

A Method of Pricing European Style Equity Options

by

David Harris

Dissertation submitted to the College of Business  
at West Virginia University  
in partial fulfillment of the requirements  
for the degree of

Doctor of Philosophy  
in  
Economics

Approved by

Ashok Abbott, Committee Chairperson  
Stratford Douglas  
Santiago Pinto  
Gerald Hobbs  
Harumi Hattori

Department of Economics

Morgantown, West Virginia  
2014

Keywords: Option Pricing, Mean Variance Finance, Cauchy Distribution

Copyright 2014 by David Harris

UMI Number: 3672868

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3672868

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

# ABSTRACT

A Method of Pricing European Style Equity Options

by: David Harris

The study of option pricing has a very short history, when compared with other elements of economics. Since the publication of a method to price European style equity options by Fischer Black and Myron Scholes in 1973 a vast amount of research on option pricing has occurred. Ultimately, the pricing of equity options depends upon the match of the model and reality. A new method to price option contracts is proposed. It is argued that the distributional assumptions of standard models are uncorrelated with nature. A new model is proposed as a start to a new class of models.

# TABLE OF CONTENTS

<b>I Front Matter</b>	<b>ix</b>
List of Figures	x
List of Tables	xi
List of Symbols	xii
Acknowledgments	xiii
<b>II Prior Theory</b>	<b>1</b>
<b>Chapter 1</b>	
<b>Introduction</b>	<b>2</b>
1.1 Reworking The Math . . . . .	4
1.2 The Background for Mean-Variance Finance . . . . .	6
1.2.1 Partial Examples . . . . .	9
1.2.2 Example–Individual Asset Allocation With Risky Assets . . . . .	10
1.2.3 First Order Conditions for an Optimum . . . . .	11
1.2.4 Moving Toward a Solution . . . . .	11
1.2.5 Example–A Continuous Time Single Asset Model . . . . .	12
1.2.6 Example–Arbitrage Pricing Theory . . . . .	13
1.2.7 Commonalities . . . . .	14
1.3 Post Markowitz and Roy . . . . .	15
1.3.1 Static/Discrete Models Post Markowitz and Roy . . . . .	15
1.3.2 Continuous Time Models Up To Black-Scholes . . . . .	15
1.3.3 Empirical Criticism . . . . .	16
1.4 Market Makers . . . . .	18
1.4.1 Diamond and Dybvig . . . . .	19
1.4.2 Options as a Risk Management Tool . . . . .	23

1.4.3	Amihud and Mendelson . . . . .	25
1.4.3.1	Liquidity During Runs . . . . .	27
1.5	Existential Risks and Cash Flows . . . . .	28
1.5.1	Estimating Existential Risks . . . . .	30
1.5.1.1	Bankruptcy . . . . .	31
1.5.1.2	Mergers . . . . .	37
1.6	Conclusion . . . . .	42

### **III Background Statistics 43**

#### **Chapter 2**

	<b>The Use of Bayesian Statistics</b>	<b>44</b>
2.1	Bayesian Versus Frequency Based Models . . . . .	45
2.1.1	Illustration . . . . .	47
2.2	Learning Through Bayesian Thinking . . . . .	49
2.2.1	Subjectivity . . . . .	50
2.2.2	Decisions by Economic Actors . . . . .	52
2.2.3	Dealing with Prior Information . . . . .	52
2.2.3.1	Introduction . . . . .	52
2.2.3.2	The Uninformative Prior . . . . .	55
2.2.3.3	The Vague Prior Distribution . . . . .	56
2.2.3.4	Informative Prior Distributions . . . . .	57
2.2.3.5	The Highly Informative Prior . . . . .	60
2.2.3.6	The Adversarial Prior . . . . .	61
2.3	A Sidestep Into Bayes Actions . . . . .	62
2.3.1	The Effect of Prior Information . . . . .	66
2.3.2	Impact on Option Pricing . . . . .	67
2.4	The Impact of Sufficiency on Prediction . . . . .	67
2.4.1	Preliminaries . . . . .	68
2.4.2	Sufficient Statistics and Information . . . . .	71
2.4.2.1	Inference . . . . .	71
2.4.2.2	Prediction . . . . .	73
2.5	The Cauchy Distribution and Its Transformations . . . . .	74
2.5.1	The Symmetric Case . . . . .	75
2.5.2	The Truncated Case . . . . .	76
2.5.3	Logarithmic Transformation . . . . .	76
2.5.4	Bayesian Methods in General . . . . .	77
2.5.5	Frequentist Methods . . . . .	78
2.5.6	Conclusion . . . . .	79

<b>Chapter 3</b>	
<b>Choosing a Likelihood Function</b>	<b>81</b>
3.1 Model Assumptions . . . . .	81
3.1.1 Difference Equations . . . . .	82
3.1.2 Böhm-Bawerk Theory . . . . .	83
3.1.3 How Bayesian and Frequentist Paradigms Affect the Equation	84
3.1.4 Scientific Modelling . . . . .	84
3.1.5 Boundary Conditions . . . . .	85
3.2 Returns . . . . .	85
3.2.1 Intuition Behind the Proof . . . . .	86
3.2.2 The Frequency Based Solution . . . . .	87
3.2.3 Impact of White's Frequentist Proof on the Bayesian Like- likelihood Function . . . . .	88
3.3 Effect on Current Theory . . . . .	89
3.3.1 Mean-Variance Finance . . . . .	89
3.3.1.1 Bayesian Interpretation . . . . .	90
3.3.1.2 The Frequentist Interpretation . . . . .	91
3.3.2 Heavy-Tailed and Econophysics Methods . . . . .	94
3.3.3 Regnault and Bachelier . . . . .	100
3.4 Special Cases—Mergers and Bankruptcy . . . . .	101
3.4.1 Bankruptcy . . . . .	101
3.4.2 Mergers . . . . .	101
3.4.2.1 Focusing on the Knowable . . . . .	102
3.4.2.2 Stock for Stock Mergers . . . . .	103
3.4.2.3 Cash for Stock Mergers . . . . .	105
3.5 Conclusion . . . . .	108
<b>IV Model Validation</b>	<b>109</b>

<b>Chapter 4</b>	
<b>Empirical Model Selection and Validation</b>	<b>110</b>
4.1 Bayesian Hypothesis Testing . . . . .	110
4.2 Bayes factor . . . . .	113
4.2.1 Composite vs. Simple Hypothesis . . . . .	115
4.2.1.1 Infinite Sets . . . . .	115
4.2.1.2 Formal Approach to Hypothesis Testing . . . . .	116
4.2.1.3 Composite Hypothesis . . . . .	117
4.2.1.4 Simple Hypothesis . . . . .	121
4.2.1.5 Properties of Maximum Likelihood Estimator Based Ratios . . . . .	122
4.3 Implementation . . . . .	124
4.3.1 The Data . . . . .	124

4.3.2	Empirical Results . . . . .	124
4.3.2.1	The Likelihood Function and the Posterior Distribution of $\sigma$ With a Normal Likelihood . . . . .	124
4.3.2.2	The Likelihood Function and the Posterior Distribution of $\sigma$ With a Log-Normal Likelihood . . . . .	127
4.4	Discussion . . . . .	129

## V Option Pricing 130

### Chapter 5

	<b>Pricing European Style Options</b>	<b>131</b>
5.1	Introduction . . . . .	131
5.2	The Game . . . . .	134
5.3	Notation . . . . .	134
5.4	Actors . . . . .	135
5.5	Actions . . . . .	135
5.6	Utility Functions . . . . .	137
5.7	Assumptions . . . . .	138
5.8	Nature of the Contracts . . . . .	140
5.9	The Profit Function for Short Puts . . . . .	141
5.9.1	The Critical Importance of the Short Put . . . . .	141
5.9.2	The Profit Function . . . . .	142
5.9.2.1	Parametric Form for Going Concern . . . . .	148
5.9.2.2	Conclusion . . . . .	149
5.10	Buying Put Contracts . . . . .	150
5.10.1	Brief Discussion . . . . .	151
5.11	No Arbitrage Equilibrium . . . . .	152
5.12	Call Options Under Various Initial Endowments . . . . .	153
5.12.1	Large Cash Endowment . . . . .	153
5.12.1.1	Derivatives of Potential Arbitrage Positions . . . . .	155
5.12.1.2	Equilibrium Pricing In Long Positions . . . . .	156
5.12.1.3	Equilibrium In Short Positions . . . . .	159
5.12.2	No Endowment . . . . .	160
5.12.3	Large Endowment of Debt . . . . .	161
5.12.4	Small Endowment of Cash . . . . .	162
5.12.5	Small Endowment of Debt . . . . .	163
5.13	Dominant Pricing . . . . .	163
5.13.1	Put Contracts . . . . .	163
5.13.2	Long Call Contracts . . . . .	164
5.13.3	Short Call Contracts . . . . .	164
5.14	Speculators . . . . .	165
5.15	Expected Profit as a Negative Bayesian Risk Function . . . . .	165

5.16 Conclusion . . . . .	167
---------------------------	-----

## Chapter 6

<b>Empirical Testing of The Proposed Option Pricing Model</b>	<b>169</b>
6.1 Introduction . . . . .	169
6.2 Bankruptcy Estimation . . . . .	171
6.2.1 Bankruptcy Data from 1925-1961 . . . . .	171
6.2.2 Bankruptcy Data from January 1962-June 1990 . . . . .	175
6.3 Merger Estimation . . . . .	180
6.3.1 Merger Data Prior to 1962 . . . . .	181
6.3.2 Merger Data 1962-June 30, 1990 . . . . .	182
6.3.2.1 Merger Risk January 1, 1962-June 30, 1990 With Missing Accounting Data . . . . .	182
6.3.2.2 Financials Versus Non-Financials . . . . .	182
6.3.2.3 Other Factors . . . . .	182
6.4 Prior Distribution of Rewards of Investing Given A Merger Will Happen . . . . .	186
6.5 Prior Distribution of Rewards from December 31, 1925-June 30, 1990 for Going Concerns . . . . .	194
6.6 Testing the Properties of the Bankruptcy Estimators . . . . .	195
6.6.1 Problem Background . . . . .	195
6.6.2 Uninformative Alternatives . . . . .	198
6.6.3 Posterior Tests . . . . .	198
6.6.3.1 Results of Posterior Tests for Bankruptcy Estimator	202
6.6.3.2 Results of Posterior Tests for the Merger Estimator	202
6.6.3.3 Results of Posterior Tests for Expected Loss Given a Merger . . . . .	203
6.6.3.4 Results of Posterior Tests of Expected Loss Given Survival . . . . .	204
6.7 Conclusion . . . . .	204

## VI Conclusions and Directions Forward 206

### Chapter 7

<b>Conclusion</b>	<b>207</b>
7.1 Mathematical Observations . . . . .	208
7.1.1 Model principles . . . . .	208
7.2 Quarternions . . . . .	209
7.3 Observations from the Data . . . . .	209
7.3.1 $\mu = 0\%$ . . . . .	209
7.3.2 $\sigma$ May Be Increasing . . . . .	210
7.4 Policy Implications . . . . .	212

7.4.1	Moving Forward . . . . .	212
7.5	Implications for Other Fields of Economics . . . . .	213
7.6	Finishing Up . . . . .	214
<b>Appendix A</b>		
	<b>Distribution of Returns Given a Merger Will Happen</b>	<b>215</b>
A.1	Preliminaries . . . . .	215
A.2	Mergers . . . . .	218
<b>Appendix B</b>		
	<b>Software Used in the Implementation of the Research</b>	<b>220</b>
B.1	Chapter 5–Comparing Competing Models . . . . .	220
B.2	Chapter 6–Bankruptcies From 1925-1961 . . . . .	227
B.3	Chapter 6–Model Selection for Bankruptcy Prior Distribution 1961-1990 . . . . .	232
	B.3.1 GDP Calculation . . . . .	232
	B.3.2 Model Selection . . . . .	244
	B.3.3 Prior Probability of Bankruptcy . . . . .	268
B.4	Prior Probability of Merger . . . . .	281
B.5	Prior Probability for Returns Given a Merger . . . . .	298
B.6	Prior Probability for $\sigma$ for a Going Concern . . . . .	307
B.7	Posterior Probability for Bankruptcy Estimator . . . . .	319
B.8	Posterior Test for Bankruptcy . . . . .	350
B.9	Posterior Test for Merger . . . . .	380
B.10	Posterior Test for Loss Given Merger . . . . .	401
B.11	Posterior Test for Loss Given Going Concern . . . . .	411
	<b>Bibliography</b>	<b>420</b>

# Part I

## Front Matter

# LIST OF FIGURES

1.1	Possible path of orders . . . . .	26
3.1	One Year Returns, Measured in Basis Points, in Stock-forStock Mergers 1925-2008 . . . . .	105
4.1	Marginal posterior distribution for the Cauchy model for $\sigma$ , measured in basis points. . . . .	126
4.2	Marginal posterior distribution for the Normal model for $\sigma$ , measured in basis points. . . . .	126
4.3	Log-Likelihood for $\sigma$ for the Log Normal Distribution . . . . .	128
6.1	Probability Density Function for the Prior Probability of Bankruptcy. . . . .	174
6.2	Prior Distribution of Returns Given a Merger . . . . .	193
6.3	Prior Distribution for Mergers from 0 to 4 . . . . .	194
6.4	Prior Distribution of $\sigma$ for Going Concerns . . . . .	195
6.5	Average Daily Loss, Given a Loss Did Occur Versus Daily Prediction of Loss Given a Flat (Neutral) Prior . . . . .	205
7.1	Drift of Sigma by Date . . . . .	211

# LIST OF TABLES

4.1	Table of Maximum Likelihood . . . . .	125
4.2	Interval and Point Estimates of the Shape Parameter of the Competing Models, Measured in Basis Points . . . . .	127
5.1	Profit Functions Given Sufficiently Large Cash Endowments . . . . .	154
5.2	Derivatives . . . . .	156
5.3	Profit Functions Given Sufficiently Large Cash Endowments . . . . .	161
5.4	Profit Functions Given Sufficiently Large Cash Endowments . . . . .	162
5.5	Endowment Specific Equilibrium Pricing for Call Options . . . . .	163
6.1	List of Ratio Classes with Members of the Classes . . . . .	176
6.2	Bankruptcy Prior for Current Ratio and Rec. Conversion Period . . . . .	178
6.3	Bankruptcy Prior for Net Margin and ROA . . . . .	180
6.4	Bankruptcy Prior When Missing Accounting Data . . . . .	180
6.5	Prior Probability of Merger Risk . . . . .	186
6.6	Table of Frequencies of Returns Given a Merger . . . . .	192

# LIST OF SYMBOLS

Significant abuse of notation occurs in this dissertation. There are two practical reasons for this. The first is that the dissertation has over two hundred equations and inequalities. They cover diverse range of fields that are normally conceptualized as distinct from one another.

This presents two problems. The first is the exhaustion of the English and Greek alphabet. The second is that in many fields letters have specific meanings. Should decision rules be denoted as  $\delta$ , or should dividends be denoted  $\delta$ . Should prices be denoted as  $p$  or should probabilities?

Symbols are defined locally and in the context of their usage. This avoids moving into the Cyrillic alphabet and permits the use of symbols in their traditional context.

# ACKNOWLEDGMENTS

I need to thank Dr. Ashok Abbott for his tremendous patience and guidance. At the beginning of the process I failed to appreciate the importance of liquidity costs, at the end I wish I had had the time to do far more work on liquidity costs.

# DEDICATION

This work is dedicated to my wife Lenore, who has kept a family in clean clothes, with warm food and lots of love. She has kept relations with friends up, made sure the mountain of little things in life were completed, and made me come to bed at reasonable hours instead of write this dissertation. She has kept the family sane while this document worked its way from drafts to completion..

In that time three children have gone from childhood to adulthood and a fourth is trying to get there. Without her this document would not have been possible.

# Part II

## Prior Theory

# INTRODUCTION

Since 1952 economics and finance have been driven by mean-variance finance, either directly due to the model being taught and used in application, regulation and law, or by driving alternative methods due to observed anomalies that would imply the various models of mean-variance finance fail to explain significant phenomena. An article, and subsequent articles following it, by Benoit Mandelbrot in 1963 should have brought an end to the discussion by empirical falsification, but did not.[60]

Mandelbrot's article was built around the premise that returns on investment followed a Cauchy distribution. As the Cauchy distribution has no mean, and as a consequence no variance, mean-variance finance should have stopped there; subject of course to further study of the distributions involved. Unfortunately a number of things prevented this.

The first difficulty was the choice of hypothesis testing method. Certain hypothesis testing methods, given the sheer quantity of data, are problematic in this setting. The second is closely related, which is the tradition of testing log returns rather than actual returns. This method comes from a time period when computational limitations made this practical. The logarithmic transformation of the data alters the mathematical properties present in ways that most would find unexpected. The third reason was the perception that no first principles reason for Mandelbrot to be correct existed, but there appeared to be a first principles reason for mean-variance finance to be correct. Indeed, two Nobel Prizes were awarded

despite known empirical difficulties. Fourth, the models were perceived to have assumptions that were too simple to model the world. This led researcher to attack the symptoms rather than the core elements. Fifth, the models made simple predictions using very well understood technologies. Ordinary least squares dominated the methodology, or variants on it, and as this is computationally simple and quick a natural resistance to the use of other technologies formed. Sixth, the methods of mean-variance finance allowed economists to drop utility from problems. Utility and more generally preferences can allow Keynes “animal spirits,” to run amok in economic models. Despite Pareto’s indifference curve, the utility function is a pointless added difficulty if it can be avoided.

Of course, the needs of empiricism can force science to face unpleasant choices. There is always a struggle between Occam’s Razor and accuracy. Models lacking a variance, and more importantly, a covariance matrix leave economists back in the pre-’50’s world with a requirement that everything be begun again from scratch. That is not trivial. Box and Draper point out,

that all models are wrong; the practical question is how wrong do they have to be to not be useful.

The answer to the question provided here is mean-variance finance carries no information about the world when understood in a Bayesian framework and results in a non-measurable set when a non-Bayesian mechanism is used. It will be shown that mean-variance finance results in a mathematical contradiction when understood in a Bayesian framework and a near complete loss of information when a non-Bayesian mechanism is used. Hence, not only is the Black-Scholes Option Pricing Model not usable as a model, it must be replaced. The beginning of the process of replacing option pricing models is contained in this work. This work does *not* provide *the* correct method to price options. It provides *a* method that can be used to price European style equity options. Definite empirical improvements are available and will be discussed.

The intuition that economics must go backward to go forward is correct, but that is not a bad thing at all. Luminiferous æther was unknowingly a step backward, or at least sideways, but physics has since not only survived but thrived. The Higgs Boson, light that becomes a solid through quantum entanglement, landing

on the moon and magnetic resonance spectroscopy show a vibrant field. Economics hasn't changed a lot since Samuelson. Tomorrow is a vibrant new day, one without a mean or a variance. Incredible natural variation is now open for economists to capture in wonderful models. Even though economics is less esoteric than physics, its problems are no less important.

## 1.1 Reworking The Math

If it is assumed that the empirical difficulties with mean-variance finance imply something is wrong in the underlying models, then it is tempting to look again at the math. If one goes back to the original articles and asks what would happen if different choices were made, then new models can be constructed. The question then regards what math to use.

The math used in this dissertation is somewhat of a reworking of the path taken by mean-variance finance. As such, it is important to understand the methods and goals of mean-variance finance, at least in generalizations. At the same time, other topics that have previously been secondary issues in economics and finance suddenly move to the forefront. In particular, liquidity costs assure the laws of probability hold for option markets where market makers permit private parties to underwrite the short position in option contracts.[76] This is a rather surprising observation. Although it is possible to assume the laws of probability into existence, the presence of an at-risk, profit-maximizing, market maker guarantees they hold.

Either a monopolist market maker or a regulated competitive market maker nearly assures an equilibrium even in the presence of risk loving actors.

Liquidity costs are not the only required but generally ignored element of option pricing. Bankruptcy is necessary for option pricing to exist in equity markets. Without the possibility of bankruptcy, such as through legalizing intergenerational slavery and indentured servitude to pay debts, there is no left boundary for the statistical distribution and as such no expected profit.<sup>1</sup> Unexpectedly, the presence of bankruptcy may be necessary for economic growth as will be shown later in the section on the data.

---

<sup>1</sup>See section 5.9 for discussion.

Merger risk is not required for an equilibrium price to exist within the option markets themselves, but the absence does raise questions on the relationship between physical capital and financial capital.

The final issue is the existence of dividends. Historically dividends have been a nuisance. Even though standard finance theory says a firm's price is equal to the discounted value of future dividend flows, in equilibrium; they have uncertain timing, uncertain payment amounts and are subject to suspension. Regardless, dividends affect option prices even for firms that do not pay dividends.

The other difference in this dissertation is using Bayesian rather than Pearson-Neyman or Fisherian methods. This is due to the methodological issues in estimation created by limiting support to the positive real numbers due to bankruptcy. Although this is a seemingly small change, that is formally accounting for survivor effects, it turns out to create economically substantive biases in estimation.

Since the 1950's a synthesis of Fisherian and Pearson-Neyman methods has happened and for purposes of this dissertation will be called Frequentist or Frequency based methods. The two schools differ in important ways, but the differences do not impact the logic or discussion here. The greatest difference between the two schools regards the distinction between statistical significance and hypothesis testing and the use of the likelihood function versus the minimum variance unbiased estimator. With one brief potential exception, the differences do not impact the content here.

In Frequency based methods the function to estimate statistics and perform inferences is critical. There is an absence of estimation tools and testing tools for the mixture type problem present here. The closest is to use the extant literature which assumes the support for returns on the entire real line. Bankruptcy makes this impossible.

Unfortunately, the minimum variance unbiased estimator extant in the literature over-estimates returns by two percent per year and underestimates risk by plus or minus four percent per year, due to the impact of bankruptcy on the distribution.[83] This absence of a pre-existing tool in the Frequency-based literature is not an issue for Bayesian methods. In any Bayesian problem, the data is passed through Bayes rule even for the strangest of all distributions. Inference is performed on the posterior distribution or as a ratio of likelihood functions.

In doing so, Bayesian estimates lack the systematic bias present in the currently available Frequentist methods.

## 1.2 The Background for Mean-Variance Finance

Historically, there have been a variety of approaches to pricing equity option contracts. This dissertation takes a new and novel approach using relatively primitive maths. The primary method currently in use in economics is based upon Itô calculus as pioneered by Louis Bachelier in 1900.[9] His work was expanded upon by Fischer Black and Myron Scholes in their article *The Pricing of Options and Corporate Liabilities*. [16] A primer in such methods is provided by Neftci.[69]

The goal of such pricing models is to determine what price, in equilibrium, would hold for option contracts. The model was originally derived as an extension of the Capital Asset Pricing Model (CAPM) and the general methodology is the continuous time expansion of that model, which is a static model. Mean-variance finance was pioneered by Harry Markowitz and A.D. Roy. This work was later expanded by Black, Sharpe, Linter, Treynor and Mossin. [15, 57, 62, 65, 89, 104, 105] An alternate model was developed by Ross as Arbitrage Pricing Theory (APT). [82] The APT differs in that it considers the factors that drive pricing.

The models have a variety of assumptions that are key to understanding them. For the continuous time models, the requirement is that the time series converges. For arbitrage pricing theory to hold an expectation and variance must exist, further through the use of principal components analysis or factor analysis additional restrictions are placed upon the error terms. For the CAPM the requirement is that an expectation, variance and covariance exist. Itô pricing models can be derived either from the CAPM framework or the no arbitrage framework.

At the core of each of these models is that the distribution of returns has both a mean and a variance. Indeed, many models go a step further and simply assume that a normal distribution is present. This is extremely reasonable given the nature of many economic models. Further, even if this isn't true, provided the underlying distribution has a mean and a variance, then the sampling distribution of the means will converge to normality due to the central limit theorem.

The central limit theorem is so named, not because of some limit at the center

of the distribution, nor due to the presence of the mean at the center, but rather due to its central importance to the field of statistics.[46] While it is central to statistics, its importance to economics is even greater. The normal distribution and the expectations operator are everywhere in the modeling of economic processes.

What very few people, other than statisticians, are aware of is that there is an important restriction in the classical central limit theorem regarding the existence of a mean and a variance. The classical central limit theorem applies to any arbitrary probability distribution with a fixed mean and variance. This requirement, if not met, causes the classical central limit theorem to be inapplicable to real world problems.

The first appearance of this restriction in the normal law of errors, as it was originally called, was in a note by Poisson. Poisson, in reviewing the theorem, noted that the distribution  $f(x) = [\pi(1 + x^2)]^{-1}$  was a counter example to the theorem, as the distribution has neither a mean nor a variance. Still, Poisson wrote,

But we shall not take this particular case into consideration; it will suffice to have remarked upon the reason for its singularity and note that we will without doubt not encounter it in practice.[96]

Independently, Bienaymé wrote an article showing that least squares regression provided the best possible mechanism to fit a line to data, in contrast to a method provided by Cauchy.[96] He had discovered that the method of ordinary least squares gave the best linear unbiased estimator. This triggered a series of articles in which Cauchy developed a distribution, the Cauchy distribution, which would force the method of ordinary least squares to fail. This distribution was of the form

$$f(\epsilon) = \frac{k}{\pi} \frac{1}{1 + k^2 \epsilon^2} \quad (1.1)$$

While nothing stops people from using ordinary least squares as an algorithm, or for that matter finding a sample mean or sample variance, the algorithm has no predictive value and inferences obtained are meaningless. Indeed, Sen notes that such a method would be perfectly inefficient when compared with valid solutions when the Cauchy distribution is present.[87]

The first appearance of the normal distribution in economics and finance appears to be a presentation by Jules Regnault in 1853.[9] He discovered empirically what Bachelier would argue theoretically in 1900. In the interim, Edgeworth, following work by Laplace, Jevons and Quetelet, would argue for applying the *law of errors* to investments in general and Bank of England notes in particular. He takes it further, seeking to unite utility theory and probability theory.[32]

Edgeworth would not be the first to attempt to do this. The first is by Bernoulli in his solution to the St. Petersburg paradox.[112] To leap into mean-variance finance one must first pass through the works of Clark[24], Böhm-Bawerk[110, 111], Veblen[109, 108], Fisher[35]. Keynes[49], Pareto and Hicks.[44]

Böhm-Bawerk, Clark, Fisher and Pareto pull together the interest rate as the marginal cost of patience. A careful read of Veblen's work on the leisure class could be read as the first work on behavioral finance. Keynes work creates an idea not possible in the classical school, inefficiency and emotion in markets. Their work stands in contrast to the combined work of Pareto and Hicks. Hicks' work is central to the classical school of thought regarding capital. It is this work that starts Markowitz down his path breaking idea of having economists measure both risk and return.[62]

While Veblen and Keynes would continue to influence future economists, the latter more than the former; it is Markowitz who would set in motion Hick's unattained goal of "an economics of risk." Although Roy simultaneously discovered the same thing, it is Markowitz's work that is remembered.[84]

Hicks appears to make two conflicting comments in his book *Value and Capital*. [44] On the one hand, he clearly argues that people include risk in their plans and prices, implying economists should measure risk. However it is also clear from his writing that the tools to measure risk do not exist.

Hicks goes on to state that economists can ignore risk because it is included in the plans and expectations of the actors. By watching actual returns we implicitly get the risk variable, hence we need not measure it.

It is improbable that Markowitz could have guessed the impact of his initial writing. The transformation is greater than formulating a trade-off scheme between risk and return, it is a way of thinking about and including statistical measures in economic thought and economic processes. A casual read of this initial work

shows a field of economics in a comparatively primitive state. Indeed, without Markowitz, this and subsequent work is impossible. Although earlier writers, such as Regnault, Edgeworth, Hicks and von Neumann bring uncertainty and risk into the discussion, Markowitz and Roy are the first to propose a mechanism of exchange between return and risk.

It is from this starting point, and that of Bachelier and Samuelson, that Black and Scholes derive the model now considered the cornerstone of most modern models.

### 1.2.1 Partial Examples

The first problem to understand with the mean-variance models is that they are perfectly valid mathematical models given the *stated* assumptions. They are not valid scientific models because it is implicit in each of these models, given they are strictly true, that the marginal investor seeks to *lose* money with every transaction. This is far from obvious and so an understanding of the goals and methods of construction of the models are necessary. John H Cochrane's *Asset Pricing* is an excellent book to use to understand these models.[25]

To understand the problem with mean-variance finance it is first critical to understand what the models are trying to do and how they are arrived at. Markowitz's original paper was designed to provide a tool for asset allocation among competing assets by individual actors. In treating prices as given, the question becomes what allocation of risky assets should be selected? The models share a variety of elements as they ultimately are inspired or even directly derived from Markowitz and Roy's original papers.[62, 84]

In some form, there is an expected return. How that expectation is arrived at is unclear in the Markowitz paper, but because there are no costs in the primary models, it is necessary that in equilibrium allocations will be driven by the expected returns and that these returns are known in some manner to everyone.

It is rather important to note that equilibrium in these models is not defined as market clearing. The quantity supplied isn't even restricted to the number of existing shares, as it is possible to short sell and in effect create an infinite number of shares. Likewise, the quantity demanded can grow without bounds as infinite

borrowing is possible. Implicitly the market clears, but there is no explicit math to support this.

Equilibrium in the CAPM, for example, would be that each asset is efficiently priced. The simplest way to explain this would be to posit that pricing errors are driven to zero from competition. If an error is present then the model cannot be in equilibrium as this would constitute an unrealized profit. This is close to an absence of arbitrage condition.

In addition, the models posit the existence of finite variance for all assets and indeed presuppose a positive definite covariance matrix for multi-asset models and a fixed variance for single asset models such as the Black-Scholes option pricing model.

Most models also include a risk-free asset. A risk-free asset is an asset that pays out the same amount in all states of nature. A risky asset's payout is state dependent. Most standard models include this. The existence or absence of a risk-free asset is without consequence to this dissertation as it has no effect on the work involved.

Most models assume the actors involved are in perfect competition with symmetric costless information and an absence of moral hazard. There is an extensive body of literature on the lifting of these implicit restrictions.

Finally, there is an implicit assumption that individuals are utility maximizing and seeking a profit from investing or that firms that are investing are profit maximizers, both subject to a chosen level of risk. Alternatively, they assume the dual problem that individuals are risk minimizers for a given chosen level of return.

### 1.2.2 Example—Individual Asset Allocation With Risky Assets

Assuming no risk free asset, a possible solution to the asset allocation problem for a single individual following the Markowitz model is as follows:

Assume an arbitrary but finite number of risky assets exist. Further assume that a vector, to be denoted  $\mu$ , exists and that this is the solution to the equilibrium required return by the market in perfect competition. Further pre-suppose that

the covariance matrix, to be denoted  $\Sigma$ , exists of the returns on investment for the universe of traded assets. For each individual actor a utility function exists. This utility function is implicitly maximized, given the risks and market prices, by choosing some total portfolio return, a scalar denoted  $\mu_p$ . From these assumptions, a solution is sought for a vector of allocations denoted  $\mathbf{s}$ .

$$\min_{\{\mathbf{s}\}} \frac{1}{2} \mathbf{s}' \Sigma \mathbf{s} \quad (1.2)$$

Subject to:

$$\boldsymbol{\mu}' \mathbf{s} = \mu_p \quad (1.3)$$

$$\mathbf{1}' \mathbf{s} = 1 \quad (1.4)$$

Placing the problem in the form of a Lagrangian equation it becomes:

$$\mathcal{L} = \frac{1}{2} \mathbf{s}' \Sigma \mathbf{s} + \lambda(\mu_p - \boldsymbol{\mu}' \mathbf{s}) + \gamma(\mathbf{1}' \mathbf{s} - 1) \quad (1.5)$$

### 1.2.3 First Order Conditions for an Optimum

Solving first order conditions to determine an optimum:

$$\mathcal{L}_s = \mathbf{s}'^* \Sigma - \lambda \boldsymbol{\mu}' - \gamma \mathbf{1}' \equiv 0 \quad (1.6)$$

$$\mathcal{L}_\lambda = \mu_p - \boldsymbol{\mu}' \mathbf{s}^* \equiv 0 \quad (1.7)$$

$$\mathcal{L}_\gamma = 1 - \mathbf{1}' \mathbf{s}^* \equiv 0 \quad (1.8)$$

Second order conditions are omitted as it is well known that the solution has a unique minimum, the second order conditions play no subsequent role in this section, they play no role in the disproof of mean-variance finance, and they consume space. For proof see any standard reference or the underlying proofs cited in the bibliography.

### 1.2.4 Moving Toward a Solution

Using the equations resulting from the first order conditions we note that we already have solved the optimal allocations given the shadow cost of changing the

expectation and the shadow cost of removing the restriction that allocations must sum to unity.

Beginning with equation 1.6 we note that post multiplying by  $\Sigma^{-1}$  results in the solution:

$$\mathbf{s}'^* = \lambda \boldsymbol{\mu}' \Sigma^{-1} + \gamma \mathbf{1}' \Sigma^{-1} \quad (1.9)$$

It becomes a matter of solving for three equations with three unknowns. Transforms of the first order conditions result in a definite solution of all unknowns. The addition of the assumption that a market portfolio exists, that is all assets are owned, and so statements about the market in general and the behavior of assets in particular are possible. The addition of a risk-free asset or a zero-covariance portfolio allow a complete solution of the market equilibrium beyond a solution for a single actor.

The solution for the Black CAPM is:

$$\mu_i - \mu_z = \beta_i (\mu_m - \mu_z), \quad (1.10)$$

where  $\mu_z$  is the expected return on a portfolio of assets with the property that the collection of assets have the relationship of having zero covariance between the chosen portfolio and the market portfolio.

### 1.2.5 Example—A Continuous Time Single Asset Model

The key to understanding Black-Scholes is in understanding the logic by which the solution is set up. Indeed, it has a long and relatively tedious solution that is in itself not very enlightening. Fundamentally, however, only a few core concepts are needed. The first is that there is to be created some contract expiring under known conditions whose value is derived from the price of an underlying security and the statistical properties of that security. The price will vary with the passage of time and is modelled as a differential equation of bounded variation.

For purposes of this dissertation, the fundamental equation that sets a solution in motion is:

$$dX_t = \mu(X_t, t)X_t dt + \sigma X_t dZ_t \quad (1.11)$$

where  $dZ_t$  is a diffusion process,  $X_t$  is a price,  $\mu$  is a function that drives price

deterministically as a function of price and time and  $\sigma$  is a constant.

Important conditions include:

$$\Pr \left( \int_0^t |\mu(X_u, u)| du < \infty \right) = 1 \quad (1.12)$$

$$\Pr \left( \int_0^t \sigma(X_u, u) du < \infty \right) = 1 \quad (1.13)$$

The key element here is that  $\sigma$  is a fixed constant. A perfectly hedged portfolio is then constructed of a European style call(or put) and some asset. This hedge removes all risk; so the logic goes that in expectation, the hedged party will receive the risk free rate. The key to understanding how this happens is through Itô's Lemma. Itô's lemma permits a solution such that at the limit, the stochastic portion converges in value to the deterministic portion driven by time.

The resulting formula becomes the difference between two cumulative normal distributions, one scaled by the price of the underlying, the other scaled by the present value of the strike price. The critical assumption is that of a convergent process and a fixed variance.

### 1.2.6 Example—Arbitrage Pricing Theory

Arbitrage pricing theory takes a completely different tack on the pricing of equity securities. It varies in key areas from Black-Scholes and the CAPM, but if its factors are complete, then all variability of return would come from the variability accounted for by the model's factors.

The fundamental idea is that each security has factors that determine its return. Further, if you demeaned those factors and placed the mean drift into a variable, then return could be predicted by knowing the factors, the drift and the effect of errors. In particular, one could construct the following equation:

$$\tilde{r}_i = \mu_i + \sum_{k=1}^K b_{ik} F_k + \epsilon_i, \quad (1.14)$$

where  $\tilde{r}_i$  is the return on some asset  $i$ ,  $\mu_i$  is the mean return on asset  $i$ ,  $\epsilon \sim \mathcal{N}(0, \sigma_i^2)$ , where  $F_k$  is a factor with mean zero, and  $b_{ik}$  is the factor loading for that return.

Ross then takes a novel approach, he asks explicitly what would the returns become in equilibrium if the investor had no money invested, that is they short sold an amount equal to the long side of the portfolio and if no arbitrage opportunities existed.

To do this three things must be present. First, there must be no money invested, that is to say:

$$\sum_{i=1}^I s_i = 0. \quad (1.15)$$

Second, the portfolio must have a large enough number of securities to drive idiosyncratic risk to zero at the limit. Third, factor analysis is used to construct orthogonal sub-portfolios so that:

$$\sum_{p=1}^P s_p b_{pk} = 0, \quad (1.16)$$

where  $p$  is a sub-portfolio,  $s_p$  is the allocation of some factor loading, and  $b_{pk}$  is the loading of that factor on that sub-portfolio. This is added together so that the portfolio has no systematic risk.

The absence of arbitrage and the absence of risk with no investment implies that the return on investment, in expectation, is zero.

Of interest is that the expected return is zero, that no systematic or idiosyncratic risk exists, and in particular the long run effect of errors do not matter at the limit.

### 1.2.7 Commonalities

The models share a handful of things. Finite variance exists, in equilibrium prices converge so that they behave as if expected returns were known ex ante, equilibrium is with reference to pricing efficiency and no actual shares exist just allocations of money, implicitly consumption needs have no consequence to the equilibrium, errors are normally distributed and implicitly there is no free lunch.

Generally accepted in economic modelling is that errors will be normally distributed. Indeed this will become the crux of the problem. An assumption of normality is generally a low risk assumption in many problems. Many distribu-

tions can be approximated with the normal distribution without a large loss of precision. Many models can be reformulated so that the difference is irrelevant. Many statistical tests are going to use an assumption of normality. This is eminently reasonable and also incorrect.

## 1.3 Post Markowitz and Roy

### 1.3.1 Static/Discrete Models Post Markowitz and Roy

Markowitz's stated motive in authoring his initial paper was to correct a statement by Hicks.[62] The most probable statement in Hicks is a sequence of comments in *Value and Capital* about risk and return. Hicks states that people consider risk when choosing a return, but also states that economists can ignore risk as we can neither see the risks as seen by the actors nor determine what information is available to the actors. However, since risk should be incorporated into return, we can simply look at returns.

What makes this change profound is not merely that we now try to measure risk, what makes this profound is that under this formulation there is a trade-off rule between risk and return. Tobin then extends this to include cash as an asset; this change starts the development of the risk-free rate as a concept in the models.[103]

The challenge at this time is one of calculation; and indeed, the idea of a Cauchy distribution being present in the data may have been better received in the modern world of super-computing. Given the limited resources available at the time the next step was by Sharpe who reduced the calculation burden substantially.[88] This led to independent work by Sharpe, Mossin, Lintner and Black.[15, 57, 65, 89]

These static models would launch the linkage between the continuous time models and option pricing creating the transformative article by Black and Scholes.[16]

### 1.3.2 Continuous Time Models Up To Black-Scholes

As mentioned earlier, Bachelier developed the ground work for Brownian motion in his dissertation in 1900. Unfortunately it would moulder on shelves for decades, while its content would have to be reinvented by Einstein and Kolmogorov.[9] It

would sit until revived by Savage, who sent postcards to economists around the world recommending they read Bachelier's dissertation.[9]

In parallel, mathematics and statistics solved many of the key subsequent challenges that would be necessary to arrive at the Black-Scholes option pricing model. From an economist's perspective, the next major innovations were Itô's development of the stochastic integral and Donsker's work to show that continuous time and discrete time models were equivalent mappings through scale invariance.[30, 45]

Although the major innovation of Bachelier was to arrive at the Brownian process with its symmetric normal distribution, this was a problem for economists as it permitted negative prices through normal shocks. The next innovation was to propose a log-normal distribution to evade the issue. It's only downside is that bankruptcy becomes impossible as the shock could never reach a zero price. In the intervening period work by Samuelson and by Osborne, a physicist, prepared the way for Fischer Black and Myron Scholes critical paper.[74, 85]

### 1.3.3 Empirical Criticism

Yilmaz provides a good discussion of the literature of criticisms of Black-Scholes and replicates them with data as well.[116] According to Yilmaz "empirical evidence shows that the classical Black-Scholes model does not describe the statistical properties of the financial time series very well." Yilmaz notes excess kurtosis, skewness, volatility clustering, the volatility smile and market completeness.

While Yilmaz provides extensive references for these, Yilmaz constructs them from data as well. This criticism goes back to Mandelbrot's criticism of the mean-variance finance in 1963.[60] While it is a truism that all models are wrong, some are more so than others. If a Cauchy distribution is present in the data, then mean based measures will be uncorrelated with nature at the limit.[87]

The thesis of this dissertation is that in the absence of boundary conditions, the distribution of returns must converge to a Cauchy distribution as time goes to infinity. If one assumes, rather than proves as is later done, that a Cauchy distribution is present then a new light is shone on the model contradictions.

While the Cauchy distribution has no measure of its kurtosis, when measured

using an assumption of normality, the Cauchy distribution has a high kurtosis. To provide an illustration, given a data set to be modelled by a Cauchy distribution and a normal distribution by setting their modes and their probable errors equal, you would find the Cauchy distribution is expected to be 310 quadrillion times more dense than the normal distribution for data with a mode of zero and a probable error of one when measured at six times the probable error.

The measured skew likely has several components. Since the standard deviation, skew and kurtosis are arrived at using normality based measures and that there are no defined moments for the Cauchy distribution, caution should be used when discussing such observations. If the models are true, then there should be no excess kurtosis or skew, except in the case the log-normal model is used. In that case, there should be no excess kurtosis or skew in log-return space.

Ignoring for a moment the problem of the measurement, assuming a Cauchy distribution is present; likely sources of skew are:

- The real world has a budget constraint
- The real world has taxation
- The skew when measured using an algorithm that requires support on the entire real line would be calculated as skewed if the data is truncated for the limitation of liability, even if the data is in fact symmetric otherwise.

Volatility clustering would be a symptom of the presence of a Cauchy distribution. Volatility clustering would result as the measure of variance over a local interval will be a random variate itself. There may also be cyclical processes at work in real data.

The volatility smile is the observation that prices paid by people in a real market imply that volatility changes non-linearly rather than linearly with time. The graphic visualization of this process is called the smile. This is a contradiction to monotone volatility with time. It is likely liquidity costs drive this, but this may also be a symptom of the absence of a fixed variance.

Finally, market completeness implies that there are the possibility of risks with no off-setting market mechanism to absorb these risks. It is difficult to understand the impact of this as an empirically testable assumption. The presence of political

risks, for example, may have no offset. How would one have priced the probability the Soviet Union would collapse? Is there a measurable process to test for revolution? What about disruptive technology? These risks may influence prices, ex ante, and event studies may capture this, but that would really be the incorporation of data as people suddenly believed it could be a real possibility. For long term options, the model cannot anticipate this.

The Bank of International Settlements in its 2012 semi-annual report on derivative contracts reports outstanding notional amounts of \$647,762 billion in over the counter contracts. They are valued at \$27,285 billion using Itô calculus based methods and if a Cauchy distribution is present then they are valued at an unknown amount. Proper pricing is a critical issue.[71]

## 1.4 Market Makers

Traditional option pricing models ignore the market makers' role in pricing option contracts. Dealing with the problem of the existence of a market maker adds a terrible complexity to the math that is not ordinarily dealt with at all. There are no frictions in the core mean-variance finance models as they disrupt the math. Yet the market maker is a key point for option pricing.

Market makers are financial intermediaries. They serve a variety of functions. They provide liquidity to the marketplace. They underwrite and create new securities where none previously existed. They provide notarial services by guaranteeing that assets and liabilities are properly assigned. They extend credit.

The unusual aspect is that they provide these services to generally anonymous counterparties. A person needing a sudden \$1,000.00 can immediately sell a like amount of securities without the need to find a buyer. The market maker acts as buyer where no market for the security currently exists. The market maker carries the asset until a buyer comes along at some future time. In essence, the market maker has loaned the market \$1,000.00 for an unspecified period of time with an uncertain pay off. This is done for a fee and in some cases for the right to create derivative securities.

In this respect, market makers are banks. Traditional deposit banks operate by taking capital and loaning it to the market place. The proceeds of those loans

are deposited creating deposits. The job of banks is to manufacture deposits. In essence, their capital is used to create money. Likewise, health insurers create a demand for health.

Without a health insurer only those with sufficient resources could pay for health care. The demand would be relatively small and the number of doctors, nurses and pharmacists would be very small. By pooling the risk and making the cost affordable, health insurers create a demand for health care professionals and health itself. Health insurance creates health.

Traditional broker-dealers place their capital at risk and in doing so create a demand for capital. By pooling risks and transferring risks to those most capable of bearing those risks, more capital is demanded.

At the heart of all financial intermediation is the providing of liquidity at the time of greatest need rather than when it would be otherwise available. Intermediaries make themselves fragile so that their customers can be flexible.

### 1.4.1 Diamond and Dybvig

The proposed model could be thought of in the framework of a Diamond and Dybvig model, but from the perspective of the bank.[28, 29, 36] To understand the Diamond and Dybvig model, one must begin with utility. The simplest way to conceptualize the model is to presume that an actor is planning consumption for the next two periods. For the next period what is not invested today is consumed tomorrow and what is invested is consumed in the following period. The added modeling element is the possibility of a liquidity shock.

In a Diamond and Dybvig model the refrigerator could break, the furnace could fail or the car could crash. This would require consuming future period investments early. The source of the unplanned illiquidity is unimportant for the model; what is important is that future liquidity needs are uncertain as to timing.

In a Diamond and Dybvig model there are three possible scenarios; they are autarky, a bond market and a banking system. The important finding of the model is that the banking system is Pareto optimal. Banks pool resources and allow a statistical transfer of resources from reserves set aside for “rainy days.” If there were no uncertainty as to the true rate of emergency then exact reserve

requirements could be constructed and optimal investment would happen in the system. The surprising result is that bond markets cannot obtain optimal risk sharing.

Empirical evidence for this understanding of broker-dealers in the various forms of active trading markets comes from a variety of sources in what are primarily before and after studies. Anand and Weaver report improved bid-ask spreads.[6] They estimated cost savings to participants in 1999 of 221 million dollars.

Tse and Zobotina went a bit further in studying the creation of a designated market maker for a specific issue. The results provide some insight into the market microstructure that is altered by the presence of a market maker. In particular Tse and Zobotina found that:

... introducing a designated market maker in the trading pit enhances competition, reduces transaction costs, and improves both liquidity and market quality. The market maker enhances the speed and efficiency of incorporating information into prices.[106]

This is also generally true of Mayhew, however Mayhew reports blocks of time in high volume issues where this statement isn't true.[64] However it is unclear the role of the market makers in those high volume markets. It may be that the volumes can be high because a market maker exists to catch the proverbial "falling knife." In that case, the market maker may be acting as an insurer of last resort permitting behavior not possible without it.

In the Diamond and Dybvig model, banks can exist in one of two states. The first state, which could be called the stable state, behaves as if all deposit withdrawals are idiosyncratic. The other state is the bank run. In the bank run, depositors view withdrawals by other depositors and may view them as unusually heavy. The perception is that there are too many withdrawals happening for it to be due to chance alone. Someone is perceived to be acting on private information and making appropriate withdrawals. Other actors, viewing this, then respond because such banks can only pay out until reserves have been paid out and then no further withdrawal is possible.

Goldstein and Kavajecz found that in periods of stress limit orders did in fact vanish and a reliance on market makers took over on the New York Stock

Exchange.[40] OHara reports the same phenomenon on the foreign exchange markets.[72] This makes use of the Diamond and Dybvig model somewhat restricted in that it is unclear how empirical support for this phenomenon impacts the model as elements of this resemble a bank run rather than a stable state, but not enough to directly apply the model.

There are a number of possible solutions to the problem of the bank run. Deposit insurance, lender of last resort policies, and banking suspensions are all strategies to manage the bank run by authorities. There are two possible types of runs facing market makers in this dissertation.

The first, and the most important seen from the perspective of moves in a game, is when an unusually large demand for shares or cash are made by a participant or many simultaneous participants. The second would be the mass exercise of option contracts destabilizing the off-setting positions. Enron was an example of this.

In the first case, for a single block order, there are a number of strategies available to the market maker. The market maker could fill the order from inventory, mark the order up or down, and gradually replace the inventory in the market. The second is to take the order off the market and quietly fill it from market orders as time passes. These orders exceed the planned reserves set aside to maintain an orderly market and could upset market prices. As such, these items are normally taken off the tape and only reported later at a weighted average price when the information can no longer impact prices.[72]

The second case is more serious. The game assumes away counterparty bankruptcy risk by sufficient collateral, but a rational bank will expand its book of business past that point. The primary role of the market maker is to be a guarantor for the liabilities of the contracts written through its customer base. In the real world, customers die, declare bankruptcy and face statistical runs. Any such risk is priced, in equilibrium, by a marked up rate when the market maker accepts the writer's contract. What is a bit more challenging to model are systemic runs on the system. In that case, many writers who were facing independent risks now face correlated systematic risks and fail. The bankruptcy cost rolls onto the market maker's balance sheet as defaults mount.

The challenge of replicating this, for the purpose of this game, is in forming a well-defined predictive function for bankruptcy. To do this, it is necessary to view

market makers as part of a macroeconomic game.

There are several potential problems with doing this here. The first is that the government controls the definition of solvency and can and does change it. Further, the government can transfer risks or losses to market makers and so the government can itself be a source of insolvency.

A second potential problem with simple prediction is the presence of a coordination problem. In the real world a variety of contract types exist. Some contracts are structured as fixed nominal contracts, others attempt to behave as variable real contracts. Some contracts can be terminated immediately, some contracts are short in duration and some are long in duration. Some contracts require prepayment, such as insurance contracts, others require intermittent payments while others require payment upon completion of the contract. In addition, financial institutions are often the nexus for contracts. An insurance company underwriting pensions and using mortgages to fund the pension payments joins together two separate types of obligations, but obligations with very different types of risk.

With a pension there are a variety of risks, including the risk the beneficiary could live too long or that the reinvestment rate is less than the required amount to meet the pension obligations. For the mortgage there is a risk of late payment or of non-payment. Further, there is the risk of interest rates falling resulting in early prepayment. If the new rate is low enough, then the pension obligations cannot be met. Some risks are correlated directly, such as interest rates, but others indirectly. Although the risk of a single annuitant living too long would have a trivial effect, many annuitants living too long would change the macroeconomic conditions. They would demand different types of services, they would remain in homes longer and so forth.

In the case of market makers in this game, there is a risk that market makers could be over-exposed to some type of event and not realize that there is an underlying correlation between a risk and a position.

In addition, the failure of a market maker would depend upon the bankruptcy law involved in the unwinding of contracts. If person A owns a call and person B underwrites the call, but the market maker does not exist, then neither contract could potentially exist.

Further, if you consider an economy with a GDP understood as  $Y_{t+1} = \gamma Y_T +$

$\epsilon, \epsilon \sim N(0, \sigma_Y^2), \gamma > 1$ , then Bayesian estimates of  $\gamma$  will follow a Cauchy distribution.[114] If the number of fixed contracts, not subject to cancellation, are large as a percentage of anticipated income after basic consumption is met, then a shock to income would result in systematic bankruptcies. The path of those contracts is not fixed, though it should be possible to game out various scenarios. As such, it is unclear whether a simple function would suffice for market makers. A combination of models for various scenarios could be created by using Bayesian networks, but because systemic crisis are relatively rare the networks would depend heavily upon the prior distributions as little data would exist for any one path.

If you extend the argument of Evgeny Slutsky that auto-regressive processes with additive errors can appear as sinusoidal processes entirely due to the summation of random shocks, then booms and busts can be thought of as statistical runs that compound on themselves.[93] Since people cannot see their own errors, or of course they would not make them, banks cannot see the errors they make when they join in the errors of market participants. As there is not a unique channel nor a unique bankruptcy process for market maker failure in the system, attempting to game this is prohibitive for the purpose of this dissertation.

Further, each contract can itself contribute to the risk of a statistical run in the forward period. Separate research on this phenomena is required to price market maker failure and participant failure risk.

### 1.4.2 Options as a Risk Management Tool

Financial intermediaries who provide services in both the spot market and the option market increase both their own liquidity and the liquidity of either marketplace. To understand why, consider a market maker who only provides spot market services.

If an order to buy securities is placed in the market, the market maker sells securities either from inventory or borrows them, often from a customer's account. To defease the risk of loss from being in the position of a short seller, the market maker has to find a seller from which to buy securities. If the market maker also underwrote positions in the derivatives market then the position would also be offset if there were customers who wanted to buy a put and another that wanted

to sell a call.

The derivatives market permits the market maker to consider two strategies unavailable to a dealer operating only in the spot market. First, the market maker could hold the short equity position and buy a long call from a customer. To do this, the customer has to be short the call, underwriting the market risk of the position. However, in doing so, the market maker's position has been transformed into a long put. This would be called a *synthetic* long put. No actual put contract is in place.

If the next order were a request for a long put, the market maker would create a contract for an actual long put and sell it to the customer. The risk would be offset by the synthetic long put. If on the other hand the next order were to sell shares, the market maker would be in the position of holding a long call. The can continue in a variety of ways until the aggregation of positions close out. As this opens up two other markets, each of which can be thought to have Poisson arrival times, the speed of order flow should increase, reducing the time risk of the maker. The probability of there being an order to gather pricing information from goes up as a function of time.

The second strategy available is to open and hold positions using the spot market and the option market to construct synthetic positions. In this role, the market maker acts more like an insurer or lender and less as a facilitator.

This creates two possible profit functions, one as a facilitator and one as a bearer of risk. In both cases the market maker underwrites the securities and assures proper transfers of ownership. The difference is in risks held.

While the market maker is indifferent to the gambles presented for the purposes of market making, it may not be indifferent to the gambles held for portfolio purposes. Indeed, Amihud and Mendelson show that a market maker will, under very mild conditions, systematically lose to informed investors. As such, portfolio positions taken as a result of market making operations could be gamed by informed counterparties if those counterparties have information unavailable to the market maker.[5]

### 1.4.3 Amihud and Mendelson

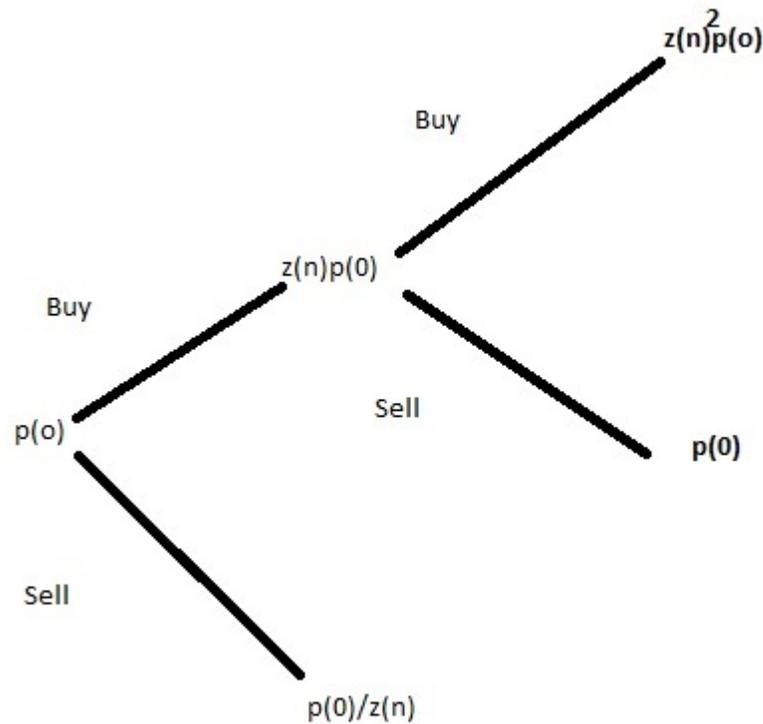
The market maker controls two things, the bid-ask spread and the institution's portfolio composition. If we assume the market maker only assumes positions for purposes of customer and market liquidity, then positions are incidental and are in essence a residue of ordinary trading behavior. Doing this permits us to transition from the more general Diamond and Dybvig style model to the narrower pricing based model of Amihud and Mendelson.[5] and from there to the operationalization of the markup/markdown process by Abbott.[1]

Abbott has shown that the bid-ask spread is a function of the half-life of a position. The larger the position a market maker must take, the greater the cost to the person entering into the market order. This satisfies the requirement from Amihud and Mendelson that some form of continuous time discounting be used.[5]

The need for continuous time discounting comes from the observation that inter-transaction arrival times are not identically distributed.[5] As such, there does not exist a convenient unit of time for which a rate can be set other than continuous time.

In addition to market orders there are limit orders. Whereas market orders drain liquidity from the market by taking it away from the market maker, limit orders add liquidity to the market. A limit order will break up the implied geometric progression of prices from a statistical run as in the figure 1.1 In a limit order, a market participant for a stock states a maximum amount they will pay for some quantity of security or the minimum amount they will accept to sell some quantity of a security. In doing so, the participant reprices the security, in essence sets a new current price, denoted  $p_t$ . If no other orders are present in the market, the market maker may fill the order from their own inventory if they feel the price is advantageous, or may simply hold the order in abeyance as a source of funds should the market move in that direction. If many limit orders are present, they provide a market maker with an estimate of both the supply and the demand curve.

If the mark-up/mark-down function were thought of as  $e^{\pm\lambda n}$ , where  $\lambda$  is the implicit cost of liquidity and  $n$  the number of shares then the market maker is going to try and maximize profits by choosing  $\lambda$ . The value of  $\lambda$  is a component of net price, As the cost of liquidity is a function of  $\lambda$ , it affects the number of shares



**Figure 1.1.** Possible path of orders

transacted, but also the frequency of shares transacted.

To see this would be the case, imagine a change in  $\lambda$  causes an increase in the quantity supplied and demanded by the market in a market with a fixed trading period. Those additional quantities must be filled in the same amount of time and must be filled sequentially as well. Two things should happen, both the average order size and the number of orders should increase. For the cost of liquidity to change without a change in the interest rate paid by the market maker, then many additional actors must be filling in the demand curve with limit orders.

In that case, the market maker can reduce the capital commitment to keep the market liquid, or could hold capital contingently as in bank letter of credit obligations.

### 1.4.3.1 Liquidity During Runs

Not approached in this game is the question of liquidity during runs. This is a pricing issue, but it is empirical and outside the scope of the data available. To understand non-equilibrium pricing, or rather multiple equilibrium pricing we must go back to the Diamond and Dybvig game.

In a Diamond and Dybvig game there are three and possibly four equilibrium states if you allow banks to sell assets into a market place. In that case, markets ride on top of the bank equilibrium reducing the phenomenon of credit rationing.[36] In the world with banks instead of markets the pricing reflects the low cost of liquidity due to its provisioning by banks. However, imagine there is some possibility of a bank run and banks cease providing market liquidity. Participants must then find buyers for their securities and new option contracts cease to be written.

In this new environment positions are no longer *ex ante* efficient. Trades become infrequent and become a function of the elasticity of supply and demand. Overall, prices should fall to reflect the higher cost of liquidity. This will place puts *in-the-money* while driving the value of calls downward. However, market makers will not be able to recover their losses by selling into the market as volume has contracted by the absence of their liquidity, exacerbating the fall. Further, collateral will be of reduced value and so more bankruptcy claims will form on the market makers. Customers with cash deposits at the market maker will need to use them to cover liquidity driving up the internal cost of liquidity for the market maker.

Is this always the case? No, it isn't. A stock that gradually loses market support will cease being supplied with derivative contracts. It could eventually make it onto the *pink sheets*, where no market maker exists. An order can be placed and may be filled months later or not at all. The market maker is a pure broker and only facilitates trading. In that case, the market maker pulls out of the market in a manner such that its departure is not evident until it is gone.

Since this state of affairs exists with positive probability, like pure bankruptcy it should be priced, but unlike pure bankruptcy where contracts are not fulfilled, here the ability to exit loss positions changes. That is  $\lambda_t \neq \lambda_T$ .

## 1.5 Existential Risks and Cash Flows

Firms are a very unusual human construction. Some firms, such as limited liability companies, usually have defined lives at which time the firm dissolves. Other firms, such as partnerships, generally dissolve when a partner dies or when the partnership agreement is terminated. Although there are older business forms extant such as the business trust or the joint stock company, most people think about the corporation when they discuss the stock market.

Corporations are unusual in that they lack, generally speaking, a predetermined date of dissolution. They can exist in perpetuity, at least in legal theory. Nonetheless, there are a number of specific types of risks to their continued existence.

The three most common existential risks are bankruptcy, liquidation and the merger. Each of these categories have an extensive legal tradition and body of law around them. Indeed they cannot be spoken of independent of the laws around them. A proper study of finance cannot be independent of the laws and institutions finance must operate in.

For purposes of the dissertation the terms are going to be reduced to relatively narrow operational definitions that are somewhat at odds with the legal definitions. Legal definitions provide clarity of the legal status of shareholders and other stakeholders but do not consider the subjective outcomes of these legal statuses on the participants.

Shareholders of a firm in bankruptcy may walk out of bankruptcy retaining economic value. Although their shares will be vacated by the court they will be issued new shares, with a diluted interest, as if a merger with the creditors had happened. Previous creditors will suddenly find themselves as shareholders in what was previously a debtor firm.

Similarly, owners of a firm in liquidation will end up with no residual resalable claim once the final liquidating dividend is paid; that is no different in practice from the shareholders of a bankrupt firm liquidated by the bankruptcy court.

A merged status also isn't a clear status. With a cash merger the pre-existing shareholders receive cash in a manner no different from a firm in liquidation. The shares of the old firm are liquidated for an agreed upon amount of cash. In other mergers, pre-existing shareholders receive shares in the buying firm, though this

is no different in status than a bankrupt firm where the pre-existing shareholders retain claims on the new entity.

Bankruptcy will be defined here as any state of existence where the future value of an equity security is zero at some future fixed date. How it arrives at that zero value is irrelevant with one important caveat. Some firms cease trading but still exist. The shares cannot be legally traded on an open exchange but can be privately sold. The value is therefore positive even if this also means the security could not be sold in most pragmatic situations. It isn't the absence of trading which triggers the bankruptcy state, it is the absence of any legal claim to an asset. It isn't the price that has gone to zero in the allocation formula of price  $\times$  quantity, it is the quantity that has gone to zero.

Under this definition, a firm issuing liquidating dividends that terminates its existence prior to some fixed period has become bankrupt by that period even though no bankruptcy filing ever happened.

This leads to two definitions.

**Definition 1.** An *allocation* at time  $t$ , for all times  $t$ , is defined as  $p_t \times q_t$ . In this definition  $p_t$  is the price of the security and  $q_t$  is the quantity of the security owned at time  $t$ .

**Definition 2.** A firm is *bankrupt* at some future time  $t'$  if at time  $t'$  the quantity,  $q_{t'}$ , in allocation  $p_{t'} \times q_{t'}$  is equal to zero. That is  $q_{t'} = 0$ .

This discussion then logically leads to a discussion of dividends since it is possible for a firm to behave as an annuity and become self-liquidating. For that purpose a definition is in order.

**Definition 3.** The *dividend rate*, to be denoted  $\delta_{t'}$ , is a cash flow to shareholders at some time  $t'$  from the owned corporation, where  $t' > t$ , where the shares were purchased at time  $t$ .

**Definition 4.** The *dividend* is defined as  $\delta_{t'} \times q_{t'}$  at time  $t' > t$  where the shares were purchased at time  $t$ .

A merger is differentiated by a change in the form of legal property owned. Just as bankruptcy regards the status of the quantity owned a merger is defined in terms of a quantity and not a price.

**Definition 5.** A *merger* is defined as a change in the legal form of a claim on equity at some future time  $t'$  from its form at time  $t$ ,  $t' > t$ . For a shareholder owning a quantity of shares in a firm to be denoted  $i$  at time  $t$ , the quantity is exchanged for a quantity of some firm to be denoted  $j$ , where  $i \neq j$ , at some time  $t'$  where  $t' > t$ . The allocation goes from  $p_t^i \times q_t^i$  to  $p_{t'}^j \times q_{t'}^j$

An important missing item here regards firms sold for property, such as cash, and shares. An example would be a firm that was merged out of existence for \$50 in cash and one share of the new firm for each share of the old firm. Under the above definition the cash would be treated as a dividend and only the new shares as the merger.

### 1.5.1 Estimating Existential Risks

This reduces the types of existential risks to two types. The first is that it will cease existing as a going concern, the second is that it will be absorbed into another entity. The first risk eliminates future cash flows, the second risk changes their character such as frequency or amount. The literature of both types of states has been from the perspective of practical users, banks that want to make loans and want to avoid bankruptcy and market participants that want to gain extra profits from anticipating future mergers. The goal of this study is limited to pricing options.

Pricing options is not a concern in the existing literature for either bankruptcy or mergers. The goals in the existing literature are more pragmatic and general. While mergers are a pragmatic risk in option pricing, the impact of mergers on option pricing has been through how the terms and conditions of the merger impact the contract to deliver shares. Is a contract to deliver 100 shares of ABCorp the same as delivering 500 shares of DE Financial post merger, if that is the terms of exchange?

The limited goal here is to find the posterior probability of bankruptcy and merger and to limit that discussion to the impact on option contracts.

### 1.5.1.1 Bankruptcy

As the goal is the development of an option pricing model, the development of a model of how and why bankruptcy happens, its legal forms and the decisions involved is outside the scope of this work. Rather the concern is developing an estimate of the probability that bankruptcy, a merger and/or a dividend will occur.

The challenge of this form of estimation comes from the nature of the data itself. Accounting data can be very challenging to work with. The variables are highly correlated by design. For a firm in equilibrium, with some predetermined stable credit policy, the firm may find that 30% of its sales are in the form of accounts receivable and the rest in cash. If the sales of that firm are known then in approximation at least two other variables are known. Correlations between accounting variables in the Compustat universe run between 60% and 96% depending on the variables involved. It is possible to have captured most of the random variation in a firm by knowing just one variable. This lack of independent information makes estimation methods difficult to construct.

The first important work on this is the publication of Altman's Z score in 1968.[3] The purpose of Altman's Z score is to produce a score that is highly indicative of the risk of bankruptcy or non-bankruptcy. It makes a rather clever use of discriminant analysis. Extreme scores indicate a high probability of bankruptcy or alternatively continuing as a going concern. Middle scores have been found not to be correlated with outcomes.[20]

The use of Altman's Z score has primarily been to make yes/no types of decisions in practice. The difficulty is that it does not translate into a probability statement about the risk of bankruptcy given accounting data.

Altman's Z score is a clever tool given the issues facing analysts in 1968. Beginning in the 1930's analysts began testing ratios for their predictive power.[3] Analysts had been extremely critical of the pre-existing rule of thumb methods then in use. Unfortunately these ratios did not seem to discriminate between going concerns and bankrupt firms.

Discriminant analysis is powerful in that it reduces the dimensionality of the problem into the number of groups the researcher is concerned with minus one. Since bankrupt versus non-bankrupt are two groups, this reduces the question down to a single dimension and looks much like a regression model in practice.

Altman made multiple computer runs on a sizeable portion of the combinations variables available and published the resulting formula he felt was the best solution. Unfortunately his publication of the criteria he used to select the best formula was unclear.

The final scoring was:

$$Z = .012X_1 + .014X_2 + .033X_3 + .006X_4 + .999X_5 \quad (1.17)$$

where:

$X_1$ =Working Capital/Total Assets
$X_2$ =Retained Earnings/Total Assets
$X_3$ =Earnings Before Interest and Taxes/Total Assets
$X_4$ =Market Value Equity/Book Value of Total Debt
$X_5$ =Sales/Total Assets
$Z$ =Overall Index

Altman split the resulting scores into three zones. The middle zone was called the “zone of ignorance” where misclassification happens.[3] It is this zone that seems to trigger the difficulties.

These difficulties are not real if used as the author intended. He provides an example:

The discriminant Z score index can be used, . . . ,as a guide in efforts to lower the costs of investigation of loan applicants. Less time and effort would be spent on companies whose Z score is very high, i.e. above 3.0, while those with low Z scores would signal a very thorough investigation.

However, as we need an option price for every security, even those securities with middling financial statements, this type of analysis won't be very helpful.

Hillegeist, et. al., in assessing Altman's Z finds that price volatility is a more accurate assessment of bankruptcy probability than other methods reviewed.[59] At a certain level this is unsurprising. A rational market maker facing the risk of their inventory being revalued to zero would want to maintain as minimal an inventory as possible. This would result in a lack of smoothing and a wide bid-ask

spread. A wide bid-ask spread will be inherently more volatile, *ceteris paribus*. Nonetheless this creates a cart before the horse problem.

This issue is brought forward in Cooley's 1975 work on Bayesian discriminant analysis. Although having many of the issues that would follow from any Frequentist method, Cooley brings up two important issues. The first is the cost function. The cost of misclassification isn't trivial and as the costs may not be symmetric, the trade between Type I error and Type II error may be misplaced by the choice of an arbitrary choice of  $\alpha$ . The second issue runs to the "cart before the horse" problem.[26]

The article was written just after the Lockheed and Penn Central collapses and the market was caught surprised at their sudden failure. As many measures depend upon the market making this decision it leaves the open the giant problem of what criteria to use.

If people in a market judge a firm as having a high probability of bankruptcy, how did they do this? How did the market maker realize this fact? How did the existing shareholders and possible future shareholders judge this? This problem is also present in Altman's Z score via the price to book ratio. While this is a good time saver for a bank loan officer it is problematic for someone attempting to price securities based on forward prospects, particularly if it is not yet common knowledge that the firm is doomed.

In a broader based criticism of the methodologies of bankruptcy, recovery and state change prediction in general, Nwogugu goes after prediction models on deeper methodological issues including probit/logit and neural network models along with discriminant analysis. He points out that none of these models possess a theoretical grounding, they ignore the political economy the firm is operating in, causal relationships, state functions and the possible unique role of government in triggering bankruptcies such as through industry deregulation or interference.[70]

An important criticism by Nwogugu is of misspecification. There are several types of misspecification possible. One is by assuming there is a single optimum level all firms are judged regardless of their markets, capital structure, or management plan. Although not mentioned by Nwogugu, another is judging firms that lack some variable or for whom the variable behaves differently than in a different industry. Finally, there are violations of assumptions in the logit/probit models in

use that make them effectively misspecified models. Included in this criticism are the hazard function models that judge it in an actuarial manner. Among the criticisms not listed above is a reliance on external raters such as bond rating services to set the rating.[70]

Nwogugu's criticism of neural network models is less well laid out. Neural networks have significant limitations, in particular, the models are subject to change, the models are difficult to verify and validate relative to the problem at hand, the coefficients cannot be interpreted and nothing similar to a significance measure or an interval estimate exist.[52]

Not well called out in his article is the role of the macroeconomy on firm specific outcomes. It is not an economics article and items that would be symptoms of macroeconomic factors are used as examples.

He also calls out the fact that the models are models of initial court filings and not of bankruptcy. That is the various models in use don't measure loss to stakeholders but rather the filing of a bankruptcy petition.

Nonetheless, he does provide reasonable criticisms of the models in that they are often used outside the context they were designed for. Many of the models were designed with a limited purpose and then generalized by the audience for such models. Unfortunately he does not provide any suggestions for better models. Indeed, we need a general model of failure not merely an attempt to estimate it using data mining.

An article which goes somewhat down the road toward thinking in terms of process and time effects is an article by Ando on the hazard term structure of failure.[7] By looking at the future time structure of the probability of failure Ando provides something other than an estimate at a fixed point in time. However this is of less interest for option pricing as the European option contract itself closes at a fixed point in time.

A promising improvement to the methods has been produced by Sun and Shenoy.[100] Some of the issues addressed by Nwogugu's critique are naturally addressed in this article through the use of Bayesian networks. A Bayesian network is somewhat of a hybrid construction. It isn't a pure computational tool nor is it a pure network as may be the case in a neural network. Their findings were that a naïve Bayesian network performed in a superior manner in both false

positives and false negatives. Still, the methodology could be justifiably open to criticism and improvement.

A Bayesian network is defined by Sun and Shenoy as,

... probabilistic graphical models that represent a set of random variables for a given problem, and the probabilistic relationships between them. The structure of a Bayesian network is represented by a directed acyclic graph, in which the nodes represent variables and the edges express the dependencies between the variables.

A model is considered a naïve model by including an assumption of independence among the predictors so that a simple Bayes factor solution can direct the outcome. Correlations are addressed through the linkages, but the prior distributions are flat. Importantly, it is assumed the relationships are linear in nature, as is the case with the other models in the literature. Their model had two layers to deal with recovering missing data in the higher order layer. They found when using other industry methods, which require complete data sets, two thirds of the data was lost. This loss is captured in the Bayesian network since Bayesian methods have a simple method of dealing with missing data.

Some of their findings were interesting. They found that turning continuous variables into discrete variables using the extended Pearson-Tukey method substantially improved performance. They also tested the bracket median method, but believed the extended Pearson-Tukey method superior because of how it handles tails.[100] The graph is designed to minimize the impact of redundant information and maximize the effect of independent information.

Unlike the other methods in use, this model requires a statement of the posterior probability of bankruptcy given whatever data is appropriate. The Sun and Shenoy article are closest in providing a probability statement. Using what amounts to the likelihood ratio test of the two hypothesis where the null is the firm will go bankrupt and the alternative is that it will not, then it is classified as bankrupt if the odds are better than one to one in favor of bankruptcy. Of course an odds ratio is just a transformation away from being a probability statement.

The two weaknesses of this article are simply the assumption of linearity and possibly the mechanism for choosing variables. There is still no underlying theory

of the various economic channels that lead a firm to bankruptcy. This is in part because bankruptcy is a complex process. Firms declare bankruptcy as a strategic move, because they are forced by their creditors, in order to protect the jobs of senior managers, to deflect the costs of previous poor decision making and due to changes in the economic environment. Bankruptcy is the result of often years or decades of choices. It is quite possible that by focusing on things such as the one year probability of bankruptcy the wrong time scale is being looked at.

Although economics lacks a general theory of bankruptcy, outside the area of prediction there is work on the role of financial ratios in the firm itself. Davis and Peles studied whether firms' ratios moved to equilibrating values or followed a random walk.[27] An equilibrating value is understood by Davis and Peles as:

an accounting ratio may have an equilibrium value if management targets a certain ratio so that any deviation from the target causes management to initiate actions that will return the ratio to the target. Also, although management may not be targeting the ratio, the interaction of management's actions with external market forces may lead to an equilibrium value.[27]

They found that when a firm experiences a shock to liquidity equilibrium forces counterbalance the impact of a little more than a third of the impact in the following period for financial ratios. For performance measures the recovery period is longer but still shows evidence of equilibrium values. Of particular interest would be where equilibrium ratios are forced by competition away from bankruptcy minimizing points.

In the case where market forces push firms to risky combinations of management decisions it may be possible to generate a substantial lead time to predict bankruptcy. This may be a very fruitful path of investigation in the future. Caution should be taken when looking at this as a long run static equilibrium. In a long run study of farming equilibrium rates, Shepard and Collins found that the technology employed altered the mix very substantially over the period so that in periods as short as a few years the values are stable, but they gradually change as technology changes the mechanisms available to produce the output desired.[90]

The other likely fruitful path is to look at the market where bankruptcy carries an entirely different impact, the bond market. Because of the fixed income nature

of the bond market the ability to adapt bond structures to changing market conditions is totally absent. Giesecke, et al, look at corporate bond default over a one hundred and fifty year period.[97]

They found that stock returns, volatility and changes in GDP are strong predictors of default rates while credit spreads were not. They further found that credit spreads are not responsive to realized default rates.[97] This latter finding is somewhat surprising, but the finding that bond defaults are a function of gross national output are unsurprising. A sudden downward shock in total output should trigger substantial increases in the number of defaults.

Outside these methods, an unbelievably wide variety of specific methodologies have been attempted to varying degrees of success. Their main goal is to improve accuracy of prediction, but do not necessarily shed a theoretical light on why bankruptcy happens. They include generalized additive models, discrete time survival trees and forests, a variety of data envelope analysis based methods, fuzzy and rough set methods, Frequentist methods assuming stochastic covariates, multidimensional scaling, asymmetric Levy' flights, support vector machines and principal components analysis.[18, 12, 22, 23, 31, 38, 50, 73, 78, 79, 80, 81, 91, 98, 99, 107, 115]

#### 1.5.1.2 Mergers

Any discussion of merger risk has to account for the critical article by Palepu on predicting which firms are takeover targets.[75] Prior to Palepu very high levels of predictive ability were reported in estimators and after Palepu the values came down substantially. Palepu's article is complex and multifaceted. Some portions apply to Bayesian methods and some are strictly important to non-Bayesian methods.

The relationship between Palepu's article and Bayesian methods is through the relationship of maximum a posteriori methods and maximum likelihood methods. As such, caution is required in understanding Palepu's criticism in light of the likelihood principle and Stein's paradox.[55, 95]

Palepu's posited two objections that had three impacts. The first was that state based sampling introduced bias both into the estimator and the test statistics that would be eliminated by random sampling. The second was that arbitrary cutoff values, without considering the context, payoffs or cost functions made the results

difficult to interpret. Both of these objections run deep into the Bayesian versus non-Bayesian debate.

The first impact, that of a biased estimator, is important in that Palepu shows that existing sampling methods would overstate the probability that a firm would be a target of merger. This is due to the fact that state based sampling will tend to push the mode of the likelihood function to the right. For individual firms whose probability is near some cut-off threshold there will be a tendency to misclassify them as targets when they are not.

This argument is important to the Bayesian method as well as it argues that the estimates will be a result of something other than the likelihood principle, that is the selection method. Palepu reports that the justification for state based sampling is that it optimizes the amount of information extracted from the data regarding the parameter.

The argument by Palepu is that the high prediction quality is an artifact of the bias being introduced and that when the bias is removed the information quality becomes so low as to have little predictive value. State based methods, as reported prior to Palepu, likely violate the conditionality principle and as a result the likelihood principle. In effect the experiment chosen is not independent of the parameter.

The Likelihood Principle is derived from two simpler ones, the Sufficiency Principle and the Conditionality Principle. The Conditionality Principle can be stated as:

**Definition 6** (Conditionality Principle). If an experiment concerning inference about  $\theta$  is chosen from a collection of possible experiments, independently of  $\theta$ , then any experiment not chosen is irrelevant to the inference.

The argument of Palepu could be restated in Bayesian terms as follows:

1. Mergers are rare events, random sampling would collect very few examples from the population as very few exist.
2. Increasing the proportion of observed mergers in the sample that represents observed mergers to maximize the information about mergers implies that the experiment was constructed in the manner that it was because the event is rare.

3. The method chosen then becomes dependent on the true value of the parameter. More common events would not use this experimental form.
4. As the Conditionality Principle does not hold, by assumption, it cannot be shown that the likelihood principle is the sole source of inference.

That said, the likelihood principle may not be the firmest foundation upon which to posit a Bayesian objection. There are several objections to the Likelihood principle, two of which occur in this dissertation. The first is that the use of risk functions for Bayesian decision theory violate the Likelihood principle, the second is that the Cauchy distribution lacks a sufficient statistic and so for the overall purposes of this dissertation the Likelihood Principle is irrelevant, even if true.[13]

A better objection may come from Cox's postulates.[46] Slightly restated from Jaynes, the relevant postulate would be:

The [decision maker] always takes into account all of the evidence it has relevant to a question. It does not arbitrarily ignore some of the information, basing its conclusions only on what remains.

In maximizing the amount of information on targets it has reduced the information on non-targets. The proportion of some occurrence is relevant information. Intentionally altering the proportions to collect more relative information about targets reduces information about the complementary group.

The second concern is not a Bayesian concern. Bayesian inference is built around the posterior distribution, assuming there is no objection to the data itself then bias is not an inherent concern of Bayesian thinking. A potential objection to Palepu is that with the large number of independent variables and the form of the inference that Palepu runs afoul of Stein's Paradox. Although not an issue in Bayesian methods with a proper prior, Stein's Paradox is an observation that the unbiased estimator can often be stochastically dominated by a biased estimator.[95] Although it is a concern of both Bayesian and Frequentist methods in how the sample was formed, the high degree of prediction may in fact be created by the introduced bias. Nonetheless, it makes testing impossible using conventional means even if it were later found to be valid.

The second objection is foundational to any decision theoretic perspective. This objection is addressed here through the use of Bayesian decision theory as a framework to solve the problem.

Indeed, if this objection were taken more seriously one would arrive at different principle for institutional portfolio selection. Palepu's paper isn't just an objection to merger estimation methods but to any arbitrary decision system that ignores cost or utility functions.

In addition to providing methodological objections Palepu did empirical tests of models then extent in the academic literature. In particular he tested six hypothesis using ten variables. The finding was that the variables and the models had little explanatory power.

This observation was from the false positive/false negative perspective. Palepu describes well the intent of these models, to predict which firm is going to be a target. The goal is to make excess returns by purchasing the firms early and selling them when the bid runs up. This is not the goal in this dissertation.

It is irrelevant to this dissertation to determine which firms will merge, rather it is the probability of merger that is of concern. An option contract that has an obligation to deliver 100 shares of a firm that merges out of existence tends to be converted into an obligation to deliver the new firm's shares at some ratio. As such, the identification problem which dominates the literature is irrelevant here. Rather, the goal is to get a low noise probability estimate.

Likewise, it is irrelevant here to determine which firms are likely acquiring firms. Such events are simply part of the ordinary management of a going concern and do not impact the terms of the option contracts. Further, the sample used by Pelepu was restricted both by industry and by time period so generalizations from significant variables need to proceed with caution. Generally, his findings were that large firms buy small firms, efficient firms buy inefficient firms, leverage is a predictor of merger and that industry specific factors matter.[75] Thirty years later these findings were echoed by Liu and Qiu in a larger cross-border study of mergers.[58]

While their variables were different and they did not test leverage but did test technology development, the overall findings were about the same with an important caveat. Liu and Qiu were not concerned with determining which firms

would merge but rather the before and after impact of mergers and acquisitions on the participants. Nonetheless the variables that were significant in terms of impact were also predictors in Palepu implying that impacted variables are also potentially the predictors. Further tests of differences between targets and non-targets showed differentiation along these variables for the use in regression on impact.[58]

Similar findings are echoed in Beccalli and Frantz on 777 deals involving European Union acquirers in the banking industry. Although the article is restricted to banking, it too found large firms buy small firms, efficient firms buy inefficient ones and that leverage matters.[11] Of importance the leverage impact was different, implying that financial institutions are different. Financial institutions are intensely leveraged and so simple comparisons between banks and industrial firms are not possible.

Jensen places Palepu's findings in the framework of the Cash Flow Theory of Mergers, that is that managers and shareholders have conflicting incentives and so firms seek targets to consume their excess cash flows from firms that benefit managers via their compensation structure.[47] Griffin and Wiggins reframe Jensen's article more formally into agency theory by providing an economic model.[41]

Extensions to the Palepu article include an article by Ambrose and Megginson to include data sets not available to this dissertation but concern ownership structure and agency issues.[4] Billett parses out the leverage effect by using credit ratings to measure the bankruptcy risk, bringing bankruptcy into the equation.[14] Agrawal appears to be the first to be concerned with takeover probability and extends Palepu to decompose the impact of a takeover threat on managerial compensation.[2] Astebro and Winter extend Palepu and Altman to include trinomial estimation of outcomes, bankruptcy, survival and merger.[8]

This framing of merger risk in terms of managerial conflicts with shareholders greatly reduces the scope of variables required to be observed in a study. The only real limitation in the merger literature is the nearly universal use of probit/logit models to estimate target risk of takeover. This contrasts sharply with the bankruptcy literature which explores an extensive range of methods.

## 1.6 Conclusion

The history of modeling used here is to determine what has been done in the past and the consequences of making different modeling choices. An important observation about science in general is that many elegant models about how the world could work exist in every discipline. Empiricism disciplines the various endeavors of science by excluding those not supported in the data. This focus on the empirical is the fundamental distinction between the humanities and the sciences. Statistical methods, as an extension of rhetoric, form the key difference; a model is nice, but a model with data is science.

## Part III

# Background Statistics

# THE USE OF BAYESIAN STATISTICS

A basic review of economic literature shows that the use of Bayesian statistics occurs in a very small percentage of research articles. Although some elements of this dissertation use frequency based statistics in addition to Bayesian statistics for completeness, the decision to primarily use Bayesian methods was made for several reasons.

The first reason is that in some portions of this work, no admissible frequency based solution exists while remaining consistent with economic theory. A decision rule is *admissible* if no other statistical rule dominates it. The most well known example of a statistical rule that is inadmissible for the purposes of making decisions is in using the method of ordinary least squares when there are three or more independent variables.[95] The James-Stein estimator stochastically dominates least squares in that case, where the method of ordinary least squares would be applicable.

This leads to a basic question of admissibility. Admissibility is an important concept in Frequentist statistics, but less so in Bayesian statistics as Wald shows that all admissible rules are either Bayesian rules or the limit form of some Bayesian rule. Under mild conditions, a Bayesian solution is always admissible.

The second reason is that not all hypothesis in this study are binary. As

Frequentist hypothesis are restricted to a binary form, this makes anything other than a binary question difficult at best.

The third reason is the ability to continually update parameter information as more data is acquired. Had a frequency based solution been chosen, parameter would have been estimated in a training set and tested over a validation set. In a Bayesian framework, the testing data also serves as training data, once they enter into the historical set. This permits an increased quality of estimation as time passes as each point of data updates posterior beliefs.

One particular weakness of this methodology, however, is the loss of a guarantee against false positives. Frequency based solutions have the virtue of providing a worst case guaranteed coverage against the possibility of false positives. For a trading desk, this could be material.

A false positive could commit significant trading resources. The issue is over the difference of the coverage between the methods. Bayesian methods do not automatically provide a protection level against false positives and hence against the cost of trading on false signals. An admissible Frequentist procedure would also permit a specification of the level of protection for a decision maker against false signals. Simultaneously, by choosing some level of significance,  $\alpha$ , false negatives are minimized.

## 2.1 Bayesian Versus Frequency Based Models

The dissertation uses a combination of Bayesian and Frequentist thinking. The necessary theorems are Frequentist in origin, but the better set of inference tools for this class of problem are Bayesian. This forces careful use of both schools of thought as they are often incompatible.

The dissertation seeks to show that the outcomes are independent of the school of thought employed. A quote from Egon Pearson seems appropriate

Controversies in the field of mathematical statistics seem largely to have arisen because statisticians have been unable to agree on how theory is to provide, in terms of probability statements, the numerical measures most helpful to those who have to draw conclusions from observational data. We are concerned here with the ways in which

mathematical theory may be put, as it were, into gear with the common processes of rational thought.[77]

It is incorrect to think that Bayesian and Frequency based models are different ways of solving the same problem. Rather they are ways of using the same data to solve *different* problems. It is often true, however, that there are no numerical differences in their estimates. The differences are then interpretive and not numerical.

For most of the 250 years of Bayesian statistics, it was called the method of inverse probability.[34] The reason is that inference was of the form,  $\Pr(\theta|\mathbf{y})$ , where  $\theta$  is a parameter or vector of parameters of interest and  $\mathbf{y}$  is the data. As such, you were inferring causes from effects as you could see the effect in the data, but could not see the causes. It was a statistical form of solving the inverse problems so common in economics.

This structural form requires that the data are given as true and therefore are fixed points and not random instantiations of a sample space. Conversely, the parameters are random variables, or more precisely beliefs about the parameters are random variables.

An hypothesis is considered a belief, so the idea that  $\mu > 5$  is one of many possible beliefs about  $\mu$ . Inference about that belief would be shown as  $\Pr(\mu > 5|\mathbf{y})$ . Beliefs about  $\mu$  change as more information arrives. So as the data set goes from  $\mathbf{y}$  to  $\mathbf{y}'$  the belief about  $\mu > 5$  goes from  $\Pr(\mu > 5|\mathbf{y})$  to  $\Pr(\mu > 5|\mathbf{y}')$ . This forces a necessarily subjective view of probability, as different viewers have access to different information. This leads to epistemic probabilities, something quite removed from the Neyman-Pearson concept of aleatory probabilities.

Frequentist, or frequency based statistics, are modeled on the long run probabilities of some event occurring. For this methodology to be used, it implies that the long run model can be known. Rather than look at past information and testing new information given prior information, frequency based measures look at the long run model and asks, "what is the probability the data looks as it does given the model is true?" That is to say  $\Pr(\mathbf{y}|\theta)$ .

In frequency based statistics, the parameters are fixed points and the data is considered random. This is the very opposite of Bayesian inference. As such, an hypothesis is true or false. It is a fixed point and cannot have probabilities of

truth or falsehood assigned to it. The data, on the other hand, are not fixed and so probability statements can be made about the likelihood of observing the data given the fixed parameter.

Using Frequentist statistics, the hypothesis  $\mu > 5$  is either true or it is false. Usually, however, if the real concern is whether or not  $\mu > 5$ , then the uninteresting and complementary hypothesis of  $\mu \leq 5$  is tested instead. Whereas Bayesian tests determine the probability a belief is true, frequency based statistics test the probability the data could not happen given a null hypothesis is true.

This probability is based upon the long run frequencies given the hypothesis and not the data alone. Whereas Bayesian statistics use only the observed data to make decisions, frequency based measures consider the samples that could have been observed according to the model.

These subtle differences can lead to rather sharp differences in the understanding of the same events. Aleatory probabilities are closely related to physical probability in the sense of dice rolls or coin tosses. Bayesian probabilities are subjective and so the tie to physical probability is looser. In a sense, it is one step removed from the physical probabilities, even for dice games.

### 2.1.1 Illustration

As a pragmatic illustration of the difference, imagine two possible dice games under perfect competition for customers.<sup>1</sup> One type of dice game is run by an honest casino and everything is fair. In the other type of dice game, con men and ex-magicians run the same game. The players do not know which type of game they are in. The house takes two die, places them in a cup, shakes them in the cup, and turns the cup upside down with the dice still covered by the cup. Players then wager against the house on whether the sum of the digits is even or odd. The house, through a croupier, rolls the dice, but the player chooses “even” or “odd.” Players pay a cover charge of one dollar in advance and can play all day for one dollar per dice roll.

Even in such a simple model of probability, the contrasts can be quite stark.

<sup>1</sup>The purpose of this illustration is to distinguish statistical schools. As such, a formal economic model is not derived. A continuation of the illustration would include equilibrium cheating and the costs of actors in an economic model. This is ignored.

In a simple sense, the Frequentist methodology has one giant advantage here, the solution will always be unique. The most natural way to approach this question is to have two hypothesis:

1. The casino does not cheat the players
2. The casino does cheat the players.

It is possible to either test the count of the wins versus losses or the percentage of times the house wins versus the house loses. For simplicity of presentation, it is easiest to choose the latter method of percentages. Setting  $\pi_{house}$  as the probability of the house winning, the most logical null hypothesis is  $\pi_{house} \leq .5$ , with the alternative hypothesis being of course  $\pi_{house} > .5$ .

The Bayesian method, however, does not automatically yield a unique answer or set of hypothesis either. The hypothesis could be the same as the frequency based method. It could be an infinite number of hypothesis, where each point on the number line is hypothesized as the true value, that is  $\pi_{house} = i, \forall i \in [0, 1]$ . It could also be any mutually exclusive and exhaustive set of hypothesis that combine intervals and points.

Likewise, the Bayesian method requires the choosing of a prior distribution for the parameter  $\pi_{house}$ . If a flat prior is used, the result will be numerically identical with the frequency based method, provided of course the same hypothesis are used. The difference would be one of interpretation. However, there is a strong economic argument and therefore statistical argument against the uniform prior. The economist is in possession of information from the model.

Competition should drive out cheating that could be detected by non-rigorous methods by casual players as it is costless to change casinos.

So, in the absence of cheating by the croupier in favor of the player, the expected value of  $\pi_{house} > .5$  in perfect competition. Since it is reasonable to believe the house is monitoring for cheating by the croupiers, with maybe a slight chance being present of cheating by croupiers for higher tips, the prior probability distribution for the estimate of  $\pi_{house}$  should be centered slightly to the right of 50%, possibly narrowly distributed and possibly skewed.

In a view quite opposite the Frequentist, this skews the outcome toward the hypothesis “the casino does cheat players,” until enough data comes in to overcome

that, if, of course, there is no cheating.

Even if the prior were centered on 50% with a variance of 25%, this additional information would reduce the required number of observations to reach a conclusion. Indeed, Jaynes shows that in similar situations, the number of required observations may be half that required for the unbiased Frequentist estimate.[46] As each observation bears a potential cost, cutting the required number of estimates in half can be meaningful.

One other difference is what is considered random by the two schools. The Frequentist school would not consider the dice rolls to be random variables as they are fixed points at the time the player calls out “even” or “odd.” Rather it is what is called out by the player that is random and hence it is the matches that are random. They are betting they can match a fixed point.

If there are too few matches, then to some degree of confidence, the result cannot be due to chance and so the fixed point of the null hypothesis is probably being manipulated causally. The Bayesian method, on the other hand, is going from effects to causes and so sees the parameter of wins as uncertain and the matching as fixed points once they occur. The Bayesian sees nothing random in the matches and non-matches that actually happened, they are the result of the parameter in use. What is uncertain is which type of game is being played, and hence the true value of the parameter.

## 2.2 Learning Through Bayesian Thinking

Bayesian methods are used in two ways in this dissertation. The first is to test the probability that a Cauchy distribution better supports the data when compared to the more commonly held normal distribution. The second use is to estimate the price of European style put option contracts with one year maturities. The Bayesian methodology is rather simple, especially when compared to frequency based methodology. Frequency based methodology, if thought through from beginning to end, is really quite complex mathematically. In most cases it is necessary to derive the asymptotically optimal procedure and then calculate sample statistics to represent the data and perform inference on the data, given an hypothesis.

Bayesian methodology is broken up into determining any prior distribution

about uncertain variables, pass the data through the likelihood function, normalize the probabilities to one, perform inference on the parameters of interest. Of these parts, only the determination of the prior has the possibility of being conceptually difficult; although normalizing the data to unity may in practice be the most challenging.

### 2.2.1 Subjectivity

A brief discussion of the subjectivity of statistics is in order. Much of the battle between Frequency and Bayesian based statistics has been over subjectivity. Some partisans have been very clear on their position on subjectivity. At the most extreme on the Bayesian side is probably Leonard Jimmie Savage.[86] He developed what he called personalistic statistics. His argument can be brought down to the idea that all statistical methods are subjective in one way or another and that because Bayesian methods do it in a disciplined way they must be the best way. This argument is echoed strongly by Harry Markowitz.[63]

This argument is philosophically pointless. Although valid as it goes, it depends strictly upon the underlying assumptions. Credible arguments in favor of Frequentist methods are just as valid and just as dependent upon underlying assumptions. This leads to two possible mechanisms to think about which tool to use.

The first mechanism is well understood in economics, that is the argument from preferences and/or costs. Some individuals simply prefer some methods. As it cannot be argued in any rational sense that a person should prefer an apple to an orange, or vice versa, it also makes no sense to argue over a preference for one method over another. The better argument is one from costs.

As most economists are trained nearly exclusively in one methodology over another, it would be costly to change methods even if the other were superior. A valid argument as to which methodology to use is from first principles. That is, “what are you trying to solve?” Solving the probability of observing a set of parameters given the data is not the same thing as the probability of observing the data, given a model. Further, there are times where one method is simply more informative than the other method. There are also times where one method

provides superior rules for decision making than the other.

Two things make Frequentist statistics objective. The first is not their relation to nature, but rather the difficulty of a researcher to tamper with the results given a chosen method and a chosen data set. This is not to say it is impossible to “lie with statistics,” but rather that for a chosen method with a given sample, the outcome of the analysis will result in a unique solution.. That is not true for Bayesian methods. The second element of objectivity comes from their evaluation of the data, given a model of the world. This second element looks objectively at one model alone.

There are two indirect Bayesian answers to these issues. The first is that basic safeguards do exist in the Bayesian methodology to prevent tampering through a careful and maybe crafty choice of prior probability distributions and hypothesis. The second is that Bayesian methods can look at a wide range of models, not simply one. Still, neither of these get to the heart of the problem that the Frequentist methodology is trying to avoid.

At the simplest level, all frequency based methods can be mapped to some Bayesian method in the sense that all calculations will result in the same statistics, ignoring any differences in philosophical understanding. Does this make that subset of Bayesian methods objective? No it does not. The fact that a Bayesian method has good frequentist properties and may indeed be numerically identical in all ways to an unbiased objective set of measurements, does not alter the subjective nature of the Bayesian tool.

A decision to numerically mimic a Frequentist solution is a purely subjective choice. It would be the best choice under the narrow circumstance of true ignorance and it was felt that the best choice of ignorance prior happens to map to a Frequentist solution.

The real goal of the economist should be to map real prior knowledge into the prior distribution. This goal is difficult and has proved elusive. Nonetheless, it can be done. Indeed, the goal of creating a prior distribution should be to accurately map knowledge into numerical calculations, even if that information is vague information.

## 2.2.2 Decisions by Economic Actors

Bayesian analysis is a form of inductive reasoning and as such is incomplete. This is not true for Frequentist statistics. The inferential goal of Frequentist statistics is to falsify some null hypothesis and as such is a statistical form of *modus tollens*. This gives a straight deductive answer to a question of interest, subject to some degree of confidence. Bayesian decision theory, on the other hand, combines traditional game theory with Bayesian analysis. It can only form a version of complete reasoning if it can exhaust the possibilities through the game. In such a case, it becomes a statistical version of *modus ponens*.

To arrive these decisions, actors observe data to form a posterior distribution. The posterior distribution is made up of three parts, the likelihood, the prior distribution and the constant of integration. Of interest to economists is the likelihood and the prior distribution.

The general assumption in mean-variance finance is that of a normally distributed likelihood function and of this dissertation of a Cauchy distributed likelihood function. Although the controversial part of this research is actually regarding the likelihood, it is usually the nature of the distribution of prior beliefs that is controversial.

## 2.2.3 Dealing with Prior Information

### 2.2.3.1 Introduction

Ideally, the distribution used to model prior knowledge would accurately describe an actor's prior knowledge of the parameter in question. In practice it is a noisy representation of prior knowledge. There has been a tendency among authors to divide prior distributions by the amount of knowledge they encode. There is no canonical list for such terms, but they are presented here as the uninformative, the vague, the informative and the highly informative prior distribution. An additional type of prior is included as well, the adversarial prior. The adversarial prior, like Fisher's null hypothesis, is designed to favor views that are in opposition to some hypothesis.

This new prior is a partial acknowledgment of the deep philosophical value of the null hypothesis. This is not a complete conversion to the Frequentist perspec-

tive as that would require placing 100% of the prior belief on one hypothesis.

The influence of Frequentist statistics in the literature is obvious. Competition with Frequentist methods has had a disciplining effect on Bayesian methods. A consequence of the Frequentist concern with unbiased estimators has been felt even in economic practice. When economists have used Bayesian statistics the bias in economics has been for the use of uninformative prior distributions, or ignorance distributions. In many respects this is unfortunate. Ignorance priors do limit the influence of a particular researcher's beliefs on inference from the data, this does throw away two sources of prior knowledge, economic theory and the knowledge of related information.

The controversy over avoiding bias in the outcomes versus including real knowledge is unavoidable. However it is important to remember that to some extent the word choice that happened to be used in the field of statistics and probability happened, in part, for polemic reasons. If the Pearson-Neyman estimator was also the unbiased estimator, then by word choice the others either conform to their model or must publish biased estimators. The same thing is true with a wide range of terms, including such terms as admissibility.

The important thing is to understand the origins of the various controversies and to use the gains from them in science. There is no reason to take sides from these often contentious controversies, rather the job of science is to grab pragmatic tools from the various mathematical systems and understand the consequences of adopting certain axioms.

Consider the relatively simple question of determining the annual bankruptcy rate of publicly traded firms. A firm is bankrupt or it is not. The question of estimation is in practice not trivial, even though this resembles a coin toss problem.

To see how and why this issue is important it can help to look at the problem from a Frequentist perspective. Frequency based methods do not concern themselves with prior information as the data is considered random and so any sample that is large enough and independently gathered should contain the same information. The key to this idea is the idea of large enough. For rare events the required sample size could be large indeed.

It isn't inference about the estimator that is of interest but rather how the estimator is constructed. The maximum likelihood estimator is simply the number

of observed occurrences of bankruptcy divided by the number of trials. If two bankruptcies are observed in 1000 observations then the proposed bankruptcy rate is 0.2%.

Bayesian methodology does not make this question so simple. Do economists have information about the bankruptcy rate, either from theory or from other data?

The simple answer is “yes,” economists do have information about the annual bankruptcy rate embedded in other data that theory says should be important. Is it exact information? The simple answer is “no.” As in all things relating to a prior distribution, it is unlikely that a researcher has precise prior knowledge. Indeed, if a researcher did have this precise prior knowledge then the research would be unnecessary.

This creates several trade-offs for the researcher. The first is model precision. Better prior distributions result in smaller credible intervals. The second is credibility with hostile audiences. The more the prior information influences the outcome, the less certain the result would be the same under differing assumptions. A third issue is the admissibility of the result. Flat prior distributions do not always create an admissible solution. From the Bayesian perspective this is why multi-dimensional ordinary least squares is inadmissible. The Frequentist least squares solution maps to the same solution a Bayesian would receive with a uniform distribution as a prior.

The uniform distribution has a distribution of:

**Definition 7.**

$$\Pr(\mathbf{X}) \propto 1$$

This results in a total mass that goes to infinity as:

$$\int_{-\infty}^{\infty} dX = \infty \quad (2.1)$$

Since the prior does not integrate to one, it isn't guaranteed the posterior distribution will integrate to one. As the number of independent variables becomes three or more, the posterior distribution does not exist if the prior distribution was uniform over the parameter space for linear regression.[39, 76] Using Bayesian

methods then requires conscious trade-offs to be made. Generally this guarantees admissibility, but also can guarantee a dispute over the quality of the research.

### 2.2.3.2 The Informative Prior

The first set of prior distributions to consider are the set of *ignorance* priors. They are so called because they have minimal impact on inference from the data. As the likelihood for bankruptcy estimation can be modeled as a binomial process, a rational prior distribution is the conjugate prior distribution, the beta distribution.

A conjugate prior has the nice property that the prior parameters can be interpreted in terms of observations. The normalized joint distribution is the posterior probability distribution of the parameter. The effect of the prior can be interpreted as though specific observations had been made.

For the binomial distribution there are three generally used ignorance prior distributions. They are the Haldane, the arcsine and the uniform distributions. Each ignorance prior distribution has different properties, even though each can model the state of a lack of knowledge.

The maximum a posteriori estimate using a Haldane prior is computationally the same as the Frequentist maximum likelihood estimator. The illusion would be that this would make the Haldane prior unbiased. The Haldane prior is an improper prior in that it does not integrate to one. Computationally, the prior is represented as  $\frac{1}{p(1-p)}$ . The distribution arose as an attempt to solve a rarely discussed, but important problem in the sciences.

For many problems in the physical sciences the sample size is one. Often the sample itself is destroyed. The classic example is to drop a substance into water to see if it is soluble. This is not a task that is repeated thirty times in order to use a z test. It has no degrees of freedom.

The Haldane prior places infinite mass at zero percent and one hundred percent. Seeing a single observation results in all of the posterior mass being at the observed value. The Haldane prior joint with the binomial distribution becomes an ordinary beta distribution as the sample size increases.<sup>2</sup>

While it is equivalent to adding no observations to the posterior distribution so that all weight in the estimate is from the data, it has a rather peculiar U shape

<sup>2</sup>When looking at improper prior distributions, it is necessary to look at their limiting forms.

where the density is minimized at a fifty percent bankruptcy rate and maximized at either a zero or one hundred percent bankruptcy rate. Although it has no impact on the posterior distribution directly, as it is the equivalent to adding no additional information, it favors very small and very large probabilities and dampens the effect of probabilities near fifty percent.

The arcsine prior is the Jeffreys' prior for the binomial distribution. A prior is a Jeffreys' prior if the results are invariant under reparameterization. For example, if instead of a rate a present value were measured then the resulting posterior calculations would not be impacted. The Jeffrey's prior for the binomial distribution is the square root of the Haldane prior. It differs from the Haldane prior in two ways.

First, it is twice as dense at the fifty percent rate. Second, it is equivalent to observing one half a bankruptcy and one half a survival. As such it is equivalent to adding a total of one observation to the data. The expectation is fifty percent.

The third commonly used ignorance prior is the uniform distribution. First suggested by Laplace and Bayes the distribution assigns equal probability to each possible value of the parameter.[46] It is a proper prior since it is bounded at zero and one in this case. While it is properly bounded, it has the peculiar property of being equivalent to adding one bankruptcy and one survival to the data set. Using the uniform prior is not uninformative even though it provides no information as to the location of the parameter other than it must be within the interval  $[0,1]$ . Further, it biases the expected value of the rate toward the fifty percent point. As the sample size becomes very large the effect is negligible, but it does go to point out that even in complete ignorance that information is provided by the prior.

### 2.2.3.3 The Vague Prior Distribution

The purpose of the vague prior is to encode real but noisy information. Information that is known from theory or other data, but not through direct research. An example more like the bankruptcy problem would be to consider using the triangular distribution to model prior bankruptcy beliefs.

Any reasonable set of economists would agree that the likely bankruptcy rate is smaller rather than larger for public firms. It is improbable that the true long term annual rate of bankruptcy is ninety-nine percent. It is more probable that

it is low. The triangular distribution, constructed as  $\Pr(\theta) \propto 1 - \theta$ , has the nice property that any point to the left is more probable than any point to the right. This is not complete ignorance and has the impact of adding two survivals and one bankruptcy to the final data set. This does bias the rate downward, but does not grant infinite weight to the extreme values as the Haldane prior does.

The vague prior has an impact on the final estimate and on the final inference, but it is very weak. Another example would be to code a normal distribution with the value expected by the researcher, but with a very wide standard deviation.

To understand why this may be of value, consider that the maximum likelihood estimator considers all values equally likely prior to seeing the data. A fork full of green beans could as easily have one million calories as five prior to observation. By using the normal distribution to code an expected value some values become so unlikely that their prior weight is nominal. By reducing the weight on the extremes of the number line, the posterior density estimate is narrower increasing both accuracy and the quality of inference.

#### 2.2.3.4 Informative Prior Distributions

While vague prior distributions encode very little additional information, prior distributions shouldn't be vague when real world information exists about the parameters. It is at this point, where information external to the data begins to seep into calculations of the data that the controversies really begin. This class of prior distributions is called the informative priors.

With an informative prior, pre-existing information impacts the estimates of the parameters and the posterior inference about those parameters. A simple and non-controversial example would be data from prior research. Someone reading a study on the spread between the LIBOR and treasury securities as a predictor of GDP from 1980-2001 could simply acquire data after 2001 and use posterior distribution of the other data as the prior distribution. This would result in no difference in calculation than had the researcher acquired the entire set of data.

Another non-controversial and simple example would be to manage data sets that do not begin at the same time. A restriction in Frequentist regression is that if there are multiple data sets with differing start dates, then regression can only be calculated from the latest start date. This is not true for Bayesian regression.

Consider a research project that intends to perform inference on CRSP, CompuStat, Treasury and Federal Reserve data. CRSP data goes back to 1925, some Treasury and Federal Reserve series are quite recent but economic theory would say people react to prices. Not knowing the value of some type of data does not mean people do not react to it, merely that the economist cannot directly see the effect without an independent variable.

Using relationships between series allows the capture of this hidden effect as uncertainty in the distribution. As new series become available their role in that uncertainty can be conditioned on prior knowledge of the known relationships up until that point. Adding each new series removes uncertainty and improves prediction. Further, rather than perform a short regression on ten years of data, nearly a century of data is at least partially encoded in the prior.

Finally, another non-controversial use of the informative prior is to manage breaks in time series. For a variety of reasons economic time series have periods of missing data, sometimes years of missing data. Frequentist regression would require choosing one set of those broken series to estimate the parameters. Bayesian regression would take the data up to the first break and create a posterior estimate. That posterior distribution would become the prior distribution once the series starts again. This process would continue until the entire data set is used.

It is here that controversy begins. Imagine a researcher who is estimating the bankruptcy rate of public firms for the United States from 1925-2012 and who has also just completed an estimate of the bankruptcy rate of British firms from 1800-2012. The question becomes “can the British data be used to estimate US parameters?”

The English and the American legal traditions are similar in most US jurisdictions, but there are other differences as well. The United Kingdom is geographically small and so the competitive environment is different. Constitutional issues and state ownership of firms differ as well at different times. How much weight should British data provide to American data?

An advantage of Frequentist methodology is that economists can ignore this type of question entirely; Bayesian methods should encode prior knowledge. If the British data is highly representative then it should code into US data as a prior. If the British experience is not representative of the American experience then little

or nothing of the British data should be encoded through the prior. If it is not representative at all, then a flat prior would be preferable.

An alternative is a mixture distribution. Mixing together a normal distribution and a uniform distribution will maintain the center of location from the British data but flatten the prior distribution so that the encoding of British data is weakened. How much should the variance be increased? That is up to the subjective viewpoint of the researcher and any referees.

While Bayesian methods are generally admissible, they do not provide a free lunch alternative to Frequentist methods. They are far from a panacea.

There are other ways to get information into a prior distribution as well. A review of stock market indices provide annualized returns. As indices have survivor effects the returns on an index should be higher than the true return as bankruptcy effects are excluded. The result from theory is that the annual bankruptcy rates must be less than the returns on the indices.

Depending on the time frame long run returns run up into the high teens. It is very safe to assume that less than twenty percent of the firms become bankrupt each year. The question is how to represent this as a prior distribution. It is important not to assign a zero probability to any value as it forces a zero percent probability into the posterior distribution. Humanity has yet to run the course of its existence and so making something impossible by assumption is dangerous.

Nonetheless, annual bankruptcy rates in excess of twenty percent are very improbable, with a much less than one percent chance. If a uniform distribution is assigned over the interval zero to twenty with ninety nine percent of the mass and a triangle distribution over the remaining range with one percent of the mass then we have a proper but arbitrary prior distribution.

The obvious question is “how is this a good idea?” The answer is the unfortunate answer that arbitrary solutions are an element of Bayesian statistics. Even to choose a conjugate prior, where one exists, is a very arbitrary choice. The most likely reason economists prefer Frequentist statistics is that any arbitrary choice that is made is hidden from the user and the arbitrary choice is usually some criteria, such as unbiasedness.

Is there anything that would be less arbitrary? No, but there may be methods that are computationally more convenient. For example, the prior could be set

as the normal distribution with a mean of zero and the 99.99<sup>th</sup> percentile at the highest ten or twenty year annualized return for a broad index.

The difference between the two is computational convenience and the appearance of exactness, even though that upper bound is itself a product of chance. Given a different way of looking at the data over a different time frame the chosen prior would have been different.

Not all the of elements of Bayesian statistics are disquieting. As the data set becomes large, regardless of the prior, the Bayesian solution and the Frequentist solution will converge.[46] There is an improper argument that since they will converge anyway, the choice of the prior does not matter. This should not be used as an excuse to be sloppy. It is true for large sets the impact of the prior is very small. This is even true for moderate size sets, but that is not the point.

The prior, as inconvenient as it is, should represent real information where it exists and ignorance where it does not.

Finally, there is the highly informative prior distribution.

### 2.2.3.5 The Highly Informative Prior

A highly informative distribution is a distribution whose effect swamps the effect of the data. There are several reasons to use a highly informative prior distribution. The first is simply because the data exists. If you have historical data and you do not use it then you are wasting information. The second can be a bit subtle.

Imagine that in some research there is a set of well researched and well understood relationships with extensive data sets. A researcher believes another unresearched variable matters as well, but the set is small. Using a highly informative prior for the parameters to be estimated that are already well understood has the nice property of swamping any idiosyncratic elements unique to the specific data set, while capturing any new information not already in the prior distribution.

The issue with highly informative prior distributions isn't quality of data, it is "why bother?" Adding ten data points to a million data points isn't likely to produce a change in parameter estimates. It is only where they are used to support a broader question that the research makes sense.

The data from the likelihood is conditioned on these prior distributions. Although Bayesian statistics does provide a method of robustness checks, that is not

the concern here, from a game theory perspective.

The concern is in getting from a prior belief, through the data, to a predictive distribution. Having formed prior beliefs into a distribution, the second half of creating parameter estimates from the data is the likelihood function. This is the contentious part of this dissertation.

### 2.2.3.6 The Adversarial Prior

A prior is considered antagonistic to the model to the extent it over-weights or favors beliefs against the proposed model. This is done in a number of ways. First, in some sections a highly informative prior is placed that would tend to discredit the model, *ex ante*. For the model to overcome this prior belief system requires substantially more evidence than a neutral observer would require. In other places, where the variance impacts posterior density testing, the prior variance is treated as substantially higher than would be the case in reality.

Adding additional prior variance makes the model appear less stable than it really is, making it more difficult to overcome adversaries' objections. Further, increasing prior variance reduces the information content of prior empirical knowledge. This adds noise where signal had previously been detected. By analogy, it is like blurring an image.

Favoring the adversary is much like null hypothesis testing, though without the closure created by *modus tollens*. Traditional Bayesian mechanisms have eschewed null hypothesis testing and have instead looked to having no favored hypothesis. While this has the benefit of neutrality, it may not change minds. Traditional null hypothesis testing concedes the argument and then begins, "given that the adversary is correct, what is the probability of observing the actual data?"

In Frequentist null hypothesis testing the concession is total. All tests are given that the null is perfectly true. That cannot function in Bayesian thinking. It cannot be the case that 100% is credited to one hypothesis as this guarantees the posterior will not be impacted by the likelihood. Instead, the only real requirement is that the prior parameters are unreasonable from the perspective of the supporters of the proposed model.

Unreasonable is subjective. It is unlikely that a dyed-in-the-wool creationist will suddenly read a research paper and become a supporter of evolution. It will always

be possible to argue that the prior distributions used were not unreasonable enough. Likewise, a strong supporter of traditional Bayesian thinking would correctly argue that the prior distributions should reflect real prior information and should not be diluted or altered.

There cannot be an answer to this question. At most, a scientific consensus could form that the conclusions are trustworthy or untrustworthy. It is unlikely anything stronger is possible to claim.

## 2.3 A Sidestep Into Bayes Actions

As actors are forward looking, they are concerned with a model of future events, given parameters. In Bayesian statistics this is called the predictive distribution. This is the distribution an actor believes will happen in the future given historical data and any prior knowledge.

The Bayesian predictive distribution can be defined as:

**Definition 8.**

$$\Pr(\mathbf{X}) = \int_{\theta \in \Theta} \Pr(\mathbf{X}|\theta) \Pr(\theta) d\theta$$

where  $\mathbf{X}$  is the future values and  $\theta$  are the parameters of interest.

Bayesian decision theory does not assume a parameter is known opening the question as how to make decisions without fixed point solutions. Bayes action is a tool to supplement decision by building a Bayesian game. Option prices are built on a predictive distribution of possible future prices. In taking expectations on a predictive distribution an implicit cost function is being accepted regarding the parameter estimates. That cost function may or may not be explicitly known to a decision maker.

As an example, an individual with one undergraduate semester in statistics may decide to use ordinary least squares to estimate a relationship between two variables. In doing so, that individual has accepted a quadratic loss function even if they are not aware they are doing so.

The skipped step, up until now, has been how to arrive at an estimate for the center of location,  $\mu$ , and a scale parameter,  $\sigma$ . This reflects the risk of looking through a rear view mirror to predict the future.

It is quite common in American economics to use American data to estimate future American data. In doing so economists ignore events that have happened in other economies such as hyper-inflation. Would the American stock market respond the same way under hyper-inflation as in other economies or would it behave as it did under prior more normal periods? There is a risk that the American economy, going forward into the distant future, will have aspects of other world economies. This implies our population estimates for  $\mu$  and  $\sigma$  bear some risk. This risk may not be fully acknowledged leaving unaccounted for risks in this model.

There have been two approaches here, Bayes action and minimax. Minimax is appealing because it doesn't need the requirement of knowing the probabilities for any state of the world that may happen before they happen.[76] Beginning with Savage this has tended to be discounted in favor of expected utility in Bayesian methodologies.

Formally, the minimax concept seeks to minimize the maximum loss from an incorrect choice of parameter estimates. It can and does happen from time to time that an action which is considered a valid Bayes action is also a valid minimax action.

The virtue of a minimax choice is that it is fundamentally conservative. The game theory equivalent model would be a model where nature moved first and is trying to purposefully deceive actors, the actors then observe their world and act on the prior observations in nature, whereupon nature would pounce. Minimax minimizes the maximum loss from a purposefully deceptive natural world.[76]

It would also minimize the risk from choosing a poor data set to model data on.

Economists cannot know how marginal actors either in the past or going forward are modeling parameters or parameter risk. We can know the normative tools taught in academia, but we cannot know with certainty how they are being implemented in the field.

A possible way for economists, regulators and institutions to sidestep the issue is through using Bayes actions. This isn't a preference choice but rather a result of the concept of admissibility.

Admissibility is more of an issue for Frequentist style statistics than for Bayesian statistics, but it is nonetheless important for Bayesian measures as well. A measure

is admissible if no other measure can stochastically dominate it.

An admissible decision rule has the same property of dominance, but with regard to a risk function. It is also closely related to Pareto optimality.

Following Parmigiani, if a utility function has already been specified then a loss function can be defined as:

**Definition 9.** A loss function  $L$  can be defined:  $u(a(\theta)) = -L_u(\theta, a)$ , where  $(\theta, a) \in (\Theta \times \mathcal{A})$

Given a prior distribution of  $\theta$ ,  $\pi(\theta)$ , a Bayes action,  $a$ , is defined thus:

**Definition 10.** An action  $a^*$  is a Bayes action if:

$$a^* = \arg \min \int_{\Theta} L(\theta, a) \pi(\theta) d\theta$$

where we define:

$$\mathcal{L}_{\pi}(a) = \int_{\Theta} L(\theta, a) \pi(\theta) d\theta$$

as the prior expected loss from an action.

This brings up the question of how to find  $a^*$ . The traditional mechanism in economics is through some formal optimization process. This may be somewhat difficult in some Bayesian problems as there is no guarantee of a unique mode or a connected credible region in which to find an optimal solution.

Wald's solution was to create a decision rule,  $\delta(x)$  with a domain of  $\mathcal{X}$  and a range of  $\mathcal{A}$ . The entire class of decision rules is  $\mathcal{D}$ . The existence of a rule to find a best action, however, then begs the question is there a risk created by the rule? Could an actor choose the wrong rule?

We need a way to define risk:

**Definition 11.** The risk function of a decision rule  $\delta$  is:

$$R(\theta, \delta) = \int_{\mathcal{X}} L(\theta, \delta) f(x|\theta) dx.$$

This risk only considers the predicted distribution. A Frequentist solution would then choose the rule for decision making that minimizes the maximum risk. The challenge with this is that there is no assurance that such a rule is admissible

for the purposes of decision making. A separate proof would have to be constructed to show the minimax rule for the circumstance involved is also an admissible rule.

Note further that the above risk function can be expanded to include the predictive distribution as the predictive distribution is fundamentally the likelihood function given a parameter set. Still, this is not quite an optimal rule yet.

Two further concepts are required, the Bayes risk and the Bayes decision rule.

**Definition 12.** The Bayes risk associated with prior distribution  $\pi$  and decision strategy  $\delta$  is:

$$r(\pi, \delta) = \int_{\Theta} R(\theta, \delta) \pi(\theta) d\theta$$

A Bayes rule that minimizes the Bayes risk is the optimal Bayesian decision rule.

**Definition 13.** A decision rule  $\delta^*$  is Bayes with respect to  $\pi$  if

$$r(\pi, \delta^*) = \inf_{\delta} r(\pi, \delta)$$

Although this gets to a best rule, it doesn't show the best rule is admissible for the purpose of making decisions. To get there two things must be done, first its necessary to define what admissibility means and second its necessary to show that we have not left out any decision rule that could be used.

The first definition we need is that of what it means to be R-better.

**Definition 14.** A decision rule  $\delta$  is called R-better than another decision rule  $\delta'$  if

$$R(\theta, \delta) \leq R(\theta, \delta')$$

and  $R(\theta, \delta) < R(\theta, \delta')$  for some  $\theta$ . It can also be said that  $\delta$  dominates  $\delta'$ .

The definition of admissibility follows as:

**Definition 15.** A decision rule  $\delta$  is admissible if there is no R-better rule.

The next segment, omitted here, is to show that the class,  $\mathcal{D}$ , is a complete class. The concern is that no rule that could be R-better is omitted.

Although the great objection to Bayesian thinking by Frequentist theorists is the prior distribution, the goal of Frequentist theorists is admissibility. Proofs

to the following theorems are omitted as it is the result that is critical for this discussion. For a discussion of Wald's Complete Class Theorem and admissibility in general, see chapter 13 in Jaynes' *Probability Theory*. [46]

**Theorem 1.** *Suppose that every Bayes rule with respect to a prior distribution  $\pi$  has the same risk function, then all these rules are admissible.*

**Theorem 2.** *Any unique Bayes estimator is admissible.*

These two theorems assure any economist, regulator or participant in the market that they can make choices based on a proper prior distribution that reflects true prior knowledge. Because Frequentist rules generally are the equivalent to multiplying a likelihood by a flat improper prior, an area where the two groups can jointly operate are in generalized Bayes rules. Generalized Bayes rules are rules where the prior distribution is improper, that is it does not integrate to one, but the posterior does so at the limit. An example of this is a uniform prior distribution with a univariate normal likelihood.

This overlap is only important for future work. In this work, it matters only that a unique Bayesian solution will, under very mild conditions, always be admissible. As such, since a properly formed posterior distribution on  $\mu$  and  $\sigma$  has no R-better solution possible, the results can be trusted.

Further, even though the mechanism by which the populace arrive at an equilibrium cannot be known, its reasonable to trust the posterior estimates in finding the basin of attractions for prices.

### 2.3.1 The Effect of Prior Information

A peculiar question often haunts Bayesian and Frequentist debates, what to do with prior information. Frequentist methods discard direct usage of prior information, using it only in study design. Prior distributions, as mentioned above, are usually sloppy. Still, this does not settle the debate.

The prior information discussion enters into the Bayesian/Frequentist debate through admissibility as well. Kale shows for distributions in the exponential family that sample statistics are not admissible when prior information exists. [48] Although not shown for other families of distributions the intuition behind this

finding is rather simple. Imagine prior data from a sample with one million data points and a current sample of one hundred data points. Discarding the prior information would result in an unbiased estimate that is also one hundred times noisier, on average, than the joint estimator.

The economist is always in possession of prior information, if nothing else that most things are not supported in the negative real numbers. Further, theory and related data usually narrow the possible range of solutions. The argument would be that no Frequentist estimator is ever admissible for an economist in almost all circumstances. Nonetheless, except where there is real prior knowledge from data it is quite common for this argument to be practically vacuous.

A numerical look at Frequentist and Bayesian estimators on sets with large enough membership will tend to converge and so using one or the other may only have theoretical significance but no practical significance. Still, it cannot be presumed there is no significance, merely that admissibility isn't truly a sufficient single condition to make decisions from without a context, data or an understanding of the prior information.

As Bayes risk resembles Pareto optimality, it is quite likely that a Bayesian decision theoretic equivalent to the Greenwald-Stiglitz theorem probably exists.

### 2.3.2 Impact on Option Pricing

It will be shown in the segment on option pricing that a profit maximizing market maker is engaging in an admissible decision function via profit maximization. If this were generalized, it would imply that entrepreneurs and business owners in general minimize risk by maximizing profit. This makes sense that it minimizes risk to maximize the premium collected for taking a risk.

## 2.4 The Impact of Sufficiency on Prediction

People are neither concerned with historical stock market prices, prior bankruptcies nor the history of mergers. They are concerned with future events and in the prediction of those events. Option contracts, as a form of insurance, depend upon the ability to forecast future liability. Sufficiency, like admissibility, is an important

concept in understanding statistical decision making.

The Cauchy distribution is used extensively throughout this dissertation and as such an understanding of the relationship between the distribution and sufficient statistics is necessary. A sufficient statistic is important in Bayesian thinking because to use any other statistic is to waste information. The Sufficiency Principle, which follows from the Neyman Factorization Theorem, shows this relationship best.

### 2.4.1 Preliminaries

Certain antecedent concepts are necessary first to discuss the Sufficiency Principle. The first is the idea of a statistic. It can be defined as:

**Definition 16.** Given a vector observations,  $\mathbf{x} = \{x_1, x_2, \dots, x_{n-1}, x_n\}$  defined on a sample space,  $\mathbf{X}$ , then some real valued function,  $t$  is said to be a statistic when  $t = t(\mathbf{x})$ .

It is important to note that the function  $t$  adds no additional information that is not already contained in the vector  $\mathbf{x}$ . So that:

$$\Pr(\mathbf{x}|\theta) = \Pr(\mathbf{x}, t|\theta) \quad (2.2)$$

The right hand side can be expanded as:

$$= \Pr(t|\theta) \Pr(\mathbf{x}|t, \theta) \quad (2.3)$$

If  $t$  does not depend upon  $\theta$  then it follows that

$$\Pr(\mathbf{x}|\theta) = \Pr(t|\theta) \Pr(\mathbf{x}|t) \quad (2.4)$$

This leads to a basic definition of a sufficient statistic for a parameter.

**Definition 17.** A statistic  $t$  is said to be a sufficient statistic for  $\theta$  if it does not depend upon  $\theta$ .

The Sufficiency Principle can be derived from the Pearson-Neyman Factorization theorem. A statement of that theorem is:

**Theorem 3** (Neyman Factorization Theorem). *A statistic  $t$  is said to be a sufficient statistic for  $\theta$  given  $\mathbf{x}$  if and only if functions  $f$  and  $g$  exist such that*

$$\Pr(\mathbf{x}|\theta) = f(t, \theta)g(\mathbf{x})$$

where  $t = t(\mathbf{x})$

*Proof.* If  $t$  is a sufficient statistic for  $\theta$  given  $\mathbf{x}$  then we can set the right side of the equation as:

$$f(t, \theta) = \Pr(t|\theta) \text{ and } g(\mathbf{x}) = \Pr(x|t) \quad (2.5)$$

In this direction the theorem follows naturally from the definition and the basic rules of probability.

In the converse direction it is necessary to assume that  $\Pr(\mathbf{x}|\theta)$  is a Lebesgue integrable probability mass function. Integrating both sides over all  $\mathbf{x}$  in  $\mathbf{X}$  such that  $t = t(\mathbf{x})$  a function  $G(t)$  can be created.

$$G(t) = \int_{\mathbf{x}:t=t(\mathbf{x})} g(\mathbf{x})d\mathbf{x} \quad (2.6)$$

Since  $\Pr(x|\theta) = \Pr(x, t|\theta)$  integrating over  $\mathbf{X}$  where  $t = t(\mathbf{x})$ , results in

$$\Pr(t|\theta) = \int_{\mathbf{x}:t=t(\mathbf{x})} \Pr(x, t|\theta)d\mathbf{x} \quad (2.7)$$

So,

$$\Pr(t|\theta) = f(t, \theta)G(t) \quad (2.8)$$

The function  $f(t, \theta)$  can be rewritten as:

$$f(t, \theta) = \frac{\Pr(t|\theta)}{G(t)} \quad (2.9)$$

Choosing any  $\mathbf{x}$  such that  $t = t(\mathbf{x})$  and substituting out  $f(t, \theta)$ , the equation in the theorem becomes

$$\Pr(\mathbf{x}|\theta) = \Pr(t|\theta)\frac{g(\mathbf{x})}{G(t)} \quad (2.10)$$

Rearranging

$$\frac{\Pr(\mathbf{x}|\theta)}{\Pr(t|\theta)} = \frac{g(\mathbf{x})}{G(t)} \quad (2.11)$$

Since

$$\frac{\Pr(\mathbf{x}|\theta)}{\Pr(t|\theta)} = \frac{\Pr(\mathbf{x}, t|\theta)}{\Pr(t|\theta)} = \Pr(\mathbf{x}|t, \theta) \quad (2.12)$$

So

$$\Pr(\mathbf{x}|t, \theta) = \frac{g(\mathbf{x})}{G(t)} \quad (2.13)$$

Since the right hand side does not depend upon  $\theta$  it follows the left hand side does not as well. As such,  $t$  is a sufficient statistic for  $\theta$ .  $\square$

It would be helpful here, both from a Bayesian and a Frequentist perspective to define the likelihood function.

**Definition 18.** The likelihood function,  $l$ , is defined as

$$l(\theta|\mathbf{x}) = \Pr(\mathbf{x}|\theta)$$

This leads to thinking of a statistic in terms of its information content and in particular, the information content of a sufficient statistic. The Sufficiency Principle leads to this discussion with ease.

**Lemma 1** (The Sufficiency Principle). *A statistic  $t$  is sufficient for  $\theta$  given  $\mathbf{x}$  if and only if*

$$l(\theta|\mathbf{x}) \propto l(\theta|t)$$

*whenever  $t = t(\mathbf{x})$  and assuming the constant of proportionality does not depend upon  $\theta$ .*

*Proof.* If  $t$  is sufficient for  $\theta$  given  $\mathbf{x}$  then

$$l(\theta|\mathbf{X}) \propto \Pr(\mathbf{x}|\theta) = \Pr(t|\theta) \Pr(\mathbf{x}|t) \propto \Pr(t|\theta) \propto l(\theta|t) \quad (2.14)$$

Conversely, if

$$l(\theta|\mathbf{x}) \propto l(\theta|t)$$

whenever  $t = t(\mathbf{x})$  then

$$\Pr(\mathbf{x}|\theta) \propto \Pr(\mathbf{x}|t), \quad (2.15)$$

by substitution. Choosing a function  $g(\mathbf{x})$  so that

$$\Pr(\mathbf{x}|\theta) = \Pr(t|\theta)g(\mathbf{x}) \quad (2.16)$$

Invoking Neyman's Factorization Theorem it is proved.  $\square$

Having shown these basic ideas, it is now possible to discuss the information content of sufficient statistics. Theorems showing the above are true for multiple parameters and for discrete distributions are omitted for brevity and because they do not illustrate additional concepts of general importance to economists.

## 2.4.2 Sufficient Statistics and Information

A key point regarding information in both Bayesian and non-Bayesian methods using the likelihood function is that a sufficient statistic regarding a parameter contains all relevant information regarding the parameter. To make this clear it needs to be noted that the posterior probability is proportionate to the likelihood times the prior probability.

**Corollary 1.** *If a statistic  $t$  is sufficient for  $\theta$  given  $\mathbf{x}$  then it follows that*

$$\Pr(\theta|\mathbf{x}) \propto \Pr(\theta|t)$$

*Proof.* If a statistic  $t$  is sufficient for  $\theta$  given  $\mathbf{x}$  then  $l(\theta|\mathbf{x}) \propto l(\theta|t)$ . If  $\Pr(\theta)$  is the prior probability of  $\theta$  then multiplying the likelihood by the prior and substituting yields:

$$\Pr(\mathbf{x}|\theta) \Pr(\theta) \propto \Pr(t|\theta) \Pr(\theta) \quad (2.17)$$

The result follows directly from Bayes' Theorem.  $\square$

This leads to a far more important observation, that inference from a sufficient statistic will be the same as for an entire sample.

### 2.4.2.1 Inference

Normally this section would concern the likelihood ratio due to the agreement among both Bayesians and Frequentists as to its power as a test. Instead the

posterior odds are going to be discussed due to their relationship to decisions. While the likelihood ratio, for simple hypothesis, gives the change in belief given by the data alone, a decision maker would rationally include any prior information as well as the sample. Further, the simpler construction highlights the key element here, that of the sufficient statistic.

**Theorem 4.** *Let  $\Theta'$  and  $\Theta''$  be disjoint regions, possibly closed and each with a neighborhood of points containing more than one point<sup>3</sup>, in closed region  $\Theta$  which is defined as the parameter space for inference about  $\theta$ , a sample  $\mathbf{x} \in \mathbf{X}$ , a posterior density function,  $f_\theta$ , that is Lebesgue integrable and statistic  $t = t(\mathbf{x})$  which is sufficient for  $\theta$  then*

$$\frac{\Pr(\theta \in \Theta'|\mathbf{x})}{\Pr(\theta \in \Theta''|\mathbf{x})} = \frac{\Pr(\theta \in \Theta'|t)}{\Pr(\theta \in \Theta''|t)}$$

*Proof.* If  $f_{\theta|\mathbf{x}}(\theta|\mathbf{x}) = \Pr(\theta|\mathbf{x})$  and  $f_{\theta|t}(\theta|t) = \Pr(\theta|t)$  then  $f_{\theta|\mathbf{x}}(\theta|\mathbf{x}) \propto f_{\theta|t}(\theta|t)$  by substitution. Since they are proportional for all  $\theta \in \Theta$  it follows that a constant  $c$  can be chosen such that  $f_{\theta|\mathbf{x}}(\theta|\mathbf{x}) = cf_{\theta|t}(\theta|t), \forall \theta \in \Theta$ .

The posterior odds of the two simple hypothesis are:

$$\frac{\int_{\theta \in \Theta'} f_{\theta|\mathbf{x}}(\theta|\mathbf{x})d\theta}{\int_{\theta \in \Theta''} f_{\theta|\mathbf{x}}(\theta|\mathbf{x})d\theta} = \frac{c \int_{\theta \in \Theta'} f_{\theta|t}(\theta|t)d\theta}{c \int_{\theta \in \Theta''} f_{\theta|t}(\theta|t)d\theta} \quad (2.18)$$

Since the constants cancel, it follows that the posterior odds ratio given the data is the same as the posterior odds ratio given a sufficient statistic.  $\square$

Note that proof for distributions with multiple unknown sufficient statistics are omitted but the proof is trivial due to marginalization of parameters.

This leads to an important corollary, that the posterior inference from a sample is the same as the posterior inference from a sufficient statistic for the parameter.

**Corollary 2.** *Given the assumptions in Theorem 4,*

$$\Pr(\theta \in \Theta'|\mathbf{x}) = \Pr(\theta \in \Theta'|t)$$

<sup>3</sup>This is to delay discussion of a *sharp* hypothesis.

*Proof.* Odds ratios can always be expressed as  $\frac{1-p}{p}$ . Since both odds are equal, then both numerators are equal to  $1 - p$  while both denominators are equal to  $p$ . Therefore the posterior probabilities of each simple hypothesis are equal.  $\square$

### 2.4.2.2 Prediction

At the beginning of this discussion on sufficiency the concern with prediction was emphasized. Mentioned earlier was the predictive distribution, in particular the probability of observing  $X_{n+1} = x_{n+1}$ . If  $t$  is a sufficient statistic for  $\theta$  then it follows that predictions based on  $t$  will be no different than predictions based on the sample.

**Theorem 5.** *If  $t$  is a sufficient statistic for  $\theta$  then the posterior beliefs about the probability of observing  $X_{n+1} = x_{n+1}$  are the same either using the data to generate the prediction or the sufficient statistic.*

*Proof.* Substituting the posterior distribution into the definition of the predictive distribution the prediction for  $x_{n+1}$  is:

$$\Pr(x_{n+1} = X_{n+1}) = \int_{\theta \in \Theta} \Pr(x_{n+1} = X_{n+1} | \theta) \Pr(\theta | \mathbf{x}) d\theta \quad (2.19)$$

From the corollary it follows that the predictions regarding future data are the same.  $\square$

Out of sample prediction is key to option pricing. Frequentist and Bayesian methods exist to project estimates out of sample. In the best circumstance, where the natural system is well behaved, the method of ordinary least squares is used. The circumstances where this happens are sufficiently common that it is often the first method projective method taught in statistics.

The attraction to the method of ordinary least squares, where all the classical assumptions are met, is that the parameter estimates are sufficient statistics, the method is admissible, it is the minimum variance unbiased estimator and it coincides with the maximum likelihood estimator. Indeed, in the absence of prior information there is a reasonable argument against Bayesian methods in favor of Frequentist methods since Frequentist methods have pre-constructed tests that

warn of violations of assumptions. Although robustness checks through sensitivity analysis exist in Bayesian methods, they are costly in that they do not exist as pre-built tools.

Violations of assumptions have led to a plethora of robust tools, but as the assumptions depart from the best case assumptions the value of ordinary least squares as a tool declines. At the extremes of the well behaved and the poorly behaved system are two statistical distributions, the normal distribution and the Cauchy distribution.

Cauchy was able to show that the Cauchy distribution leads to the failure of the method of ordinary least squares to estimate parameters. The Cauchy distribution lacks the nice properties of the normal.

## 2.5 The Cauchy Distribution and Its Transformations

The Cauchy distribution is problematic for a variety of reasons. It lacks a mean and hence has no variance. There is no sufficient statistic for the location or scale parameters. Most statistical tools are undefined with the Cauchy distribution and hence are inadmissible.

As a consequence of this absence of sufficiency, the maximum likelihood estimator does not produce a sufficient statistic. As the maximum a posteriori estimator is nothing more than the maximum likelihood estimator conditioned on prior information, it also does not produce a sufficient statistic. The prior information can be thought of as the posterior of data and an ignorance prior and as such the maximum a posteriori estimator is the maximum likelihood estimator of the joint estimate of prior and current information.

This does not mean the maximum likelihood estimator is a poor estimator for the parameter, merely that there is still information held in the sample  $\mathbf{x}$  that remains unused. The maximum likelihood estimator and its Bayesian cousin waste information. This leads to two remaining projective methodologies, that of the minimum variance unbiased estimator of the parameters and the Bayesian method of marginalization to construct a predictive distribution.

It should be noted that the maximum likelihood estimator can be conditioned on the scale parameter, as an ancillary statistic to construct intervals and to perform inference. In that case, they are jointly minimally sufficient.[33] Nonetheless, a projection outside the data is not sufficient and wastes information.

Where the Cauchy distribution is present, it is usually operated on with transformations. As such, the properties of using various transformations need discussed.

### 2.5.1 The Symmetric Case

The symmetric case is the true Cauchy distribution. In certain respects its properties are superior to the truncated case in that the population median and the mode are the same. In the population the median and the location parameter are located at the same point. The twenty-fifth and seventy-fifth percentiles mark the points of inflection, the half maximum point and the distance between them is twice the scale parameter. This property makes the order statistics well defined and very useful.

Order statistics are always sufficient statistics. Given these nice properties, order statistics are not only sufficient, but they minimize a linear cost function and hence are admissible, at least assuming the true cost function is linear.

Since real world data is not symmetric this would seem to be incidental, but that is not the case. Of the Frequentist methods, Theil's method of regression, combined with Rothenberg's estimator for the median slope can usually be treated as a symmetric case.[83, 101] Noting that the median minimizes the absolute loss function,

$$L(\theta, a) = |\theta - a| \quad (2.20)$$

it should also be noted that it is an acceptable solution to the all-or-nothing loss function which is minimized by the mode,

$$L(\theta, a) = \begin{cases} 0 & \text{if } |\theta - a| \leq \epsilon \\ 1 & \text{if } |\theta - a| > \epsilon, \end{cases} \quad (2.21)$$

when  $\epsilon$  is sufficiently wide so that the distance between the median and the mode

is small enough then the median approximation satisfies both loss functions.

As Rothenberg's estimator is the minimum variance unbiased estimator of the median and as Theil's method provides an unbiased estimator of the median slope of the set of all possible slopes, Theil's method provides all of the nice properties desired in a Frequentist estimate of the population parameter. Further, any intervals are well defined in Theil.[101]

### 2.5.2 The Truncated Case

The truncated case happens when returns are defined as future value divided by present value. The mode and the median no longer match, the interquartile range is no longer located at the points of inflection. The median no longer is located at the position of the location parameter. In the data set used here, the difference is approximately two percent per annum. To understand the difference, had the growth rate of India over the twentieth century been two percent greater per capita income would be seven time higher by the end of the century and India's per capita income would be around the same scale of Spain or Portugal today.

Though it is a biased estimator, median based inference is still valid. Although the order statistics are shifted from where they would be in the symmetric case, this is basically a different distribution. It follows that order based measures are valid for inference, though not necessarily for finding the basin of attraction.

Since the median is no longer located at  $\mu$  it can be argued that the median estimator is inadmissible, leaving only Bayesian methods. This may not be true. As with articles designed for small samples using quantiles, there may be a way to recover  $\mu$  and  $\sigma$  from the truncated case making clever use of quantiles.[21]

### 2.5.3 Logarithmic Transformation

The logarithmic transformation has two cases. The first case is captured purely by taking the log of the data, the second by running regression on the log of the data. In the first case, where no regression is run, the data will follow the hyperbolic secant distribution:[56]

**Definition 19** (The Hyperbolic Secant Distribution).

$$\Pr(x) = \frac{1}{2} \operatorname{sech} \left( \frac{\pi (x - \mu)}{2\sigma} \right)$$

Regressing on a log-log transformation of the Cauchy distribution results, generally, in a normal distribution as the set becomes very large. As long as the parameter estimate avoids the explosive root problem, the transformation will yield a well behaved equation. More generally, although leptokurtic, the Hyperbolic Secant distribution has finite variance and transformations that can be seen as sums drawn from the hyperbolic secant distribution are well behaved.[10] When carefully constructed the distributions have finite variance the classical central limit theorem holds and so estimators of  $\beta$  should converge to normality.

Further, as many macroeconomic models are built in logarithmic space, this has all the nice properties attributed to the normal distribution without the problems of the Cauchy distribution. It does not solve the problem more commonly seen by financial economists where the concern is in terms of prices and quantities. The logarithmic transformation is distorting in that case as the mean of the logs maps to the median of the underlying distribution, which in the case of the truncated Cauchy distribution has all of the distortion problems listed above.

#### 2.5.4 Bayesian Methods in General

The likelihood function for the truncated and the symmetric case is proportionate over the mutually supported space. The difference is the constant of integration. The form of the Cauchy distribution is:

$$\frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x - \mu)^2}, \quad (2.22)$$

while the form of the truncated distribution is:

$$\frac{2}{\pi + 2 \tan^{-1} \left( \frac{\mu}{\sigma} \right)} \frac{\sigma}{\sigma^2 + (x - \mu)^2} \quad (2.23)$$

For both Frequentists and Bayesians the interquartile range no longer is an unbiased estimator for the shape parameter, but the two points where the values are

at half the maximum value of the likelihood function still have a distance of twice the shape parameter. This fact may be useful to Frequentist estimation.

Except in some very specific applications where the constant of integration impacts the solution, as may happen in some circumstances with Bayes factors, the constant of integration can be dropped.

Since the Bayesian predictive distribution uses all of the data, by construction, both the questions of admissibility and sufficiency are sidestepped entirely.

### 2.5.5 Frequentist Methods

Although the problem of estimators could be evaded entirely using only Bayesian estimators, this fails to resolve three important problems. Indeed, despite all other discussions, these three issues form a key element of the Bayesian and the Frequentist debate and cannot be ignored.

The first goes to the core of the problem of the concept of admissibility. Subject to very mild conditions, Bayesian solutions are admissible. This isn't a plus.

It is a plus for a Frequentist measure, but it isn't a plus for a Bayesian measure. To understand why consider a series of experiments by a Creationist designed to test variety of components of evolution. Assume they are well designed from the viewpoint of biologist both methodologically and in their scope.

Also assume that the null hypothesis is that evolution has no effect, that is, " $H_0 : \mu = 0$ ," for each effect. Assume each t-test rejects Creationism and that the Bayes factors for each test reject Creationism. Also assume that while the Creationist distribution is non-degenerate it is highly prejudiced.

Although neutral tests reject Creationism, it is still possible for the Creationist to reject evolution as the posterior density may still properly include the "no effect" hypothesis. That is a fully admissible solution.

The Frequentist methodology would correctly exclude Creationism as the Bayesian method would for people with only mild biases toward Creationism. In the presence of the Creationist's prior beliefs the Frequentist methods would automatically fail to be admissible, assuming all tests used distributions in the exponential family.

Admissibility came out of the Frequentist school as a criterion for statistics. It is valuable and should not be ignored, but it is quite a limited concept. Indeed

it hides the problem of the prior, as it fails to discriminate between Bayesian solutions.

Consider two possible proper vague prior distributions for some experiment. It is unclear which prior best encodes the information. It is not correct to conclude that since either solution is admissible then it follows that it does not matter which solution is chosen. That defeats common sense and indeed defeats the underlying postulates in the various axiomatic systems of Bayesian thinking.

Admissibility does not permit the rascal to escape the hard work behind the problem by seeking refuge in Bayesian methods. If anything, it makes the it more difficult rather than less as it is a non-criterion for Bayesian methodology.

Bayesian methods are not unbiased methods. Even if the method is inadmissible, most Frequentist methods are unbiased. Inadmissibility, the negative criterion, should exclude only once the reason for inadmissibility has been determined.

The second reason is computational feasibility. Although most Bayesian solutions are mathematically simple, they may not be computationally simple. Although modern computational methods have ameliorated this issue, they have not eliminated them.

The third issue gets to the heart of a divisive issue, the sharp or point null hypothesis. The correct solution to solving a problem of the form  $\Pr(\mathbf{x}|\theta)$  is to use a Frequentist methodology. Their's regression, with slight modification to optimize estimates of the median, has many of the desired nice properties of an estimator in most circumstances. It cannot consider prior knowledge, but otherwise has nice statistical properties and is usually computationally feasible. It can also answer the null hypothesis  $\Pr(\mathbf{x}|\beta = 0)$ . The Bayesian method is painfully challenged by that question.

## 2.5.6 Conclusion

Three statistical concepts are of importance to this work. They follow from the nature of the problem and the data set in use.

In this work, there exists very substantial prior information from a massive data set, there are no sufficient statistics for the Cauchy distribution and the issue of admissibility is evaded through Bayesian marginalization of the posterior over

the parameter space to project onto the future sample space.

Although it cannot, logically, be concluded that the Bayesian method dominates the Frequentist to the point it should be excluded, it does simplify the solution of the problem. To solve the problem from the Frequentist perspective one would solve a very large optimization problem with multiple nuisance parameters. This is challenging with well behaved distributions. Combined with the other elements this favors Bayesian methods over Frequentist methods.

As such, the bias of the document is toward Bayesian methods and away from Frequentist methods. Although this bias is present, for a different problem in economics this bias would not be acceptable.

# CHOOSING A LIKELIHOOD FUNCTION

The debate over heavy or thin tailed distributions has been going on since Mandelbrot first noted that the distribution of returns did not match economic theory.[60] Although the argument is being made that under very mild assumptions, the distribution of returns in both the Bayesian and Frequentist paradigms must converge to a Cauchy distribution, this should not be construed as arguing that returns follow a Cauchy distribution. Rather, in the blackboard economics generally used in finance and economics, returns must converge to a Cauchy distribution. Adding in very simple economic constraints can have a large impact on the distribution observed in nature and also confirm that the Cauchy distribution is in fact a reasonable likelihood function when compared with the normal distribution.

## 3.1 Model Assumptions

This dissertation has a number of relatively simple assumptions that should be non-controversial. In particular, the model assumes that the Böhm-Bawerk and marginalist paradigms, generally accepted for over a hundred years, are valid. The model adopts the mean-variance assumption that future wealth equals current wealth times a reward plus a random shock. This research further generalizes

this and argues that the static model is the same as an auto-regressive of degree one process, without a loss of generality. It assumes that both the Bayesian and the Frequentist models of probability and statistics, when viewed separately, are completely valid understandings of their fields. Finally, it assumes that scientific models have at least two properties; that is that models are mathematically coherent and that measurable inference can be performed on a model.

This last assumption is little more than a reworking of Cox's postulates for a narrow purpose.[46]

### 3.1.1 Difference Equations

Key to understanding the various models is the structure of the equations used to make them. Implicitly or explicitly, the models use difference or differential equations. Stochastic economic models can be divided into three groups: static models, discrete time models and continuous time models. The relationship between discrete and continuous time models is through scale invariance.[30] The relationship between static models and discrete time models in economics comes from the proposition that, subject to a model's assumptions, economic models are statements of general economic principles that hold across time.

For example, if a model contains  $x_1 = f(x_0)$  and  $x_t = f(x_{t-1})$ , then by induction it can be shown that  $x_{t+1} = f(x_t), \forall t \in \mathbb{N}$ . So static models of the form  $\tilde{w} = R\bar{w} + \epsilon$ , where  $\tilde{w}$  is an uncertain future wealth,  $R$  is a parameter, and  $\epsilon$  is a random variable, could be re-written, without a loss of generality, as  $w_{t+1} = Rw_t + \epsilon_{t+1}$ .

The equation  $\tilde{w} = R\bar{w} + \epsilon$ , it should be noted, is the basis of an ill-posed problem as used by economists. Gauss reminds us that it is only in the limiting form of a well posed mathematical process that any real discussion of the properties of  $\tilde{w} = R\bar{w} + \epsilon_{t+1}$  can begin.[46] If  $\tilde{w}$  and  $w_{t+1}$  were not treated as being equivalent constructions, then indeed it would have a most peculiar case.

While it is quite possible to imagine single gambles which have no economic consequence in the future, this is not what is generally discuss in economics. That said, this does not preclude the existence of multiple limiting models. This proof is one such model, but it is believed that it fully encompasses the range of behaviors possible in a mean-variance finance proof.

**Assumption 1.** The equation

$$\tilde{w} = R\bar{w} + \epsilon$$

can be expressed as

$$w_{t+1} = R w_t + \epsilon_{t+1},$$

without loss of generality. In the above equation,  $\epsilon$  is drawn from a distribution with finite variance and is centered on zero. As well,  $\epsilon_t \perp \epsilon_{t+1}, \forall t$ .

### 3.1.2 Böhm-Bawerk Theory

The author has run into a rather unexpected argument in the course of this paper that while purely technical and not reasonable within the context of economics is nonetheless a key element for the existence of heavy tailed distributions. The argument is that finance theories do not explicitly require that the marginal actor is trying to make a profit from investing. Technically, this is true. The assumption is usually implicit.

Regardless, in the late 19<sup>th</sup> and early 20<sup>th</sup> century significant work was done on capital and interest rates; this work underpins all modern thinking. In particular, the work of Eugen von Böhm-Bawerk on the *agio*, or premium, theory of interest rates and the writing of James Bates Clark on marginalism come together in the writings of Irving Fisher and later in J.R. Hicks.[24][35][44][110][111]

Of importance to this paper is the idea of an investor requiring an anticipated premium for deferring consumption. This implies that for Frequentist models, that  $R > 1$  and the center of location of  $R$  for Bayesian models is greater than one.

Showing this is true is rather simple. Ignoring issues of uncertainty for a moment, a utility maximizer will prefer a positive return if the alternative is a zero return on nominal money. Under uncertainty some funds may be maintained in money if there is some minimum level of consumption required in following time period or under strong risk aversion.

Alternatively, one could ask the counter-factual question, “what if the reward for investing was anticipated to be a loss in every period, ignoring shocks?” The capital stock in a finite resource environment would go to zero. This would imply

no spears, no seeds, no machinery. This implies extinction so systemic losses are excluded for the parameter  $R$ . If humans lived in the state  $R = 1$  then while capital could form it couldn't be partitioned to allow for different prices for different risks. As such,  $R > 1$  is the only available option.

**Assumption 2.** The anticipated return for investing by the marginal actor must be positive.

### 3.1.3 How Bayesian and Frequentist Paradigms Affect the Equation

In both models only the vector  $\mathbf{w}_t$  is observable. The error term,  $\epsilon$  and the reward for investing,  $R$ , are unobservable and of course  $w_{t+1}$  is yet to be observed. What differs between the Bayesian and the Frequentist paradigms, is what is a random variable and what is a fixed point.

In the Frequentist model  $R$  is a fixed point. It has a degenerate distribution. The vector  $\mathbf{w}_{t+1}$  and  $\epsilon$  are random variates. Although  $R$  does not have a distribution with density, there is a distribution of  $\hat{R} - R$ . Indeed, these differences are thought of as *errors* as the true value is a fixed point.

In the Bayesian model  $R$  and  $\epsilon$  are random variates and the vector  $\mathbf{w}_t$  is fixed. The future value,  $w_{t+t}$  has not been observed and so remains a random variate until seen.

**Frequentist Assumption 1.** *In the equation,  $w_{t+1} = Rw_t + \epsilon_{t+1}$ ,  $R > 1$*

**Bayesian Assumption 1.** *In the equation,  $w_{t+1} = Rw_t + \epsilon_{t+1}$ , the center of location of  $R$ ,  $\mu_R$ , is greater than one.*

### 3.1.4 Scientific Modelling

A definition of what constitutes a scientific model is necessary here. It seems to require at least two parts. The first part is mathematical coherence. This only requires that the models follow the standard rules of mathematics unless some axiom or postulate is added to create differences. Any standard regularity conditions assumed by economists may be included implicitly. Fundamentally, the connections must be logical and consistent with the rules of mathematics.

The second portion is that the variables and/or parameters of interest are measurable and inference on those parameters is possible. If some aspect of the model could not be measured, then it fails the second criteria.

### 3.1.5 Boundary Conditions

Neither the models of mean-variance finance, nor other economic models with stochastic difference and differential equations tend to include boundary conditions. It is possible that prices could be infinitely negative where a normal distribution is used and there is no upper bound in resources. The consequences of this are not necessarily trivial.

There are two potentially large consequences of boundary conditions being absent.

The first is that frequency based statistics tend to explicitly or implicitly depend upon rank statistics in order to perform significance testing when a Cauchy distribution is present. If the Cauchy distribution is truncated on the left at zero, but the center of location and scale parameters are unknown, then the rank measures are shifted an unknown amount. Many estimators depend upon the median being the center of location. With truncation, the median and the mode no longer match. The mode, as the basin of attraction, is now the center of location.

The second has to do with thin tails and market failure. If one posits that a future budget constraint exists, then there exists a positive probability that the constraint will be to the left of the market clearing price causing a market to fail. This both skews the distribution and thins the tails from the tails expected by a Cauchy distribution. Not accounting for bankruptcy on the left and potential market failure on the right results in a truncated, skewed distribution without finite variance and possibly without known analytic properties.

## 3.2 Returns

One of the large challenges in financial economics has been explaining and modeling the presence of heavy tails in the distribution of returns. While many difficult models have been proposed, they are based on the fit to the data and not on

beliefs about how humans must behave in an economic system. A difficulty in finding a solution is that the Bayesian solution and the frequency based solution are not the same at all.

### 3.2.1 Intuition Behind the Proof

Bayesian statistics are a form of case-based reasoning. Each data point is an individual case and the goal is to extract any relevant information from each data point. This happens through the likelihood function. Looking at the data on a point-by-point basis, the question is whether a natural likelihood function exists for  $R$  from which to extract information.

The intuition behind the proof for the distribution of returns can be constructed from a far simpler method already used by economists; that is, to divide the realized future value by the present value. In this case, since Bayesian methodology permits viewing one data point at a time for information, that process will be adopted here. Given any one observation at an arbitrarily chosen time  $t$  and given the earlier equations, a specific observation  $R_t$  can be thought of thus:

$$R_t \equiv \frac{w_{t+1}}{w_t}, \Pr(w_t \neq 0) = 1 \quad (3.1)$$

$R_t$  is now data and so each value of  $R_t$  is treated as a fixed point. Bayesian statistics has a construction called a predictive distribution; since  $R_{t+1}$  is yet to be seen, it is a random variable until it is observed, drawn from a distribution with a center of location to be defined as  $\mu_R$ . Once observed,  $R_1 \dots R_t$  become fixed points from which inference about  $\mu_R$  can be performed.

For any observation about  $R_t$ , its important to note that:

$$w_{t+1} = \mu_R w_t + \epsilon_{t+1} \quad (3.2)$$

where  $w_t$  is

$$w_t = \mu_R w_{t-1} + \epsilon_t, \quad (3.3)$$

and this resolves to:

$$R_t = \frac{\mu_R^2 w_{t-1} + \mu_R \epsilon_t + \epsilon_{t+1}}{\mu_R w_{t-1} + \epsilon_t} \quad (3.4)$$

This simplifies to

$$R_t = \mu_R + \frac{\epsilon_{t+1}}{\mu_R w_{t-1} + \epsilon_t}. \quad (3.5)$$

Since  $\mu_R w_{t-1}$  is a constant,  $R_t$  is a function of the ratio of two random variables. The question is “what is the distribution of the shock?” If the ubiquitous answer in economics is used, which is that  $\epsilon$  converges to a normal distribution, then by well known theorem the distribution of  $R$  about its center of location across time is a Cauchy distribution.[37, 42]

On the other hand, accepting a basic tenets of mean-variance finance, that of many buyers and sellers and noting the presence of a double auction so that in equilibrium there will be no winner’s curse, it follows that the rational choice is to bid or ask the expected appraisal value. The distribution of expected bid and ask appraisals will, by the central limit theorem, converge to normality as the bids are in fact the expected future sample means of each actor’s distribution of appraisal values. In that case also, the Cauchy distribution will be present for the returns.

Under very mild assumptions, the likelihood for  $R$  should converge to a Cauchy distribution in each static period in a Bayesian framework. This intuition permits the transition from an estimator of  $\hat{R}$  to  $\hat{R}|w_{t+1} = R w_t + \epsilon_{t+1}$ . The best way to do this is to begin with the Frequentist proof by White.

### 3.2.2 The Frequency Based Solution

Frequency based statistics are a form of deductive reasoning. The goal is to create a statistical form of modus tollens. An hypothesis is created and then the data is tested as if the hypothesis were true. If the test rejects the hypothesis, then to some degree of confidence, the hypothesis is false. The concern here is the construction of a test which could falsify an hypothesis.

Noting that  $R$  is a fixed point, the goal is to construct a test which could be based upon an hypothesized  $R$  and an estimator  $\hat{R}$ . White notes that from prior research, the maximum likelihood estimator for  $\hat{R}$  given that  $w_{t+1} = R w_t + \epsilon_{t+1}$  is the least squares estimator, for all possible values of  $R$ . [114] Normalizing the scale parameter to 1, he notes that the limiting distribution of  $\hat{R} - R$  is the Cauchy distribution, where  $\epsilon$  follows any distribution with finite variance and is centered on zero. It is also assumed that  $\epsilon_t \perp \epsilon_{t+1}$ .

### 3.2.3 Impact of White's Frequentist Proof on the Bayesian Likelihood Function

A Bayesian solution could follow directly from White's proof for two reasons. First, the form of the proof has a Bayesian interpretation; but secondly, under suitable regularity conditions the asymptotic posterior can be estimated from the Fisher information and the maximum likelihood estimate.[102]

While the Bayesian method has made use of the method of maximum likelihood since at least Laplace and Gauss, it is used as a special case of the method of maximum a posteriori. Bayesian methods require a prior distribution for the parameters of interest. If that prior distribution is the uniform distribution then the two methods are computationally identical. This is important as it also means the distributions are identical, although White was solving a Frequentist problem. While White was solving a different type of problem, his proof happens to have a Bayesian interpretation.

White solves for the distribution by normalizing the distribution of the difference between the estimated value and the true value of the center of location with the square root of Fisher information. In Bayesian statistics, the square root of Fisher information is known as the Jeffreys' prior.[55] Although the Cauchy distribution has no Jeffreys' prior, the likelihood estimator of  $R$  given the difference equation does have one. For all finite samples of fixed size  $T$ , it is a constant.

The Jeffreys' prior is an uninformative prior that is invariant under transformation of the data. By multiplying the distribution about the estimate by the Jeffreys' prior, it added no information to the posterior distribution and only the information contained in the likelihood function passed into the posterior.

There is a question then about the likelihood function. White's proof indirectly addresses this. In White's proof it is observed that product of the Jeffreys' prior and the distribution of the error maps to the product of the Jeffreys' prior and the distribution of the ratio of two normal random variates. This ratio is shown to converge to a Cauchy distribution. This ratio is the likelihood function.

Effectively what White has shown is that the product of the likelihoods, also known as Bayesian updating, has the same distribution as the ratio distribution of a future value and a present value. Since the product of a series of Cauchy distri-

butions is a Cauchy distribution, and White shows that for  $\epsilon$  of any distribution which admits a mean of zero and finite variance, the distribution of  $R$  about the true value is a Cauchy distribution. The predictive distribution of returns is also a Cauchy distribution.

The question becomes then, “is an uninformative prior reasonable?” At time zero, before humans invented capital there was no information about the value of capital. As time goes to infinity, that value becomes more certain. Since there was no information at time zero about its value, then it is reasonable to use an uninformative prior. As the likelihood function, though not its value, can be assumed to be invariant across time, then it is reasonable to apply a Cauchy likelihood function to the data.

### 3.3 Effect on Current Theory

The effect of the Cauchy distribution on existing theory depends, of course, on what part of theory . For some areas of finance and economics, the use of a mean or a variance was only a convenience and the results would be approximately the same on a distribution free basis. For others, the problems are more extensive.

#### 3.3.1 Mean-Variance Finance

There are three principle normative models in mean-variance finance: the Capital Asset Pricing Model (CAPM), the Arbitrage Pricing Theory (APT), and the Black-Scholes Option Pricing Model and related Itô calculus based methods (OPM). It is simplest to begin the discussion with the CAPM as Black-Scholes can be derived from it. The form of the Black CAPM is the simplest in that it has the fewest number of assumptions.[15] The goal is to choose a portfolio of securities while minimizing the portfolio’s variance by choosing a desired level of return. The dual problem of maximizing return while choosing the variance would have the same mathematical outcome. The form of the Black CAPM is:

$$\min_{\{s'\}} s' \Sigma s \quad (3.6)$$

subject to:

$$\mathbf{s}'\mathbf{1} = 1 \quad (3.7)$$

and

$$E(\mathbf{s}'\mathbf{R} + \mathbf{s}'\mathbf{\Lambda}) = \mu_{portfolio} \quad (3.8)$$

In these equations:  $\mathbf{s}'$ , is the vector of allocations;  $\mathbf{\Sigma}$  is a covariance matrix;  $\mathbf{1}$  is a vector of ones;  $\mathbf{\Lambda}$  is a vector of normally distributed errors; and  $R$  is an unobserved true growth rate. Two implicit assumptions of mean-variance finance are brought out here.

**Assumption 3.** In models of mean-variance finance, an expected return on investment (or alternatively expected reward) exists.

**Assumption 4.** In models of mean-variance finance, a variance of returns exists. For multi-asset models, a positive definite covariance matrix of returns exists.

### 3.3.1.1 Bayesian Interpretation

The Bayesian interpretation of this formulation would have the vector of returns to be drawn from Cauchy distributions. The share of the portfolio for any given asset is not stochastic and as such can be treated as a constant for the purposes of forming the expectation. What does need to be solved is the predictive expectation of  $R_i|w_1 \dots w_t$ , for each asset  $i$ . Given the most general form of the Cauchy distribution, the expected return is:

$$E(R_i|\mu_i, \sigma_i, w_1 \dots w_t) = \int_{-\infty}^{\infty} \frac{R_i \sigma_i}{\pi(\sigma_i^2 + (R_i - \mu_i)^2)} dR_i \quad (3.9)$$

$$= \left[ \frac{\sigma_i \log(\mu_i^2 - R_i \mu_i + \sigma_i^2 + R_i^2) - 2 \tan^{-1} \left( \frac{\mu_i - R_i}{\sigma_i} \right)}{2\pi} \right]_{-\infty}^{\infty} \quad (3.10)$$

$$= \infty - \infty + 0 \quad (3.11)$$

It follows that  $E(R_i|\mu, \sigma_i, w_1 \dots w_t)$  does not exist, for any  $i$ . This contradicts the above assumption that it does exist. The CAPM is false by contradiction.

Similar assumptions about returns are present in the APT and the OPM. Since the mean does not exist, the variance about the mean does not exist. Nothing about

the CAPM is mathematically coherent in Bayesian statistics. Since the math is not valid, it cannot be a valid scientific model.

### 3.3.1.2 The Frequentist Interpretation

The Frequentist interpretation of the same set of equations is quite different. In the Bayesian interpretation, none of the necessary expectations for the model to function exist. In the Frequentist interpretation they must exist as they are non-random fixed points. Anyone can construct a model made up almost entirely of fixed points. The expectation operator only has the effect of getting rid of the diffusion term as the drift term is fixed though unknown. The question isn't "can such a model be constructed," but rather "can the data falsify it?"

At this point, it is important to be careful how to interpret this model of fixed but unknown points. There are a number of dangerous statistical traps to be found in this construction. Consider, for example, "how do people find the equilibrium conditions?" Whereas Bayesian methods could be interpreted as a tool for the search algorithm, Frequentist methods posit finding the equilibrium as true by assumption.

There are two paths possible. One leads to the idea of fiducial statistics and the other to perfect foreknowledge. While fiducial statistics is a largely discredited topic, research on the field still continues.[43] The alternative, perfect foreknowledge has a deus ex machina element to it.

The attempt to construct fiducial statistics by R.A. Fisher was based on a very simple observation. In performing a significance test on an hypothesis, say  $\mu = 5$ , it should be possible to perform a significance test for every value on the real number line, not merely at five. This collection of tests does not end up forming a proper density function. As tempting as fiducial statistics is, it turns out to not be valid.

The mechanism to arrive at the equilibrium is unclear; it only matters that it is assumed that the arrival happens. Although this creates some philosophical discomfort, it is necessary discomfort. It must be posited that the model of fixed points is true. The mechanics of the process remain a mystery.

Noting that  $\hat{R}_i - R_i$  is drawn from a Cauchy distribution and that  $R_i$  is a fixed point, it follows that  $\hat{R}_i$  is drawn from a Cauchy distribution. It was noted earlier

that prior proofs have shown that the maximum likelihood estimator for  $R_i$  is the least squares estimator.

The least squares estimator is the estimator for the expectation for the slope. The algorithmic solution for the least squares estimator represents the effect of the sample on the test. The question, however, is about the ability to perform inference on the CAPM. Can it be shown as true or false?

What are the properties of any significance test of the CAPM (or any standard mean-variance model), given the mathematical properties of the model(s) are strictly true?

As precision is defined as the reciprocal of the variance, one can find the precision of a test by finding its asymptotic variance about a point. For all  $R_i$  the precision of the test for a sample is estimated knowing that  $\hat{R}_i$  is drawn from a Cauchy distribution.

Although a variance is a form of expectation, in order to construct this, the Cauchy principal value will be used instead as no variance about the mean can exist.

$$Var(\hat{R}_i - R_i) = \lim_{c \rightarrow \infty} \int_{R_i - c}^{R_i + c} \frac{\sigma_i}{2\pi} \frac{\hat{R}_i^2}{\sigma_i^2 + (\hat{R}_i - R_i)^2} d\hat{R}_i \quad (3.12)$$

$$\begin{aligned} &= \sigma \left[ R_i \log(R_i^2 - 2R_i\hat{R}_i + \sigma_i^2 + \hat{R}_i^2) + \left( \frac{R_i^2}{\sigma} - \sigma \right) \tan^{-1} \left( \frac{\hat{R}_i - R_i}{\sigma_i} \right) + \hat{R}_i \right]_{-\infty}^{\infty} \\ &= \infty \end{aligned} \quad (3.13)$$

Therefore, at the limit, any significance test is of precision zero even with an infinite amount of data. The CAPM is immeasurable in the Frequentist paradigm. While by construction it must be a valid mathematical model, it is not a valid scientific model as the CAPM and any other mean-variance model cannot be constructed with valid measures as written.

It is important to note that there is a valid methodology when dealing with the Cauchy distribution in both frequency based and Bayesian statistics, but to go to those methods is to assume mean-variance finance is false.

A separate estimation issue occurs when economists estimate the CAPM and related mean-variance finance tools by directly taking market returns, subtracting

the risk-free rate and using that difference to form the standard *beta*-based solution.

If the returns are treated as data, then they become random variates and from this random variate is subtracted a constant, the risk-free rate. As a constant by assumption, the risk-free rate cannot have a distribution associated with it at any time  $t$ . As in the Bayesian intuitive solution, returns will converge to a Cauchy distribution and be translated by an amount equal by the risk free rate. The question then becomes “if we use the least squares method as an approximation, what impact on the interpretation of results should occur?”

Fortunately, this is already answered in the literature. Sen finds that the asymptotic relative efficiency of the method of least squares is zero compared to any median based method.[87] To adopt a median based method is to abandon the mean-variance method. If it is used as an approximation, then anyone using Theil’s method of regression would gain an immediate advantage over the mean based method as Theil’s method, especially if augmented with other median estimation tools, has the highest known efficiency. That being known, it should be possible to form a statistical arbitrage process over mean-variance users and systematically win. Standard economic theory rules that out, so this approximation should be excluded by both statistical theory and economic theory. Why would someone knowingly adopt a perfectly inefficient tool or even a tool which is perfectly inefficient on a relative basis when standard tools exist that are efficient?

Finally, there is the log difference approximation. This one is a bit more challenging to address. There are two reasons to use the logarithmic transformation of prices to arrive at an approximation of return. The first is to linearize the data to make it easier to work with. The other is to use it for reproducibility with older studies. Older studies took the differences in the logs of the prices as an approximation due to poor computing power.

There are two real issues with this latter usage. First, the underlying theory makes no sense in logarithmic space. People do not purchase  $\log(5000 \text{ shares})$  for  $\log(\$5 \text{ per share})$ . Second, using a distorting approximation simply because the last person did so defeats reason. Things do not gain validity simply from tradition or age. The originators of the practice did it from computational necessity. That constraint no longer exists.

The first case, linearizing the data, is a valid goal. Nonetheless, the use the

logarithmic transformation is not problem free. The logarithmic transformation trims the tails so that the distribution is no longer heavy tailed because the reward (or return) on prices is no longer what is being measured. To understand why, note that systematic rewards must be greater than one but also less than  $e$ , the base of natural logarithms. Transformed into logs if  $\frac{p_1}{p_0} = 1.05$  then  $\log(p_1) - \log(p_0) \approx .05$ . If this is systematically true, then the regression estimator of return will be between zero and one. So by the results of Mann and Wald, it follows that returns will converge to normality as the sample size becomes very large.[61]

Unfortunately, it is no longer possible to determine what the coefficients mean. The value of the regression constant is now multiplicative, when theory says it should be near zero and additive. The allocations are no longer allocations and it isn't clear what they have become. Finally,  $\beta$  should map onto Theil's regression if the various components of the stock market are strictly independent, but they are not.

It is not invalid to use the logarithmic transformation, but this doesn't support mean-variance finance either. Indeed, it is somewhat difficult to determine what is being supported. There is an information loss in the logarithmic transformation, but it isn't clear what that implies for human behavior acting in markets.

### 3.3.2 Heavy-Tailed and Econophysics Methods

Although the Cauchy distribution lacks an expectation, certain utility functions will have an expectation. If wealth is drawn from a Cauchy distribution and the utility function is logarithmic utility, then expected wealth is a function of the Bose-Einstein distribution. The indefinite integral for the expected utility of wealth is:

$$\begin{aligned}
 E(U(\tilde{w})) = & \frac{\sigma \left[ -\text{Li}_2\left(\frac{w}{\mu - \sqrt{-\sigma}}\right) + \text{Li}_2\left(\frac{w}{\mu + \sqrt{-\sigma}}\right) \right]}{2\sqrt{-\sigma}} \\
 & + \frac{\sigma \log(w) \left[ \log\left(1 - \frac{w}{\mu + \sqrt{-\sigma}}\right) - \log\left(1 + \frac{w}{\sqrt{-\sigma} - \mu}\right) \right]}{2\sqrt{-\sigma}} \\
 & + \text{constant}
 \end{aligned} \tag{3.14}$$

The  $\text{Li}_2$  operator is the dilogarithm, a special case of the polylogarithm. Although the polylogarithm can be defined as a series, it can also be defined as the Bose-Einstein distribution divided by the gamma function. This would bring equity securities into Bose-Einstein statistics by simply solving the above problem with reference to logarithmic utility. If the same problem were solved using zero as the lower bound for wealth and infinity as the upper bound the problem simplifies to:

$$E(U(\tilde{w})) = \frac{1}{4}i\sqrt{\sigma} \left[ \log^2 \left( -\frac{1}{\mu + i\sqrt{\sigma}} \right) - \log^2 \left( \frac{1}{-\mu + i\sqrt{\sigma}} \right) \right] \quad (3.15)$$

Another important utility function has been exponential utility. The indefinite integral for the expected utility of wealth, where  $U(\tilde{w}) = -e^{-\alpha w}$  is:

$$E(U(\tilde{w})) = -\frac{i(e^{2i\alpha\sigma}\text{Ei}(\alpha(-w + \mu - i\sigma)) - \text{Ei}(\alpha(-w + \mu + i\sigma)))}{2e^{\alpha(\mu+i\sigma)}} \quad (3.16)$$

The definite integral from zero to infinity is:

$$E(U(\tilde{w})) = \frac{1}{2}i(e^{-\alpha(\mu-i\sigma)}(\text{Ei}(\alpha(\mu - i\sigma)) + \log(-\mu + i\sigma) - \log(\mu - i\sigma)) - e^{-\alpha(\mu+i\sigma)}(\text{Ei}(\alpha(\mu + i\sigma)) - \log(\mu + i\sigma) + \log(-\mu - i\sigma))). \quad (3.17)$$

The Ei operator is the exponential integral operator used in neutron transfer and interstellar heat problems. Very quickly, simple models of *rational expectations* turn into deep physics problems.

Alternatively, there is the observation from behavioral finance that the utility function should be concave in gains and convex in losses. Although a complicated model could be constructed, a simplified model has interesting implications. A function that naturally is convex on the left and concave on the right is the arctangent. The arctangent is also the cumulative density function of the Cauchy distribution. Behavioral finance implies that losses are weighted more than gains, but ignoring that for a second, one can note that unweighted arctangential utility is risk neutral.

Unfortunately, there isn't a known analytic solution for a general form of expected arctangential utility, but there is one if the utility of wealth is centered on

$\mu$ . In that case, if utility is:

$$U(\tilde{w}) = \alpha \tan^{-1} \left( \frac{w - \mu}{\sigma} \right), \quad (3.18)$$

where  $\alpha$  is a weighting over some segment of the function. This allows for piecewise integration to meet the needs of behavioral finance. Under this set of assumptions, the indefinite integral for expected arctangential utility becomes:

$$E(U(\tilde{w})) = \alpha \left[ \tan^{-1} \left( \frac{w - \mu}{\sigma} \right) \right]^2 + constant \quad (3.19)$$

This is a very simple mathematical function with the interesting property that it is the square of the cumulative density function. In all three of the above cases with definite integrals, it should be possible to construct allocation models using the Envelope Theorem as they are all functions of the portfolio mode and probable error.

What may not be obvious is that the Cauchy distribution is intimately linked to complex numbers, the logarithm and the trigonometric functions. Another aspect that may not be apparent is the absence of a Taylor expansion. Because the cumulative normal distribution lacks an analytic form, it is common in economics to perform estimates around a point. This is not an issue for the Cauchy distribution, but as the Cauchy distribution lacks moments, if a Taylor expansion were needed, it does not exist. These seemingly mild changes have significant consequences for standard modeling tools.

Heavy tailed studies have generally discarded the Cauchy distribution in empirical studies since the broader class of four parameter stable distributions provides a better fit and also have no defined variance as implemented. This gives the possibility to explain certain elements of the Keynesian/Classical split.

Consider a purchase on January 3<sup>rd</sup>, 2011 of 100 shares of IBM stock at \$147.50 per share from a family's endowment of cash. The changes in prices seen in the market are changes in the endowments in the budget constraint of the various households and firms as time unwinds. One year later, the position is closed at \$187.00 per share. The cash is moved from the endowment of cash of another household or firm to the current budget constraint of the family in question.

Decisions about what to do with the funds are part of a constrained optimization problem for the household. The question in forming the model by the economist is “what does the model need to do?” A choice of distributions then determines the model, or alternatively, choosing a model determines the nature of the possible distribution in use.

Assuming that the two models are rough equivalents in the sense that they both map onto the observed sample of returns. The two models can be treated as equivalent in that it isn't obvious one model is better than another. One model features a possible set of returns that are drawn from the best fit four parameter stable distribution. The other model features a possible set of returns from a mixture distribution. Both models are supported only on the non-negative real numbers to allow for bankruptcy. The mixture model is a mixture of a Cauchy distribution, as above, with a distribution for constraints on the budget constraint.

Although not directly observable, external constraints determine the frontier of the family's budget. As an example, bank reserve requirements, legal lending limits as a function of bank capital, prudential regulation and loss reserve requirements all play a role in the limitations on the capacity of the family to access liquidity.

The model with the stable distribution implicitly has no form of borrowing constraint. It is conceptually possible to borrow infinite sums. The distribution is skewed, but the sources of skew are not part of the model. Had those sources been separately modeled, then the distribution of returns, subject to those sources, would become a symmetric Cauchy distribution. Skew in the data warns of the possible existence of information not accounted for in the model.

Now consider a relatively simple model that includes planetary product as well as a constraint on what portion of planetary income could be spent on investment activities. With  $Y$  being planetary product,  $\gamma$  the growth rate, and  $\varepsilon$  a normal error, let the difference equation for planetary product be:

$$Y_{t+1} = \gamma Y_t + \varepsilon_{t+1}, \gamma > 1 \quad (3.20)$$

Let the same equation of value persist,  $w_{t+1} = R w_t + \epsilon_{t+1}$ , but with an added constraint:

$$w_{t+1} \leq \alpha Y_{t+1}, 0 < \alpha < 1 \quad (3.21)$$

Positing that  $\alpha$  is a prudential or political constraint that limits the external costs from over investment. In this model there is no money supply, although it would result in a similar outcome if a money supply were used. In this circumstance,  $\alpha$  is a non-market constraint on the budget constraint.

From Bayes law it follows:

$$\Pr(w_{t+1}|w_{t+1} \leq \alpha Y_{t+1}) \propto \Pr(w_{t+1} \leq \alpha Y_{t+1}|w_{t+1}) \Pr(w_{t+1}) \quad (3.22)$$

Since the distribution for  $\alpha Y_{t+1}$  must be a Cauchy distribution, given the assumptions, it follows that the probability of choosing a value for  $w_{t+1}$  such that it is also less than or equal to  $\alpha Y_{t+1}$  is the cumulative density function from 0 to  $w_{t+1}$ . Further, since it is truncated at 0, this probability is:

$$\Pr(w_{t+1} \leq \alpha Y_{t+1}|w_{t+1}) = \frac{\pi - 2 \tan^{-1} \left( \frac{w_{t+1} - \mu_{\alpha Y}}{\sigma_{\alpha Y}} \right)}{\pi + 2 \tan^{-1} \left( \frac{\mu_{\alpha Y}}{\sigma_{\alpha Y}} \right)} \quad (3.23)$$

With the unconditional distribution truncated at zero,  $w_{t+1}$  has the density function:

$$\Pr(w_{t+1}) = \frac{\sigma_w}{\left( \frac{\pi}{2} + \tan^{-1} \left( \frac{\mu_w}{\sigma_w} \right) \right) (\sigma_w^2 + (w_{t+1} - \mu_w)^2)} \quad (3.24)$$

So the density function for  $\Pr(w_{t+1}|w_{t+1} \leq \alpha Y_{t+1})$  is the product of the two terms, divided by the constant of integration. Currently, the constant of integration is unknown. An analytic solution is yet to be found. Nonetheless, numerical methods to estimate it exist.

If the value of physical capital,  $k_{t+1}$  is modeled using the same autoregressive of degree one explosive process as elsewhere, then it will be independent of  $w_{t+1}$ , even if both processes depend upon  $k_t$ . Then its possible to talk about binding constraints on the capital markets not permitting prices to reach a free market clearing price in the short run. In such a model, the present value of cash flows from physical capital, the price of capital and the price of the financial capital representing it should be equal.

The classic Keynesian prescription to lift the constraints would either to be to relax  $\alpha$  or to increase  $Y$ . However, it isn't clear this is the correct solution.

How the constraint is set should matter. All that is posited is that a constraint exists. It may serve prudential goals to protect the broader society. It could also serve purely political goals to protect elected officials and regulators.

When the regulatory constraint is binding, it is quite possible that the liquidity available would prevent the market price of financial capital from equalling the discounted present value of the physical capital in the system.

Since physical capital is a slowly decaying stock, compared to the speed of capital market trades, it can be completely unaffected by capital market errors in the short run. However, if the constraint is systematically binding for some time, the real economy can be impacted as there are two channels through which new physical capital is formed.

The first channel is through the reinvestment of cash flows. The second is through the formation of new capital. Although there are no explicit loans in this model, if one disaggregated the components of  $w_{t+1}$ , then both equity IPO's and loans could be made.

Assuming that both financial capital and physical capital can be purchased by firms. If the yield on financial capital is higher than the yield on physical capital, to some degree of probability, then the capital stock should fall to meet the market value of the capital stock in the capital markets. In the classical model this would be a very desirable response. Actors who over-built would find the market respond adversely and the excess physical capital would depreciate out of existence. The market would adjust on its own. No activity from the constraint setting body would make sense.

In this case, a prudential regulation serves an efficiency purpose.

If the appropriate value of  $\alpha$  is uncertain set then the problem of a binding constraint is multi-fold. If a prudential constraint is reached, should it be relaxed if it is actually prudential?

This triggers two possible cases. If it is believed the value of  $\alpha$  was correctly set then it should not be changed. If the value was incorrectly set, then it should be altered up or down to the appropriate prudential level.

What if the constraint is a political constraint instead of a prudential one, such as maintaining employment? Then it is quite possible the government should expand spending to increase the value of  $Y$  to make the constraint slack in the fol-

lowing period. Alternatively, the terms of the constraint could be relaxed, possibly through open market operations to support the value of capital.

This begs the question of the distribution involved. In the four parameter alpha stable model the budget constraint and constraints on the budget constraint, such as reserve requirements, are implicit. This can only be a classical style model.

In its Keynesian form the above mixture distribution depends upon the structure of such regulation. Of course if the constraint is non-linear, the above distribution would be a poor fit. Another distribution should be modeled.

Although the above is less than a toy model of the economy, it is an attempt to point out that one should not assume distributions into existence. Distributions should follow from the models employed.

### 3.3.3 Regnault and Bachelier

So why did Regnault and Bachelier observe what they observed?[9] They were studying the price movements, over short periods of time, of *rentes*, a fixed income investment. While an investment in stocks has an uncertain future value, an investment in bonds does not. Each cash flow discounted from face value for a fixed rate bond will mature with fixed value, assuming that payments are made as agreed. The risk at time zero is in appraising the probability of a failure to pay. Although there are a number of ways to model that probability, one of the simplest is the normal distribution.

Indeed, with all things except the probability of payment being fixed and certain, the only appraisal risk comes in a failure to estimate bankruptcy risk. While the reinvestment risk for the portfolio of cash flows probably does converge to a Cauchy distribution, each specific bond has an upper bound payment, a lower bound at zero and a probability for each intermediate cash flow. An expected value exists and it is usually associated with a finite period of time, not the unbounded life of the equity of a corporation.

## 3.4 Special Cases—Mergers and Bankruptcy

Bankruptcy and mergers create a special case for the likelihood function. In both cases, the firm ceases to exist. Unlike the perpetual existence case described above, both existential states provide for conditions different from those predicated on perpetual existence. In the case of bankruptcy, the value of the allocation goes to zero. Indeed, given that a future bankruptcy state is true, the return is exactly -100%. For mergers, this also creates special circumstances. Although mergers superficially resemble the continuous case, there are boundary conditions from economic theory which would make them different. As such, the likelihood function of either should be explored.

### 3.4.1 Bankruptcy

A Bayesian consideration of bankruptcy returns would look at the product of probable returns, given bankruptcy, by the probability of bankruptcy. Returns in bankruptcy are conditioned at -100%. The returns collapse to a fixed point multiplied by either a beta distribution or its trinomial Dirichlet extension.

It is important to note that there is no specific reason that bankruptcy must be modeled using the beta distribution. Logit and probit models as well as other approximations should be perfectly valid, if the underlying assumptions can be met. The article by Nwogugu should be consulted first.[70]

### 3.4.2 Mergers

Mergers have certain properties not shared by going concerns. First, acquiring firms would not wish to lose value by purchasing a firm. So they would seek firms where the value of the financial capital is less than or equal to the value of the physical capital. As this should not be the case in equilibrium, these are non-equilibrium transactions. Indeed, it isn't possible to discuss mergers in equilibrium as transaction costs would preclude their existence.

This implies that the center of location for mergers should be greater than the center of location for going concerns, at least before transaction costs are accounted for.

Second, for a variety of reasons it is reasonable to assume there is no defined variance. In the case of a stock for stock merger the returns after the acquisition should converge to the Cauchy distribution with a gain in value for the acquiring firm triggered by the capture of undervalued physical and human capital. It is, however, this period leading up to the merger that is of interest for one year option contracts. A return is still a future value divided by a present value but with the added dimension that the acquired asset should be undervalued relative to its value as physical capital.

Third, mergers take place over time periods generally of less than one year, often much shorter periods than a year. As such, the distribution should be a mixture distribution of a before and an after announcement distribution.

Fourth, from the literature it has been observed that larger firms tend to acquire smaller firms. Abbott has shown, with the exception of the largest firms, that liquidity costs tend to decline with size. The prior model had constructed liquidity costs implicitly into  $R$ . As such, liquidity costs were effectively treated as a constant with differences showing up in the error term. This effect should further shift the distribution to the right as selling shareholders should see an increase in overall liquidity.

#### 3.4.2.1 Focusing on the Knowable

There are a wide variety of types of mergers. There are offers payable in cash, offers payable in the shares of the acquiring firm, combination offers and also offers for property. An example of an offer involving cash, shares and property could be in the acquisition of a restaurant chain.

Assuming the existing shareholder owned the land and buildings as well as the remaining property and equipment, a risk averse buyer may offer ownership in the acquiring firm, but also return the land and buildings to the pre-existing shareholders with cash. This would make the existing shareholders future landlords for the acquiring firm and would also receive dividends from the broader and more diverse acquiring firm. If the restaurant fails the shareholders of the acquired firm still hold the land and would continue to receive the dividends of the acquiring firm, albeit likely reduced due to the failure of a component of the merged firm. If the restaurant continues as a going concern then the landlords get the dividends,

but they also collect rents from the acquiring firm possibly giving them a higher return than had the acquiring firm taken all the risks.

Such a transaction is problematic for this study. There is no way to know such deal terms given the properties of the CRSP database. Further, valuation of the property received isn't possible. A convention is used for non-cash only transactions in valuing the merger. A selling firm valued at \$10 per share being acquired by a firm valued at \$50 per share would be treated, for purposes of calculating return, as a five to one swap of shares. By assumption, the selling shareholders could have marketed whatever cash and property they had and bought the shares of the acquiring firm. That is on the merger day, the selling shareholders taking stock and other property would not accept a loss and would demand approximate parity with a stock for stock transaction.

This may not be true. It stands as a definite limitation on this research, but does not adversely impact option pricing either way as any tests of option prices would be scaled by the same error as the data. A scaling error would be undetectable for the purposes used here.

### 3.4.2.2 Stock for Stock Mergers

It is the overall contention of this dissertation that distributions exist for a reason. That is they should follow for underlying economic reasons and should not be chosen arbitrarily simply because they fit well. Although it may be the case that the true distribution is unknown, research into what distributions should be present should proceed. Although a best fit, but arbitrary, stable distribution would be a valid candidate, such a model lacks interpretive features. It would be difficult, at best, to investigate. Further, generic stable distributions lack closed form solutions and so are difficult to model.

Still, many parameterized stable distributions exist and serve a variety of roles. A good candidate distribution for the mixture of returns would be a mixture of the Landau distribution and the Cauchy distribution.

The Landau distribution is used as a distribution for energy transfers on rare collisions. Similarly, there is an transfer of unrealized value from one set of shareholders to another. Under boundary conditions it is the more general Vavilov distribution. The danger of doing this is that it is reasoning by analogy. Firms are

not elementary particles. Mergers are not collisions.

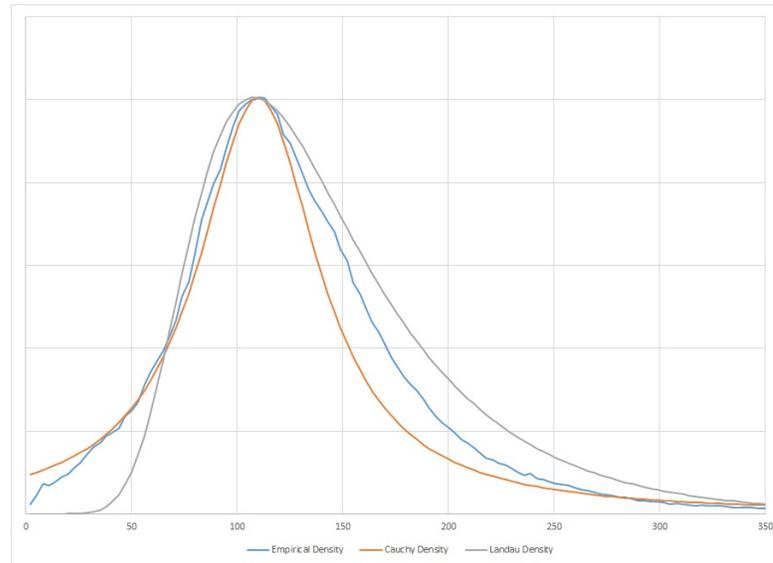
If it is a Landau or Vavilov distribution the important question is why that should be the case. The Landau distribution is derived with respect to the Bethe equation for the change in energy with respect to distance travelled. The Landau distribution can be thought of as a limiting case of the Vavilov distribution. There is nothing in the Bethe equation that should lend itself to financial economics except the concept of a tender offer as having a stopping power over the time series. Caution should be exercised unless it becomes clear such a relationship exists with natural first principle reference to economic behavior.

As with the Cauchy distribution, the proper distribution should be a ratio distribution, but in stock for stock mergers certain special boundary conditions should first exist. It is the mathematical formulation of the boundary condition that has prevented the author from a derivation.<sup>1</sup> The first boundary condition is that the proposed buyer has subjectively judged the market prices far enough below the fair market value of its capital that it is more valuable to buy out the firm than it is to buy additional physical capital. Only the proposed buyer knows their hidden evaluation of the equilibrium valuation. The second boundary is that there is sufficient space to the right of the current market price to accomplish two distinct things. The first is that the buy-out price is far enough below the equilibrium to justify the transaction costs, second that the price is sufficiently large compared to the current price to gain operational control of the firm.

This depends upon the subjective review of the buyer and the subjective alternatives of the seller. If enough potential sellers refuse to sell because the price is too low then the potential buyer loses their evaluation and proposal costs. On the other hand, if the price is large enough to allow the buyer to take control then unwilling sellers are faced with two bad alternatives. The first is to sell at a price that is possibly well below fair market value. The second is to retain the shares as a minority shareholder losing any control of dividends, operations and cash flows. Further, if the firm is no longer traded then the potential seller may own a nearly perfectly illiquid investment that may never again provide a cash flow.

This implies that for one boundary condition the current price is far below the center of location. Of course this is subject to both Type I and Type II errors.

<sup>1</sup>See **Appendix A** for a partial derivation under simplified assumptions



**Figure 3.1.** One Year Returns, Measured in Basis Points, in Stock-for-Stock Mergers 1925-2008

The second boundary condition appears to be contingent upon a threshold function that may consider past prices. If many buyers purchased the firm at prices that are low relative to buyout price then there may be many motivated sellers. On the other hand, if there is a large insider with good knowledge of the fair market value there may be substantial resistance to a merger.

For equity options that will be subject for a stock for stock merger there should be a mixture of two distributions, the proposed Landau distribution and the Cauchy distribution. During the period from the start of the contract until the contractual acceptance of the merger the Landau distribution should be present as the private information of the actors should govern the distribution. From the moment that the merger is considered a “*done deal*” then a Cauchy distribution should be present as the new firm is just a going concern with a proposed infinite life.

### 3.4.2.3 Cash for Stock Mergers

Cash for stock mergers are a special limiting case. Whereas stock for stock provides the selling shareholders a contingent claim, cash for stock provides perfect liquidity. As cash is expensive, a cash purchase should have additional properties.

As with the stock for stock mergers, some stable skew distribution is reasonable, again the Landau distribution is proposed as a candidate distribution, but this is not reasonable once the merger is certain or even very likely. At that point a distribution with finite variance would be preferable. If it is assumed that the rates of return are efficiently priced once perfect liquidation is certain then traditional economic tools should solve this problem. Going back to equation 3.1 note that the future value is now fixed.

For a *single* period discount bond with a total return at time  $t$  of  $R_t^r$  in the paid state and a \$0 return in the loss state provides a simple mechanism to evaluate the bond. If the bond has a probability of no payout equal to  $B$  and a probability of full payout equal to  $1-B$ . The expected utility of wealth becomes  $E[(1-B)U(R_t^R) + BU(0)]$ .

Because the numerator in a single period discount bond is fixed and not subject to variability, the purchase price can be thought of as a discount to a face value which could be normalized to unity. So,

$$R_t^r = \frac{w_T}{w_t} = \frac{1}{e^{-\log(R_t^r)}} \quad (3.25)$$

If it is assumed that actors are concerned with their rate of return a uniform mechanism of comparison is required. The instantaneous rate of return provides a uniform measurement then the concern is with  $\log(R_t^r)$ . This would permit simple multiplication in the exponent to rescale a reward.

The problem of the actors, but not the economist, is to estimate the probability of bankruptcy. The actor needs to calculate a required expected return given a bankruptcy rate.

If it is assumed that each actor makes estimates of price given a measure of risk, denoted by a scale parameter  $\sigma$  that varies from risk to risk and is subjectively determined then the actors could be thought of as solving:

$$\Pr(\log(R^r)|\sigma_i), \forall i \in I \quad (3.26)$$

The economist, on the other hand, usually only sees events after transactions actually happen and almost never get to see the information available to the actors. Whereas the actors are anticipating net returns, economists only see realizations.

The actors are attempting to set prices given risks, but the economist is attempting to estimate risk given prices. While the actor is concerned with an expected total return of  $(1 - B)R_t^r$ , the only concern here is the distribution *given that the bond pays off*. While the actor is concerned with both states, this study only cares about the successful state. Further, errors would have to be estimated with reference to other transactions and not actor information. The economist's measure has its own error distribution. If that error is denoted  $\epsilon$  and it is thought to be small relative to  $\sigma$  and independent of the errors of the actors, then the economist's estimate of what a return should be can be modeled as:

$$\log(R^m) = \log(R^r) + \epsilon, \quad (3.27)$$

where  $R^m$  is the modeled return.

If

$$\Pr(\log(R^r)|\sigma) = \Pr((\log(R^r)|\sigma) \quad (3.28)$$

and

$$q(\epsilon) = \Pr(\epsilon) \quad (3.29)$$

then Landon[46, 51] notes that

$$f(\log(R^m)) = \int p(\log(R^m) - \epsilon|\sigma)q(\epsilon)d\epsilon. \quad (3.30)$$

If the additional condition is imposed on the economist's error term that the expectation of the economist's error is zero, then expanding the integral brings the Fokker-Planck equation, the result of such a derivation has been shown to be the Gaussian distribution.[46, 51] This implies that the distribution of total return for a single period discount bond should be the log-normal distribution. That is to say:

$$\Pr(\log(R^r)) \sim \mathcal{N}(\mu, \sigma^2) \quad (3.31)$$

It follows that prior to announcement that prices should fluctuate with a stable distribution lacking variance, probably the Landau, but should fluctuate as a log-normal distribution following the general acceptance of the merger's likely completion.

It may be useful to note that successful and failed cash-for-stock offers may inform the general question of the necessary boundary to overcome seller resistance.

### 3.5 Conclusion

Standard models assume a very simple, very gentle statistical distribution exists to model option prices. Bankruptcy does not exist, mergers do not impact pricing and variance ends up being the only concern.

Bayesian methods are built around uncertainty rather than errors due to chance. Bayesian distributions are not distributions of shocks, innovations, model errors or chance events. Bayesian distributions are a quantifications of uncertainty. They describe the impact of incomplete information on a decision. Chance effects are contained inside this uncertainty, but the uncertainty is wider than the impact of chance.

The surface argument here is that mean-variance finance is false by contradiction, but the deeper argument is that statistical distributions exist for a reason. Preferably they would never be assumed into existence, but rather determined by the nature of the problem faced by the decision maker.

Of course all of this is of no practical importance if the data support the existing model better than the proposed likelihood functions. It isn't enough to make a reasonable argument, it is necessary to perform empirical analysis. Model adequacy is a question of reasonableness of both reasoning and empirical outcome. Fortunately methods exist to test empirical model adequacy.

When reasoning conflicts, math disciplines the discussion giving three possible outcomes. The first is that the existing models are clearly better than proposed models. Unlike Frequentist methods, Bayesian methods provide a method to weight the adequacy of models. The second is that proposed are clearly inferior to existing models. The third is the dreadful state of being unable to choose among models. In that dreadful state science should go down both paths to explore the reasoning involved and to gather more data.

## Part IV

# Model Validation

# EMPIRICAL MODEL SELECTION AND VALIDATION

To move mathematics from being in the realm of opinion to the realm of science, it is necessary to test logical assertions to determine their validity. Up to this point, everything has been a series of logical assertions. It is fundamental to collect and test data against propositions. The methodology chosen is Bayesian, but subject to some additional qualifications, a Frequentist solution is provided as well. As Bayesian methods tend to be unfamiliar to economists, a general explanation is provided.

## 4.1 Bayesian Hypothesis Testing

Bayesian methods have two nearly equivalent methodologies to solve inference related problems. The methods are either to integrate the posterior density function over the region of each hypothesis or to construct odds ratios from either the posterior density or the likelihood functions.

The advantage of integrating over the posterior distribution is that the posterior probability always contains the exact probability an hypothesis is true given the data and the prior information. Its disadvantage is that it requires a region to integrate over. If the hypothesis can be stated as a region of points then a solutions

exists, but if the hypothesis can only be constructed as individual points then no solution for the integral exists.

Frequentist methods generally handle point or “sharp” hypothesis well. Bayesian methods do not.

The weakness of using a sharp hypothesis becomes apparent as the sample size becomes very large, even for Frequentist methods. As the sample size grows, the power of Frequentist tests against trivial but real departures from the model in the sharp hypothesis have caused critics to note that unless the hypothesized model is the exact true model in nature then all null hypothesis are guaranteed to be false once the sample size becomes large enough.

There are two partial Bayesian solutions to the sharp hypothesis problem. They are the Bayes factor, due to Alan Turing and I.J. Good, and Lindley’s method which in many respects resembles a Pearson-Neyman acceptance region.[46, 55]

Although Lindley’s method superficially resembles the Pearson-Neyman acceptance region, this is only literally the case where a binary hypothesis is mutually exclusive and exhaustive of possibilities. Still, there are important conceptual differences. An analogy would illustrate the differences.

It has become common in medicine to provide a placebo as a control against a medicine that is to be tested. In cases such as this, Fisher’s “no effect” hypothesis, often implemented as  $\mu = 0$ , can break down in large samples. The reason is that no physician really believes that a placebo, which is a sugar pill, has no effect on the human body. Rather the belief is that either the effect is small enough to be unnoticed, uncorrelated with the effects of the drug, or result in psychological effects that swamp any physiological effects created by the sugar. Lindley’s method resembles the Pearson-Neyman acceptance region in that it creates a region around  $\mu = 0$  and determines if the high density region encloses zero. This usually differs from Frequentist methods due to the presence of non-binary hypothesis. Under many circumstances it would suffer from the same problems that Frequentist tests possess.

Lindley’s method, to some extent, evades the sharp hypothesis by encapsulating the sharp point in a region, in essence a region indeterminably different as from if the true value were strictly zero.

The Bayes factor is the ratio of the likelihoods of one hypothesis versus another

hypothesis. Data drawn from a binomial likelihood provides a simple example. Given an hypothesis that  $p = i, \forall i \in [0, 1]$  a question arises as to the relative probability of any two points in the supported region. There are an infinite number of hypothesis, but consider a concrete example were there were two successes viewed and one hundred and ninety eight failures, and two important hypothesized points of  $p = .01$  or  $p = .02$ .

The likelihood ratio would be

$$\frac{.01^2 \times .99^{198}}{.02^2 \times .99^{198}} = 1.866. \quad (4.1)$$

This implies a 1% rate is 1.866 times more probable than a 2% rate. At an odds ratio either greater than 19 or less than  $\frac{1}{19}$  there is some comfort that one hypothesis is substantially more likely than another. In this case, the evidence for either hypothesis over another would be considered pretty weak. Using Fisher's five percent criterion a nineteen to one ratio could be considered significant. The range for this would be the solution to:

$$\frac{.01^2 \times .99^{198}}{p^2(1-p)^{198}} \geq 19 \quad (4.2)$$

The interval probable values for the parameter then becomes (.00093,.03757). Effectively the set of point null hypothesis have been turned into three regions. A region of likely solutions, a region of solutions that are too small to be probable and a region of points too large to be probable given the data. It is also important to note that this region differs from the posterior region with a beta prior distribution, either uniform or Haldane. Under the uniform prior the 95% high density region is (.00309,.0355).

For the purpose of testing a Cauchy model versus a normal model, Bayes factors are used for reasons that will become apparent. Caution should be used in interpreting Bayes factors for non-binary hypothesis as they share many of the issues reported as problems with using p-values for inference.[53]

## 4.2 Bayes factor

Bayes factors take advantage of the structure of Bayes theorem and is an odds ratio rather than a probability. The Bayes factor describes how rational scientists should change their beliefs given only the models and the data. Bayesian analysis, being subjective, roots itself in the subjective views of the scientists allowing scientists to hold a wide variety of beliefs *prior* to seeing the data through the eyes of the models, but transforms that belief numerically into new beliefs.

A highly partisan and fully rational scientist faced with a very large Bayes factor in the direction opposite of their pre-existing beliefs would cease being a partisan scientist. In the case of this research, the Bayes factors are so large as to foreclose any possible debate. Even a scientist that accepted mean-variance finance with a prior probability of 99.9999% probability that it is true would have to reject mean-variance finance overwhelmingly.

The Bayes factor can be defined in two ways. The first way, as the ratio of beliefs before and after seeing the data, is totally subjective. The definition is:

**Definition 20.** The Bayes factor, denoted  $B$ , is

$$B = \frac{p_n \pi_c}{p_c \pi_n},$$

where  $p_n$  is the posterior probability the mean-variance model is true, given the data and prior. The posterior probability the Cauchy model is true is denoted  $p_c$ , given the prior for the Cauchy distribution.  $\pi_n$  and  $\pi_c$  are the prior probabilities for the normal and the Cauchy model being true, respectively.

The advantage of this definition is that it describes how a rational scientist should adjust the odds of their beliefs based on the observations. To provide an example, imagine some scientist was 99% sure some model was better than another model, prior to collecting the data. The prior odds would be  $.99/.01 = 99 : 1$ . If after seeing the data through the eyes of the competing models the posterior belief was 75% belief in the model, then the posterior odds would be  $.75/.25 = 3 : 1$ .

The Bayes factor would be

$$B = \frac{.75 \times .01}{.25 \times .99} = .0\bar{3} \quad (4.3)$$

This constitutes 1:33 odds. This implies that the evidence runs strongly against the original beliefs of the scientist. Note that the scientist still favors the original belief. Why? Because this was only a single experiment and the scientist is basing this not only on the one experiment but a lifetime of experience.

The scientist's beliefs are no longer firm, but one experiment was not enough to sway belief even though the evidence was reasonably strong. On the other hand, an indifferent scientist, one who had not decided on a model and gave both of them even odds would assign a 97% probability of truth to the alternative model. An indifferent scientist should become partisan.

On the surface it would appear that this method of testing an hypothesis between two models would be terribly subjective, until one realizes that the Bayes factor is a constant for a given set of data. The alternative definition of the Bayes factor is strictly data and model dependent.

**Definition 21.** The Bayes factor, denoted  $B$ , is

$$B = \frac{\Pr(\text{data}|\text{model}_1)}{\Pr(\text{data}|\text{model}_2)}$$

Note that both the numerator and the denominator are of the form of a frequency based hypothesis test had an optimization process been included.

The relationship to Bayes theorem can be more readily seen by setting the two definitions equal to each other and rearranging the parts as:

$$\frac{p_n}{p_c} = \frac{\Pr(\text{data}|\text{normal model})\pi_n}{\Pr(\text{data}|\text{Cauchy model})\pi_c} \quad (4.4)$$

Then if both sides are multiplied by the posterior probability of the Cauchy model, you arrive at Bayes theorem:

$$p_n = \frac{\Pr(\text{data}|\text{normal model})\pi_n}{\Pr(\text{data})} \quad (4.5)$$

This second definition is powerful because it gives the odds one model is true over another, given the data, by looking at the ratio of the probabilities the data would appear as it did given that those models were true. The difference between this method and a test such as the Kolmogorov-Smirnov test is that the

Kolmogorov-Smirnov test is complete for purposes of deduction in that one of the two hypothesis must be true. Neither hypothesis need be true for the Bayesian test.

### 4.2.1 Composite vs. Simple Hypothesis

A significant issue missed in the above definitions is the distinction between a simple and a composite hypothesis. The above definitions leave two issues untouched. The first is “how are the model parameters known?” The second issue is all the other possible models that could have been considered but were not, such as using a Poisson jump or skewed stable distributions?

#### 4.2.1.1 Infinite Sets

Frequentist methods have a simple solution to the problem of a set of uncountably many possible mutually exclusive and exhaustive hypothesis by creating a binary hypothesis. To use Fisher’s initial example, if it is assumed to be true that Mendel’s laws have no effect in nature, then the natural alternative hypothesis would be that Mendel’s laws do have an effect in nature. Then it follows from falsification of the null that Mendel’s laws hold.[117]

This excludes all possible models of nature that do not include Mendel’s laws including Darwin’s own explanation of inheritance, Creationism, Intelligent Design and everything else to some degree of confidence.

The problem of formal hypothesis construction and the distinction of Frequentist methods as a truth telling engine versus Bayesian methods as an explanation sorting engine should be important to any economic inquiry. Frequentist tests such as the Kolmogorov-Smirnov test or Lilliefors’ test partition the world into a set of only one possible distribution and a complementary set of all possible other distributions. The complement is a set of infinite membership.

A similar Bayesian solution should not exist as the set of all possible distributions, including mixture distributions, would of necessity result in infinitismals as prior probabilities.

This highlights the incomplete nature of Bayesian reasoning in unbounded and uncountable problems. The Bayesian method must be restricted to at most the set

of existing ideas. Bayesian probabilities are built around beliefs. The Bayesian solution sorts existing explanations. Possibilities still unknown in human imagination are outside the framework of Bayesian inference, but not Frequentist inference.

#### 4.2.1.2 Formal Approach to Hypothesis Testing

A more formal structure is required to discuss this and so each one will be explored. It is necessary to first define some logical assertions.

$S =$  set of all posited statistical distributions and models representing returns. (4.6)

$S_1 =$  Returns converge in probability to a normal distribution. (4.7)

$S_2 =$  Returns converge in probability to a Cauchy distribution. (4.8)

$S_3 = S \wedge \neg S_1 \wedge \neg S_2$  (4.9)

$\theta_1 = \begin{bmatrix} \mu_1 \\ \sigma_1 \end{bmatrix} = \begin{bmatrix} \mu_n \\ \sigma_n \end{bmatrix}$  (4.10)

$\theta_2 = \begin{bmatrix} \mu_2 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} \mu_c \\ \sigma_c \end{bmatrix}$  (4.11)

The data from a random variate  $\tilde{\mathbf{y}}$  is  $\mathbf{y}$ .

Noting that  $\Pr(S) = 1$  it follows that:

$$\Pr(S|\mathbf{y}) = 1, \quad (4.12)$$

regardless of the relationship between  $S$  and  $\tilde{\mathbf{y}}$ . Rearranging, it becomes

$$\Pr(S_1 + S_2 + S_3|\mathbf{y}) = \Pr(S_1|\mathbf{y}) + \Pr(S_2|\mathbf{y}) + \Pr(S_3|\mathbf{y}) = 1. \quad (4.13)$$

Only two of the possible logical propositions are of interest,  $S_1$  and  $S_2$ . Suppose there exists some constant,  $k$ , such that  $\Pr(S_1|\mathbf{y}) + \Pr(S_2|\mathbf{y}) = k \leq 1$ . The odds ratio,

$$\text{Odds}(S_1, S_2|\mathbf{y}) = \frac{\Pr(S_1|\mathbf{y})}{k - \Pr(S_1|\mathbf{y})}, \Pr(S_1|\mathbf{y}) \neq k.$$

As an odds ratio can always be constructed in the form  $\frac{p}{1-p}$ , it then follows that:

$$\text{Odds}(S_1, S_2|\mathbf{y}) = \frac{\Pr(S_1|\mathbf{y})}{k - \Pr(S_1|\mathbf{y})} = \frac{p_1}{1 - p_1}. \quad (4.14)$$

Cross multiplying:

$$\Pr(S_1|\mathbf{y}) - p_1 \Pr(S_1|\mathbf{y}) = p_1 k - p_1 \Pr(S_1|\mathbf{y}) \quad (4.15)$$

$$p_1 = \frac{\Pr(S_1|\mathbf{y})}{k} \quad (4.16)$$

Defining  $p_2$  as

$$p_2 = 1 - p_1 \quad (4.17)$$

rearranging

$$p_2 = \frac{\Pr(S_2|\mathbf{y})}{k} \quad (4.18)$$

$$\frac{p_1}{p_2} = \frac{\Pr(S_1|\mathbf{y})}{\Pr(S_2|\mathbf{y})} \quad (4.19)$$

Since  $p_1 + p_2 = 1$ , the odds ratio is a valid measure of the *relative* probabilities that either model is true.

#### 4.2.1.3 Composite Hypothesis

A composite hypothesis has two unknowns; in this case, the distribution and the parameters are unknown. The parameters of the distribution are a nuisance parameter in that the actual value is irrelevant to the matter at hand. To test two distributions for their probability, it is necessary to derive parameters. The two hypothesis for this are:

$$H_0 : S_1 \text{ is true.} \quad (4.20)$$

$$H_A : S_2 \text{ is true.} \quad (4.21)$$

From Bayes theorem:

$$\Pr(S_1|\mathbf{y}, \boldsymbol{\theta}_1) = \frac{\Pr(\mathbf{y}, \boldsymbol{\theta}_1|S_1) \Pr(S_1)}{\Pr(\mathbf{y})} \quad (4.22)$$

$$\Pr(\mathbf{y}, \boldsymbol{\theta}_1 | S_1) = \Pr(\mathbf{y} | \boldsymbol{\theta}_1, S_1) \Pr(\boldsymbol{\theta}_1 | S_1) \quad (4.23)$$

$$\Pr(\boldsymbol{\theta}_1 | S_1) = \frac{\Pr(\boldsymbol{\theta}_1, S_1)}{\Pr(S_1)} \quad (4.24)$$

The likelihood function for this equation is:

$$\mathcal{L}(S_1, \boldsymbol{\theta}_1 | \mathbf{y}) = \Pr(\mathbf{y} | \boldsymbol{\theta}_1, S_1) \frac{\Pr(\boldsymbol{\theta}_1, S_1)}{\Pr(S_1)} \quad (4.25)$$

The prior distribution is  $\Pr(S_1)$  and the normalizing constant is  $\Pr(\mathbf{y})$ . Of importance, the prior distribution is now included in the likelihood function. Without it, the weight of the evidence would depend entirely upon the data and the model. The prior distribution normalizes the likelihood due to the compound nature of the hypothesis.[55]

With a subscript change, the same is true for inference about  $S_2$ . As they both have the same denominator, the concern is not the exact value, but rather a proportion. While the inclusion of the prior probability into the likelihood function removes its objectivity, it also provides a range of weights times the evidence that must be met to overcome some level of objection. A person 99% sure of mean-variance finance will increase the weight by about one percent on the objective portion of the mean-variance likelihood function, but will increase one hundred fold the weight of the Cauchy objective portion. An indifferent scientist will have only the objective portion to consider as the prior probabilities will cancel to unity.

$$\Pr(S_1 | \mathbf{y}, \boldsymbol{\theta}_1) \propto \Pr(S_1) \mathcal{L}(S_1, \boldsymbol{\theta}_1 | \mathbf{y}) \quad (4.26)$$

$$\Pr(S_2 | \mathbf{y}, \boldsymbol{\theta}_2) \propto \Pr(S_2) \mathcal{L}(S_2, \boldsymbol{\theta}_2 | \mathbf{y}) \quad (4.27)$$

It is not known what parameter is valid, so the parameters are integrated out over the set of all possible values. The factor becomes:

$$B = \frac{\int_{\boldsymbol{\theta}_1 \in \Theta_1} \Pr(\mathbf{y} | \boldsymbol{\theta}_1, S_1) \Pr(\boldsymbol{\theta}_1, S_1) d\boldsymbol{\theta}_1 \Pr(S_2)}{\int_{\boldsymbol{\theta}_2 \in \Theta_2} \Pr(\mathbf{y} | \boldsymbol{\theta}_2, S_2) \Pr(\boldsymbol{\theta}_2, S_2) d\boldsymbol{\theta}_2 \Pr(S_1)} \quad (4.28)$$

The fundamental lesson of this derivation is that a partisan actor should discount favorable evidence as it is obviously true and be deeply disturbed by contrary evidence. Another pragmatic solution exists given the peculiar nature of the data

and the problem at hand.

It is not required to solve the problem in terms of a single parameter. Although only the model is specifically in question, Bayes theorem allows for solving the problem jointly and then to remove the nuisance parameters through marginalization. In particular, an obvious feature of the data is that the mode is clearly one. This is true regardless of the granularity used to measure this. Further, this is observable in the data by the end of the first year of data.

As it is clearly increasing on the left and decreasing on the right and as the distribution is truncated, then it follows that it is harmless to condition both the normal distribution and the Cauchy distribution on a center of location precisely equal to one. This reduces the uncertainty in the posterior distribution and reduces the computational complexity.

Rather than solve for  $\Pr(S_1|\theta_1, \mathbf{y})$  the problem can be solved in the following manner.

$$\Pr(S_1, \theta_1|\mathbf{y}) = \frac{\Pr(\mathbf{y}|S_1, \theta_1) \Pr(S_1, \theta_1)}{\Pr(\mathbf{y})} \quad (4.29)$$

The prior can be decomposed as:

$$\Pr(S_1, \theta_1) = \Pr(\theta_1|S_1) \Pr(S_1) \quad (4.30)$$

As  $\mu$  has been conditioned to be equal to one:

$$\Pr(\mu_n = 1) = 1 \quad (4.31)$$

This simplifies the prior distribution as:

$$\Pr(S_1, \sigma_n) = \Pr(\sigma_n|S_1) \Pr(S_1) \quad (4.32)$$

Two simple conventions for the prior distribution can be acquired from economics and rhetoric. Because both the normal and the Cauchy distribution are symmetric distributions, certain shape parameters are unreasonable due to the implied economic behavior. This very unreasonableness is quite valuable for setting a prior distribution.

The real concern with setting a prior distribution is that the prior could so

influence the posterior distribution as to be able to rig the experiment so that the outcome is determined by the strength of the prior. As the location of the shape parameter is unknown, the goal is to make the prior distribution sufficiently flat over the reasonable range, probably make it flat in the somewhat unreasonable range, and vanishing over the incredible range.

Noting the center is one, a shape parameter of one would imply one-third of all capital transactions result in greater than a one hundred percent return on invested capital under the Cauchy model and one fifth for the normal model. Further, it implies an incredible instability of daily price. Prices would still have one as the basin of attraction, but with prices that oscillate wildly. Further, as the Cauchy distribution has no variance, using it in a prior is perfectly imprecise. So one could set the prior as:

$$\Pr(\sigma_{n,c}|S_{n,c}) = \begin{cases} \frac{2}{2+\pi} & : 0 < \sigma_{n,c} \leq 1 \\ \frac{2}{(2+\pi)(1+(\sigma_{n,c}-1)^2)} & : \sigma_{n,c} > 1 \end{cases} \quad (4.33)$$

The second convention is an extension of Fisher's original inspiration for the Frequentist school of statistics. The null hypothesis should always concede the point trying to be proved as false by assumption. By falsifying the null, the subject of controversy is proven to be true, to some degree of confidence.

A similar concession should be made here. That is the other side should be given extraordinary prior weight, indeed, nearly degenerate prior weight. Nonetheless, there is a scale of probability so small that it isn't meaningful to speak of it in terms of human perception of differences in probability. In order to support this principle, granting a prior relative probability of:

$$\Pr(S_1) = 999,999 \times \Pr(S_2), \Pr(S_1) + \Pr(S_2) \leq 1, \quad (4.34)$$

is very adversarial to the proposed model.

This is eminantly unreasonable and therefore a good test. For a test of two hypothesis to overcome such a profound prejudice effectively excludes all remaining support for the normal distribution. A highly prejudiced prior serves roughly the same function as Fisher's null hypothesis and the use of minimax distributions to evaluate hypothesis.

This should result in the joint estimate:

$$\Pr(S_1, \sigma_n | \mathbf{y}) = \frac{\Pr(\mathbf{y} | S_1, \sigma_n) \Pr(\sigma | S_1) \Pr(S_1)}{\Pr(\mathbf{y})} \quad (4.35)$$

Since the estimator of  $\sigma_{n,c}$  is only of incidental significance to the question at hand, the next step would be to marginalize out the shape parameter.

$$\int_0^\infty \Pr(S_1, \sigma_n | \mathbf{y}) d\sigma_n = \Pr(S_1 | \sigma_n, \mathbf{y}) = \frac{\int_0^\infty \Pr(\mathbf{y} | S_1, \sigma_n) \Pr(\sigma_n | S_1) d\sigma_n \Pr(S_1)}{\Pr(\mathbf{y})} \quad (4.36)$$

This still pulls the prior into the likelihood function, but over the region of interest it is sufficiently flat as to have no material impact on the Bayes factor.

The advantage of this methodology is that it directly considers the set of all possible parameters. If each possible combination of parameters can be thought of as a model, then it is averaging the solution giving trivial weight extremely improbable parameters and giving greatest weight to the most probable parameters. The uncertainty in parameter selection is removed by considering the entire set.

#### 4.2.1.4 Simple Hypothesis

For a simple hypothesis, the only unknown is the statistical distribution. The parameters are chosen by the researcher. This increases the number of hypothesis since there is an infinite plane of possible parameters for the normal and Cauchy distribution. The parameter uncertainty is gone because the researcher has excluded it by design. The hypothesis goes from being of a choice between two models to being a choice between a model and chosen parameters versus another model and its chosen parameters. Calculations become increasingly simple, but there is now a fuzziness to the probability statement. It is no longer based upon the shape of the curve, it is now based on two specifically defined curves that is curves with parameters.

The hypothesis become:

$$H_0 : S_1 \text{ and } \boldsymbol{\theta}_1 \text{ is true.} \quad (4.37)$$

$$H_A : S_2 \text{ and } \boldsymbol{\theta}_2 \text{ is true.} \quad (4.38)$$

The math becomes far simpler. The posterior probability statement for the null hypothesis becomes:

$$\Pr(S_1, \theta_1 | \mathbf{y}_1) = \frac{\Pr(\mathbf{y} | S_1, \theta_1) \Pr(S_1)}{\Pr(\mathbf{y})} \quad (4.39)$$

There then become two reasonable possible methods to choose parameters. The first is to choose the maximum likelihood estimate, the second is to choose scientifically interesting parameters. They may of course be the same.

#### 4.2.1.5 Properties of Maximum Likelihood Estimator Based Ratios

There is an enormous advantage to choosing the maximum likelihood estimator for the parameters of either distribution. By choosing the likelihoods that maximize the numerator and the denominator, it becomes a test of best case versus best case, given the data actually observed.

Had the alternative been considered complementary to the null then this would have been identical to the Frequentist Likelihood Ratio Test. It could be argued that both the normal and the Cauchy distributions are the result of a summation process and are the only two options possible. The question would be “are these the only two limiting forms for returns?”

The three big candidates are jump diffusion models, the log-normal distribution and the broader class of stable distributions. The jump diffusion model can be excluded by White’s proof.[114] White only assumed finite variance and errors with a zero mean. The reason to force a Poisson jump into the error term is to try and replicate the large sudden departures in price observed in the data. There is no theoretical basis for it. As this is simply a mixture of normals then as time goes to infinity, it follows the test statistic will be the Cauchy distribution.

The log-normal distribution is a rather interesting special case. It is skewed and the real data is skewed. Under proper transformation, the returns could become log-normally distributed if the raw data is the Cauchy distribution. Still, it should be possible to rule out the log-normal due to nature of the error terms.

The log-normal is the result of a single error and not a dual error. The log-normal should appear in discounting situations, but there is no reason to see it with equity securities other following the approval of a cash for stock merger. The

log-normal distribution will receive its own special discussion in its own section.

This leaves the normal distribution. Although the normal distribution should be excluded by prior proof, it is the basis of the contention. Economists normally assume data is normally distributed in Frequentist tests.

The broad class of stable distributions includes both the normal and the Cauchy distribution. There are three reasons to exclude the broad class of stable distributions from consideration.

The first is that the trinary hypothesis of using a Cauchy distribution, a normal distribution, or some other stable distribution will run afoul of Lindley's paradox. The normal and the Cauchy distribution are the result of specific parameterizations of the broad class of stable distributions. As the parameter in question for the stable distribution is a subset of the real numbers and the Cauchy and the normal are precise values, the prior probability of a random variable being exactly equal to a fixed number is zero. Its another form of the infinitismals problem again.

The second is more grounded in economics. It should be the case that distributions exist for a reason. They are the result of behavior. Distributions should be chosen as a result of consequences. Although it may have to be tolerated as a reason for a time, the "just because," solution to an unknown phenomenon should be unsatisfactory.

The third is that prior research on stable distributions have been from a Frequentist perspective. This is quite important as Frequentist measures can be very sensitive to matching underlying assumptions. Truncation is a massive departure for stable distributions.

An extreme example of this is a study by Lee and Lee.[54] Although this work is troubled in other ways, there is no reason to believe the math itself is suspect. Not discussed in the article is that the parameter estimates for the Korean stock exchange imply that probabilities are not additive in South Korea as the distribution is not stable. Assuming the formulas were correctly implemented, this incredible result can only be an artifact of the method used.

A more prosaic and simple example is the truncated normal distribution. Standard tests of skewness will show a perfectly non-skewed but truncated distribution as skewed. The minimum variance unbiased estimator of the center of location for the Cauchy distribution, using Rothenberg's method is biased by two per cent per

annum in this data set.[83]

Using the odds ratio the problem becomes:

$$O = \frac{\sup \mathcal{L}(S_1, \boldsymbol{\theta}_1 | \mathbf{y})}{\sup \mathcal{L}(S_2, \boldsymbol{\theta}_2 | \mathbf{y})} \quad (4.40)$$

From the Neyman-Pearson lemma, this implementation is the uniformly most powerful test. If this is a binary hypothesis, then the result of this test excludes normality in both the Bayesian and the Frequentist paradigms.

## 4.3 Implementation

### 4.3.1 The Data

The data chosen was all end of day data from the Center for Research in Security Prices (CRSP), a commonly used data set. Each security was equally weighted. An alternative would have been to weight the returns by volume. Given liquidity costs as well as asymmetric volumes at the start and finish, this is impractical. This is equivalent to purchasing an equal dollar amount of each security each day and holding it for the designated holding period. Transaction dates run from December 31, 1925 to December 31, 2013.

### 4.3.2 Empirical Results

#### 4.3.2.1 The Likelihood Function and the Posterior Distribution of $\sigma$ With a Normal Likelihood

The likelihood function is a unnormalized distribution function. It measures the exact probability of observing the precise set of events appearing in the sample. With a uniform prior distribution the likelihood distribution is a scaling of the posterior distribution.

In order to find the maximum likelihood estimator for both models, a fine mesh of one-tenth of a basis point was created to cover the densest region. The portion of the data considered a going concern was used due to the nature of the underlying theory and the mixture nature of the distributions involved. There were 49,645,521

observations.

Survivorship effects are handled in a mixed way here. It appears in the likelihood as a truncated distribution, see **Eq. 2.23**. The constant of integration is adjusted so that support is over the range of zero to infinity rather than negative infinity to positive infinity.

Using the truncated likelihood function is necessary because it does shift the parameter estimates, otherwise. Further, the shift is different between the normal and the Cauchy distribution.

On the other hand estimates of the probability of merger or bankruptcy are dropped as they would appear identically in the numerator and denominator of the Bayes factor and cancel out. The distribution from merged firms was excluded because it is so very skewed and on theoretical grounds as well.

The mean-variance models assume that the markets are in equilibrium. Mergers can only happen when securities are under valued, or at least perceived to be undervalued, as the alternative is to buy physical capital and expand, issue a dividend, or give pay raises to management. Its not rational to engage in mergers when a market is in equilibrium. Purchases in equilibrium should only happen as false positives.

The purpose of this section is very narrow, that is to justify the use of the Cauchy model empirically rather than the normal model or log normal model of Black-Scholes for going concerns. As standard models ignore bankruptcy and require information to be fully incorporated into prices at the time of option purchase, it would be unfair to the models of mean-variance finance to include elements known to violate the assumptions of the models.

The dispute is narrow. Should models of return have a variance or not have a variance? That is the open question.

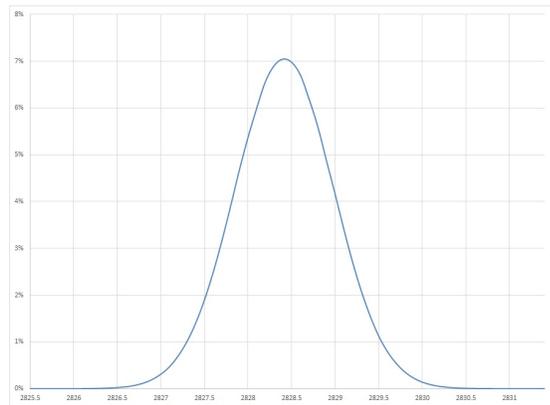
Model	Likelihood of Observing $\sigma$ at Most Likely Location	Most Likely Location of $\sigma$ in Basis Points
Cauchy Model	$10^{-15,055,291.23}$	2828.4
Normal Model	$10^{-23,713,281.08}$	9962.0

**Table 4.1.** Table of Maximum Likelihood

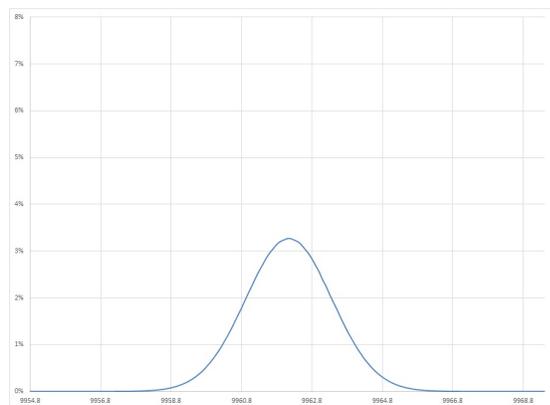
The Likelihood Ratio Test excludes the normal model with an odds ratio of  $10^{8,657,989.86} : 1$ . If it could be shown that these two models are the complementary

set of hypothesis, given the sources of skew would impact each distribution in the same manner and cancel out, then in both Frequentist and Bayesian methodologies, mean-variance finance is excluded.

A strictly Bayesian solution is the summation over all possible values, but within just a short distance from the maximum point probabilities fall off with over 90 orders of magnitude. As such, summation over the real number line is impossible given the discrete nature of computation. Conversion of the likelihood into a posterior density is a bit easier to understand. Although the posterior is the normalized product of the likelihood and the prior, for the regions involved the prior distributions were flat.



**Figure 4.1.** Marginal posterior distribution for the Cauchy model for  $\sigma$ , measured in basis points.



**Figure 4.2.** Marginal posterior distribution for the Normal model for  $\sigma$ , measured in basis points.

The posterior density function for the normal distribution is more diffuse than the Cauchy distribution. This results in a Bayes factor of  $10^{8.657,990.19}$ , which makes the Cauchy model about 1.4 times as likely as the normal model when compared to the likelihood ratio. However, given the magnitudes of the falsification, this is of no import.

Model	Lower Bound	MLE	Upper Bound
Cauchy	2826.9	2828.4	2829.9
Normal	9958.9	9962.0	9965.3

**Table 4.2.** Interval and Point Estimates of the Shape Parameter of the Competing Models, Measured in Basis Points

The ninety-nine percent highest density region for the Cauchy model is only three basis points wide with a peak at 28.284%, a very reasonable measure of uncertainty in returns. Simple integration implies that the interquartile range in the data should be between a loss of 26.4% and a gain of 26.4%.

The 99% highest density region for the normal distribution is economically unreasonable. While it is 6.4 basis points wide, that is not enough to bring down the estimate to reasonable levels. Peaking at 99.62%, the implication is that nearly one-fifth of all returns are greater than 100% per annum. The interquartile range for the data implied from the parameters should range from a 33.6% loss to an 80.1% gain. Even without performing a statistical test, the normal model can be excluded.

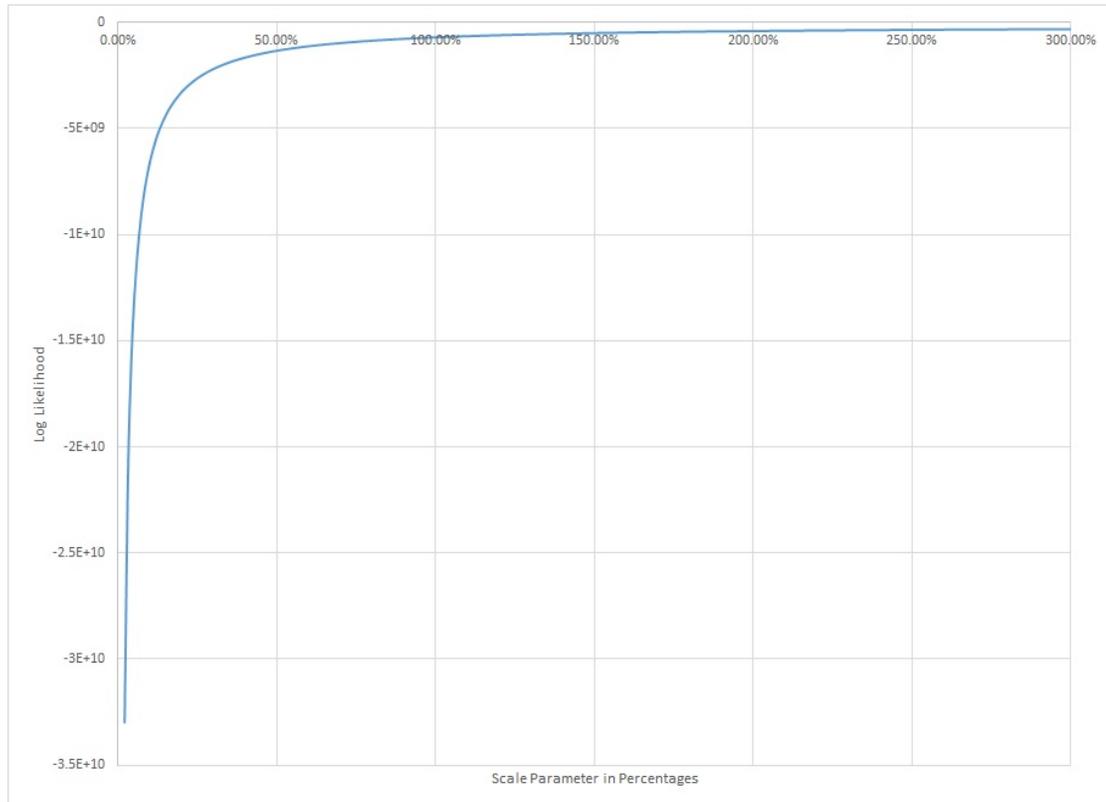
#### 4.3.2.2 The Likelihood Function and the Posterior Distribution of $\sigma$ With a Log-Normal Likelihood

The 99% highest density region for the log-normal likelihood was not found. A search was made for  $\sigma$  over the range of .01 to 3. What was found was that it is very likely no proper posterior distribution exists for  $\sigma$  given the data.

Treating the likelihood measurements as data points, the log-likelihood of  $\sigma$  over the range was approximately  $-1/x$  times a constant. This implies the likelihood would maximize when  $\sigma$  went to infinity.

On economic grounds alone the log normal distribution can be excluded.

If the log-likelihood is hyperbolic over the parameter space, then the proposed



**Figure 4.3.** Log-Likelihood for  $\sigma$  for the Log Normal Distribution

posterior distribution could not integrate to unity. This is visually obvious in **Figure 4.3**

It was totally unexpected that the likelihood would fail to integrate to one. Although it was expected that the Log Normal distribution would be a poor fit, it wasn't expected that the fit could not be measured.

One factor that may have played a role is the observation that the mode is unity. This was included in the estimate as it was for the normal and the Cauchy model. This may have restricted the possible values for  $\sigma$ . This forces the identity:

$$\exp\{\mu - \sigma^2\} = 1 \quad (4.41)$$

This leads to the obvious identity:

$$\mu = \sigma^2 \quad (4.42)$$

It may be true that estimates for  $\mu$  and  $\sigma$  may have been realistic had the mode not been conditioned to unity.

## 4.4 Discussion

The information loss due to using a normal distribution is nearly total. Although not calculated as the magnitudes are already so large, the Kullback-Leibler divergence is closely related to the log-odds in an expectational form. The normal distribution is sometimes used as an approximation for the Cauchy distribution, particularly when computational resources are limited. This would be very unfortunate here as the information loss would be catastrophic. The normal model is unrealistic and can be excluded. By moving to the Cauchy model, it is clear that an improvement has been made.

The data imply fundamentally different sets of human behavior. Because the data set is so large and represents such a long period of time, it is clear that the normal distribution should not be used and this is true almost regardless of the prior distribution for the parameter set. To get the normal distribution to provide a good fit to the data a distribution approaching the Dirac functional would have to serve as a prior distribution. Such a prior would be considered degenerate by any ordinary person.

It is also obvious that the log-normal distribution cannot be used to model returns. This may not be true if log return is mapped to another log return using standard regression models. The open issue is what the covariance matrix would mean. Prices do not covary, but they do appear to co-move. The mean of the log return maps to the median of the raw returns. It is an open question as to what the covariance between two sets of log returns would imply. Although the log return should converge to the hyperbolic secant distribution, regression mappings should converge to normality.

It is open to debate as to the interpretation of the covariance of instantaneous rates of return when returned to raw data.

# Part V

## Option Pricing

# PRICING EUROPEAN STYLE OPTIONS

The exclusion of Itô based methods for option pricing requires a fundamental second look at prior research and models. By excluding Frequentist methods due to admissibility issues it becomes possible to narrow the tools to Bayesian methods; this is still too large a class of possible solutions to consider.

It will be shown that a profit maximizing market maker is choosing a risk minimizing strategy, therefore it becomes possible to focus on profit maximization as a bounding condition for a point estimate of equilibrium prices. Still, it is wise to look again at prior attempts to price option contracts in order to construct a model.

## 5.1 Introduction

In order to price options, it is necessary to frame the model inside some greater framework. The primary logical framework is the study of financial asset pricing.

The study of financial asset pricing is relatively short. The first such academic study is by Bachelier in his doctoral thesis. Leonard Jimmie Savage mailed postcards out to leading economists and mathematicians to read Bachelier's dissertation. His work was derived from work on *rentes* and was preceded empirically

by Regnault.[9] This set off a flurry of work on Brownian motion and also option pricing. Work by Osborne extended the Brownian motion argument which was followed on by Samuelson, Boness and Sprenkle.[17, 74, 85, 94] As this dissertation is fundamentally a reworking of the math the approach resembles their work in some ways but not in others.

Like Bachelier, the study is using the limiting distribution of returns, but instead of the normal distribution, the limiting distribution is the Cauchy distribution. Like Boness, this dissertation assumes non-negativity of prices, but unlike Boness who used the log-normal distribution to evade the issue this work uses the Cauchy distribution truncated at zero for bankruptcy. Like Sprenkle who discounted at the expected rate of return for the stock price, this work considers the terminal value of the security, but discounts at the subjective opportunity cost of funds. Like Black and Scholes, this work considers the no arbitrage equilibrium but does not use Brownian motion to calculate the option price. Like the mean-variance models, the game in which the price is constructed provides actors with complete historical information and there is no informational asymmetry. Unlike these articles, this study includes liquidity costs, merger risk and the probability of bankruptcy.

Because European options have a definable terminal value without the possibility of early exercise, the simplest solution is to look at the distribution of possible terminal values. For put prices, the values are bounded at zero and at  $k$ , the strike price. An expected value therefore exists regardless of the distribution chosen. Ignoring time values and liquidity costs, the simplest mechanism to value a put contract is to integrate over the set of possible final values. Everything else becomes setting values to present value and adjusting for the costs of liquidity. With call options there can be no expected value as the future value diverges without bounds. However, it is possible to price calls through equilibrium pricing.

Certain things then become necessary to price European style option contracts. The first of these is an interest rate. In the real world there are a wide range of available interest rates. In practice there are offers that are available nationally, such as those from purchasing a bond over an organized national exchange, and there are those that are only available in certain localities or to certain people, such as from credit unions.

In this model there are two rates, a deposit rate and a loan rate. The existence of the spread isn't for realism. It is for coherence under de Finetti's Coherence Principle. From the Coherence Principle it is possible to derive Kolmogorov's axioms of probability. If there isn't a deposit and loan rate spread, then the most basic laws of probability cannot be assured to hold.

This split can be thought of as similar to a sales commission and indeed would encourage market makers and dealers to seek out sales staff in order to capture as much of this spread as possible by creating deals.

The second of these is the bankruptcy rate. Although it may be surprising, it simply isn't possible to solve the problem without a legal bankruptcy process. The political process precedes the economic process. Without either enforceable bankruptcy contract provisions or existing statutory provisions there is no way to find the expected value of a European style equity put option contract.

The strike price is the right boundary, bankruptcy creates the left boundary at zero. It is only between these two boundaries that an expected value can form.

The third required aspect is the bid-ask spread. Like the deposit-loan spread required for coherence, the bid-ask spread is required for coherence. Coherence permits the existence of call contracts even though no expected value could exist.

The fourth required tool for pricing options are initial endowments. Actors with different endowments may face different decisions. An actor indebted by \$100,000 is in a different position than one flush with cash.

There is a fifth factor not required for pricing, except of course in the real world, but which is quite useful. That is to include a probability of merger.

Mergers are often bankruptcy alternatives and a firm that is merging is really selling out its underlying physical capital. As financial capital is a claim on physical capital, by prohibiting mergers under any circumstance inefficiencies would form in the primary markets for physical capital and, through competition, labor. One would have to wonder at such an arrangement. It is unsurprising that this is not seen in the real world.

One other thing is helpful, heterogeneous preferences. Although not necessary, they permit a spectrum of risk premiums and do not require a wide range of necessary random events to converge properly in order for the supply curve and the demand curve to meet at least somewhere.

## 5.2 The Game

This game proposes a simple option pricing model, but it is not *the* model of option pricing. Fortunately, through the use of Bayes factors, it should be possible in relatively short order to find nearly optimal solutions for pricing by changing the information inputs. The data is not available for this study to attempt this.

Actors in the game engage in several actions prior to deciding on portfolio allocations. They set a prior distribution, estimate parameters using a cost function, construct a predictive distribution and use that distribution to engage in decisions regarding securities positions.

The game exists as a limited game inside a set of games, one game for each possible security and at one subsection of time in an infinitely repeated game.

## 5.3 Notation

Depending on the circumstances, most of the appropriate notation is suppressed. This notation will vary slightly from section to section and will be annotated in the text. For example, the bid price for a call option would be  $\frac{\psi}{z(n)}$  while the asked price would be  $z(n)\psi$ . While this is relatively mild, the complete representation a call as a function of the variables that map onto it would be:

$$\psi = \psi(p_t, k, t, T, \Delta t, i_D|\Delta t, i_L|\Delta t, \delta|[t, T], n, B, M) \quad (5.1)$$

Likewise the full notation for a put would be:

$$\phi = \phi(p_t, k, t, T, \Delta t, i_D|\Delta t, i_L|\Delta t, \delta|[t, T], n, B, M) \quad (5.2)$$

In addition, there are multiple interest rates in the game. For some calculations a deposit or a loan rate is specifically important and would be denoted  $i_D$  or  $i_L$ . In some places the interest rate is denoted  $i_X$  where  $X \in \{D, L\}$ .

## 5.4 Actors

The game has three types of actors; they are nature, market makers and participants. At time zero, nature makes an initial move for each actor endowing them with resources. Participants are endowed with  $\bar{m}_i$  of liquid wealth, such that  $-\infty < \bar{m}_i < \infty$  and  $\bar{m}_i \in \mathfrak{R}, \forall i \in I$ .  $I$  is the index of participants. In addition, nature endows market makers and participants with  $\omega_i$  in risky wealth. Resources are endowed so that  $\sum_{i \in I} \bar{m}_i = 0$  and  $\sum_{i \in I} \omega_i = \omega_{universe}$ . Market makers are assumed to be near their equilibrium balance sheets. Likewise, participants are near or at their equilibrium balance sheets.

## 5.5 Actions

Securities are in decision theoretic terms, lotteries. Buying  $n$  shares of an asset has a different probability distribution than buying  $n$  call contracts when seen in reward space. As such, they are different lotteries with different density functions. Some combinations of lotteries map to the same probability distribution as another lottery. Buying an equity security is the same as buying a call option and selling a put contract in the sense they have the same density function. The game is restricted to decisions to be executed at time  $t$ . Although actors have access to a wide range of risky assets, all assets except the one of interest are collectively held as an individual's portfolio designated  $\omega_i$ . That portfolio contains all other risky wealth. This serves two functions.

First, instead of having the notation  $p_t$  each firm would have to be identified and so there would end up with some notation such as  $p_t^f$  for each firm  $1 \dots f$  in  $F$ . This would add no information. Second, the purpose of this article is to formulate a model of option pricing based on a single security. While it is true that there may be demand for a set of options from many different firms, that is a portfolio effect and should be already included in the prices of the underlying security. Derivatives derive their value from the pricing of the underlying.

Actors have several possible actions, denoted  $a$  or sometimes  $a'$  etc, they can take on several possible lotteries. The actions are:

- Place market order to:

- Buy
  - \* shares in the underlying security
  - \* contracts in a European style option on the underlying security (also called going long)
- Sell
  - \* shares in the underlying security
  - \* contracts in a European style option on the underlying security (also called writing)
- Place limit order to:
  - Buy
    - \* shares in the underlying security
    - \* contracts in a European style option on the underlying security
  - Sell
    - \* shares in the underlying security
    - \* contracts in a European style option on the underlying security
- Place multiple orders in a convex combination of the above order types
- Do nothing

The set of all possible actions over all possible permitted combinations of securities is denoted  $\mathcal{A}$ .

Some actions are functionally excluded by dominance. For example, buying a call and selling a put while short selling the underlying security would be equivalent to “do nothing,” except that there would be transaction costs. Hence, do nothing dominates doing something, where that something is more costly than a simpler solution. Although the move is not excluded in the game by rule, it is dominated and so including “Do Nothing” permits a pragmatic rationality-based bounding for the set  $\mathcal{A}$ .

For most practical purposes, limit orders are the same as “do nothing.” If a security is trading at \$10 per share and a sell limit order for \$15 is place, this is generally no different than placing a market order and so will be treated as

a market order. However, a security trading at \$10 per share when a sell limit order of \$5 is issued will result in no action. As only one static moment is under consideration here, though it is a generic moment, limit orders will either map to a *market order* or orders to *do nothing*.

This general rule undergoes further discussion and elaboration in the segment on liquidity.

There are five possible simple lotteries. They are:

- Equity security
- Call option
- Put option
- Deposit contract
- Loan contract

Compound lotteries are made up of several simple lotteries.

The outcome of the deposit and loan contracts are known with certainty and so result in the same outcome in all states of nature, except in the case of the bankruptcy of the obligor. If a buy order would exhaust and exceed all possible liquid assets then a loan would automatically be granted. Likewise, any net cash generated by transactions would automatically create a deposit contract to the extent the cash exceeded any outstanding debt.

## 5.6 Utility Functions

Utility functions can be represented in one of two manners. The first and the most common in economic models is the utility of some uncertain variable, such as wealth. This is often represented as  $U(\tilde{w})$ . This is also sometimes represented in terms of consumption in economic models. The realized utility will depend upon the final state of  $w$ .

The other possibility is to use the utility of the action chosen. This would be represented as  $U(a)$  or  $U(a')$ . Once an action is chosen, alternative lotteries would no longer matter. In this form, *regret utility* is being used. Regret utility, not in its

expectational form, is the utility of an action in a given state of nature,  $\theta$ , where  $\Theta$  is the set of all possible states. It can be expressed as:

$$\mathcal{U}(a, \theta) = u(a, \theta) - \sup_{a'(\theta)} u(a', \theta) \quad (5.3)$$

## 5.7 Assumptions

There are six explicit prices in the game, four of which are exogenous. They are  $p_t$  and  $p_T$  which are the current and future price of some security;  $i_D$  and  $i_L$ , which are the interest rates on deposits and loans respectively at time  $t$ ; and  $\phi$  and  $\psi$ , which are the current prices for put and call options if liquidity costs are ignored. They are endogenous variables. There is also an implicit price  $\lambda$ , which is used to price the cost of liquidity. The cost of liquidity is treated as exogenous.

For participants all moves happen at time  $t$ , denoted  $t$ . The consequences of these actions are paid out at time  $T$ . The interval  $T - t = \Delta t$ . For the market makers, it is possible for moves to be made in continuous time over the interval  $[t, T]$  and implicitly there will be other participants available as counter-parties over the interval.

In addition to the current exogenous market prices, there are three other variables which affect the price of option contracts. They are the strike price, denoted  $k$ , the number of contracts or shares purchased denoted  $n$ , and  $\delta$ , which is the future value of dividends over the interval  $(t, T]$  marked up to future value at the deposit rate  $i_D$ . As is standard in economic notation, a variable noted with a \* is at the equilibrium quantity.

Although securities are priced at  $p_t, p_T, \psi$ , and  $\phi$ , they are marked up or down to the bid and ask price by requiring a liquidity premium. The liquidity premium, from Abbott[1], is being modeled as a function  $z(n)$  such that:

$$z(n) = e^{n\lambda} \quad (5.4)$$

In addition to price variability, securities are subject to existential risks. In particular, the probability of bankruptcy or merger whose probability is denoted  $B$  and  $M$ .

Both  $\psi$  and  $\phi$  are unknown functions of the above variables. For purposes of notation, most or all of the function variables are being suppressed to either  $\psi$  or  $\psi(n)$  and either  $\phi$  or  $\phi(n)$ .

Interest rates and the future value of dividends are, of course, dependent on the interval of time over which they are to happen. Likewise, to simplify notation

$$i_D = i_D|\Delta t, t \quad (5.5)$$

and

$$i_L = i_L|\Delta t, t \quad (5.6)$$

and

$$\delta = \delta|[t, T] \quad (5.7)$$

Market makers and participants that write options are assumed to be profit maximizers. Participants that are option buyers are assumed to be either profit maximizers or utility maximizers. Utility maximizers are assumed to have strictly concave utility with heterogeneous preferences.

Participant actors have heterogeneous endowments. A consequence of this is that different actors face different interest rates. A result of this is that they have different reservation prices.

All participants are assumed to have access to sufficient credit facilities that they could engage in any profitable transaction up to and past the point of profitability.

Participants in the game have perfect knowledge of all relevant historical data and the data set is very large. Actors base their actions on the predictive distribution created from the data from the beginning of the data set to time  $t - 1$ . As the set is very large, it is assumed that the differences in parameter estimates are less than the number of significant digits and so in a discrete space are equal. Implicitly this presumes no actor holds a degenerate prior.

## 5.8 Nature of the Contracts

Participants are able to either write or buy an unlimited number of European style put or call options. A European style equity option is an option on an underlying security that can only be exercised at maturity. In this case maturity is at time  $T$ . In addition, or in lieu of, they could purchase or sell an unlimited quantity of the underlying security.

A put contract grants the buyer the right, but not the obligation to require the writer to buy a security at a predetermined price denoted  $k$  and called the *strike price*. A call contract grants the buyer the right, but not the obligation to require the writer to sell a security at a previously chosen strike price, denoted  $k$ .

Unlike equity securities or bonds, which are fixed in quantity over the short run, option contracts can exist without limit. Financial intermediaries create financial contracts and these contracts are flexible in quantity. This mechanism is similar to the manner in which banks create money.

When a new bank forms the equity is loaned out, these loans cause the recipients of the money to deposit it in banks. This money is then loaned out again, creating new deposits in the process, until some contractual or regulatory limit is reached.

Similarly, market makers in this game insure the market against adverse movements by writing options and by making a market for those parties that wish to absorb those risks. This assures the market that participants can become underwriters of those risks. In a sense, the market maker acts as a Lloyd's association does in insurance when combined with a reinsurer to cover risk of failure by the contract writers. For purposes of the game, only market makers can create derivative securities.

Like a bank certificate of deposit, no option obligation exists until a market maker agrees to open an account and create one. Likewise, a risky option position cannot be closed early without a counter-party willing to absorb the risk. By insuring the primary markets against certain types of risks, it makes it possible to increase the size of the primary markets by permitting risks considered unacceptable to one party to be sold to another party.

As a consequence, market makers are at risk with every potential transaction. Consider, for example, a party that wishes to go long 1 call option in ABCorp.

The market maker posts a price. If that price is acceptable to the counter-party, the position is opened. Another party comes along wishing to write a call option on ABCorp. The market maker posts a price. If it is acceptable to that party then they write the contract.

In the first position the market maker is short one contract. In the second position the market maker is long one contract. On the surface the market maker bears little or no risk. If the second party declares bankruptcy and defaults on the written call, should it be exercised, it is the market maker who is obligated to fulfil the long contract.

In the absence of counter-party failure there is in practice only one open contract with the market maker acting primarily as a clearing institution. In the real world, two contracts exist. In this game, each separate position will be accounted for, even if participants effectively clear out the position by absorbing the market maker's risk.

## 5.9 The Profit Function for Short Puts

### 5.9.1 The Critical Importance of the Short Put

The capacity to solve any element of the system revolves around the short put. It is the only portion of the system where an expected profit exists. As such, puts are necessary for financial stability. Although the European style short put seems like an esoteric concept, it is in fact a simpler mathematical construction than a bank deposit. It is quite possible bank deposits exist in this sea of instability because of the nature of the put contract.

To understand the relative simplicity of an equity put contract when compared to a bank deposit or a bank loan, it is important to think about what a bank deposit grants the actor. Depositors receive a debt obligation and a long put that they didn't pay a premium for. If interest rates increase enough, then a depositor will remove their deposit and redeposit the money at the higher rate. If interest rates fall, then the depositor can continue to receive the higher rate.

Banks create a bond, often at a fixed rate of interest, and includes a long American style put. Implicitly the bank would loan the customer the money for

the premium through a discounted deposit rate and possibly a penalty for early withdrawal. For a time deposit, the depositor receives a rate  $i_D$  from time  $t$  to time  $T$  unless a higher rate appears. Then the depositor is free to choose the higher of the two rates. In the absence of a penalty for early withdrawal, the depositor is guaranteed the supremum of the available rates up to some maximal time period.

This is mathematically far more complex than a European style put on an equity security. The obligation is only at the end of the period. It defines a payoff only at that time. The writer has no automatic future obligation to the buyer, such as the obligation to write another put. A bank, on the other hand, as a common carrier, must agree to accept a rate marked up deposit from the same customer who cancelled the prior agreement.

When one considers that central banks attempt to control the static volume of the real supply of deposits, there exists an implication that the private production of a public good by banks includes the mathematical elements of that provisioning of public goods. A bank contract is a good deal more complicated than the goal of pricing European style equity options.

Still, this relatively simple contract is the building block of all American style options, European style call options and standard banking products.

### 5.9.2 The Profit Function

The profit functions<sup>1</sup> for participants are such that revenues are marked down by an exponentially growing cost of liquidity in  $n$  and costs are marked up by an exponentially growing cost of liquidity in  $n$ . The long option positions are cost plus profit positions and as such maximize utility by insuring against risk rather than generating a profit. By assumption, at the margin, actors are risk averse and so for purposes of this game there are no speculators. The presence of speculators in the market can have an impact to be discussed later in the section on speculators.

As market makers could be the permanent holders of all short positions, it is the market maker's self interest to assure not only that sufficient premiums are being collected, but also that the option writers are profitable.

Two of the three possible single contract short positions can have no expected

---

<sup>1</sup>Economists usually define the profit function in terms of a maximization. That assumption is relaxed until the end of the chapter.

profitability as the expectation diverges and therefore does not exist. Expected gains and losses for short selling a stock or call option cannot be defined. As it happens, this is not a difficulty in equilibrium.

As will be shown later, the call price, given a large cash endowment, in equilibrium is  $\psi(n) = \phi(n) + p_t \frac{1-i_L}{1+i_D} - \frac{k+\delta}{1+i_D} z(n)$ . Every variable on the right hand side is well defined except  $\phi$  and  $\delta$ . Setting aside considerations of  $\delta$  for a moment, the question becomes “is there an optimal put premium given an obligation by the market maker to write  $n$  put contracts without a volume limitation?”

The market participants will avoid writing put contracts if the price is set too low, sticking the maker with the entire inventory. Likewise, market participants could rush in to take up inventory if the price were too high. In an equilibrium there is no pressure for prices to change. The concern is with equilibrium profitability.

The goal is to set a price such that the volume sold produces maximal profitability. The profit function for the short put contract is:

$$\Delta\Pi_S^P(n) = \frac{n}{z(n)}\phi(1+i_X) - n \max\left(0, k - \frac{p_T}{z(n)}\right) \quad (5.8)$$

In this equation  $x \in \{D, L\}$ . As the goal is to get participants to voluntarily choose optimal volume, it must first be determined what is an optimal volume.

The function  $\max\left(0, k - \frac{p_T}{z(n)}\right)$  has three possible existential states of nature. They are

1. The firm is bankrupt
2. The firm has been merged out of existence
3. The firm is a going concern

In the bankrupt state, the writer pays  $k$  since  $p_T = 0$ . This happens with probability  $B$ .  $B$  is understood in a Bayesian sense as  $B|\iota$ , where  $\iota$  is information. This is also true for the other existential states.

The expected profit, given bankruptcy, is:

$$E(\Delta\Pi_S^P(n)|Bankruptcy) = \left(\frac{n}{z(n)}\phi(1+i_X) - nk\right) E(B) \quad (5.9)$$

Without bankruptcy, it is not possible to calculate option premiums as the integrals would hopelessly diverge in all existential states.

Mergers are often a bankruptcy substitute, as such, it is reasonable to believe that post merger returns may have different parameters than for firms which are a going concern. In the non-bankruptcy states the Cauchy distribution has two parameters in each remaining state. These are based on the posterior estimates from the existing data set.

If  $G$  is thought of as the probability of a firm continuing as a going concern and  $M$  the probability of merger, then there are two ways to handle the relationship between  $G$ ,  $M$ , and  $B$ . They are:

1.  $G+M+B=1$
2. Or by having both of the following conditions be true:
  - (a)  $G|\text{not bankrupt} + M|\text{not bankrupt} = 1$
  - (b)  $(G|\text{not bankrupt} + M|\text{not bankrupt})(1 - B) + B = 1$

The method used is the multinomial choice of  $G + M + B = 1$ .

The profit function in the merged state comes from a distribution without known analytic properties. The method of histograms permits an approximate solution for this problem.[13] Fundamentally, predicted values are based on the probability that the final price,  $p_T$  will be inside a particular partition. This can only be an approximate solution and so refinement of a solution using first principles would be a significant step forward.

Noting that in equation 5.8 that a loss would happen anywhere  $p_T < kz(n)$ , implies that out of the money options, that is those greater than the strike price should be exercised due to liquidity costs. One partition, should then cover the no loss region of  $p_T \geq kz(n)$ . The remaining partitions should be optimally chosen to minimize information loss. The optimal number of partitions is assumed to be  $S + 1$ , where  $S \in \mathbb{Z}^+$ .

The probability of being in a given slice, given a merger will happen in the contract period, is unknown, but is estimated here using the multivariate normal of dimension  $S+1$ . Each partition is mutually exclusive and therefore independent.

The beliefs regarding the probability of being in a particular in-the-money slice,  $s \in \{1 \dots S\}$ , is:

$$\Pr\left(\frac{s-1}{S}kz(n) \leq p_t < \frac{s}{S}kz(n) | \text{Merged}\right) \sim \mathcal{N}(\mu_s, \sigma_s^2)M \quad (5.10)$$

and for the out of the money slice:

$$\Pr(z(n)p_t \geq k | \text{Merged}) \sim \mathcal{N}(\mu_{S+1}, \sigma_{S+1}^2)M \quad (5.11)$$

A multinomial distribution could have been used as well.

As partitions are, by this construction, of equal width, this creates an expected profit function of:

$$\begin{aligned} E(\Delta\Pi_S^P(n) | \text{Merged}) &= E(M) \frac{n}{z(n)} \phi(1 + i_X) \\ &\quad - nkE(M) \left\{ (1 - \mu_{S+1}) - \sum_{s=1}^S \mu_s \frac{(s - \frac{1}{2})}{S} \right\} \end{aligned} \quad (5.12)$$

The second  $z(n)$  from equation 5.8 vanishes because  $z(n)$  scales the partitions so that each partition is  $\frac{kz(n)}{S}$  wide, but the midpoint is discounted by  $\frac{1}{z(n)}$ . Likewise, as  $k$  scales the width of the histogram and appears as the paid out strike price,  $k$  gets pulled out to the side.

In the going concern state of nature, the expected profit function is:

$$\begin{aligned} E(\Delta\Pi_S^P(n) | \text{Going Concern}) &= E(G) \left[ \frac{n}{z(n)} \phi(1 + i_x) - 0 - \right. \\ &\quad \left. \frac{2n}{\pi + 2 \tan^{-1}(\frac{\mu_G}{\sigma_G})} \int_0^{kz(n)} \left( k - \frac{p_T}{z(n)} \right) \frac{\sigma_G}{\sigma_G^2 + (p_T - \mu_G)^2} dp_T \right] \end{aligned} \quad (5.13)$$

For simplicity, it was assumed the contract was an all or nothing contract. It will marginally overstate costs if this is not true for small values of  $n$ . The center of location is  $\mu_G$  and the parameter of spread is  $\sigma_G$ . Note however that the current spot price is information and that  $\mu_G$  and  $\sigma_G$  are notationally shortened from  $\mu_G|p_t$  and  $\sigma_G|p_t$ . It should also be noted that additional information beyond the strike

price could be included such as accountancy data or dividend payments.

Because the distribution is truncated at zero due to the existence of bankruptcy, the coefficient of integration is  $\frac{2}{\pi + \tan^{-1}(\frac{\mu}{\sigma})}$  instead of  $\pi^{-1}$ .

Prior to evaluating the parametric form for the going concern, it may be valuable to consider using the distribution free form, as in equation 5.12, for the going concern. Then the joint profit function for all three states of nature, in approximation, becomes:

$$E(\Delta\Pi_S^P(n)) = \frac{n}{z(n)}\phi(1 + i_x) - nk \left[ E(B) + E(M) \left\{ (1 - \mu_{S+1}^M) - \sum_{s=1}^S \mu_s^M \frac{(s - \frac{1}{2})}{S} \right\} + E(G) \left\{ (1 - \mu_{S+1}^G) - \sum_{s=1}^S \mu_s^G \frac{(s - \frac{1}{2})}{S} \right\} \right] \quad (5.14)$$

Setting a function  $\Lambda(k)$  as the net unrecoverable loss function, as:

$$\Lambda(k) = k \left[ E(B) + E(M) \left\{ (1 - \mu_{S+1}^M) - \sum_{s=1}^S \mu_s^M \frac{(s - \frac{1}{2})}{S} \right\} + E(G) \left\{ (1 - \mu_{S+1}^G) - \sum_{s=1}^S \mu_s^G \frac{(s - \frac{1}{2})}{S} \right\} \right], \quad (5.15)$$

the problem simplifies to the more visually tractable:

$$E(\Delta\Pi_S^P(n)|n; p_t; k) = \frac{n}{z(n)}\phi(1 + i_x) - n\Lambda(k) \quad (5.16)$$

Since  $p_t$  and  $k$  are exogenous, this permits a solution for  $\phi$  given an optimal value for  $n$ .

First order conditions for this form are:

$$\frac{dE(\Delta\Pi_S^P(n))}{dn} = \frac{\phi(1 + i_x)}{z(n)} - \frac{\phi(1 + i_x)\lambda n}{z(n)} - \Lambda(k) \equiv 0 \quad (5.17)$$

With a little manipulation, the equation can be brought into product-log form and  $n^*$  can be arrived at thus:

$$n^* = \frac{\left(1 - W\left(\frac{e\Lambda(k)}{\phi(1+i_X)}\right)\right)}{\lambda}, \quad (5.18)$$

where  $W(x)$  solves:

$$x = W(x) \exp(W(x)) \quad (5.19)$$

Unfortunately,  $\phi$  has yet to be solved for. Still, it illustrates the important inverse relationship between optimal volume and  $\lambda$ . Since volume, and not  $\phi$  can be observed in the market,  $\phi$ , is solved for as:

$$\phi = \frac{\Lambda(k)z(n^*)}{(1+i_X)(1-n^*\lambda)} \quad (5.20)$$

For completeness, second order conditions support a maximum when:

$$n\lambda < 2 \quad (5.21)$$

The profit function to write a short put, by substitution, is:

$$\sup \Delta \Pi_S^P(n) = \frac{n^*\Lambda(k)}{1-\lambda n^*} - n^* \max\left(0, k - \frac{p_T}{z(n^*)}\right), \quad (5.22)$$

the expectation for which is:

$$E(\sup \Delta \Pi_S^P(n)) = \frac{n^*\Lambda(k)}{1-\lambda n^*} - n^*\Lambda(k) \quad (5.23)$$

Although this is an approximation, it has a nice form. Writers receive a percentage mark-up over costs. In this model,  $\phi$ , is a function of  $n^*$  and not  $n$ . The alternative would be for  $\phi$  to vary directly with  $n$ . There is an important conceptual difference that goes to the core of banking. If  $\phi$  is a constant, then all mark-ups and mark-downs are taken by the market maker. The market maker's role is to absorb the volume. It represents motion along the supply curve. It is a liquidity cost and not a size effect. It represent timing and the ability to maintain stable supply and demand curves.

On the other hand,  $\phi(n)$  represents a shift of the curve. It represents a movement along a different supply curve. This is different from a simple repricing of the underlying security, from  $p_t$  to  $p'_t$ . This would imply the existence of a market participant with pricing power. Of course, in equilibrium this should not occur in the world of many competing actors, but could be imagined in a world where a principal market maker failed and another actor was willing to underwrite the missing contracts. It could also be the case where a market maker wanted to exit a line of business and sell its book of business to another actor.

For completeness, this state is provided. Equation 5.8 becomes:

$$\Delta\Pi_S^P(n) = \frac{n}{z(n)}\phi(n)(1 + i_X) - n \max\left(0, k - \frac{p_T}{z(n)}\right) \quad (5.24)$$

Taking expectations and setting the first derivative to zero generates a differential equation for  $\phi(n)$  whose solution is:

$$\phi(n) = \frac{\Lambda(k)z(n)}{1 + i_X} + \frac{cz(n)}{n}, \quad (5.25)$$

where  $c \in \mathfrak{R}^{++}$ .

The expected profit function is:

$$E(\sup \Delta\Pi_S^P(n)) = c(1 + i_X) \quad (5.26)$$

The writer is recapturing all costs and collecting a flat fee, independent of  $n$ ,  $\lambda$ , and  $k$ . For that to be the case, the market maker has to transfer capital and liquidity profits to remove the risk from its books.

### 5.9.2.1 Parametric Form for Going Concern

The parametric form of the going concern profit function shown in equation 5.27 is far less tame. Indeed, a simple visual inspection would cause anyone to doubt that the first and second derivative would go anywhere simple or useful. Far more important, the pattern of  $\frac{n}{z(n)}a - nb$ , is hopelessly broken here. Although it is the parametric form of the close, distribution-free approximation and at its core the same pattern must hold, it isn't obvious how that would be arrived at.

The simple pattern where  $z(n)$  and  $k$  scale the cost function is obscured by the nature of the integration. The first and second derivative are contained in an appendix. The fundamental lessons are the same, except that there isn't an observed analytic solution for the second derivative test.

$$\begin{aligned}
E(\Delta\Pi_S^P(n)|\text{Going Concern}) = E(G) & \left[ \frac{n}{z(n)}\phi(1+i_x) - 0 \right. \\
& - \frac{2n(k-z(n)\mu)\tan^{-1}\left(\frac{k-z(n)\mu}{\sigma z(n)}\right)}{\pi+2\tan^{-1}\left(\frac{\mu}{\sigma}\right)} \\
& - \frac{2n(k-z(n)\mu)\tan^{-1}\left(\frac{\mu}{\sigma}\right)}{\pi+2\tan^{-1}\left(\frac{\mu}{\sigma}\right)} \\
& \left. + nz(n)\sigma \left\{ \log(\mu^2+\sigma^2) - \log\left(\frac{k^2}{z(2n)} - \frac{2k\mu}{z(n)} + \mu^2 + \sigma^2\right) \right\} \right]
\end{aligned} \tag{5.27}$$

The primary issue that makes the problem difficult is that it is rational to exercise out-of-the-money option contracts when, adjusted for the cost of liquidity, they are pragmatically in the money. A person holding an option for 10,000 shares of ABCorp with a strike at 50 when the current price is 49.75 would not exercise under the Black-Scholes model, but must exercise here if the shares are still desired if the market order to buy the shares in the open market would drive it over \$50, which it surely would. Unfortunately, liquidity costs are marked up and down using an exponential cost function. This is what leads to the difficult functional form.

### 5.9.2.2 Conclusion

The writer of a contract is the insurer. If the insurer fails, the market maker assumes the liabilities. It is in the self-interest of the market maker for the writer to be profitable and have adequate assets to meet obligations. The price observed in this derivation is the *minimum* reserve price that the writer should receive. Market maker fees must be on top of that minimum reserve. A profit maximizing writer, as will be shown later, is a risk minimizing writer. This also minimizes the

risk to the market maker. Should a risk loving writer appear, the market maker must exclude bids below the minimum reserve as it increases the risk to the writer and to the market maker.

## 5.10 Buying Put Contracts

It seems the most likely way to attack this problem is through the indifference curve. In equilibrium, the marginal actor should be indifferent among the choices of investments and/or insurance options. As such, in equilibrium it should be true that:

$$\int_0^\infty \int_{\omega \in \Omega} U(\omega + np_T) \Pr(\omega) d\omega \Pr(p_T) dp_T \equiv \int_0^\infty \int_{\omega \in \Omega} U(\omega + \max(0, \frac{p_T}{z(n)} - k) + (\bar{m} - n\psi(n))i_X) \Pr(\omega) d\omega \Pr(p_T) dp_T \equiv \int_{\omega \in \Omega} U(\omega + \bar{m}i_X) \Pr(\omega) d\omega \quad (5.28)$$

where  $\omega$  is other uncertain wealth not allocated to this asset with a domain of possible values  $\Omega$ .

Further it is probable that attacking the problem up to the strike price and from the strike price to infinity may hold promise. Numerical solutions do seem possible, but are outside the data available for this research.

The other possibility is to use regret utility functions. An interesting observation on regret utility is that long option contracts are never the supremum of a set of actions for any given state of nature in n-space. This implies that under regret utility long contracts always have negative utility in profit space over volume. This is rather interesting in that such a contract makes the amount of regret absolutely certain since the contract guarantees a worst possible state of nature.

Regret utility factors in the cost of false positives, that is the performing action  $a$  when the optimal action, known only after the fact, is action  $a'$ . The existence of long options imply the absence of a dominant strategy in all states of nature. That is to say, no contract can stochastically dominate another. Under regret utility, the stock market lacks a regret-free solution. At most, all strategies will

bring about equal marginal regret. Indeed, the “do nothing” strategy is also never the supremum strategy. Even what is traditionally called the risk free strategy is regret filled.

### 5.10.1 Brief Discussion

A brief discussion is in order as this mechanism for pricing puts is slightly different from what would be expected in a mean-variance framework. Note, for example, there are no dividends in the formula for a put contract. This does not mean dividend payments do not impact the price of puts.

A simple example would be a firm paying liquidating dividends over a period of years. An option on the current price would almost certainly be guaranteed to be *in the money* far enough into the future. This would differ from a firm paying dividends from profits or no dividend at all.

How should one incorporate such a dividend? It should be in the likelihood function, subject to any prior information about dividend payments on prices. It should appear in  $\Pr(p_T|\delta)$  and not as a correcting factor outside the expected cost function. It is inherently true that dividends are uncertain. A Board of Directors, as in any legislature, is subject to time inconsistency. That is, the games are subgame imperfect. This forces inductive reasoning to properly estimate the role of either announced or historical dividends on future prices.

If prices were not conditioned on information that included dividends, then the effect of dividends is disbursed into the general uncertainty of future prices. In essence, ignoring dividends increases uncertainty, but that does not inherently mean the gain in information is worth the computational costs. It simply means the effects of dividends become hidden in the uncertainty about price changes.

Another missing element is the relation between the current spot price and the strike price. This missing information is captured in  $\mu_G, \mu_M, \sigma_G, \sigma_M$  as they are really  $\mu_X|p_t$  and  $\sigma_X|p_t$ . So the strike price vanishes into the posterior via the likelihood function as well.

It is reasonably certain, but not perfectly certain, that a stock currently priced at \$50 per share with a put option with a strike price at \$100 per share with one year to run in the contract will expire in the money in the absence of information

that would cause one to believe that the current price is far from the equilibrium price.

One other slight difference from mean-variance finance is that calls are priced actuarially while puts acquire their value through put-call parity. This is reversed as calls have no expected value.

## 5.11 No Arbitrage Equilibrium

Noted in the literature is that a condition that must be met is the absence of arbitrage opportunities. This can be justified under a number of possible assumptions or as a consequence of rationality concepts. However, one of the simplest is de Finetti's coherence principle. [76] de Finetti set about an axiomatization of probability theory in 1937 built around the concept of gambling. Shortcomings in this approach are noted in Shimony, Janes and Nau. [46, 68, 92] de Finetti's Coherence Principle can be stated as;

**Assumption 5.** A bookmaker's betting odds are coherent if a client cannot place a bet or a combination of bets such that no matter what outcome occurs, the bookmaker will lose money.

An open question to this assumption, of course, is do rational actors have to use coherent probabilities. Although this issue is covered by Ramsey and Savage, it can simply be excluded here by the assumption that the market maker is a profit maximizer and could simply choose not to engage in transactions that result in a sure loss, when the alternative was a zero change in profits.[76, 86]

The practical implication, here, is that an infinite number of possible prices are not possible. Depending upon perspective, the presence of a book maker assures either the existence of a no arbitrage equilibrium or as an alternative construction, perfect competition subject to liquidity costs.

The binding rule for the bookmaker is that the bookmaker will accept any and all gambles as long as they are finite in number at the posted price.

The rule, when combined with the Coherence Principle create a set of binding conditions on the bookmaker.

If two different portfolios exist, denoted  $\Xi_1$  and  $\Xi_2$ , with a common payoff function of  $\Pi_1 = \Pi_2 = \Pi_{\Xi}$ , and with posted market prices  $\xi_1$  and  $\xi_2$  respectively, then if  $\xi_1 \neq \xi_2$  then it is possible to guarantee the bookmaker can lose money if participants can freely buy or sell portfolios. The proof is simple. Assume, without loss of generality, that  $\xi_1 > \xi_2$  and that a participant can purchase or sell  $n > 0$  portfolios at the different prices. The participant sells  $n$  portfolios of  $\Xi_1$  at  $\xi_1$  and purchases  $n$  portfolios of  $\Xi_2$  at  $\xi_2$ . At inception of the position the market participant receives  $n(\xi_1 - \xi_2)$  from the market maker, up to the total capital of the market maker. At the payoff time the participant receives  $n\Pi_{\Xi}$  and pays  $n\Pi_{\Xi}$  for a net profit of  $n(\xi_1 - \xi_2)$ . The market maker is guaranteed a loss of  $n(\xi_1 - \xi_2)$ . This violates the assumption of the market maker being a profit maximizer as the market maker could have set the prices equal and received a higher payoff, though that payoff would also have been zero. With liquidity costs, equal prices would have guaranteed a positive profit to the bookmaker.

As a side note, from the combination of the above rule with the Coherence Principle it is possible to derive as theorems the Kolmogorov axioms of probability.

## 5.12 Call Options Under Various Initial Endowments

As different possible actors could approach the market maker with different reservation prices, different possible no arbitrage equilibrium prices exist. It is important to not read an equilibrium price as *the* equilibrium price but rather as the equilibrium price conditional upon a state of nature. Each possible state of nature is dependent upon the subjective conditions of the actors approaching the market maker to enter into positions.

### 5.12.1 Large Cash Endowment

In this state of nature, it is assumed that  $\bar{m} \gg 0$  and that after the position is entered into sufficient cash exists to maintain the position without borrowing. The

profit function, should no action be taken to enter into a position, is:

$$\Pi_{\bar{m} \gg 0} = i_D \bar{m} \quad (5.29)$$

The profit functions in table 5.1 actually represent a change in profit from the *do nothing* choice. As such, they are prefixed with a  $\Delta$  to make that clear. The profit functions of this state of nature are in table 5.1. Some elements of these

Type of Position	Formula for Profit Function
Long Call	$\Delta \Pi_L^C(n) = n \max(0, \frac{p_T}{z(n)} - k) - nz(n)\psi(1 + i_D)$
Long Put	$\Delta \Pi_L^P(n) = n \max(0, k - z(n)p_T) - nz(n)\phi(1 + i_D)$
Long Equity Position	$\Delta \Pi_L^S(n) = \frac{n}{z(n)}p_T - nz(n)p_t(1 + i_D) + n\delta$
Short Call	$\Delta \Pi_S^C(n) = \frac{n}{z(n)}\psi(1 + i_D) - n \max(0, z(n)p_T - k)$
Short Put	$\Delta \Pi_S^P(n) = \frac{n}{z(n)}\phi(1 + i_D) - n \max(0, k - \frac{p_T}{z(n)})$
Short Equity Position	$\Delta \Pi_S^S(n) = \frac{n}{z(n)}p_t(1 - i_L) - nz(n)p_T - n\delta$

**Table 5.1.** Profit Functions Given Sufficiently Large Cash Endowments

profit functions should be made explicit.

The profit is the net profit at the terminal date of the contract. These are not present values, but nominal future profits. For the long position, although no cost earns interest, the cash used would otherwise have been on deposit and so the change in profit includes the lost interest. The profit function for the position itself is:

$$\Pi_L^C(n) = n \max\left(0, \frac{p_T}{z(n)} - k\right) - nz(n)\psi \quad (5.30)$$

Also of note is the interest adjustment to the short equity position, which is  $1 - i_L$ .

There are multiple ways in which a broker-dealer can manage both option and short equity positions. Some firms charge a special borrowing fee which is really a mark-up of the bid-ask spread, some charge interest on the value of the initial position and some charge interest on the continuous balance. For purposes of this game interest is paid on the initial balance. In the United States interest is not paid on the cash received for the short sale. It is held as collateral and used by the broker-dealer until repaid giving the dealer an additional reward in the form of an interest free loan.

A long European style call option without liquidity costs is valued at  $\max(0, p_T - k)$ . This ignores an important element of profitability, that is realiza-

tion. The cash to cash cycle is exactly that, a cash to cash cycle. The contract delivers  $n$  shares of some firm which can be sold at the bid price of  $\frac{p_T}{z(n)}$ . The owner of the contract does have to deliver  $k$  to exercise the contract.

To see the difference it is best to do so in a world without transaction costs, interest or dividends. A shareholder endowed with 100 shares of ABCorp at time zero who continued to hold the same security at time one and with the same 100 shares made no profit. They started with 100 shares and ended with 100 shares. Even if the price went from \$10 to \$20 no realization happened, the property position is the same, that is the property owner holds 100 shares. If at time three the price were \$3 per share no loss would happen unless the shareholder sold and then it would be based on the initial price and not the interim non-realized prices.

One other important difference between the game and the real world is the absence of taxes. There are no income, property or inheritance taxes. Taxes could alter the equilibrium prices, particularly with differential tax rates.

There is also a difference from the normal construction of a profit function in economics. Normally a profit function could be constructed as price, given a quantity, times that quantity minus the costs to produce that quantity. This isn't quite the construction here. The short call price is bid price which is a composite function of some unknown function and a liquidity adjustment written as  $\frac{\psi}{z(n)}$ . Likewise the cost function contain loss mitigation revenues. The writer of a short put must pay the strike price, but recovers from a total loss by selling the received shares in the open market. As a consequence  $\frac{\psi}{z(n)}$  is subject to the law of demand, but  $\psi$  may not be, provided it increases slower than  $z(n)$  decreases. Costs may be convex in volume, but the cost function is multiplicative with revenues in some places and subtractive in others.

Without this construction, discussing the three parties to transactions, the writer, the buyer and the market maker is very difficult. Still, it results in a less than traditional construction from what would be seen in industrial production economics.

#### 5.12.1.1 Derivatives of Potential Arbitrage Positions

The derivatives for the arbitrage position  $call = put + equity$  are shown below. In the special case where the final price,  $p_T$  is less than the strike price,  $k$ , it is not

necessary to separately close the long position while simultaneously purchasing a short position as implied by the separate equations. The joint position becomes:

$$\Delta \Pi_L^{P+S}(n) = nk - nz(n)\phi(1 + i_D) - nz(n)p_t(1 + i_D) + n\delta \quad (5.31)$$

The put owner delivers the shares, already purchased at time zero, and delivers them for  $k$  as per the contract..

The table of derivatives for the change in profit functions of the two positions with respect to  $p_T$  is as follows:

Type	Price	Derivative of Call	Derivative of Put Plus Equity
Long	$p_T < k$	0	0
	$p_T = k$	undefined	undefined
	$p_T > k$	$\frac{n}{z(n)}$	$\frac{n}{z(n)}$
Short	$p_T < k$	0	0
	$p_T = k$	undefined	undefined
	$p_T > k$	$-nz(n)$	$-nz(n)$

**Table 5.2.** Derivatives

Relatively simple math will show that as the only differences between the cases are the interest rate, the derivatives will be the same for all possible outcomes and all states of nature, given either a long or short position. As such, for brevity, the derivatives with respect to  $p_T$  are not shown for the other endowed states.

It follows that if the change in profitability of calls equals the change in profitability of puts and equity positions then the *no arbitrage* requirement is met. The method to determine the equilibrium price is to determine if it is invariant over changes in  $p_T$ . If it is not, then the differences in the anti-derivatives may not be zero.

### 5.12.1.2 Equilibrium Pricing In Long Positions

For each position one should note that there are three cases. As liquidity costs exist, it could matter if the interval of the bid and the ask prices cover the strike price. As mentioned earlier, an option to buy 10,000 shares of stock at \$10 per share. Now imagine the current prices is \$10.50 per share to sell 100 shares. Exercising enough contracts to close out 100 shares would cost \$1,000 and generate revenue

of \$1050.00. Now imagine that selling 10,000 shares would result in a net price of \$9.50 per share. The exercising party would pay \$100,000 but only receive \$95,000 in revenue. It would be unwise to exercise very many of those contracts.

**CASE I:**  $\frac{p_T}{z(n)} \geq k$

The change in the profit functions of the call versus put plus equity position must equal, as such:

$$\frac{p_T}{z(n)} - k - z(n)\psi(1 + i_D) = -z(n)\phi(1 + i_D) + \frac{p_T}{z(n)} - z(n)p_t(1 + i_D) + \delta \quad (5.32)$$

This reduces down to:

$$\psi = \phi + p_t - \frac{k + \delta}{z(n)(1 + i_D)} \quad (5.33)$$

To extract a little more intuition out of the equilibrium, consider the case of the *at the money* call price. In that circumstance the strike price is also the same as the stock price so  $k = p_t$ . In that case the formula becomes:

$$\psi = \phi + p_t \frac{z(n)(1 + i_D) - 1}{z(n)(1 + i_D)} - \frac{\delta}{z(n)(1 + i_D)} \quad (5.34)$$

For a long position the interpretation is that the price of a call option is equal to the price of a put option plus the carrying cost of buying the initial shares, marked down to present value and adjusted for liquidity costs minus the present value of dividends missed by holding the call position, again adjusted for liquidity costs.

Two other features are important here. First, the formula is independent of any value of  $p_T$  and so no uncertainty is present. Second, although  $\delta$  does not bear the usual notation for an estimator as would be the case if it were represented as  $\hat{\delta}$ ; this is done for convenience as dividends, like liquidity costs, appear everywhere.

**CASE II:**  $k \geq p_T$

The equation for this case is:

$$-z(n)\psi(1 + i_D) = k - z(n)\phi(1 + i_D) - z(n)p_t(1 + i_D) + \delta \quad (5.35)$$

This reduces down to:

$$\psi = \phi + p_t - \frac{k + \delta}{z(n)(1 + i_D)} \quad (5.36)$$

**CASE III:**  $kz(n) > p_T > k$

This case has two subcases, one in which the contracts are exercised as an *all or nothing* execution and those that permit *partial execution* of the total position.

**CASE IIIa: All or Nothing Execution**

If the contract required all or nothing execution then the contract would not be executed. This is the above example of having a nominally quoted price greater than  $k$ , but when marked down for volume results in a net price below the strike. For a simple formal proof, let  $p_T = kz(n) - \xi z(n)$ ,  $\xi > 0$ , then it follows that the profit function for a call option is  $-z(n)\psi(1 + i_D)$  due to the fact that the value of the contract would be  $\max(0, -\xi) = 0$ .

This leads to a somewhat surprising result in academic models that ignore liquidity, that is that the put option should be exercised even though the contract is out of the money.

Again, subject to the overall restriction, let  $p_T = kz(n) - \xi z(n)$ ,  $\xi > 0$ . It follows that the value of the equivalent position is:

$$k - \frac{kz(n) - \xi z(n)}{z(n)} - z(n)p_t(1 + i_D) + \delta, \quad (5.37)$$

as this is greater than the non-exercised profit by an amount  $\xi$ , the option must be exercised for maximal profitability.

The equilibrium is the same as for the low price equilibrium, which is:

$$\psi = \phi + p_t - \frac{k + \delta}{z(n)(1 + i_D)} \quad (5.38)$$

**CASE IIIB: Partial Execution**

In the case where some, but not all contracts could be executed profitably, it is assumed there exists a quantity  $n'$  such that  $0 < n' < n$  and that the execution of  $n'$  contracts is profit maximizing. Although the existence of a profit maximizing quantity has yet to be shown, it is assumed that it exists here.

The terminal profit function,  $\Delta\Pi_L^C$ , becomes:

$$\Delta\Pi_L^C = n' \left[ \left( \frac{p_T}{z(n')} - k \right) - z(n)\psi(1 + i_D) \right] - (n - n')z(n)\psi(1 + i_D) \quad (5.39)$$

This reduces to:

$$\Delta\Pi_L^C = n' \left( \frac{p_T}{z(n')} - k \right) - nz(n)\psi(1 + i_D) \quad (5.40)$$

The terminal profit function,  $\Delta\Pi_L^{P+S}(n)$  becomes:

$$\begin{aligned} \Delta\Pi_L^{P+S}(n) = n' & \left[ \frac{p_T}{z(n')} - z(n)p_t(1 + i_D) + \delta - z(n)\phi(1 + i_D) \right] \\ & + (n - n')[k - z(n)\phi(1 + i_D) - z(n)p_t(1 + i_D) + \delta] \end{aligned} \quad (5.41)$$

In equilibrium, this reduces to:

$$-nz(n)\psi(1 + i_D) = nk - nz(n)p_t(1 + i_D) + n\delta - nz(n)\phi(1 + i_D) \quad (5.42)$$

Which is:

$$\psi = \phi + p_t - \frac{k + \delta}{z(n)(1 + i_D)} \quad (5.43)$$

For all long positions, where the endowment of cash is positive and sufficient to cover the cost of the positions, the equilibrium condition is:

$$\psi = \phi + p_t - \frac{k + \delta}{z(n)(1 + i_D)} \quad (5.44)$$

For subsequent cases, calculations of *Case III* are omitted as it is simply a variation of coefficients from the above case due to different interest rates.

### 5.12.1.3 Equilibrium In Short Positions

Although the profit equations for long and short option positions are the additive inverse of each other, this is not true for the long and short equity position. This difference results in a bid-ask spread even without liquidity costs, such as where  $\lambda = 0$ . The difference between the gross amount of a call price the market would

be willing to pay,  $\psi_L$ , is greater than the insurer requires,  $\psi_S$ . This violates the law of one price. It could be interpreted, however, that in equilibrium, the spread is sufficient to support commissions to an insurance agent to link the parties together.

Proper understanding of this difference is that *if* this condition is the equilibrium condition, then neither long nor short participants can form an arbitrage position against the market maker *if* the market maker keeps the spread between the prices.

**CASE I:**  $z(n)p_T \geq k$

There are slight mathematical differences created by going from long to short, but it is the short case which is critical from a policy making perspective. It is the option writer that needs sufficient reserves to support the system. As the buyer has no method to inspect the writers, indeed, in the American over the counter market exercise by long holders is exercised by random assignment, it is dependent upon the market makers to set adequate reserve and collateral requirements.

The equilibrium condition is:

$$\psi \frac{1 + i_D}{z(n)} + k - z(n)p_T = \phi \frac{1 + i_D}{z(n)} + p_t \frac{1 - i_L}{z(n)} - z(n)p_T - \delta \quad (5.45)$$

This resolves to:

$$\psi = \phi + p_t \frac{1 - i_L}{1 + i_D} - \frac{k + \delta}{1 + i_D} z(n) \quad (5.46)$$

**CASE II&III** For economy of space, the calculations are omitted for the other cases as they also resolve to equation 5.46.

### 5.12.2 No Endowment

In this state of nature, it is assumed that  $\bar{m} = 0$ . Further it is assumed that the participants have sufficient access to credit as to be able to make purchases at an interest cost of  $i_L$ . The profit function, should no action be taken to enter into a position, is:

$$\Pi_{\bar{m}=0} = 0 \quad (5.47)$$

The profit functions of this state of nature are in table 5.3.

There are slight differences in this state of nature from the large endowment state. In order to enter into a long position the participant has to borrow funds

Type of Position	Formula for Profit Function
Long Call	$\Delta\Pi_L^C(n) = n \max(0, \frac{p_T}{z(n)} - k) - nz(n)\psi(1 + i_L)$
Long Put	$\Delta\Pi_L^P(n) = n \max(0, k - z(n)p_T) - nz(n)\phi(1 + i_L)$
Long Equity Position	$\Delta\Pi_L^S(n) = \frac{n}{z(n)}p_T - nz(n)p_t(1 + i_L) + n\delta\frac{1+i_L}{1+i_D}$
Short Call	$\Delta\Pi_S^C(n) = \frac{n}{z(n)}\psi(1 + i_D) - n \max(0, z(n)p_T - k)$
Short Put	$\Delta\Pi_S^P(n) = \frac{n}{z(n)}\phi(1 + i_D) - n \max(0, k - \frac{p_T}{z(n)})$
Short Equity Position	$\Delta\Pi_S^S(n) = \frac{n}{z(n)}p_t(1 - i_L) - nz(n)p_T - n\delta\frac{1+i_L}{1+i_D}$

**Table 5.3.** Profit Functions Given Sufficiently Large Cash Endowments

and so costs are marked up by the interest rate. This is opposite the short side where any money received goes to a deposit account. There is also a difference in both equity positions.

In both equity positions  $\delta$  is a future value. In a long equity position the receipt of dividends would pay down the debt from the purchase and so must be discounted back to present value so it can earn the commercial loan rate of interest. In the short equity position passed dividends are no longer paid from an endowment of cash. As such, passed dividends must be paid from borrowings at the commercial loan rate. As  $\delta$  is defined with reference to the deposit rate, it must first be discounted back to the present value to be costed out at the commercial loan rate.

Basic algebra confirms that the results will be of the same form, but with different coefficients. For the long position, all cases result in the equilibrium formula:

$$\psi = \phi + p_t - \frac{k}{z(n)(1 + i_L)} - \frac{\delta}{z(n)(1 + i_D)} \quad (5.48)$$

For the short position, all cases result in the formula:

$$\psi = \phi + p_t \frac{1 - i_L}{1 + i_D} - \frac{k(1 + i_D) + \delta(1 + i_L)}{(1 + i_D)^2} z(n) \quad (5.49)$$

### 5.12.3 Large Endowment of Debt

In this state of nature, it is assumed that  $\bar{m} \ll 0$ . It is assumed that the participants have sufficient access to credit facilities as to be able to make purchases at an interest cost of  $i_L$ . It is further assumed that no cash revenue is sufficient to entirely pay down the debt to a positive cash position. The profit function, should

no action be taken to enter into a position, is:

$$\Pi_{\bar{m} \leq 0} = -i_L \bar{m} \quad (5.50)$$

The profit functions of this state of nature are in table 5.4.

Type of Position	Formula for Profit Function
Long Call	$\Delta \Pi_L^C(n) = n \max(0, \frac{p_T}{z(n)} - k) - nz(n)\psi(1 + i_L)$
Long Put	$\Delta \Pi_L^P(n) = n \max(0, k - z(n)p_T) - nz(n)\phi(1 + i_L)$
Long Equity Position	$\Delta \Pi_L^S(n) = \frac{n}{z(n)}p_T - nz(n)p_t(1 + i_L) + n\delta \frac{1+i_L}{1+i_D}$
Short Call	$\Delta \Pi_S^C(n) = \frac{n}{z(n)}\psi(1 + i_L) - n \max(0, z(n)p_T - k)$
Short Put	$\Delta \Pi_S^P(n) = \frac{n}{z(n)}\phi(1 + i_L) - n \max(0, k - \frac{p_T}{z(n)})$
Short Equity Position	$\Delta \Pi_S^S(n) = \frac{n}{z(n)}p_t(1 - i_L) - nz(n)p_T - n\delta \frac{1+i_L}{1+i_D}$

**Table 5.4.** Profit Functions Given Sufficiently Large Cash Endowments

This state of nature is distinguished by all transactions either paying down debt or increasing debt. As such, the deposit rate only appears in the discounting of dividends. The equilibrium equation for the long position in this state is:

$$\psi = \phi + p_t - \frac{k}{z(n)(1 + i_L)} - \frac{\delta}{z(n)(1 + i_D)} \quad (5.51)$$

The equation for the short position is:

$$\psi = \phi + p_t \frac{1 - i_L}{1 + i_L} - z(n) \frac{k}{1 + i_L} - z(n) \frac{\delta}{1 + i_D} \quad (5.52)$$

This state of nature is important as it describes the state of nature for the hedge fund industry.

#### 5.12.4 Small Endowment of Cash

In the case where the participant begins with a small endowment of cash, but wishes to make purchases that require the acquisition of debt for profit maximization, it is best to think in terms of the marginal transaction.

This transaction is a single transaction and so is at a single price. The value of the position has to be sufficient to make it worth going into debt to accomplish the purchase. This permits two ways to think about the problem. One would be

to blend the profit function between the deposit and the loan rates. The other would be to concern oneself only with the marginal transaction.

The blended method, while a correct profit function, doesn't represent the last dollar spent. The marginal long position is the same as the wholly indebted state, while the short position is like the large endowment of cash state.

At the margin, this is no different than the no endowment state, that is  $\bar{m} = 0$ . As such, the equilibrium conditions are the same as for that state.

### 5.12.5 Small Endowment of Debt

In this case, the participant begins with a small endowment of debt. Long purchases result in greater debt, but short positions more than pay off the debt. In that case, the participant is actually deciding that the overall position is no longer valuable enough to warrant carrying debt to maintain it.

As in the small endowment of cash state, at the margin, the result is the same as the no endowment state, that is  $\bar{m} = 0$ .

## 5.13 Dominant Pricing

The equilibrium prices for the various endowments for call options are shown in table 5.5.

Endowment	Short Formula	Long Formula
$\bar{m} \gg 0$	$\psi = \phi + p_t \frac{1-i_L}{1+i_D} - \frac{k+\delta}{1+i_D} z(n)$	$\psi = \phi + p_t - \frac{k+\delta}{z(n)(1+i_D)}$
$\bar{m} \approx 0$	$\psi = \phi + p_t \frac{1-i_L}{1+i_D} - \frac{k(1+i_D)+\delta(1+i_L)}{(1+i_D)^2} z(n)$	$\psi = \phi + p_t - \frac{k}{z(n)(1+i_L)} - \frac{\delta}{z(n)(1+i_D)}$
$\bar{m} \ll 0$	$\psi = \phi + p_t \frac{1-i_L}{1+i_L} - z(n) \frac{k}{1+i_L} - z(n) \frac{\delta}{1+i_D}$	$\psi = \phi + p_t - \frac{k}{z(n)(1+i_L)} - \frac{\delta}{z(n)(1+i_D)}$

**Table 5.5.** Endowment Specific Equilibrium Pricing for Call Options

### 5.13.1 Put Contracts

The formula for  $\phi$  is simpler and less diverse than the call option. As those who have an endowment of debt have a higher discount rate, then they are willing to accept a lower price. As such, the lowest price wins the contract on the short side. The simple interpretation is that those who are willing to accept significant

amounts of debt on their balance sheets are willing to accept smaller premiums to pay for risks.

For those endowed with cash to write a contract at the higher price either implies that those participants endowed with debt have reached their equilibrium balance sheet and/or credit restrictions prevent further underwriting.

This implies that hedge funds and financial institutions should dominate the market for writing put contracts.

The long put is priced by symmetry as  $z(n)\phi$ . As the price in the normal state is low, all types of endowments should be willing to buy long put contracts.

### 5.13.2 Long Call Contracts

For those willing to carry debt to buy long call contracts, the reservation price is greater than the reservation price for those unwilling to carry debt. This, of course, makes sense. Someone carrying debt would find insurance to be of greater value. A leveraged loss is magnified by the proportion the balance sheet is leveraged.

### 5.13.3 Short Call Contracts

The reservation price for  $\phi$  for those not heavily indebted is greater than for those who are heavily indebted. Since coherence would require participants willing to enter into either side of the position, either call or put plus a loan, the presence of heavily indebted parties willing to underwrite contracts should preclude the other two sets of endowment pricing from becoming operative. This is due to the unwillingness to compete with indebted parties to underwrite puts.

If the case exists where heavily indebted parties are no longer willing to underwrite call or put contracts, then those near zero would offer lower prices than those heavily endowed with cash. Only in the case where parties are unwilling to incur increased debt in order to underwrite contracts, as those near zero are required to do when shorting stocks and shorting dividends, will those heavily endowed with cash be able to get their reservation price. Of course the market maker, as an indebted party, would lose money on each transferred risk. So either this state does not exist, or it appears in the run state. Cash rich investors appear when there is blood in the street or when the government bails out the system.

## 5.14 Speculators

Allowing risk loving actors into the game is not disruptive on the long side as a risk loving actor would be willing to pay a higher premium than a risk averse person for the right to purchase a gamble. It is disruptive on the short side of the market.

Risk loving individuals will pay a premium for each gamble, in effect guaranteeing that their wealth will go to zero given enough time. This implies that risk loving individuals, given enough time, will default on option contracts at high rates.

In markets with nationally insured banks a market maker would love to accept the higher than normal option prices from risk loving buyers and sell options at unusually large discounts to risk loving sellers. Until the market craters the makers will make unusually wide profits. The collapse of the writers would result in the collapse of the market makers, but this no longer is the case under too big to fail doctrines. In that case, the shareholders of the market makers keep the unusually wide profits, but get recapitalized to do it again from risk averse tax payers.

Two remedies to this are to require minimum regulatory option premiums with prudential regulations similar in form to that found in the insurance and reinsurance industry. The alternative is to have a no bailout provision in constitutional law. The challenge of the no bailout provision is that it could be an incredibly costly solution.

## 5.15 Expected Profit as a Negative Bayesian Risk Function

The expected profit of the short put has a representation as a negative Bayesian risk function. Focusing on the expected profit of the short put can also be thought of as part of a decision rule.

A decision rule that states, subject to  $n$ , the market maker will buy or sell contracts or the underlying instruments at a stated price implies the market maker is indifferent among the choices.

Such a decision implies a rule exists whereupon the market maker maintains a free substitution among actions,  $a, a'$ .

This warrants another look at the profit function. Note that an equivalent proof could be constructed using the Envelope Theorem.

The change in the profit function created by underwriting for the writer of the short put is:

$$\Delta\Pi_S^P(n) = \frac{n}{z(n)}\phi(n)(1 + i_X) - n \max\left(0, k - \frac{p_T}{z(n)}\right) \quad (5.53)$$

The market maker must accept no less than this minimum amount, although in most circumstances, the market maker received more revenue. This is the minimum net revenue function for the market maker and the revenue function for a writer.

Redefining the change in profit function as the loss function for a decision rule and noting the dependence on the parameters it is:

$$L(\boldsymbol{\theta}, \delta, n) = -\Delta\Pi_S^P(n, \boldsymbol{\theta}) \quad (5.54)$$

Defining  $f(p_T|\boldsymbol{\theta})$  as the predictive distribution for a given set of parameters, the risk function from definition 11 becomes nothing more than the negative of the expected profit function as the risk is:

$$R(\boldsymbol{\theta}, \delta, n) = \int_0^\infty L(\boldsymbol{\theta}, \delta, n) f(p_T|\boldsymbol{\theta}) dp_T \quad (5.55)$$

Noting that the option prices were priced with reference for chosen values of the parameter set, it must be noted that the true values of the parameter set are unknown. Instead, market makers price contracts based on their beliefs about the true values. So, while it is necessary for the underwriter to price with regard to  $n$ , they must do so based on their beliefs which are random variables.

If  $\pi(\boldsymbol{\theta})$  is the posterior density function then the Bayes risk associated with the decision strategy is:

$$r(\pi, \delta, n) = \int_{\Theta} \int_0^\infty [L(\boldsymbol{\theta}, \delta, n) f(p_T|\boldsymbol{\theta}) dp_T] \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (5.56)$$

Note that the random variate,  $\boldsymbol{\theta}$  no longer exists as a variable as it has been marginalized out. Notice also that utility of choosing which specific action to choose is gone as well. Further, the final price no longer influences decisions. Only

three variables remain, the decision rule, the posterior density and  $n$ . For a given data set,  $\pi$  has a fixed functional form. It can be treated as an invariant function at any arbitrarily chosen time,  $t$ .

So, given time  $t$ , the Bayes risk transforms into:

$$r(\delta, n). \quad (5.57)$$

A decision rule where the decision maker always chooses the value of  $n^*$  that maximizes profitability also always minimizes Bayes risk. A couple of observations also appear to be in order regarding the admissibility of  $n^*$ . First, the posterior should be unimodal in the individual parameters, although that isn't strictly necessary. Further, if a profit function exists then it provides a positive expected profit. This is just a requirement of rationality, otherwise the market maker would withdraw and no activity would happen. From the formula above, however, note that both sides of the profit function are multiplied by:

$$\frac{n}{z(n)} \quad (5.58)$$

This precludes infinite profits as:

$$\lim_{n \rightarrow 0} \frac{n}{z(n)} = 0 \text{ and } \lim_{n \rightarrow \infty} \frac{n}{z(n)} = 0 \quad (5.59)$$

So  $n^*$  is a positive real number and a Bayes estimator. If there are multiple values that maximize profit then they have equivalent risk functions and as the risk is minimized they are all admissible. It is improbable that the estimate of  $n^*$  will be multimodal, given whatever data there is, and as such is likely a unique Bayes estimator.

## 5.16 Conclusion

Leveraged institutions will, in most circumstances, set the price. Although this model would allow for an array of prices, it should be leveraged firms that are the price setters.

This model is rather simple. Integrate over the area at risk, find the expecta-

tion, subtract expected value from the strike price. Discount that value to present value, account for dividends, and set the price of calls as an equilibrium with puts. None of that is actually very difficult as concepts. The implementation is challenging, but the concepts are simple.

Pricing American style options requires the knowledge of as yet unknown mathematics. There isn't a continuous time model for paths that look like the models presented here. That should not be a problem as there is a very simple discrete solution.

The American style option allows exercise any time within the contract period. For a one year contract, the only requirement is estimating the probability of early exercise for each day in the contract time period. For a one year option contract, given there are around 255 trading days, this would be the same as calculating 255 options, each one day longer than the last and adjusting for the likelihood of early exercise.

Asian style options require greater statistical theory than may be present. Asian style options are based on the sample mean of prices over the period. As the underlying has no mean, and as the sample mean is a Cauchy distributed random variable, significant work will need to be performed. By not fixing the strike price, but setting it at the average price a significant challenge is posed in attempting to price them.

A further challenge comes in pricing interest rate options, commodities and more exotic items such as weather derivatives. Not touched were the prices for futures and forwards contracts. A lot of work needs to be performed to complete this set of option contracts.

# EMPIRICAL TESTING OF THE PROPOSED OPTION PRICING MODEL

## 6.1 Introduction

Empirical testing of the proposed model requires some boundary conditions on the sampling period and commentary on both the data and any prior distributions. Only the estimation of one-year at-the-money equity put contracts were considered as the alternative would have required substantial additional estimation. Because of the inability to access certain information due to its non-public nature, the test is necessarily restricted.

In particular, liquidity information regarding option contracts is proprietary information of the market makers and the U.S. Securities and Exchange Commission. Attempts to acquire such data were rebuffed. Further, actual trading data on option prices is not available. As such, only a test of future values given the implicit liquidity costs and theoretical option prices can be constructed.

As the model only requires the estimation of expected values over bounded regions a substantial amount of estimation normally required in Bayesian methodologies can be avoided. Some important limitations were introduced for compu-

tational simplicity. In particular, many tools in standard use in the field, such as logistic regression, were avoided due to the computational cost this creates for Bayesian updating. This implies that better models could be created for any significant component of the model; but, as computational time and memory allocation were important, choices were made in favor of computational efficiency over mathematical efficiency. Indeed the author chose to follow the advice of Frederick Mosteller regarding computational efficiency.

Until recently mathematical statisticians have spent a great deal of effort developing “efficient statistics” and “most powerful tests.” This concentration of effort has often led to neglect of questions of economy. Indeed some may have confused the meaning of technical statistical terms “efficient” and “efficiency” with the layman’s concept of their meaning. No matter how much energetic activity is put into analysis and computation, it seems reasonable to inquire whether the output of information is comparable in value to the input measured in dollars, man-hours or otherwise.[66]

Necessarily then this implies that the test of the model should be noisier than a more efficient and better model. Although the model is admissible when compared to an equivalent Frequentist model, the model may not be admissible when compared to a superior Bayesian or Frequentist model. Further, as is necessary in any Bayesian model, choices regarding prior distributions and their hyperparameters were arbitrary. That said, they are an honest attempt to model the views of the author. It is necessary to model bankruptcy risk, merger risk and the density function for returns given the various possible states of the world that could be present in the future.

If a proof that the sample mean was a sufficient statistic for the expectation of any Lebesgue-integrable, bounded, but unknown distribution then a distribution-free solution naturally follows. The author was unable to find such a proof and did not spend any time trying to create such a proof. Nonetheless, if it does or can exist, then a wide range of possible alternatives exists.

## 6.2 Bankruptcy Estimation

Bankruptcy represents the simplest of the estimation problems present in the model. Because of how bankruptcy is defined, a party is either bankrupt or they are not. This makes the likelihood function the binomial distribution. Because the data set is so large this provides two alternative ways to model the prior and the likelihood.

First, the normal approximation to the binomial could be used and a normal or uniform prior could be used. As the only concern is the expectation, either tool would result in the same solution with the difference only being one of computational complexity. Second, the beta distribution could be used as the prior distribution resulting in a beta distribution as the posterior distribution. As the maximum a posteriori estimator of the expectation would be the same either way, the beta posterior was used due to some simplifications available regarding model selection.

The data used was the Compustat data set of quarterly and annual financial data, the CRSP data set for security prices and the final disposition of a security, and Gross Domestic Product data for the United States as produced by the Bureau of Economic Analysis. As bankruptcy data goes back to the beginning of the CRSP data in 1925, Compustat data does not become available until January 1962. As such a prior was created based on CRSP data alone, unconditioned on either accounting or Gross Domestic Product data.

### 6.2.1 Bankruptcy Data from 1925-1961

For most of the period, the CRSP data set primarily included only the largest firms in the United States. As such, these firms were also the least likely to go bankrupt and so a prior distribution based only upon this data would tend to skew the prior probability downward. This forces one of two possible choices.

The first choice is to discard the data as non-representative of the data. The challenge with this choice is that it requires discarding the data from the Great Depression and World War II. Although these are not representative events in American history, ignoring them requires ignoring periods of tremendous financial and social stress. The second choice is to include the data, but to reduce its weight

in some manner that can easily be objectionable. There is no way to weight the value of this data objectively as there is no alternative reference set to norm it against.

Nonetheless, some general principles were used to alter the weight of the prior distribution, though the subjective nature of the decisions involved in dosing is unavoidable. A goal of minimizing the impact of objections, while preserving information was used.

Alternatives to weighting including creating a mixture distribution of the beta distribution and the uniform distribution, or mixing the beta distribution with data from the bond markets as to bankruptcy rates for the period. The later choice has to be excluded for empirical reasons. As mentioned before, bond data does not seem responsive to actual realizations of default rates and so there is no evidence that bond data will be adequate to gather short run bankruptcy data.[97] The former choice results in an objectionable choice of weights.

There is a non-trivial problem that could be created by ignoring the data from the Depression and using some alternative, weakly informative prior such as the triangular distribution, the Haldane distribution or the uniform distribution. That problem is that some combinations of accounting data should result in a prior sample with absolutely no observations of bankruptcy. Although this post 1961 group of firms represents real data, from the viewpoint of pre-1962 data, especially the Great Depression, it seems implausible that the resulting prior distributions would be credible to anyone.

Consider them in order of the implied number of observations, remembering that extreme combinations of observations are relatively rare and that accounting data is only quarterly data at best. Under the Haldane prior any observation of no bankruptcies in a subgroup during the period prior to the sample period but after 1961 would result in an expectation of a zero percent chance of bankruptcy. This implies that there are firms where it is impossible for bankruptcy to happen, even during periods of extreme financial stress. Because some combinations of accounting data are rare, it may not be surprising to find zero bankruptcies in one hundred observations. That does not mean that the true bankruptcy rate is zero, merely that none were observed.

For the uniform prior, with one hundred observations, the implicit bankruptcy

rate for the rarely observed group would be 0.99%. For the triangular prior the expectation would be 0.97%. The role of the prior would be very strong and would determine the implicit rate of bankruptcy.

For subgroups with few observations, the impact of the prior upon the posterior distribution is unavoidable. For the author, the choice came down to discarding data from periods of extreme stress and choosing a reasonable but weak informative prior distribution.

The author chose a rather simple method, similar in substance to a maximum entropy method. Two rules were used. The first rule is that the expectation of the observed bankruptcies would be preserved in close approximation. The second rule is that the variance would be maximized, while maintaining a proper density, by adopting the rule for the parameters of the beta distribution for the unmixed data thus:

$$\frac{1}{\beta} = \frac{\text{number of observed bankruptcies}}{\text{number of survivals}}, \alpha \equiv 1 \quad (6.1)$$

The maximum entropy method, unconditioned on prior information would be the uniform distribution. Values of  $\alpha, \beta < 1$  guarantee negative entropy values. However the method of maximum entropy is built around constraints that are verifiable in data. This differs from the method of maximum entropy primarily by enforcing the use of integers instead of the gamma function. Factorial based solutions are not difficult in SAS, the gamma function generalizes factorials and is computationally expensive. The resulting prior distribution has less entropy than the true maximum entropy value. The second reason the difference matters less is that this element of the prior distribution is only weighted ten percent, to be discussed below.

There were 10,827,796 observed firm-days in the period. Each day a recorded firm existed in the data set is counted as one firm-day. There were 34,310 observed firm days where bankruptcy is observed within one year of the observation date. Rounded to the nearest integer, this implies a beta distribution where  $\alpha = 1$  and  $\beta = 315$ .

This gives a density function for the largest firms of:

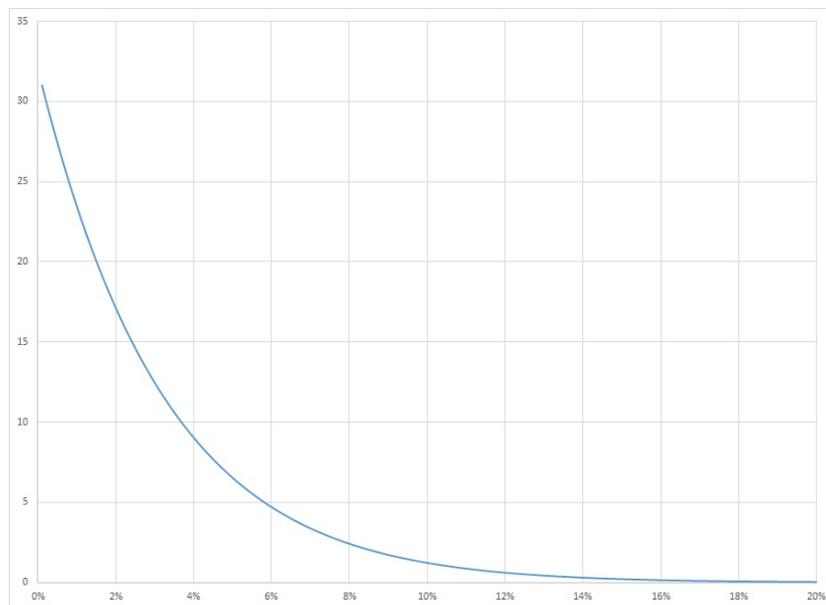
$$\Pr(\text{bankruptcy}) = 315(1 - p)^{314} \quad (6.2)$$

The author chose to give this prior data a weight of ten percent and a uniform distribution the weight of ninety percent. Although there is some arbitrariness in the choice, it was selected by keeping the properties of other data in the future sample in mind. In particular, the sample size would grow to about ten times the original sample size in the data. Although the average weighting should be less, there is another property of the sample that needed attended to, that of size. As time has passed since the passage of the Securities Acts during the Great Depression and due to technological changes in trading and accounting costs, the relative size of the firm traded in the capital markets has fell pretty dramatically.

As smaller firms have greater exposure to regional crisis they should be at greater risk of failure, how much greater is unclear. Further, partitioning on accounting and economic data should result in important differences in rates. Using these rules results in a beta prior distribution of:

$$\Pr(\text{bankruptcy}) = 32(1 - p)^{31} \quad (6.3)$$

Graphically, the density function for the probability density of the prior distribution of bankruptcy, unconditioned upon accounting data is peaked at zero and falls off quickly to the asymptote.



**Figure 6.1.** Probability Density Function for the Prior Probability of Bankruptcy.

In the form of an expectation, this results in an expected bankruptcy rate of approximately 3.03%. For the above example of one hundred observations of no bankruptcies the expected rate would be 0.76%. While this is substantially lower than the prior expectation used, it is about twice the historical rate observed during the period. It also has the virtue of granting at least some weight out into the low twenty percent region. If the accounting data is in fact a valuable predictor of bankruptcy rates, then some large rates should be expected.

Further, if the prior distribution is inappropriate, then this could be an important factor in the falsification of the proposed model.

### 6.2.2 Bankruptcy Data from January 1962-June 1990

Beginning in 1962 both Gross Domestic Product(GDP) data and accounting data were incorporated into the estimation of bankruptcy rates. As with the study by Sun and Shenoy, Pearson-Tukey groups were constructed. Unlike the Sun and Shenoy research missing data were not reconstructed using an alternate layer of data that was correlated with the underlying data. Rather, the models with the largest posterior density were used where the data was available and then combined using model averaging. If information sufficient for one model but not the other existed then only one model was used. If the accounting information for the models with greatest posterior density was missing then estimation was done without accounting data. If no accounting data was available, then only GDP was used to estimate the rate of bankruptcy. The models were constructed using Pearson-Tukey groupings of change in quarterly GDP and pairings of dissimilar accounting ratios.

This methodology has a number of weaknesses. First, industry groups were ignored. While this was done for computational simplicity, it is quite probable that larger industry groupings may have been an important variable. Second, the model choice was frozen on June 30, 1990 to minimize computation. This prohibits drift in the model when drift may be called for. Third, it presumes a unique equilibrium for all industries and that this equilibrium is the same; implying the true model does not change with changing technologies and social circumstances. Fourth, it ignores security prices. While this avoids the “cart before the horse,” problem it

ignores an important form of data.

Because of both inflation and the increase in firm diversity with time, raw accounting data is difficult to compare over time. Accounting ratios normalize the data with respect to inflation, but not with respect to the increasing diversity. Restructuring accounting ratios into quantile data reduces the magnitude of the diversity effect. By grouping the quantiles of the ratios into Pearson-Tukey groups the tails and the body are accounted for.

Because accounting data is strongly correlated by design, there is little independent information in various combinations of data. The variables used carry correlations, measured as Pearson product moment correlation coefficients, of between 0.60 and 0.96. By partitioning the data into mutually exclusive sets the problem of orthogonality is resolved.

Four groups of accounting ratios were used with a rule that no model would use a pairing within the same class of accounting ratios. The list of classes and ratios are in table 6.1. Since there is no reason to believe any combinations of ratios are any better than any other, a uniform prior is set over each of the seventy-eight combinations. Generally speaking Bayesian methods should exhaust the model space and as such, sets with one ratio each would normally be considered.

Class of Ratios	Ratio
Profit Margins	Net Margin
	Gross Margin
Return Ratios	Return on Equity
	Return on Assets
	Return on Net Assets
Debt Ratios	Debt to Equity
	Debt Ratio
	Long Term Debt to Equity
Liquidity Ratios	Acid Ratio
	Cash Ratio
	Current Ratio
Turnover Ratios	Earning Power Ratio
	Asset Turnover
	Receivable Conversion

**Table 6.1.** List of Ratio Classes with Members of the Classes

This was not done as the implications for firms reporting just enough data

to only calculate a single ratio have doubtful accounting data anyway. It is also possible to remove GDP as an element of the data, however this was not done for theoretical reasons.

Although it would be unsurprising to see accounting data vary with GDP data, the accounting data has been contemporaneously ranked. As such, the effect of a low ratio in one period may very well be considered a high ratio in another. This would be lost with accounting ranks. GDP is captured to reflect uncertainty and errors of actors outside the firm and as such is a latent variable.

Bayesian model averaging was used among the highest probability models, selecting the smallest number of models to cover the densest 99% region of the model space. Bayesian model averaging works by weighting a parameter of interest, in this case the probability of bankruptcy, by the probability that the model is the true model, normalized to one.

For example, if the first model predicts a 5% chance of bankruptcy and the second model predicts a 10% chance of bankruptcy, the probability of bankruptcy would be weighted by the probability either model is the true model. If there were a 60% chance the first model was correct then the expectation of the two models is:

$$.05 \times .6 + .10 \times .4 = .07 \quad (6.4)$$

Since the true model is unknown, information is captured from both models through a blended posterior density function. As this dissertation is only concerned with expectations, this somewhat more robust view is lost.

For observations where the accounting ratios are missing for both models, then only GDP is used as a factor.

The model containing the current ratio and the accounts receivable conversion period had a posterior probability of 53.163%. This was marginally better than the model containing the net profit margin and the return on assets, which had a posterior probability of 46.829%. The probability that one of the two models is a fair representation of the true model, given the model restrictions and the data, is 99.992%. As such, the probability that any one of the remaining seventy-six models is a fair representation of the true model is only one-one hundred and twenty-fifth of one percent. As this is outside the credible set, they are excluded from the model construction.

The resulting partitions and their associated parameter estimates, where  $\alpha$  and  $\beta$  are the prior parameters,  $k$  is the number of bankruptcies observed and  $n$  is the total number of observations are presented in **Table 6.2**. Fundamentally, this model deals with two things, the ability to fund current operations and the speed with which delayed payments are received. Interestingly, one must look at the empirical data to make good predictions. For example, a firm with a poor ability to finance its current operations from its *own* assets, with slow repayment by its own customers and facing a sharp general economic decline has a relatively low bankruptcy rate.

Current Ratio	Accounts Receivable Conversion Period	Change in GDP	$\alpha$	$\beta$	k	n	Expected Probability of Bankruptcy
Low	Low	Strong Downward Shock	1	32	0	540	0.17%
Low	Low	Midrange	1	32	4919	100699	4.88%
Low	Low	Strong Upward Shock	1	32	459	6042	7.57%
Low	Medium	Strong Downward Shock	1	32	450	6480	6.92%
Low	Medium	Midrange	1	32	44130	456465	9.67%
Low	Medium	Strong Upward Shock	1	32	3272	40840	8.01%
Low	High	Strong Downward Shock	1	32	0	90	0.81%
Low	High	Midrange	1	32	4908	34568	14.19%
Low	High	Strong Upward Shock	1	32	331	2691	12.19%
Medium	Low	Strong Downward Shock	1	32	450	11165	4.03%
Medium	Low	Midrange	1	32	34859	890758	3.91%
Medium	Low	Strong Upward Shock	1	32	2912	54570	5.33%
Medium	Medium	Strong Downward Shock	1	32	6030	124957	4.83%
Medium	Medium	Midrange	1	32	319793	9234218	3.46%
Medium	Medium	Strong Upward Shock	1	32	22514	486126	4.63%
Medium	High	Strong Downward Shock	1	32	810	3960	20.31%
Medium	High	Midrange	1	32	29325	366148	8.01%
Medium	High	Strong Upward Shock	1	32	1546	20652	7.48%
High	Low	Strong Downward Shock	1	32	0	900	0.11%
High	Low	Midrange	1	32	2093	65369	3.20%
High	Low	Strong Upward Shock	1	32	181	5392	3.35%
High	Medium	Strong Downward Shock	1	32	450	9396	4.78%
High	Medium	Midrange	1	32	21836	1005458	2.17%
High	Medium	Strong Upward Shock	1	32	2369	67300	3.52%
High	High	Strong Downward Shock	1	32	0	90	0.81%
High	High	Midrange	1	32	1555	64726	2.40%
High	High	Strong Upward Shock	1	32	90	2183	4.11%

**Table 6.2.** Bankruptcy Prior for Current Ratio and Rec. Conversion Period

This somewhat counter-intuitive result isn't surprising if one assumes the management of such firms are proactive in managing their assets. While such a firm

should be struggling just to meet payroll, the management would rationally seek a line of credit to smooth out current operations and this very line of credit permits a firm to make only minimal payments from their own operations. It isn't safe to stereotype combinations of data without looking at why that data likely exists. The ability to get others to finance current operations allows management to extend longer credit to customers and to turn over cash and its equivalents quickly.

The alternative model uses two highly correlated measures, the net profit margin and the return on assets. Certain combinations of data are impossible in this pairing. It is impossible for a firm with the highest operating losses to also have the highest returns on assets. Because of this internal correlation, it is the differences between the measures that are informative. The sole difference is various forms of leverage. A firm can have a low net margin but an extremely high return on assets if the firm is highly leveraged. The model implicitly measures profitability as a function of leverage.

The prior distribution for the alternative model can be found in **Table 6.3**. Finally, the prior for the distribution for the case where no accounting data exists is shown in **Table 6.4**.

Net Margin	Return on Assets	Change in GDP	$\alpha$	$\beta$	k	n	Expected Probability of Bankruptcy
Low	Low	Sharp Decline	1	32	990	4505	21.84%
Low	Low	Midrange	1	32	98013	637046	15.38%
Low	Low	Sharp Increase	1	32	12101	77972	15.51%
Low	Medium	Sharp Decline	1	32	270	900	29.05%
Low	Medium	Midrange	1	32	16578	166359	9.96%
Low	Medium	Sharp Increase	1	32	270	3916	6.86%
Medium	Low	Sharp Decline	1	32	720	3240	22.03%
Medium	Low	Midrange	1	32	18509	160367	11.54%
Medium	Low	Sharp Increase	1	32	457	4594	9.90%
Medium	Medium	Sharp Decline	1	32	6210	140662	4.41%
Medium	Medium	Midrange	1	32	346706	11002330	3.15%
Medium	Medium	Sharp Increase	1	32	20487	561672	3.65%
Medium	High	Sharp Decline	1	32	270	11160	2.42%
Medium	High	Midrange	1	32	18702	900047	2.08%
Medium	High	Sharp Increase	1	32	1458	54068	2.70%
High	Low	Midrange	1	32	272	1273	20.90%
High	Medium	Sharp Decline	1	32	180	5940	3.03%
High	Medium	Midrange	1	32	11604	660855	1.76%
High	Medium	Sharp Increase	1	32	725	45895	1.58%
High	High	Sharp Decline	1	32	180	4230	4.25%
High	High	Midrange	1	32	10068	451823	2.23%
High	High	Sharp Increase	1	32	910	24899	3.65%

**Table 6.3.** Bankruptcy Prior for Net Margin and ROA

Change in GDP	$\alpha$	$\beta$	k	n	Expected Probability of Bankruptcy
Sharp Decline	1	32	1286	18813	6.83%
Midrange	1	32	110171	8596956	1.28%
Sharp Increase	1	32	4276	357635	1.20%

**Table 6.4.** Bankruptcy Prior When Missing Accounting Data

## 6.3 Merger Estimation

Sample based objections to merger estimation that were applicable to Bankruptcy estimation are equally applicable to merger estimation. As with bankruptcy estimation, logit/probit style models were not used. Rather, a similar mechanism as

was used for bankruptcy estimation was used for merger rate estimation. This was due to the computational expense created by high frequency Bayesian updating of probit or logit models. Any generalized linear model is computationally expensive in SAS when Bayesian methods are used. They almost always involve Markov Chain Monte Carlo methods and create a non-analytic posterior. As such, there is no convenient way to program the past posterior distributions into becoming future prior distributions. This is a significant limitation. It is hard to argue that a better model could not be formed. There is no dispute here of that claim. The question here is the noise sufficiently large as to make the estimation method unreliable.

### 6.3.1 Merger Data Prior to 1962

Prior to 1962 there were 78,110 observed firm-days within 365 days of the merger date. The complementary set had 10,749,686 firm-days of observations. Using the same rule as in the bankruptcy model, holding  $\alpha$  constant to one and rounding  $\beta$  to the nearest integer to allow the use of factorials instead of the gamma function to permit computational ease, the resulting posterior distribution should be:

$$\Pr(\text{merger}) = 139(1 - p)^{138} \quad (6.5)$$

However, as with the bankruptcy data, the firms in this period are the large compared to the sample which includes non-exchange traded securities. As such, it was weighted 10% with 90% going to a uniform distribution resulting in a prior for the post-1961 period of

$$\Pr(\text{merger}) = 16(1 - p)^{15} \quad (6.6)$$

This results in an expected merger rate of 6.25%. As it turns out, this very substantially overstates the risk of merger in most circumstances. This overstatement is somewhat ameliorated by the shape of the beta distribution which favors smaller probabilities over larger probabilities. Nonetheless, it tends toward overstatement. Because of the shape of the prior, the likelihood and the prior do very substantially overlap, but the expectation given the 1962-1990 sample is overstated. The effect

is generally small due to sample size in the 1962-1990 period.

### **6.3.2 Merger Data 1962-June 30, 1990**

Mergers were either calculated conditioned only on pre-1962 counts or conditioned on prior information and accounting data. If the full set of accounting data was present then it was used with pre-1962 count data as represented in **Eq. 6.6**. If there was missing accounting information then only the count based prior was used from **Eq. 6.6**.

#### **6.3.2.1 Merger Risk January 1, 1962-June 30, 1990 With Missing Accounting Data**

The count of firm-days happening within 365 days of a merger were 396,243 where accounting data was missing. The complementary set had 9,598,954 firm-days. The posterior expected probability of a merger, given the prior and the existence of missing accounting data, was 3.964%. Given that the unit of accuracy is one tenth of one basis point the maximum likelihood estimator of the mean and the posterior expectation are indistinguishable.

#### **6.3.2.2 Financials Versus Non-Financials**

SIC codes are an industry classification system designed by the Bureau of the Census to make coding and classification simple. The CRSP data set capture SIC code data as daily data. As firms are free to change their classification at any time, CRSP updates it on a daily basis.

As the literature specifies that financials and non-financials should expect a reversed impact of leverage, the data is partitioned using SIC codes into financial and non-financial firms.

#### **6.3.2.3 Other Factors**

Three other variables were used to estimate merger risk. The factors are firm size, asset efficiency and leverage.

The book value of common equity was used as a proxy for firm size. There are other possible candidates for the estimation of firm size including gross assets, total

revenue, total net income, total pre-tax income and market capitalization. Except for market capitalization which takes share price into account, the remaining methods are accounting based methods. As such, they are subject to any distortions created by the rules of accounting. Market capitalization can be distorted by bull and bear markets and so shares the problems of accounting based measures in that systematic distortions can occur. The accounting value of common equity does have one advantage, in the end, it is the segment of the balance sheet that directly impacts shareholders in the long run. The book value of equity changes slowly and so tends to rank firms in about the same location from year to year relative to the other firms. The other measures are far less stable.

Return on assets was used as a measure of asset efficiency. As with all accounting measures distortions will happen, however, it is a proxy for the quality of firm management. A firm that fails to utilize all of its assets will have inferior returns as measured by the ratio. The measure also ignores the direct effect of leverage, though it is impacted by diminishing returns from scale. Return on net assets or return on common equity directly capture leverage effects in the measure.

Finally, the debt-to-equity ratio captures leverage directly for most types of leverage. Again, the measure is far from perfect. Firms will move assets and liabilities off their balance sheet for a variety of legitimate and illegitimate reasons.

Each of these measures were divided into Pearson-Tukey groups.

The prior distribution of the probability of merger given accounting data is:

Group	DE	ROA	TEQ	Alpha	Beta	k	n	Expected Probability of Merger
Financial	-1	-1	-1	1	15	501	10159	4.93%
Financial	-1	-1	0	1	15	0	4691	0.02%
Financial	-1	-1	1	1	15	0	0	6.25%
Financial	-1	0	-1	1	15	416	17663	2.36%
Financial	-1	0	0	1	15	954	24505	3.89%
Financial	-1	0	1	1	15	0	0	6.25%
Financial	-1	1	-1	1	15	0	2113	0.05%
Financial	-1	1	0	1	15	361	8621	4.19%

Group	DE	ROA	TEQ	Alpha	Beta	k	n	Expected Probability of Merger
Financial	-1	1	1	1	15	0	0	6.25%
Financial	0	-1	-1	1	15	0	10590	0.01%
Financial	0	-1	0	1	15	1126	49744	2.26%
Financial	0	-1	1	1	15	0	1652	0.06%
Financial	0	0	-1	1	15	92	15118	0.61%
Financial	0	0	0	1	15	18150	701014	2.59%
Financial	0	0	1	1	15	4021	137886	2.92%
Financial	0	1	-1	1	15	0	2878	0.03%
Financial	0	1	0	1	15	1463	69985	2.09%
Financial	0	1	1	1	15	91	2653	3.45%
Financial	1	-1	-1	1	15	0	7242	0.01%
Financial	1	-1	0	1	15	340	27600	1.23%
Financial	1	-1	1	1	15	0	3768	0.03%
Financial	1	0	-1	1	15	277	8746	3.17%
Financial	1	0	0	1	15	23915	967363	2.47%
Financial	1	0	1	1	15	5765	285635	2.02%
Financial	1	1	-1	1	15	91	781	11.54%
Financial	1	1	0	1	15	93	1647	5.65%
Financial	1	1	1	1	15	0	0	6.25%
Non-Financial	-1	-1	-1	1	15	5457	162033	3.37%
Non-Financial	-1	-1	0	1	15	276	83548	0.33%
Non-Financial	-1	-1	1	1	15	0	458	0.21%
Non-Financial	-1	0	-1	1	15	10916	278967	3.91%
Non-Financial	-1	0	0	1	15	10299	619024	1.66%

Group	DE	ROA	TEQ	Alpha	Beta	k	n	Expected Probability of Merger
Non-Financial	-1	0	1	1	15	0	16330	0.01%
Non-Financial	-1	1	-1	1	15	1943	72718	2.67%
Non-Financial	-1	1	0	1	15	5739	331738	1.73%
Non-Financial	-1	1	1	1	15	0	6888	0.01%
Non-Financial	0	-1	-1	1	15	3732	187174	1.99%
Non-Financial	0	-1	0	1	15	20279	975423	2.08%
Non-Financial	0	-1	1	1	15	563	25897	2.18%
Non-Financial	0	0	-1	1	15	4328	380492	1.14%
Non-Financial	0	0	0	1	15	350565	12645401	2.77%
Non-Financial	0	0	1	1	15	33135	1925713	1.72%
Non-Financial	0	1	-1	1	15	1426	102961	1.39%
Non-Financial	0	1	0	1	15	33133	1636463	2.02%
Non-Financial	0	1	1	1	15	2170	141263	1.54%
Non-Financial	1	-1	-1	1	15	1368	63585	2.15%

Group	DE	ROA	TEQ	Alpha	Beta	k	n	Expected Probability of Merger
Non-Financial	1	-1	0	1	15	1810	74778	2.42%
Non-Financial	1	-1	1	1	15	0	0	6.25%
Non-Financial	1	0	-1	1	15	1951	72803	2.68%
Non-Financial	1	0	0	1	15	15202	415382	3.66%
Non-Financial	1	0	1	1	15	661	14361	4.60%
Non-Financial	1	1	-1	1	15	0	10242	0.01%
Non-Financial	1	1	0	1	15	410	21803	1.88%
Non-Financial	1	1	1	1	15	0	823	0.12%

Table 6.5: Prior Probability of Merger Risk

## 6.4 Prior Distribution of Rewards of Investing Given A Merger Will Happen

Given a merger will happen, a mixture distribution should be present. This mixture should be different depending upon whether it was a merger for cash or for equity. Mergers with a mix of cash and equity were treated as stock-for-stock mergers. The logic being that the cash could be used to buy equity or could be treated as a dividend.

As the distribution includes a skewed stable distribution without variance, the author has two choices. The first is to describe the density function in terms of series. This would be a computational nightmare in SAS, particularly given that

the parameter values would have to vary.

The second solution takes advantage of the nature of the put contract. Its expectation from zero to the strike price is approximated by the sample mean over the same period. However, lacking an appropriate theorem, this useful observation isn't useful in Bayesian methodologies. Bayesian methodologies use estimates of parameters and the likelihood function is unknown. Nonetheless, a solution does present itself through discrete approximations of the integral.

By cutting the range into a histogram of small enough partitions, it is possible to estimate the densities over the range of potential loss. Two potential mechanisms exist to estimate the loss given both a merger happened and a loss happened. The solutions are either to use a Dirichlet prior with a multinomial likelihood with each histogram block being one partition of the multinomial distribution, or a Normal-Wishart approximation.

The data used was CRSP data from December 31, 1925- June 30, 1990. A choice then exists on the prior for this set of data. Use of a uniform distribution does not make a lot of sense as it is clear that large losses should be rare and that losses following announcement of the merger should also be rare. Further, the Landau distribution would favor positive returns. As mergers should result from purchases of undervalued assets returns should at least be slightly positive and the variability is clearly unknown.

The Cauchy distribution where  $\mu = 1.05$  and  $\sigma = 1$  multiplied by  $2/3^{\text{rds}}$  to reduce the weight of losses and truncated at zero was used to create the prior normal mean of each partition. The standard deviation, to allow for simple updating despite truncation, was taken as centered on  $\mu/3.3$  to place enough distance from boundary conditions to minimize the distortion created in Bayesian updating.

One hundred rectangular partitions were created in the reward space from zero to one. Although a polygon would slightly improve the estimate the partitions are quite small and the effect is believed to be well below the unit of precision required.

The Multivariate Normal-Wishart distribution is a compound distribution of the multivariate normal distribution and the Wishart distribution, which is a generalization of the Gamma distribution. The joint distribution is a four parameter distribution.

Within the framework of the rules of probability, the joint prior probability of

the mean and the precision  $\boldsymbol{\mu}$  and where  $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$  is:

$$\Pr(\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) = \Pr(\boldsymbol{\mu}|\boldsymbol{\Lambda}^{-1}) \Pr(\boldsymbol{\Lambda}) \quad (6.7)$$

The distribution of possible values of the mean, given a precision matrix, is treated as the normal distribution, while the Wishart distribution is the distribution of possible precision matrices, the inverse of the covariance matrix. The two remaining parameters can be thought of as a weight for the relative number of observations that the prior is treated as representing and the number of dimensions minus one. The prior was given a weight equivalent to one observation while the number of dimensions is 100 so the degrees of freedom are 99.

The prior for the probability an observation will be in a particular partition was set with a mean of:

$$\mu_i = \frac{2 \tan^{-1} \left( \frac{i}{100} - 1.05 \right) - \tan^{-1} \left( \frac{i-1}{100} - 1.05 \right)}{\frac{\pi}{2} + \tan^{-1} (1.05)}, \quad (6.8)$$

a standard deviation of

$$\sigma_{ii} = \frac{\mu_i}{3.3}, \quad (6.9)$$

and a covariance of

$$\sigma_{ij} = 0, \forall i \neq j \quad (6.10)$$

The covariance has no distribution as it is conditioned to be zero by construction.

The observed frequencies were:

Return	Prior Mean	Count of Observations	Posterior Mean
0.005	0.20%	708	0.14%
0.015	0.20%	348	0.07%
0.025	0.20%	488	0.10%
0.035	0.21%	189	0.04%
0.045	0.21%	507	0.10%
0.055	0.21%	810	0.16%
0.065	0.21%	743	0.15%
0.075	0.22%	545	0.11%
0.085	0.22%	681	0.13%
0.095	0.22%	352	0.07%
0.105	0.22%	402	0.08%
0.115	0.22%	397	0.08%
0.125	0.23%	399	0.08%
0.135	0.23%	209	0.04%
0.145	0.23%	427	0.08%
0.155	0.23%	531	0.11%
0.165	0.24%	426	0.08%
0.175	0.24%	354	0.07%
0.185	0.24%	395	0.08%
0.195	0.24%	351	0.07%
0.205	0.25%	399	0.08%
0.215	0.25%	325	0.06%
0.225	0.25%	386	0.08%
0.235	0.25%	312	0.06%
0.245	0.25%	502	0.10%
0.255	0.26%	448	0.09%
0.265	0.26%	350	0.07%
0.275	0.26%	427	0.08%
0.285	0.26%	516	0.10%
0.295	0.27%	503	0.10%

Return	Prior Mean	Count of Observations	Posterior Mean
0.305	0.27%	359	0.07%
0.315	0.27%	365	0.07%
0.325	0.28%	369	0.07%
0.335	0.28%	553	0.11%
0.345	0.28%	501	0.10%
0.355	0.28%	421	0.08%
0.365	0.29%	491	0.10%
0.375	0.29%	602	0.12%
0.385	0.29%	484	0.10%
0.395	0.29%	676	0.13%
0.405	0.30%	462	0.09%
0.415	0.30%	513	0.10%
0.425	0.30%	600	0.12%
0.435	0.30%	544	0.11%
0.445	0.31%	573	0.11%
0.455	0.31%	506	0.10%
0.465	0.31%	560	0.11%
0.475	0.32%	485	0.10%
0.485	0.32%	542	0.11%
0.495	0.32%	1107	0.22%
0.505	0.32%	504	0.10%
0.515	0.33%	576	0.11%
0.525	0.33%	547	0.11%
0.535	0.33%	722	0.14%
0.545	0.33%	685	0.14%
0.555	0.34%	652	0.13%
0.565	0.34%	547	0.11%
0.575	0.34%	721	0.14%
0.585	0.35%	656	0.13%
0.595	0.35%	691	0.14%
0.605	0.35%	770	0.15%

Return	Prior Mean	Count of Observations	Posterior Mean
0.615	0.35%	840	0.17%
0.625	0.36%	875	0.17%
0.635	0.36%	928	0.18%
0.645	0.36%	962	0.19%
0.655	0.36%	892	0.18%
0.665	0.37%	1172	0.23%
0.675	0.37%	914	0.18%
0.685	0.37%	1012	0.20%
0.695	0.37%	1079	0.21%
0.705	0.38%	1062	0.21%
0.715	0.38%	1297	0.26%
0.725	0.38%	1191	0.24%
0.735	0.38%	1160	0.23%
0.745	0.38%	1435	0.28%
0.755	0.39%	1355	0.27%
0.765	0.39%	1270	0.25%
0.775	0.39%	1385	0.27%
0.785	0.39%	1535	0.30%
0.795	0.39%	1676	0.33%
0.805	0.40%	1455	0.29%
0.815	0.40%	1562	0.31%
0.825	0.40%	1842	0.36%
0.835	0.40%	1950	0.39%
0.845	0.40%	1673	0.33%
0.855	0.40%	1836	0.36%
0.865	0.41%	1800	0.36%
0.875	0.41%	1798	0.36%
0.885	0.41%	1986	0.39%
0.895	0.41%	1913	0.38%
0.905	0.41%	1992	0.39%
0.915	0.41%	2157	0.43%

Return	Prior Mean	Count of Observations	Posterior Mean
0.925	0.41%	2312	0.46%
0.935	0.41%	2688	0.53%
0.945	0.42%	2595	0.51%
0.955	0.42%	2650	0.52%
0.965	0.42%	2778	0.55%
0.975	0.42%	3053	0.60%
0.985	0.42%	3473	0.69%
0.995	0.42%	6129	1.21%

Table 6.6: Table of Frequencies of Returns Given a Merger

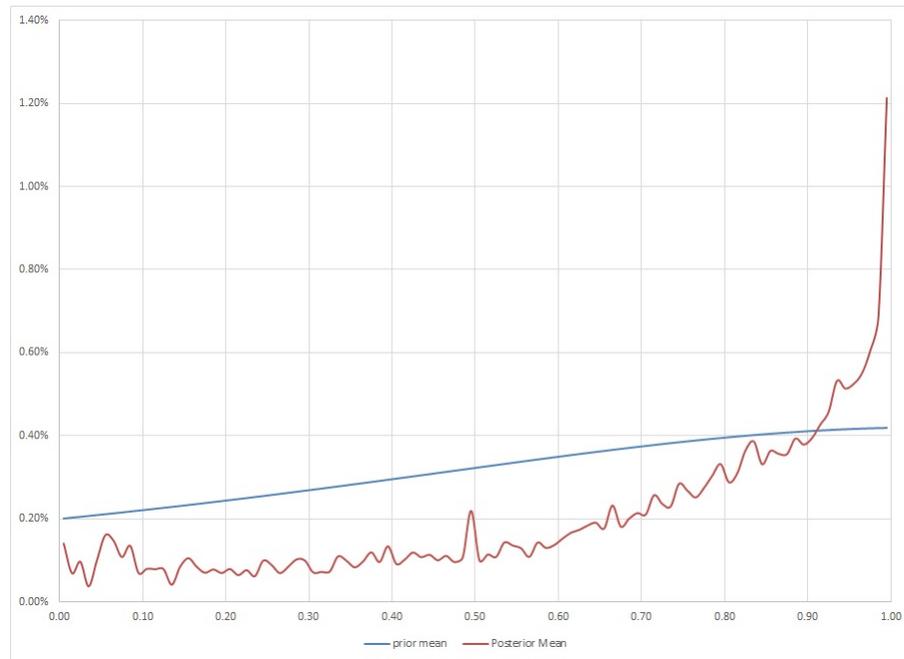
Generally, the prior overestimated the volume of loss, however as the number of observations in each category is quite large, and the prior had the effect of only being one observation in each category, the posterior and the maximum likelihood estimate are identical out to the accuracy level desired. The difference between a uniform prior and the Cauchy prior ended up being of the order of  $10^{-7}$  or less.

Because  $\sigma_{ij} = 0$  updating the posterior mean collapsed to the simple formula:

$$\mu'_i = \frac{s_i^2}{n\sigma_i^2 + s_i^2}\mu_i + \frac{\sigma_i^2}{n\sigma_i^2 + s_i^2}\bar{x}_i, \quad (6.11)$$

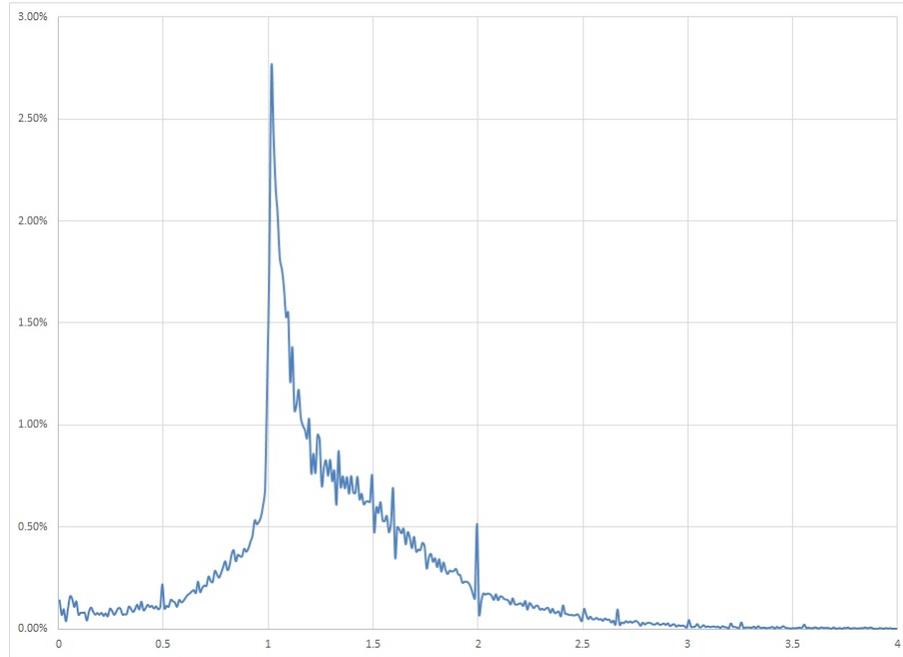
where  $s_i$  is the sample standard deviation for the partition and  $\bar{x}_i$  is the sample mean for the partition.

Visually the posterior means of the merger data, with its Cauchy prior means from zero to one graphically appear as:



**Figure 6.2.** Prior Distribution of Returns Given a Merger

However if the prior were extended to out-of-the-money puts to include puts at three hundred percent above the current price, then quite a different image appears.



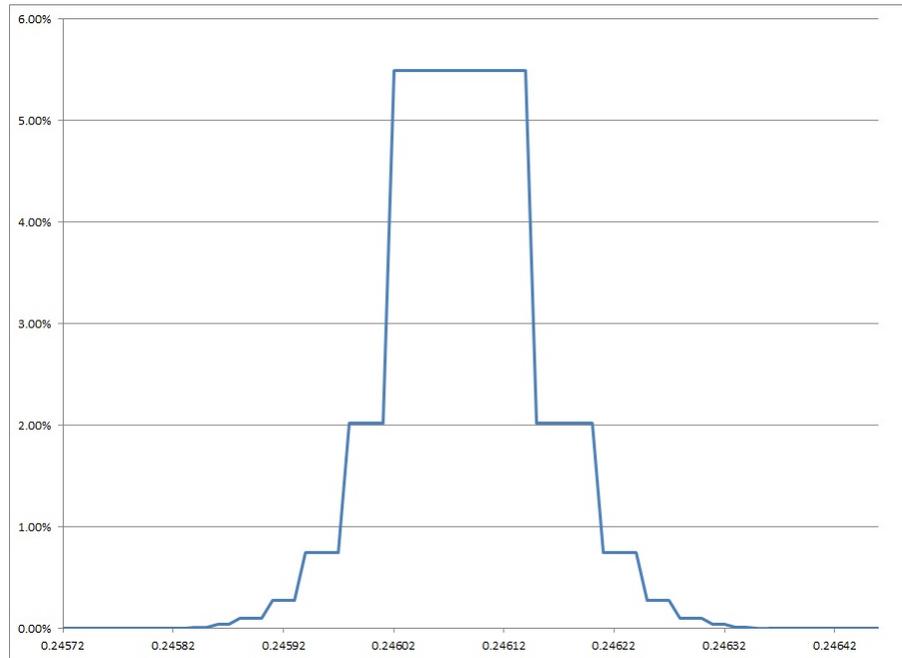
**Figure 6.3.** Prior Distribution for Mergers from 0 to 4

Its this image that makes the fact that it is a mixture distribution quite clear.

## 6.5 Prior Distribution of Rewards from December 31, 1925-June 30, 1990 for Going Concerns

Because it is so clear from 1926 forward that the mode is one, the likelihood function is conditioned on  $\mu = 1$ . This leaves open only the question of the value of  $\sigma$ . A search algorithm was created to place fine screen over the densest region. The mode was found to  $\frac{1}{10}$ <sup>th</sup> of one basis point of  $\sigma$ . The 99% credible interval is (0.24590,0.24627) The maximum a posteriori estimator is tied. The region of highest density is (0.24602,0.24614).

Unfortunately, the standard method of updating the process over a parameter



**Figure 6.4.** Prior Distribution of  $\sigma$  for Going Concerns

space was not computationally feasible using daily Bayesian updating. Daily updating requires treating each day from July 1, 1990 to December 31, 2007 and treat it as a sample. Bayes theorem is then applied daily. The uncertainty would be marginalized out on a predictive distribution and an expected loss would form. To accomplish this goal, quite a number of computational methods were tried. Ultimately the same method as used for the unknown merger distribution was used as it was computationally feasible without needing to substantially reduce the sample size.

## 6.6 Testing the Properties of the Bankruptcy Estimators

### 6.6.1 Problem Background

The bankruptcy estimators provide a predicted rate of bankruptcy for each date in the sample period. As bankruptcy in **definition 2** implies a binary state. If  $q_t^i$  is redefined as  $k \times (1 - \{0 \text{ or } 1\})$  where one is one bankruptcy and zero is a going

concern, then a simple binomial model appears.

Creating a binary variable  $x_t^i$  which refers to the state of firm  $i$  on day  $t$ , then the expectation of  $x_t^i$  from the data available at time  $t - 1$  can be denoted as  $E_{t-1}(x_t^i)$ . It is also useful to note that, due to model construction,  $E(x_t^i) \equiv E(x_t^j)$  for many values of  $i$  and  $j$  allowing joint updating. Further, the gap between reality and the expectation can be denoted by  $g_t^i$  where:

$$g_t^i = x_t^i - E_{t-1}(x_t^i) \quad (6.12)$$

The next Bayesian question is to determine the likelihood function for  $g_t^i$ . Noting the concern here is for the forward value of the expectation against individual observations, it quickly becomes obvious that the solution can be reduced to the difference between a sample mean and a prior expectation. Given finite variance, the sampling distribution of the means and the expectations will be normal. Noting that the difference of two normals is a normal distribution itself, it is possible to posit a likelihood for the difference as being generated from a normal distribution.[113] A slightly more tenuous assumption was made that the bankruptcy models' properties were constant over time, yielding:

$$g_t^i \sim \mathcal{N}(\delta, \omega), \forall i, t. \quad (6.13)$$

Although any possible prior that either represents real knowledge, beliefs of supporters of the proposed model, or the beliefs of adversaries could be used, there is a computational advantage to express those beliefs using a conjugate prior distribution.

One of the possible conjugate prior distributions for the normal mean and variance is the normal-inverse chi-squared distribution mixture distribution. As will be shown, it is a more than reasonable representation of prior beliefs and so the computational efficiencies outweigh any attempt to set a numerical and arbitrary prior onto the model forcing some other mechanism such as acceptance-rejection testing to determine the constant of integration and the marginal posterior testing of hypothesis. The formulas here are provided by Lee.[55] For a more complete exposition on the analysis of conjugate prior distributions for the normal distribution, see Murphy.[67]

The question of the prior distribution would have the potential to be contentious, except that the data set is so large that it will swamp the prior to the required degree of accuracy. Nonetheless, an appropriate prior is called for. Although the model expectation is a biased expectation, there isn't a method to determine if that bias is greater than or less than zero. The other question is how much weight should the prior distribution have on the posterior distribution.

While it is reasonable to use prior variance estimates from the pre-1990 sample data, another mechanism was chosen. Treating the prior as carrying only one observation's strength and choosing a conjugate prior, the question becomes what is the most extreme possible value for the variance, prior to seeing any prior data or sample data. The answer is one. The largest possible variance on a Bernoulli trial based on a prediction would be one, ignoring Bessel's correction to prevent the value from exploding.

The use of the largest possible variance weakens any centering effects placing a center of zero, which is a very reasonable center. It also increases the proposed risk of the model triggering an increase in the probability of falsification to the extent the variance is of concern; though, due to the size of the data set, the impact is quite small.

The model requires two prior parameters and two hyperparameters. A hyperparameter is a parameter that does not appear in the likelihood function, but which does impact the calculation of the posterior.

The prior parameter estimates, for a normal-inverse chi-squared prior distribution, are:

$$\delta_0 = 0, \omega_0 = 1. \quad (6.14)$$

As the prior reflects very little actual knowledge, it is reasonable to give it a sample size of one. This implies one degree of freedom for the prior estimator of the mean and zero degrees of freedom for the prior estimate of the variance.

The hyperparameters are:

$$n_0 = 1, \nu_0 = 0 \quad (6.15)$$

### 6.6.2 Uninformative Alternatives

Given a reasonable center of location for the difference estimator, this is the most unreasonable possible choice of parameters that also bears some semblance of realism. The two other conceptually leading choices would be a reference prior where,

$$\Pr(\delta, \omega) \propto \omega^{-1}, \quad (6.16)$$

or a flat prior, which would technically be a proper prior due to boundary conditions, of

$$\delta = 0, \omega = \lim_{k \rightarrow \infty} k, n = 0. \quad (6.17)$$

Although the reference prior maximizes the value of the evidence and can, in a sense, be thought of as being the view of someone with no strong prior beliefs, it permits prior beliefs that somewhat favor the proposed model. Because the values are bounded, due to the binary nature, the reference prior would favor realistic to smaller than realistic variances. Due to the low probability of bankruptcy and the truncation, the reference prior could be viewed as favorable to the proposed model. As a consequence, the reference prior was not chosen.

The latter choice was excluded as it does ignore real information. It is also an improper prior, or at least the limiting form of proper prior distributions, were boundary conditions not present. Improper prior distributions are avoided due to dimensionality reasons.[95]

### 6.6.3 Posterior Tests

To simplify notation and computation,  $m$  will be used to denote the number of bankruptcies observed in the sample;  $n$  will be used to denote the number of firms surviving as going concerns; and as many firms share the same expectation  $E_m$  will denote shared expectations for a particular part of a model.

If  $x_t^i = 1$  when a firm will be bankrupt within one year and  $x_t^i = 0$  when firm survives for at least one year, for a given day and shared expectation, then the sample mean of the differences is:

$$\bar{g} = \frac{\sum_{i \in M} (x_t^i - E_M(x_t^i))}{m + n} \quad (6.18)$$

This simplifies to:

$$\bar{g} = \frac{m}{m+n} - E_M(x_t^i) \quad (6.19)$$

The sample variance is:

$$s^2 = \frac{\sum_{i=1}^{m+n} (x_t^i - E_M(x_t^i) - \bar{g})^2}{m+n-1} \quad (6.20)$$

This simplifies to:

$$s^2 = \frac{\sum_{i=1}^{m+n} (x_t^i - \frac{m}{m+n})^2}{m+n-1} \quad (6.21)$$

This further simplifies to:

$$s^2 = \frac{mn}{(m+n)(m+n-1)} \quad (6.22)$$

The posterior then needs to answer two distinct questions. The first question is whether or not the estimator will trigger systematic losses to the market maker due to being a systematic underestimate. The second question is whether or not the estimator is too risky to use. The first is a question of the posterior distribution of the mean, the second of the posterior distribution of the variance.

There are nearly 100,000 combinations of days and model subsets. That is a non-trivial reporting issue, unless a simpler solution exists, which it does. It will be shown that each separate test can be combined in a disciplined manner, permitting a joint test even though the projected expectation changes with information.

As it is easier to define the MAP estimates of the parameters than to reproduce the proof, the posterior parameters and hyperparameters are:

$$n_1 = n_0 + m + n \quad (6.23)$$

$$\nu_1 = \nu_0 + m + n \quad (6.24)$$

$$\delta_1 = \frac{n_0 \delta_0 + (m+n) \bar{g}}{n_1} \quad (6.25)$$

The posterior sum of the squares, denoted  $S_1$  for the posterior,  $S$  for the sample, and  $S_0 = \nu_0 \sigma_0^2$ , is

$$S_1 = S_0 + S + (n_0^{-1} + (m + n)^{-1})^{-1}(\delta_0 - \bar{g})^2 \quad (6.26)$$

The posterior MAP estimate of the variance is:

$$\omega_1^2 = S_1/\nu_1 \quad (6.27)$$

Substituting for the chosen prior values:

$$n_1 = 1 + m + n \quad (6.28)$$

$$\nu_1 = m + n \quad (6.29)$$

$$\delta_1 = \frac{m - (m + n)E_M(x_t^i)}{m + n + 1} \quad (6.30)$$

$$S_1 = \frac{mn}{m + n} + \frac{m + n}{m + n + 1} \left( \frac{m}{m + n} - E_M(x_t^i) \right)^2 \quad (6.31)$$

$$\omega_1^2 = \left( \frac{\frac{mn}{m+n} + \frac{m+n}{m+n+1} \left( \frac{m}{m+n} - E_M(x_t^i) \right)^2}{m + n} \right) \quad (6.32)$$

$$\omega_1^2 = \frac{[E_M(x_t^i)]^2 n + m[(E_M(x_t^i) - 1)^2 + n]}{(m + n)(m + n + 1)} \quad (6.33)$$

The posterior distribution, from Lee, is:[55]

$$\Pr(\delta, \omega | \mathbf{x}) = \omega^{-(\nu_1+1)/2-1} \exp \left[ -\frac{n_1 \delta^2 - 2n_1 \delta \delta_1 + n_1 \delta_1^2 + S_1}{2\omega} \right] \quad (6.34)$$

Then, from marginalization, it follows that:

$$\Pr(\delta | \mathbf{x}) = \int_0^\infty \Pr(\delta, \omega) d\omega \quad (6.35)$$

Murphy provides a simple derivation.[67] The marginal distribution of the mean is Student's t-distribution, so that:

$$\Pr(\delta | \bar{g}) \propto t_{\nu_1}(\delta | \delta_1, \omega_1^2/n_1) \quad (6.36)$$

Because  $\nu_1$  is so large, it is possible to substitute z-scores for raw values. The final z-score is simply the weighted average z-score of the entire body of tests. This reduces a large scale, multi-parameter problem down to one parameter. This permits the option writer to evaluate the model with the simple hypothesis:

**Hypothesis (1a).**  $z \leq 0$

**Hypothesis (1b).**  $z > 0$

While the market maker's only concern is that there is no systematic loss, the buyer of the option has the concern that the mark up is not too large. However, as the risk premiums are unknown, this can only be stated as requiring the model difference to be inside the interval  $(-c,0)$ , where  $c$  is a constant.

The second set of hypothesis is whether or not the variance is small enough. Small enough, of course, depends upon the degree of leverage. A firm leverage 1:1 can take larger losses without concern than a firm leveraged 25:1.

A one percent standard deviation would be large for highly leveraged firms, so the writer would find the model too risky if the posterior variance were greater than one basis point, or 1/10,0000.

**Hypothesis (2a).**  $\omega^2 \leq .0001$

**Hypothesis (2b).**  $\omega^2 > .0001$

The formula for the distribution of possible posterior variances is from Lee, but see Murphy for a complete proof.[55, 67] The marginal distribution of the variance is:

$$\Pr(\omega^2|\mathbf{x}) = S_1\chi^{-2}(\nu_1) \quad (6.37)$$

Because the degrees of freedom are so large, it is better to work with the precision rather than the variance for computational purposes. For test purposes, the test is based on the precision, the multiplicative inverse of variance, rather than the variance itself.

$$\Pr(S_1/\omega^2|\mathbf{x}) \sim \chi^2(\nu_1) \quad (6.38)$$

There is a question of how much risk could be absorbed by the writer. The cut-off point chosen for the hypothesis was purely arbitrary. A better solution

would be to make it less than some value chosen on a firm by firm basis, using the cost function and leverage of the firm as the mechanism to set a value.

### 6.6.3.1 Results of Posterior Tests for Bankruptcy Estimator

The maximum a posteriori estimator for the mean difference between the model and the actual observed outcomes was -0.007997. The maximum a posteriori estimator for the variance of the differences was 0.000125271. There were 34,366,629 observations and 34,366,628 degrees of freedom for the test of the precision. The sum of the squares was 4,305.14.

Given the very large sample size, the z-score for the hypothesis test is -4181.25. The posterior probability of the writer not taking a loss is sufficiently close to unity that it cannot be calculated. The probability is sufficient to accept hypothesis 1a, that the model is adequate from the writer's perspective. It will require research into risk premiums to determine the adequacy from the view of the buyer.

The 99.99% Bayesian credible interval for the average difference between the model and reality is (-.008004,-.007989).

Likewise, given the very large sample size, the  $\chi^2$  test is also conclusive and rejects hypothesis 2a. Given a required standard deviation of 0.01 or less, the model is rejected. As this was chosen arbitrarily as a cut-off, the 99.99% credible interval is provided as well.

As the  $\chi^2$  distribution goes to the normal distribution as the sample size goes to infinity, the normal approximation was used. The z-score was -1062.03 on the precision. Using precision reverses the sign of the hypothesis and soundly rejects 2a.

The 99.99% highest density region credible interval for  $\omega$  is (0.0112065,0.0112176) with a maximum a posteriori estimate of 0.011212. The standard deviation is approximately 12% larger than the hypothesized level. Whether or not this is a good or bad level depends upon the alternatives and the capital necessary.

### 6.6.3.2 Results of Posterior Tests for the Merger Estimator

The maximum a posteriori estimator for the mean difference between the model and the actual observed outcomes was 0.018530. The maximum a posteriori es-

estimator for the variance of the differences was .002706. There were 42,337,719 observations and 42,337,719 degrees of freedom for the test of the precision. The sum of the squares was 2,202,565.09.

Given the very large sample size, the z-score for the hypothesis test is 2,318.65. The posterior probability of the writer taking a loss is sufficiently close to unity that it cannot be calculated. The probability is sufficient to reject hypothesis 1a, that the model is adequate from the writer's perspective.

The 99.99% Bayesian credible interval for the average difference between the model and reality is (0.01839,0.01867).

The 99.99% highest density region credible interval for  $\omega$  is (0.0507,0.0533) with a maximum a posteriori estimate of 0.052. The standard deviation is very large and implies the need for research into this field. The z-score was 325.83 on the precision.

There were more observations in the merger set than in the bankruptcy set due to the use of different databases, along with effect of more rules to reject an observation as suspect in the bankruptcy case due to the differing nature of the data. This had the effect of removing more observations from the bankruptcy data than the merger data.

### 6.6.3.3 Results of Posterior Tests for Expected Loss Given a Merger

The maximum a posteriori estimator for the mean difference between the model and the actual observed outcomes was .008341105. The maximum a posteriori estimator for the variance of the differences was .030224. There were 180,681 observations and 180,681 degrees of freedom for the test of the precision. The sum of the squares was 5,460.67.

The number of observations was relatively small due to the absence of offer data. Offer prices were reconstructed by taking the first trade after merger and dividing it by the last trade before the merger, with the proviso that a trade occurred within the last week of existence and the first week post merger. As many firms being merged out of existence have limited trading, many trades were excluded by the inability to estimate the terms of exchange. The presence of a deal term database or more liberal terms for estimation would have substantially increased the number of observations.

The z-score for the hypothesis test is 20.3944. The posterior probability of the writer taking a loss is sufficiently close to unity that it can be treated as zero. The probability is sufficient to reject hypothesis 1a, that the model is adequate from the writer's perspective. Even with a time drift for  $\sigma$ , the model is inadequate.

The 99.99% Bayesian credible interval for the average difference between the model and reality is (0.007288,0.009395).

Given the high natural variability in the data, a second scale parameter based hypothesis was not used. Rather, the distribution of the scale is provided instead. The 99.99% highest density region credible interval for  $\omega$  is (.026978,.03347).

#### 6.6.3.4 Results of Posterior Tests of Expected Loss Given Survival

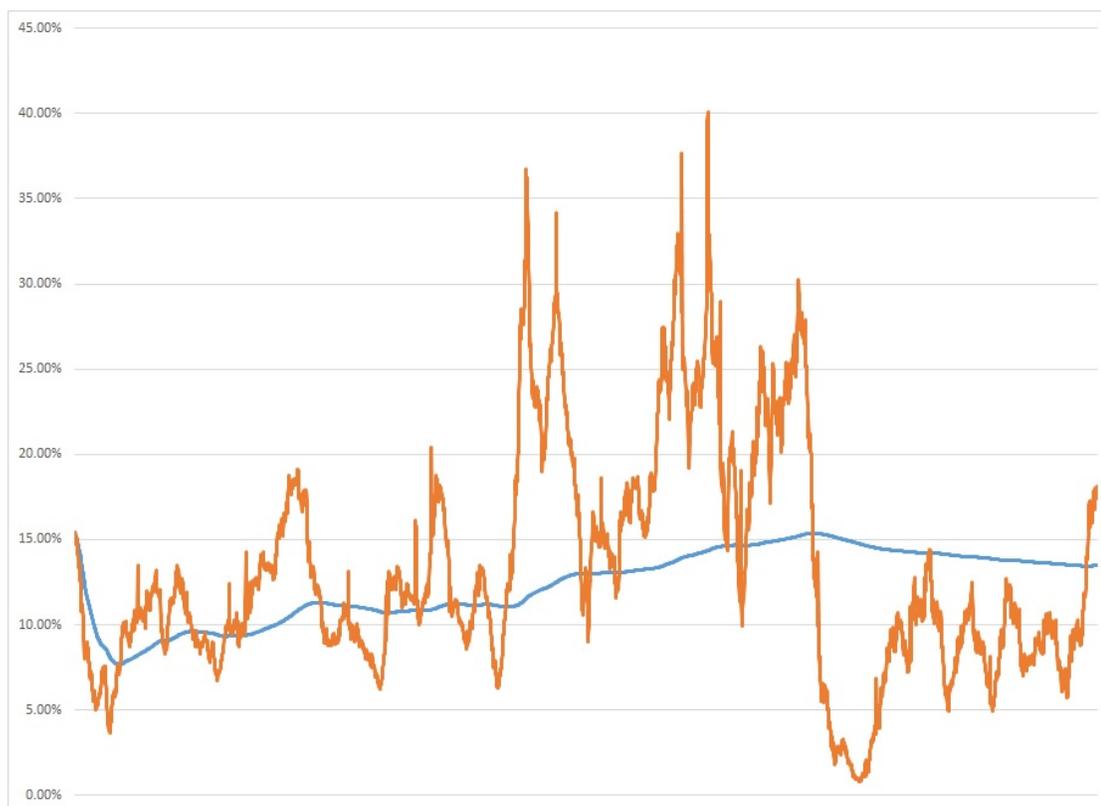
The maximum a posteriori estimator for the mean difference between the model and the actual observed outcomes was .015058. The maximum a posteriori estimator for the variance of the differences was 0.0459931. There were 20,421,083 observations and 20,421,083 degrees of freedom for the test of the precision. To be included as an observation, a security had to trade between 352 days and 365 days from the date of the original trade. The sum of the squares was 939,236.13.

The z-score for the hypothesis test is 317.301. The posterior probability of the writer not taking a loss is sufficiently close to unity that it cannot be calculated. The probability is sufficient to reject hypothesis 1a, that the model is adequate from the writer's perspective. Ignoring the posterior tests for a moment, a four percent drift in  $\sigma$  would be enough to trigger the model error alone. If the drift is accounted for, this likely an excellent mechanism to estimate option loss premiums.

The 99.99% Bayesian credible interval for the average difference between the model and reality is (0.01487,.01524). As with returns given a merger, a second hypothesis test was not performed. As above, the credible interval was provided. The 99.99% highest density region credible interval for  $\omega$  is (0.04599,0.046) with a maximum a posteriori estimate of .0459931.

## 6.7 Conclusion

Except for the bankruptcy estimator, results were disappointing. In the case of merger analysis, there is too little theory present to form a good model. A merger



**Figure 6.5.** Average Daily Loss, Given a Loss Did Occur Versus Daily Prediction of Loss Given a Flat (Neutral) Prior

is a state and there are many paths to the state. Like having a fever, there are many paths to the body creating the physiological process known as fever. Current explanations revolve around the principal-agent problem, but there are certainly other sources of merger risk. In the case of merged returns, a theoretical case needs to be made, in a strong form, for the form of the distribution. As to the return for going returns, existing methods should work given an increase in resource availability.

If the hypothesis were changed so that the standard deviation was less than or equal to the standard deviation used in the Fama-French Three Factor Model, or their broader four or five factor model, then the best model has a 2.01 percent standard deviation. Using the broader measure, the bankruptcy and merger risk estimator are better, although they are in a different subject field. Actual costing of risk should be performed to determine the subjective quality of the various models for different actors.

## Part VI

# Conclusions and Directions Forward

## CONCLUSION

This work is a start towards option pricing. Not touched were American or Asian style option contracts. Untouched was work on commodities and various other contracts on underlying assets. Significant empirical work needs performed on estimating merger risks. A better understanding of the determinants of bankruptcy risk are needed. There is no work here on dividend estimation. Returns on equity investments were not conditioned on interest rates. The risk premium wasn't estimated.

This work is incomplete. On the other hand, general principles have been presented that will allow competing models to be tested. The author lacked access to trade data for option contracts. It should be determined if the mark-up/mark-down on option contracts converges to liquidity costs of the underlying so that actors are indifferent between direct ownership and the purchase of derivatives.

Fundamentally, this is a model of capital. As such it touches macroeconomic models. These models have been separated since mean-variance finance came to the fore. Logically, the value of physical capital must equal the financial capital, in equilibrium. The principles of financial capital and physical capital cannot diverge.

It would be advised to return financial economics to macro- and micro-economics.

## 7.1 Mathematical Observations

### 7.1.1 Model principles

It is time to move forward. In terms of broad principles, certain things should be present in models that have not been present in the past. The market maker, if one exists, should operate under some form of a coherence principle. Liquidity costs assure coherence and Pareto optimality. Dividends need to exist in models since they exist in reality. Bankruptcy needs to be in the model for option pricing. The likelihood function should be driven by the model and not convenience.

On the other hand, there isn't a unique axiomization of Bayesian decision theory. The axioms chosen should fit the needs of the problem. Axioms have consequences. It is quite possible a more complete solution could be arrived at using a different construction.

Under de Finetti, prices and probabilities are strictly interchangeable. Probabilities and not just preferences are revealed by prices in this system. As such, separations of probabilities and preferences are a problem. Conversely, Savage's or Ramsey's method separates out pure probabilities from preferences, but this is not a free ride either as it requires knowledge of internal states.[68]

Additional research needs performed on the distribution under mergers and on the impact of boundary conditions on the likelihood function. Research also needs performed on some version of continuous time methods.

A way forward may be through cost functions. If an expectation is turned around, so that it isn't a tool but rather a solution to a cost function problem, then it may be possible to restructure anticipation in terms of cost functions.

If the expected value were thought of as being the solution to the anticipated value, then it should be possible to have a variety of solutions that include the median or the mode as a solution.

Continuous time finance uses expectations for a variety of things, but depends upon it for differentiation. If the mode or the median are sufficiently smooth as they move through time, then the differentiation no longer depends upon an expectation. A general anticipation operator should then choose appropriately among the mean, median or mode as the appropriate anticipated solution.

## 7.2 Quarternions

Another opportunity to improve predictions may be using Quarternions. Pioneering work regarding the use of the complex plane to model planar Brownian motion, which projects onto the real numbers as the Cauchy distribution began with work by Krzysztof Burdzy.[19] It may be possible to extend this work through the use of quarternions. Returns can be thought of as:

$$R_t = R_t(p_t, p_T) \quad (7.1)$$

Returns, as described here, are the function of two types of data, fast data and slow data.

Liquidity costs are the fast data. Accounting and macroeconomic data are slowly changing data. If liquidity costs are a function of a variable  $\lambda$  and accounting data is represented as  $\iota$ , then returns could be described as:

$$R_t = R_t(p_t(\lambda_t, \iota_t), p_T(\lambda_T, \iota_T)) \quad (7.2)$$

It is common in mathematics and statistics to model the Cauchy distribution as a projection from planar Brownian motion in the complex plane. Adding two dimensions to each price results in a four dimensional error. The natural model for this is probably the quarternion. Quarternions arise as the quotient of two vectors. As a ratio function of two dimensional errors, quarternions may provide superior computational options.

## 7.3 Observations from the Data

### 7.3.1 $\mu = 0\%$

The most striking component of the data is that the center of location for going concerns is very clearly 0% plus dividends across the entire data set. The interpretation is that risks are paid for through the dividend process. Prices adjust when dividends are inadequate. This, of course, raises the question of what the predictive dividend for a security is. Indeed, it also implies prices fail to fully impound

information, but dividend policy joint with prices may. This is an open question.

There are two mechanisms to think about the dividend. The first is to note that a present value, a future value and an annuity payment are equivalents. As such, the absence of a current cash payment by a firm reinvesting its profits could be thought of in terms of a current cash payment. The second method is to calculate the dividend given information. In that scenario there is never a zero dividend level.

Which works better is an empirical question.

Still, there are consequences to this construction. The first is that the marginal actor demands payment in cash for risks and that these funds are no longer available to management. Both bankruptcy risk and inflation risk are compensated through dividend payments. While the marginal actor receives payment in cash, the average actor receives cash plus growth. The average actor is paid the median return, which is 2%. The median return is the growth rate of capital and is a result of the fact that the median and the mode are different.

The median return differs from the modal return due to the existence of bankruptcy law. The limitation of liability truncates the distribution. As the mean of the log return will be the median raw return, the growth rate is entirely due to the limitation of liability.

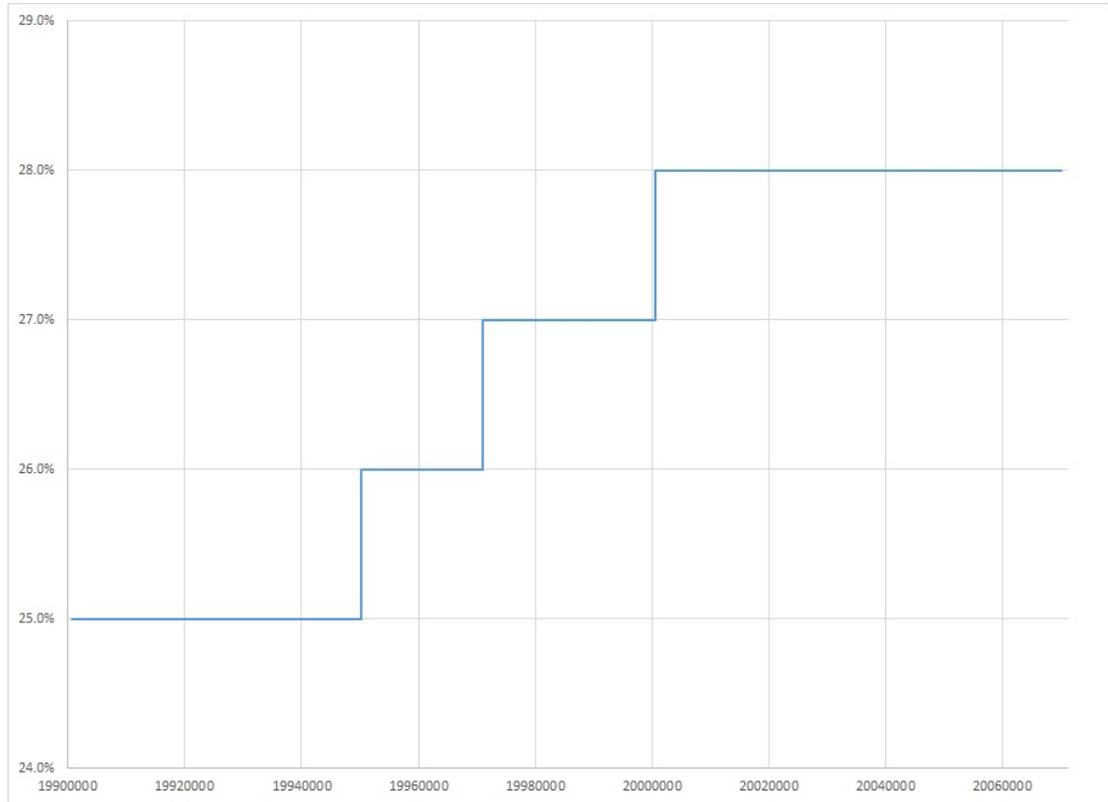
The implication is that strong bankruptcy protection triggers higher economic growth because adequate funds must be set aside as reserves. It is these reserves that are the source of growth.

If the shape parameter is not fixed, then things that drive the shape parameter drive the national growth rate.

### 7.3.2 $\sigma$ May Be Increasing

The value of  $\sigma$  at the beginning of the period is not the value at the end. Because of operational limitations to estimate movement of the posterior a wide mesh was placed over returns as a function of time and the maximum a posteriori estimator was plotted. The finest mesh that appeared to be operationally possible was 100 basis points. While quite crude, provide a picture of a time drift in  $\sigma$  because it was only measured over the sample period it isn't possible to determine when the

drift began, or how consistent it is.



**Figure 7.1.** Drift of Sigma by Date

The difficulty of this graphic is that if it were a linear process, that is:

$$\Pr(x|\mu; \beta_1; \beta_0; \tau) = \frac{\beta_1(\tau - t_0) + \beta_0}{(\beta_1(\tau - t_0) + \beta_0)^2 + (x - \mu)^2}, \quad (7.3)$$

then risk is drifting for some unknown reason. This is actually quite a bit of drift over the interval. There is no reason to believe risk is a function of time, so some investigation needs performed. Unfortunately, the sample period also corresponds to the period known as the Great Moderation. Inadequate information exists for the reason for drift, given the information base. It may simply be an artifact of the data measurement process. Nothing is clear about this, other than that it appears to have happened.

## 7.4 Policy Implications

In the world of blackboard economics, the limiting distribution of returns for going returns is the Cauchy distribution. Nonetheless, mean-variance finance models are the normative models of economics. They are taught at all levels, they are on doctoral comprehensive exams, they are used in industry, and they underlie regulatory models explicitly or implicitly. This paper requires the abandonment of mean-variance finance.

Generally Accepted Accounting Principles use mean-variance methods. Textbooks carry the CAPM and have students calculate the cost of capital based on the model. Real firms evaluate management and projects on what is an improper algorithm. Regulatory models that use a normal distribution rather than a Cauchy distribution to calculate required capital for financial intermediaries will tend to undercapitalize those institutions. Hedge funds using Itô calculus based methods are using methods uncorrelated with the true model.

The profession has discussed this issue for fifty years. It is time to put it to bed.

### 7.4.1 Moving Forward

Knowing the distributions going forward permits a new range of policy mechanisms. For example, as utility maximization, when combined with Bayesian analysis is admissible, it should be possible to construct Deming style processes for pension fund management. Fund managers should be able to state, ex ante, the predictive distribution of their policies once a better understanding of the variables involved are studied in greater depth. Trustees should be able to state their assumptions and the costs associated with differing outcomes. Any improvement in the management of financial capital improves the well being of labor, as labor competes with physical capital.

Historically, there have been significant improvements in the utilization of labor. This permits significant improvements in the utilization of capital. It is hoped this will result in improvements in the real wage.

The challenge with moving forward is that there is so much to do. Mutual funds should stop reporting average returns. Accounting ratios, as ratio data, need an

analysis of the true underlying distribution performed. Old articles need dusted off and checked for the mathematical principles used. New methods of integration and differentiation need to be created in stochastic calculus. Undergraduate education needs at least a minor tweaking. Dealing with uncertainty needs to be moved into an undergraduate forum.

It is important to remember that a very short time ago, sequencing DNA was a strictly doctoral level skill. Now it is an undergraduate level skill. It does not follow that dealing with uncertainty is a doctoral level skill. Bayesian decision theory is very accessible.

The list goes on. The way forward is to look at past ideas and to vet them once again for the model construction, the statistical tests used and the intuition. Regulatory structures have had the efficient market hypothesis built into their very construction using ideas that are no longer supportable.

Finally, Bayes factors and posterior testing allow a contraction of the model space. Economics is wonderful at generating models with statistical significance, but not so good at picking and choosing among the crowd.

Fortunately, individuals will pick a problem, solve it and move on. Self interest guarantees the large problems will be solved in relatively short order. There are many dissertations to publish in the next decade. This is just a first work in a chain of works passing through Eugene Fama and Benoit Mandelbrot and backwards to Augustin Cauchy and even further back to Maria Agnesi and Fermat.

## 7.5 Implications for Other Fields of Economics

Any deferral of consumption for the purpose of improving well being in the future will trigger a Cauchy distribution being somewhere in the mixture of distributions. The study of marriage, religion, child rearing and so forth in all the social sciences will likely be impacted in significant ways. There is no average marital experience. There is no average religious experience and there is no average time spent with a child growing up.

## 7.6 Finishing Up

This work is a start. It provides a frustrating sense of incompleteness. Seven chapters are too short. Had dividends been tackled there would have been at least two further chapters. Had the impact of interest rates been attempted then a book on the yield curve would need to be constructed and then a volume on prices given the curve. Bankruptcy and merger risk could have each held their own chapter. Liquidity costs and counter-party failure risks could have covered at least two additional estimators and three or four additional chapters as well. A chapter or two on moral hazard and adverse selection when informational asymmetries are included should be present. A chapter on time inconsistency for boards of directors and for legislative inputs should be required. What is the government risk built into a contract? Finally, an appendix or two on Bose-Einstein statistics, planar Brownian motion and quaternions may end up being helpful.

This work is a start, but just barely.

# DISTRIBUTION OF RETURNS GIVEN A MERGER WILL HAPPEN

To understand a possible derivation for the distribution of returns *given* that a merger will happen, it is helpful first to derive the distribution of returns, given that the firm will be a going concern over the entire period. To arrive at the Cauchy distribution, it is necessary to set the underlying conditions.

## A.1 Preliminaries

In economic discussions, the sample would be made up of pairs of prices, such that any given price is in the set:

$$\{p_t : 0 < p_t < \infty\}, \quad (\text{A.1})$$

and returns would be derived from a pair of prices, the buying and selling price, such that

$$(p_\tau, p_T) \in \mathfrak{R}^2 \quad (\text{A.2})$$

In statistics and mathematics, it is more common to place this ordered pair in the complex plane rather than in  $\mathfrak{R}^2$ . The differences are subtle and there is no loss of generality created here in using  $\mathfrak{R}^2$ .

There are advantages in using the complex plane when the Cauchy distribution is present. They do not matter in this derivation. In particular, the complex plane isn't a total ordering, and all analytic functions are infinitely differentiable. Any point can be thought of as  $(0, 0)$  without the need of a transformation.

This derivation of the Cauchy distribution and the distribution of returns, given a merger will happen, will not be stated in price and return space. Because of this, there will be some loss in specificity. The gain is in notational simplicity and generality.

It is a small matter to add the minor complications created by bounded real prices, but it is not a small matter of presentation. Further, this removes time from the equations and simply allows the subjects in question to be an ordered pair of numbers centered for computational convenience around zero.

The first goal is to construct a derivation of the Cauchy distribution that can be used with the added information that a merger will happen.

In order to facilitate this, it is important to note the sample space is simply:

$$\Omega \subseteq \{(x, y) : (x, y) \in \mathbb{R}^2 \text{ and } x \neq 0\} \quad (\text{A.3})$$

This does not show how these two variables relate to each other. So an event space, based on a functional relationship between the variables is necessary. The set of all possible events is defined as:

$$\mathcal{F} \subseteq \left\{ S : \frac{y}{x} \rightarrow S, \forall (x, y) \in \Omega \right\} \quad (\text{A.4})$$

What is missing is a function to map  $\mathcal{F}$  onto the closed interval  $[0, 1]$ . Extending the assumption of normality, so ubiquitous in economics, to  $x$  and  $y$  it is convenient to define each as a random variables whose marginal distributions are:

$$f_x(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}x^2 \right\}, \quad (\text{A.5})$$

and

$$f_y(y) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}y^2 \right\} \quad (\text{A.6})$$

This use of the standard normal distribution is not without a slight loss in

generality. By forcing the variances to be equal the errors will be circular rather than elliptical. This computational convenience does mean that the result will be the standard Cauchy and the standard merger distribution, rather than a more general one that allowed the variance to systematically change over time.

Further, it is assumed that the variables are independent. This is very reasonable and creates a joint marginal distribution of:

$$f_{xy}(x, y) = f_x(x)f_y(y) = \frac{1}{2\pi} \exp \left\{ -\frac{x^2 + y^2}{2} \right\} \quad (\text{A.7})$$

The next step is to relate this marginal distribution to the relationship between  $x$  and  $y$ . The importance of  $S$  is that removes the importance of the level of  $x$  or  $y$  in the same manner that returns remove the importance of levels of prices. This relationship, can, of course be inverted. For any line  $S$ , the value of  $y$  can be known if  $x$  is known using the simple equation of a line,

$$y = Sx \quad (\text{A.8})$$

The challenge then is to find the statistical distribution of the lines implied by the set  $\mathcal{F}$ . Since it is the entire line that is of interest and not a point-wise solution, it is logical to convert lines into angles and segments rather than pairings of points through the transformation:

$$x = r \cos \theta \text{ and } y = r \sin \theta \quad (\text{A.9})$$

Noting that each individual line is of measure zero, the cdf of the lines will be taken in a neighborhood around  $S$ . Since a change in  $S$  implies a change in slope, then as the distance from the origin goes to infinity, the small shift will be quite a large distance between two points taken at the same radius on each line.

Transforming the density function to polar coordinates, the improper integral for the cumulative density function is:

$$\frac{2}{2\pi} \int \int e^{-\frac{r^2}{2}} r dr d\theta \quad (\text{A.10})$$

The *extra*  $r$  in the equation comes from the need to scale the disturbance in  $S$  as

$r$  becomes large. The *extra* 2 comes from the fact that there is an identical area above and below the x-axis.

If the cumulative density function of  $S$  over a small neighborhood is considered, then this is equivalent to:

$$F_S(S)dS = \Pr[s \in [S, S + dS]] \quad (\text{A.11})$$

Differentiating, this allows a statement of the pdf. Noting that

$$\tan \theta = \frac{y}{x} = S, \quad (\text{A.12})$$

it obviously follows that,

$$\theta = \tan^{-1}(S) \quad (\text{A.13})$$

The proper cumulative density function, noting the above relationship, for  $S$  is:

$$\Pr(0 < S \leq s) = F_S(0 < S \leq s) = \frac{1}{\pi} \int_0^{\tan^{-1}(S)} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta \quad (\text{A.14})$$

The interior integral

$$\int_0^{\infty} e^{-\frac{r^2}{2}} r dr = 1, \quad (\text{A.15})$$

so,

$$\Pr(0 < S \leq s) = F_S(0 < S \leq s) = \frac{1}{\pi} \int_0^{\tan^{-1}(S)} d\theta, \quad (\text{A.16})$$

leaving

$$\Pr(0 < S \leq s) = F_S(0 < S \leq s) = \frac{1}{\pi} \tan^{-1}(S), \quad (\text{A.17})$$

as the cdf.

Differentiating, this leaves the standard Cauchy distribution of:

$$f_S(S) = \frac{1}{\pi} \frac{1}{1 + S^2} \quad (\text{A.18})$$

## A.2 Mergers

Except for the party initiating the merger and, at some point, the management of the acquired firm, no one in the general public knows the terms and conditions of

the merger or acquisition prior to announcement.

In order to acquire a large enough percentage, if not all outstanding shares, the acquiring party must increase the price to disgorge the shares from existing owners. The buyer has to pay legal fees, fees for due diligence and to move along the supply curve enough to acquire the desired percentage of shares. So if (0,0) is the equilibrium point, in the absence of a merger, it must be the case that the new equilibrium will be at (0,k). Again if the appraisal errors are normal about the center of location k, this leads to a new distribution.

$y = Sx$  is not redefined by the merger; it still passes through (0,0); the appraisal errors no longer are centered there. This gives an interesting boundary condition for value investing. Because the owners of a firm acquired by another firm can be forced into a perfect state of illiquidity, the acquiring firm may not need to go all the way to the equilibrium point. A value investor should consider a margin of safety to avoid complete illiquidity or a loss in realized value due to a forced acquisition.

As it is the future price that is affected, the distribution that will shift along the y-axis. This leaves density function of:

$$f_{xy}(x, y) = \frac{1}{2\pi} \exp \left\{ -\frac{x^2 + y^2 - 2yk + k^2}{2} \right\}. \quad (\text{A.19})$$

Translating back into polar coordinates, the density function becomes:

$$f_S(S) = \frac{1}{2\pi} \exp \left\{ -\frac{r^2 - 2rk \sin \theta + k^2}{2} \right\}. \quad (\text{A.20})$$

Solving as before by integrating over the density function using the neighborhoods around the lines and differentiating, the standard pdf is:

$$\frac{d}{dS} \frac{\exp\left(-\frac{k^2}{2}\right) \int_0^{\tan^{-1}(S)} \left( \int_0^\infty r \exp\left(-\frac{1}{2}(r^2 - 2kr \sin(\theta))\right) dr \right) d\theta}{2\pi} = \frac{e^{-\frac{k^2}{2}} \left( \frac{\sqrt{\frac{\pi}{2}} k S e^{\frac{k^2 S^2}{2(S^2+1)}} \left( \operatorname{erf}\left(\frac{kS}{\sqrt{2}\sqrt{S^2+1}}\right) + 1 \right)}{\sqrt{S^2+1}} + 1 \right)}{2\pi (S^2 + 1)} \quad (\text{A.21})$$

# SOFTWARE USED IN THE IMPLEMENTATION OF THE RESEARCH

## B.1 Chapter 5–Comparing Competing Models

```
/*options mprint symbolgen mlogic;*/  
/*options obs=100000;*/  
  
/*gathers all price data, in raw form, and calculates maximum and minimum  
values of dates*/  
proc sql;  
create view gather as  
select permno,date,prc,cfacpr,4*(year(date)-1929)+qtr(date)  
as quarter,min(date) as dmin, max(date) as dmax  
from crsp.dsf  
where prc>0 and ^missing(prc)  
group by permno  
order by permno,date;  
quit;
```

```

/*restricts data to common shares only and assigns final status code*/
data common_only/view=common_only;
set crsp.dseall(keep=permno dlstcd shrcd dlamt nwperm
where=(^missing(dlstcd)));
by permno;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
      else if substr(left(shrcd),2,1)='5' then delete;
      else if substr(left(shrcd),2,1)='8' then delete;
      else if substr(left(shrcd),1,1)='2' then delete;
      else if substr(left(shrcd),1,1)='4' then delete;
      else if substr(left(shrcd),1,1)='7' then delete;
operational=0;
merge=0;
cash_buyout=0;
private=0;
bankrupt=0;
if dlstcd=100 then operational=1;
      else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
      else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;
      else if dlstcd in(502,510,513,514,516,517,519,520,573,575,580,581,
582,583,584) then private=1;
      else if dlstcd in(574) then bankrupt=1;
      else if dlstcd in(234,242,243,251,252,261,262,300,332,341,342,400,450
,460,470,500) then do;
          if dlstcd in(234,500) then do;
              if nwperm>0 then merge=1;
              else if nwperm=0 and dlamt>0 then cash_buyout=1;
              else private=1;
          end;
          else if dlstcd in (242,243,251,341,342) then do;
              if nwperm>0 then merge=1;

```

```

        else cash_buyout=1;
    end;
else if dlstcd in (252,261,262,450) then do;
    if nwperm>0 then merge=1;
    else if dlamt>0 then cash_buyout=1;
    else delete;
end;
else if dlstcd in(300) then do;
    if nwperm>0 then merge=1;
    else delete;
end;
else if dlstcd in (332) then do;
    if nwperm>0 then merge=1;
    else private=1;
end;
else if dlstcd in (400,460,470) then do;
    if dlamt>0 then cash_buyout=1;
    else bankrupt=1;
end;
end;
run;

/*links price data with list of common data and final states*/
proc sql;
create view link1 as
select a.*,b.operational,b.merge,b.cash_buyout,b.private,b.bankrupt
from gather as a, common_only as b
where a.permno=b.permno
order by permno;
quit;

/*updates status to so that final status only happens in final trading year*/
data link2/view=link2;

```

```

set link1;
if operational=0 then do;
if dmax-date>365 then do;
operational=1;
bankrupt=0;
merge=0;
private=0;
cash_buyout=0;
end;
end;
bankrupt=bankrupt+private;
drop private;
run;

/*creates a framework of permno-dates that include weekends and holidays*/
data frame1/view=frame1;
set gather(keep=permno dmin dmax);
by permno;
run;

/*removes duplicate permno*/
data frame2/view=frame2;
set frame1;
by permno;
lpermno=lag(permno);
run;

/*creates a calendar of dates*/
data frame3;
set frame2(where=(permno NE lpermno and dmax-dmin>730));
by permno;
do date=dmin to dmax;
output;

```

```

end;
run;

/*splits the data into operational, merger
and cash. bankruptcy is not needed as it goes to zero*/
data operational1/view=operational1;
set link2(where=(operational=1));
by permno;
run;

/*joins trading calendar with the full calendar*/
proc sql;
create view operational2 as
select a.*,b.prc,b.cfacpr,b.quarter
from frame3 as a
left join
operational1 as b
on a.permno=b.permno and a.date=b.date;
quit;

proc sort data=operational2 out=operational3;
by permno descending date;
run;

/*creates lead prices*/
%macro lead;

data operational4/view=operational4;
set operational3;
by permno descending date;
%do n=365 %to 352 %by -1;

```

```

lprc&n=lag&n(prc);
lcfacpr&n=lag&n(cfacpr);
%if %eval(&n)=365 %then %do;
lpermno&n=lag&n(permno);
%end;
%end;
run;

proc sort data=operational4 out=operational5;
by permno date;
run;

proc datasets library=work;
run;

/*creates return*/
/*not normalized as trade data for dissertation purposes is in nominal form*/
data operational6/view=operational6;
set operational5(where=(^missing(prc)));
by permno date;
if permno=lpermno365 then do;
if ^missing(lprc365) then return=lprc365/prc*cfacpr/lcfacpr365;
else if ^missing(lprc364) then return=lprc364/prc*cfacpr/lcfacpr364;
else if ^missing(lprc363) then return=lprc363/prc*cfacpr/lcfacpr363;
else if ^missing(lprc362) then return=lprc362/prc*cfacpr/lcfacpr362;
else if ^missing(lprc361) then return=lprc361/prc*cfacpr/lcfacpr361;
else if ^missing(lprc360) then return=lprc360/prc*cfacpr/lcfacpr360;
else if ^missing(lprc359) then return=lprc359/prc*cfacpr/lcfacpr359;
else if ^missing(lprc358) then return=lprc358/prc*cfacpr/lcfacpr358;
else if ^missing(lprc357) then return=lprc357/prc*cfacpr/lcfacpr357;
else if ^missing(lprc356) then return=lprc356/prc*cfacpr/lcfacpr356;
else if ^missing(lprc355) then return=lprc355/prc*cfacpr/lcfacpr355;
else if ^missing(lprc354) then return=lprc354/prc*cfacpr/lcfacpr354;

```

```

else if ^missing(lprc353) then return=lprc353/prc*cfacpr/lcfacpr353;
else if ^missing(lprc352) then return=lprc352/prc*cfacpr/lcfacpr352;
else return=.;
end;
run;

data operational7/view=operational7;
set operational6(where=(^missing(return)));
keep return;
run;

%mend;
%lead;

/*calculates the log likelihood*/
%macro lfunction;

data return1/view=return1;
set operational7;
%do i=28200 %to 28400;
%let shape=0.00001*&i;
Cauchy_log_likelihood&i=LOG(&shape/(((&shape)**2)+(return-1)**2))
-LOG(constant('pi')/2+atan(1/(&shape)));
%end;
%do i=99500 %to 99700;
%let shape=&i/100000;
normal_log_likelihood&i=-((return-1)**2)/(2*(&shape)**2)
-log(1-cdf('normal',0,1,&shape))-log(&shape)-1/2*log(2*constant('pi'));
%end;
run;

%mend;

```

```

%lfunction;

proc summary data=return1;
var Cauchy_log_likelihood28200-Cauchy_log_likelihood28400
    normal_log_likelihood99500-normal_log_likelihood99700;
output out=return2 sum=;
run;

proc export
data=return2
outfile='/home/wvu/deharris/new/bayes_factors/bayes_factor.csv'
dbms=csv;

```

This data was exported into Excel where very simple numerical integration was performed.

## B.2 Chapter 6-Bankruptcies From 1925-1961

```

options fullstimer;

/*This program sets prior distributions for transactions that
happened before the start of the compustat database*/

/*Since the compustat database begins on 1/1/62 the last
purchase can be on 12/31/61 with a final close of 12/31/62*/

proc summary data=crsp.dsf(where=((prc>0))) n max min;
by permno;
var date;
output out=work.dates n=number max=max_date min=min_date;
run;

```

```

/*this excludes rarely traded securities*/
/*this also deletes securities first started after the
compustat start date*/
data work.dates;
set work.dates (where=( _FREQ_>510 and min_date<MDY(1,1,1962)));
by permno;
keep permno max_date min_date;
run;

proc sort data=work.dates out=sorted_dates;
by permno;
run;

proc datasets library=work;
delete dates;
run;

data common_only/view=common_only;
merge sorted_dates(in=aa) crsp.dseall(keep=permno dlstcd shrccd
  dlamt nwperm where=(^missing(dlstcd)) in=bb);
by permno;
if aa and bb then do;
    if substr(left(shrccd),2,1)='3' then delete;
    else if substr(left(shrccd),2,1)='4' then delete;
    else if substr(left(shrccd),2,1)='5' then delete;
    else if substr(left(shrccd),2,1)='8' then delete;
    else if substr(left(shrccd),1,1)='2' then delete;
    else if substr(left(shrccd),1,1)='4' then delete;
    else if substr(left(shrccd),1,1)='7' then delete;
    else do;
        operational=0;
        bankrupt=0;
    end;
end;

```

```

                cash=0;
                merge=0;
                private=0;
            end;
        end;
    else delete;
    drop shrcd;
    run;

data active/view=active;
set common_only(where=(dlstcd=100 or max_date>mdy(12,31,1961)));
operational=mdy(12,31,1961)-min_date+1;
run;

data inactive/view=inactive;
set common_only(where=(dlstcd^=100 and max_date<mdy(1,1,1962)));
run;

data merger/view=merger;
set inactive (where=(dlstcd in(200,231,232,241,244,301,331,343)));
operational=max_date-min_date-365+1;
merge=365;
run;

data cash/view=cash;
set inactive (where=(dlstcd in(233,235,271,333,334,361)));
operational=max_date-min_date-365+1;
cash=365;
run;

data private/view=private;
set inactive (where=(dlstcd in(502,510,513,514,516,517,519,520,

```

```

573,575,580,581,582,583,584))) );
operational=max_date-min_date-365+1;
private=365;
run;

data bankrupt/view=bankrupt;
set inactive (where=(dlstcd in(574)));
operational=max_date-min_date-365+1;
bankrupt=365;
run;

data mixed/view=mixed;
set inactive(where=(dlstcd in(234,242,243,251,252,261,262,300,
332,341,342,400,450,460,470,500)));
operational=max_date-min_date-365+1;
if dlstcd in(234,500) then do;
if nwperm>0 then merge=365;
else if nwperm=0 and dlamt>0 then cash_payout=365;
else private=365;
end;
if dlstcd in (242,243,251,341,342) then do;
if nwperm>0 then merge=365;
else cash=365;
end;
if dlstcd in (252,261,262,450) then do;
if nwperm>0 then merge=365;
else if dlamt>0 then cash=365;
else delete;
end;
if dlstcd in(300) then do;
if nwperm>0 then merge=365;
else delete;
end;

```

```

if dlstcd in (332) then do;
if nwperm>0 then merge=365;
else private=365;
end;
if dlstcd in (400,460,470) then do;
if dlamt>0 then cash=365;
else bankrupt=365;
end;
run;

data combined/view=combined;
set active merger cash private bankrupt mixed;
run;

proc summary data=combined sum;
var operational merge cash private bankrupt;
output out=state_summary sum(operational)=s_op
sum(merge)=s_m sum(cash)=s_cash sum(private)=s_p sum(bankrupt)=s_b;
run;

proc datasets library=work;
delete sorted_dates;
run;

data posterior/view=posterior;
set state_summary;
total=s_op+s_m+s_cash+s_p+s_b;
going=s_op;
merge=s_m+s_cash;
bankrupt=s_p+s_b;
m_going=going/total;
v_going=(going*(total-going))/((total+1)*total**2);
m_merge=merge/total;

```

```

v_merge=(merge*(total-merge))/((total+1)*total**2);
m_bankrupt=bankrupt/total;
v_bankrupt=bankrupt*(total-bankrupt)/((total+1)*total**2);
label total="Total number of observations"
going="Total number of going concerns"
merge="Total number of mergers"
bankrupt="Total number of bankruptcies"
m_going="Posterior mean of going concerns"
v_going="Posterior variance of going concerns"
m_merge="Posterior mean of mergers"
v_merge="Posterior variance of mergers"
m_bankrupt="Posterior mean of bankruptcies"
v_bankrupt="Posterior variance of bankruptcies"
;
keep total going merge bankrupt m_going v_going m_merge v_merge
m_bankrupt v_bankrupt;
run;

proc print data=posterior;
run;

```

## B.3 Chapter 6-Model Selection for Bankruptcy Prior Distribution 1961-1990

### B.3.1 GDP Calculation

```

libname gdp '/home/wvu/deharris/new/state2';

/*January 1, 1929 is first day of quarter 1*/
/*in order to calculate the quarter take (year-1929)+current

```

```
quarter*/

data raw_data;
input quarter 1-3 gdp 4-11;
datalines;
  4 103.6
  8 91.2
 12 76.5
 16 58.7
 20 56.4
 24 66
 28 73.3
 32 83.8
 36 91.9
 40 86.1
 44 92.2
 48 101.4
 52 126.7
 56 161.9
 60 198.6
 64 219.8
 68 223
 72 222.2
 73 237.2
 74 240.4
 75 244.5
 76 254.3
 77 260.3
 78 267.3
 79 273.8
 80 275.1
 81 269.9
 82 266.2
```

83	267.6
84	265.2
85	275.2
86	284.5
87	301.9
88	313.3
89	329
90	336.6
91	343.5
92	347.9
93	351.2
94	352.1
95	358.5
96	371.4
97	378.4
98	382
99	381.1
100	375.9
101	375.2
102	376
103	380.8
104	389.4
105	402.6
106	410.9
107	419.4
108	426
109	428.3
110	434.2
111	439.2
112	448.1
113	457.2
114	459.2
115	466.4

116 461.5  
117 453.9  
118 458  
119 471.7  
120 485  
121 495.5  
122 508.5  
123 509.3  
124 513.2  
125 527  
126 526.2  
127 529  
128 523.7  
129 528  
130 539  
131 549.5  
132 562.6  
133 576.1  
134 583.2  
135 590  
136 593.3  
137 602.5  
138 611.2  
139 623.9  
140 633.5  
141 649.6  
142 658.9  
143 670.5  
144 675.6  
145 695.7  
146 708.1  
147 725.2  
148 747.5

149 770.8  
150 779.9  
151 793.1  
152 806.9  
153 817.8  
154 822.3  
155 837  
156 852.7  
157 879.8  
158 904.1  
159 919.3  
160 936.2  
161 960.9  
162 976.1  
163 996.3  
164 1004.5  
165 1017.1  
166 1033.1  
167 1050.5  
168 1052.7  
169 1098.1  
170 1118.8  
171 1139.1  
172 1151.4  
173 1190.1  
174 1225.6  
175 1249.3  
176 1286.6  
177 1335.1  
178 1371.5  
179 1390.7  
180 1431.8  
181 1446.5

182 1484.8  
183 1513.7  
184 1552.8  
185 1569.4  
186 1605  
187 1662.4  
188 1713.9  
189 1771.9  
190 1804.2  
191 1837.7  
192 1884.5  
193 1938.5  
194 2005.2  
195 2066  
196 2110.8  
197 2149.1  
198 2274.7  
199 2335.2  
200 2416  
201 2463.3  
202 2526.4  
203 2599.7  
204 2659.4  
205 2724.1  
206 2728  
207 2785.2  
208 2915.3  
209 3051.4  
210 3084.3  
211 3177  
212 3194.7  
213 3184.9  
214 3240.9

215 3274.4  
216 3312.5  
217 3381  
218 3482.2  
219 3587.1  
220 3688.1  
221 3807.4  
222 3906.3  
223 3976  
224 4034  
225 4117.2  
226 4175.7  
227 4258.3  
228 4318.7  
229 4382.4  
230 4423.2  
231 4491.3  
232 4543.3  
233 4611.1  
234 4686.7  
235 4764.5  
236 4883.1  
237 4948.6  
238 5059.3  
239 5142.8  
240 5251  
241 5360.3  
242 5453.6  
243 5532.9  
244 5581.7  
245 5708.1  
246 5797.4  
247 5850.6

248 5846  
249 5880.2  
250 5962  
251 6033.7  
252 6092.5  
253 6190.7  
254 6295.2  
255 6389.7  
256 6493.6  
257 6544.5  
258 6622.7  
259 6688.3  
260 6813.8  
261 6916.3  
262 7044.3  
263 7131.8  
264 7248.2  
265 7307.7  
266 7355.8  
267 7452.5  
268 7542.5  
269 7638.2  
270 7800  
271 7892.7  
272 8023  
273 8137  
274 8276.8  
275 8409.9  
276 8505.7  
277 8600.6  
278 8698.6  
279 8847.2  
280 9027.5

281 9148.6  
282 9252.6  
283 9405.1  
284 9607.7  
285 9709.5  
286 9949.1  
287 10017.5  
288 10129.8  
289 10165.1  
290 10301.3  
291 10305.2  
292 10373.1  
293 10498.7  
294 10601.9  
295 10701.7  
296 10766.9  
297 10887.4  
298 11011.6  
299 11255.1  
300 11414.8  
301 11589.9  
302 11762.9  
303 11936.3  
304 12123.9  
305 12361.8  
306 12500  
307 12728.6  
308 12901.4  
309 13161.4  
310 13330.4  
311 13432.8  
312 13584.2  
313 13758.5

```
314 13976.8
315 14126.2
316 14253.2
317 14273.9
318 14415.5
319 14395.1
320 14081.7
321 13893.7
322 13854.1
323 13920.5
324 14087.4
325 14277.9
326 14467.8
327 14605.5
328 14755
329 14867.8
330 15012.8
331 15176.1
332 15319.4
333 15454
;
run;

/*this segment creates lagged variables*/
data lag_gdp/view=lag_gdp;
set raw_data;
lag_gdp=lag(gdp);
lag_quarter=lag(quarter);
run;

/*this segment represents the change in gdp per quarter*/
data gdpratio/view=gdpratio;
set lag_gdp(where=(^missing(lag_gdp)));
```

```

ratio=(gdp/lag_gdp)**(1/(quarter-lag_quarter));
run;

/*this initializes the final data set*/
data final;
set gdpratio;
rank_ratio=1000;
label rank_ratio="Empirical percentile of observed change in GDP"
;
run;

%macro percentiles;

%do i=129 %to 333;

proc rank data=gdpratio(where=(quarter<=&i)) percent ties=mean out=ranking;
var ratio;
ranks rank_ratio;
run;

data rank_ratio;
set ranking(where=(quarter=&i));
call symput('transfer',rank_ratio);
run;

data final;
set final;
if quarter=&i then rank_ratio=&transfer;
run;

proc datasets library=work;

```

```
delete rank_ratio ranking;
run;

%end;

%mend percentiles;

%percentiles;

/*convert to tukey groups*/
data final;
set final(where=(rank_ratio<1000) keep=quarter rank_ratio);
if rank_ratio>10 and rank_ratio<90 then rank_ratio=0;
else if rank_ratio LE 10 then rank_ratio=-1;
else rank_ratio=1;
run;

data gdp.storage;
set final;
run;

proc datasets library=gdp;
run;

proc contents data=gdp.storage;
run;

proc print data=gdp.storage;
run;
```

### B.3.2 Model Selection

```

libname gdp '/home/wvu/deharris/new/state2';

/*options mprint symbolgen mlogic obs=1000000;*/

/*gather gvkey permno linkages*/
data link/view=link;
set crspa.ccmxpf_lnkused(keep=ugvkey ulinkdt apermno ulinkenddt
ulinktype where=(ulinktype in("LC","LU","LS")))
rename=(ugvkey=gvkey apermno=permno));
by permno;
if missing(ulinkenddt) then ulinkenddt=mdy(12,31,2013);
if missing(ulinkdt) then delete;
if missing(gvkey) then delete;
drop ulinktype;
run;

/*find boundary dates*/
proc summary data=crsp.dsf(where=(prc>0 and ^missing(prc)));
by permno;
var date;
output out=minimax min(date)=dmin max(date)=dmax;
run;

proc summary data=crsp.dsf;
by permno;
var date;
output out=end_date max(date)=end_date;
run;

/*merge*/
data crsp1/view=crsp1;
merge minimax(in=aa drop=_type_ _freq_)

```

```

end_date(in=bb drop=_type_ _freq_);
by permno;
if aa and bb;
run;

/*merge*/
data crsp2/view=crsp2;
merge crsp1(in=aa where=(dmax-dmin>730)) link;
by permno;
if aa;
run;

/*determine final disposition*/
data state_space/view=state_space;
set crsp.dseall(keep=permno dlstcd shrcd dlamt
nwperm where=(^missing(dlstcd)));
by permno;
operational=0;
merge=0;
cash_buyout=0;
private=0;
bankrupt=0;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
else if substr(left(shrcd),2,1)='5' then delete;
else if substr(left(shrcd),2,1)='8' then delete;
else if substr(left(shrcd),1,1)='2' then delete;
else if substr(left(shrcd),1,1)='4' then delete;
else if substr(left(shrcd),1,1)='7' then delete;
else do;
if dlstcd=100 then operational=1;
else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;

```

```

else if dlstcd in(502,510,513,514,516,517,519,520,573,
575,580,581,582,583,584) then
private=1;
else if dlstcd in(574) then bankrupt=1;
      else if dlstcd
in(234,242,243,251,252,261,262,300,
332,341,342,400,450,460,470,500) then do;
      if dlstcd in(234,500) then do;
          if nwperm>0 then merge=1;
          else if nwperm=0 and dlamt>0 then
              cash_buyout=1;
          else private=1;
      end;
      else if dlstcd in (242,243,251,341,342) then do;
          if nwperm>0 then merge=1;
          else cash_buyout=1;
      end;
      else if dlstcd in (252,261,262,450) then do;
          if nwperm>0 then merge=1;
          else if dlamt>0 then cash_buyout=1;
          else delete;
      end;

end;

end;
else if dlstcd in(300) then do;
      if nwperm>0 then merge=1;
      else delete;

end;
else if dlstcd in (332) then do;
      if nwperm>0 then merge=1;
      else private=1;

end;

```

```

        else if dlstcd in (400,460,470) then do;
            if dlamt>0 then cash_buyout=1;
            else bankrupt=1;
        end;
    end;
end;
bankrupt=private+bankrupt;
if bankrupt=1 then operational=0;
else operational=1;
drop shrcd nwperm dlamt dlstcd cash_buyout private merge;
run;

/*merge*/
data crsp3/view=crsp3;
merge crsp2(in=aa) state_space(in=bb);
by permno;
if aa and bb;
run;

/*find blank gvkey*/
data crsp4/view=crsp4;
set crsp3;
by permno;
test=length(gvkey);
run;

/*for items with missing gvkey*/
data nogvkey1/view=nogvkey1;
set crsp4(where=(test=1));
by permno;
drop gvkey test ulinkdt ulinkenddt;
run;

/*items with gvkey*/

```

```

data gvkey1/view=gvkey1;
set crsp4(where=(test>1));
by permno;
drop test;
run;

/*sort by gvkey*/
proc sort data=gvkey1 out=gvkey2;
by gvkey;
run;

proc datasets library=work;
run;

/*gather acct data*/
data acct_data/view=acct_data;
set comp.fundq(keep=datadate gvkey IBQ CSHOQ SALEQ COGSQ
NIQ CEQQ ATQ PPENTQ INVTQ RECTQ APQ DLCQ DLTTQ OIADPQ ACTQ
LCTQ CHQ CHEQ
LTQ DLTTQ where=(^missing(date) and date>mdy(12,31,1961)
and date<mdy(7,1,1990))rename=(datadate=date));
by gvkey date;
if missing(NIQ) then NIQ=IBQ;
if missing(CHEQ) then cash=chq;
else cash=CHEQ;
quarter=4*(year(date)-1929)+qtr(date)+1;
label cash="Cash"
       quarter="Quarter"
;
drop IBQ CHEQ CHQ;
run;

```

```

/*convert to ratios*/
data ratios1/view=ratios1;
/*create financial ratios*/
set acct_data;
by gvkey date;
/* Gross margin*/
if (~missing(SALEQ) and ~missing(COGSQ) and
(SALEQ NE 0)) then gm=(SALEQ-COGSQ)/SALEQ;
    else gm=.;
/*Net Margin*/
if (~missing(niq) and ~missing(SALEQ) and
(SALEQ NE 0)) then nm=niq/SALEQ;
    else nm=.;
/*Return on Equity*/
if (~missing(NIQ) and ~missing(CEQQ) and
(CEQQ NE 0)) then roe=NIQ/CEQQ;
    else roe=.;
/*Return on Assets*/
if (~missing(NIQ) and ~missing(ATQ) and
(ATQ NE 0)) then roa=NIQ/ATQ;
else roa=.;
/*return on net assets*/
if (~missing(NIQ) and ~missing(PPENTQ) and
~missing(INVTQ) and ~missing(RECTQ) and
~missing(APQ) and (PPENTQ+INVTQ+RECTQ-APQ NE 0)) then
rona=NIQ/(PPENTQ+INVTQ+RECTQ-APQ);
    else rona=.;
/*Debt to Equity ratio*/
if (~missing(DLCQ) and ~missing(DLTTQ) and ~missing(CEQQ)
and (CEQQ NE 0)) then de=(DLCQ+DLTTQ)/CEQQ;
    else de=.;
/*Earning power ratio*/
if (~missing(OIADPQ) and ~missing(ATQ)

```

```

and (ATQ NE 0)) then ep=OIADPQ/ATQ;
    else ep=.;
/*Current ratio*/
if (^missing(ACTQ) and ^missing(LCTQ) and
(LCTQ NE 0)) then cr=ACTQ/LCTQ;
    else cr=.;
/*acid ratio*/
if (^missing(ACTQ) and ^missing(INVTQ) and
^missing(LCTQ) and (LCTQ NE 0)) then acid=(ACTQ-INVTQ)/LCTQ;
    else acid=.;
/*Note that XPPQ is prepaid expense but is
null in the entire data set and so was not used.*/
/*cash ratio*/
if (^missing(cash) and ^missing(LCTQ)
and (LCTQ NE 0)) then cashratio=cash/LCTQ;
    else cashratio=.;
/*asset turnover*/
if (^missing(SALEQ) and ^missing(ATQ) and
(ATQ NE 0)) then turnover=SALEQ/ATQ;
    else turnover=.;
/*receivables conversion*/
if (^missing(RECTQ) and ^missing(SALEQ)
and (SALEQ NE 0)) then rec_conv=RECTQ/SALEQ;
    else rec_conv=.;
/*debt ratio*/
if (^missing(LTQ) and ^missing(ATQ)
and (ATQ NE 0)) then debt_r=LTQ/ATQ;
    else debt_r=.;
/*long term debt to equity*/
if (^missing(DLTTQ) and ^missing(CEQQ)
and (CEQQ NE 0)) then ltd=DLTTQ/CEQQ;
    else ltd=.;
label acid="Acid ratio"

```

```

    cashratio="Cash ratio"
    cr="Current ratio"
    de="Debt to Equity ratio"
    debt_r="Debt ratio"
    ep="Earning power ratio"
    gm="Gross margin"
    ltd="Long term debt to equity"
    nm="Net margin"
    rec_conv="Receivables conversion rate"
    roa="Return on assets"
    roe="Return on equity"
    rona="Return on net assets"
    turnover="Asset turnovers"
;
drop cshoq SALEQ COGSQ NIQ CEQQ ATQ PPENTQ INVTQ
    RECTQ APQ DLCQ DLTTQ OIADPQ ACTQ LCTQ LTQ DLTTQ cash;
run;

/*sort by quarter to rank*/
proc sort data=ratios1 out=ratios2;
by quarter;
run;

/*rank data*/
proc rank data=ratios2 percent ties=mean out=ratios3;
by quarter;
var gm nm roe roa rona de debt_r ltd cr acid
    cashratio ep turnover rec_conv;
ranks rgm rnm rroe rroa rrona rde rdebt_r
    rltd rcr racid rcashratio rep rturnover rrec_conv;
run;

```

```

/*assings ratios to macro variables*/
%let a1=gm;
%let a2=nm;
%let a3=roe;
%let a4=roa;
%let a5=rona;
%let a6=de;
%let a7=debt_r;
%let a8=ltd;
%let a9=cr;
%let a10=acid;
%let a11=cashratio;
%let a12=ep;
%let a13=turnover;
%let a14=rec_conv;

%macro tukey;

data ratios4/view=ratios4;
set ratios3;
by quarter;
%do i=1 %to 14;
if ~missing(r&&a&i) then do;
if r&&a&i>10 and r&&a&i<90 then &&a&i=0;
else if r&&a&i LE 10 then &&a&i=-1;
else &&a&i=1;
end;
%end;
drop r&a1 r&a2 r&a3 r&a4 r&a5 r&a6 r&a7 r&a8
r&a9 r&a10 r&a11 r&a12 r&a13 r&a14;
run;

```

```

data ratios5/view=ratios5;
merge ratios4(in=aa) gdp.storage;
by quarter;
if aa;
run;

%mend;
%tukey;

proc sort data=ratios5 out=ratios6;
by gvkey;
run;

/*join crsp and compustat data*/
data gvkey3/view=gvkey3;
merge gvkey2(in=aa) ratios6;
by gvkey;
if aa;
run;

/*delete out of sample data*/
data gvkey4/view=gvkey4;
set gvkey3;
by gvkey;
if date>dmax then delete;
if date<dmin then delete;
if date>ulinkenddt then delete;
if date<ulinkdt then delete;
if dmax>mdy(6,30,1990) then dmax=mdy(6,30,1990);
if dmin<mdy(1,1,1961) then dmin=mdy(1,1,1960);
run;

```

```

/*sort by permno quarter*/
proc sort data=gvkey4 out=gvkey5;
by permno quarter;
run;

/*split data to create date data*/
proc sort data=gvkey5(keep=permno dmin
dmax end_date operational bankrupt)
nodupkey out=gvkey_left1;
by permno;
run;

/*create calendar*/
data gvkey_left2;
set gvkey_left1;
by permno;
do date=dmin to dmax;
quarter=4*(year(date)-1929)+qtr(date);
output;
end;
run;

/*create registry that items are sorted*/
proc sort data=gvkey_left2 presorted out=gvkey_left3;
by permno quarter;
run;

/*create quarter specific data*/
data gvkey_right1/view=gvkey_right1;
set gvkey5(keep=permno quarter gvkey ulinkdt ulinkenddt
&a1 &a2 &a3 &a4 &a5 &a6 &a7 &a8 &a9 &a10 &a11 &a12 &a13 &a14
rank_ratio);
by permno quarter;

```

```

run;

proc sort data=gvkey_right1 presorted out=gvkey_right2;
by permno quarter;
run;

/*remerge data*/
data gvkey6/view=gvkey6;
merge gvkey_left3 gvkey_right2;
by permno quarter;
run;

/*only allow for 365 days of bankruptcy*/
data gvkey7/view=gvkey7;
set gvkey6;
by permno quarter;
if bankrupt=1 then do;
if end_date-dmax>365 then operational=1;
end;
drop bankrupt;
run;

/*create filter to remove items without accounting
data to create models*/
%macro filter;

data gvkey8/view=gvkey8;
set gvkey7;
by permno quarter;
count=0;
%do i=1 %to 2;
    %do j=3 %to 14;
        if ^missing(&&a&i) and

```

```

^missing(&&a&j) then count=count+1;
%end;
%end;
%do i=3 %to 5;
    %do j=6 %to 14;
        if ^missing(&&a&i) and
            ^missing(&&a&j) then count=count+1;
    %end;
%end;
%do i=6 %to 8;
    %do j=9 %to 14;
        if ^missing(&&a&i) and
            ^missing(&&a&j) then count=count+1;
    %end;
%end;
%do i=9 %to 11;
    %do j=12 %to 14;
        if ^missing(&&a&i) and ^
            missing(&&a&j) then count=count+1;
    %end;
%end;
run;

%mend;

%filter;

/*create set with accounting data*/
data gvkey9/view=gvkey9;
set gvkey8(where=(count>0));
by permno quarter;
drop count;
run;

```

```

/*create set without accounting data*/
data gdponly/view=gdponly;
set gvkey8(where=(count=0));
drop count;
run;

%macro models;
%let model=0;

%do i=1 %to 2;
    %do j=3 %to 14;
%let model=%eval(&model+1);

proc freq data=gvkey9(where=(^missing(&&a&i) and
^missing(&&a&j) and ^missing(rank_ratio) and ^missing(operational)))
    noprint;
tables &&a&i*&&a&j*rank_ratio*operational/out=model&model;
run;

data g;
do i=-1 to 1;
do j=-1 to 1;
do k=-1 to 1;
do m=0 to 1;
&&a&i=i;
&&a&j=j;
rank_ratio=k;
operational=m;
output;

```

```

end;
end;
end;
end;
run;

```

```

data h;
merge g(drop=i j k m) model&model(drop=percent);
by &&a&i &&a&j rank_ratio operational;
if missing(count) then count=0;
run;

```

```

data model&model;
set h;
retain alpha 0;
if operational=0 then alpha=count;
else beta=count;
drop count;
run;

```

```

data model&model;
set model&model(where=(operational=1));
drop operational;
k=alpha;
n=alpha+beta;
alpha=1;
beta=32;
run;

```

```

proc datasets library=work;
delete g h;
run;

```

```

        %end;
    %end;
%do i=3 %to 5;
    %do j=6 %to 14;
        %let model=%eval(&model+1);

proc freq data=gvkey9(where=(^missing(&&a&i) and
    ^missing(&&a&j) and ^missing(rank_ratio) and
    ^missing(operational))) noprint;
tables &&a&i*&&a&j*rank_ratio*operational/out=model&model;
run;

data g;
do i=-1 to 1;
    do j=-1 to 1;
        do k=-1 to 1;
            do m=0 to 1;
                &&a&i=i;
                &&a&j=j;
                rank_ratio=k;
                operational=m;
                output;
            end;
        end;
    end;
end;

run;

data h;
merge g(drop=i j k m) model&model(drop=percent);
by &&a&i &&a&j rank_ratio operational;
if missing(count) then count=0;

```

```

run;

data model&model;
set h;

retain alpha 0;
if operational=0 then alpha=count;
else beta=count;
drop count;

run;

data model&model;
set model&model(where=(operational=1));
drop operational;
k=alpha;
n=alpha+beta;
alpha=1;
beta=32;
run;

proc datasets library=work;
delete g h;

run;

%end;

%end;

%do i=6 %to 8;
%do j=9 %to 14;
%let model=%eval(&model+1);

proc freq data=gvkey9(where=(^missing(&a&i)
and ^missing(&a&j) and ^missing(rank_ratio)
and ^missing(operational))) noprint;

```

```

tables &&a&i*&&a&j*rank_ratio*operational/out=model&model;
run;

data g;
do i=-1 to 1;
    do j=-1 to 1;
        do k=-1 to 1;
            do m=0 to 1;
                &&a&i=i;
                &&a&j=j;
                rank_ratio=k;
                operational=m;
                output;
            end;
        end;
    end;
end;

run;

data h;
merge g(drop=i j k m) model&model(drop=percent);
by &&a&i &&a&j rank_ratio operational;
if missing(count) then count=0;
run;

data model&model;
set h;
retain alpha 0;
if operational=0 then alpha=count;
    else beta=count;
drop count;
run;

```

```

data model&model;
set model&model(where=(operational=1));
drop operational;
k=alpha;
n=alpha+beta;
alpha=1;
beta=32;
run;

proc datasets library=work;
    delete g h;
run;

%end;
%end;
%do i=9 %to 11;
    %do j=12 %to 14;
        %let model=%eval(&model+1);

proc freq data=gvkey9(where=(^missing(&&a&i) and
^missing(&&a&j) and ^missing(rank_ratio) and
^missing(operational))) noprint;
tables &&a&i*&&a&j*rank_ratio*operational/out=model&model;
run;

data g;
do i=-1 to 1;
    do j=-1 to 1;
        do k=-1 to 1;
            do m=0 to 1;
                &&a&i=i;

```



```
proc datasets library=work;
    delete g h;
run;

    %end;
%end;
run;

%mend;

%models;

%macro identify;

%do i=1 %to 78;

proc summary data=model&i;
var n k;
output out=uniform&i sum=;
run;

data uniform&i;
set uniform&i(drop=_TYPE_ _FREQ_);
alpha=1;
beta=32;
run;

data uniform&i;
set uniform&i;
```

```

model=&i;
if n>0 then do;
    likelihood=0;
    do i=n to n+alpha+beta-1;
        likelihood=likelihood-log(i);
    end;
    if n ne k then do;
        do i=n-k to n-k+beta-1;
            likelihood=likelihood+log(i);
        end;
    end;
    else do;
        do i=1 to beta-1;
            likelihood=likelihood+log(i);
        end;
    end;
    do i=1 to alpha+beta-1;
        likelihood=likelihood+log(i);
    end;
    do i=1 to beta-1;
        likelihood=likelihood-log(i);
    end;
end;
else likelihood=0;
run;

proc print data=uniform&i;
run;

data model&i;
set model&i;
model=&i;
if n>0 then do;

```

```

likelihood=0;
do i=n to n+alpha+beta-1;
    likelihood=likelihood-log(i);
end;
if n ne k then do;
do i=n-k to n-k+beta-1;
    likelihood=likelihood+log(i);
end;
end;
else do;
do i=1 to beta-1;
likelihood=likelihood+log(i);
end;
end;
do i=1 to alpha+beta-1;
likelihood=likelihood+log(i);
end;
do i=1 to beta-1;
likelihood=likelihood-log(i);
end;
end;
else likelihood=0;
run;

proc print data=model&i;
run;

proc summary data=model&i;
var model likelihood;
output out=g&i mean(model)= sum(likelihood)=;
run;

%end;

```

```
%mend;

%identify;

data h;
set g1-g78;
run;

proc print data=h;
run;

proc sort data=h;
by likelihood;
run;

proc print data=h;
run;

data h;
set h;
by likelihood;
probability=EXP(likelihood);
dummy=1;
drop _TYPE_ _FREQ_ likelihood;
run;

proc sort data=h presorted out=hh;
by dummy;
run;

proc summary data=h;
```

```

var probability;
output out=bayes_denominator sum(probability)=denominator;
run;

data bayes_denominator;
set bayes_denominator;
dummy=1;
keep denominator dummy;
run;

proc sort data=bayes_denominator presorted out=bayes_denominator2;
by dummy;
run;

data posterior;
merge hh(in=aa) bayes_denominator2(keep=denominator dummy);
by dummy;
if aa;
posterior_probability=probability/denominator;
run;

proc print data=posterior;
run;

```

### B.3.3 Prior Probability of Bankruptcy

```

libname gdp '/home/wvu/deharris/new/state2';

/*options mprint symbolgen mlogic obs=1000000;*/

```

```

/*gather gvkey permno linkages*/
data link/view=link;
set crspa.ccmxpf_lnkused(keep=ugvkey ulinkdt apermno ulinkenddt
ulinktype where=(ulinktype in("LC","LU","LS"))
rename=(ugvkey=gvkey apermno=permno));
by permno;
if missing(ulinkenddt) then ulinkenddt=mdy(12,31,2013);
if missing(ulinkdt) then delete;
if missing(gvkey) then delete;
drop ulinktype;
run;

/*find boundary dates*/
proc summary data=crsp.dsf(where=(prc>0 and ^missing(prc)));
by permno;
var date;
output out=minimax min(date)=dmin max(date)=dmax;
run;

data crsp1/view=crsp1;
set minimax(drop=_type_ _freq_);
by permno;
run;

/*merge*/
data crsp2/view=crsp2;
merge crsp1(in=aa where=(dmax-dmin>730)) link;
by permno;
if aa;
run;

/*determine final disposition*/
data state_space/view=state_space;

```

```

set crsp.dseall(keep=permno dlstcd
shrcd dlamt nwperm where=(^missing(dlstcd)));
by permno;
operational=0;
merge=0;
cash_buyout=0;
private=0;
bankrupt=0;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
else if substr(left(shrcd),2,1)='5' then delete;
else if substr(left(shrcd),2,1)='8' then delete;
else if substr(left(shrcd),1,1)='2' then delete;
else if substr(left(shrcd),1,1)='4' then delete;
else if substr(left(shrcd),1,1)='7' then delete;
else do;
if dlstcd=100 then operational=1;
else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;
else if dlstcd in(502,510,513,514,516,517,519,520,573,575,580,581,582,583,584)
then
private=1;
else if dlstcd in(574) then bankrupt=1;
else if dlstcd
in(234,242,243,251,252,261,262,300,332,341,342,400,450,460,470,500) then do;
if dlstcd in(234,500) then do;
if nwperm>0 then merge=1;
else if nwperm=0 and dlamt>0 then cash_buyout=1;
else private=1;
end;
else if dlstcd in (242,243,251,341,342) then do;
if nwperm>0 then merge=1;
else cash_buyout=1;

```

```

end;
else if dlstcd in (252,261,262,450) then do;
    if nwperm>0 then merge=1;
    else if dlamt>0 then cash_buyout=1;
    else delete;
end;

end;

else if dlstcd in(300) then do;
    if nwperm>0 then merge=1;
    else delete;

end;
else if dlstcd in (332) then do;
    if nwperm>0 then merge=1;
    else private=1;

end;
else if dlstcd in (400,460,470) then do;
    if dlamt>0 then cash_buyout=1;
    else bankrupt=1;

end;

end;

bankrupt=private+bankrupt;
if bankrupt=1 then operational=0;
else operational=1;
drop shrcd nwperm dlamt dlstcd cash_buyout private merge;
run;

/*merge*/
data crsp3/view=crsp3;
merge crsp2(in=aa) state_space(in=bb);
by permno;
if aa and bb;

```

```

run;

/*find blank gvkey*/
data crsp4/view=crsp4;
set crsp3;
by permno;
test=length(gvkey);
run;

/*for items with missing gvkey*/
data nogvkey1/view=nogvkey1;
set crsp4(where=(test=1));
by permno;
drop gvkey test ulinkdt ulinkenddt;
run;

data nogvkey2;
set nogvkey1;
by permno;
do date=dmin to dmax;
if dmax-date>365 then operational=1;
else if bankrupt=1 then operational=0;
quarter=4*(year(date)-1929)+qtr(date)+1;
output;
end;
run;

proc sort data=nogvkey2 out=nogvkey3;
by quarter;
run;

data nogvkey4/view=nogvkey4;
merge nogvkey3 gdp.storage;

```

```

by quarter;
if missing(rank_ratio) then delete;
if missing(operational) then delete;
run;

proc sort data=nogvkey4 out=nogvkey5;
by permno date;
run;

/*items with gvkey*/
data gvkey1/view=gvkey1;
set crsp4(where=(test>1));
by permno;
drop test;
run;

/*sort by gvkey*/
proc sort data=gvkey1 out=gvkey2;
by gvkey;
run;

proc datasets library=work;
run;

/*gather acct data*/
data acct_data/view=acct_data;
set comp.fundq(keep=datadate gvkey SALEQ ATQ RECTQ ACTQ
LCTQ NIQ IBQ where=(~missing(date) and date>mdy(12,31,1961) and
date<mdy(7,1,1990))rename=(datadate=date));
by gvkey date;
if missing(NIQ) then NIQ=IBQ;
quarter=4*(year(date)-1929)+qtr(date)+1;

```

```

label quarter="Quarter";
run;

/*convert to ratios*/
data ratios1/view=ratios1;
/*create financial ratios*/
set acct_data;
by gvkey date;
/*Net Margin*/
if (^missing(niq) and ^missing(SALEQ) and (SALEQ NE 0)) then nm=niq/SALEQ;
    else nm=.;
/*Return on Assets*/
if (^missing(NIQ) and ^missing(ATQ) and (ATQ NE 0)) then roa=NIQ/ATQ;
else roa=.;
/*Current ratio*/
if (^missing(ACTQ) and ^missing(LCTQ) and (LCTQ NE 0)) then cr=ACTQ/LCTQ;
    else cr=.;
/*receivables conversion*/
if (^missing(RECTQ) and ^missing(SALEQ)
and (SALEQ NE 0)) then rec_conv=RECTQ/SALEQ;
    else rec_conv=.;
label    cr="Current ratio"
        nm="Net margin"
        rec_conv="Receivables converstion rate"
        roa="Return on assets"
;
drop ibq saleq NIQ ATQ RECTQ ACTQ LCTQ;
run;

/*sort by quarter to rank*/
proc sort data=ratios1 out=ratios2;
by quarter;

```

```

run;

/*rank data*/
proc rank data=ratios2 percent ties=mean out=ratios3;
by quarter;
var nm roa cr rec_conv;
ranks rnm rroa rcr rrec_conv;
run;

/*assings ratios to macro variables*/
%let a1=nm;
%let a2=roa;
%let a3=cr;
%let a4=rec_conv;

%macro tukey;

data ratios4/view=ratios4;
set ratios3;
by quarter;
%do i=1 %to 4;
if ^missing(r&&a&i) then do;
if r&&a&i>10 and r&&a&i<90 then &&a&i=0;
else if r&&a&i LE 10 then &&a&i=-1;
else &&a&i=1;
end;
%end;
drop r&a1 r&a2 r&a3 r&a4;
run;

data ratios5/view=ratios5;
merge ratios4(in=aa) gdp.storage;

```

```

by quarter;
if aa;
run;

%mend;
%tukey;

proc sort data=ratios5 out=ratios6;
by gvkey;
run;

/*join crsp and compustat data*/
data gvkey3/view=gvkey3;
merge gvkey2(in=aa) ratios6;
by gvkey;
if aa;
run;

/*delete out of sample data*/
data gvkey4/view=gvkey4;
set gvkey3;
by gvkey;
if date>dmax then delete;
if date<dmin then delete;
if date>ulinkenddt then delete;
if date<ulinkdt then delete;
if dmax>mdy(6,30,1990) then dmax=mdy(6,30,1990);
if dmin<mdy(1,1,1961) then dmin=mdy(1,1,1960);
run;

/*sort by permno quarter*/
proc sort data=gvkey4 out=gvkey5;

```

```

by permno quarter;
run;

/*split data to create date data*/
proc sort data=gvkey5(keep=permno dmin dmax
operational bankrupt) nodupkey out=gvkey_left1;
by permno;
run;

/*create calendar*/
data gvkey_left2;
set gvkey_left1;
by permno;
do date=dmin to dmax;
quarter=4*(year(date)-1929)+qtr(date);
output;
end;
run;

/*create registry that items are sorted*/
proc sort data=gvkey_left2 presorted out=gvkey_left3;
by permno quarter;
run;

/*create quarter specific data*/
data gvkey_right1/view=gvkey_right1;
set gvkey5(keep=permno quarter gvkey ulinkdt
ulinkenddt &a1 &a2 &a3 &a4 rank_ratio);
by permno quarter;
run;

proc sort data=gvkey_right1 presorted out=gvkey_right2;
by permno quarter;

```

```

run;

/*remerge data*/
data gvkey6/view=gvkey6;
merge gvkey_left3 gvkey_right2;
by permno quarter;
run;

/*only allow for 365 days of bankruptcy*/
data gvkey7/view=gvkey7;
set gvkey6;
by permno quarter;
if bankrupt=1 then do;
if dmax-date>365 then operational=1;
end;
drop bankrupt;
run;

data gvkey8/view=gvkey8;
set gvkey7(where=(^missing(dmax) and
^missing(rank_ratio)));
by permno quarter;
run;

data both1 nmroa1 arrec1 none1;
set gvkey8;
by permno quarter;
if ^missing(&a1) and ^missing(&a2) and
^missing(&a3) and ^missing(&a4) then output both1;
else if ^missing(&a1) and ^missing(&a2)
then output nmroa1;
else if ^missing(&a3) and ^missing(&a4)

```

```
then output arrec1;  
else output none1;  
run;
```

```
proc contents data=nmroa1;  
run;
```

```
proc contents data=arrec1;  
run;
```

```
proc contents data=none1;  
run;
```

```
proc contents data=both1;  
run;
```

```
proc contents data=nogvkey5;  
run;
```

```
proc freq data=both1 noprint;  
tables &a1*&a2*rank_ratio*operational/out=nm_model1;  
run;
```

```
proc freq data=both1 noprint;  
tables &a3*&a4*rank_ratio*operational/out=ar_model1;  
run;
```

```
proc freq data=arrec1 noprint;  
tables &a3*&a4*rank_ratio*operational/out=ar_model2;  
run;
```

```
proc freq data=nmroa1 noprint;
tables &a1*&a2*rank_ratio*operational/out=nm_model2;
run;

proc freq data=nogvkey5 noprint;
tables rank_ratio*operational/out=nogvkey6;
run;

proc print data=nogvkey6;
run;

proc freq data=none1 noprint;
tables rank_ratio*operational/out=none2;
run;

proc print data=none2;
run;

proc sql;
create table nm_model as
select a.&a1,a.&a2,a.rank_ratio,
a.operational, a.count+b.count
as totals
from nm_model1 as a, nm_model2 as b
where a.&a1=b.&a1 and a.&a2=b.&a2 and
a.rank_ratio=b.rank_ratio and a.operational=b.operational;
quit;

proc sql;
```

```

create table ar_model as
select a.&a3, a.&a4,a.rank_ratio, a.operational,
a.count+b.count as totals
from ar_model1 as a, ar_model2 as b
where a.&a3=b.&a3 and a.&a4=b.&a4 and
a.rank_ratio=b.rank_ratio and a.operational=b.operational;
quit;

```

```

proc sql;
create table gdp_bankruptcy as
select a.rank_ratio, a.operational,
a.count, a.count+b.count as totals
from nogvkey6 as a left join none2 as b
on a.rank_ratio=b.rank_ratio
and a.operational=b.operational;
quit;

```

```

data gdp_bankruptcy;
set gdp_bankruptcy;
if missing(totals) then totals=count;
drop count;
run;

```

```

proc print data=gdp_bankruptcy;
run;

```

## B.4 Prior Probability of Merger

```

/*options mprint symbolgen mlogic obs=1000000;*/

```

```

/*gather acct data*/

```

```

data acct_data/view=acct_data;
set comp.fundq(keep=datadate gvkey ceqq ATQ NIQ IBQ
where=(^missing(date)
and date>mdy(12,31,1961) and
date<mdy(7,1,1990))rename=(datadate=date));
by gvkey date;
if missing(NIQ) then NIQ=IBQ;
quarter=4*(year(date)-1929)+qtr(date)+1;
label quarter="Quarter";
drop ibq;
run;

/*convert to ratios*/
data ratios1/view=ratios1;
/*create financial ratios*/
set acct_data;
by gvkey date;
/*Return on Assets*/
if (^missing(NIQ) and ^missing(ATQ) and (ATQ NE 0))
then roa=NIQ/ATQ;
else roa=.;
/*Debt to equity*/
if (^missing(atq) and ^missing(ceqq) and (ceqq NE
0)) then
de=atq/ceqq-1;
      else de=.;
label    de="Debt to Equity"
        roa="Return on Assets"
;
keep gvkey date roa de ceqq quarter;
run;

data ratios1a/view=ratios1a;

```

```

set ratios1(where=(^missing(roa) and ^missing(de) and
^missing(ceqq) and ^missing(date)));
run;

/*sort by quarter to rank*/
proc sort data=ratios1a out=ratios2;
by quarter;
run;

/*rank data*/
proc rank data=ratios2 percent ties=mean out=ratios3;
by quarter;
var roa ceqq de;
ranks rroa rceqq rde;
run;

/*assings ratios to macro variables*/
%let a1=roa;
%let a2=ceqq;
%let a3=de;

%macro tukey;

data ratios4/view=ratios4;
set ratios3;
by quarter;
%do i=1 %to 3;
if ^missing(r&&a&i) then do;
if r&&a&i>10 and r&&a&i<90 then &&a&i=0;
else if r&&a&i LE 10 then &&a&i=-1;
else &&a&i=1;

```

```
end;
%end;
drop r&a1 r&a2 r&a3;
run;

%mend;
%tukey;

proc sort data=ratios4 out=ratios5;
by gvkey descending date;
run;

data ratios6/view=ratios6;
set ratios5;
by gvkey descending date;
lgvkey=lag(gvkey);
ldate=lag(date);
run;

proc sort data=ratios6 out=ratios7;
by gvkey date;
run;

data ratios8;
set ratios7(where=(ldate-date<367 or
missing(ldate)));
by gvkey date;
k2=gvkey;
k3=&a1;
k4=&a2;
k5=&a3;
label k1="Date"
```

```

k2="GVKEY"
k3="Return on Assets"
k4="Total Equity"
k5="Debt to Equity"
;
if gvkey NE lgvkey then do;
do k1=date to date+90;
output;
end;
end;
else do;
do k1=date to ldate;
output;
end;
end;
keep k1-k5;
run;

proc datasets library=work;
delete ratios2 ratios3 ratios5 ratios7;
modify ratios8;
rename k1=date k2=gvkey k3=roa k4=teq k5=de;
run;

proc sort data=ratios8 presorted out=ratios8a;
by gvkey date;

/*gather gvkey permno linkages*/
data link1;
set crspa.ccmxpf_lnkused(keep=ugvkey ulinkdt apermno
  ulinkenddt ulinktype where=(ulinktype in("LC","LU",
  "LS"))) rename=(ugvkey=gvkey apermno=permno));
by permno;

```

```

if missing(ulinkenddt) then ulinkenddt=mdy(12,31,2013);
if missing(ulinkdt) then delete;
else if missing(gvkey) then delete;
else if ulinkenddt<mdy(1,1,1962) then delete;
else if ulinkdt>mdy(6,30,1990) then delete;
else do;
if ulinkenddt>mdy(6,30,1990) then ulinkenddt=mdy(6,30,1990);
if ulinkdt<mdy(1,1,1962) then ulinkdt=mdy(1,1,1962);
do date=ulinkdt to ulinkenddt;
output;
end;
end;
drop ulinktype ulinkdt ulinkenddt;
run;

```

```

proc sort data=link1 out=link2;
by permno date;
run;

```

```

data link3/view=link3;
set link2;
by permno date;
lpermno=lag(permno);
ldate=lag(date);
run;

```

```

data link4/view=link4;
set link3;
if permno=lpermno and date=ldate then delete;
drop ldate lpermno;
run;

```

```

proc sort data=link4 out=link5;
by gvkey date;
run;

proc datasets library=work;
delete link2;
run;

data link6/view=link6;
merge ratios8a(in=aa) link5(in=bb);
by gvkey date;
if aa and bb;
run;

proc sort data=link6 out=link7;
by permno;
run;

proc datasets library=work;
delete link5;
run;

/*determine final disposition*/
data state_space1/view=state_space1;
set crsp.dseall(keep=permno dlstcd shrcd dlamt
nwperm where=(^missing(dlstcd)));
by permno;
operational=0;
merge=0;
cash_buyout=0;
private=0;
bankrupt=0;

```

```

if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
else if substr(left(shrcd),2,1)='5' then delete;
else if substr(left(shrcd),2,1)='8' then delete;
else if substr(left(shrcd),1,1)='2' then delete;
else if substr(left(shrcd),1,1)='4' then delete;
else if substr(left(shrcd),1,1)='7' then delete;
else do;
if dlstcd=100 then operational=1;
else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;
else if dlstcd in(502,510,513,514,516,517,519,520,573,575,580,
581,582,583,584) then
private=1;
else if dlstcd in(574) then bankrupt=1;
      else if dlstcd
in(234,242,243,251,252,261,262,300,332,341,342,400,450,460,470,500) then do;
          if dlstcd in(234,500) then do;
              if nwperm>0 then merge=1;
              else if nwperm=0 and dlamt>0 then cash_buyout=1;
              else private=1;
          end;
          else if dlstcd in (242,243,251,341,342) then do;
              if nwperm>0 then merge=1;
              else cash_buyout=1;
          end;
          else if dlstcd in (252,261,262,450) then do;
              if nwperm>0 then merge=1;
              else if dlamt>0 then cash_buyout=1;
              else delete;
          end;
      end;
end;

```

```

end;
else if dlstcd in(300) then do;
            if nwperm>0 then merge=1;
            else delete;
        end;
    else if dlstcd in (332) then do;
            if nwperm>0 then merge=1;
            else private=1;
        end;
    else if dlstcd in (400,460,470) then do;
            if dlamt>0 then cash_buyout=1;
            else bankrupt=1;
        end;
    end;
end;
merge=merge+cash_buyout;
if merge=1 then operational=0;
else operational=1;
drop shrcd nwperm dlamt dlstcd cash_buyout private bankrupt merge;
run;

data state_space2;
merge link7(in=aa) state_space1(in=bb);
by permno;
if aa and bb;
drop gvkey;
run;

proc sort data=state_space2 out=state_space3;
by permno date;
run;

```

```

data sic1/view=sic1;
set crsp.dsf(keep=hsiccd permno date rename=(hsiccd=sic));
by permno date;
if sic>1999 and sic<4000 then category=4;
else if sic>5999 and sic<6800 then category=8;
else if sic GE 6800 and sic<9000 then category=9;
else if sic GE 4000 and sic<5000 then category=5;
else if sic >999 and sic<1500 then category=2;
else if sic GE 5200 and sic<6000 then category=7;
else if sic GE 5000 and sic<5200 then category=6;
else if sic ge 1500 and sic<1800 then category=3;
else if sic GE 100 and sic LE 999 then category=1;
else if sic ge 9000 and sic< 9730 then category=10;
else category=11;
run;

```

```

data sic2/view=sic2;
set sic1(where=(^missing(sic) and date>mdy(12,31,1961)
and date<mdy(7,1,1990)));
by permno date;
if category=8 then group="F";
else group="N";
drop category;
run;

```

```

data sic3/view=sic3;
merge state_space3(in=aa) sic2(in=bb);
by permno date;
if aa and bb;
run;

```

```

proc summary data=sic3;
by permno;

```

```
var date;
output out=sic3a min(date)=dmin max(date)=dmax;
run;
```

```
data sic4;
set sic3a(drop=_FREQ_ _TYPE_);
by permno;
do date=dmin to dmax;
output;
end;
run;
```

```
proc sort data=sic4 presorted out=sic5;
by permno date;
run;
```

```
%macro weekends;
```

```
data sic6/view=sic6;
merge sic3 sic5;
by permno date;
run;
```

```
data sic7/view=sic7;
set sic6;
by permno date;
%do i=1 %to 7;
lde&i=lag&i(de);
lgroup&i=lag&i(group);
loperational&i=lag&i(operational);
lpermno&i=lag&i(permno);
lroa&i=lag&i(roa);
lsic&i=lag&i(sic);
```

```

lteq&i=lag&i(teq);
%end;
run;

data sic8/view=sic8;
set sic7;
by permno date;
if missing(sic) then do;
test=0;
%do i=1 %to 7;
if test=0 and permno=lpermno&i and
^missing(lsic&i) then do;
de=lde&i;
group=lgroup&i;
operational=loperational&i;
roa=lroa&i;
sic=lsic&i;
teq=lteq&i;
test=1;
end;
%end;
end;
run;

data sic9/view=sic9;
set sic8(where=(^missing(sic)))keep= permno date
de group operational roa sic teq);
by permno date;
test=1;
run;

%mend;
%weekends;

```

```

proc summary data=crsp.dsf(where=(prc>0 and ^missing(prc)));
by permno;
var date;
output out=minimax1 min(date)=dmin max(date)=dmax;
run;

data minimax2/view=minimax2;
set minimax1(drop=_FREQ_ _TYPE_ where=(dmax-dmin>730));
by permno;
run;

data sic10/view=sic10;
merge sic9 minimax2(in=bb keep=permno);
by permno;
if bb;
run;

proc sort data=sic10 out=sic11;
by permno date;
run;

data frame1;
set minimax2 (where=(dmax>mdy(1,1,1962) and dmin<mdy(6,30,1990)));
by permno;
if dmin<mdy(1,1,1962) then dmin=mdy(1,1,1962);
if dmax>mdy(6,30,1990) then dmax=mdy(6,30,1990);
do date=dmin to dmax;
output;
end;
drop dmin dmax;
run;

```

```
proc sort data=frame1 out=frame2 presorted;
by permno date;
run;

proc datasets library=work;
delete frame1;
run;

data frame3/view=frame3;
merge sic11 frame2;
by permno date;
run;

data frame4a/view=frame4a;
set frame3(where=(missing(operational)));
by permno;
run;

data frame4b/view=frame4b;
merge frame4a(drop=operational in=aa) state_space1(in=bb);
by permno;
if aa and bb;
run;

data frame4c/view=frame4c;
set frame3(where=(^missing(operational)));
by permno;
run;

data frame4d/view=frame4d;
set frame4b frame4c;
run;
```

```
proc sort data=frame4d out=frame5;
by permno;
run;

data frame6/view=frame6;
merge frame5(in=aa) minimax2;
by permno;
if aa;
run;

data frame7/view=frame7;
set frame6(where=(^missing(date)));
by permno;
if dmax-date>365 and operational=0 then operational=1;
run;

proc freq data=frame7(where=(missing(sic))) noprint;
tables operational/out=no_data1;
run;

proc freq data=frame7(where=(^missing(sic))) noprint;
tables group*de*roa*teq*operational/out=with_data1;
run;

proc print data=no_data1;
run;

proc print data=with_data1;
run;

data no_data2;
set no_data1(drop=percent);
retain alpha;
```

```
if operational=0 then alpha=count;
else do;
percent=alpha/(alpha+count);
output;
end;
run;

proc print data=no_data2;
run;

data with_data2;
do i=1 to 2;
if i=1 then group="F";
else group="N";
do de=-1 to 1;
do roa=-1 to 1;
do teq=-1 to 1;
do operational=0 to 1;
output;
end;
end;
end;
end;
end;
run;

proc sort data=with_data2 out=with_data3 presorted;
by group de roa teq operational;
run;

data with_data4/view=with_data4;
merge with_data3 with_data1;
by group de roa teq operational;
```

```
run;

data with_data5/view=with_data5;
set with_data4;
if missing(count) then count=0;
run;

data with_data6/view=with_data6;
set with_data5;
retain alpha 0;
if operational=0 then alpha=count;
if operational=1 then
merger_expectation=(alpha+1)/(alpha+count+16);
run;

proc print data=with_data6;

data with_data7/view=with_data7;
set with_data6;
if operational=0 then delete;
run;

proc print data=with_data7;
run;

proc export data=with_data7
outfile='/home/wvu/deharris/new/state2/merge_counts.csv'
dbms=csv
replace;
run;
```

## B.5 Prior Probability for Returns Given a Merger

```

/*options mprint symbolgen mlogic;*/
/*options obs=1000000;*/

/*gathers all price data, in raw form, and calculates maximum and
minimum values of dates*/
proc sql;
create view gather as
select permno,date,prc,cfacpr,min(date) as dmin, max(date)
as dmax
from crsp.dsf
where prc>0 and ^missing(prc)
group by permno
order by permno,date;
quit;

/*restricts data to common shares only and assigns final status
code*/
data common_only/view=common_only;
set crsp.dseall(keep=permno dlstcd shrcd dlamt nwperm
where=(^missing(dlstcd)));
by permno;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
      else if substr(left(shrcd),2,1)='5' then delete;
      else if substr(left(shrcd),2,1)='8' then delete;
      else if substr(left(shrcd),1,1)='2' then delete;
      else if substr(left(shrcd),1,1)='4' then delete;
      else if substr(left(shrcd),1,1)='7' then delete;
else do;
operational=0;
merge=0;

```

```

cash_buyout=0;
private=0;
bankrupt=0;
if dlstcd=100 then operational=1;
    else if dlstcd in(200,231,232,241,244,301,331,343)
        then merge=1;
    else if dlstcd in(233,235,271,333,334,361)
        then cash_buyout=1;
    else if dlstcd in(502,510,513,514,516,517,519,
520,573,575,580,581,582,583,584) then private=1;
    else if dlstcd in(574) then bankrupt=1;
    else if dlstcd in(234,242,243,251,252,261,262,
300,332,341,342,400,450,460,470,500) then do;
        if dlstcd in(234,500) then do;
            if nwperm>0 then merge=1;
            else if nwperm=0 and dlamt>0 then cash_buyout=1;
            else private=1;
        end;
    else if dlstcd in (242,243,251,341,342) then do;
        if nwperm>0 then merge=1;
        else cash_buyout=1;
    end;
    else if dlstcd in (252,261,262,450) then do;
        if nwperm>0 then merge=1;
        else if dlamt>0 then cash_buyout=1;
        else delete;
    end;
    else if dlstcd in(300) then do;
        if nwperm>0 then merge=1;
        else delete;
    end;
    else if dlstcd in (332) then do;
        if nwperm>0 then merge=1;

```

```

                else private=1;
            end;
        else if dlstcd in (400,460,470) then do;
            if dlamt>0 then cash_buyout=1;
            else bankrupt=1;
        end;
    end;
end;
run;

/*links price data with list of common data and final states*/
proc sql;
create view link1 as
select a.*,b.merge,b.cash_buyout,b.dlamt,b.nwperm
from gather as a, common_only as b
where a.permno=b.permno
order by permno;
quit;

/*updates status to so that final status only happens in
final trading year*/
data link2/view=link2;
set link1;
if cash_buyout=1 or merge=1 then do;
if dmax-date>365 then delete;
end;
run;

data cash_return/view=cash_return;
set link2(where=(cash_buyout=1 and date<mdy(7,1,1990)));
return=dlamt/prc;
keep return;
run;

```

```
proc sort data=link2(where=(dmax-dmin>730))
nodupkey out=merge1;
by permno;
run;

data merge2/view=merge2;
set merge1;
old_permno=permno;
permno=nwperm;
low=dmax+1;
high=dmax+365;
keep permno old_permno low high;
run;

proc sort data=merge2 out=merge3;
by permno;
run;

data merge4/view=merge4;
merge merge3(in=aa) gather(drop=dmax);
by permno;
if aa;
run;

data merge5/view=merge5;
set merge4(where=(date le high and date ge low));
permno=old_permno;
drop old_permno high low;
run;

data merge6/view=merge6;
```

```
set merge5 link2(where=(merge=1));
prc=prc/cfacpr;
drop cfacpr;
run;

proc sort data=merge6 out=merge7;
by permno date;
run;

data merge8/view=merge8;
set merge7;
by permno date;
ldate=lag(date);
lprc=lag(prc);
run;

data merge9/view=merge9;
merge merge8 gather(keep=permno dmax);
by permno;
run;

data merge10/view=merge10;
set merge9;
by permno;
if ldate=dmax then test=1;
keep permno date prc test lprc;
run;
```

```
data merge11/view=merge11;
set merge10(where=(test=1));
by permno;
if date-ldate<8 then converstion_factor=lprc/prc;
else delete;
run;
```

```
data merge12/view=merge12;
merge merge7 merge11(in=aa);
by permno;
if aa;
run;
```

```
data merge13/view=merge13;
set merge12;
by permno;
if date>dmax then prc=prc*converstion_factor;
keep permno date prc;
run;
```

```
proc sort data=merge13 nodupkey out=merge14;
by permno date;
run;
```

```
proc summary data=merge14;
by permno;
var date;
output out=frame1 min(date)=dmin max(date)=dmax;
```

```
run;

data frame2;
set frame1(keep=permno dmin dmax);
by permno;
do date=dmin to dmax;
output;
end;
drop dmin dmax;
run;

proc sort data=frame2 presorted out=frame3;
by permno date;
run;

proc datasets library=work;
delete frame1;
run;

data frame4;
merge frame3 merge14;
by permno date;
run;

proc datasets library=work;
delete frame2;
run;

%macro lead;
```

```

proc sort data=frame4 out=lead0;
by permno descending date;
run;

data lead1/view=lead1;
set lead0;
by permno descending date;
%do n=365 %to 352 %by -1;
lprc&n=lag&n(prc);
%if %eval(&n)=365 %then %do;
lpermno&n=lag&n(permno);
%end;
%end;
run;

proc sort data=lead1 out=lead2;
by permno date;
run;

proc datasets library=work;
run;

/*Note that the closing transaction can only happen
on day 365, as such, returns are not normalized
to reflect the true number of days between the
hypothetical transaction*/

data lead3/view=lead3;
set lead2(where=(^missing(prc)));
by permno date;
cfacpr=1;
%do n=352 %to 365;

```

```

lcfacpr&n=1;
%end;
if permno=lpermno365 then do;
if ^missing(lprc365) then return=lprc365/prc*cfacpr/lcfacpr365;
else if ^missing(lprc364) then return=lprc364/prc*cfacpr/lcfacpr364;
else if ^missing(lprc363) then return=lprc363/prc*cfacpr/lcfacpr363;
else if ^missing(lprc362) then return=lprc362/prc*cfacpr/lcfacpr362;
else if ^missing(lprc361) then return=lprc361/prc*cfacpr/lcfacpr361;
else if ^missing(lprc360) then return=lprc360/prc*cfacpr/lcfacpr360;
else if ^missing(lprc359) then return=lprc359/prc*cfacpr/lcfacpr359;
else if ^missing(lprc358) then return=lprc358/prc*cfacpr/lcfacpr358;
else if ^missing(lprc357) then return=lprc357/prc*cfacpr/lcfacpr357;
else if ^missing(lprc356) then return=lprc356/prc*cfacpr/lcfacpr356;
else if ^missing(lprc355) then return=lprc355/prc*cfacpr/lcfacpr355;
else if ^missing(lprc354) then return=lprc354/prc*cfacpr/lcfacpr354;
else if ^missing(lprc353) then return=lprc353/prc*cfacpr/lcfacpr353;
else if ^missing(lprc352) then return=lprc352/prc*cfacpr/lcfacpr352;
else return=.;
end;
run;

data mreturn1/view=mreturn1;
set lead3(keep=return date where=(^missing(return) and
date<mdy(7,1,1990))) cash_return;
do i=.01 to 1 by .01;
if return<i and return=>i-.01 then return_class=i-.005;
end;
if return=>1 then return_class=1.5;
drop return;
run;

%mend;

```

```

%lead;

proc freq data=mreturn1;
tables return_class/out=distribution;
run;

proc print data=distribution;
run;

proc export data=work.distribution
outfile="/home/wvu/deharris/new/state2/merger_prior.csv"
dbms=csv
replace;
run;

```

## B.6 Prior Probability for $\sigma$ for a Going Concern

```

/*options mprint symbolgen mlogic;*/
/*options obs=10000;*/

/*gathers all price data, in raw form, and calculates maximum and
minimum values of dates*/
proc sql;
create view gather as
select permno,date,prc,cfacpr,min(date) as dmin, max(date)
as dmax
from crsp.dsf
where prc>0 and ^missing(prc)
group by permno

```

```

order by permno,date;
quit;

/*restricts data to common shares only and assigns final status
code*/
data common_only/view=common_only;
set crsp.dseall(keep=permno dlstcd shrcd dlamt nwperm
where=(^missing(dlstcd)));
by permno;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
      else if substr(left(shrcd),2,1)='5' then delete;
      else if substr(left(shrcd),2,1)='8' then delete;
      else if substr(left(shrcd),1,1)='2' then delete;
      else if substr(left(shrcd),1,1)='4' then delete;
      else if substr(left(shrcd),1,1)='7' then delete;
else do;
operational=0;
merge=0;
cash_buyout=0;
private=0;
bankrupt=0;
if dlstcd=100 then operational=1;
      else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
      else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;
      else if dlstcd in(502,510,513,514,516,517,519,520,573,575,580,
581,582,583,584) then private=1;
      else if dlstcd in(574) then bankrupt=1;
      else if dlstcd in(234,242,243,251,252,261,262,300,332,341,342,
400,450,460,470,500) then do;
            if dlstcd in(234,500) then do;
                  if nwperm>0 then merge=1;
                  else if nwperm=0 and dlamt>0 then cash_buyout=1;

```

```

        else private=1;
    end;
else if dlstcd in (242,243,251,341,342) then do;
    if nwperm>0 then merge=1;
    else cash_buyout=1;
end;
else if dlstcd in (252,261,262,450) then do;
    if nwperm>0 then merge=1;
    else if dlamt>0 then cash_buyout=1;
    else delete;
end;
else if dlstcd in(300) then do;
    if nwperm>0 then merge=1;
    else delete;
end;
else if dlstcd in (332) then do;
    if nwperm>0 then merge=1;
    else private=1;
end;
else if dlstcd in (400,460,470) then do;
    if dlamt>0 then cash_buyout=1;
    else bankrupt=1;
end;
end;
end;
run;

/*links price data with list of common data and final states*/
proc sql;
create view link1 as
select a.*,b.operational,b.merge,b.cash_buyout,b.private,b.bankrupt
from gather as a, common_only as b
where a.permno=b.permno

```

```

order by permno;
quit;

/*updates status to so that final status only happens in
final trading year*/
data link2/view=link2;
set link1;
if operational=0 then do;
if dmax-date>365 then do;
operational=1;
bankrupt=0;
merge=0;
private=0;
cash_buyout=0;
end;
end;
bankrupt=bankrupt+private;
merge=merge+cash_buyout;
drop private cash_buyout;
run;

/*creates a framework of permno-dates that include
weekends and holidays*/
data frame1/view=frame1;
set gather(keep=permno dmin dmax);
by permno;
run;

/*removes duplicate permno*/
data frame2/view=frame2;
set frame1;
by permno;
lpermno=lag(permno);

```

```

run;

/*creates a calendar of dates*/
data frame3;
set frame2(where=(permno NE lpermno and dmax-dmin>730));
by permno;
do date=dmin to dmax;
output;
end;
run;

/*splits the data into operational and merger.  bankruptcy
is not needed as it goes to zero*/
data operational1/view=operational1;
set link2(where=(operational=1));
by permno;
run;

data merge1/view=merge1;
set link2(where=(merge=1));
by permno;
run;

proc sql;
create view operational2 as
select a.*,b.prc,b.cfacpr
from frame3 as a
left join
operational1 as b
on a.permno=b.permno and a.date=b.date;
quit;

```

```
proc sort data=operational2 out=operational3;
by permno descending date;
run;
```

```
%macro lead;
```

```
data operational4/view=operational4;
set operational3;
by permno descending date;
%do n=365 %to 352 %by -1;
lprc&n=lag&n(prc);
lcfacpr&n=lag&n(cfacpr);
%if %eval(&n)=365 %then %do;
lpermno&n=lag&n(permno);
%end;
%end;
run;
```

```
proc sort data=operational4 out=operational5;
by permno date;
run;
```

```
proc datasets library=work;
run;
```

```
/*Note that the closing transaction can only happen on day 365,
as such, returns are not normalized
to reflect the true number of days between the hypothetical
transaction*/
```

```
data operational6/view=operational6;
```

```

set operational5(where=(^missing(prc)));
by permno date;
if permno=lpermno365 then do;
if ^missing(lprc365) then return=lprc365/prc*cfacpr/lcfacpr365;
else if ^missing(lprc364) then return=lprc364/prc*cfacpr/lcfacpr364;
else if ^missing(lprc363) then return=lprc363/prc*cfacpr/lcfacpr363;
else if ^missing(lprc362) then return=lprc362/prc*cfacpr/lcfacpr362;
else if ^missing(lprc361) then return=lprc361/prc*cfacpr/lcfacpr361;
else if ^missing(lprc360) then return=lprc360/prc*cfacpr/lcfacpr360;
else if ^missing(lprc359) then return=lprc359/prc*cfacpr/lcfacpr359;
else if ^missing(lprc358) then return=lprc358/prc*cfacpr/lcfacpr358;
else if ^missing(lprc357) then return=lprc357/prc*cfacpr/lcfacpr357;
else if ^missing(lprc356) then return=lprc356/prc*cfacpr/lcfacpr356;
else if ^missing(lprc355) then return=lprc355/prc*cfacpr/lcfacpr355;
else if ^missing(lprc354) then return=lprc354/prc*cfacpr/lcfacpr354;
else if ^missing(lprc353) then return=lprc353/prc*cfacpr/lcfacpr353;
else if ^missing(lprc352) then return=lprc352/prc*cfacpr/lcfacpr352;
else return=.;
end;
run;

data operational7;
set operational6(keep=return date where=(^missing(return) and
date<mdy(7,1,1990)));
drop date;
run;

%mend;
%lead;

proc iml;

```

```

start log_likelihood(values,g);
ll_likelihood=J(nrow(values),1,0);
do i=1 to nrow(values);
if i<=nrow(values) then ll_likelihood[i,1]=log(g[1,1]/(g[1,1]#g[1,1]+
(values[i,1]-1)##2))-log(1/2+atan(1/g[1,1])/constant('pi'));
end;
total=ll_likelihood[+,];
return(total);
finish log_likelihood;

```

```

start bounds_by_granularity(lower,upper,level,return);
scale=do(lower,upper,level)';
l_likelihood=scale;
z=0;
do i=lower to upper by level;
z=z+1;
if z <=nrow(scale) then l_likelihood[z,1]=
log_likelihood(return,scale[z,1]);
end;
print scale l_likelihood;
boundary=max(l_likelihood)-log(1000000);
index_value=scale;
do i=1 to nrow(l_likelihood);
if boundary>l_likelihood[i,1] then index_value[i,1]=0;
else index_value[i,1]=1;
end;
test=0;
do i=1 to nrow(index_value);
if test=0 then do;
if index_value[i,1]=1 then do;
test=test+1;
lower=scale[i,1]-level/10*9;
end;
end;

```

```

end;
else if test=1 then do;
if index_value[i,1]=0 then do;
test=test+1;
upper=scale[i-1,1]+level/10*9;
end;
end;
end;
finish bounds_by_granularity;

start final_bounds(lower,upper,level,return);
    scale=do(lower,upper,level)';
    l_likelihood=scale;
    z=0;
    do i=lower to upper by level;
        z=z+1;
        if z <=nrow(scale) then l_likelihood[z,1]=
            log_likelihood(return,scale[z,1]);
    end;
    print scale l_likelihood;
    boundary=max(l_likelihood)-log(1000000);
    index_value=scale;
    do i=1 to nrow(index_value);
if boundary>l_likelihood[i,1] then index_value[i,1]=0;
        else index_value[i,1]=1;
    end;
    test=0;
do i=1 to nrow(index_value);
if test=0 then do;
if index_value[i,1]=1 then do;
test=test+1;
lower=scale[i,1];
end;
end;

```

```

end;
else if test=1 then do;
if index_value[i,1]=0 then do;
test=test+1;
upper=scale[i,1];
end;
end;
end;
finish final_bounds;

use operational7;
read all var{return} into return;
close operational7;

level=.1;
lower=.2;
upper=.5;

do granularity=1 to 3;
if granularity<=3 then do;
level=level/10;
call bounds_by_granularity(lower,upper,level,return);
end;
end;

level=level/10;
call final_bounds(lower,upper,level,return);

sigma=do(lower,upper,level)';
likelihood=sigma;
zz=0;

```

```
do i=lower to upper by level;
zz=zz+1;
    if zz <=nrow(scale) then likelihood[zz,1]=
    log_likelihood(return,sigma[zz,1]);
end;

print sigma likelihood;

create posterior var{sigma likelihood};
append;
close posterior;

quit;

data posterior;
set posterior;
dummy=1;
run;

proc summary data=posterior;
var likelihood;
output out=mle max(likelihood)=max;
run;

data mle;
set mle;
dummy=1;
run;

data transform;
merge posterior mle;
by dummy;
likelihood=likelihood-max;
```

```
probability=exp(likelihood);
drop _TYPE_ _FREQ_ max;
run;

proc summary data=transform;
var probability;
output out=const sum(probability)=sum;
run;

data const;
set const;
dummy=1;
drop _TYPE_ _FREQ_;
run;

data prob;
merge transform const;
by dummy;
probability=probability/sum;
run;

proc print prob;

proc export data=prob
    outfile='/home/wvu/deharris/new/state2/going_return_posterior.csv'
    dbms=csv
    replace;
run;
```

## B.7 Posterior Probability for Bankruptcy Estimator

```

libname gdp '/home/wvu/deharris/new/state3';
libname bankrupt '/home/wvu/deharris/new/state3';

/*options mprint symbolgen mlogic obs=100000;*/

/*gathers all price data, in raw form, and calculates maximum and
minimum values of dates*/
proc sql;
create view gather1 as
select permno,min(date) as dmin, max(date) as dmax
from crsp.dsf
where prc>0 and ^missing(prc)
group by permno;
quit;

data gather2/view=gather2;
set gather1(where=(dmax-dmin>730));
by permno;
run;

/*determine final disposition*/
data state_space/view=state_space;
set crsp.dseall(keep=permno dlstcd shrcd dlamt nwperm
where=(^missing(dlstcd)));
by permno;
operational=0;
merge=0;
cash_buyout=0;

```

```

private=0;
bankrupt=0;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
else if substr(left(shrcd),2,1)='5' then delete;
else if substr(left(shrcd),2,1)='8' then delete;
else if substr(left(shrcd),1,1)='2' then delete;
else if substr(left(shrcd),1,1)='4' then delete;
else if substr(left(shrcd),1,1)='7' then delete;
else do;
if dlstcd=100 then operational=1;
else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;
else if dlstcd in(502,510,513,514,516,517,519,520,573,575,580,581,
582,583,584) then
private=1;
else if dlstcd in(574) then bankrupt=1;
      else if dlstcd
in(234,242,243,251,252,261,262,300,332,341,342,400,450,460,470,500) then do;
          if dlstcd in(234,500) then do;
              if nwperm>0 then merge=1;
              else if nwperm=0 and dlamt>0 then cash_buyout=1;
              else private=1;
          end;
          else if dlstcd in (242,243,251,341,342) then do;
              if nwperm>0 then merge=1;
              else cash_buyout=1;
          end;
          else if dlstcd in (252,261,262,450) then do;
              if nwperm>0 then merge=1;
              else if dlamt>0 then cash_buyout=1;
              else delete;
          end;
      end;
end;

```

```

end;
else if dlstcd in(300) then do;
    if nwperm>0 then merge=1;
    else delete;
end;
else if dlstcd in (332) then do;
    if nwperm>0 then merge=1;
    else private=1;
end;
else if dlstcd in (400,460,470) then do;
    if dlamt>0 then cash_buyout=1;
    else bankrupt=1;
end;
end;
end;
bankrupt=private+bankrupt;
if bankrupt=1 then operational=0;
else operational=1;
drop shrcd nwperm dlamt dlstcd cash_buyout private merge bankrupt;
run;

data gather3/view=gather3;
merge gather2(in=aa) state_space(in=bb);
by permno;
if aa and bb;
run;

/*gather gvkey permno linkages*/
data link/view=link;
set crspa.ccmxpf_lnkused(keep=ugvkey ulinkdt apermno ulinkenddt
ulinktype where=(ulinktype in("LC","LU","LS"))) rename=(ugvkey=

```

```

gvkey apermno=permno));
by permno;
if missing(ulinkenddt) then ulinkenddt=mdy(12,31,2013);
if missing(ulinkdt) then delete;
else if missing(gvkey) then delete;
else if length(gvkey)=1 then delete;
drop ulinktype;
run;

proc sql;
create view gather4 as
select a.*,b.gvkey,b.ulinkdt,b.ulinkenddt
from gather3 as a
left join
link as b
on a.permno=b.permno
order by permno;
quit;

data gather5/view=gather5;
set gather4(where=(^missing(dmin) and ^missing(dmax)));
by permno;
run;

/*splits the data into groups with accounting and groups
without accounting data*/
data no_accounting_data1 accounting_data1;
set gather5;
by permno;
do date=dmin to dmax;
if missing(ulinkdt) then do;
quarter=4*(year(date)-1929)+qtr(date);
output no_accounting_data1;

```

```

end;
else do;
if date GE ulinkdt and date LE ulinkenddt then
do;
quarter=4*(year(date)-1929)+qtr(date);
output accounting_data1;
end;
else delete;
end;
end;
end;
run;

/*drops unused fields*/
data no_accounting_data2/view=no_accounting_data2;
set no_accounting_data1(drop=gvkey ulinkdt ulinkenddt);
by permno;
run;

data accounting_data2/view=accounting_data2;
set accounting_data1(drop=ulinkdt ulinkenddt);
by permno;
run;

proc sort data=accounting_data2 out=accounting_data3;
by gvkey quarter;
run;

proc datasets library=work;
modify accounting_data3;
index create gvkey;
delete accounting_data1;
run;

```

```

/*This is the end of the permno based gathering and
combining of data*/

/*gather acct data*/
data acct_data/view=acct_data;
set comp.fundq(keep=datadate gvkey SALEQ ATQ RECTQ ACTQ
LCTQ NIQ IBQ where=(^missing(date))rename=
(datadate=date));
by gvkey date;
if missing(NIQ) then NIQ=IBQ;
quarter=4*(year(date)-1929)+qtr(date);
label quarter="Quarter";
run;

/*convert to ratios*/
data ratios1/view=ratios1;
/*create financial ratios*/
set acct_data;
by gvkey date;
/*Net Margin*/
if (^missing(niq) and ^missing(SALEQ) and (SALEQ NE 0))
then nm=niq/SALEQ;
else nm=.;
/*Return on Assets*/
if (^missing(NIQ) and ^missing(ATQ) and (ATQ NE 0))
then roa=NIQ/ATQ;
else roa=.;
/*Current ratio*/
if (^missing(ACTQ) and ^missing(LCTQ) and (LCTQ NE 0))

```

```

then cr=ACTQ/LCTQ;
      else cr=.;
/*receivables conversion*/
if (^missing(RECTQ) and ^missing(SALEQ) and (SALEQ NE 0))
then rec_conv=RECTQ/SALEQ;
      else rec_conv=.;
label   cr="Current ratio"
        nm="Net margin"
        rec_conv="Receivables conversion rate"
        roa="Return on assets"
;
drop ibq saleq NIQ ATQ RECTQ ACTQ LCTQ date;
run;

/*sort by quarter to rank*/
proc sort data=ratios1 out=ratios2;
by quarter;
run;

/*rank data*/
proc rank data=ratios2 percent ties=mean out=ratios3;
by quarter;
var nm roa cr rec_conv;
ranks rnm rroa rcr rrec_conv;
run;

proc datasets library=work;
delete ratios2;
run;

/*assings ratios to macro variables*/
%let a1=nm;

```

```

%let a2=roa;
%let a3=cr;
%let a4=rec_conv;

%macro tukey;

data ratios4/view=ratios4;
set ratios3;
by quarter;
%do i=1 %to 4;
if ^missing(r&&a&i) then do;
if r&&a&i>10 and r&&a&i<90 then &&a&i=0;
else if r&&a&i LE 10 then &&a&i=-1;
else &&a&i=1;
end;
%end;
drop r&a1 r&a2 r&a3 r&a4;
run;

%mend;
%tukey;

proc sort data=ratios4 out=ratios5 nodupkey;
by gvkey quarter;
run;

proc datasets library=work;
modify ratios5;
index create gvkey;
delete ratios2 ratios3
run;

proc sql;

```

```

create view accounting_data4 as
select a.*, b.nm,b.roa,b.cr,b.rec_conv
from accounting_data3 as a
left join
ratios5 as b
on a.gvkey=b.gvkey and a.quarter=b.quarter;
quit;

```

```

data accounting_data5/view=accounting_data5;
set accounting_data4(drop=gvkey);
run;

```

```

data total_set1/view=total_set1;
set no_accounting_data2 accounting_data5;
if operational=0 then do;
if dmax-date>365 then operational=1;
end;
run;

```

```

data total_set2/view=total_set2;
set total_set1(where=(^missing(quarter)));
run;

```

```

proc sql;
create view total_set3 as
select a.*, b.rank_ratio
from total_set2 as a left join gdp.storage as b
on a.quarter=b.quarter;
quit;

```

```

data total_set4/view=total_set4;
set total_set3(keep=permno date nm roa cr rec_conv rank_ratio
operational);

```

```

if ^missing(nm) and ^missing(roa) and ^missing(cr) and
^missing(rec_conv) then category=1;
    else if ^missing(nm) and ^missing(roa)
    then category=2;
        else if ^missing(cr) and ^missing(rec_conv)
        then category=3;
            else category=4;

run;

proc sort data=total_set4 out=total_set5;
by date;
run;

proc datasets library=work;
delete accounting_data3 no_accounting_data1 ratios5;
modify total_set5;
index create date;
run;

proc freq data=total_set5(where=((date>mdy(12,31,1961) and
date<mdy(7,1,1990))and (category=1 or category=2))) noprint;
tables nm*roa*rank_ratio*operational/out=modela1;
run;

data modela1;
set modela1;
date=mdy(6,30,1990);
run;

proc freq data=total_set5(where=((date>mdy(6,30,1990) and

```

```

    date<mdy(1,1,2008))and (category=1 or category=2))) noprint;
tables date*nm*roa*rank_ratio*operational/out=modela2;
run;

```

```

data modela3;
do date=mdy(6,30,1990) to mdy(12,31,2007);
do nm=-1 to 1;
    do roa=-1 to 1;
        do rank_ratio=-1 to 1;
            do operational=0 to 1;
                output;
            end;
        end;
    end;
end;
end;
run;

```

```

data modela4/view=modela4;
set modela1 modela2;
run;

```

```

proc sort data=modela3 out=modela5 presorted;
by date nm roa rank_ratio operational;
run;

```

```

proc sort data=modela4 out=modela6;
by date nm roa rank_ratio operational;
run;

```

```

data modela7;
merge modela5 modela6;
by date nm roa rank_ratio operational;

```

```
if missing(count) then count=0;
if operational=0 then bankrupt=count;
else going_concern=count;
drop count;
run;

proc datasets library=work;
delete modela1-modela3 modela5-modela6;
run;

data modela8a/view=modela8a;
set modela7(where=(operational=0));
by date nm roa rank_ratio;
drop going_concern operational;
run;

data modela8b/view=modela8b;
set modela7(where=(operational=1));
by date nm roa rank_ratio;
drop bankrupt operational;
run;

data modela9/view=modela9;
merge modela8a modela8b;
by date nm roa rank_ratio;
if date=mdy(6,30,1990) then do;
bankrupt=bankrupt+1;
going_concern=going_concern+32;
end;
run;
```

```

%macro cumulative;

%let count=0;

data modela10/view=modela10;
set modela9;
by date nm roa rank_ratio;
retain alpha1-alpha27 beta1-beta27 0;
%do i=-1 %to 1;
%do j=-1 %to 1;
%do k=-1 %to 1;
%let count=%eval(&count+1);
%if %eval(&i)=-1 & %eval(&j)=-1 & %eval(&k)=-1 %then %do;
if nm=%eval(&i) and roa=%eval(&j) and rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;
%end;
%else %do;
else if nm=%eval(&i) and roa=%eval(&j) and rank_ratio=%eval(&k)
then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;

%end;
%end;
%end;
%end;
run;

```

```
%mend;
```

```
%cumulative;
```

```
data modela11/view=modela11;
set modela10(where=(bankrupt+going_concern>0)
drop=alpha1-alpha27 beta1-beta27 percent);
by date nm roa rank_ratio;
run;
```

```
proc freq data=total_set5(where=((date>mdy(12,31,1961)
and date<mdy(7,1,1990))and (category=1 or category=3)))
noprint;
tables cr*rec_conv*rank_ratio*operational/out=modelb1;
run;
```

```
data modelb1;
set modelb1;
date=mdy(6,30,1990);
run;
```

```
proc freq data=total_set5(where=((date>mdy(6,30,1990) and
date<mdy(1,1,2008))and (category=1 or category=3))) noprint;
tables date*cr*rec_conv*rank_ratio*operational/out=modelb2;
run;
```

```
data modelb3;
```

```

do date=mdy(6,30,1990) to mdy(12,31,2007);
do cr=-1 to 1;
    do rec_conv=-1 to 1;
        do rank_ratio=-1 to 1;
            do operational=0 to 1;
                output;
            end;
        end;
    end;
end;

end;
end;
run;

data modelb4/view=modelb4;
set modelb1 modelb2;
run;

proc sort data=modelb3 out=modelb5 presorted;
by date cr rec_conv rank_ratio operational;
run;

proc sort data=modelb4 out=modelb6;
by date cr rec_conv rank_ratio operational;
run;

data modelb7;
merge modelb5 modelb6;
by date cr rec_conv rank_ratio operational;
if missing(count) then count=0;
if operational=0 then bankrupt=count;
else going_concern=count;
drop count;
run;

```

```
proc datasets library=work;
delete modelb1-modelb3 modelb5-modelb6;
run;
```

```
data modelb8a/view=modelb8a;
set modelb7(where=(operational=0));
by date cr rec_conv rank_ratio;
drop going_concern operational;
run;
```

```
data modelb8b/view=modelb8b;
set modelb7(where=(operational=1));
by date cr rec_conv rank_ratio;
drop bankrupt operational;
run;
```

```
data modelb9/view=modelb9;
merge modelb8a modelb8b;
by date cr rec_conv rank_ratio;
if date=mdy(6,30,1990) then do;
bankrupt=bankrupt+1;
going_concern=going_concern+32;
end;
run;
```

```
%macro cumulative;
```

```
%let count=0;
```

```
data modelb10/view=modelb10;
```

```

set modelb9;
by date cr rec_conv rank_ratio;
retain alpha1-alpha27 beta1-beta27 0;
%do i=-1 %to 1;
%do j=-1 %to 1;
%do k=-1 %to 1;
%let count=%eval(&count+1);
%if %eval(&i)=-1 & %eval(&j)=-1 & %eval(&k)=-1 %then %do;
if cr=%eval(&i) and rec_conv=%eval(&j) and rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;
%end;
%else %do;
else if cr=%eval(&i) and rec_conv=%eval(&j) and rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;

%end;
%end;
%end;
%end;
run;

%mend;

%cumulative;

```

```

data modelb11/view=modelb11;
set modelb10(where=(bankrupt+going_concern>0)
drop=alpha1-alpha27 beta1-beta27 percent);
by date cr rec_conv rank_ratio;
run;

```

```

proc freq data=total_set5(where=((date>mdy(12,31,1961) and
date<mdy(7,1,1990))and (category=4))) noprint;
tables rank_ratio*operational/out=modelc1;
run;

```

```

data modelc1;
set modelc1;
date=mdy(6,30,1990);
run;

```

```

proc print data=modelc1;
title"verification that there are no missing dates";
run;

```

```

proc freq data=total_set5(where=((date>mdy(6,30,1990) and
date<mdy(1,1,2008))and (category=4))) noprint;
tables date*rank_ratio*operational/out=modelc2;
run;

```

```

data modelc3;
do date=mdy(6,30,1990) to mdy(12,31,2007);
    do rank_ratio=-1 to 1;

```

```
                do operational=0 to 1;
                    output;
                end;
            end;
        end;
    end;
run;

data modelc4/view=modelc4;
set modelc1 modelc2;
run;

proc sort data=modelc3 out=modelc5 presorted;
by date rank_ratio operational;
run;

proc sort data=modelc4 out=modelc6;
by date rank_ratio operational;
run;

data modelc7;
merge modelc5 modelc6;
by date rank_ratio operational;
if missing(count) then count=0;
if operational=0 then bankrupt=count;
else going_concern=count;
drop count;
run;

proc datasets library=work;
delete modelc1-modelc3 modelc5-modelc6;
run;

data modelc8a/view=modelc8a;
```

```

set modelc7(where=(operational=0));
by date rank_ratio;
drop going_concern operational;
run;

```

```

data modelc8b/view=modelc8b;
set modelc7(where=(operational=1));
by date rank_ratio;
drop bankrupt operational;
run;

```

```

data modelc9/view=modelc9;
merge modelc8a modelc8b;
by date rank_ratio;
if date=mdy(6,30,1990) then do;
bankrupt=bankrupt+1;
going_concern=going_concern+32;
end;
run;

```

```
%macro cumulative;
```

```
%let count=0;
```

```

data modelc10/view=modelc10;
set modelc9;
by date rank_ratio;
retain alpha1-alpha3 beta1-beta3 0;
%do k=-1 %to 1;
%let count=%eval(&count+1);
%if %eval(&k)=-1 %then %do;

```

```

if rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;
%end;
%else %do;
else if rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;

%end;
%end;
run;

%mend;

%cumulative;

data modelc11/view=modelc11;
set modelc10(where=(bankrupt+going_concern>0)
drop=alpha1-alpha3 beta1-beta3 percent);
by date rank_ratio;
run;

proc freq data=total_set5(where=((date>mdy(6,30,1990) and
date<mdy(1,1,2008))and (category=1))) noprint;

```

```

tables date*nm*roa*cr*rec_conv/out=sample1a;
run;

proc sql;
create view sample1b as
select a.date,a.nm,a.roa,a.cr,a.rec_conv,
sum(b.expectation*.53163,c.expectation*.46829)
  as expectation
from sample1a as a, modela11 as b, modelb11 as c
where a.date=b.date and a.date=c.date and
a.nm=b.nm and a.roa=b.roa and a.cr=c.cr and
a.rec_conv=c.rec_conv and a.date>mdy(6,30,1990)
order by date;
quit;

proc print data=sample1b(obs=1000);
title"posterior expected bankruptcy rate";
run;

proc freq data=total_set5(where=(mdy(6,30,1990)
and date<mdy(1,1,2008) and category=1)) noprint;
tables date*nm*roa*cr*rec_conv*operational/out=sample1c;
run;

data sample1d/view=sample1d;
set sample1c;
by date nm roa cr rec_conv operational;
if operational=0 then bankrupt=count;
else going_concern=count;
run;

data sample1e/view=sample1e;
set sample1d(where=(operational=0));

```

```
by date nm roa cr rec_conv;
drop operational percent;
run;
```

```
data sample1f/view=sample1f;
set sample1d(where=(operational=1));
by date nm roa cr rec_conv;
drop operational percent;
run;
```

```
data sample1g/view=sample1g;
merge sample1e sample1f;
by date nm roa cr rec_conv;
run;
```

```
proc means data=sample1g n nmiss;
title"count of missing bankruptcies and
going concerns(if any)";
run;
```

```
data sample1h/view=sample1h;
set sample1g;
by date nm roa cr rec_conv;
if missing(bankrupt) then bankrupt=0;
if missing(going_concern) then going_concern=0;
run;
```

```
proc sql;
create view sample1i as
select a.*,b.bankrupt,b.going_concern
from sample1b as a, sample1h as b
where a.date=b.date and a.nm=b.nm and
a.roa=b.roa and a.cr=b.cr and a.rec_conv=b.rec_conv;
```

```

run;

data sample1j/view=sample1j;
set sample1i(drop=nm roa cr rec_conv);
run;

proc freq data=total_set5(where=((date>mdy(6,30,1990)
and date<mdy(1,1,2008))and (category=2))) noprint;
tables date*nm*roa/out=sample2a;
run;

proc sql;
create view sample2b as
select a.date,a.nm,a.roa,b.expectation
from sample2a as a, modela11 as b
where a.date=b.date and a.nm=b.nm and
a.roa=b.roa and a.date>mdy(6,30,1990)
order by date;
quit;

proc print data=sample2b(obs=100);
title"posterior expected bankruptcy rate2";
run;

proc freq data=total_set5(where=(mdy(6,30,1990)
and date<mdy(1,1,2008) and category=2)) noprint;
tables date*nm*roa*operational/out=sample2c;
run;

data sample2d/view=sample2d;

```

```

set sample2c;
by date nm roa operational;
if operational=0 then bankrupt=count;
else going_concern=count;
run;

data sample2e/view=sample2e;
set sample2d(where=(operational=0));
by date nm roa;
drop operational percent;
run;

data sample2f/view=sample2f;
set sample2d(where=(operational=1));
by date nm roa;
drop operational percent;
run;

data sample2g/view=sample2g;
merge sample2e sample2f;
by date nm roa;
run;

proc means data=sample2g n nmiss;
title"count of missing bankruptcies and
going concerns(if any) 2";
run;

data sample2h/view=sample2h;
set sample2g;
by date nm roa;
if missing(bankrupt) then bankrupt=0;
if missing(going_concern) then going_concern=0;

```

```

run;

proc sql;
create view sample2i as
select a.*,b.bankrupt,b.going_concern
from sample2b as a, sample2h as b
where a.date=b.date and a.nm=b.nm
and a.roa=b.roa;
run;

data sample2j/view=sample2j;
set sample2i(drop=nm roa);
run;

proc freq data=total_set5(where=((date>mdy(6,30,1990)
and date<mdy(1,1,2008))and (category=3))) noprint;
tables date*cr*rec_conv/out=sample3a;
run;

proc sql;
create view sample3b as
select a.date,a.cr,a.rec_conv,b.expectation
from sample3a as a, modelb11 as b
where a.date=b.date and a.cr=b.cr and
a.rec_conv=b.rec_conv and a.date>mdy(6,30,1990)
order by date;
quit;

proc print data=sample3b(obs=100);
title"posterior expected bankruptcy rate3";

```

```

run;

proc freq data=total_set5(where=(mdy(6,30,1990)
and date<mdy(1,1,2008) and category=3)) noprint;
tables date*cr*rec_conv*operational/out=sample3c;
run;

data sample3d/view=sample3d;
set sample3c;
by date cr rec_conv operational;
if operational=0 then bankrupt=count;
else going_concern=count;
run;

data sample3e/view=sample3e;
set sample3d(where=(operational=0));
by date cr rec_conv;
drop operational percent;
run;

data sample3f/view=sample3f;
set sample3d(where=(operational=1));
by date cr rec_conv;
drop operational percent;
run;

data sample3g/view=sample3g;
merge sample3e sample3f;
by date cr rec_conv;
run;

proc means data=sample3g n nmiss;
title"count of missing bankruptcies and going

```

```

concerns(if any)3";
run;

data sample3h/view=sample3h;
set sample3g;
by date cr rec_conv;
if missing(bankrupt) then bankrupt=0;
if missing(going_concern) then going_concern=0;
run;

proc sql;
create view sample3i as
select a.*,b.bankrupt,b.going_concern
from sample3b as a, sample3h as b
where a.date=b.date and a.cr=b.cr and
a.rec_conv=b.rec_conv;
run;

data sample3j/view=sample3j;
set sample3i(drop=cr rec_conv);
run;

proc freq data=total_set5(where=((date>mdy(6,30,1990)
and date<mdy(1,1,2008))and (category=4))) noprint;
tables date/out=sample4a;
run;

proc sql;
create view sample4b as
select a.date,b.expectation

```

```
from sample4a as a, modelc11 as b
where a.date=b.date and a.date>mdy(6,30,1990)
order by date;
quit;

proc print data=sample4b(obs=100);
title"posterior expected bankruptcy rate4";
run;

proc freq data=total_set5(where=(mdy(6,30,1990)
and date<mdy(1,1,2008) and category=3)) noprint;
tables date*operational/out=sample4c;
run;

data sample4d/view=sample4d;
set sample4c;
by date operational;
if operational=0 then bankrupt=count;
else going_concern=count;
run;

data sample4e/view=sample4e;
set sample4d(where=(operational=0));
by date;
drop operational percent;
run;

data sample4f/view=sample4f;
set sample4d(where=(operational=1));
by date;
drop operational percent;
run;
```

```
data sample4g/view=sample4g;
merge sample4e sample4f;
by date;
run;

proc means data=sample4g n nmiss;
title"count of missing bankruptcies and
going concerns(if any)";
run;

data sample4h/view=sample4h;
set sample4g;
by date;
if missing(bankrupt) then bankrupt=0;
if missing(going_concern) then going_concern=0;
run;

proc sql;
create view sample4i as
select a.*,b.bankrupt,b.going_concern
from sample4b as a, sample4h as b
where a.date=b.date;
run;

/*redundent statement for programming simplicity*/
data sample4j/view=sample4j;
set sample4i;
run;

data joint_sample1/view=joint_sample1;
set sample1j sample2j sample3j sample4j;
run;
```

```

proc print data=joint_sample1(obs=10);
title "verification of variables in final set";
run;

data joint_sample2/view=joint_sample2;
set joint_sample1;
retain number omega 1 df delta 0;
g_bar=bankrupt/(bankrupt+going_concern)-
expectation;
if bankrupt+going_concern=1 then s_squared=0;
else s_squared=bankrupt*going_concern/
(bankrupt+going_concern)/(bankrupt+going_concern-1);
sum_of_squares=omega*df+s_squared*
(bankrupt+going_concern-1)+1/
(1/number+1/(bankrupt+going_concern))*
((delta-g_bar)**2);
delta=(number*delta+
(bankrupt+going_concern)*g_bar)/
(number+bankrupt+going_concern);
number=number+bankrupt+going_concern;
df=df+bankrupt+going_concern;
omega=sum_of_squares/df;
run;

proc summary data=joint_sample2;
var number;
output out=joint_sample3 max(number)=max;
run;

proc sql;
create table joint_sample4 as
select a.*
from joint_sample2 as a, joint_sample3 as b

```

```

where a.number=b.max;
run;

proc print data=joint_sample4;
run;

```

## B.8 Posterior Test for Bankruptcy

```

libname gdp '/home/wvu/deharris/new/state3';
libname bankrupt '/home/wvu/deharris/new/state3';

/*options mprint symbolgen mlogic obs=100000;*/

/*gathers all price data, in raw form, and calculates maximum and minimum values
proc sql;
create view gather1 as
select permno,min(date) as dmin, max(date) as dmax
from crsp.dsf
where prc>0 and ^missing(prc)
group by permno;
quit;

data gather2/view=gather2;
set gather1(where=(dmax-dmin>730));
by permno;
run;

/*determine final disposition*/
data state_space/view=state_space;
set crsp.dseall(keep=permno dlstcd shrcd dlamt nwperm where=(^missing(dlstcd)));

```

```

by permno;
operational=0;
merge=0;
cash_buyout=0;
private=0;
bankrupt=0;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
else if substr(left(shrcd),2,1)='5' then delete;
else if substr(left(shrcd),2,1)='8' then delete;
else if substr(left(shrcd),1,1)='2' then delete;
else if substr(left(shrcd),1,1)='4' then delete;
else if substr(left(shrcd),1,1)='7' then delete;
else do;
if dlstcd=100 then operational=1;
else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;
else if dlstcd in(502,510,513,514,516,517,519,520,573,575,580,581,582,583,584) th
private=1;
else if dlstcd in(574) then bankrupt=1;
      else if dlstcd
in(234,242,243,251,252,261,262,300,332,341,342,400,450,460,470,500) then do;
          if dlstcd in(234,500) then do;
              if nwperm>0 then merge=1;
              else if nwperm=0 and dlamt>0 then cash_buyout=1;
              else private=1;
          end;
          else if dlstcd in (242,243,251,341,342) then do;
              if nwperm>0 then merge=1;
              else cash_buyout=1;
          end;
          else if dlstcd in (252,261,262,450) then do;
              if nwperm>0 then merge=1;

```

```

                else if dlamt>0 then cash_buyout=1;
                else delete;
            end;
        end;

end;

else if dlstcd in(300) then do;
        if nwperm>0 then merge=1;
        else delete;
    end;
    else if dlstcd in (332) then do;
        if nwperm>0 then merge=1;
        else private=1;
    end;
    else if dlstcd in (400,460,470) then do;
        if dlamt>0 then cash_buyout=1;
        else bankrupt=1;
    end;
end;

bankrupt=private+bankrupt;
if bankrupt=1 then operational=0;
else operational=1;
drop shrcd nwperm dlamt dlstcd cash_buyout private merge bankrupt;
run;

data gather3/view=gather3;
merge gather2(in=aa) state_space(in=bb);
by permno;
if aa and bb;
run;

/*gather gvkey permno linkages*/

```

```

data link/view=link;
set crspa.ccmxpf_lnkused(keep=ugvkey ulinkdt apermno ulinkenddt
ulinktype where=(ulinktype in("LC","LU","LS")) rename=(ugvkey=gvkey apermno=permno)
by permno;
if missing(ulinkenddt) then ulinkenddt=mdy(12,31,2013);
if missing(ulinkdt) then delete;
else if missing(gvkey) then delete;
else if length(gvkey)=1 then delete;
drop ulinktype;
run;

```

```

proc sql;
create view gather4 as
select a.*,b.gvkey,b.ulinkdt,b.ulinkenddt
from gather3 as a
left join
link as b
on a.permno=b.permno
order by permno;
quit;

```

```

data gather5/view=gather5;
set gather4(where=(^missing(dmin) and ^missing(dmax)));
by permno;
run;

```

```

/*splits the data into groups with accounting and groups without accounting data*/
data no_accounting_data1 accounting_data1;
set gather5;
by permno;
do date=dmin to dmax;
if missing(ulinkdt) then do;
quarter=4*(year(date)-1929)+qtr(date);

```

```

output no_accounting_data1;
end;
else do;
if date GE ulinkdt and date LE ulinkenddt then do;
quarter=4*(year(date)-1929)+qtr(date);
output accounting_data1;
end;
else do;
quarter=4*(year(date)-1929)+qtr(date);
output no_accounting_data1;

end;
end;
end;
run;

/*drops unused fields*/
data no_accounting_data2/view=no_accounting_data2;
set no_accounting_data1(drop=gvkey ulinkdt ulinkenddt);
by permno;
run;

data accounting_data2/view=accounting_data2;
set accounting_data1(drop=ulinkdt ulinkenddt);
by permno;
run;

proc sort data=accounting_data2 out=accounting_data3;
by gvkey quarter;
run;

proc datasets library=work;
modify accounting_data3;

```

```

index create gvkey;
delete accounting_data1;
run;

/*This is the end of the permno based gathering and combining of data*/

/*gather acct data*/
data acct_data/view=acct_data;
set comp.fundq(keep=datadate gvkey SALEQ ATQ RECTQ ACTQ
LCTQ NIQ IBQ where=(^missing(date))rename=(datadate=date));
by gvkey date;
if missing(NIQ) then NIQ=IBQ;
quarter=4*(year(date)-1929)+qtr(date);
label quarter="Quarter";
run;

/*convert to ratios*/
data ratios1/view=ratios1;
/*create financial ratios*/
set acct_data;
by gvkey date;
/*Net Margin*/
if (^missing(niq) and ^missing(SALEQ) and (SALEQ NE 0)) then nm=niq/SALEQ;
    else nm=.;
/*Return on Assets*/
if (^missing(NIQ) and ^missing(ATQ) and (ATQ NE 0)) then roa=NIQ/ATQ;
else roa=.;
/*Current ratio*/
if (^missing(ACTQ) and ^missing(LCTQ) and (LCTQ NE 0)) then cr=ACTQ/LCTQ;
    else cr=.;

```

```

/*receivables conversion*/
if (~missing(RECTQ) and ~missing(SALEQ) and (SALEQ NE 0)) then rec_conv=RECTQ/SAL
    else rec_conv=.;
label    cr="Current ratio"
         nm="Net margin"
         rec_conv="Receivables conversion rate"
         roa="Return on assets"
;
drop ibq saleq NIQ ATQ RECTQ ACTQ LCTQ date;
run;

/*sort by quarter to rank*/
proc sort data=ratios1 out=ratios2;
by quarter;
run;

/*rank data*/
proc rank data=ratios2 percent ties=mean out=ratios3;
by quarter;
var nm roa cr rec_conv;
ranks rnm rroa rcr rrec_conv;
run;

proc datasets library=work;
delete ratios2;
run;

/*assings ratios to macro variables*/
%let a1=nm;
%let a2=roa;
%let a3=cr;
%let a4=rec_conv;

```

```

%macro tukey;

data ratios4/view=ratios4;
set ratios3;
by quarter;
%do i=1 %to 4;
if ^missing(r&&a&i) then do;
if r&&a&i>10 and r&&a&i<90 then &&a&i=0;
else if r&&a&i LE 10 then &&a&i=-1;
else &&a&i=1;
end;
%end;
drop r&a1 r&a2 r&a3 r&a4;
run;

%mend;
%tukey;

proc sort data=ratios4 out=ratios5 nodupkey;
by gvkey quarter;
run;

proc datasets library=work;
modify ratios5;
index create gvkey;
delete ratios2 ratios3
run;

proc sql;
create view accounting_data4 as
select a.*, b.nm,b.roa,b.cr,b.rec_conv
from accounting_data3 as a

```

```

left join
ratios5 as b
on a.gvkey=b.gvkey and a.quarter=b.quarter;
quit;

data accounting_data5/view=accounting_data5;
set accounting_data4(drop=gvkey);
run;

data total_set1/view=total_set1;
set no_accounting_data2 accounting_data5;
if operational=0 then do;
if dmax-date>365 then operational=1;
end;
run;

data total_set2/view=total_set2;
set total_set1(where=(^missing(quarter)));
run;

proc sql;
create view total_set3 as
select a.*, b.rank_ratio
from total_set2 as a left join gdp.storage as b
on a.quarter=b.quarter;
quit;

proc sort data=total_set3 out=total_set3a;
by permno date;
run;

data total_set3b/view=total_set3b;
set total_set3a;

```

```

by permno date;
lpermno=lag91(permno);
lnm=lag91(nm);
lroa=lag91(roa);
lcr=lag91(cr);
lrec_conv=lag91(rec_conv);
run;

```

```

data total_set3c/view=total_set3c;
set total_set3b;
by permno date;
if permno=lpermno then do;
nm=lnm;
roa=lroa;
cr=lcr;
rec_conv=lrec_conv;
end;
drop lpermno lnm lroa lcr lrec_conv;
run;

```

```

data total_set4/view=total_set4;
set total_set3c(keep=permno date nm roa cr rec_conv rank_ratio
operational);
if ^missing(nm) and ^missing(roa) and ^missing(cr) and ^missing(rec_conv) then ca
    else if ^missing(nm) and ^missing(roa) then category=2;
    else if ^missing(cr) and ^missing(rec_conv) then category=3;
    else category=4;

run;

```

```

proc sort data=total_set4 out=total_set5;

```

```
by date;
run;
```

```
proc datasets library=work;
delete accounting_data3 no_accounting_data1 ratios5;
modify total_set5;
index create date;
run;
```

```
proc freq data=total_set5(where=((date>mdy(12,31,1961) and date<mdy(7,1,1990))and
tables nm*roa*rank_ratio*operational/out=modela1;
run;
```

```
data modela1;
set modela1;
date=mdy(6,30,1990);
run;
```

```
proc freq data=total_set5(where=((date>mdy(6,30,1990) and date<mdy(1,1,2008))and
tables date*nm*roa*rank_ratio*operational/out=modela2;
run;
```

```
data modela3;
do date=mdy(6,30,1990) to mdy(12,31,2007);
do nm=-1 to 1;
do roa=-1 to 1;
do rank_ratio=-1 to 1;
do operational=0 to 1;
output;
end;
end;
```

```
end;

end;
end;
run;

data modela4/view=modela4;
set modela1 modela2;
run;

proc sort data=modela3 out=modela5 presorted;
by date nm roa rank_ratio operational;
run;

proc sort data=modela4 out=modela6;
by date nm roa rank_ratio operational;
run;

data modela7;
merge modela5 modela6;
by date nm roa rank_ratio operational;
if missing(count) then count=0;
if operational=0 then bankrupt=count;
else going_concern=count;
drop count;
run;

proc datasets library=work;
delete modela1-modela3 modela5-modela6;
run;

data modela8a/view=modela8a;
set modela7(where=(operational=0));
by date nm roa rank_ratio;
```

```

drop going_concern operational;
run;

data modela8b/view=modela8b;
set modela7(where=(operational=1));
by date nm roa rank_ratio;
drop bankrupt operational;
run;

data modela9/view=modela9;
merge modela8a modela8b;
by date nm roa rank_ratio;
if date=mdy(6,30,1990) then do;
bankrupt=bankrupt+1;
going_concern=going_concern+32;
end;
run;

%macro cumulative;

%let count=0;

data modela10/view=modela10;
set modela9;
by date nm roa rank_ratio;
retain alpha1-alpha27 beta1-beta27 0;
%do i=-1 %to 1;
%do j=-1 %to 1;
%do k=-1 %to 1;
%let count=%eval(&count+1);

```

```

%if %eval(&i)=-1 & %eval(&j)=-1 & %eval(&k)=-1 %then %do;
if nm=%eval(&i) and roa=%eval(&j) and rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;
%end;
%else %do;
else if nm=%eval(&i) and roa=%eval(&j) and rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_conce
expectation=alpha&count/(alpha&co
end;

%end;
%end;
%end;
%end;
run;

%mend;

%cumulative;

data modela11/view=modela11;
set modela10(where=(bankrupt+going_concern>0) drop=alpha1-alpha27 beta1-beta27 pe
by date nm roa rank_ratio;
run;

```

```
proc freq data=total_set5(where=((date>mdy(12,31,1961) and date<mdy(7,1,1990))and
tables cr*rec_conv*rank_ratio*operational/out=modelb1;
run;
```

```
data modelb1;
set modelb1;
date=mdy(6,30,1990);
run;
```

```
proc freq data=total_set5(where=((date>mdy(6,30,1990) and date<mdy(1,1,2008))and
tables date*cr*rec_conv*rank_ratio*operational/out=modelb2;
run;
```

```
data modelb3;
do date=mdy(6,30,1990) to mdy(12,31,2007);
do cr=-1 to 1;
    do rec_conv=-1 to 1;
        do rank_ratio=-1 to 1;
            do operational=0 to 1;
                output;
            end;
        end;
    end;
end;
end;
end;
run;
```

```
data modelb4/view=modelb4;
set modelb1 modelb2;
run;
```

```
proc sort data=modelb3 out=modelb5 presorted;
```

```
by date cr rec_conv rank_ratio operational;  
run;
```

```
proc sort data=modelb4 out=modelb6;  
by date cr rec_conv rank_ratio operational;  
run;
```

```
data modelb7;  
merge modelb5 modelb6;  
by date cr rec_conv rank_ratio operational;  
if missing(count) then count=0;  
if operational=0 then bankrupt=count;  
else going_concern=count;  
drop count;  
run;
```

```
proc datasets library=work;  
delete modelb1-modelb3 modelb5-modelb6;  
run;
```

```
data modelb8a/view=modelb8a;  
set modelb7(where=(operational=0));  
by date cr rec_conv rank_ratio;  
drop going_concern operational;  
run;
```

```
data modelb8b/view=modelb8b;  
set modelb7(where=(operational=1));  
by date cr rec_conv rank_ratio;  
drop bankrupt operational;  
run;
```

```
data modelb9/view=modelb9;
```

```

merge modelb8a modelb8b;
by date cr rec_conv rank_ratio;
if date=mdy(6,30,1990) then do;
bankrupt=bankrupt+1;
going_concern=going_concern+32;
end;
run;

%macro cumulative;

%let count=0;

data modelb10/view=modelb10;
set modelb9;
by date cr rec_conv rank_ratio;
retain alpha1-alpha27 beta1-beta27 0;
%do i=-1 %to 1;
%do j=-1 %to 1;
%do k=-1 %to 1;
%let count=%eval(&count+1);
%if %eval(&i)=-1 & %eval(&j)=-1 & %eval(&k)=-1 %then %do;
if cr=%eval(&i) and rec_conv=%eval(&j) and rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;
%end;
%else %do;
else if cr=%eval(&i) and rec_conv=%eval(&j) and rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_conce

```

```

expectation=alpha&count/(alpha&co
end;

%end;
%end;
%end;
%end;
run;

%mend;

%cumulative;

data modelb11/view=modelb11;
set modelb10(where=(bankrupt+going_concern>0) drop=alpha1-alpha27 beta1-beta27 pe
by date cr rec_conv rank_ratio;
run;

proc freq data=total_set5(where=((date>mdy(12,31,1961) and date<mdy(7,1,1990))and
tables rank_ratio*operational/out=modelc1;
run;

data modelc1;
set modelc1;
date=mdy(6,30,1990);
run;

```

```
proc print data=modelc1;
title"verification that there are no missing dates";
run;

proc freq data=total_set5(where=((date>mdy(6,30,1990) and date<mdy(1,1,2008))and
tables date*rank_ratio*operational/out=modelc2;
run;

data modelc3;
do date=mdy(6,30,1990) to mdy(12,31,2007);
    do rank_ratio=-1 to 1;
        do operational=0 to 1;
            output;
        end;
    end;
end;
run;

data modelc4/view=modelc4;
set modelc1 modelc2;
run;

proc sort data=modelc3 out=modelc5 presorted;
by date rank_ratio operational;
run;

proc sort data=modelc4 out=modelc6;
by date rank_ratio operational;
run;

data modelc7;
merge modelc5 modelc6;
```

```
by date rank_ratio operational;
if missing(count) then count=0;
if operational=0 then bankrupt=count;
else going_concern=count;
drop count;
run;

proc datasets library=work;
delete modelc1-modelc3 modelc5-modelc6;
run;

data modelc8a/view=modelc8a;
set modelc7(where=(operational=0));
by date rank_ratio;
drop going_concern operational;
run;

data modelc8b/view=modelc8b;
set modelc7(where=(operational=1));
by date rank_ratio;
drop bankrupt operational;
run;

data modelc9/view=modelc9;
merge modelc8a modelc8b;
by date rank_ratio;
if date=mdy(6,30,1990) then do;
bankrupt=bankrupt+1;
going_concern=going_concern+32;
end;
run;
```

```

%macro cumulative;

%let count=0;

data modelc10/view=modelc10;
set modelc9;
by date rank_ratio;
retain alpha1-alpha3 beta1-beta3 0;
%do k=-1 %to 1;
%let count=%eval(&count+1);
%if %eval(&k)=-1 %then %do;
if rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;
%end;
%else %do;
else if rank_ratio=%eval(&k) then do;
alpha&count=alpha&count+bankrupt;
beta&count=beta&count+going_conce
expectation=alpha&count/(alpha&co
end;

%end;
%end;
run;

%mend;

%cumulative;

```

```

data modelc11/view=modelc11;
set modelc10(where=(bankrupt+going_concern>0) drop=alpha1-alpha3 beta1-beta3 perc
by date rank_ratio;
run;

```

```

proc freq data=total_set5(where=((date>mdy(6,30,1990) and date<mdy(1,1,2008))and
tables date*nm*roa*cr*rec_conv/out=sample1a;
run;

```

```

proc sql;
create view sample1b as
select a.date,a.nm,a.roa,a.cr,a.rec_conv,sum(b.expectation*.53163,c.expectation*
from sample1a as a, modela11 as b, modelb11 as c
where a.date=b.date and a.date=c.date and a.nm=b.nm and a.roa=b.roa and a.cr=c.cr
order by date;
quit;

```

```

proc print data=sample1b(obs=1000);
title"posterior expected bankruptcy rate";
run;

```

```

proc freq data=total_set5(where=(mdy(6,30,1990) and date<mdy(1,1,2008) and catego
tables date*nm*roa*cr*rec_conv*operational/out=sample1c;
run;

```

```

data sample1d/view=sample1d;
set sample1c;
by date nm roa cr rec_conv operational;
if operational=0 then bankrupt=count;

```

```

else going_concern=count;
run;

data sample1e/view=sample1e;
set sample1d(where=(operational=0));
by date nm roa cr rec_conv;
drop operational percent;
run;

data sample1f/view=sample1f;
set sample1d(where=(operational=1));
by date nm roa cr rec_conv;
drop operational percent;
run;

data sample1g/view=sample1g;
merge sample1e sample1f;
by date nm roa cr rec_conv;
run;

proc means data=sample1g n nmiss;
title"count of missing bankruptcies and going concerns(if any)";
run;

data sample1h/view=sample1h;
set sample1g;
by date nm roa cr rec_conv;
if missing(bankrupt) then bankrupt=0;
if missing(going_concern) then going_concern=0;
run;

proc sql;
create view sample1i as

```

```

select a.*,b.bankrupt,b.going_concern
from sample1b as a, sample1h as b
where a.date=b.date and a.nm=b.nm and a.roa=b.roa and a.cr=b.cr and a.rec_conv=b.
run;

```

```

data sample1j/view=sample1j;
set sample1i(drop=nm roa cr rec_conv);
run;

```

```

proc freq data=total_set5(where=((date>mdy(6,30,1990) and date<mdy(1,1,2008))and
tables date*nm*roa/out=sample2a;
run;

```

```

proc sql;
create view sample2b as
select a.date,a.nm,a.roa,b.expectation
from sample2a as a, modela11 as b
where a.date=b.date and a.nm=b.nm and a.roa=b.roa and a.date>mdy(6,30,1990)
order by date;
quit;

```

```

proc print data=sample2b(obs=100);
title"posterior expected bankruptcy rate2";
run;

```

```

proc freq data=total_set5(where=(mdy(6,30,1990) and date<mdy(1,1,2008) and catego
tables date*nm*roa*operational/out=sample2c;
run;

```

```

data sample2d/view=sample2d;

```

```

set sample2c;
by date nm roa operational;
if operational=0 then bankrupt=count;
else going_concern=count;
run;

data sample2e/view=sample2e;
set sample2d(where=(operational=0));
by date nm roa;
drop operational percent;
run;

data sample2f/view=sample2f;
set sample2d(where=(operational=1));
by date nm roa;
drop operational percent;
run;

data sample2g/view=sample2g;
merge sample2e sample2f;
by date nm roa;
run;

proc means data=sample2g n nmiss;
title"count of missing bankruptcies and going concerns(if any) 2";
run;

data sample2h/view=sample2h;
set sample2g;
by date nm roa;
if missing(bankrupt) then bankrupt=0;
if missing(going_concern) then going_concern=0;
run;

```

```
proc sql;
create view sample2i as
select a.*,b.bankrupt,b.going_concern
from sample2b as a, sample2h as b
where a.date=b.date and a.nm=b.nm and a.roa=b.roa;
run;
```

```
data sample2j/view=sample2j;
set sample2i(drop=nm roa);
run;
```

```
proc freq data=total_set5(where=((date>mdy(6,30,1990) and date<mdy(1,1,2008))and
tables date*cr*rec_conv/out=sample3a;
run;
```

```
proc sql;
create view sample3b as
select a.date,a.cr,a.rec_conv,b.expectation
from sample3a as a, modelb11 as b
where a.date=b.date and a.cr=b.cr and a.rec_conv=b.rec_conv and a.date>mdy(6,30,1
order by date;
quit;
```

```
proc print data=sample3b(obs=100);
title"posterior expected bankruptcy rate3";
run;
```

```
proc freq data=total_set5(where=(mdy(6,30,1990) and date<mdy(1,1,2008) and catego
tables date*cr*rec_conv*operational/out=sample3c;
```

```
run;

data sample3d/view=sample3d;
set sample3c;
by date cr rec_conv operational;
if operational=0 then bankrupt=count;
else going_concern=count;
run;

data sample3e/view=sample3e;
set sample3d(where=(operational=0));
by date cr rec_conv;
drop operational percent;
run;

data sample3f/view=sample3f;
set sample3d(where=(operational=1));
by date cr rec_conv;
drop operational percent;
run;

data sample3g/view=sample3g;
merge sample3e sample3f;
by date cr rec_conv;
run;

proc means data=sample3g n nmiss;
title"count of missing bankruptcies and going concerns(if any)3";
run;

data sample3h/view=sample3h;
set sample3g;
by date cr rec_conv;
```

```

if missing(bankrupt) then bankrupt=0;
if missing(going_concern) then going_concern=0;
run;

```

```

proc sql;
create view sample3i as
select a.*,b.bankrupt,b.going_concern
from sample3b as a, sample3h as b
where a.date=b.date and a.cr=b.cr and a.rec_conv=b.rec_conv;
run;

```

```

data sample3j/view=sample3j;
set sample3i(drop=cr rec_conv);
run;

```

```

proc freq data=total_set5(where=((date>mdy(6,30,1990) and date<mdy(1,1,2008)))and
tables date/out=sample4a;
run;

```

```

proc sql;
create view sample4b as
select a.date,b.expectation
from sample4a as a, modelc11 as b
where a.date=b.date and a.date>mdy(6,30,1990)
order by date;
quit;

```

```

proc print data=sample4b(obs=100);
title"posterior expected bankruptcy rate4";
run;

```

```
proc freq data=total_set5(where=(mdy(6,30,1990) and date<mdy(1,1,2008) and catego
tables date*operational/out=sample4c;
run;

data sample4d/view=sample4d;
set sample4c;
by date operational;
if operational=0 then bankrupt=count;
else going_concern=count;
run;

data sample4e/view=sample4e;
set sample4d(where=(operational=0));
by date;
drop operational percent;
run;

data sample4f/view=sample4f;
set sample4d(where=(operational=1));
by date;
drop operational percent;
run;

data sample4g/view=sample4g;
merge sample4e sample4f;
by date;
run;

proc means data=sample4g n nmiss;
title"count of missing bankruptcies and going concerns(if any)";
run;
```

```

data sample4h/view=sample4h;
set sample4g;
by date;
if missing(bankrupt) then bankrupt=0;
if missing(going_concern) then going_concern=0;
run;

```

```

proc sql;
create view sample4i as
select a.*,b.bankrupt,b.going_concern
from sample4b as a, sample4h as b
where a.date=b.date;
run;

```

```

/*redundent statement for programming simplicity*/
data sample4j/view=sample4j;
set sample4i;
run;

```

```

data joint_sample1/view=joint_sample1;
set sample1j sample2j sample3j sample4j;
run;

```

```

proc print data=joint_sample1(obs=10);
title "verification of variables in final set";
run;

```

```

data joint_sample2/view=joint_sample2;
set joint_sample1;
retain number omega 1 df delta 0;
g_bar=bankrupt/(bankrupt+going_concern)-expectation;
if bankrupt+going_concern=1 then s_squared=0;
else s_squared=bankrupt*going_concern/(bankrupt+going_concern)/(bankrupt+going_co

```

```

sum_of_squares=omega*df+s_squared*(bankrupt+going_concern-1)+1/(1/number+1/(bankr
delta=(number*delta+(bankrupt+going_concern)*g_bar)/(number+bankrupt+going_concern
number=number+bankrupt+going_concern;
df=df+bankrupt+going_concern;
omega=sum_of_squares/df;
run;

proc summary data=joint_sample2;
var number;
output out=joint_sample3 max(number)=max;
run;

proc sql;
create table joint_sample4 as
select a.*
from joint_sample2 as a, joint_sample3 as b
where a.number=b.max;
run;

proc print data=joint_sample4;
run;

```

## B.9 Posterior Test for Merger

```
/*options mprint symbolgen mlogic obs=100000;*/
```

```
/*gathers all price data, in raw form, and calculates maximum and minimum values
proc sql;
create view gather1 as
select permno,min(date) as dmin, max(date) as dmax

```

```

from crsp.dsf
where prc>0 and ^missing(prc)
group by permno;
quit;

data gather2/view=gather2;
set gather1(where=(dmax-dmin>730));
by permno;
run;

/*determine final disposition*/
data state_space/view=state_space;
set crsp.dseall(keep=permno dlstcd shrcd dlamt nwperm where=(^missing(dlstcd)));
by permno;
operational=0;
merge=0;
cash_buyout=0;
private=0;
bankrupt=0;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
else if substr(left(shrcd),2,1)='5' then delete;
else if substr(left(shrcd),2,1)='8' then delete;
else if substr(left(shrcd),1,1)='2' then delete;
else if substr(left(shrcd),1,1)='4' then delete;
else if substr(left(shrcd),1,1)='7' then delete;
else do;
if dlstcd=100 then operational=1;
else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;
else if dlstcd in(502,510,513,514,516,517,519,520,573,575,580,581,582,583,584) then
private=1;

```

```

else if dlstcd in(574) then bankrupt=1;
    else if dlstcd
in(234,242,243,251,252,261,262,300,332,341,342,400,450,460,470,500) then do;
        if dlstcd in(234,500) then do;
            if nwperm>0 then merge=1;
            else if nwperm=0 and dlamt>0 then cash_buyout=1;
            else private=1;
        end;
        else if dlstcd in (242,243,251,341,342) then do;
            if nwperm>0 then merge=1;
            else cash_buyout=1;
        end;
        else if dlstcd in (252,261,262,450) then do;
            if nwperm>0 then merge=1;
            else if dlamt>0 then cash_buyout=1;
            else delete;
        end;
    end;

end;
else if dlstcd in(300) then do;
        if nwperm>0 then merge=1;
        else delete;
    end;
    else if dlstcd in (332) then do;
        if nwperm>0 then merge=1;
        else private=1;
    end;
    else if dlstcd in (400,460,470) then do;
        if dlamt>0 then cash_buyout=1;
        else bankrupt=1;
    end;
end;
end;

```

```
merge=merge+cash_buyout;
if merge=1 then operational=0;
else operational=1;
drop shrcd nwperm dlamt dlstcd cash_buyout private merge bankrupt;
run;
```

```
data gather3/view=gather3;
merge gather2(in=aa) state_space(in=bb);
by permno;
if aa and bb;
run;
```

```
/*gather gvkey permno linkages*/
data link/view=link;
set crspa.ccmxpf_lnkused(keep=ugvkey ulinkdt apermno ulinkenddt
ulinktype where=(ulinktype in("LC","LU","LS")) rename=(ugvkey=gvkey apermno=permno)
by permno;
if missing(ulinkenddt) then ulinkenddt=mdy(12,31,2013);
if missing(ulinkdt) then delete;
else if missing(gvkey) then delete;
else if length(gvkey)=1 then delete;
drop ulinktype;
run;
```

```
proc sql;
create view gather4 as
select a.*,b.gvkey,b.ulinkdt,b.ulinkenddt
from gather3 as a
left join
link as b
on a.permno=b.permno
order by permno;
```

```

quit;

data gather5;
set gather4(where=(^missing(dmin) and ^missing(dmax) and dmax>mdy(12,31,1961) and
by permno;
if dmin<mdy(1,1,1962) then dmin=mdy(1,1,1962);
if dmax>mdy(12,31,2007) then dmax=mdy(12,31,2007);
do date=dmin to dmax;
quarter=4*(year(date)-1929)+qtr(date);
output;
end;
run;

data no_accounting_data1;
set gather5(where=(missing(ulinkdt)));
by permno;
run;

proc sort data=gather5(where=(^missing(ulinkdt) and date ge ulinkdt and date le u
by permno date;
run;

data subset1/view=subset1;
set no_accounting_data1(keep=permno date) accounting_data1(keep=permno date);
dummy=1;
run;

proc sort data=subset1 out=subset2;
by permno date;
run;

proc sort data=gather5 presorted out=subset3;
by permno date;

```

```
run;

proc datasets library=work;
delete gather5;
run;

data subset4/view=subset4;
merge subset3 subset2;
by permno date;
run;

data no_accounting_data2/view=no_accounting_data2;
set no_accounting_data1 subset4(where=(missing(dummy)));
drop gvkey ulinkdt ulinkenddt dummy;
run;

proc sort data=no_accounting_data2 out=no_accounting_data3;
by permno date;
run;

data accounting_data2/view=accounting_data2;
set accounting_data1(drop=ulinkdt ulinkenddt);
by permno;
run;

proc sort data=accounting_data2 out=accounting_data3;
by gvkey quarter;
run;

proc datasets library=work;
```

```

modify accounting_data3;
index create gvkey;
delete accounting_data1;
run;

/*This is the end of the permno based gathering and combining of data*/

/*gather acct data*/
data acct_data/view=acct_data;
set comp.fundq(keep=datadate gvkey ceqq ATQ NIQ IBQ
where=(^missing(date)
and date>mdy(12,31,1961) and
date<mdy(1,1,2008))rename=(datadate=date));
by gvkey date;
if missing(NIQ) then NIQ=IBQ;
quarter=4*(year(date)-1929)+qtr(date);
label quarter="Quarter";
drop ibq;
run;

/*convert to ratios*/
data ratios1/view=ratios1;
/*create financial ratios*/
set acct_data;
by gvkey date;
/*Return on Assets*/
if (^missing(NIQ) and ^missing(ATQ) and (ATQ NE 0)) then roa=NIQ/ATQ;
else roa=.;
/*Debt to equity*/
if (^missing(atq) and ^missing(ceqq) and (ceqq NE
0)) then
de=atq/ceqq-1;

```

```

        else de=.;
label    de="Debt to Equity"
        roa="Return on Assets"
;
keep gvkey date roa de ceqq quarter;
run;

data ratios1a/view=ratios1a;
set ratios1(where=(^missing(roa) and ^missing(de) and ^missing(ceqq) and ^missing
run;

/*sort by quarter to rank*/
proc sort data=ratios1a out=ratios2;
by quarter;
run;

/*rank data*/
proc rank data=ratios2 percent ties=mean out=ratios3;
by quarter;
var roa ceqq de;
ranks rroa rceqq rde;
run;

/*assings ratios to macro variables*/
%let a1=roa;
%let a2=ceqq;
%let a3=de;

%macro tukey;

data ratios4/view=ratios4;

```

```

set ratios3;
by quarter;
%do i=1 %to 3;
if ^missing(r&&a&i) then do;
if r&&a&i>10 and r&&a&i<90 then &&a&i=0;
else if r&&a&i LE 10 then &&a&i=-1;
else &&a&i=1;
end;
%end;
drop r&a1 r&a2 r&a3 date;
run;

%mend;
%tukey;

proc sort data=ratios4 out=ratios5;
by gvkey quarter;
run;

proc sql;
create view accounting_data4 as
select a.*, b.de,b.roa,b.ceqq
from accounting_data3 as a
left join
ratios5 as b
on a.gvkey=b.gvkey and a.quarter=b.quarter;
quit;

data accounting_data5/view=accounting_data5;
set accounting_data4(drop=gvkey where=(date>mdy(13,31,1961) and date<mdy(1,1,2008)
run;

```

```

data total_set1/view=total_set1;
set no_accounting_data2 accounting_data5;
if operational=0 then do;
if dmax-date>365 then operational=1;
end;
run;

data total_set2/view=total_set2;
set total_set1(where=(^missing(quarter)));
run;

data sic1/view=sic1;
set crsp.dsf(keep=hsiccd permno date rename=(hsiccd=sic) where=(^missing(sic) and
by permno date;
if sic>5999 and sic<6800 then group="F";
    else group="N";
run;

proc sql;
create view total_set3 as
select a.*, b.group
from total_set2 as a left join sic1 as b
on a.permno=b.permno and a.date=b.date
order by permno, date;
quit;

%macro weekends;

data total_set4/view=total_set4;

```

```

set total_set3;
by permno date;
lpermno=lag91(permno);
lde=lag91(de);
lroa=lag91(roa);
lceqq=lag91(ceqq);
run;

data total_set5/view=total_set5;
set total_set4;
by permno date;
if permno=lpermno then do;
de=lde;
roa=lroa;
ceqq=lceqq;
end;
drop lpermno lde lroa lceqq;
run;

data total_set6/view=total_set6;
set total_set5;
by permno date;
%do i=1 %to 7;
lde&i=lag&i(de);
lgroup&i=lag&i(group);
lpermno&i=lag&i(permno);
lroa&i=lag&i(roa);
lceqq&i=lag&i(ceqq);
%end;
run;

data total_set7/view=total_set7;
set total_set6;

```

```

by permno date;
if missing(group) then do;
    test=0;
    %do i=1 %to 7;
%if %eval(&i)=1 %then %do;
    if test=0 and permno=lpermno&i and ^missing(lgroup&i) then do;
        de=lde&i;
        group=lgroup&i;
        roa=lroa&i;
        ceqq=lceqq&i;
        test=1;
    end;
%end;
%else %do;
else if test=0 and permno=lpermno&i and ^missing(lgroup&i) then do;
    de=lde&i;
    group=lgroup&i;
    roa=lroa&i;
    ceqq=lceqq&i;
    test=1;
end;
%end;
    %end;
end;
run;

%mend;
%weekends;

data full_model1 empty_model1;
set total_set7(keep=permno date de roa ceqq group operational);
if ^missing(de) and ^missing(group) and ^missing(roa) and ^missing(ceqq) then do;
output full_model1;

```

```
end;  
else do;  
output empty_model1;  
end;
```

```
proc sort data=full_model1 out=full_model2;  
by date;  
run;
```

```
proc datasets library=work;  
modify full_model2;  
index create date;  
run;
```

```
proc sort data=empty_model1 out=empty_model2;  
by date;  
run;
```

```
proc datasets library=work;  
modify empty_model2;  
index create date;  
run;
```

```
proc datasets library=work;  
delete empty_model1 full_model1;  
run;
```

```
proc freq data=full_model2(where=((date>mdy(12,31,1961) and date<mdy(7,1,1990))))  
tables de*roa*ceqq*group*operational/out=modela1;  
run;
```

```

data modela1;
set modela1;
date=mdy(6,30,1990);
run;

proc freq data=full_model2(where=((date>mdy(6,30,1990) and date<mdy(1,1,2008))))
tables date*de*roa*ceqq*group*operational/out=modela2;
run;

data modela3;
do date=mdy(6,30,1990) to mdy(12,31,2007);
do de=-1 to 1;
do roa=-1 to 1;
do ceqq=-1 to 1;
do i=1 to 2;
if i=1 then group="F";
else group="N";
do operational=0 to 1;
output;
end;
end;
end;
end;
end;
run;

data modela4/view=modela4;
set modela1 modela2;
run;

proc sort data=modela3 out=modela5 presorted;

```

```
by date de roa ceqq group operational;  
run;
```

```
proc sort data=modela4 out=modela6;  
by date de roa ceqq group operational;  
run;
```

```
data modela7;  
merge modela5 modela6;  
by date de roa ceqq group operational;  
if missing(count) then count=0;  
if operational=0 then merged=count;  
else going_concern=count;  
drop count;  
run;
```

```
proc datasets library=work;  
delete modela1-modela3 modela5-modela6;  
run;
```

```
data modela8a/view=modela8a;  
set modela7(where=(operational=0));  
by date de roa ceqq group;  
drop going_concern operational;  
run;
```

```
data modela8b/view=modela8b;  
set modela7(where=(operational=1));  
by date de roa ceqq group;  
drop merged operational;  
run;
```

```
data modela9/view=modela9;
```

```

merge modela8a modela8b;
by date de roa ceqq group;
if date=mdy(6,30,1990) then do;
merged=merged+1;
going_concern=going_concern+15;
end;
run;

%macro cumulative;

%let count=0;

data modela10/view=modela10;
set modela9;
by date de roa ceqq group;
retain alpha1-alpha54 beta1-beta54 0;
%do i=-1 %to 1;
%do j=-1 %to 1;
%do k=-1 %to 1;
%do m=1 %to 2;
%let count=%eval(&count+1);
%if %eval(&i)=-1 & %eval(&j)=-1 & %eval(&k)=-1 & %eval(&m)=1 %then %do;
if de=%eval(&i) and roa=%eval(&j) and ceqq=%eval(&k) and group="F" then do;
if alpha&count+beta&count>0 then do;
expectation=alpha&count/(alpha&count+beta&count);
end;
alpha&count=alpha&count+merged;
beta&count=beta&count+going_concern;
expectation=alpha&count/(alpha&count+beta&count);
end;

```

```

%end;
%else %do;
else if de=%eval(&i) and roa=%eval(&j) and ceqq=%eval(&k) and group="F" then do;
    if alpha&count+beta&count>0 then
        expectation=alpha&count/(
end;
    alpha&count=alpha&count+m
    beta&count=beta&count+goi
    expectation=alpha&count/(
end;
else if de=%eval(&i) and roa=%eval(&j) an
    if alpha&count+beta&count>0 then
        expectation=alpha&count/(
end;
    alpha&count=alpha&count+m
    beta&count=beta&count+goi
    expectation=alpha&count/(
end;

%end;
%end;
%end;
%end;
%end;
run;

%mend;

%cumulative;

```

```

data modela11/view=modela11;
set modela10(where=(merged+going_concern>0) drop=alpha1-alpha54 beta1-beta54 perc
by date de roa ceqq group;
run;

```

```

proc freq data=empty_model2(where=((date>mdy(12,31,1961) and date<mdy(7,1,1990)))
tables operational/out=modelb1;
run;

```

```

data modelb1;
set modelb1;
date=mdy(6,30,1990);
run;

```

```

proc freq data=empty_model2(where=((date>mdy(6,30,1990) and date<mdy(1,1,2008))))
tables date*operational/out=modelb2;
run;

```

```

data modelb3;
do date=mdy(6,30,1990) to mdy(12,31,2007);
            do operational=0 to 1;
                output;
            end;
end;
run;

```

```

data modelb4/view=modelb4;
set modelb1 modelb2;
run;

```

```

proc sort data=modelb3 out=modelb5 presorted;
by date operational;

```

```
run;

proc sort data=modelb4 out=modelb6;
by date operational;
run;

data modelb7;
merge modelb5 modelb6;
by date operational;
if missing(count) then count=0;
if operational=0 then merged=count;

else going_concern=count;
drop count;
run;

proc datasets library=work;
delete modelb1-modelb3 modelb5-modelb6;
run;

data modelb8a/view=modelb8a;
set modelb7(where=(operational=0));
by date;
drop going_concern operational;
run;

data modelb8b/view=modelb8b;
set modelb7(where=(operational=1));
by date;
drop merged operational;
run;

data modelb9/view=modelb9;
```

```
merge modelb8a modelb8b;
by date;
if date=mdy(6,30,1990) then do;
merged=merged+1;
going_concern=going_concern+15;
end;
run;
```

```
data modelb10/view=modelb10;
set modelb9;
by date;
retain alpha beta 0;
if alpha+beta>0 then do;
expectation=alpha/(alpha+beta);
end;
alpha=alpha+merged;
beta=beta+going_concern;
run;
```

```
data modelb11/view=modelb11;
set modelb10(where=(merged+going_concern>0) drop=alpha beta percent);
by date;
run;
```

```

data combined_sample/view=combined_sample;
set modela11(keep=date merged going_concern expectation) modelb11(keep=date merge
going_concern expectation);
if missing(merged) then merged=0;
if missing(going_concern) then going_concern=0;
if missing(expectation) then delete;
run;

```

```

data solution/view=solution;
set combined_sample(where=(date>mdy(6,30,1990) and date<mdy(1,1,2008)));
retain count number omega 1 df delta 0;
count=count+1;
g_bar=merged/(merged+going_concern)-expectation;
if merged+going_concern=1 then s_squared=0;
else s_squared=merged*going_concern/(merged+going_concern)/(merged+going_concern-
sum_of_squares=omega*df+s_squared*(merged+going_concern-1)+1/(1/number+1/(merged+
delta=(number*delta+(merged+going_concern)*g_bar)/(number+merged+going_concern);
number=number+merged+going_concern;
df=df+merged+going_concern;
omega=sum_of_squares/df;
run;

```

```

proc sql;
    create view final1 as
    select *, max(count) as maximum_count
    from solution;
quit;

proc print data=final1(where=(count=maximum_count));
run;

```

## B.10 Posterior Test for Loss Given Merger

```

/*options mprint symbolgen mlogic;*/

/*
data filter1;
set crsp.dsf(keep=permno date prc cfacpr where=(date<mdy(1,1,12008) and rand('uni
by permno date;
run;
*/

data filter1/view=filter1;
set crsp.dsf(keep=permno date prc cfacpr where=(date<mdy(1,1,12008)));
by permno date;
run;

proc summary data=filter1;
by permno;
var date;
output out=d_end max(date)=d_end;
run;

proc summary data=filter1(where=(prc>0 and ~missing(prc)));
by permno;
var date;
output out=date_bounds min(date)=dmin max(date)=dmax;
run;

/*restricts data to common shares only and assigns final status code*/
data common_only/view=common_only;

```

```

set crsp.dseall(keep=permno dlstcd shrcd dlamt nwperm where=(^missing(dlstcd)));
by permno;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
      else if substr(left(shrcd),2,1)='5' then delete;
      else if substr(left(shrcd),2,1)='8' then delete;
      else if substr(left(shrcd),1,1)='2' then delete;
      else if substr(left(shrcd),1,1)='4' then delete;
      else if substr(left(shrcd),1,1)='7' then delete;
else do;
operational=0;
merge=0;
cash_buyout=0;
private=0;
bankrupt=0;
if dlstcd=100 then operational=1;
      else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
      else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;
      else if dlstcd in(502,510,513,514,516,517,519,520,573,575,580,581,582,583)
      else if dlstcd in(574) then bankrupt=1;
      else if dlstcd in(234,242,243,251,252,261,262,300,332,341,342,400,450,460)
            if dlstcd in(234,500) then do;
                  if nwperm>0 then merge=1;
                  else if nwperm=0 and dlamt>0 then cash_buyout=1;
                  else private=1;
            end;
      else if dlstcd in (242,243,251,341,342) then do;
            if nwperm>0 then merge=1;
            else cash_buyout=1;
      end;
      else if dlstcd in (252,261,262,450) then do;
            if nwperm>0 then merge=1;
            else if dlamt>0 then cash_buyout=1;

```

```

                else delete;
            end;
        else if dlstcd in(300) then do;
            if nwperm>0 then merge=1;
            else delete;
        end;
    else if dlstcd in (332) then do;
        if nwperm>0 then merge=1;
        else private=1;
    end;
    else if dlstcd in (400,460,470) then do;
        if dlamt>0 then cash_buyout=1;
        else bankrupt=1;
    end;
end;
end;
run;

data filter2/view=filter2;
merge filter1(in=aa) d_end(keep=permno d_end) date_bounds(in=cc keep=permno dmin
by permno;
if aa and bb and cc;
run;

data filter3/view=filter3;
set filter2(where=(d_end-dmax<366 and d_end-date<366));
by permno;
if prc>0 then lower_distance=dmax-date;
run;

proc sql;
create view filter4 as
select a.*,min(a.lower_distance) as minimum_lower_distance

```

```

from filter3 as a
where ^missing(lower_distance)
group by permno
order by permno;
quit;

```

```

data filter5/view=filter5;
set filter4(where=(minimum_lower_distance<7 and prc>0));
by permno;
prc=prc/cfacpr;
drop cfacpr;
run;

```

```

proc sort data=filter5 out=filter6 nodupkey;
by permno;
run;

```

```

proc sql;
create view new_firm1 as
select a.permno, a.date, a.prc,a.cfacpr,b.permno as old_perm, b.d_end,b.dmax
from crsp.dsf as a right join filter6 as b
on a.permno=b.nwperm and a.date>b.d_end and a.date<b.d_end+365;
quit;

```

```

data new_firm2/view=new_firm2;
set new_firm1(where=(prc>0 and ^missing(prc)));
upper_distance=date-d_end;

```

```
run;

proc sql;
create view new_firm3 as
select a.*, min(a.upper_distance) as minimum_upper_distance
from new_firm2 as a
group by permno
order by permno;
quit;

data new_firm4/view=new_firm4;
set new_firm3(where=(minimum_upper_distance<8));
by permno;
prc=prc/cfacpr;
drop cfacpr;
run;

data extract_upper_price/view=extract_upper_price;
set new_firm4(where=(upper_distance=minimum_upper_distance));
upper_price=prc;
run;

data extract_lower_price/view=extract_lower_price;
set filter5(where=(date=dmax));
lower_price=prc;
run;

data combine_set1/view=combine_set1;
set filter5 new_firm4;
run;

proc sql;
create view combine_set2 as
```

```

select a.*,b.lower_price,c.upper_price
from combine_set1 as a, extract_lower_price as b, extract_upper_price as c
where a.permno=b.permno and a.permno=c.permno;
quit;

```

```

data combine_set3/view=combine_set3;
set combine_set2(keep=date dlamt permno prc cash_buyout d_end
lower_price merge upper_price);
if cash_buyout = 1 and date>d_end then delete;
else if date>d_end then prc=prc*lower_price/upper_price;
drop lower_price upper_price;
run;

```

```

data cash_return/view=cash_return;
set combine_set3(where=(cash_buyout=1));
return=dlamt/prc;
drop cash_buyout dlamt prc;
run;

```

```

proc sort data=combine_set3 (where=(date<mdy(1,1,2008))) out=storage366 nodupkey;
by permno date;
run;

```

```

proc datasets library=work;
modify storage366;
index create permno;
run;

```

```

%macro conditional_merge;

```

```

%do i=365 %to 352 %by -1;
%let j=%eval(&i+1);

data dummy&i/view=dummy&i;
set storage&j;
by permno date;
date=date-&i;
prc&i=prc;
drop prc;
run;

proc sort data=dummy&i out=lead&i presorted;
by permno date;
run;

data lead&i/view=lead&i;
merge storage&j(in=aa) dummy&i(in==bb);
by permno date;
if aa and bb;
run;

data first_join&i/view=first_join&i;
merge storage&j(in=aa where=(prc>0)) lead&i(where=(prc&i>0));
by permno date;
if aa and ^missing(prc&i) then return=prc&i/prc;
else return=.;
run;

data return&i/view=return&i;
set first_join&i(where=(^missing(return)));
by permno date;
run;

```

```
data storage&i/view=storage&i;
set first_join&i(where=(missing(return)));
by permno date;
run;

%end;

%mend;

%conditional_merge;

data joint_return1/view=joint_return1;
set return352-return365 cash_return;
keep date return;
run;

proc sort data=joint_return1 out=joint_return2;
by date;
run;

proc datasets library=work;
modify joint_return2;
index create date;
delete date_bounds d_end filter1 filter6 storage366 lead352-lead365;
run;

%macro loss_class;

data loss_class1/view=loss_class1;
set joint_return2;
```

```

by date;
if return GE 1 then do;
loss0=1;
end;
%do i=1 %to 100;
else if 1-return<&i/100 and 1-return GE (&i-1)/100 then do;
loss&i=1;
end;
%end;
run;

proc summary data=loss_class1;
by date;
var loss0-loss100;
output out=loss_class2 sum(loss0-loss100)=/autoname;
run;

data loss_class3/view=loss_class3;
set loss_class2;
by date;
retain cumulative_loss0-cumulative_loss100 total 0;
%let k=_Sum;
%do i=0 %to 100;
cumulative_loss&i=sum(cumulative_loss&i,loss&i&k);
%end;
total=sum(total,_FREQ_);
run;

data loss_class4/view=loss_class4;
set loss_class3(keep=date cumulative_loss0-cumulative_loss100 total);
by date;
%do i=0 %to 100;
p&i=cumulative_loss&i/total;

```

```

%end;
keep date p0-p100;
run;

data expectation1/view=expectation1;
set loss_class4;
by date;
expectation=0;
%do i=1 %to 100;
expectation=(&i/100-.005)*p&i+expectation;
%end;
keep date expectation;
run;

data expectation2/view=expectation2;
set joint_return2;
by date;
if return ge 1 then loss=0;
else loss=1-return;
run;

data expectation4/view=expectation4;
merge expectation1(keep=date expectation) expectation2(keep=date loss);
by date;
if date<mdy(7,1,1990) then delete;
else if date>mdy(12,31,2007) then delete;
else delta=loss-expectation;
run;

%mend;
%loss_class;

```

```

data expectation5/view=expectation5;
set expectation4(keep=date delta);
by date;
if missing(delta) then delete;
sigma=delta*delta;
run;

proc summary data=expectation5;
var delta sigma;
output out=posterior1 mean(delta)=x_bar sum(sigma)=s_squares n(delta)=number;
run;

data posterior2;
set posterior1;
n_one=1+number;
mu=number*x_bar/n_one;
s1=1+s_squares+x_bar*x_bar/(1+1/number);
std=SQRT(s1/number);
t_score=mu/(std/SQRT(n_one));
chi_square=(s1-n_one)/SQRT(2*n_one);
run;

proc print data=posterior2;
run;

```

## B.11 Posterior Test for Loss Given Going Concern

```

/*options mprint symbolgen mlogic;*/
/*options nosource nonotes;*/

```

```

/*
data filter1;
set crsp.dsf(keep=permno date prc cfacpr where=(date<mdy(1,1,12008) and rand('uni
by permno date;
run;
*/

data filter1/view=filter1;
set crsp.dsf(keep=permno date prc cfacpr where=(date<mdy(1,1,12008)));
by permno date;
run;

proc summary data=filter1;
by permno;
var date;
output out=d_end max(date)=d_end;
run;

/*restricts data to common shares only and assigns final status code*/
data common_only/view=common_only;
set crsp.dseall(keep=permno dlstcd shrcd dlamt nwperm where=(^missing(dlstcd)));
by permno;
if substr(left(shrcd),2,1)='3' then delete;
else if substr(left(shrcd),2,1)='4' then delete;
      else if substr(left(shrcd),2,1)='5' then delete;
      else if substr(left(shrcd),2,1)='8' then delete;
      else if substr(left(shrcd),1,1)='2' then delete;
      else if substr(left(shrcd),1,1)='4' then delete;
      else if substr(left(shrcd),1,1)='7' then delete;
else do;
operational=0;

```

```

merge=0;
cash_buyout=0;
private=0;
bankrupt=0;
if dlstcd=100 then operational=1;
    else if dlstcd in(200,231,232,241,244,301,331,343) then merge=1;
    else if dlstcd in(233,235,271,333,334,361) then cash_buyout=1;
    else if dlstcd in(502,510,513,514,516,517,519,520,573,575,580,581,582,583)
    else if dlstcd in(574) then bankrupt=1;
    else if dlstcd in(234,242,243,251,252,261,262,300,332,341,342,400,450,460)
        if dlstcd in(234,500) then do;
            if nwperm>0 then merge=1;
            else if nwperm=0 and dlamt>0 then cash_buyout=1;
            else private=1;
        end;
    else if dlstcd in (242,243,251,341,342) then do;
        if nwperm>0 then merge=1;
        else cash_buyout=1;
    end;
    else if dlstcd in (252,261,262,450) then do;
        if nwperm>0 then merge=1;
        else if dlamt>0 then cash_buyout=1;
        else delete;
    end;
    else if dlstcd in(300) then do;
        if nwperm>0 then merge=1;
        else delete;
    end;
    else if dlstcd in (332) then do;
        if nwperm>0 then merge=1;
        else private=1;
    end;
    else if dlstcd in (400,460,470) then do;

```

```

                if dlamt>0 then cash_buyout=1;
                else bankrupt=1;
            end;
        end;
    end;
end;
run;

data filter2/view=filter2;
merge filter1(in=aa) d_end(keep=permno d_end) common_only(in=bb);
by permno;
if aa and bb;
run;

data filter3/view=filter3;
set filter2(where=(~missing(prc) and prc>0));
by permno;
if operational=0 and d_end-date<366 then delete;
else prc=prc/cfacpr;
keep date permno prc;
run;

proc sort data=filter3 (where=(date<mdy(1,1,2008))) out=storage366 nodupkey;
by permno date;
run;

proc datasets library=work;
modify storage366;
index create permno;
run;

```

```

%macro conditional_merge;

%do i=365 %to 352 %by -1;
%let j=%eval(&i+1);

data dummy&i/view=dummy&i;
set storage&j;
by permno date;
date=date-&i;
prc&i=prc;
drop prc;
run;

proc sort data=dummy&i out=lead&i presorted;
by permno date;
run;

data lead&i/view=lead&i;
merge storage&j(in=aa) dummy&i(in==bb);
by permno date;
if aa and bb;
run;

data first_join&i/view=first_join&i;
merge storage&j(in=aa where=(prc>0)) lead&i(where=(prc&i>0));
by permno date;
if aa and ^missing(prc&i) then return=prc&i/prc;
else return=.;
run;

data return&i/view=return&i;
set first_join&i(where=(^missing(return)));
by permno date;

```

```
run;

data storage&i/view=storage&i;
set first_join&i(where=(missing(return)));
by permno date;
run;

%end;

%mend;

%conditional_merge;

data joint_return1/view=joint_return1;
set return352-return365;
keep date return;
run;

proc sort data=joint_return1(where=(^missing(date) and ^missing(return))) out=joi
by date;
run;

proc datasets library=work;
modify joint_return2;
index create date;
delete date_bounds d_end filter1 filter6 storage366 lead352-lead365;
run;

%macro loss_class;
```

```

data loss_class1/view=loss_class1;
set joint_return2;
by date;
if return GE 1 then do;
loss0=1;
end;
%do i=1 %to 100;
else if 1-return<&i/100 and 1-return GE (&i-1)/100 then do;
loss&i=1;
end;
%end;
run;

proc summary data=loss_class1;
by date;
var loss0-loss100;
output out=loss_class2 sum(loss0-loss100)=/autoname;
run;

data loss_class3/view=loss_class3;
set loss_class2;
by date;
retain cumulative_loss0-cumulative_loss100 total 0;
%let k=_Sum;
%do i=0 %to 100;
cumulative_loss&i=sum(cumulative_loss&i,loss&i&k);
%end;
total=sum(total,_FREQ_);
run;

data loss_class4/view=loss_class4;
set loss_class3(keep=date cumulative_loss0-cumulative_loss100 total);
by date;

```

```

%do i=0 %to 100;
p&i=cumulative_loss&i/total;
%end;
keep date p0-p100;
run;

```

```

data expectation1/view=expectation1;
set loss_class4;
by date;
expectation=0;
%do i=1 %to 100;
expectation=(&i/100-.005)*p&i+expectation;
%end;
keep date expectation;
run;

```

```

data expectation2/view=expectation2;
set joint_return2;
by date;
if return ge 1 then loss=0;
else loss=1-return;
run;

```

```

data expectation4/view=expectation4;
merge expectation1(keep=date expectation) expectation2(keep=date loss);
by date;
if date<mdy(7,1,1990) then delete;
else if date>mdy(12,31,2007) then delete;
else delta=loss-expectation;
run;

```

```

%mend;
%loss_class;

data expectation5/view=expectation5;
set expectation4(keep=date delta);
by date;
if missing(delta) then delete;
sigma=delta*delta;
run;

proc summary data=expectation5;
var delta sigma;
output out=posterior1 mean(delta)=x_bar sum(sigma)=s_squares n(delta)=number;
run;

data posterior2;
set posterior1;
n_one=1+number;
mu=number*x_bar/n_one;
s1=1+s_squares+x_bar*x_bar/(1+1/number);
std=SQRT(s1/number);
t_score=mu/(std/SQRT(n_one));
chi_square=(s1-n_one)/SQRT(2*n_one);
run;

proc print data=posterior2;
run;

```

# BIBLIOGRAPHY

- [1] Ashok Abbott. *Valuation Handbook*, chapter Measures of Discount for Lack of Marketability and Liquidity, pages 474–507. Wiley Finance, Hoboken, NJ, 2009.
- [2] Anup Agrawal and Charles R. Knoeber. Managerial compensation and the threat of takeover. *Journal of Financial Economics*, 47(2):219–239, February 1998.
- [3] Edward I. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609, Sep 1968.
- [4] Brent W. Ambrose and William L. Megginson. The role of asset structure, ownership structure and takeover defenses in determining acquisition likelihood. *Journal of Financial and Quantitative Analysis*, 27(4):575–589, December 1992.
- [5] Yakov Amihud and Haim Mendelson. Dealership market: Market making with inventory. *Journal of Financial Economics*, 8:31–53, 1980.
- [6] Amber Ananad and Daniel G. Weaver. The value of the specialist: Empirical evidence from the cboe. *Journal of Financial Markets*, 9:100–118, 2006.
- [7] Tomohiro Ando. Bayesian inference for the hazard term structure with functional predictors using bayesian predictive information criteria. *Computational Statistics and Data Analysis*, 53(6):1925–1939, 15 April 2009.
- [8] Thomas Astebro and Joachim K. Winter. More than a dummy: The probability of failure, survival and acquisition of firms in financial distress. *European Management Review*, 9(1):1–17, Spring 2012.
- [9] Louis Bachelier, M.H.A Davis, and Alison Etheridge. *Théorie de la spéculation. English: Louis Bachelier's theory of speculation the origins of modern finance*. Princeton University Press, Princeton, NJ, 2006.

- [10] W.D. Baten. The probability law for the sum of  $n$  independent variables, each subject to the law  $(\frac{1}{(2h)})\operatorname{sech}(\pi x/(2h))$ . *Bulletin of the American Mathematical Society*, 40(4):284–290, April 1934.
- [11] Elena Beccalli and Pascal Frantz. The determinants of mergers and acquisitions in banking. *Journal of Financial Services Research*, 43(2013):265–291, June 2013.
- [12] Daniel Berg. Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23(2):129–143, 2007.
- [13] James O. Berger. *Statistical Decision Theory, Foundations, Concepts, and Methods*. Springer Series in Statistics. Springer-Verlag, New York, 1980.
- [14] Matthew T. Billett. Targeting capital structure: The relationship between risky debt and the firm’s likelihood of being acquired. *The Journal of Business*, 69(2):173–192, April 1996.
- [15] Fischer Black. Capital market equilibrium with restricted borrowing. *Journal of Business*, 45(3):444–455, 1972.
- [16] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, May-June 1973.
- [17] A.J. Boness. Elements of a theory of stock option value. *Journal of Political Economy*, 72:163–75, April 1973.
- [18] Imad Bou-Hamad, Denis Larocque, and Hatem Ben-Ameur. Discrete-time survival trees and forests with time-varying covariates; application to bankruptcy data. *Statistical Modelling*, 11(5):429–446, 2011.
- [19] Krzysztof Burdzy. *Excursions of Complex Brownian Motion*. University of California, Berkeley, 1984.
- [20] Robert B. Carton. *Measuring Organizational Performance: An Exploratory Study*. PhD thesis, University of Georgia, Athens, Georgia, 2004.
- [21] Lai K. Chan. Linear estimation of the location and scale parameters of the cauchy distribution based on sample quantiles. *Journal of the American Statistical Association*, 65(330):851–859, June 1970.
- [22] Z. H. Che, H. S. Wang, and Chih-Ling Chuang. A fuzzy ahp and dea approach for making bank loan decisions for small and medium enterprises in taiwan. *Expert Systems with Applications*, 37(10):7189–7199, October 2010.

- [23] Der-Jang Chi, Ching-Chiang Yeh, and Ming-Cheng Lai. A hybrid approach of dea, rough set theory and random forests for credit rating. *International Journal of Innovative Computing Information and Control*, 7(8):4885–4897, August 2011.
- [24] John Bates Clark. *The Distribution of Wealth: A Theory of Wages, Interest and Profits*. The MacMillan Company, New York, 1908.
- [25] John H. Cochrane. *Asset Pricing*. Princeton University Press, Princeton, NJ, revised edition edition, 2005.
- [26] Philip L. Cooley. Bayesian and cost considerations for optimal classification with discriminant analysis. *The Journal of Risk and Insurance*, 42(2):277–287, June 1975.
- [27] Harry Zvi Davis and Yoram C. Peles. Measuring equilibrating forces of financial ratios. *The Accounting Review*, 68(4):725–747, October 1993.
- [28] Douglas W. Diamond. Banks and liquidity creation: A simple exposition of the diamond-dybvig model. *Economic Quarterly*, 93(2):189–200, Spring 2007.
- [29] Douglas W Diamond and Philip H Dybvig. Bank runs, deposit insurance, and liquidity. *The Journal of Political Economy*, 91(3):401–419, 6 1983.
- [30] MD Donsker. An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society*, 6, 1951-52.
- [31] Darrell Duffie, Leandro Saita, and Ke Wang. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3):635–665, March 2007.
- [32] F Y Edgeworth. The mathematical theory of banking. *Journal of the Royal Statistical Society*, 51(1):113–127, 1888.
- [33] Eugene F. Fama and Richard Roll. Some properties of symmetric stable distributions. *Journal of the American Statistical Association*, 63(323):pp. 817–836, Sep 1968.
- [34] Stephen E. Fienberg. When did bayesian inference become “bayesian?”. *Bayesian Analysis*, 1(1):1–40, 2006.
- [35] Irving Fisher. *The Theory of Interest: As Determined by Impatience to Spend Income and Opportunity to Invest It*. The MacMillan Company, New York, 1930.

- [36] Xavier Freixas and Jean-Charles Rochet. *Microeconomics of Banking*. Massachusetts Institute of Technology, Cambridge, MA, 1997.
- [37] R.C. Geary. The frequency distribution of the quotient of two normal variates. *Journal of the Royal Statistical Society*, 93(3):442–446, 1930.
- [38] Adrian Gepp, Kuldeep Kumar, and Sukanto Bhattacharya. Business failure prediction using decision trees. *Journal of Forecasting*, 29(6):536–555, September 2010.
- [39] John Geweke. *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc., Hoboken, NJ, 2005.
- [40] Michael A. Goldstein and Kenneth A. Kavajecz. Trading strategies during circuit breakers and extreme market movements. *Journal of Financial Markets*, 7:301–333, 2004.
- [41] James M. Griffin and Steven N. Wiggins. Takeovers-managerial incompetence or managerial shirking. *Economic Inquiry*, 30(2):355–370, April 1992.
- [42] John Gurland. Inversion formulae for the distribution of ratios. *The Annals of Mathematical Statistics*, 19(2):228–237, June 1948.
- [43] Frank Hampel. The proper fiducial argument. Seminar für Statistik–ETH Zentrum, February 2003.
- [44] J.R. Hicks. *Value and Capital: an inquiry into some fundamental principles of economic theory*. Clarendon Press, Oxford, 1939.
- [45] Kiyoshi Itô. Stochastic integral. In *Proceedings of the Imperial Academy of Tokyo*, volume 20, pages 519–524, 1944.
- [46] E. T. Jaynes. *Probability Theory: The Language of Science*. Cambridge University Press, Cambridge, 2003.
- [47] Michael C. Jensen. Takeovers: Their causes and consequences. *The Journal of Economic Perspectives*, 2(1):21–48, Winter 1988.
- [48] B.K. Kale. Inadmissibility of the maximum likelihood estimator in the presence of prior information. *Canadian Mathematical Bulletin*, 13(3):391–393, September 1970.
- [49] John Maynard Keynes. *The General Theory of Employment, Interest and Money*. Harcourt, Brace, New York, 1936.
- [50] R. F. Krampf and J. D. Williams. Multidimensional scaling as a research tool - explanation and application. *Journal of Business Research*, 2(2):157–175, 1974.

- [51] Vernon D. Landon. The distribution of amplitude with time in fluctuation noise. *Proceedings IRE*, 29(1):50–54, 1941.
- [52] Brian J. Taylor Laura L. Pullum and Marjorie A. Darrah. *Guidance for the Verification and Validation of Neural Networks*. Wiley IEEE Computer Society, March 2007.
- [53] Michael Lavine and Mark J Schervish. Bayes factors: What they are and what they are not. *The American Statistician*, 53(2):119–122, May 1999.
- [54] Kyuon Eun Lee; and Jae Woo Lee. Scaling properties of price changes for korean stock indices. *Journal of the Korean Physical Society*, 44(3):668–671, March 2004.
- [55] Peter M. Lee. *Bayesian Statistics: An Introduction*. Arnold; Wiley, London; New York, 3rd edition, 2004.
- [56] Lawrence M. Leemis and Jacquelyn T. McQueston. Univariate distribution relationships. *The American Statistician*, 62(1):45–53, February 2008.
- [57] John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1):13–37, 1965.
- [58] Qing Liu and Larry D. Qiu. Characteristics of acquirers and targets in domestic and cross-border mergers and acquisitions. *Review of Development Economics*, 17(3):474–493, 2013.
- [59] Stephen A Hillegeist.; Elizabeth K. Keating; Donald P. Cram; Kyle G. Lundstedt. Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9:5–34, 2004.
- [60] Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.
- [61] HB Mann and A Wald. On the statistical treatment of linear stochastic difference equations. *Econometrica*, 11:173–200, 1943.
- [62] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, March 1952.
- [63] H.M. Markowitz and N. Usemen. The likelihood of various stock market return distributions, part 1: principles of inference. *Journal of Risk and Uncertainty*, 13:207–219, 1996.
- [64] Stewart Mayhew. Competition, market structure, and bid-ask spreads in stock option markets. *The Journal of Finance*, 57(2):931–958, April 2002.

- [65] Jan Mossin. Equilibrium in a capital asset market. *Econometrica*, 34(4):768–783, 1966.
- [66] Frederick Mosteller. On some useful “inefficient” statistics. *The Annals of Mathematical Statistics*, 17(4):377–408, 1946.
- [67] Kevin P Murphy. Conjugate bayesian analysis of the gaussian distribution. <http://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf>, October 2007.
- [68] Robert F. Nau. De finetti was right: Probability does not exist. *Theory and Decision*, 51:89–124, 2001.
- [69] Salih N Neftci. *An introduction to the mathematics of financial derivatives*. Academic Press, San Diego, 2000.
- [70] Michael Nwogugu. Decision-making, risk and corporate governance: A critique of methodological issues in bankruptcy/recovery prediction models. *Applied Mathematics and Computation*, 185:178–196, 2007.
- [71] Bank of International Settlements. Bis quarterly review, september 2012. Technical report, Bank of International Settlements, 2012.
- [72] Maureen O’Hara. Overview: Market structure issues in market liquidity. BIS Papers 2, Bank for International Settlements, 2000.
- [73] Yohei Okada and Hiroshi Konno. Failure discrimination by semi-definite programming using a maximal margin ellipsoidal surface. *Journal of Computational Finance*, 12(3):63–77, Spring 2009.
- [74] M.F.M. Osborne. Brownian motion in the stock market. *Operations Research*, 7(2):145–173, March/April 1959.
- [75] Krishna G. Palepu. Predicting takeover targets: A methodological and empirical analysis. *Journal of Accounting and Economics*, 8:3–35, 1986.
- [76] Giovanni Parmigiani and Lurdes Inoue. *Decision Theory: Principles and Approaches*. Wiley Series in Probability and Statistics. Wiley, Chichester, West Sussex, 2009.
- [77] E.S. Pearson. Statistical concepts in the relation to reality. *Journal of the the Royal Statistical Society. Series B(Methodological)*, 17(2):204–207, 1955.
- [78] Boris Podobnik, Davor Horvatic, Alexander M. Petersen, Branko Urošević, and H. Eugene Stanley. Bankruptcy risk model and empirical tests. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43):18325–18330, OCT 26 2010.

- [79] Boris Podobnik, Aljosa Valentincic, Davor Horvatic, and H. Eugene Stanley. Asymmetric levy flight in financial ratios. *Proceedings of the National Academy of Sciences of the United States of America*, 108(44):17883–17888, November 2011.
- [80] I. M. Premachandra, Gurmeet Singh Bhabra, and Toshiyuki Sueyoshi. Dea as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research*, 193(2):412–424, March 2009.
- [81] I. M. Premachandra, Yao Chen, and John Watson. Dea as a tool for predicting corporate failure and success: A case of bankruptcy assessment. *Omega-International Journal of Management Science*, 39(6):620–626, December 2011.
- [82] Stephen Ross. Arbitrage theory of capital asset pricing. *Journal of Economic Theory*, pages 341–360, 1976.
- [83] T. J. Rothenberg, F. M. Fisher, and C. B. Tilanus. A note on estimation from a cauchy sample. *Journal of the American Statistical Association*, 59(306):460–463, 1964.
- [84] A.D. Roy. Safety first and the holding of assets. *Econometrica*, 20:431–439, 1952.
- [85] Paul A. Samuelson. Rational theory of warrant pricing. *Industrial Management Review*, 6:13–31, 1965.
- [86] L. J. Savage. *The Foundations of Statistics*. John Wiley & Sons, New York, 1954.
- [87] Pranab Kumar Sen. Estimates of the regression coefficient based on kendall's tau. *Journal of the American Statistical Association*, 63(324):1379–1389, December 1968.
- [88] William E. Sharpe. A simplified model of portfolio analysis. *Management Science*, 9:277–293, 1963.
- [89] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, September 1964.
- [90] Lawrence E. Shepard and Robert A. Collins. Why do farmers fail? farm bankruptcies 1910-1978. *American Journal of Agricultural Economics*, 64(4):609–615, November 1982.

- [91] Udaya Shetty, T. P. M. Pakkala, and T. Mallikarjunappa. A modified directional distance formulation of dea to assess bankruptcy: An application to it/ites companies in india. *Expert Systems with Applications*, 39(2):1988–1997, February 2012.
- [92] Abner Shimony. Coherence and the axioms of confirmation. *Journal of Symbolic Logic*, 20:1–28, 1955.
- [93] Eugen Slutsky. The summation of random causes as the source of cyclic processes. *Econometrica*, 5(2), Apr. 1937.
- [94] C. Sprenkle. Warrant prices as an indicator of expectation. *Yale Economic Essays*, 1:412–474, 1964.
- [95] Charles Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 197–206. University of California Press, 1956.
- [96] Stephen M Stigler. Studies in the history of probability and statistics. xxxiii: Cauchy and the witch of agnesi: An historical note on the cauchy distribution. *Biometrika*, 61(2):375–380, 1974.
- [97] Kay Giesecke; Francis A Longstaff; Stephen Schaefer; Ilya Strebulaev. Corporate bond default risk: a 150-year perspective. *Journal of Financial Economics*, 102:233–250, July 2011.
- [98] Toshiyuki Sueyoshi and Mika Goto. Dea-da for bankruptcy-based performance assessment: Misclassification analysis of japanese construction industry. *European Journal of Operational Research*, 199(2):576–594, December 1 2009.
- [99] Toshiyuki Sueyoshi and Mika Goto. Methodological comparison between dea (data envelopment analysis) and dea-da (discriminant analysis) from the perspective of bankruptcy assessment. *European Journal of Operational Research*, 199(2):561–575, December 1 2009 2009.
- [100] Lili Sun and Prakash P. Shenoy. Using bayesian networks for bankruptcy prediction: Some methodological issues. *Computing, Artificial Intelligence and Information Management*, 180:738–753, 2007.
- [101] H. Theil. A rank invariant method of linear and polynomial regression (i,ii,iii). In *Proceedings of the Koninklijke Nederlandse Academie van Wetenschappen. Series A. Mathematical Sciences.*, pages 386–392, 521–525, 1397–1412, 1950.

- [102] Gary Koop; Dale J. Poirier; Justin L Tobias. *Bayesian Econometric Methods*, volume 7 of *Econometric Exercises*. Cambridge University Press, New York, 2007.
- [103] James Tobin. Liquidity preference as a behavior toward risk. *Review of Economic Studies*, 25:65–86, 1958.
- [104] Jack Treynor. Toward a theory of market value of risky assets. 1962.
- [105] Jack Treynor. Implications for the theory of finance. 1963.
- [106] Yiuman Tse and Tatyana Zabolina. Do designated market makers improve liquidity in open-outcry futures markets? *The Journal of Futures Markets*, 24(5):479–502, 2004.
- [107] Che-Tsung Tung and Yu-Je Lee. A novel approach to construct grey principal component analysis evaluation model. *Expert Systems with Applications*, 36(3):5916–5920, April 2009. PT: J; TC: 4; UT: WOS:000263817100022.
- [108] Thorstein Veblen. *The Theory of the Leisure Class: An Economic Study in the Evolution of Institutions*. The Macmillan Company, New York, 1899.
- [109] Thorstein Veblen. *The Theory of the Business Enterprise*. C. Scribner's Sons, New York, 1904.
- [110] Eugen von Böhm-Bawerk. *Capital and Interest: A Critical History of Economical Theory*. MacMillan and Co., London, 1890.
- [111] Eugen von Böhm-Bawerk. *The Positive Theory of Capital*. MacMillan and Co., London, 1891.
- [112] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 3rd edition, 1953.
- [113] Eric W. Weisstein. Normal difference distribution. From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/NormalDifferenceDistribution.html>.
- [114] John S. White. The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics*, 29(4):1188–1197, Dec 1958.
- [115] Ching-Chiang Yeh, Der-Jang Chi, and Ming-Fu Hsu. A hybrid approach of dea, rough set and support vector machines for business failure prediction. *Expert Systems with Applications*, 37(2):1535–1541, March 2010. PT: J; TC: 18; UT: WOS:000272432300073.

- [116] Büşra Zeynep Yılmaz. Completion, pricing and calibration in a levy market model. Master's thesis, The Institute of Applied Mathematics of Middle East Technical University, September 2010.
- [117] Stephen T. Ziliak and Deirdre N. McCloskey. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice and Lives*. University of Michigan Press, Ann Arbor, 2012.

## Vita

David Harris

1213 River Road  
Fairmont, WV 26554

### Education

- Ph.D. Economics, West Virginia University, 2014.
- M.B.A., West Virginia University, 2003.
- B.A. Psychology, West Virginia University, 1990.

### Experience

- Ball State University, Visiting Instructor, Spring 2012
- Fairmont State University/Pierpont Community & Technical College, Adjunct Professor, 2005-2011
- West Virginia University, School of Engineering, Instructor, Fall 2006
- Institute for Software Research, now West Virginia High Technology Consortium Foundation, Project Manager and Senior Research Analyst, 2000-2005

### Teaching Assistant

- Teaching Assistant, West Virginia University, Department of Economics, Professor: Ronald Balvers
- Teaching Assistant, West Virginia University, Department of Economics, Professor: Brian Cushing

### Fields of Research Interest

Financial economics, financial intermediation, econometrics

### Teaching Interests

Macroeconomics, microeconomics, decision making under risk, investments, financial institutions, and business statistics

## Research

### Articles

- Why Heavy Tails. Southwestern Economic Review. 2014. pp.127-151  
A mystery for the last fifty years, since Benoit Mandelbrot published his paper *The Variation of Certain Speculative Prices*, has been why the distribution of financial returns have heavy tails. A result of which is that many normative models generate empirical contradictions. Using both Bayesian and Frequentist methodologies, it is shown why heavy tails mathematically must be present in returns. As a consequence, it is shown that the models of mean-variance finance do not follow from first principles.

### Monograph

- Greater Than Unit Root Processes  
Co-Author: Timothy Glatzer  
Editor: Amy Blake  
Institute for New Economic Thinking  
New York, New York  
2012

### Working Papers

#### An Empirical Model Test

Harris, following Mandelbrot, asserts that mean-variance finance models are invalid as scientific models. Most prior empirical tests of the various models of mean-variance finance have been performed in the Frequentist methodology. This methodology tests the probability the data appears as it does, given that the underlying model in question is true. Bayesian statistics permit a direct one to one test of models through the use of Bayes factors. Bayes factors are an odds ratio that test the probability one model is true versus the probability another model is true. Results of this test exclude mean-variance finance in favor of heavy tailed distributions with a probability near one.

#### Estimating the Price for European Style Equity Option Contracts

Given that Itô calculus based methods are unavailable for the pricing of options methods are limited in calculating the equilibrium value of option prices. A simple Bayesian methodology is presented to price contracts, given equity and accounting data using parametric and semi-parametric methods. Extensions of this model to American and Asian style options are discussed.

#### Why Practitioners Should Use Bayesian Statistics (Temporarily)

An unexpected side effect of many recent changes in the understanding of

time series data, where boundary conditions exist such as bankruptcy, is that standard Frequentist tools for hypothesis testing do not yet exist. While lower powered tools, such as rank statistics, do exist for purposes of inference, Bayesian statistics offers both an unbiased estimator and a valid method of inference which should temporarily dominate long run frequency based tools.

**Bankruptcy Estimation** The author tests 78 potential bankruptcy models, based in part on the work of Sun and Shenoy. Rather than using Bayesian networks, the author proposes model averaging and contingent estimation. In addition, rather than predicting which firms will go bankrupt, the model provides a direct probability a firm will go bankrupt within one year.

## Early Stage Working Papers

### Generalizing Expectations: Introducing Three New Operators

The expectations operator is not a valid operator in many financial and macroeconomic modeling situations. This paper introduces three new operators denoted the ordinarily, the usually and the anticipation operator. Each of these operators are designed to find the basin of attraction when the expectation operator does not exist.

Author's notes:

- Each operator is induced from a loss function
- The *anticipation* operator is designed to overcome a perceived flaw in Bayesian decision theory, that is it has an *ad hoc* nature at times. The goal of the anticipation operator is to induce a tool resulting from loss minimization where multiple distributions are present with different conflicting properties under standard loss functions. It is believed that this set of rules will remove the perceived *ad hoc* nature of some elements of Bayesian decision theory.
- This paper will probably be broken into multiple papers as it includes such properties as tests for monotonicity, linearity and so forth.
- It is not clear that the paper will be capable of addressing both Frequentist and Bayesian concerns. The paper may have to be written twice, once for Frequentist methodologies and once for Bayesian methodologies
- This paper is also being planned for 2015
- This paper is also joint with Tim Glatzer, a mathematician

### Creating Formulas for Stable Distributions

Most stable distributions have no formula, although there are a variety of special cases that are analytic. The author proposes that this lack of analyticity comes from the role the cumulative normal distribution plays in

arriving in forming stable distributions. Once the cumulative normal portion is accounted for, the rest of the formulas should fall out from simple functions.

- For models with one source of error, it is argued that stable distributions can be modeled as projections from the complex plane.
- For models with two sources of error, such as from prices estimated with pricing errors due to models of accounting data and pricing errors due to models of liquidity, it is argued that stable distributions can be modeled as projections from a quaternion hyperspace onto the real numbers.
- A possible distribution of returns for mergers is arrived at. It is also the distribution for value investing. It is a Benjamin Graham/Warren Buffett distribution.
- A general rule for Bayesian analysis of any equity security, in or out of equilibrium is currently in production, for the special case of a portfolio with a fixed termination date and no possibility of reinvestment.

## Conference Presentations

- *Financial Instability*  
Institute for New Economic Thinking: Bridging Silos, Breaking Silences: New Responses to Instability and Inequality  
New York. 2011
- *A Necessary Limitation for Models of Capital and Capital Markets*  
Eastern Economic Association 37<sup>th</sup> Annual Conference  
New York. 2011
- *A Necessary Limitation for Models of Capital and Capital Markets*  
Western Economic Association 85<sup>th</sup> Annual Conference  
Portland. 2010
- *An Ordered Market Hypothesis: Evidence Against Accepted Theory*  
Academy of Accounting and Financial Studies  
Las Vegas. 2005

## Teaching

- Ball State University Spring 2012
  - Introduction to Microeconomics
  - Introduction to Statistics

- Fairmont State University/Pierpont Community & Technical College 2005-2011
  - Introduction to Macroeconomics
  - Financial Literacy
  - Real Estate Principles
  - Economics (non-major service course)
  - Special Topics in MIS (Joint with Defense Acquisition University and Defense Systems Management College)
- West Virginia University, Department of Economics Summer/Fall 2010
  - Macroeconomics(MBA)
  - Introduction to Statistics (undergraduate)
- West Virginia University, College of Engineering Fall 2006
  - Engineering Economy

## Field Research Experience

- Senior Research Analyst/Project Manager(2001-2006).  
Center for Technical Leadership Project. Institute for Software Research.  
Sponsored by Office of the Secretary of Defense, Defense Acquisition University, Defense Systems Management College

## Grants

- Develop Monograph to Cover Existing Literature on Greater than Unit Root Processes  
Institute for New Economic Thinking  
Principal Investigator  
Co-Principal Investigator: Timothy Glatzer  
Total Amount \$93,242.00  
Dates:2011-2012

## Professional Societies

- The American Economic Association 2011-
- Eastern Economic Association 2010-

- Royal Economic Society 2010-
- The Econometric Society 2010-2012
- Western Economic Association 2010-
- Association for Psychological Type 2000-2006
- Institute for Electrical and Electronic Engineers–Engineering Management Society 2004-2007
- National Association of Life Underwriters 1993-1996

## Computer Languages and Operating Systems

**Current** computer languages include (i) SAS (ii) MATLAB (iii) eViews (iv) Python , **previously** used languages include (v) C and C++ (vi) BASIC (vii) FORTRAN77 (viii) Stata (ix) S ,**familiar** operating systems include (xi) Unix (xii) Windows (xiii) Apple .

Last updated: December 12, 2014