,000 replications.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Mean</th>
<th>Mean (NS)</th>
<th>Dev</th>
<th>Dev(NS)</th>
<th>Beta</th>
<th>Beta (NS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5%ile</td>
<td>-0.002</td>
<td>—</td>
<td>0.156</td>
<td>—</td>
<td>0.771</td>
<td>—</td>
</tr>
<tr>
<td>50%ile</td>
<td>0.009</td>
<td>—</td>
<td>0.160</td>
<td>—</td>
<td>0.820</td>
<td>—</td>
</tr>
<tr>
<td>97.5%ile</td>
<td>0.020</td>
<td>—</td>
<td>0.165</td>
<td>—</td>
<td>0.869</td>
<td>—</td>
</tr>
</tbody>
</table>

Panel A: No Performance Cut

| 2.5%ile | 0.003 | -0.016 | 0.149 | 0.166 | 0.670 | 0.612 |
| 50%ile  | 0.013 | -0.003 | 0.154 | 0.179 | 0.784 | 0.919 |
| 97.5%ile| 0.025 | 0.005  | 0.160 | 0.192 | 0.901 | 1.215 |

Panel B: 5% Performance Cut

| 2.5%ile | 0.007 | -0.012 | 0.146 | 0.160 | 0.620 | 0.651 |
| 50%ile  | 0.017 | 0.000  | 0.152 | 0.169 | 0.794 | 0.850 |
| 97.5%ile| 0.030 | 0.008  | 0.161 | 0.179 | 0.966 | 1.050 |

Panel C: 10% Performance Cut

| 2.5%ile | 0.013 | -0.007 | 0.143 | 0.157 | 0.557 | 0.703 |
| 50%ile  | 0.023 | 0.004  | 0.153 | 0.163 | 0.843 | 0.812 |
| 97.5%ile| 0.039 | 0.013  | 0.167 | 0.170 | 1.112 | 0.925 |
Figure I
Total Return Variance vs. Beta for Funds

In Figure I, we plot the BGIR (1992) relationship between fund beta and total return variance. A scaled plot of the distribution of fund betas is also presented. Figure II contains the same information for our simulation of stock returns.

Figure II
Total Return Variance vs. Beta for Stocks
1.3.2 Explanatory Power of Beta

In this section, we consider the effect of survivorship bias on the ability of beta to explain the cross-section of stock returns, via the methodology of Fama & MacBeth (1973). Our main finding is a confirmation of the Affleck-Graves & Bradfield (1993), Chan & Lakonishok (1993) and Malkiel (1997) conjectures that the Fama-MacBeth test has very low power, even when the CAPM holds. We also find that the power of the test is virtually unaffected by the amount of survivorship bias.

We begin with a discussion of the methodology. First, we estimate betas for all 1000 stocks based on the first 36 months of return data. For each month (beginning with month 37 and ending with month 72), we regress the cross-section of monthly stock returns on the beta estimates. Since, by construction, the betas are constant for the entire sample period, we do not use rolling estimates.

Fama & MacBeth (1973) and Fama & French (1992) use portfolio betas instead of individual stock betas because the latter tend to be noisy. To guard against potential “errors-in-variables” problems, we repeat the analysis exploiting our knowledge of the true individual stock betas. The second analysis provides a benchmark for interpreting our results. Kim (1995) and Shanken (1992) demonstrate that traditional statistical inference in a Fama-MacBeth framework tends to overstate the precision of slope estimates, due to estimation errors in beta, though Jagannathan & Wang (1998) dispute this view. Kothari, Shanken & Sloan (1995) demonstrates that inferences about the explanatory power of beta can be influenced by using beta estimates obtained from annual versus monthly data. We can avoid all of these complications by using the true betas.

In both analyses, we perform a simple \( t \)-test on the average of the 36 regression coefficients that result. Table II presents, for both analyses, the fraction of \( t \)-statistics that are greater than two over all 20,000 replications for various performance cuts.

Notice that beta has no explanatory power more than 87% of the time, even in a market where the static CAPM holds. This confirms the Affleck-Graves & Bradfield (1993) finding obtained through a similar analysis. It also provides a convincing counterexample to Grauer (1999) who suggests that “the only way the regression coefficients can lead us to believe
that the model is false is if we employ a proxy for the market portfolio.” Interestingly, a
perfect solution to the errors-in-variables problem improves the power, but only slightly. The
improvement is never more than .5%.

How does survivorship bias effect the results? The answer is little, if at all. Table III
document
docs a modest increase in the median t-statistic as survivorship bias increases, with
more dramatic increases when the true betas are used in the regression. Our explanation
of this result closely parallels our argument for why average beta increases as a function of
survivorship bias. As the performance cut deepens, surviving stocks tend to become more
homogenous in their return variances. The result is a sample of stocks with comparable
return variances but different betas. Affleck-Graves & Bradfield (1993) demonstrate that
beta does a better job explaining returns in such a sample. Table II shows the effect on the
power of the test is negligible, however.
Table II
Power of the Fama-MacBeth Regression Test

72-months of stock return data for 1000 stocks is simulated according to the market model in equation (5). For each market replication, we estimate the betas for all 1000 stocks based on the first 36 months of return data. We carry out Fama-MacBeth regressions on the remaining 36 months of data using the estimated betas. We calculate the t-statistic in a test that the average slope coefficient in the 36 regressions is zero. We present the fraction of replications with t-statistics greater than 2 for various performance cuts based on 20,000 replications. Results from an analysis in which the true betas are used as independent variables in the Fama-MacBeth regressions are also presented for comparison.

<table>
<thead>
<tr>
<th>Performance Cut</th>
<th>% of t-statistics &gt; 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated Beta</td>
</tr>
<tr>
<td>None</td>
<td>11.22</td>
</tr>
<tr>
<td>5%</td>
<td>12.23</td>
</tr>
<tr>
<td>10%</td>
<td>12.19</td>
</tr>
<tr>
<td>20%</td>
<td>12.00</td>
</tr>
</tbody>
</table>

Table III
Survivorship Bias and the Fama-MacBeth Regression Results

72-months of stock return data for 1000 stocks is simulated according to the market model in equation (5). For each market replication, we estimate the betas for all 1000 stocks based on the first 36 months of return data. We carry out Fama-MacBeth regressions on the remaining 36 months of data using the estimated betas. We calculate the t-statistic in a test that the average slope coefficient in the 36 regressions is zero. We present the 2.5, 50, and 97.5%iles of the sampling distribution of the t-statistics for various performance cuts based on 20,000 replications. Results from an analysis in which the true betas are used as independent variables in the Fama-MacBeth regressions are also presented for comparison.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>No Cut</th>
<th>5% Cut</th>
<th>10% Cut</th>
<th>20% Cut</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>True</td>
<td>Est.</td>
<td>True</td>
</tr>
<tr>
<td>2.5%ile</td>
<td>-1.275</td>
<td>-1.239</td>
<td>-0.735</td>
<td>-0.707</td>
</tr>
<tr>
<td>50%ile</td>
<td>0.755</td>
<td>0.750</td>
<td>0.983</td>
<td>0.987</td>
</tr>
<tr>
<td>97.5%ile</td>
<td>2.833</td>
<td>2.834</td>
<td>2.762</td>
<td>2.766</td>
</tr>
</tbody>
</table>

1.3.3 Short-Term Positive Autocorrelation

In this section, we study the effect of survivorship bias on studies of stock price momentum, via the methodology of DeBondt & Thaler (1985) and Jegadeesh & Titman (1993). Our simulations confirm the Lo & MacKinlay (1990) and Conrad & Kaul (1998) result that momentum strategies can earn positive returns even when stock idiosyncratic returns are
serially and cross-sectionally uncorrelated by construction. Survivorship bias effects the results in two ways. First, it makes momentum portfolio returns appear larger. Second, and more importantly, it makes the estimated beta of the momentum strategy appear lower than its true beta.

As before, we begin with a discussion of the methodology. We calculate the 36-month cumulative excess return of all 1000 stocks and label the top decile “winners” and the bottom decile “losers.” Excess returns are defined as the raw difference between a stock’s return and the return of the market index, in accordance with DeBondt & Thaler (1985).

We buy the winners, sell the losers, and hold the portfolio for 36 months. We observe the $k$-month cumulative excess returns on the portfolio ($k = 3, 6, 9, 12, 18, 24, 30, 36$).

Table IV presents the 2.5, 50, and 97.5%iles of the $k$-month cumulative excess portfolio return for various performance cuts based on 20,000 replications.

First, note that the median cumulative excess return to the strategy is positive. In fact, the mean cumulative excess return is significantly positive for all $k$ and across all performance cuts. No $t$-statistic is less than 17.

We can use the analysis of Lo & MacKinlay (1990), Jegadeesh & Titman (1993), and Conrad & Kaul (1998) to interpret the positive returns. The analysis centers on a trading strategy that weights stocks at time $t$ by the amount of their past $k$-period excess return with respect to an equally weighted index. Lo & MacKinlay (1990) demonstrate that the future $k$-period profit $\pi_k$ from this weighted relative strength strategy (WRSS) - a strategy which is similar to buying winners and selling losers - equals

$$\pi_k = \sum_{i=1}^{N} \omega_{it}R_{it+k}$$

$$= \sum_{i=1}^{N} \frac{1}{N}(R_{it-k} - R_{mt-k})R_{it+k}$$

which has expected value

$$E[\pi_k] = \sigma^2_{\mu_t} - \text{Cov}[R_{mt-k}, R_{mt+k}] + \text{Cov}[R_{it-k}, R_{it+k}]$$

(1.8)

The second and third terms indicate the amount of profit in the WRSS that comes from
serial correlation in market returns and in individual stock returns, respectively. In our simulation, we have set them equal to zero by construction. The only remaining source of positive profit, therefore, is the first term, which measures the amount of cross-sectional dispersion in expected returns.

Jegadeesh & Titman (1993) dismiss this explanation for positive momentum strategy returns for two reasons: First, the post-ranking beta of the strategy is negative in their dataset. Second, momentum portfolio returns are not significantly smaller in their size- and beta-sorted subsamples.

We take up the issues in reverse order. First, we claim that size- and beta-sorting has an ambiguous effect on momentum portfolio returns. We agree that if size and beta are related to expected returns, then momentum profits in size- and beta-sorted subsamples are smaller ceteris paribus because the sorting procedure reduces $\sigma_{\mu_i}^2$. However, equation (8) shows that momentum profits depend on the serial covariance in the market return $\text{Cov}[R_{mt-k}, R_{mt+k}]$ and the average covariance in individual stock returns $\overline{\text{Cov}}[R_{it-k}, R_{it+k}]$ as well.

It is not obvious that these terms are unaffected by the sorting procedure. Jegadeesh & Titman (1993) find that momentum profits are usually high in small-size subsamples. Equation (8) shows that the result can be due to high positive average covariance in small stock returns (perhaps due to infrequent trading) or large negative covariance in the small stock index. Given that Jegadeesh & Titman (1993) find “that the serial covariance of 6-month returns of the equally weighted index [of NYSE/AMEX stocks] is negative” either situation can apply. Since increases in the second and third terms of equation (8) can mask reductions in the first term, the sorting results do not argue unambiguously against a role for $\sigma_{\mu_i}^2$ in explaining momentum profits.

Now, we claim that the post-ranking momentum portfolio beta does not proxy for the cross-sectional dispersion in mean returns. Conrad & Kaul (1998) provide an excellent counterexample. They estimate the three terms of equation (8) directly on a sample of NYSE/AMEX stocks during the 1926-1989 period. Their analysis includes a 1968-1989 subperiod which overlaps with the 1965-1989 period studied by Jegadeesh & Titman (1993).

They consider momentum strategies with $k$-month formation periods and $k$-month holding periods ($k = 3, 6, 9, 12, 18, 24, 36$). For all $k$, they obtain estimates of $\sigma_{\mu_i}^2$ and $\hat{P} =$
\[-\text{Cov} [R_{mt-k}, R_{mt+k}] + \overline{\text{Cov}} [R_{it-k}, R_{it+k}],\] where \(\hat{P}\) can be thought of as the total profit from "predictability."

In the 1968-1989 subperiod, for all \(k\), the average \(\sigma_{\mu_i}^2\) is significantly greater than zero. More importantly, if Jegadeesh & Titman (1993)'s post-ranking momentum portfolio beta of -.08 implies that cross-sectional dispersion in mean returns is insignificant, then momentum profits are due entirely to \(\hat{P}\). However, Conrad & Kaul (1998) find that the average \(\hat{P}\) is significantly negative for all \(k\). Certainly, a small post-ranking portfolio beta does not argue unambiguously against a role for the first term in equation (8).

Yet, one still expects to see momentum strategies with positive betas if their returns are positive on average. This leads us to ask whether survivorship-induced biases in the estimation of beta can at least partially account for the low post-ranking momentum portfolio beta in Jegadeesh & Titman (1993).

We explore the question more fully in Tables V. There we display the 2.5, 50, and 97.5%iles of the sampling distribution of the true and estimated momentum strategy betas for the 36-month formation period. Table V shows that the estimated portfolio beta is smaller than the true portfolio beta across all non-zero performance cuts. The differences between the estimates and the true values are significantly negative as well (maximum \(t\)-statistic: -9.5). The tables also show that the negative bias in the estimated portfolio beta increases as a function of survivorship bias.
Table IV
Survivorship and the Performance of Relative Strength Portfolios
(36-month Formation Period)

72-months of stock return data for 1000 stocks is simulated according to the market model in equation (5). For each market replication, stocks are ranked based on their past 36-month cumulative total return. An equally-weighted long-short portfolio of the best and worst performing 10% of the stocks is formed at month 36 and held for the remaining 36 months. We present the 2.5, 50, and 97.5%iles of the sampling distribution of the k-month cumulative residual of the long-short portfolio for various performance cuts based on 20,000 replications.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>k = 3</th>
<th>k = 6</th>
<th>k = 9</th>
<th>k = 12</th>
<th>k = 18</th>
<th>k = 24</th>
<th>k = 30</th>
<th>k = 36</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5%ile</td>
<td>-0.184</td>
<td>-0.253</td>
<td>-0.298</td>
<td>-0.331</td>
<td>-0.408</td>
<td>-0.463</td>
<td>-0.513</td>
<td>-0.572</td>
</tr>
<tr>
<td>50%ile</td>
<td>0.008</td>
<td>0.015</td>
<td>0.022</td>
<td>0.031</td>
<td>0.047</td>
<td>0.061</td>
<td>0.078</td>
<td>0.095</td>
</tr>
<tr>
<td>97.5%ile</td>
<td>0.233</td>
<td>0.341</td>
<td>0.429</td>
<td>0.525</td>
<td>0.667</td>
<td>0.805</td>
<td>0.917</td>
<td>1.052</td>
</tr>
<tr>
<td>Panel B: 5% Performance Cut</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5%ile</td>
<td>-0.146</td>
<td>-0.181</td>
<td>-0.191</td>
<td>-0.201</td>
<td>-0.237</td>
<td>-0.243</td>
<td>-0.264</td>
<td>-0.275</td>
</tr>
<tr>
<td>50%ile</td>
<td>0.017</td>
<td>0.034</td>
<td>0.051</td>
<td>0.071</td>
<td>0.107</td>
<td>0.143</td>
<td>0.178</td>
<td>0.217</td>
</tr>
<tr>
<td>97.5%ile</td>
<td>0.213</td>
<td>0.309</td>
<td>0.381</td>
<td>0.439</td>
<td>0.604</td>
<td>0.711</td>
<td>0.847</td>
<td>0.968</td>
</tr>
<tr>
<td>Panel C: 10% Performance Cut</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5%ile</td>
<td>-0.143</td>
<td>-0.168</td>
<td>-0.172</td>
<td>-0.160</td>
<td>-0.200</td>
<td>-0.200</td>
<td>-0.218</td>
<td>-0.214</td>
</tr>
<tr>
<td>50%ile</td>
<td>0.022</td>
<td>0.042</td>
<td>0.064</td>
<td>0.088</td>
<td>0.129</td>
<td>0.176</td>
<td>0.217</td>
<td>0.266</td>
</tr>
<tr>
<td>97.5%ile</td>
<td>0.213</td>
<td>0.307</td>
<td>0.380</td>
<td>0.430</td>
<td>0.601</td>
<td>0.712</td>
<td>0.862</td>
<td>0.973</td>
</tr>
<tr>
<td>Panel D: 20% Performance Cut</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5%ile</td>
<td>-0.165</td>
<td>-0.190</td>
<td>-0.183</td>
<td>-0.163</td>
<td>-0.204</td>
<td>-0.183</td>
<td>-0.208</td>
<td>-0.203</td>
</tr>
<tr>
<td>50%ile</td>
<td>0.027</td>
<td>0.053</td>
<td>0.080</td>
<td>0.111</td>
<td>0.165</td>
<td>0.222</td>
<td>0.275</td>
<td>0.334</td>
</tr>
<tr>
<td>97.5%ile</td>
<td>0.243</td>
<td>0.340</td>
<td>0.420</td>
<td>0.468</td>
<td>0.655</td>
<td>0.779</td>
<td>0.954</td>
<td>1.072</td>
</tr>
</tbody>
</table>
Table V
True and Estimated Beta of Relative Strength Portfolios
(36-month Formation Period)

72-months of stock return data for 1000 stocks is simulated according to the market model in equation (5). For each market replication, stocks are ranked based on their past 36-month cumulative total return. An equally-weighted long-short portfolio of the best and worst performing 10% of the stocks is formed at month 36 and held for the remaining 36 months. We present the 2.5, 50, and 97.5%iles of the sampling distribution of the true and estimated portfolio betas for various performance cuts based on 20,000 replications.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>No Cut</th>
<th>5% Cut</th>
<th>10% Cut</th>
<th>20% Cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5%ile</td>
<td>-1.001</td>
<td>-1.008</td>
<td>-0.545</td>
<td>-0.558</td>
</tr>
<tr>
<td>50%ile</td>
<td>0.594</td>
<td>0.597</td>
<td>0.669</td>
<td>0.664</td>
</tr>
<tr>
<td>97.5%ile</td>
<td>1.777</td>
<td>1.781</td>
<td>1.442</td>
<td>1.452</td>
</tr>
</tbody>
</table>

1.4 Conclusion

Our findings are speculative. They can not be interpreted unless we decide how much we believe that our simulations imitate the real world. In this section, we take the most optimistic view and discuss the implications of our results for studies of the risk/return tradeoff and return predictability. In so doing, we are free to suggest new directions for further research while allowing the reader to assess the expected utility under his own beliefs.

1.4.1 Means, Standard Deviations, and Betas

Survivorship bias influences statistical inference in Finance, especially when first, second, and cross-moments are the objects of inference. Section 3.1 demonstrates that survivorship bias has a larger effect on stock studies than on mutual fund studies because stock returns are more variable than mutual fund returns. Since we restrict ourselves to a very special model of survivorship - one based on removing poor performers only, we can not say if the result holds in the real world as well. An analysis of the actual stock market returns of survivors along the lines of Malkiel (1995) would be useful in this regard.
1.4.2 Risk/Return Tradeoff

The Sharpe-Lintner-Black Captial Asset Pricing Model predicts that expected returns on stocks are a linear function of their market betas and that market betas suffice to describe the cross-section of expected returns. Believers in the CAPM usually take the view of Roll & Ross (1994) or Kandel & Stambaugh (1995) when confronted by Fama & French (1992). They argue that there are (slightly) non-efficient market proxies that produce betas which are uncorrelated with expected returns. When the market portfolio in the test is misspecified in this manner, variables other than the misspecified betas may have explanatory power.

Fama & French (1996) facetiously chide the believers as “awaiting the coming of M,” the true market portfolio. M’s arrival will yield market betas that are linearly related to expected returns and that can explain returns on their own (if the CAPM is true).

Our results suggest another possible (partial) justification for the believers. “M” is already here, but we are distracted by other m’s and can’t see it. More formally, section 3.2 suggests that the CAPM prediction that expected returns on stocks are a linear function of their market betas is not identical to the statistical statement that cross-sectional regressions of returns on beta yield t-statistics greater than 2.

However, the finding if true does not by itself save the CAPM. As Fama & French (1996) state “there is now compelling evidence that β does not suffice to explain expected returns.” On this point, we don’t have anything new to say. But, we can point to work that makes the following point: the CAPM prediction that market betas suffice to describe the cross-section of expected returns is not identical to the statistical statement that no non-beta independent variable in the world yields a t-statistic greater than 2 in cross-sectional regressions. In fact, Foster, Smith & Whaley (1997) and Kan & Zhang (1999) show that if we allow the number of potential non-beta regressors to grow without bound, our chances of finding three that explain returns is doomed to grow to 100% even if the CAPM is true. Implicit in their work, of course, is a call for more (theory-based) discipline to constrain the set of possible non-beta regressors.

Finally, there may be an upper-bound on the value of research that explains the explanatory power of non-beta regressors by appealing to their correlation with beta. Section 3.2
suggests that even non-beta regressors that are perfectly correlated with beta have a good chance of explaining nothing.

1.4.3 Return Predictability

According to Fama (1998) and Fama & French (1996) short-term continuation of returns documented by Jegadeesh & Titman (1993) is an "open puzzle" in the field of Finance. Our analysis suggests we should reconsider explanations based on the cross-sectional dispersion in expected returns. Unless we choose to believe that the "true" capital asset pricing model implies that all stocks have identical expected returns, equation (8) demonstrates that momentum strategies yield positive returns in the absence of large amounts of negative serial-and/or positive cross-correlation. Among the questions that remain for those adopting this view is the following: How does $\hat{P}$ behave as a function of time?

Section 3.3 also suggests that survivorship bias can make momentum returns appear higher and momentum risk appear lower than it is in reality. If true, future researchers who identify momentum strategies yielding high returns with low risk should be particularly careful to rule out survivorship bias as a partial explanation.
Chapter 2

A Role for the Theory of Tournaments in Studies of Mutual Fund Behavior

Abstract

We develop a two-period model of a Brown, Harlow & Starks (1996) mutual fund tournament in which two fund managers with unequal midyear performances compete for new cash inflows. When one of the managers is an exogenous benchmark, a winning manager indexes and a losing manager gambles. However, when both managers are active, a winning manager is more likely to gamble - especially when the midyear performance gap is high or when stocks offer high returns and low volatility. Empirical evidence that winning managers gamble has been documented in Chevalier & Ellison (1997), who find a positive correlation between past performance and increases in unsystematic risk, and Busse (1998), who finds that likely winners increase total risk more than likely losers. With a dataset of weekly returns from 1984-1996 for 660 mutual funds, we test for the other managerial gaming effects predicted by the model. Our results suggest a role for the theory of tournaments in studies of mutual fund behavior.

2.1 Introduction

Every mutual fund prospectus contains some form of the following statement: "Returns are based on past results and are not an indication of future performance."

Whether the statement is true seems to be an open question. For instance Hendricks, Patel & Zeckhauser (1993) find that an active portfolio of top-octile funds (on the basis of their relative returns in the most recent year) significantly out-performs a passive index of
fellow mutual funds. However, Brown & Goetzmann (1995) find that while in most years winners repeat, occasionally the effect is dramatically reversed, and Malkiel (1995) finds that a strategy of buying past winners would not have beat the market during the 1980’s.

Similarly, there are many different explanations for the positive returns to chasing winners. Carhart (1997) suggests that mutual fund performance persistence disappears in a market where expected returns are characterized by a Fama-French (1996) three-factor model plus an additional momentum factor. Brown, Goetzmann, Ibbotson & Ross (1992) suggest that a small amount of survivor bias is sufficient to create the appearance of persistence in mutual fund returns. Grinblatt, Titman & Wermers (1995) suggest that winning fund managers actively exploit the Jegadeesh & Titman (1993) momentum anomaly.

As academics ponder whether past performance indicates future results, fund investors act as if they do. Sirri and Tufano (1998) examine the performance-flow relationship with a dataset of 690 equity funds during 1971-1990. In each year, the authors divide the sample into 20 performance bins and calculate the average growth rate in net new money to the funds in the bin. They find that “exceptionally high performing funds reap large rewards, but poorer performance is generally not penalized.” Similar highly asymmetric “winner-take-all” patterns are found in Ippolito (1992) and Goetzmann and Peles (1997).

The performance-flow relationship is similar when performance is measured in other ways. However, Patel, Zeckhauser, and Hendricks (1994) show that the “inclusion of ranks drives out the statistical significance of other performance measures.” Verifying this anecdotally is Capon, Fitzsimons, and Prince (1996), who survey 3,386 fund-owning households. They find that these households consider published performance rankings to be the most important source of information about the funds they chose.1.

The question of this paper is: Does the “winner-take-all” characteristic of fund inflows have any effect on fund managers? Our approach is inspired by the work of Brown, Harlow & Starks (1996; hereafter referred to as BHS), Chevalier & Ellison (1997; hereafter referred to as CE), and Busse (1998), who study the risk-taking behavior of fund managers at midyear

---

1 Other sources of information included advertising, commission-based financial advisors, seminars, recommendations from friends/family, recommendations from business associates, fee-based financial advisors, books and direct mail.
as they compete for new cash inflows.

Important to the BHS, CE and Busse studies is the notion that mutual fund management companies act to maximize total assets under management, since they are typically paid at a rate based on a percentage of those assets. This is a view prevalent in the industry as well, as evidenced by the following quote from the Goldman Sachs Investment Management Industry Group in 1995.

Managing money is not the true business of the money management industry. Rather, it is gathering and retaining assets. Good money management skills are only one of many important tools essential to the business of attracting and retaining assets.

All three authors test whether managers manipulate risk in response to their midyear performance to increase the odds of bringing large cash inflows to the fund. BHS find that managers whose midyear performance ranks below the median increase total risk more than managers whose performance ranks above. They take this finding as support for the view that the mutual fund industry is similar to a tournament, with past losers behaving differently from past winners. Busse repeats the BHS analysis with higher frequency data but finds the opposite result, a finding he takes as refuting the tournament hypothesis. CE estimate the relationship between flows and past performance in excess of the market. Using the relationship they identify, they predict that extreme winners and moderate losers increase residual risk more than extreme losers and moderate winners. But, when they examine time series data, they find a positive correlation between excess returns and residual risk increases.

We build on these studies by developing a two-period model of a mutual fund industry in which two fund managers with unequal midyear performances compete for new cash inflows at the end of the year. We attempt to reconcile past empirical contradictions in the literature by considering explicitly the difference between competing against an exogenous benchmark

\footnote{We do not consider the effects of incentive compensation plans which provide increased compensation or reduced payments depending on the relationship between the fund’s performance and that of a comparable market index. These plans are studied in Grinblatt & Titman (1989), Starks (1987) and Modigliani & Pogue (1975). According to Golec (1992), Heinkel & Stoughton (1994), and Das & Sundaram (1998), they are relatively rare.}
such as the S&P 500 and competing in a tournament where the benchmark is endogenous (e.g. a "Top 10" list).

We find that when one of the managers is an exogenous benchmark, the model predicts that winning managers index to lock in their lead and losers gamble to catch up. This is basically the prediction tested by BHS and Busse.

However, in the endogenous case where both managers are active, a winning manager is more likely to gamble to maintain his lead. As in the exogenous case, the winner indexes, but now with respect to his expectation of what the loser will do. Since he expects the loser to gamble, especially when far behind or when stocks offer high returns and low volatility, the winner gambles with high probability. The loser anticipates this reasoning and chooses the opposite strategy with high probability in equilibrium.

Empirical evidence that winning managers gamble is documented in CE (1997), Busse (1998), Carhart (1997), Koski & Pontiff (1999), and Chen & Pennacchi (1999). So, with a dataset of weekly returns from 1984-1996 for 660 mutual funds, we test for the other managerial gaming effects predicted by the model. For the exogenous case, we test whether winning funds index. For the endogenous case, we test for a negative relationship between market volatility and increased risk-taking by winning mutual funds. Our results provide more support for the predictions of the endogenous case, suggesting a role for the theory of tournaments in studies of mutual fund behavior.

The structure of the paper is as follows: In the next section, we present the model. In section 3.0, we present empirical results. Section 4.0 concludes.

2.2 Model

In this section, we first develop a model for studying the actions of the managers as they compete for new fund inflows. We then present solutions to the model in relevant specific cases.
2.2.1 Structure

Like BHS (1996), CE (1997) and Busse (1998), we are interested in the problem faced by competing portfolio managers in a tournament in the middle of the year.

Assumption 1 There are two periods, the midyear and the end of the year.

The two-period structure and the choice of the midyear as the single epoch of strategy change is based on the constructions of previous researchers. For instance, CE (1997) consider a situation where managers have the opportunity to change their portfolio strategy at the beginning of the 4th quarter of every year. BHS (1996) examine five potential epochs of strategy change – the end of April, May, June, July, and August – and find that the data support the notion that “managers revise their investment strategies within the month following the release of second quarter performance rankings.”

At midyear, the managers observe their year-to-date performances and simultaneously choose a portfolio to hold for the rest of the year which maximizes expected management fees for their management company. The midyear performances are exogenous to our model, but can be thought of as an implication of Palomino (1998). Palomino studies the equilibrium trading strategies of risk-neutral relative performance maximizing managers at the beginning of the year. He finds that when the managers are heterogeneously informed, there is a unique symmetric equilibrium where each manager plays a strategy which is linear in his signal. As we shall see, it is better for a manager to arrive at the midyear evaluation with high year-to-date performance relative to his competitors. Therefore, we can endogenize the exogenous midyear performances by considering our model - and the analysis of BHS (1996), CE (1997) and Busse (1998) - as the second stage of the situation considered by Palomino (1998).

Managers are restricted to a single portfolio revision at midyear which they must live with for the rest of the year. In actuality of course, opportunities to revise portfolio holdings arrive continuously. Our model can be thought of as a single stage in a repeated game that begins once once there are differences in the year-to-date performance of the managers.

Assumption 2 There are two managers, a winner and a loser.

The two manager setting is also a construction of previous researchers. BHS (1996) and Busse (1998) consider the average behavior of “winners” and “losers” - defined by the
Authors as those funds with above median or below median performance, respectively. Lazear & Rosen (1981) develop the central intuition of tournament models in a two-person setting. Our two-manager model is the simplest non-trivial setting in which to study the effects of competition for new fund inflows.

We consider two different games: In the first game, one manager is an active manager and the other is an exogenous market index. In the second game, both managers are active managers. The winner is the manager who has achieved the highest midyear return \( m \).

**Assumption 3** Manager \( i \) chooses a combination of risky and riskless asset for the second half of the year. The proportion of the portfolio in the risky asset is \( \alpha_i \in [0, 1] \).

The restriction of portfolio choice to the interval [0,1] is not important, but it is important that the managers' choices be constrained to a bounded interval. The assumption mimics the limits on fund borrowing and short selling in a mutual fund prospectus and it rules out uninteresting corner solutions.

The restriction to a single risky and riskless asset also creates the simplest possible setting in which to study the risk-taking behavior of the managers. As we shall see, the implications of the model, though literally applicable to market timers, also apply to stock pickers.

**Assumption 4** The risky asset pays a random return \( x \sim N(\mu, \sigma^2) \) and the riskless asset pays \( r \) for certain. The return distributions of the assets are known to each manager.

We assume that all information about the economy becomes public knowledge at midyear. By doing so, we abstract from a discussion of the equilibrium implications of delegated portfolio management (see Brennan, 1993 or Kyle, 1985) and from the strategic implications of possessing superior knowledge (see Huberman & Kandel, 1993).

Our view is that these effects are likely to be small in comparison with the incentive effects of a manager's own compensation plan. First, the empirical evidence for the existence of superior fund managers in the real world is relatively weak. Second, the incentives that such managers have in a more general model may be offsetting. For instance in Kyle (1985), an informed manager restrains his trading in the risky asset to minimize the price impact. In contrast, an informed manager in Huberman & Kandel (1993) exaggerates his trading in the risky asset to signal skill to investors who evaluate him based on realized returns.
**Assumption 5** Manager $i$ is risk neutral and receives compensation at the end of the year in the form a management fee, which is a fixed percentage ($k$) of his total assets under management.

Assuming that the initial size of manager $i$’s fund is $s_i$, his compensation is given by:

$$C_i[\alpha_i, \alpha_j] = ks_i \{1 + m_i + \alpha_i x + (1 - \alpha_i)r + W\Pi_i[\alpha_i, \alpha_j]\}$$  \hspace{1cm} (2.1)

where $\Pi_i[\alpha_i, \alpha_j] = 1$ iff $m_i + \alpha_i x + (1 - \alpha_i)r > m_j + \alpha_j x + (1 - \alpha_j)r$

$$\Pi_i[\alpha_i, \alpha_j] = 0 \text{ otherwise}$$

In equation (1), we see that manager $i$ can increase his total assets under management by generating internal growth through his second-half portfolio choice ($\alpha_i$) or by receiving fund inflows as result of out-performing the index or his peer, manager $j$.

Since each manager is risk-neutral, each chooses a portfolio to maximize his expected compensation. The compensation plan takes the form of a flat-fee arrangement. Golec (1992), Stoughton & Heinkel (1994) and Das & Sundaram (1998) find that this is the most common way a mutual fund board pays the management company. Though Bhattacharya & Pfleiderer (1985), Dybvig, Farnsworth & Carpenter (1999), and Stoughton (1993) characterize an optimal contract for portfolio management that solves the problems of adverse selection and moral hazard, we do not have any evidence that their specifications are being used.\(^3\) Note that the choice of this compensation plan assumes that the portfolio manager has the same interests as his management company.

The compensation of each manager is influenced therefore, not only by the outcome of his portfolio choice but by the behavior of investors through the winner-take-all performance-flow relationship ($W\Pi$) described in the introduction. In games of the first type, out-performance of a market index yields a large and positive amount of new net inflow to the fund, i.e. a “winner’s prize.” In games of the second type, the winner’s prize goes to the active manager

\(^3\)Recently, the SEC has considered requiring mutual funds to disclose manager pay. See Skakun (1996) and Institutional Investor (1995) for a discussion.
with the highest return. Games of the second type are clearly related to the tournaments of Lazear & Rosen (1981), Green & Stokey (1983), Nalebuff & Stiglitz (1983), Rosen (1986), Bronars (1987), and McLaughlin (1987). In those models and in ours, the compensation of each agent is a function of the rank-order of his output in a tournament of similar agents.

Within the framework above, it is possible to study the optimal actions of a fund manager in each type of game. In the following section we do so, beginning with games of the first type.

### 2.2.2 Analysis of Type I Games

In the Type I game, an active manager competes against a publicly observable index. Since manager $i$’s compensation is effected by the portfolio return of the index, we are interested in the properties of the following function:

$$
\Re(\alpha_j) = \arg \max_{\alpha_i} E \{ C_i(\alpha_i, \alpha_j) \} 
$$

(2.2)

Equation (2) gives the “best response” portfolio choice of manager $i$ to the index portfolio $\alpha_j$. Under the assumed distribution for $x$, equation (2) has the same solution as the following problem:

$$
\Re(\alpha_j) = \arg \max_{\alpha_i} \left\{ ks_i(1 + m_i + \alpha_i \mu + (1 - \alpha_i)r + W\Pi_i[\alpha_i, \alpha_j]) \right\} 
$$

(2.3)

$$
\Pi_i = \begin{cases} 
1 - N \left( \frac{T - \mu}{\sigma} \right) & \text{if } \alpha_i > \alpha_j \\
N \left( \frac{T - \mu}{\sigma} \right) & \text{if } \alpha_i < \alpha_j \\
1 & \text{if } \alpha_i = \alpha_j \text{ and } m_i > m_j \\
0 & \text{otherwise}
\end{cases}
$$

(2.4)

$$
T = \frac{m_j - m_i}{\alpha_i - \alpha_j} + r.
$$

(2.5)

The threshold $T$ gives the return that must be realized on the risky asset in order for manager $i$’s total return for the year, $\alpha_i x + (1 - \alpha_i)r + m_i$, to exceed the total return of the index. $\Pi_i$ is the probability that manager wins the game, given manager $j$’s portfolio choice.

---

4In what follows, $N(z)$ is the standard normal cumulative distribution function and $n(z)$ is the associated standard normal density.
Before we discuss the solution to (3), there are insights to be gained by considering the game variable \( \Pi \) separately, in the absence of borrowing constraints. By inspection of equations 4-5, we have the following proposition:

**Proposition 1** For a winning manager, indexing yields the highest probability of winning the game. For a losing manager, gambling yields the highest probability of winning.

By choosing the index, the winner assures himself of winning the contest regardless of the return on the risky asset. A strategy of mimicry is suicidal for a losing manager, however. So, equation (4) leads him to choose either \( \alpha_i \to \infty \) or \( \alpha_i \to -\infty \). In the presence of a positive risk premium, \( \alpha_i \to \infty \) is optimal.

Proposition 1 captures a common intuition about how winning and losing managers are likely to behave in this type of game. However, it is not a general solution to (3). Though a closed-form solution to (3) can not be obtained, plots of the best response function for winning and losing managers are included in Figures I and II for a set of realistic parameter values.

We use \( e \equiv \mu - r = .043 \) and \( \sigma = .141 \) as the 6-month risk premium and standard deviation of the risky asset, based on historical Ibbotson Associates estimates. In the top panel of each figure, the performance gap \( g \equiv m_i - m_j \) is the average absolute difference between the median midyear performance of the top 10% and bottom 90% of the funds in the sample. In the second panel of Figure I (Figure II), \( g \) is the average absolute difference between the median midyear performance of funds that out-performed (under-performed) the index and the midyear performance of the index. The growth rate in net new money to the ultimate winner fund \( (W = .35) \) is based on Sirri & Tufano (1998). They found that the average rate of cash inflow to funds that rank in the top 10% is about 45%. All other funds receive average fund inflows on the order of 10%.

Figures I and II demonstrate that the main driving force in the solution is a tradeoff between the value of the risk premium and the value of winning the game. For instance, the winner never indexes in the truest sense of the word (unless the index is \( \alpha_j = 1 \)) because he is always willing to trade a small probability of losing for additional risk premium. Figure I
also shows that the winner is less willing to trade probability for premium when he is close to his competitor at midyear.

Similarly, the loser doesn't always gamble because, if the index is "risky," he is willing to trade premium for a larger chance of winning. Figure II also shows that the loser is less willing to gamble when he is close to his competitor at midyear.
Parameter values were $e = .043$, $g = .106$, $\sigma = .141$, and $W = .35$. 
Parameter values were $e = .043$, $g = .106$, $\sigma = .141$, and $W = .35$.

Parameter values were $e = .043$, $g = .030$, $\sigma = .141$, and $W = .35$.

2.2.3 Analysis of Type II Games

In the Type II game, an active manager competes against another active manager whose portfolio choices are not observable. This is a standard two-person non-cooperative simultaneous-move game. To find closed form Nash equilibria in the game, we observe that in Figure II, the losing manager, broadly speaking, chooses one of two types of portfolios - "risky" or
“safe”. We see from Figure I that the winning manager’s best response to these choices by the loser is “risky” and “safe”, respectively. It would appear, therefore, that we can capture much of the essential nature of the tournament by considering two possible strategies for each manager, \( \alpha = 0 \) and \( \alpha = 1 \).\(^5\)

When we solve for the equilibrium in the simpler two-person two-strategy game, we obtain the following proposition:

**Proposition 2** Winning managers are more likely to choose a risky portfolio than losing managers. A large midyear performance gap \( (m_{\text{winner}} - m_{\text{loser}}) \) and a high risk premium \( (\mu - r) \) increases this difference in risk-taking likelihoods, while a large volatility \( (\sigma) \) decreases it.

In the simplified Type II game, the managers face the following payoff matrix:

<table>
<thead>
<tr>
<th>(W, L)</th>
<th>Riskless</th>
<th>Risky</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riskless</td>
<td>( ks(R_w + W), ksR_l )</td>
<td>( ks(R_w + WN(\frac{R_w - R_l}{\sigma})), ks(R_l + e + W \left[1 - N(\frac{R_w - R_l}{\sigma})\right]) )</td>
</tr>
<tr>
<td>Risky</td>
<td>( ks(R_w + e + W \left[1 - N(\frac{R_w - R_l}{\sigma})\right]), ks(R_l + WN(\frac{R_w - R_l}{\sigma})) )</td>
<td>( ks(R_w + e + W), ks(R_l + e) )</td>
</tr>
</tbody>
</table>

where the following substitutions have been made for simplicity:

\[
\begin{align*}
    s_w &= s_l = s \\
    m_{\text{winner}} - m_{\text{loser}} &= g \quad \text{(performance gap)} \\
    \mu - r &= e \quad \text{(risk premium)} \\
    1 + m_i + r &= R_i
\end{align*}
\]

Examination of the payoff matrix reveals that there are two important cases:

- \( e > WN(\frac{g - e}{\sigma}) \)

  This case corresponds to the situation in which more is gained, in expectation, from holding the risky portfolio than from winning the contest. If this is the case, holding the risky portfolio is a strictly dominant strategy for each manager.

- \( e \leq WN(\frac{g - e}{\sigma}) \)

  This case corresponds to the situation in which both manager’s decisions are driven by the desire to win the contest. If this is the case, there is a unique mixed-strategy equilibrium given by the expressions below.

---

\(^5\)In computer simulations, the results are unchanged when the choice set is not restricted in this way.
\[ p = \frac{\frac{e}{W} + 1 - N\left(\frac{g-e}{\sigma}\right)}{1 - N\left(\frac{g-e}{\sigma}\right) + N\left(-\frac{g-e}{\sigma}\right)} \] (2.6)

\[ q = \frac{-\frac{e}{W} + N\left(-\frac{g-e}{\sigma}\right)}{1 - N\left(\frac{g-e}{\sigma}\right) + N\left(-\frac{g-e}{\sigma}\right)} \] (2.7)

Here \( p (q) \) is the probability that a winning (losing) manager chooses the risky portfolio.\(^6\) Plots of \( p \) (the top line) and \( q \) as a function of the risk premium \( e \), performance gap \( g \), and volatility \( \sigma \) are shown in Figure III for a set of realistic parameters. Note that the difference between the fraction of winning and losing managers that choose a risky portfolio is positive and that it increases with \( g \) and \( e \) and decreases with \( \sigma \).

It is also important to note the range of parameter values for which the mixed strategy outcome is irrelevant i.e. when \( e > WN\left(-\frac{g-e}{\sigma}\right) \). For instance, when the performance gap between winners and losers is extremely high, it is not worthwhile for the managers to pay attention to the contest. Similarly, when the volatility of the risky asset is extremely low, any performance gap is, in a sense, “locked in” and the managers ignore the contest. A large risk premium makes it worthwhile to hold the risky asset at all times, regardless of the ramifications of this strategy in the contest. Finally, a contest with a low winner’s prize is not worth entering.

Though somewhat puzzling at first, the intuition for why winners are more likely to choose risky portfolios comes from the analysis of Type I games. There we see that the winner maintains his lead by mimicking his opponent. When his opponent’s strategy is unknown, however, the winner develops a belief about what it is likely to be. Since he expects his losing opponent to gamble, especially when far behind or when stocks offer high returns and low volatility, the winner gambles with high probability. In equilibrium, the loser correctly anticipates this reasoning and chooses the opposite strategy with high probability.

Propositions 1 and 2 suggest two useful ways to test for managerial gaming effects in the

---

\(^6\)Later, we interpret \( p \) as the fraction of winning managers who choose a risky portfolio (in the cross section).
data in addition to those already examined in the literature. In the following section, we construct statistical tests for these effects.

Figure III - Mixed Strategy Equilibria as a Function of Various Parameters
Parameter values were $e = .043$, $g = .106$, $\sigma = .141$, and $W = .35$.

2.3 Empirical Analysis

Before discussing the structure of the tests, it is useful to discuss the characteristics of the dataset.

2.3.1 Data Description

The dataset includes weekly returns on mutual funds classified as "growth" by Lipper Analytical Services, Inc. By restricting our sample to growth funds, we hope to control for differences in investment objectives and other institutional characteristics that may affect risk-taking behavior. As a result, it becomes more reasonable to attribute cross-sectional differences in fund risk-taking behavior to the incentives we are studying. Also, the emphasis on growth funds makes it reasonable to compare the performance of these funds to all-equity benchmarks such as the S&P 500 or the CRSP value-weighted index of NYSE/AMEX/NASDAQ stocks.

The weekly return frequency is a valuable feature of the dataset because test statistics of
interest include estimates of beta over periods as short as 6 months.\textsuperscript{7} The returns we study are net of management fees. They are calculated under the assumption that all dividends and capital gains distributions are reinvested on the ex-dividend date at the ex-dividend net asset value.

A fund is included in the analysis beginning in the first year that it reports returns for the entire year. For 1996, the last full year of data, there were 660 such funds. Appendix B contains a full listing of these funds.

One concern with the data is that it is survivorship biased. Once a fund merges with another or is liquidated, Lipper removes its return history from the file. There is a large and active literature on survivorship-induced patterns in mutual fund returns inspired by Brown, Goetzmann, Ibbotson and Ross (1992), Malkiel (1995), and Carpenter & Lynch (1999). However, since our central predictions are confirmed in datasets largely free of survivorship bias, we do not address this issue here.

In Table I, we have reported for each year, the number of funds that reported weekly returns for the entire year, the average annual total return of these funds, the annual total return on the S&P 500 index, the fraction of funds out-performed by the S&P 500 index, and the median midyear performances of winning funds (top 10%) and losing funds (bottom 90%). This partition of the data is not special, though it is in accordance with the findings of Sirri & Tufano (1998). Note that in 1984 it was possible to under-perform the S&P 500 and still “win” while in 1992, over 170 funds beat the S&P 500 and “lost,” according to the Sirri & Tufano (1998) results.

\textsuperscript{7}The CRSP Survivor Bias Free Mutual Fund Database provides returns at a monthly frequency.
Table I  
Descriptive Statistics

In each tournament year, the number of funds reporting returns for the full year is listed. The average full-year total return for all funds is presented, along with the full-year total return on the S&P 500 index. For comparison, we have included the fraction of funds out-performed by the S&P 500 in each year. For winners, defined here as those funds whose midyear total return ranks in the top 10% of all funds, the median return is presented. The same information is also presented for losers, defined here as all funds not in the top 10%.

<table>
<thead>
<tr>
<th>Year</th>
<th>Num.</th>
<th>Avg</th>
<th>S&amp;P</th>
<th>S&amp;P%ile</th>
<th>(m_w)</th>
<th>(m_l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>114</td>
<td>-4.4</td>
<td>6.3</td>
<td>91</td>
<td>-5.04</td>
<td>-14.8</td>
</tr>
<tr>
<td>1985</td>
<td>126</td>
<td>27.6</td>
<td>31.8</td>
<td>74</td>
<td>25.8</td>
<td>17.7</td>
</tr>
<tr>
<td>1986</td>
<td>137</td>
<td>17.3</td>
<td>18.7</td>
<td>59</td>
<td>25.5</td>
<td>13.7</td>
</tr>
<tr>
<td>1987</td>
<td>152</td>
<td>2.9</td>
<td>5.3</td>
<td>66</td>
<td>38.3</td>
<td>28.1</td>
</tr>
<tr>
<td>1988</td>
<td>180</td>
<td>14.9</td>
<td>16.6</td>
<td>58</td>
<td>22.9</td>
<td>9.5</td>
</tr>
<tr>
<td>1989</td>
<td>189</td>
<td>26.6</td>
<td>31.7</td>
<td>76</td>
<td>34.3</td>
<td>22.6</td>
</tr>
<tr>
<td>1990</td>
<td>201</td>
<td>-4.3</td>
<td>3.1</td>
<td>54</td>
<td>13.4</td>
<td>4.4</td>
</tr>
<tr>
<td>1991</td>
<td>222</td>
<td>33.4</td>
<td>30.5</td>
<td>44</td>
<td>35.1</td>
<td>19.4</td>
</tr>
<tr>
<td>1992</td>
<td>254</td>
<td>12.4</td>
<td>7.6</td>
<td>22</td>
<td>13.0</td>
<td>2.3</td>
</tr>
<tr>
<td>1993</td>
<td>319</td>
<td>11.0</td>
<td>10.1</td>
<td>48</td>
<td>14.4</td>
<td>2.3</td>
</tr>
<tr>
<td>1994</td>
<td>413</td>
<td>-1.7</td>
<td>1.3</td>
<td>77</td>
<td>2.5</td>
<td>-4.3</td>
</tr>
<tr>
<td>1995</td>
<td>532</td>
<td>30.3</td>
<td>37.6</td>
<td>88</td>
<td>36.5</td>
<td>23.6</td>
</tr>
<tr>
<td>1996</td>
<td>660</td>
<td>21.0</td>
<td>23.0</td>
<td>63</td>
<td>8.3</td>
<td>2.5</td>
</tr>
<tr>
<td>Avg</td>
<td>269</td>
<td>14.4</td>
<td>16.7</td>
<td>63</td>
<td>20.4</td>
<td>10.6</td>
</tr>
</tbody>
</table>

2.3.2 Methodology

Since the propositions are developed in a setting with one winning and one losing manager, we take them to describe general tendencies of winning and losing managers as groups. The statistics we study, therefore, are based on the average behavior of a group of funds instead of the behavior of individual funds.

Throughout, we assume that fund returns are generated by the following market model:

\[
x_{it} = r_f + \beta_i (r_{mt} - r_f) + \epsilon_{it} \quad \text{with}
\]

\[
\sigma^2_{x_i} = \beta^2_i \sigma^2_{r_m} + \sigma^2_{\epsilon_i}.
\]

In the Type I case, the propensity of winners to index in the second-half of the year can be tested directly by testing whether regressions of second-half returns on the index yield higher \(R^2\)'s for winners than losers. Winners are those funds whose first-half total return is higher
than the first-half total return of the CRSP value-weighted index of NYSE/AMEX/NASDAQ stocks.

Note that a strict interpretation of the model implies that both loser and winner $R^2$'s are equal to 1, since there is a single risky asset. However, the essence of the tournament prediction is that the winner maximizes his chance of winning by holding the same portfolio as the index and the loser maximizes his chance of winning by holding a portfolio that is “different” from the index. In more realistic settings where there are multiple risky assets, the loser portfolio can differ from the index via stock selection as well as market timing. A test based on $R^2$ captures this realism and is in the spirit of the model.

The Type II case implies that the difference in risk-taking likelihoods between winners and losers is positive and is positively related to the first-half market risk premium and midyear performance gap, and negatively related to the first-half market standard deviation. Here, winners are those funds whose first-half total return is in the top 10\%ile of all funds in the sample that year.

There are a number of issues that make it difficult to test these predictions directly. To see why, define $\Delta$ as the difference between the fraction of winning funds ($f_w$) and the fraction of losing funds ($f_l$) with betas greater than the median beta of all funds in the second half. By comparing the betas of the funds to the median beta, we are able to test the central intuition of the Type II case: winners expect the losers to gamble and so they choose higher beta portfolios to maintain their lead. By using $f_w$ and $f_l$ for $p$ and $q$, we are operationalizing the mixed strategy equilibrium of Proposition 2.

The difficulty arises when we consider the properties of $\Delta$ under the reasonable null hypothesis that the managers choose a portfolio at the beginning of the year and hold it throughout the entire year. If the market goes up in the first half, high beta funds are likely to be the winners and low beta funds are likely to be the losers. Since the managers don’t change their portfolios, a higher fraction of winner funds will have betas greater than the median beta of all funds in the second-half i.e. $\Delta$ is more likely to be positive than negative.\footnote{During 1984-1996, the market was up in the first-half every year except 1984 and 1994. $\Delta$ was significantly positive in 8 of the 13 years.} As for the regressions, an estimate of the first-half market risk premium or the
midyear performance gap is essentially a proxy for how well the market did in the first half. Both estimates are positively related to $\Delta$ in the absence of any managerial gaming effects.

However, the relationship between $\Delta$ and the market standard deviation should be positive under the null in contrast with the negative relationship predicted by the model. If the market is less volatile, a higher proportion of each fund’s total risk is accounted for by residual risk, according to the market model. Therefore, high non-systematic returns can make low beta stocks winners and can decrease $\Delta$ in the process. $\Delta$ is positively related to the market standard deviation in the absence of any managerial gaming effects.

2.3.3 Results

In Table II, we report for each year the outcome of a Wilcoxon Rank Sum test on the $R^2$'s from a regression of winner and loser funds on the market. The distribution of the test statistic - a standardized sum of the winner ranks - is approximately normal when the number of funds grows large. High values of the statistic imply that the $R^2$'s for winner funds tend to have higher ranks than those of loser funds. In 11 of the 13 tournament years studied however, the value of this statistic was either significantly negative or insignificant. There is statistically significant evidence that winners indexed in the second-half of 1985 and 1990, but there are three years - 1988, 1991 and 1995 - where the same could be said for losers.

Though there are only two years where winners indexed, the value of the test statistic in the first-half of both years is insignificant, suggesting that managers actually changed their portfolios in the second half. There are additional years where the test statistics increase, but remain negative or insignificant. To control for the behavior of the managers in the first half, every year we perform a cross-sectional regression of each fund’s second-half $R^2$ on its first-half $R^2$ and a winner dummy variable. Though the $t$-statistic of the time series average first-half $R^2$ coefficients is 12.2, the $t$-statistic of the average winner coefficients is only .4. We conclude that it would be difficult to argue with this dataset that indexing is prevalent

---

9 Regression results verify this in simulations with no managerial gaming effects and in our actual sample.
10 Survivorship would tend to bias the results towards a positive relationship between $\Delta$ and the market standard deviation, in so far as it results in the removal of high risk losers from the sample.
among winners in the second-half of the year.\footnote{We also test for a negative relationship between the midyear performance gap between the median winner and the index, and the second-half test statistics in Table II. The results are insignificant.}

In Table III, we report slope coefficients, standard errors, and $R^2$'s for two bivariate regressions of $\Delta$ on the first-half market volatility ($\sigma$) and the first-half $\Delta$ as a control. The first regression is based on a simulation where fund managers choose their portfolios at the beginning of the year and hold them for the entire year. In this setting, it is possible to see what relationship between $\Delta$ and $\sigma$ obtains in the absence of managerial effects.

We construct a sequence of mutual fund industry returns according to the parameters of the Brown, Goetzmann, Ibbotson & Ross (1992) study. The methodology is as follows: First, we choose 660 $\beta$'s according to the BGIR parameters. A choice of $\beta$ fixes the variance in non-systematic returns. We then generate a 52-week return series according to a one-factor model parameterized to match actual market returns. In each simulated year, the market standard deviation for that year is chosen from a normal distribution parameterized to match the sample first-half standard deviations of the CRSP index. In the middle of each simulated year, we designate the top 10% "winners" and observe the fraction of winning and losing managers with above-median betas in the second-half. The SPlus code for the procedure is included in Appendix B. The entire procedure is repeated 10,000 times and the regression is performed. The second regression is based on the actual fund data.

As expected, there is a significantly positive relationship between the first-half market standard deviation and the second-half $\Delta$ in the absence of managerial gaming effects. But, in the actual data, there is a significantly negative relationship as predicted by the model.
Table II
Wilcoxon Rank-Sum Test Results

Do Winners Index?

In each tournament year, funds are classified as winners or losers based on their total return through the first half. Winners are those funds whose first-half total return is higher than the first-half total return of the CRSP value-weighted index of NYSE/AMEX/NASDAQ stocks. The $R^2$ in a regression of a fund's second-half returns on the second-half returns of the index is obtained for all winner and loser funds. A Wilcoxon Rank-Sum test is performed on these values. The test statistic, a standardized version of the winner rank sums, is approximately normally distributed for large sample sizes. The value of the test statistic $Z_2$ in each year is reported. The value of the statistic in the first half $Z_1$ of each year is reported for comparison.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number</th>
<th>Winners</th>
<th>$Z_1$</th>
<th>$Z_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>114</td>
<td>30</td>
<td>-0.98</td>
<td>-0.61</td>
</tr>
<tr>
<td>1985</td>
<td>126</td>
<td>45</td>
<td>0.01</td>
<td>2.48</td>
</tr>
<tr>
<td>1986</td>
<td>137</td>
<td>54</td>
<td>1.81</td>
<td>1.68</td>
</tr>
<tr>
<td>1987</td>
<td>152</td>
<td>70</td>
<td>2.30</td>
<td>1.74</td>
</tr>
<tr>
<td>1988</td>
<td>180</td>
<td>69</td>
<td>-5.10</td>
<td>-3.87</td>
</tr>
<tr>
<td>1989</td>
<td>189</td>
<td>85</td>
<td>0.75</td>
<td>-0.26</td>
</tr>
<tr>
<td>1990</td>
<td>201</td>
<td>158</td>
<td>1.33</td>
<td>3.00</td>
</tr>
<tr>
<td>1991</td>
<td>222</td>
<td>111</td>
<td>-3.49</td>
<td>-3.29</td>
</tr>
<tr>
<td>1992</td>
<td>254</td>
<td>71</td>
<td>-1.23</td>
<td>0.14</td>
</tr>
<tr>
<td>1993</td>
<td>319</td>
<td>108</td>
<td>-3.50</td>
<td>-1.29</td>
</tr>
<tr>
<td>1994</td>
<td>413</td>
<td>168</td>
<td>1.94</td>
<td>0.94</td>
</tr>
<tr>
<td>1995</td>
<td>532</td>
<td>254</td>
<td>-3.55</td>
<td>-4.93</td>
</tr>
<tr>
<td>1996</td>
<td>660</td>
<td>257</td>
<td>-0.55</td>
<td>-0.69</td>
</tr>
</tbody>
</table>
Table III
Regression Results (Type II)

In each tournament year, funds are classified as winners or losers based on their total return through the first half. Winners are those funds whose first-half total return is in the top 10\%ile of all funds in the sample that year. Two bivariate regressions of $\Delta = f_w - f_l$ are performed with the first-half $\Delta$ and the market standard deviation ($\sigma$) as independent variables. We define $f_w$ ($f_l$) as the fraction of winners (losers) with second-half betas larger than the median manager's. The first-half $\Delta$ is calculated similarly using first-half betas. The first regression is based on a simulation of fund returns with no managerial gaming effects. The second regression is based on our actual data. The estimated coefficients, standard errors and $R^2$'s are presented in each case.

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Dependent Variable: $\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>simulated</td>
</tr>
<tr>
<td>Intercept</td>
<td>-.0046</td>
</tr>
<tr>
<td></td>
<td>(.0011)</td>
</tr>
<tr>
<td>First-half $\Delta$</td>
<td>.9762</td>
</tr>
<tr>
<td></td>
<td>(.0009)</td>
</tr>
<tr>
<td>Standard Deviation ($\sigma$)</td>
<td>.4692</td>
</tr>
<tr>
<td></td>
<td>(.0659)</td>
</tr>
<tr>
<td>$N$</td>
<td>10000</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.99</td>
</tr>
</tbody>
</table>

2.4 Conclusion

Brown, Harlow, and Starks (1996) suggest that “viewing the mutual fund market as a tournament in which all funds having comparable investment objectives compete with one another provides a useful framework for a better understanding of the portfolio management decision-making process.” We argue that once a formal model of tournaments is considered, our empirical analysis as well as that of previous researchers supports this suggestion.

One interesting implication of these findings is that fund managers think about what their competition is doing. Competition as a non-informational motive for trade may have a role to play in explaining why Lakonishok & Smidt (1986) find turnover is higher for winning stocks, why Brown & Goetzmann (1995) find winning funds pursue common management strategies with high total risk, and why Carhart, Kaniel, Musto & Reed (1999) find winning funds deliberately cause price shifts with buy orders at quarter/year end.
Competition also may have a role to play in solving the moral hazard problems that potentially arise in the delegated portfolio management settings studied by Admati & Pfleiderer (1997). It may be that the winner-take-all behavior of fund investors and an asset-based management fee is the best way to make sure that fund managers serve their shareholder's interests.
Chapter 3

Are Stock Returns Stable?

Abstract
We revisit earlier studies of the marginal distribution of stock returns and the properties of stable laws. We apply recent advances in bootstrapping techniques to test whether stock returns exhibit stability under addition. Statistically, we are unable to reject stability for the indexes and large stocks in our sample. However, we find significant economic rejections of the stable hypothesis for all assets.

Co-Authors: Blake LeBaron and Andrew Lo

3.1 Introduction

What is the probability that the stock market - as measured by the S&P 500 - will drop by more than 19% tomorrow? What is the largest “typical loss” one might expect over a period of a month on a given portfolio of derivative securities? How much equity should a company hold to protect its operations from the effects of adverse market movements over the next quarter?

Questions of this type are becoming increasingly important for the management of risk in financial institutions. Meaningful answers to them ultimately require a suitable model of the stock return generating process, and, in particular, the distribution of stock returns conditional on past information.

In the special case that the conditional distribution is independent of past prices, many
questions of interest can be answered by studying the marginal distribution of stock returns.\footnote{A description of the marginal distribution of returns does not have economic content, however. An institution may value a 5\% chance of a $100$ million loss differently, depending on whether it happens in good times or bad. Ait-Sahalia and Lo (1997) develop the implications of this for Value at Risk (VaR).} Though Lo & MacKinlay (1988) and others document significant departures from the random walk in practice, properties of the marginal distribution may still merit study if the actual amount of dependence in the data is not economically significant.

In this paper, we focus on an especially important characteristic of the marginal distribution from the standpoint of risk management. Though stock returns appear normally distributed, large magnitude returns occur more frequently than a normal distribution would predict. To illustrate, consider a market analyst who on October 16, 1987 used the mean and standard deviation of all S&P 500 daily returns since 1962 to parameterize a normal distribution as his model for the marginal distribution of stock returns. If he were asked to predict the number of days during the next 8 years on which the stock market would drop in excess of 3, 5, or 19\%, his answer in all cases would be zero. Yet, as we know by now, there were 11, 4, and 1 such days, respectively.

Though a normal distribution does not capture the excess kurtosis present in actual stock returns, it has a number of properties that make it useful for theoretical work in Finance. Most important of these is stability, i.e. a sum of normal random variables is also normal. In Finance, this property is particularly important. Returns on a set of stocks are added cross-sectionally to form portfolio returns. Sequences of returns on a given stock are added over time to form holding period returns. If the distributions of portfolio and holding period returns differ from the summands, theoretical results are more difficult to obtain.

To preserve the appealing theoretical properties of the normal distribution while attempting to explain the excess kurtosis in the data, Mandelbrot (1960, 1961, 1963) and Fama (1965a) introduced the family of stable distributions to the field of financial economics. The normal is a member of this family, but the other members exhibit higher degrees of kurtosis giving them potential to better fit the data.

Early testing supported the use of non-normal stable laws. Inspired by this evidence, Fama (1965b) and Samuelson (1967) discuss the implications of non-normal stable laws for...
capital asset pricing models. McCulloch (1996) extends these discussions to the pricing of derivative securities.

However recent testing has been less supportive. For instance, Officer (1972), Blattberg & Gonedes (1974), Fielitz & Rozelle (1983), and Akgiray & Booth (1988) all find evidence of non-stability in stock returns.

In this paper, we revisit earlier studies of the marginal distribution of stock returns and the properties of stable laws. Our central question is the following: Do non-normal stable laws provide a statistically and economically accurate description of the marginal distribution of stock returns?

To answer the question, we apply recent advances in bootstrapping techniques to evaluate the quality of a commonly used test for the invariance of the characteristic exponent over sums of stable random variables, i.e. the stability-under-addition test. According to Baillie (1993), Lau & Lau (1993), and Deo (1996), previous rejections of stability may be due in part to inadequate statistical inference. Statistically, we are unable to reject stability for the indexes and large stocks in our sample. However, we find significant economic rejections of the stable hypothesis for all assets.

The paper proceeds as follows. In Section 2, we review the statistical properties of stable distributions. Section 3 describes our implementation of the stability-under-addition test. Section 4 presents the empirical results. Section 5 discusses the economic significance of the results, and section 6 concludes.

### 3.2 A Review of Stable Distributions

The statistical theory of stable distributions is well developed and has been reviewed extensively in several references (see for example, Levy (1925) and Feller (1971)). Hence, we shall provide only a brief review of their definitions and properties here.

Although a stable random variable \( X \) is characterized by its stability under addition, a more formal definition is given by its characteristic function \( \phi_x(t) \equiv E[\exp(itx)] \):
\[ \phi_z(t) = \exp \left[ i \mu t - \gamma |t|^{\alpha} (1 + i \beta \text{sgn}(t) \omega(t, \alpha)) \right] \quad (3.1) \]

where \( t \) is any real number, \( i = \sqrt{-1} \), and

\[
\omega(t, \alpha) = \begin{cases} 
\tan \left( \frac{\pi \alpha}{2} \right) & \text{if } \alpha \neq 1 \\
\frac{2}{\pi} \log |t| & \text{if } \alpha = 1 
\end{cases}
\]

and \( \alpha \in (0, 2) \) and \( \beta \in [-1, 1] \).

The characteristic equation above shows that stable distributions are characterized by four parameters: the characteristic exponent \( \alpha \), the location parameter \( \mu \), the skewness parameter \( \beta \), and the scale parameter \( \gamma \). The parameter \( \mu \) is the mean of \( X \) when \( \alpha > 1 \). When \( \alpha \leq 1 \), the mean is infinite, and \( \mu \) must be interpreted as a location parameter such as the median, e.g., when \( \beta = 0 \). When \( \beta < 0 \), this yields a distribution skewed to the left, and \( \beta > 0 \) yields a distribution skewed to the right. When \( \beta = 0 \), the distribution is symmetric. The scale parameter \( \gamma \) is one-half the variance of \( X \) when \( \alpha = 2 \). However, when \( \alpha < 2 \), the variance is infinite, and \( \gamma \) must be interpreted differently, such as one-half the interquartile range, e.g., when \( \alpha = 1 \) and \( \beta = 0 \).

The characteristic exponent \( \alpha \) is the parameter of most interest since it governs the total probability contained in the tails of the distribution. Stable distributions with smaller values of \( \alpha \) have more probability in their extreme tails. Moreover, the only finite moments \( X \) possesses are moments \( E[X^q] \) where \( q < \alpha \).

Finally, there are only three sets of parameter values for which simple expressions for the PDF of \( X \) can be obtained in closed form:

\[ \beta = 0, \quad \alpha = 1 \Rightarrow \text{Cauchy} \]
\[ \alpha = 2 \Rightarrow \text{Normal} \]
\[ \alpha = \frac{1}{2} \Rightarrow 1/\chi_1^2 \]
3.3 Test of Stability – Stability Under Addition

One of the defining properties of stable random variables is the fact that the distribution of a sum of IID stable random variables is also stable and has the same value of $\alpha$ and $\beta$ as the component summands. Moreover, Fama (1965) shows that the invariance of the characteristic exponent also holds for independent summands that differ in scale $\gamma$ and location $\mu$. Thus, ARMA$(p,q)$ processes, which can be be expressed as sums of independent stable random variables, also exhibit stability under addition.

A heteroscedastic process will not generally exhibit this property because its observations can only be written as products of independent random variables, which, as de Vries (1991) and Ghose & Khroner (1995) demonstrate, are stable only in very special cases. However, Bidarkota & McCulloch (1996) demonstrate that a GARCH process can still possess conditionally stable increments, while being unconditionally non-stable.

The stability property implies that under the null hypothesis that the data are independent symmetric stable random variables with the same value of $\alpha$, the original series and derivative series of $k$-observation sums should yield identical estimates of $\alpha$ for any $k$.

3.3.1 Implementation of the Test Statistic

Tests of stability based on $\alpha$ depend critically on methods for estimating the characteristic exponent. The fact that simple expressions for the density exist in only three cases makes standard estimation methods such as maximum likelihood difficult to implement\(^2\), though Nolan (1996) shows that progress is being made in this area.

Nevertheless, two popular methods have emerged for estimating $\alpha$. The first method, due to Fama & Roll (1971) and extended by McCulloch (1986), is based on sample order statistics. The second method, due to Press (1972) and Koutrouvelis (1980, 1981), is based on a regression involving the empirical log-characteristic function of the stable distribution. A comparison of these two methods is performed by Akgiray & Lamoureux (1989) in which they conclude that the second method is the more reliable one. Our test, therefore, relies on the $\alpha$-estimation procedure of Koutrouvelis (1980, 1981).

The estimation procedure of Koutrouvelis is based on the following observation:

\[ \log \left[ -\log |\phi(t)|^2 \right] = \log(e^\alpha) + \alpha \log |t| \]  

(3.2)

Equation 2 depends only on the characteristic exponent \( \alpha \) and the scale parameter \( c \), suggesting an estimate of \( \alpha \) can be obtained by regressing the empirical characteristic function evaluated at a set of appropriate values of \( t \) on \( \log |t| \). The size of the set and the appropriate values of \( t \) are dependent on the actual values of \( \alpha \) and \( c \) in the data. Koutrouvelis (1980, 1981) recommends obtaining a preliminary estimate of these parameters via Fama & Roll (1971), and gives guidance on the appropriate set of \( t \)-values. The Splus Code implementing the Koutrouvelis (1980, 1981) estimator is included in Appendix C.

Thus, the entire procedure is as follows: we form 50 time-aggregated return series' from the original data set and obtain an \( \hat{\alpha} \) for each series. Under the null-hypothesis, all of these \( \hat{\alpha} \)'s should be equal. To make the requisite multiple comparisons, we fit a regression line to the estimates. Our notion of "equality" is an estimated slope "close" to zero.

To assess statistical significance, we obtain the sampling distribution of the slope of this best-fit line. We use both the IID bootstrap of Efron (1979) and the moving block bootstrap (block size = 50) of Künsch (1989) and Liu & Singh (1992). The block bootstrap enables us to replicate time series dependence in the data, if any. The comparison of block and IID bootstraps allows us to get a sense of the actual amount of dependence in the data. Liu & Singh (1995) show that confidence intervals delivered by the usual IID bootstrap are conservative in a non-IID setting. Thus, wider confidence intervals for the block bootstrap are consistent with the presence of dependence in the data.

\[ ^3 \text{The choice of block size is, of course, a matter of judgement. Politis & Romano (1994) and Lahiri (1995) discuss the properties of the stationary block bootstrap, in which the block size is a random variable according to some distribution. Our choice of 50 represents, in a sense, a strongly held prior as to the appropriate length of the block size. We have chosen 50 based on our view that 2 months of positive autocorrelations in return volatility is not unreasonable. The results do not change appreciably for other reasonable block sizes.} \]
3.3.2 Bootstrap Methods and Stable Distributions

Introduced by Efron (1979), bootstrap methods have proven to be a useful tool in the analysis of many statistical problems. Through the use of extensive computer simulations they allow us to analyze the reliability of estimated statistics by resampling from the original data set. Among the many strengths of the bootstrap are its ability to estimate standard errors and confidence regions for any arbitrary estimator. For many of the estimators used here, deriving analytical expressions for these errors would be exceedingly complicated if not impossible. Another advantage is that the bootstrap may yield better small sample properties for some of the estimators, thereby improving on asymptotic approximations.

Although the bootstrap appears to have general applicability, and its strengths are well documented, it should not be applied without caution. There are some well known cases in which it can fail. Especially relevant to the work here are cases concerning stable distributions. Whereas bootstrap distributions generally converge to the true asymptotic distribution as the sample size goes to infinity, Hall (1990) shows that for certain statistics and distributions this consistency does not hold. Of great concern to our work are Hall's results which demonstrate that the bootstrap distribution of the sample mean does not converge to the true distribution if the variance does not exist. This is troubling for the tests performed here, since we are often operating subject to a null hypothesis of a stable law without second moments for the underlying stock return distributions.

However, we believe this result does not apply to the types of estimators used in this paper. Bickel & Freedman (1981) derive a fairly broad class of estimators under which the bootstrap consistency holds. It is believed that the function considered in this paper fall into this class.

To see why, we consider the Koutrouvelis estimator first. The first stage of this estimator is a von Mises functional, more specifically a u-statistic. It is built from sums of functions of the observed series,

\[
\begin{align*}
  w_{n,j} &= \frac{1}{n} \sum_{i=1}^{n} \sin(t_jX_i) \\
  y_{n,j} &= \frac{1}{n} \sum_{i=1}^{n} \cos(t_jX_i).
\end{align*}
\]
Therefore, second moments exist for the individual functions, and their means $w$ and $y$. The estimator is built from regressions using

$$z_{n,j} = -\log(w_{n,j}^2 + y_{n,j}^2) \text{ regressed on}$$

$$u_j = \log|t_j|$$

for selected $t_j$. Given that the regressand is a continuous function of an object which is converging in distribution, it converges as well.

The results of this regression yield $\alpha$ estimates which are then used in another regression to test their stability over temporal aggregation. Though it would be difficult to prove convergence to the correct distribution at each level of the entire test, we observe that at each stage we have a smooth function of an object whose convergence properties are well known. We therefore believe that the Koutrouvellis estimator is covered by the Bickel & Friedman (1981) results.

### 3.4 Empirical Results

#### 3.4.1 Data and Implementation

We use the daily returns of three indexes and ten individual securities that are continuously listed over the entire sample period from July 2, 1962 through December 31, 1995. Summary statistics for these return series are shown in Table 1a. A descriptive comparison of each asset's return distribution versus an appropriately calibrated normal distribution is shown in Table 1b.
Table 1a
Stock Market Returns, 1962 to 1995

Summary statistics for daily returns (in percent) of CRSP value- and equal-weighted stock indexes and ten individual securities continuously listed over the entire sample period from July 2, 1962 to December 31, 1995.

<table>
<thead>
<tr>
<th>Security</th>
<th>Mean</th>
<th>Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>VW Index</td>
<td>.044</td>
<td>0.82</td>
<td>-1.33</td>
<td>34.92</td>
<td>-18.10</td>
<td>8.87</td>
</tr>
<tr>
<td>EW Index</td>
<td>.073</td>
<td>0.76</td>
<td>-0.93</td>
<td>26.03</td>
<td>-14.19</td>
<td>9.83</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>.043</td>
<td>0.80</td>
<td>-1.71</td>
<td>45.50</td>
<td>-18.80</td>
<td>8.31</td>
</tr>
<tr>
<td>IBM</td>
<td>.039</td>
<td>1.42</td>
<td>-0.18</td>
<td>12.48</td>
<td>-22.96</td>
<td>11.72</td>
</tr>
<tr>
<td>General Signal</td>
<td>.054</td>
<td>1.66</td>
<td>0.01</td>
<td>3.35</td>
<td>-13.46</td>
<td>9.43</td>
</tr>
<tr>
<td>Wrigley</td>
<td>.072</td>
<td>1.45</td>
<td>-0.00</td>
<td>11.03</td>
<td>-18.67</td>
<td>11.89</td>
</tr>
<tr>
<td>Interlake</td>
<td>.043</td>
<td>2.16</td>
<td>0.72</td>
<td>12.35</td>
<td>-17.24</td>
<td>23.08</td>
</tr>
<tr>
<td>Raytech</td>
<td>.050</td>
<td>3.39</td>
<td>2.25</td>
<td>59.40</td>
<td>-57.90</td>
<td>75.00</td>
</tr>
<tr>
<td>Ampco-Pittsburgh</td>
<td>.053</td>
<td>2.41</td>
<td>0.66</td>
<td>5.02</td>
<td>-19.05</td>
<td>19.18</td>
</tr>
<tr>
<td>Energen</td>
<td>.054</td>
<td>1.41</td>
<td>0.27</td>
<td>5.91</td>
<td>-12.82</td>
<td>11.11</td>
</tr>
<tr>
<td>General Host</td>
<td>.070</td>
<td>2.79</td>
<td>0.74</td>
<td>6.18</td>
<td>-23.53</td>
<td>22.92</td>
</tr>
<tr>
<td>Garan</td>
<td>.079</td>
<td>2.35</td>
<td>0.72</td>
<td>7.13</td>
<td>-16.67</td>
<td>19.07</td>
</tr>
<tr>
<td>Continental Materials</td>
<td>.143</td>
<td>5.24</td>
<td>0.93</td>
<td>6.49</td>
<td>-26.92</td>
<td>50.00</td>
</tr>
</tbody>
</table>
Table 1b
Summary Statistics for Non-Overlapping Multi-Day S&P 500 Index Returns

Summary statistics for daily and non-overlapping multiperiod continuously compounded returns of the S&P 500 from July 2, 1962 to December 31, 1995. For comparison, percentiles of the normal distribution with identical mean and standard deviation are also given. All entries are reported as percentages.

<table>
<thead>
<tr>
<th>Return</th>
<th>Mean</th>
<th>S.D.</th>
<th>1%</th>
<th>2.5%</th>
<th>5%</th>
<th>95%</th>
<th>97.5%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Day</td>
<td>0.04</td>
<td>0.86</td>
<td>-2.10</td>
<td>-1.60</td>
<td>-1.28</td>
<td>1.34</td>
<td>1.74</td>
<td>2.27</td>
</tr>
<tr>
<td>Normal</td>
<td>0.04</td>
<td>0.86</td>
<td>-1.96</td>
<td>-1.64</td>
<td>-1.37</td>
<td>1.46</td>
<td>1.73</td>
<td>2.05</td>
</tr>
<tr>
<td>5-Day</td>
<td>0.22</td>
<td>2.03</td>
<td>-5.21</td>
<td>-3.81</td>
<td>-2.93</td>
<td>3.32</td>
<td>3.85</td>
<td>5.10</td>
</tr>
<tr>
<td>Normal</td>
<td>0.22</td>
<td>2.03</td>
<td>-4.50</td>
<td>-3.76</td>
<td>-3.12</td>
<td>3.55</td>
<td>4.19</td>
<td>4.93</td>
</tr>
<tr>
<td>10-Day</td>
<td>0.43</td>
<td>2.91</td>
<td>-7.17</td>
<td>-5.52</td>
<td>-4.37</td>
<td>4.65</td>
<td>6.08</td>
<td>7.71</td>
</tr>
<tr>
<td>Normal</td>
<td>0.43</td>
<td>2.91</td>
<td>-6.33</td>
<td>-5.27</td>
<td>-4.35</td>
<td>5.22</td>
<td>6.13</td>
<td>7.20</td>
</tr>
<tr>
<td>15-Day</td>
<td>0.65</td>
<td>3.54</td>
<td>-8.60</td>
<td>-7.02</td>
<td>-5.37</td>
<td>6.03</td>
<td>7.42</td>
<td>8.81</td>
</tr>
<tr>
<td>Normal</td>
<td>0.65</td>
<td>3.54</td>
<td>-7.59</td>
<td>-6.29</td>
<td>-5.17</td>
<td>6.47</td>
<td>7.59</td>
<td>8.89</td>
</tr>
<tr>
<td>20-Day</td>
<td>0.87</td>
<td>4.25</td>
<td>-10.91</td>
<td>-8.96</td>
<td>-6.36</td>
<td>7.66</td>
<td>9.06</td>
<td>11.31</td>
</tr>
<tr>
<td>Normal</td>
<td>0.87</td>
<td>4.25</td>
<td>-9.01</td>
<td>-7.45</td>
<td>-6.12</td>
<td>7.85</td>
<td>9.19</td>
<td>10.75</td>
</tr>
<tr>
<td>25-Day</td>
<td>1.08</td>
<td>4.66</td>
<td>-12.67</td>
<td>-8.55</td>
<td>-6.83</td>
<td>7.80</td>
<td>9.28</td>
<td>10.68</td>
</tr>
<tr>
<td>Normal</td>
<td>1.08</td>
<td>4.66</td>
<td>-9.76</td>
<td>-8.05</td>
<td>-6.58</td>
<td>8.74</td>
<td>10.21</td>
<td>11.92</td>
</tr>
<tr>
<td>30-Day</td>
<td>1.30</td>
<td>5.12</td>
<td>-13.27</td>
<td>-9.68</td>
<td>-7.71</td>
<td>8.84</td>
<td>11.54</td>
<td>14.55</td>
</tr>
<tr>
<td>Normal</td>
<td>1.30</td>
<td>5.12</td>
<td>-10.61</td>
<td>-8.73</td>
<td>-7.12</td>
<td>9.72</td>
<td>11.33</td>
<td>13.21</td>
</tr>
<tr>
<td>35-Day</td>
<td>1.50</td>
<td>5.38</td>
<td>-16.26</td>
<td>-10.47</td>
<td>-6.46</td>
<td>9.25</td>
<td>10.84</td>
<td>14.46</td>
</tr>
<tr>
<td>Normal</td>
<td>1.50</td>
<td>5.38</td>
<td>-11.01</td>
<td>-9.04</td>
<td>-7.35</td>
<td>10.35</td>
<td>12.05</td>
<td>14.02</td>
</tr>
<tr>
<td>40-Day</td>
<td>1.72</td>
<td>5.84</td>
<td>-13.08</td>
<td>-8.49</td>
<td>-6.44</td>
<td>10.32</td>
<td>12.88</td>
<td>13.88</td>
</tr>
<tr>
<td>Normal</td>
<td>1.72</td>
<td>5.84</td>
<td>-11.88</td>
<td>-9.74</td>
<td>-7.89</td>
<td>11.33</td>
<td>13.17</td>
<td>15.31</td>
</tr>
<tr>
<td>45-Day</td>
<td>1.96</td>
<td>6.41</td>
<td>-13.58</td>
<td>-10.10</td>
<td>-7.89</td>
<td>10.70</td>
<td>12.09</td>
<td>13.33</td>
</tr>
<tr>
<td>Normal</td>
<td>1.96</td>
<td>6.41</td>
<td>-12.96</td>
<td>-10.61</td>
<td>-8.59</td>
<td>12.50</td>
<td>14.52</td>
<td>16.87</td>
</tr>
<tr>
<td>50-Day</td>
<td>2.15</td>
<td>6.48</td>
<td>-17.35</td>
<td>-12.11</td>
<td>-8.82</td>
<td>10.58</td>
<td>12.73</td>
<td>14.54</td>
</tr>
<tr>
<td>Normal</td>
<td>2.15</td>
<td>6.48</td>
<td>-12.92</td>
<td>-10.55</td>
<td>-8.51</td>
<td>12.80</td>
<td>14.84</td>
<td>17.21</td>
</tr>
</tbody>
</table>
The indexes are CRSP NYSE-AMEX market-return indexes, while the stocks are randomly selected representatives of the continuously listed stocks in each CRSP size-decile. The daily returns are continuously compounded and include dividends. For each asset, there are generally 8431 observations. However, some series are smaller, but never less than 8427, reflecting missing observations in the CRSP file.

For indexes and individual securities, we perform our stability-under-addition test twice, first using sums formed from non-overlapping observations and second from overlapping observations. The estimations based on overlapping sums have the advantage of using \( n-k+1 \) observations versus \( n/k \) for the non-overlapping case. By using overlapping \( k \)-sums, we obtain a more efficient estimates of \( \alpha \) and hence a more powerful test. However, the results for both sets of tests are similar. Since the results are easier to interpret for the non-overlapping case, we report these only.

Figure 1 contains estimated sample paths for the actual data.
Figure 1
Estimates of $\alpha$ for Selected Daily U.S. Stock Market Prices

Estimates of $\hat{\alpha}(k)$ of stable parameter $\alpha$ using non-overlapping $k$-day continuously compounded returns for various securities and indexes, from July 2, 1962 to December 31, 1995 (8,431 daily observations), using the method of Koutrouvelis (1980). For comparison, the plot of the regression $\hat{\alpha}(k) = \delta_0 + \delta_1 k + \varepsilon_k$ is displayed.

3.4.2 Results of the Stability Under Addition Test

Results For Stock Indexes

The first three rows of Table 2 report the slope and intercept of a regression of estimated $\alpha$ on sum-size ($k$) and a confidence interval for the slope coefficient obtained via the IID and
block bootstrap.

Similar to the findings of previous studies, the point estimates are generally below 2 and they exhibit a tendency to increase with sum-size. According to Figure 1 and the regression results in Table 2, all 50 point estimates are generally below 2 and they exhibit a tendency to increase with sum-size, as well.

The tendency for the point estimate to increase with sum-size is seen most clearly in the results for the equal-weighted index. For this asset, the slope coefficient is $1.79 \times 10^{-3}$ versus $-0.19 \times 10^{-3}$ and $0.11 \times 10^{-3}$ for the value-weighted and S&P 500 indexes, respectively. The equal-weighted index includes a higher proportion of small stocks. As we shall see in the next section, these stocks generally tend to produce lower point estimates which increase dramatically with sum-size.

In contrast with previous studies, the bootstrap confidence intervals suggest that there is not sufficient evidence here for a rejection of stability. The intervals include zero in all cases except the IID/equal-weighted case.

**Results For Individual Securities**

Rows 4-13 of Table 2 report regression coefficients and confidence intervals for individual stocks. Again the point estimates are lower than 2 and they exhibit a tendency to increase with sum-size.

The point estimates of $\alpha$ are generally lower for the individual stocks than for the indexes, as evidenced by average point estimates of $\alpha$ that are on average lower for individual stocks than for indexes (1.63 versus 1.75). Moreover, the tendency to increase with sum-size is more pronounced for individual stocks than for indexes, as evidence by slope coefficients for individual stocks which are on average higher than those for indexes ($4.29 \times 10^{-3}$ versus $0.70 \times 10^{-3}$).

The point estimates of $\alpha$ also tend to decrease with the market capitalization of the stock, though this pattern is not monotonic. Starting with IBM, the largest of the ten stocks in terms of market capitalization, the average point estimate of $\alpha$ is 1.88, while for Continental Materials, the smallest stock by the same measure, it is 1.40. Moreover, the tendency for estimates of $\alpha$ to increase with sum-size is more pronounced for the smaller stocks, though
this pattern too is not monotonic. The slope coefficient for IBM is \(0.44 \times 10^{-3}\) while for Continental Materials it is \(10.40 \times 10^{-3}\).

This time, confirming the observations are the results of the bootstrapped confidence intervals. They demonstrate that a slope of zero is an outcome likely to happen no more than 2.5% of the time for all of the stocks except the largest - IBM. The IID bootstrap rejects General Signal and Wrigley, but the block bootstrap does not. Table 2 also indicates that the IID bootstrap intervals are often too conservative to include the point estimates, as expected when the data are dependent.

### Table 2
Estimates of \(\alpha\) for Selected Daily U.S. Stock Market Prices

Estimates \(\hat{\alpha}(k)\) of stable parameter \(\alpha\) using non-overlapping 4-day continuously compounded returns for various securities and indexes, from July 2, 1962 to December 31, 1995 (8,431 daily observations), using the method of Koutrouvelis (1980). \(\hat{\delta}_0\) and \(\hat{\delta}_1\) are coefficient estimates from the regression \(\hat{\alpha}(k) = \delta_0 + \delta_1 k + \epsilon_k\) and IID and block bootstrap percentiles are based on 1,000 replications.

<table>
<thead>
<tr>
<th>Security</th>
<th>(\hat{\delta}_0)</th>
<th>(\hat{\delta}_1 \times 10^{-3})</th>
<th>IID-BS (\hat{\delta}_1 \times 10^{-3})</th>
<th>Block-BS (\hat{\delta}_1 \times 10^{-3})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5%</td>
<td>50%</td>
<td>97.5%</td>
<td>2.5%</td>
</tr>
<tr>
<td>VW Index</td>
<td>-0.52</td>
<td>1.62</td>
<td>3.35</td>
<td>-2.08</td>
</tr>
<tr>
<td>EW Index</td>
<td>0.59</td>
<td>2.89</td>
<td>4.84</td>
<td>-1.19</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>-0.67</td>
<td>1.52</td>
<td>3.23</td>
<td>-1.89</td>
</tr>
<tr>
<td>IBM</td>
<td>-0.71</td>
<td>1.34</td>
<td>2.72</td>
<td>-1.05</td>
</tr>
<tr>
<td>General Signal</td>
<td>0.05</td>
<td>1.92</td>
<td>3.32</td>
<td>-0.23</td>
</tr>
<tr>
<td>Wrigley</td>
<td>1.05</td>
<td>3.19</td>
<td>4.81</td>
<td>-0.54</td>
</tr>
<tr>
<td>Interlake</td>
<td>4.38</td>
<td>6.33</td>
<td>7.97</td>
<td>0.18</td>
</tr>
<tr>
<td>Raytech</td>
<td>3.05</td>
<td>5.96</td>
<td>8.19</td>
<td>3.13</td>
</tr>
<tr>
<td>Ampco-Pittsburgh</td>
<td>0.74</td>
<td>2.56</td>
<td>3.90</td>
<td>2.71</td>
</tr>
<tr>
<td>Energen</td>
<td>0.09</td>
<td>1.98</td>
<td>3.31</td>
<td>0.25</td>
</tr>
<tr>
<td>General Host</td>
<td>0.48</td>
<td>2.44</td>
<td>3.88</td>
<td>0.69</td>
</tr>
<tr>
<td>Garan</td>
<td>1.69</td>
<td>3.87</td>
<td>5.39</td>
<td>2.72</td>
</tr>
<tr>
<td>Continental Materials</td>
<td>4.21</td>
<td>5.98</td>
<td>7.48</td>
<td>6.05</td>
</tr>
</tbody>
</table>
3.5 Stable Laws and Risk Management

Previous sections have shown statistically that the stock returns can be distinguished from the stable distributions. However, the tests are primarily statistical, and do not offer an intuitive magnitude of the deviation from stability in practice. The question remains whether stable laws, although not exactly correct, are a good approximation for stock returns.

We explore this question from the standpoint of estimating tail probabilities. If stock returns follow a stable law, then estimating tail probabilities, and distribution quantiles for long horizon returns is an easy task. The scaling properties of stable distributions allow estimation to proceed at the daily horizon, and then be expanded to whatever horizon is needed. This scaling property make stable distributions very desirable in practical uses.

We assume that the researcher has already estimated the Koutrouvelis value for $\alpha$ and will use this in all scaling procedures. We use an algorithm developed by Nolan (1996) to convert our estimates into quantile levels. Table 3 gives the sum of the estimated tail probabilities in the left and right 5 percent tails - $\Pr(|r| > Q_{.95})$ - using the estimated quantiles for 20- and 60-day series'. Under the stable null hypothesis these should be 10 percent at both horizons since they have been scaled using the appropriate scaling procedure.

In most cases in the table, we observe the tail probabilities dropping off significantly as the horizon is lengthened. For example, Garan Inc. has a negligible number of observations in the estimated 60-day tails. More than 2% of the observations fall in the estimated 20-day tail, however. IBM is an exception. Its 20- and 60-day tail probabilities are nearly identical.

To judge significance the bootstrap is used. Two different bootstrap procedures are used. There first is an independent bootstrap where the daily returns series is redrawn individually, and the second is a block bootstrap where the new returns series is built by sampling points in blocks of 50.

The numbers $Q_{.025}$ and $Q_{.975}$ are the two-tailed 95% confidence interval of the fraction of observations in the estimated 10% tail from the 1000 bootstrap simulations. The column labeled $\Pr(tail > 0.10)$ indicates the fraction of bootstraps with estimated tail probabilities greater than 0.10. Our assumption is that a risk manager is likely to be concerned about underestimating the frequency of large magnitude events. A large (small) $\Pr(tail > 0.10)$
implies that stable approximations of \( Q_{0.25} \) and \( Q_{0.75} \) often underestimate (overestimate) the frequency of large magnitude events.

The confidence intervals and probabilities show that the critical values generated from the stable null hypothesis are far off target in most cases. For many of the individual stock series, \( \text{Pr}(\text{tail} > 0.10) \) is zero. For the 20-day aggregated Wrigley series using the blocked bootstrap, the confidence region’s right-hand boundary is at 4.96% with a negligible percentage of the runs generating tail probabilities larger than 10 percent. One exception is again IBM, where the confidence region includes 10 percent. However, it should be noted that the distribution is still skewed to the left a little, with only 6 percent of the simulations giving a tail probability greater than 10 percent.

For comparison, the final rows of the table present results from a simulated stable distribution with \( \alpha = 1.6 \) and sample size equal to that of the stock return series. It is clear here that there is no drop off in tail probability. It shows an estimated tail probability at the 60-day aggregation level close to 10%.

The results in Table 3 can be taken as evidence that the deviations from actual return distributions the stable law probably are economically significant\(^4\). Using a stable scaling rule could greatly overestimate the probability of large tail events at the long horizon in many cases. It also should be remembered here that the gaussian distribution is in the stable class and scales with \( \alpha = 2 \). Approximations with sample gaussians would also miss the tail probabilities at longer horizons as well.

\(^4\)Of course, the complete case for economic significance would require full modeling of the decision making process and objective functions.
Table 3
Actual Tail Probability vs. Stable Law Estimates

Estimated \( \alpha \) and scale parameters are converted to sample quantiles using the Nolan (1996). These are then used to estimate the probability \( \Pr(|r| > Q_{95}) \) for 20- and 60-day aggregation levels. Bootstrap percentiles are presented below each estimated value (in percentages). The first uses an IID bootstrap resample and the second, a block resample with blocksize set to 50. Also noted is \( \Pr(tail > 0.10) \), the fraction of bootstrap tail probabilities greater than 0.10.

<table>
<thead>
<tr>
<th>Security</th>
<th>20-day Aggregation</th>
<th>60-day Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( Q_{0.05} )</td>
<td>( Q_{0.95} )</td>
</tr>
<tr>
<td>VW Index</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>3.56 7.60</td>
<td>0.00</td>
</tr>
<tr>
<td>EW Index</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>3.33 7.60</td>
<td>0.00</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>3.80 8.31</td>
<td>0.00</td>
</tr>
<tr>
<td>IBM</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>5.94 11.16</td>
<td>12.40</td>
</tr>
<tr>
<td>General Signal</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>3.56 7.84</td>
<td>0.00</td>
</tr>
<tr>
<td>Wrigley</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>1.19 4.28</td>
<td>0.00</td>
</tr>
<tr>
<td>Interlake</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Raytech</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>0.00 0.71</td>
<td>0.00</td>
</tr>
<tr>
<td>Amoco-Pittsburgh</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>2.61 6.18</td>
<td>0.00</td>
</tr>
<tr>
<td>Energen</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>5.46 10.45</td>
<td>5.00</td>
</tr>
<tr>
<td>General Host</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>3.80 8.31</td>
<td>0.00</td>
</tr>
<tr>
<td>Garan</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>0.48 2.85</td>
<td>0.00</td>
</tr>
<tr>
<td>Continental Materials</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Stable (( \alpha = 1.6 ))</td>
<td>IID</td>
<td>BLOCK</td>
</tr>
<tr>
<td></td>
<td>5.70 11.16</td>
<td>11.60</td>
</tr>
</tbody>
</table>
3.6 Conclusion

The marginal distribution of stock returns can be an important object of study, particularly when there is very little dependence in the data. However, our results suggest that not only are the stable laws ill-equipped to handle problems of economic interest, but there is enough dependence in the data to call into question the value of studying the marginal distribution alone.

Whether there is enough dependence to allow for profitable trading strategies is another question, but clearly it is important to consider time series predictability in future studies of the properties of stock returns.
Appendix A

Code and Derivations for Survivorship Analysis

A.1 Splus Code for Means, Variances, Beta & Momentum Strategies

"survivormaker" <-
function()
{
  index <- rnorm(72, mean=.0071, sd=.057)
  betas <- rnorm(1000, mean=.82, sd=.79)
  residuals <- matrix(rnorm(72000, mean=0,
                     sd=.01*(2.55*(betas^2)-1.25*betas+12.71)),
                     ncol=1000, byrow=TRUE)
  returns <- .0031 + matrix(index, ncol=1)%%matrix(betas, nrow=1) + residuals
  cumexcess36 <- apply(returns[1:36,] - matrix(rep(index[1:36], 1000),
                     ncol=1000), 2, sum)
  excessreturns <- returns[37:72,] - matrix(rep(index[37:72], 1000),
                     ncol=1000)
  avgreturns <- apply(excessreturns, 2, mean)
  stdreturns <- sqrt(apply(excessreturns, 2, var))
  betaestimates <- matrix(lsfit(index[1:36], returns[1:36,])$coef[2,], nrow=1)
  future3 <- apply(excessreturns[1:3,], 2, sum)
  future6 <- apply(excessreturns[1:6,], 2, sum)
  future9 <- apply(excessreturns[1:9,], 2, sum)
  future12 <- apply(excessreturns[1:12,], 2, sum)
  future18 <- apply(excessreturns[1:18,], 2, sum)
  future24 <- apply(excessreturns[1:24,], 2, sum)
  future30 <- apply(excessreturns[1:30,], 2, sum)
  future36 <- apply(excessreturns[1:36,], 2, sum)
  yearllycumreturns <- rbind(apply(excessreturns[1:12,], 2, sum),
                       apply(excessreturns[13:24,], 2, sum),
                       apply(excessreturns[25:36,], 2, sum),
                       apply(excessreturns[37:48,], 2, sum),
                       apply(excessreturns[49:60,], 2, sum),
                       apply(excessreturns[61:72,], 2, sum))

72
is95 <- whosurvives(yearlycumreturns, .05)
is90 <- whosurvives(yearlycumreturns, .10)
is80 <- whosurvives(yearlycumreturns, .20)
winners100 <- matrix(quantile(cumexcess36, c(.90,.10)))
winners95 <- matrix(quantile(cumexcess36[is95], c(.90,.10)))
winners90 <- matrix(quantile(cumexcess36[is90], c(.90,.10)))
winners80 <- matrix(quantile(cumexcess36[is80], c(.90,.10)))
is100win <- cumexcess36 >= winners100[1]
is100lose <- cumexcess36 <= winners100[2]
is95win <- is95 & (cumexcess36 >= winners95[1])
is95lose <- is95 & (cumexcess36 <= winners95[2])
is90win <- is90 & (cumexcess36 >= winners90[1])
is90lose <- is90 & (cumexcess36 <= winners90[2])
is80win <- is80 & (cumexcess36 >= winners80[1])
is80lose <- is80 & (cumexcess36 <= winners80[2])
winnercumexcess100 <- apply(excessreturns[,is100win], 1, mean)
losercumexcess100 <- apply(excessreturns[,is100lose], 1, mean)
winnercumexcess95 <- apply(excessreturns[,is95win], 1, mean)
losercumexcess95 <- apply(excessreturns[,is95lose], 1, mean)
winnercumexcess90 <- apply(excessreturns[,is90win], 1, mean)
losercumexcess90 <- apply(excessreturns[,is90lose], 1, mean)
winnercumexcess80 <- apply(excessreturns[,is80win], 1, mean)
losercumexcess80 <- apply(excessreturns[,is80lose], 1, mean)
write(c(sum(winnercumexcess100[1:3]) - losercumexcess100[1:3]),
      sum(winnercumexcess100[1:6] - losercumexcess100[1:6]),
      sum(winnercumexcess100[1:9] - losercumexcess100[1:9]),
      sum(winnercumexcess100[1:12] - losercumexcess100[1:12]),
      sum(winnercumexcess100[1:18] - losercumexcess100[1:18]),
      sum(winnercumexcess100[1:24] - losercumexcess100[1:24]),
      sum(winnercumexcess100[1:30] - losercumexcess100[1:30]),
      sum(winnercumexcess100[1:36] - losercumexcess100[1:36])),
      file="/home/jdtaylor/winlose100", ncolumns=8, append=T)
write(c(sum(winnercumexcess95[1:3]) - losercumexcess95[1:3]),
      sum(winnercumexcess95[1:6] - losercumexcess95[1:6]),
      sum(winnercumexcess95[1:9] - losercumexcess95[1:9]),
      sum(winnercumexcess95[1:12] - losercumexcess95[1:12]),
      sum(winnercumexcess95[1:18] - losercumexcess95[1:18]),
      sum(winnercumexcess95[1:24] - losercumexcess95[1:24]),
      sum(winnercumexcess95[1:30] - losercumexcess95[1:30]),
      sum(winnercumexcess95[1:36] - losercumexcess95[1:36])),
      file="/home/jdtaylor/winlose95", ncolumns=8, append=T)
write(c(sum(winnercumexcess90[1:3]) - losercumexcess90[1:3]),
      sum(winnercumexcess90[1:9] - losercumexcess90[1:9]),
      sum(winnercumexcess90[1:12] - losercumexcess90[1:12]),
      sum(winnercumexcess90[1:18] - losercumexcess90[1:18]),
      sum(winnercumexcess90[1:24] - losercumexcess90[1:24]),
      sum(winnercumexcess90[1:30] - losercumexcess90[1:30]),
      sum(winnercumexcess90[1:36] - losercumexcess90[1:36])),
      file="/home/jdtaylor/winlose90", ncolumns=8, append=T)
write(c(sum(winnercumexcess80[1:3]) - losercumexcess80[1:3]),
      sum(winnercumexcess80[1:9] - losercumexcess80[1:9]),
      sum(winnercumexcess80[1:12] - losercumexcess80[1:12]),
sum(winnercumexcess80[1:18] - losercumexcess80[1:18]),
sum(winnercumexcess80[1:24] - losercumexcess80[1:24]),
sum(winnercumexcess80[1:30] - losercumexcess80[1:30]),
sum(winnercumexcess80[1:36] - losercumexcess80[1:36]),
file="/home/jdtaylor/winlose80", ncolumns=8, append=T)
write(c(mean(avgreturns), mean(avgreturns[is95]), mean(avgreturns[is90]), mean(avgreturns[is80]),
mean(avgreturns[is80]), mean(avgreturns[is90]), mean(avgreturns[is95]),
mean(avgreturns[is95]), mean(avgreturns[is90]), mean(avgreturns[is95]),
mean(stdreturns[is90]), mean(stdreturns[is80]), mean(stdreturns[is80]), mean(betas),
mean(betas[is90]), mean(betas[is95]), mean(betas[is95]), mean(betas[is95]), mean(betas[is80]),
mean(betas[is80]), mean(betas[is80]), mean(betas[is90]), mean(betas[is90]),
file="/home/jdtaylor/descriptive_stats", ncolumns=21, append=T)
write(c(mean(betas[is100win]) - mean(betas[is100lose]),
mean(betas[is100win]) - mean(betas[is100lose])),
file="/home/jdtaylor/errors100", ncolumns=2, append=T)
write(c(mean(betas[is95win]) - mean(betas[is95lose]),
mean(betas[is95win]) - mean(betas[is95lose])),
file="/home/jdtaylor/errors95", ncolumns=2, append=T)
write(c(mean(betas[is90win]) - mean(betas[is90lose]),
mean(betas[is90win]) - mean(betas[is90lose])),
file="/home/jdtaylor/errors90", ncolumns=2, append=T)
write(c(mean(betas[is80win]) - mean(betas[is80lose]),
mean(betas[is80win]) - mean(betas[is80lose])),
file="/home/jdtaylor/errors80", ncolumns=2, append=T)


"whosurvives"<-
function(object, criteria)
{
current <- matrix(rep(T, 1000), nrow = 1)
for(i in 1:6) {
    current <- object[i, ] >=
        quantile(object[i, current], criteria) & current
}
current
}
A.2 Splus Code for Cross-Sectional Regressions

"famamcbeth"

function()
{
index <- matrix(rnorm(72, mean = .0071, sd = .057), ncol = 1)
betas <- rnorm(1000, mean = .82, sd = .79)
residuals <- matrix(rnorm(72000, mean = 0,  
                       sd = (.01*(2.55*(betas^2) - 1.25*betas+12.71))),  
                       ncol = 1000, byrow = TRUE)
returns <- .0031 + index*matrix(betas, nrow = 1) + residuals
betaestimates <- matrix(lsfit(index[1:36], returns[1:36,]$coef[2,], ncol = 1)
yearlycumreturns <- rbind(apply(returns[1:12,], 2, sum), apply(returns[13:24,], 2, sum),  
                          apply(returns[25:36,], 2, sum), apply(returns[37:48,], 2, sum),  
                          apply(returns[49:60,], 2, sum), apply(returns[61:72,], 2, sum))
is95 <- whosurvives(yearlycumreturns, .05)
is90 <- whosurvives(yearlycumreturns, .10)
is80 <- whosurvives(yearlycumreturns, .20)
a100 <- lsfit(betaestimates, t(returns[37:72,])$coef[2,])
b95 <- lsfit(betaestimates[is95], t(returns[37:72, is95])$coef[2,])
c90 <- lsfit(betaestimates[is90], t(returns[37:72, is90])$coef[2,])
d80 <- lsfit(betaestimates[is80], t(returns[37:72, is80])$coef[2,])
write(c(mean(a100)/(sqrt(var(a100))/sqrt(length(a100)) ),  
       mean(b95)/(sqrt(var(b95))/sqrt(length(b95)) ),  
       mean(c90)/(sqrt(var(c90))/sqrt(length(c90)) ),  
       mean(d80)/(sqrt(var(d80))/sqrt(length(d80)) ),  
       file = "/home/jdtaylor/averagetstat", ncolumns = 4, append = T) 
}

75
A.3 Derivation of the Total Variance of Stock and Mutual Fund Returns

Consider the model of asset returns given in equation (5):

\[ x_{it} = r_f + \beta_i(r_{mt} - r_f) + \epsilon_{it}. \]

We are interested in the total variance of stock and mutual fund returns implied by our model and BGIR (1992). In general, the total variance is given by:

\[ \text{var}(x) = E[\text{var}(x|\beta_i)] + \text{var}[E(x|\beta_i)] \]

We first obtain \( \text{var}[E(x|\beta_i)] \) in both models. For BGIR (1992),

\[
\begin{align*}
E(x|\beta_i) & = 0.0058 + 0.0036\beta_i \quad \text{which has variance} \\
\text{var}[E(x|\beta_i)] & = 0.0072^2 \text{var}(\beta_i) = 3.2 \times 10^{-6}
\end{align*}
\]

For our model, we have

\[
\begin{align*}
E(x|\beta_i) & = 0.0031 + 0.0032\beta_i \quad \text{which has variance} \\
\text{var}[E(x|\beta_i)] & = 0.0071^2 \text{var}(\beta_i) = 3.1 \times 10^{-5}
\end{align*}
\]

The remaining component of the total variance for BGIR (1992) is given by

\[
\begin{align*}
\text{var}(x|\beta_i) & = 0.0036\beta_i^2 + 0.0045(1 - \beta_i)^2 \quad \text{which has expectation} \\
E[\text{var}(x|\beta_i)] & = 0.081E(\beta_i^2) + 0.0045 - 0.0090E(\beta_i) \\
& = 0.038
\end{align*}
\]

For our model, we have

\[
\begin{align*}
\text{var}(x|\beta_i) & = 0.0032\beta_i^2 + \left\{0.01(2.55\beta_i^2 - 1.25\beta_i + 12.71)\right\}^2 \quad \text{which has expectation} \\
E[\text{var}(x|\beta_i)] & = 0.0007E(\beta_i^4) - 0.0006E(\beta_i^3) + 0.0098E(\beta_i^2) - 0.0032E(\beta_i) + 0.0162 \\
& = 0.0279
\end{align*}
\]

The total variances for BGIR (1992) and our model are 0.038 and 0.0279, respectively.
A.4 Conditional Expectation of a Truncated $N(0, \sigma^2)$ Random Variable

The conditional density of a $N(0, \sigma^2)$ random variable $x$ when $x \geq k$, is given by:

$$f_{x|x \geq k}(x|x \geq k) = \frac{1}{\sqrt{2\pi\sigma(1 - \Phi(\frac{k}{\sigma}))}} e^{\frac{-x^2}{2\sigma^2}}, \quad k \leq x \leq \infty, \quad k < \infty.$$

The conditional expectation of $x$ given that $x \geq k$ becomes

$$E(x|x \geq k) = \int_{x=k}^{\infty} x f_{x|x \geq k}(x|x \geq k) dx$$

which simplifies to

$$\frac{1}{\sqrt{2\pi\sigma(1 - \Phi(\frac{k}{\sigma}))}} \int_{\frac{k^2}{2\sigma^2}}^{\infty} e^{-u\sigma^2} du$$

with the substitutions $u = \frac{x^2}{2\sigma^2}$ and $du = \frac{x}{\sigma^2} dx$.

Integration yields the result

$$E(x|x \geq k) = \frac{\sigma}{\sqrt{2\pi(1 - \Phi(\frac{k}{\sigma}))}} e^{\frac{-k^2}{2\sigma^2}}$$
Appendix B

Code & Fund Information for Tournament Analysis

B.1 Splus Code for Tournament Analysis

"theorybreaker"<- function()
{
  stddev <- abs(rnorm(1, mean=0.01638756, sd=0.003427511))
  index <- rnorm(52, mean=0.00155985, sd=stddev)
  betas <- rnorm(660, mean=.95, sd=.25)
  residuals <- matrix(rnorm(34320, mean=0, sd=sqrt((.05349*(1-betas)^2)/52)),
                      ncol=660, byrow=TRUE)

  returns <- 0.001220396 + matrix(index,ncol=1)%*%matrix(betas,nrow=1) + residuals

  star <- apply(returns[1:30,],2,sum) > quantile(apply(returns[1:30,],2,sum),.9)

  delta1 <- sum(lsfit(index[1:30], returns[1:30,star])$coef[2,] >
                median(lsfit(index[1:30], returns[1:30,])$coef[2,])/sum(star) -
                sum(lsfit(index[1:30], returns[1:30,!star])$coef[2,] >
                median(lsfit(index[1:30], returns[1:30,])$coef[2,])/sum(!star))

  delta2 <- sum(lsfit(index[31:52], returns[31:52,star])$coef[2,] >
                median(lsfit(index[31:52], returns[31:52,])$coef[2,])/sum(star) -
                sum(lsfit(index[31:52], returns[31:52,!star])$coef[2,] >
                median(lsfit(index[31:52], returns[31:52,])$coef[2,])/sum(!star))

  gap <- median(apply(returns[1:30,star], 2, sum)) -
         median(apply(returns[1:30,!star], 2, sum))

  sigma <- sqrt(var(index[1:30]))

  premium <- mean(index[1:30])

  write(c(delta1, delta2, gap, premium, sigma), file="occam_data", ncolumns=5, append=T)
}
B.2 Lipper-Classified Growth Funds

“Growth funds normally invest in companies with long-term earnings expected to grow significantly faster than the earnings of the stocks represented in a major unmanaged stock index. These funds will normally have an above-average price-to-earnings ratio, price-to-book ratio, and three-year earnings growth figure.”
Lipper Analytical Services

1784FundsGrowthFund  FranklinCustGroAdv
1stSrcMngrmDvsmfdrtl  FranklinCustGroI
59Wall1StUSEquity     FranklinCustGroII
AALFundsCapitalGrA     FranklinEquityAdv
AALFundsCapitalGrB     FranklinEquityI
AARPCapitalGrowth      FranklinEquityII
AHAIInvDvdrEquity      FranklinStrB1Chp
AIMEqWngartenRtlA      FranklinStrCAGroI
AIMEqWngartenRtlB      FremontGrowth
AIMSummitFund          FrontierEquityFund
AIMGrowthA             FundmgGrowthFinI
AIMGrowthB             GAMAmericaCapitalA
AIMValueA              GEFundsPremGroEqA
AIMValueB              GEFundsPremGroEqB
AITVisionUSEquity      GEFundsPremGroEqC
AMCAPFund              GEFundsPremGroEqD
AMCOREVintageAggrGro   GMOCoreFundII
AMCOREVintageEquity    GMOCoreFundIII
APITrustYrktnClVal     GMOGrowthFundIII
AcademyValueFund       GMOTobacco-FrCoeIII
AccessorGrowth         GMUSsorFundsFundIII
AchievementEquityA     GTGlobalAmMc
AchievementEquityInst  GTGlobalAmMc
AdvantusHorizonA       GabelliAsset
AdvantusHorizonB       GabelliGrowth
AdvantusHorizonC       GalaxyEqyGroRtlA
AetnaGrowthAdv         GalaxyEqyGroRtlB
AetnaGrowthSel         GalaxyEqyGroTr
AlgerRetGrowth         GatewayCincinnatiFund
AlgerGrowthB           GintelFund
AllianceFundA          GlenmedeEquityFund
AllianceFundAdv        GlenmedeLargeCapValue
AllianceFundB          GlobaltGrowthFund
AllianceFundC          GoldmanEqCapeB
AlliancePortGrowthA    GoldmanEqCapeB
AlliancePortGrowthAdv  GoldmanEqCapeCore
AlliancePortGrowthB    GoldmanEqCapeCoreEquity
AlliancePortGrowthC    GoldmanEqCapeCoreEquityII
AlliancePremierGrA     GoldmanEqCapeCoreEquityInst
AlliancePremierGrAd    GoldmanEqCapeCoreEquityInst
AlliancePremierGrB     GradisonGrtEstabVal
AlliancePremierGrC     GranumValueFund
OneGroupLgCoGroB       OppenheimerGrowthA
OneGroupLgCoGroFid     OppenheimerGrowthB
OppenheimerGrowthC     OppenheimerGrowthY
OppenheimerDiscValA    OppenheimerDiscValB
OppenheimerDiscValC    OppenheimerDiscValY
OppenmrQuestcvp1A     OppenmrQuestcvp1B
OppenmrQuestcvp1C     OsterweisFund
PBHGCesGrowthPBHG      PBHGLargeCap20PBHG
PBHGLargeCapGroPBHG    PICGrowth
PICIppicPinnacleGrowth PIMCOCoreEquityAdmn
PIMCOCoreEquityInst   PIMCOGrowthA
PIMCOGrowthB          PIMCOGrowthC
PacCapGrowthStkInst   PacCapGrowthStkRet
PacHnmBluechipA       PainewbrGrowthA
PainewbrGrowthB       PainewbrGrowthC
PainewbrPaceLrgco     PappAmerica-Abroad
ParkstoneAggrAllInst  ParkstoneLrgCapA
ParkstoneLrgCapB      ParkstoneLrgCapC
ParkstoneLrgCapInst   ParnassusFund
PasadenaGrowthA       PasadenaGrowthB
PasadenaGrowthC       PasadenaNiftyFiftyA

79
CommonSenseGroB
CompassLgCpGrInst
CompassLgCpGrInvA
CompassLgCpGrInvB
CompassLgCpGrInvC
CompassLgCpGrSvc
CompositeNorthwestA
CompositeNorthwestB
CorefundGrowthEqA
CorefundGrowthEqY
CrabbeHusonEquityInst
CrabbeHusonEquityPrm
CrestfedsCapAppInvA
CrestfedsCapAppTr
Croft-LeominsterValue
CrowleyGrowth&Income
DGIvEquity
DWRetSrSamsrValue
DWRetSrSampGrowth
DavidLBabsonGrowth
DavisNYVentureA
DavisNYVentureB
DavisNYVentureC
DavisNYVentureY
DavisGrowthOpptyA
DavisGrowthOpptyB
DeanWitterAmerValB
DeanWitterCapGroB
DeanWitterMktLeadB
DelafielhFund
DelawareGrAggrGroA
DelawareGrAggrGroB
DelawareGrAggrGroC
DelawareGrGroStkA
DelawareGrGroStkB
DelawareGrGroStkC
DelawareQuantumA
DelawareQuantumB
DelawareQuantumC
DelawareQuantumInst
DelawareUSGrowthA
DelawareUSGrowthB
DelawareUSGrowthC
DelawareUSGrowthInst
DepositorsBoston
DiversificationFund
DominioSocialEquity
DominionInsightGrowth
DreyfusAppreciationFd
DreyfusCoreValueInst
DreyfusCoreValueInv
DreyfusCoreValueR
DreyfusGroOpportunity
DreyfusLargeCoGrowth
MFSGrowthOpptyA
MFSGrowthOpptyB
MFSInstlResearch
MFSMassInvestGroA
MFSMassInvestGroB
MFSResearchA
MFSResearchB
MFSResearchC
MFSResearchI
MFSStrategicGrowthA
MFSStrategicGrowthB
MFSStrategicGrowthC
MFSStrategicGrowthI
MFSUnionStdEquity
Mairs&PowerGrowth
Manning&NapierTaxmgd
MarketvestEquity
MarquisGroEqtyA
MarquisGroEqtyB
MasonStrGrowthStkA
MasonStrGrowthStkB
MastersSelInvSelEqty
MasterworksGrowthStk
MathersFund
MatrixGrowthFund
MatterhornGrowth
Matthew25Fund
MentorCapGrothA
MentorCapGrothBl
MeridianValueFund
MerrillAsstGroppA
MerrillAsstGroppB
MerrillAsstGroppC
MerrillAsstGroppD
MerrillFdtomorrowA
MerrillFdtomorrowB
MerrillFdtomorrowC
MerrillFdtomorrowD
MerrillFdmntlGroA
MerrillFdmntlGroB
MerrillFdmntlGroC
MerrillFdmntlGroD
MerrillGroFndA
MerrillGroFndB
MerrillGroFndC
MerrillGroFndD
MerrimanCapApprec
MinervaEquity
MnstryInstlGroEqInst
MnstryInstlGroEqServ
MonettaLargeCapEquity
MonitorGrowthInv
MonitorGrowthTr
Montag&CaldwellGroI
SchwabCapAnalytics
ScoutStockFund
ScudderClassicGrowth
ScudderLargeCoGrowth
ScudderLargeCoValue
ScudderPathwayGrowth
ScudderValueFund
SeafirstBlueChipFund
SecondFidExchange
SecurityEqEquityA
SecurityEqEquityB
SecurityEqSocAwarA
SecurityEqSocAwarB
SeftonFdsTrEqtyVal
SeligmanGrothA
SeligmanGrothB
SeligmanGrothD
SentinelGrowthA
SentryFund
SequioaFund
SextantGrowth
SierraAssetCapGroA
SierraAssetCapGroB
SierraGrothA
SierraGrothB
SierraGroth1
SierraGrothS
SmBarneyApprecA
SmBarneyApprecB
SmBarneyApprecC
SmBarneyFdmntlValA
SmBarneyFdmntlValB
SmBarneyFdmntlValC
SmBarneyInvMgdGrA
SmBarneyInvMgdGrB
SmBarneyInvMgdGrC
SoundShore
SouthtrustVulcanStock
StagecoachGroIncA
StagecoachGroIncB
StagecoachGro&IncInst
StandyerWoodEquity
StarGrothEquity
StarRelativeValue
StateFrmGroth
SteadmanInvestment
SteadmanTech&Growth
SteinRoeAdvGrowth
SteinRoeAdvSpecial
SteinRoeAdvYoung
SteinRoeGrothStock
SteinRoeSpecialF
SteinRoeYoungInvestor
StratusCapApprec
Appendix C

Splus and C Code for Estimation and Bootstrap


"alphacalc" <-
function(filename)
{

X <- scan(filename, what = numeric())

X <- (X - mean(X, 0.25))/((quantile(X, 0.72) - quantile(X, 0.28))/
1.654)

counter <- length(X)

fama <- ((quantile(X, 0.96) - quantile(X, 0.04)) * 0.827)/(quantile( X, 0.72) - quantile(X, 0.28))

fama <- -259.9068 + 0.01830014 * fama^7 - 0.5446718 * fama^6 + 6.748905 *
    fama^5 - 45.23456 * fama^4 + 177.3666 * fama^3 - 407.0392 *
    fama^2 + 505.6875 * fama

if(fama < 1.1)
    fama <- 1
if(fama > 2)
    fama <- 2

famachart <- round(fama, digits = 1)
counterchart <- round(counter, digits = -2)
if(famachart == 2)
    famachart <- 1.9
if(counterchart > 1600)
    counterchart <- 1600

85
if(counterchart < 200)
  counterchart <- 200
K <- chart$z[(10 * (famachart - 1)) + 1, (counterchart/100) - 1]
K <- round(K, digits = 0)

  t <- matrix((3.14159265 * c(1:K))/25, ncol = 1)
alpha <- log(t)
phi <- apply(exp((t * (1i)) %% t.default(X)), 1, mean)
phi <- log(- log(Mod(phi)^2))
kout <- lsfit(alpha, phi, intercept = T)$coef[2]
c(8431 %% counter, kout, fama, counter, K)
}

"bootstrap"<-
function()
{
  unix("blocker", rawfile, output = F)
  unix("datamaker", paste(rawfile, "S", sep = ""), output = F)
  OUT <- t.default(apply(filelist, 1, alphacalc))
  write(t.default(OUT), paste(rawfile, "S_results", sep = ""),
        ncol = 5, T)
}

C.2.1 C Code to Resample Random Blocks of 50 ("blocker")

#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <time.h>
define MAX 26000
define BLOCK_SIZE 50

/* Function definitions */
int data_points(double *, char *);
void block(double *, double *, int, int);
void sample_out(double *, int, char *);

int main()
{
  /* Variable and Array Definitions */
double x[MAX];
double *y;
char file_n[30];
int sample_size;

  /* Main Program Logic Begins Here */
  printf("please enter name of file\n");
  scanf("%s", file_n);

  sample_size = data_points(x, file_n);
y = calloc(sample_size, sizeof(double));

  block(x, y, sample_size, (int) (time(0) - 84530400));
  sample_out(y, (sample_size/BLOCK_SIZE)*BLOCK_SIZE, file_n);
  return(0);
}

/* Function reads sample data from a file into an array */

int data_points(double *x, char *file_n)
{  
  /* Define variables for this function */

  FILE *fpi;
  int i, test;

  /* Open the file defined by the user */

  fpi=fopen(file_n, "r");
  if (fpi == NULL)
  {
    printf("error in opening %s", file_n);
    exit(-1);
  }

  /* Until an EOF is reached, read data points into an array of size MAX */

  for (i=0, test=0; (test != EOF) && (i < MAX); i++)
    test = fscanf(fpi, "%lf", &x[i]);

  return(i-1);
}

void block(double *x, double *y, int n, int seed)  
{  
  int i=0, j=0, l=0;
  srand(seed);

  while (i < n/BLOCK_SIZE)
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <time.h>
#define MAX 26000
#define BLOCK_SIZE 50

  /* Function definitions */
  int data_points(double *, char *);
  void block(double *, double *, int, int);
  void sample_out(double *, int, char *);

  int main()
  {
  /* Variable and Array Definitions */
  double x[MAX];
  double *y;
  char file_n[30];
  int sample_size;

  /* Main Program Logic Begins Here */
  printf("please enter name of file\n");
  scanf("%s", file_n);

  sample_size = data_points(x, file_n);
y = calloc(sample_size, sizeof(double));

block(x, y, sample_size, (int)(time(0) - 84530400));

sample_out(y, (sample_size/BLOCK_SIZE)*BLOCK_SIZE, file_n);
return(0);
}

/* Function reads sample data from a file into an array */

int data_points(double *x, char *file_n)
{
    /* Define variables for this function */

    FILE *fpi;
    int i, test;

    /* Open the file defined by the user */

    fpi=fopen(file_n, "r");
    if (fpi == NULL)
    {
        printf("error in opening \%s", file_n);
        exit(-1);
    }

    /* Until an EOF is reached, read data points into an array of size MAX */

    for (i=0, test=0; (test != EOF) && (i < MAX); i++)
    {
        test = fscanf(fpi, "%lf", &x[i]);
    }

    return(i-1);
}

void block(double *x, double *y, int n, int seed)
{
    int i=0, j=0, l=0;
    srand(seed);

    while (i < n/BLOCK_SIZE)
    {
        l = (int)((double)rand()/RAND_MAX * (double)((n-1) - (BLOCK_SIZE-1)));
        for (j=i*BLOCK_SIZE; j< i*BLOCK_SIZE + BLOCK_SIZE; j++)
        {
            y[j] = x[l + j - (i*BLOCK_SIZE)];
        }
        i++;
    }
}

void sample_out(double *y, int sample_size, char *file)
{
    /* Define variables for this function */

    /* Sample data... */
}
FILE *fpi;
int i, test;
char s[30];
char p;

i = 0; p = file[0];
while (p != '0') {s[i] = p; i++; p = file[i];}
s[i] = 'S';
s[++i] = p;

fpi=fopen(s, "w");
if (fpi == NULL)
{
printf("error in opening %s", s);
exit(-1);
}

/*@ Until end of data, writes data points to file */

for (i=0, test=0; (test != EOF) && (i < sample_size); i++)
test = fprintf(fpi, "%f\n", y[i]);

printf("%d data points written to %s\n", sample_size, s);
}

C.2.2  C Code to Form 50 Non-Overlapping Series’ ("datamaker")

#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#define MAX 260000

/*@ Function definitions */
int data_points(double *, char *);
void convert(double *, int);
int differencing(double *, double *, int, int);
void print_out(double *, int, int, char *);

int main()
{

/*@ Variable and Array Definitions */

double x[MAX];
char file_n[30];
double *y;
int n, sample_size;
int i=1;

/*@ Main Program Logic Begins Here */
printf("please enter name of file\n");
scanf("%s", file_n);

n=data_points(x, file_n);
convert(x, n);
while (i <= 50)
{
    y = calloc(n, sizeof(double));
    sample_size = differencing(x,y,n,i);
    print_out(y, sample_size,i, file_n);
    free(y);
    i++;
}

return(0);
}

/* *************** Functions Follow *************** */
/* Function reads sample data from a file into an array */

int data_points(double *x, char *file_n)
{
    /* Define variables for this function */

    FILE *fpi;
    int i, test;

    /* Open the file defined by the user */

    fpi=fopen(file_n, "r");
    if (fpi == NULL)
    {
        printf("error in opening %s", file_n);
        exit(-1);
    }

    /* Until an EOF is reached, read data points into an array of size MAX */
    #include <stdio.h>
    #include <math.h>
    #include <stdlib.h>
    #define MAX 26000

    /* Function definitions */
    int data_points(double *, char *);
    void convert(double *, int);
    int differencing(double *, double *, int, int);
    void print_out(double *, double *, int, int, char *);

    int main()
    {

        91
/* Variable and Array Definitions */

double x[MAX];
char file_n[30];
double *y;
int n, sample_size;
int i=1;

/* Main Program Logic Begins Here */

printf("please enter name of file\n");
scanf("%s", file_n);

n=data_points(x, file_n);

convert(x, n);
while (i <= 50)
{
  y = calloc(n, sizeof(double));
  sample_size = differencing(x, y, n, i);
  print_out(y, sample_size, i, file_n);
  free(y);
  i++;
}

return(0);
}

/* ******************* Functions Follow ******************* */
/* Function reads sample data from a file into an array */

int data_points(double *x, char *file_n)
{
/* Define variables for this function */

  FILE *fpi;
  int i, test;

  /* Open the file defined by the user */

  fpi=fopen(file_n, "r");
  if (fpi == NULL)
  {
    printf("error in opening %s", file_n);
    exit(-1);
  }

  /* Until an EOF is reached, read data points into an array of size MAX */

  for (i=0, test=0; (test != EOF) && (i < MAX); i++)
    test = fscanf(fpi, "%lf", &x[i]);
return(i-1);
}

/* Function takes the raw data and converts it to log differences */
void convert(double *x, int n)
{
    int i=0;
    double temp=0.;
    while (i < n)
    {
        temp = x[i];
        x[i] = log(temp + 1);
        i++;
    }
}

/* Function takes the log differences & the desired differencing interval,
   creates another array & returns its length */
int differencing(double *x, double *y, int n, int diff)
{
    /* int diff; */
    double sum=0.;
    int i=0;
    int j=0;
    int k=0;

    /* printf("enter desired differencing interval\n"); */
    /* scanf("%d", &diff); */
    n = n / diff;
    while (i < n)
    {
        while (k < diff)
        {
            sum = x[j + k] + sum;
            k++;
        }
        j = j + diff;
        y[i] = sum;
        k = 0;
        sum = 0.;
        i++;
    }
    return(n);
}

/* Function prints the properly differentiated array in Splus-readable format */
void print_out(double *y, int sample_size, int diff, char *file)
{
    /* Define variables for this function */

    FILE *fpi;
    int i, test;
    /* char file[30]; */
char s[30];
char p,t,o;

/* Open the file defined by the user */

/* printf("enter desired output filename\n"); */
/* scanf("%s", &file[0]); */

int t = '0';
o = '0';

if (diff%10 == 0) {o = '0';}
if (diff%10 == 1) {o = '1';}
if (diff%10 == 2) {o = '2';}
if (diff%10 == 3) {o = '3';}
if (diff%10 == 4) {o = '4';}
if (diff%10 == 5) {o = '5';}
if (diff%10 == 6) {o = '6';}
if (diff%10 == 7) {o = '7';}
if (diff%10 == 8) {o = '8';}
if (diff%10 == 9) {o = '9';}

if (diff/10 == 0) {t = '0';}
if (diff/10 == 1) {t = '1';}
if (diff/10 == 2) {t = '2';}
if (diff/10 == 3) {t = '3';}
if (diff/10 == 4) {t = '4';}
if (diff/10 == 5) {t = '5';}

i=0; p=file[0];
while (p != ' \0') {s[i] = p; i++; p = file[i];}
s[i] = t;
s[++i] = o;
s[++i] = p;

fp1=fopen(s, "w");
if (fp1 == NULL)
{
    printf("error in opening %s", s);
    exit(-1);
}

/* Until end of data, writes data points to file */

for (i=0, test=0; (test ! = EOF) && (i < sample_size); i++)
    test = fprintf(fp1, "%f\n", y[i]);

printf("%d data points written to %s\n", sample_size, s);
Bibliography


