Abstract

Estimating Marginal Survival in the Presence of Dependent and Independent Censoring: With Applications to Dividend Initiation Policy

by

Gretchen Abigail Fix

In many survival analysis settings, the assumption of non-informative (i.e. independent) censoring is not valid. Zheng and Klein (1995, 1996) develop a copula-based method for estimating the marginal survival functions of bivariate dependent competing risks data. We expand upon this earlier work and adapt their method to data in which there are three competing risks representing both dependent and independent censoring. Specifically, our extension allows for the estimation of the survival functions of dependent competing risks $X$ and $Y$ in the presence of a third independent competing risk $Z$. An application to dividend initiation data is presented.
Acknowledgments

Many people have helped me through the journey of graduate school. First, I would like to thank my committee members. All made unique contributions throughout the process. James Weston introduced me to the dividend initiation topic and gave great direction for the finance side of the dissertation. James Thompson, in his role as undergraduate advisor for statistics majors, recruited me for the graduate program. Xuelin Huang helped immensely with the development of the survival analysis component of the dissertation. In addition to his statistical expertise, he offered extraordinary accessibility and helpfulness. Finally, Kathy Ensor served as the glue that held everything together throughout my graduate school career. Kathy kept me on track with the big picture in sight when details were bogging me down.

I would also like to thank other members of the Rice statistics department. Diane Brown and Leticia Gonzales do a great job of keeping things running on a day-to-day basis. Scott Baggett has been a great resource for me with SAS help. Peter Olofsson deserves a measure of gratitude for getting me through measure theory. I will always remember to start with indicators.

I couldn’t have made it through without the help of many of my fellow graduate students. Blair Christian provided computing help countless times. Li Deng was a great roommate for two years, aside from the fishhead soup incident. Mike Lecocke and Rick Ott were here when it all started. Finally, I would like to extend special
thanks to Chad Bhatti. He has been an invaluable friend. Chad can always be counted on for a Diet Dr. Pepper and a dose of Midwestern sensibility.

Sean Bolks has filled many roles. He has been a boss, an academic mentor, a running partner, and a friend. His forthrightness, insight, and wisdom are always valued.

I would be remiss not to mention Howard and Lucinda LeDet. They have become a second set of parents to me. I couldn’t have survived in Houston without them. I owe their washing machine a debt of thanks.

Lastly, I owe the greatest amount of thanks to my parents, Robert and Susan. At an early age, they instilled in me a love of learning. As I grew up, they provided me with every opportunity to explore my interests and create my own path in life. A debt of such magnitude can, in truth, never be repaid.
# Contents

Abstract  
Acknowledgments  
List of Figures  
List of Tables  
1 Introduction  
2 Financial Background  
   2.1 Dividend policy hypotheses  
      2.1.1 Signaling hypothesis  
      2.1.2 Pecking order hypothesis  
      2.1.3 Lintner's early empirical work  
   2.2 Current research on dividends and related topics  
      2.2.1 Fama and French. “New Lists: Fundamentals and Survival Rates”  
      2.2.2 Fama and French. “Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?”  
      2.2.3 Grullon and Michaely. “Dividends, Share Repurchases, and the Substitution Hypothesis”
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.4 Deshmukh. “Dividend Initiations and Asymmetric Information: A Hazard Model”</td>
<td>16</td>
</tr>
<tr>
<td>2.2.5 DeAngelo, et al. “Are Dividends Disappearing? Dividend Concentration and the Consolidation of Earnings”</td>
<td>18</td>
</tr>
<tr>
<td>2.2.6 Remarks</td>
<td>19</td>
</tr>
<tr>
<td>2.3 Life cycle hypothesis of dividend initiation</td>
<td>20</td>
</tr>
<tr>
<td>3 Introduction to Survival Analysis and Copulas</td>
<td>23</td>
</tr>
<tr>
<td>3.1 Introduction to survival analysis concepts</td>
<td>23</td>
</tr>
<tr>
<td>3.1.1 Basics</td>
<td>24</td>
</tr>
<tr>
<td>3.1.2 The Kaplan-Meier estimator</td>
<td>26</td>
</tr>
<tr>
<td>3.1.3 The Cox Proportional Hazards model</td>
<td>27</td>
</tr>
<tr>
<td>3.1.4 Remarks</td>
<td>29</td>
</tr>
<tr>
<td>3.2 Introduction to copulas</td>
<td>30</td>
</tr>
<tr>
<td>3.2.1 Definition</td>
<td>30</td>
</tr>
<tr>
<td>3.2.2 Archimedean copulas</td>
<td>32</td>
</tr>
<tr>
<td>3.2.3 Remarks</td>
<td>36</td>
</tr>
<tr>
<td>4 Exploring Dividend Initiation Policy: Tracking the Disappearing Dividends</td>
<td>37</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>38</td>
</tr>
<tr>
<td>4.2 Proposed model and statement of hypotheses</td>
<td>42</td>
</tr>
<tr>
<td>4.2.1 Consideration of age in a model of dividend initiation</td>
<td>42</td>
</tr>
<tr>
<td>4.2.2 Stratification by exchange</td>
<td>44</td>
</tr>
</tbody>
</table>
4.3 Data ................................................................. 46
  4.3.1 Simple statistics of the CRSP sample ....................... 47
  4.3.2 Simple statistics of the covariates .......................... 49
4.4 Consideration of age in a model of dividend initiation ............ 51
  4.4.1 Survival curve estimation ................................... 51
  4.4.2 Hazard modeling .............................................. 53
4.5 Analysis of exchange-specific dividend initiation behavior .......... 61
  4.5.1 Exchange-specific survival curve estimation ............... 61
  4.5.2 Exchange-specific hazard modeling ......................... 64
4.6 Conclusions ....................................................... 68

5 Data Structure and Dependent Censoring .......................... 71
  5.1 Data structure and key result ................................... 73
  5.2 The self-consistent estimator ................................. 74
  5.3 The copula-graphic estimator ................................... 76
  5.4 Martingale extension of the copula-graphic estimator ......... 80
  5.5 Remarks on the copula approach ............................... 84

6 The Extended Copula-Graphic Estimator .......................... 86
  6.1 Trivariate partial competing risks data structure .............. 86
  6.2 Recovery of quantities from TPCR data ........................ 88
    6.2.1 Estimation of \( k(t) \) .................................. 88
    6.2.2 Estimation of \( p_1(t) \) .................................. 89
6.3 Implementation ......................................................... 91
  6.3.1 Function 1: \texttt{g.km.xy} ........................................ 94
  6.3.2 Function 2: \texttt{g.ipcw} ........................................ 96
  6.3.3 Function 3: \texttt{g.emp.prob-maker} ............................ 97
  6.3.4 Function 4: \texttt{g.cg.est} ....................................... 98

7 Simulation Study of the ECGE ........................................ 102
  7.1 Measures of association ............................................ 102
    7.1.1 Spearman's rho .................................................. 102
    7.1.2 Kendall's tau .................................................... 103
  7.2 Simulating copula data ............................................ 105
    7.2.1 Simulating from the Gumbel-Hougaard copula ............... 106
    7.2.2 Simulating from the Gamma Frailty copula .................. 109
    7.2.3 Simulating from the Frank copula ............................ 110
    7.2.4 Simulating from the independence copula ................. 111
  7.3 Simulations .......................................................... 111
    7.3.1 Simulation 1: Assessing performance and variance ......... 112
    7.3.2 Simulation 2: Investigating sensitivity to the choice of copula or implied strength of association .................... 116
    7.3.3 Simulation 3: Investigating the Frank copula and negative dependence ....................................................... 119

8 Application: Revisiting Dividend Initiation ....................... 123
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Re-estimating the survival curves</td>
<td>125</td>
</tr>
<tr>
<td>8.2</td>
<td>Sensitivity to tie breaking</td>
<td>133</td>
</tr>
<tr>
<td>9</td>
<td>Conclusion</td>
<td>137</td>
</tr>
<tr>
<td>A</td>
<td>Filtering and Formatting the Data</td>
<td>141</td>
</tr>
<tr>
<td>A.1</td>
<td>Creating the CRSP sample</td>
<td>141</td>
</tr>
<tr>
<td>A.2</td>
<td>Creating the Compustat sample</td>
<td>142</td>
</tr>
<tr>
<td>A.3</td>
<td>Formatting the data for survival analysis</td>
<td>144</td>
</tr>
<tr>
<td>B</td>
<td>S-Plus Code for the ECGE</td>
<td>146</td>
</tr>
<tr>
<td>B.1</td>
<td>g.km.xy</td>
<td>146</td>
</tr>
<tr>
<td>B.2</td>
<td>g.ipcw</td>
<td>148</td>
</tr>
<tr>
<td>B.3</td>
<td>g.emp.prob.maker</td>
<td>152</td>
</tr>
<tr>
<td>B.4</td>
<td>g.cg.est</td>
<td>154</td>
</tr>
<tr>
<td>B.5</td>
<td>copula.eqns.for.x</td>
<td>156</td>
</tr>
<tr>
<td>B.6</td>
<td>nm.copula.for.x</td>
<td>157</td>
</tr>
<tr>
<td>C</td>
<td>S-Plus Code for Copula Simulation</td>
<td>159</td>
</tr>
<tr>
<td>D</td>
<td>Deriving the Gamma Frailty Copula</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>164</td>
</tr>
</tbody>
</table>
List of Figures

4.1 All Firms: Kaplan-Meier Survival Curves for Dividend Initiation. . . 52
4.2 All Firms: Graphical View of Coefficients of the Cox PH Model. . . 57
4.3 Projected Survival Curves Obtained from the Cox PH Model. . . . 60
4.4 NYSE Firms: Kaplan-Meier Survival Curves for Dividend Initiation. . 62
4.5 Nasdaq Firms: Kaplan-Meier Survival Curves for Dividend Initiation. 63

6.1 Sample Output from g.km.xy. ................................. 95
6.2 Sample Output from g.ipcw. ................................. 97
6.3 Sample Output from g.cg.est. ................................. 101

7.1 Simulation 1: Performance of the ECGE under Assumption of the
Correct Dependence Structure. ................................. 114
7.2 Simulation 1: Variance of the ECGE. ................................. 115
7.3 Simulation 2: Performance of the ECGE under Assumption of the
Correct Dependence Structure. ................................. 117
7.4 Simulation 2: Performance of the ECGE under Assumption of the
Incorrect Copula and Varying Strengths of Dependence. .............. 118
7.5 Simulation 3: Performance of the ECGE under Assumption of the Correct Dependence Structure. .......................... 120

7.6 Simulation 3: Performance of the ECGE under Assumption of the Correct Copula and Varying Strengths of Dependence. ............. 121

8.1 All Firms: ECGE Survival Curves for Dividend Initiation .................. 126

8.2 All Firms: Confidence Limits for ECGE Survival Curves for Dividend Initiation ........................................ 127

8.3 NYSE Firms: ECGE Survival Curves for Dividend Initiation .......... 128

8.4 NYSE Firms: Confidence Limits for ECGE Survival Curves for Dividend Initiation ........................................ 130

8.5 Nasdaq Firms: ECGE Survival Curves for Dividend Initiation ....... 131

8.6 Nasdaq Firms: Confidence Limits for ECGE Survival Curves for Dividend Initiation ........................................ 132

8.7 Differences in Survival Estimates from Two Tie Breaking Approaches 135
List of Tables

4.1 Summary Statistics for Age at Listing .......................... 47
4.2 Composition by Exchange ........................................... 48
4.3 Percentage of Payers vs. Non-Payers ............................. 48
4.4 Summary Statistics for Profitability, Investment Opportunity, and Size 50
4.5 Tests of Equality of Survival Curves; All Firms ................. 52
4.6 Cox PH Model for Dividend Initiation; All Firms ............... 56
4.7 Tests of Equality of Survival Curves; NYSE Firms ............... 62
4.8 Tests of Equality of Survival Curves; Nasdaq Firms ............. 63
4.9 Cox PH Model for Dividend Initiation; NYSE Firms ............. 65
4.10 Cox PH Model for Dividend Initiation; Nasdaq Firms .......... 66

7.1 Spearman's Rho and Kendall's Tau for Implemented Copulas .... 105

8.1 Distribution of Event Types ......................................... 124
8.2 Maximum Differences in Survival Estimates ....................... 134
Chapter 1

Introduction

The statistical research presented in this dissertation was motivated by a topic raised in the finance literature. Some time ago, at the suggestion of Dr. James Weston, I read an article by Fama and French (2001) entitled "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?" This article captured my interest on two levels. First, due to recent changes in tax legislation that reduced the double taxation burden on dividends, the study of dividend prevalence is extremely timely and relevant.

Secondly, I was interested in the methodology being used to explore the topic. Coming from a statistical background, it seemed to me that the dividend data structure lent itself perfectly to the survival analysis framework. The finance practitioners were relying primarily on traditional regression techniques (such as logistic regression) to analyze the data.

Therefore, I set out to conduct my own analysis of dividend initiation using survival analysis tools. While I was confident that the topic could effectively be analyzed
using the survival analysis framework, I quickly realized that the data did not satis-
ify one of the usual assumptions of survival analysis—that of independent censoring.
The most prevalent source of (non-administrative) censoring in the data, bankruptcy,
should be considered informative for dividend initiation.

To further complicate the issue, there remained sources of censoring in the data
that were best considered to be non-informative for dividend initiation. Although
there have been several techniques developed for estimating marginal survival in the
presence of dependent censoring, the co-presence of independent censoring deemed
the data incompatible with those methods.

The challenge of dealing with the dependent and independent censoring in the
data gave me the opportunity to engage in statistical research with an immediate
application for my efforts. Specifically, I set about to expand an existing survival
estimator (which used a copula-based method to address dependent censoring) to
incorporate the effects of independent censoring as well.

This dissertation is organized as follows. In Chapter 2, we provide the statistically-
oriented reader with historical background on dividend theory and an introduction to
current research on dividend initiation. Chapter 3 gives an introduction to survival
analysis and copulas. Chapter 4 contains our first efforts at addressing the “disap-
ppearing dividends” phenomenon for a finance audience. The presence of dependent
censoring is not rigorously addressed. We discuss our data structure and existing
copula-based methods for addressing dependent censoring in Chapter 5. Chapter 6
details the development of the Extended Copula-Graphic Estimator (ECGE). The ECGE is our copula-based estimator for recovering marginal survival functions when both dependent and independent censoring are present. The results of simulation studies evaluating the performance of the ECGE are shown in Chapter 7. In Chapter 8, we revisit the dividend initiation data and analyze it using the ECGE. Chapter 9 concludes.
Chapter 2

Financial Background

Before embarking upon a detailed exploration of dividend behavior, we shall first give a formal definition of dividends. Frankfurter and Wood (2003, p. 6) state that “[d]ividends are commonly defined as the distribution of earnings (past or present) in real assets among the shareholders of the firm in proportion to their ownership.”

One of the most basic questions surrounding dividends is why they even exist. Currently in the United States, dividends have the dubious distinction of being the only source of income that is subject to double taxation. Dividends are paid from a firm’s after-tax income and are also taxed as personal income for the recipient. (Although recently enacted tax legislation has reduced the personal tax burden on dividends, this legislation rests on politically shaky ground and could be reversed.) Therefore, dividends can be viewed as “an involuntary tax liability to the owners of a firm imposed on a marginal liquidation of their ownership” (Frankfurter and Wood, 2003, p. 6).

It would seem, then, that shareholders would prefer that firms put earnings toward
something other than dividend payouts, such as reinvestment in the firm or share repurchases. However, empirical studies have found that stock prices tend to increase upon the announcement of an (unexpected) dividend increase and tend to decrease upon the announcement of an (unexpected) dividend decrease (Asquith and Mullins, 1983; Grullon et al., 2002). Therefore, the empirical evidence suggests that:

- dividend announcements bring information to the market.
- dividends are viewed favorably by investors.

### 2.1 Dividend policy hypotheses

Every quarter (or year) a firm must make a decision regarding its dividend policy. Depending on its present dividend-paying status, it must decide whether to initiate dividends, change the level of current dividends, cease dividend payment, or maintain the status quo. Since a firm operates in a state of constantly changing information, opportunity, and outlook, dividend policy is dynamic in nature.

Several theories have been formulated to explain what factors drive dividend policy. A few of these theories are introduced below. The purpose of this section is to give the reader a flavor for academic thought on dividend policy. It is not meant to be a thorough or complete overview of dividend policy research.

#### 2.1.1 Signaling hypothesis

Signaling theory for dividends was formally introduced by Bhattacharya (1979). The crux of this hypothesis is that potential investors possess incomplete information
about a firm’s profitability. Signaling theory assumes that the profit making machinery of the firm will stay in place longer than the current investors hold their stakes in the firm. These current investors will eventually desire to transfer their ownership to new investors. However, as stated above, potential investors possess incomplete information about the firm’s profitability and prospects. Therefore, firms use dividends as a signal of future expected cash flows, with high dividends signaling strong future profitability. Current shareholders are willing to accept the cost of receiving dividends since a strong dividend stream makes their investment appear more valuable to potential shareholders.

2.1.2 Pecking order hypothesis

Pecking order is another theory based on the concept of asymmetric information. The theory was named and presented by Myers (1984), although he notes that its concepts have been present in the finance literature since (at least) the early 1960s. The name “pecking order” was used because the theory rests upon the assumption that firms have a strict hierarchy of financing sources for funding investment projects. For example, the first (most preferred) source is a firm’s internally generated cash flow. Last on the list is external financing, such as debt, hybrid securities (e.g. convertible bonds) or equity.

Myers (1984, p. 582) explains pecking order theory from the viewpoint of a firm that must raise money to finance a positive net present value (NPV) project. In this case, the firm is unable to internally finance the project and must resort to external
financing. The firm's management knows both the NPV of the investment, $y$, and the value of the firm if the project is foregone. Investors outside the firm do not know these values. If the firm undertakes the investment, its value will increase by $y$. However, since outside investors have imperfect information, the firm may have to finance the project by selling the securities (debt or equity) for less than they are actually worth. Suppose that the firm is able to sell the securities necessary to finance the project at a market value of $N$, but that the firm's management knows that the true value of the securities is $N_1$. Let $\Delta N = N_1 - N$ represent the amount by which the securities are undervalued. If the amount by which the securities are undervalued exceeds the NPV of the project (i.e. if $\Delta N > y$), the firm will not pursue the project.

If the firm could have retained enough of its internally generated cash flow, then it would not have been put in the position of asymmetric information possibly forcing it not to pursue a positive NPV project. Since dividends are an outlet for a firm's internally generated cash flow, the pecking order theory suggests that the higher the level of asymmetric information, the lower the level of dividends. This is directly opposite the effect suggested by signaling theory.

2.1.3 Lintner's early empirical work

Lintner (1956) conducted ground breaking empirical work in the field of dividend policy. His work was based on an extensive study of 28 firms over the post-war years of 1947 to 1953. For each firm, Lintner performed a financial analysis and conducted interviews with multiple members of management (presidents, vice presidents
of finance, treasurers, controllers, or directors). A summary of his findings is given below.

Based on his observations of the managerial decision making process, Lintner believes that dividend policy is formulated using the current level of dividends as a starting point. The relevant question to firm management is not "What should the newly established level of dividends be?" Instead, it is "What is the appropriate change in the level of current dividends?"

Lintner (1956, p. 102) finds that "the relationship between current earnings and the existing dividend rate was very generally much the most important single factor determining the amount of any change in dividends decided upon." Lintner also identifies a predominant strategy practiced by two-thirds of the companies in his sample. These firms have an established target payout ratio and an established adjustment rate which reflects the speed at which the firm adjusts dividend rates to reach the target payout ratio. For example, if earnings in the current period are double that of the previous period, a firm would be unlikely to suddenly double the size of its dividend. Instead, it would be more likely to gradually increase the dividend over the coming periods in accordance with an established adjustment rate. This method of "dividend smoothing" is supported by the common beliefs of management that

- investors and markets prefer relatively stable or gradually increasing dividend payout rates.

- dividend cuts are looked upon unfavorably by investors; it is unwise to make
drastic increases in dividends to levels which may not be sustainable.

Finally, Lintner offers a model for dividend policy based on the smoothing behavior observed in his sample. For firm $i$, the dividend payout at time $t$, $D_{it}$, is determined in terms of the change in dividend payment, $\Delta D_{it}$, measured from the previous period to the current period, as follows:

$$\Delta D_{it} = a_i + c_i \left( D^*_t - D_{it-1} \right) + u_{it}$$  \hspace{1cm} (2.1)

where $D^*_t = r_i P_t$ (where $r_i$ is the target payout ratio and $P_t$ is the current year’s post-tax profits) and $u_{it}$ is the error term. According to Lintner (1956, p. 107), the parameter $c_i$ represents “the fraction of the difference between this “target” dividend $D^*_t$ and the actual payment made in the preceding year $D_{it-1}$, which the company will intend on the average to reflect in its current year’s dividend as an increase (or decrease) from the previous year’s payment.” The value of $a_i$ is usually observed to be positive, as it reflects the preference of firms to raise (rather than reduce) dividends.

Lintner asserts that the model is reasonably fit to his sample data. Twenty-six (of 28) companies had well defined values of $r_i$ (“established as a matter of policy” (Lintner, 1956, p. 108)), and twenty companies had well defined values of $c_i$. He states that “about 85 per cent of the company-years of dividend action studied in this group of twenty-eight companies can be explained in terms of the model with only moderate discrepancies” (Lintner, 1956, p. 108).

Lintner does not directly discuss why firms elect to pay dividends in the first place. The closest he comes to addressing this issue is to state that the general view
of management is a responsibility to disperse a fair share of earnings to stockholders, especially when there is a substantial increase in earnings. He does not address the taxation disadvantage of dividends.

2.2 Current research on dividends and related topics

The purpose of the next section is to give the reader an introduction to recent empirical work on dividend policy. Five papers are presented. The first paper, “New Lists: Fundamentals and Survival Rates,” by Fama and French (2003) provides background information on changing firm characteristics in the US market, but does not directly apply to the dividend topic. The second paper, “Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?” also by Fama and French (2001), documents a sharp decline in the proportion of dividend-paying firms over the time period 1978 to 1998. They hypothesize that this decline was caused by changing firm characteristics and by a reduction in the “propensity to pay” dividends. Grullon and Michaely (2002), in “Dividends, Share Repurchases, and the Substitution Hypothesis,” argue that firms have begun replacing dividend payouts with share repurchases. In “Dividend Initiations and Asymmetric Information: A Hazard Model,” Deshmukh (2003) provides a recent contribution to the literature which is relevant for both its content and methodology. Deshmukh uses survival analysis techniques to test the signaling and pecking-order hypotheses. In the final paper, “Are Dividends Disappearing? Dividend Concentration and the Consolidation of Earnings,” DeAn-
gelo et al. (2003) present evidence that although the proportion of dividend-paying firms has declined, real and nominal dividends increased over the period 1978 to 2000. They attribute the simultaneous decline in the number of payers and increase in the total dividend payout to changes in the distribution of earnings.


The main topic of this Fama and French paper is not dividend policy. However, the focus of this paper, changes in the long-term viability of newly listed firms, will prove relevant to our discussion.

Fama and French report that 1979 marked a jump in the rate at which new firms list on US exchanges. In the years before 1979, new lists averaged 160 per year. In the years following 1979, new lists averaged 550 per year.

Furthermore, after 1979, the characteristics of new lists changed. The distribution of profitability became more left-skewed (toward lower profitability), and the distribution of growth became more right-skewed (toward higher growth). Fama and French measure profitability as the ratio of earnings before interest to assets, $E_t/A_t$, and growth as the growth rate of assets, $(A_t - A_{t-1})/A_t$. The authors assert that these changing firm characteristics negatively impacted new list survival. In 1973, the probability that a new list survived at least ten years was 61%; by 1991, this probability fell to 37% (Fama and French, 2003, p. 3). Further analysis attributes this decline in the survival rate not to mergers, but to delistings caused by poor
performance.

2.2.2 Fama and French. (2001). "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?"

In 1973, 52.8% of publicly traded, non-financial, non-utility (hereafter "industrial") firms paid dividends. In 1978, this figure reached its peak at 66.5%. By 1999, the proportion of dividend-paying industrial firms declined to just 20.8% (Fama and French, 2001, p. 4). In this paper, Fama and French pose three main questions of interest:

1. What are the characteristics of firms that choose to pay dividends?

2. Is the decline in the number and percentage of payers caused by a decline in the prevalence of the characteristics identified above?

3. Have firms possessing the characteristics historically typical of dividend-payers become less likely to pay dividends?

Fama and French identify three characteristics of firms that are relevant to their decision to pay dividends: profitability, investment (growth) opportunity, and size. Former payers (firms who once paid dividends but have since ceased) are best characterized as financially distressed. Never payers tend to be more profitable than former payers and tend to have abundant investment opportunities. Dividend-paying firms are generally more profitable than never payers, but possess fewer investment opportunities. Also, dividend-paying firms tend to be larger than non-paying firms.
With regard to the second question, Fama and French mention the surge of new lists that began flooding the market in 1979. As discussed previously, these new lists tended to possess low profitability, strong growth prospects, and small size. Therefore, these new lists helped shift the population of firms toward characteristics that reduce the likelihood of paying dividends.

Fama and French assert that regardless of their characteristics, firms have become less likely to pay dividends. In other words, firms have developed a lower propensity to pay dividends. To establish this lower propensity to pay effect, Fama and French use a logistic regression approach.

For their analysis, the authors establish 1963–1977 as the base period and 1978–1998 as the forecast period. For each firm-year observation in the base period, the response variable, $y_t$, is set to 1 if a firm pays a dividend in year $t$ and 0 otherwise. They first fit the logistic regression model to the base period data using size, profitability, and investment opportunity as covariates (independent variables). They then use the coefficient vector obtained from the base period fit and the covariate data from the forecast period to estimate the expected percentage of dividend-payers for each year from 1978 to 1998.

Variation in the annual expected percentage of dividend-payers reflects the changing firm characteristics of the post 1977 era. Fama and French find a definite downward trend in the annual expected percentage of dividend-payers, implying that firms are increasingly tending toward the characteristics of non-dividend-payers.
The difference between the expected percentage of payers and the actual percentage of payers reflects changes in the propensity to pay. For every year after 1977, the actual percentage of payers falls below the expected percentage of payers. This establishes the decreased propensity to pay effect.

Finally, the authors discuss the role of share repurchase activity. They acknowledge that some researchers believe that firms are substituting share repurchases for dividends and that increased share repurchase activity is the reason that fewer firms are paying dividends. Fama and French dismiss this argument, maintaining that share repurchases are largely the domain of dividend-paying firms.

2.2.3 Grullon and Michaely. (2002). "Dividends, Share Repurchases, and the Substitution Hypothesis"

Despite the relative tax advantage of capital gains over ordinary (i.e. dividend) income, US firms have seemed to favor dividends over share repurchases when undertaking cash payouts. However, according to Grullon and Michaely, share repurchase activity has been on the rise over the past twenty years. Over the period 1980–2000, share repurchase expenditures grew at an average annual rate of 26.1%, while dividends grew at an average annual rate of 6.8% (Grullon and Michaely, 2002, p. 1649).

In this paper, Grullon and Michaely identify three questions of interest:

1. Has there been a change in payout policy? Specifically, have firms become more likely to engage in share repurchase activity?

2. Are firms funding share repurchases with money they otherwise would have
used for dividends?

3. If firms are substituting share repurchases for dividends, why did they not start earlier?

With regard to a change in payout policy, Grullon and Michaely assert that since the mid 1980s, more firms have opted to initiate share repurchases than dividends. Also, the percentage of firms that pay only dividends (relative to the total number of firms distributing cash to equity holders) fell from 69% in 1972 to 20% in 2000 (Grullon and Michaely, 2002, p. 1660).

Grullon and Michaely believe that share repurchases are being funded by dollars that otherwise would have gone toward increasing or initiating dividends. They believe that this is especially true for large, established firms. According to their analysis, the market reaction to the announcement of a dividend reduction has been significantly less negative for firms that engage in share repurchases. This supplies evidence for the substitution effect.

Grullon and Michaely’s findings are in contrast to those presented by Fama and French (2001). However, Grullon and Michaely believe that Fama and French’s analysis is tainted by the use of an incorrect accounting measure of share repurchase activity.

The authors note a marked increase in share repurchases after the passage of Rule 10b-18 in 1982. Prior to this, there were no definite guidelines set by the Securities and Exchange Commission (SEC) to regulate share repurchase activity. In such an
environment, repurchasing firms faced the risk of triggering an SEC investigation into market manipulation practices. Grullon and Michaely hypothesize that this lack of regulation was a deterrent to share repurchase activity.


In this paper, Deshmukh models the decision of firms to initiate dividends. His sample consists of firms that went public between 1990 and 1997 and contains 1,371 firms and 6,171 firm-years of observations. A firm is tracked from its initial public offering (IPO) until dividend initiation or the end of the study (December 31, 2000), whichever comes first. Firms are considered to be censored if they have not initiated a dividend by the end of the study. Deshmukh makes no mention of other potential sources of censoring in his data, such as bankruptcies and mergers. Given the time period and population under consideration (i.e. new firms of the 1990s), it seems likely that there are many bankruptcy censorings in his sample. Deshmukh’s analysis is conducted under the assumption of independent censoring.

Deshmukh states that his paper makes contributions to the literature on three fronts. First, it is the first attempt at modeling the transition of a firm from non-dividend-payer to dividend-payer. Furthermore, Deshmukh constructs his sample using newly listed firms. As such, these firms are likely to be in a state of high growth and low cashflow and in need of external funding. Also, as young firms, they are likely to be facing high levels of asymmetric information. This sample provides an
opportunity to test the signaling and pecking-order hypotheses described in Section 2.1.

Secondly, the author notes that his methodology allows him to include non-dividend-paying firms in the analysis. Previous empirical work has ignored them, possibly leading to selection bias in the sample and yielding results that should not be generalized.

Finally, Deshmukh remarks that the use of the survival analysis framework, namely hazard modeling, allows him to study dividend policy from a dynamic point of view.

Deshmukh identifies four factors that are relevant to the dividend initiation decision and included as covariates in his models. They are

1. agency costs of external equity, represented by the log of the number of common shareholders of the firm.

2. cash flow, calculated as the ratio of earnings before interest, depreciation, and taxes (EBIDT) to the book value of assets.

3. growth opportunity, represented by the ratio of the market value of assets to the book value of assets.

4. asymmetric information, measured by firm size.

Deshmukh fits three different models to his data. He first takes a discrete-time approach and uses a logistic regression model. Next, he switches to continuous time and an exponential regression model. Finally, he uses a Cox Proportional Hazards
The Cox PH model allows the hazard of dividend initiation to be a function of time. This will be a much emphasized point when we develop our own hypothesis and model of dividend initiation.

All three models yield consistent and robust results. The hazard rate is positively related to firm size and cash flow and negatively related to growth opportunity. The positive relationship between firm size (the proxy for asymmetric information) and the hazard of dividend initiation provides support for the pecking-order hypothesis and against the signaling hypothesis.

2.2.5 DeAngelo, et al. (2003). "Are Dividends Disappearing? Dividend Concentration and the Consolidation of Earnings"

The work of Fama and French (2001) shows a decrease in the number and percentage of dividend-paying industrial firms over the period 1978–1998. However, DeAngelo et al. (2003) report that the levels of real and nominal dividends paid by industrial firms increased over this period. Furthermore, they find that the reduction in the number and percentage of payers came mainly from the loss of firms that pay a small dividend. In fact, the largest payers significantly increased dividends. The authors state that the "increase in real dividends paid by firms at the top of the dividend distribution swamps the dividend reduction associated with the loss of many small payers at the bottom" (DeAngelo et al., 2003, p. 2).

In addition, DeAngelo et al. study the high and increasing concentrations of both the dividend supply and earnings and examine the linkages between these two factors.
For example, in 2000, just 75 firms paid 75% of aggregate industrial dividends. In the same year, 56 firms with over $500 million in earnings were responsible for 86.2% of aggregate industrial earnings and 61.4% of aggregate industrial dividends (DeAngelo et al., 2003, p. 2).

The authors provide dollar amounts that help put the scale and growth of dividend payouts in perspective. In 1978, aggregate industrial nominal dividends were $31.3 billion. By 2000, this number increased 207.3% to $96.2 billion. Aggregate industrial real dividends measured $36.4 billion in 2000, an increase of 16.3% over 1978 levels (DeAngelo et al., 2003, p. 5). In 1978, the (firm-level) mean and median dividend payout (in real terms) were $14.4 million and $1.4 million, respectively. By 2000, these figures were $39.2 million and $3.6 million (DeAngelo et al., 2003, p. 6). The increasing gap between mean and median is evidence of increased dividend concentration.

2.2.6 Remarks

The five papers presented in this overview contrast and complement one another. Much of the evidence for Fama and French's 2001 argument that firm characteristics changed during the 1980s and 1990s is drawn upon again in the context of new firm survivorship in their 2003 paper. Grullon and Michaely directly refute the claim of Fama and French that share repurchase activity has not played a role in the decreasing proportion of dividend-paying firms. While the number and percentage of dividend-paying firms has dropped and amid the claim that dividends are "disappearing,"
DeAngelo et al. remind us that in both nominal and real terms, dividends are on the rise.

Deshmukh's paper is an important contribution to the literature. However, his discussion of the statistics underlying his models is very limited, and he fails to address the possible existence of dependent censoring. Based on the findings of Fama and French, dependent censoring (as brought about by bankruptcy) is likely to be an important component of an analysis containing newly listed firms of the 1990s. While Deshmukh correctly identifies survival analysis as a suitable framework for the analysis, we believe the application can be strengthened.

2.3 Life cycle hypothesis of dividend initiation

Recent empirical research has identified a common set of covariates used to model dividend policy decisions. Fama and French (2001) use profitability, investment opportunity, and size. Grullon and Michaely (2002) base their model on that of Lintner (1956) and use earnings and the market value of equity in their study of the substitution effect. (Their findings also suggest that repurchase activity should be considered when modeling dividend policy.) Deshmukh (2003) finds that cash flow, investment opportunity, and size contribute significantly to his models of dividend initiation.

Like Deshmukh, we restrict our analysis to the decision of firms to initiate dividends. Based on the literature, a model of dividend initiation should account for profitability (or earnings or cash flow), investment opportunity, size, and perhaps, repurchase activity. However, we believe that another important input to dividend
policy decisions, thus far ignored in the literature, is the age of the firm.

As mentioned previously, current tax laws dictate that it is theoretically more advantageous for firms to reinvest earnings in positive NPV projects than to use earnings to pay dividends. However, as firms and their industries mature, they naturally begin to run out of positive NPV projects in which to invest. Therefore, as firms age, they naturally develop a higher tendency to turn to dividends as a method for disbursing earnings.

We refer to this as the “life cycle hypothesis” for dividend policy. We explain our hypothesis in more detail in Chapter 4, Section 2. Young firms with abundant growth opportunities will refrain from paying dividends. Mature firms will structure their payout policy around the fact that growth opportunities are dwindling.

By setting up a model of dividend initiation using the Cox PH framework, Deshmukh (2003) allows the hazard of dividend initiation to depend upon time. In his study, time is measured relative to the listing year of a firm. Therefore, Deshmukh’s model allows the hazard of dividend initiation to be a function of a firm’s age measured from listing. However, given the life cycle hypothesis outlined above, we believe that incorporation-age, rather than listing-age, is a more relevant input to dividend policy.

We note that the market conditions of the 1980s and 1990s encouraged firms to go public relatively early in their life cycles. Furthermore, an increase in the annual number of new lists began in 1979. Fama and French (2003) report that prior to
1979, approximately 160 new firms listed on US exchanges each year. After 1979, this figure jumped to approximately 550 per year. We wish to ascertain whether the decrease in the proportion of dividend-paying firms over the period 1978 to 1998 was partially due to the fact that the market was flooded each year with record numbers of new lists. Furthermore, these new lists were significantly less mature, by incorporation-age standards, than the new lists of previous decades. In this manner, part of the reduction that Fama and French attribute to the decreased propensity to pay effect may be another example of shifting firm characteristics. In this case, the shift was toward younger, more immature firms which had not yet begun to exhaust their growth opportunities and develop a higher tendency to initiate dividends.
Chapter 3

Introduction to Survival Analysis and Copulas

3.1 Introduction to survival analysis concepts

We have chosen to analyze the topic of dividend initiation using a statistical methodology known as survival analysis. Survival analysis goes by many other names, often depending upon the discipline or context in which it is being applied. Social scientists refer to it as event history analysis. Engineers refer to it as reliability or failure time analysis.

Survival analysis is a collection of statistical methods for data analysis when the response of interest is time until an event occurs. As Kalbfleisch and Prentice (2002, p. 1) note "[s]uch events are generically refered to as failures, although the event may, for instance, be the performance of a certain task in a learning experiment in psychology or a change of residence in a demographic study." In our analysis, the event of interest is dividend initiation. Firms are considered to fail when they initiate dividends. In other words, we are analyzing how long firms survive as non-dividend-
Survival analysis is often used in biostatistical or industrial settings. In biostatistics, it is frequently used to study treatment effectiveness. In this case, the outcome variable of interest may be the survival or remission times of patients undergoing various treatment regimens. In an industrial setting, one may be interested in studying the reliability of a system. In this case, the outcome variable of interest may be the time to failure of a machine.

3.1.1 Basics

There are two main functions of interest in survival analysis, the survival function and the hazard function. The survival function, $S(t)$, gives the probability that a subject survives longer than some specified time $t$. There is a simple relation between the survival function and the commonly used cumulative distribution function, $F(t)$.

$$S(t) = \Pr(T > t) = 1 - \Pr(T \leq t) = 1 - F(t).$$

A few properties of the survival function follow:

- $S(t)$ is non-increasing. (This follows directly from the fact that $F(t)$ is non-decreasing.)

- $S(0) = 1$. At the start of a study, all subjects are alive.

- $S(\infty) = 0$. If a study were allowed to continue indefinitely, eventually no subjects would be left alive.
The hazard function, $\lambda(t)$, is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}. \quad (3.2)$$

The hazard function can be interpreted as the instantaneous potential per unit time for failure given that an observation has survived up to time $t$. It should be emphasized that the hazard function is a conditional failure rate, not a conditional failure probability. The range of the hazard function is not restricted to be between 0 and 1.

It can be shown that there is a one-to-one correspondence between the survival function and the hazard function. If one is known, the other can be derived. Let $f(t)$ be the probability density function (pdf) of the continuous random variable $T$. From Eqn. (3.1), we have

$$S(t) = \int_t^\infty f(x)dx. \quad (3.3)$$

From Eqn. (3.2) and the following definition of the pdf

$$f(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t} \quad (3.4)$$

it follows that

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} \quad (3.5)$$

Integrating with respect to $t$ and using the fact that $S(0) = 1$, we obtain

$$-\int_0^t \lambda(s)ds = \log S(t). \quad (3.6)$$
Equivalently,

$$\exp\{-\int_0^t \lambda(s)ds\} = S(t).$$  \hspace{1cm} (3.7)

Most datasets used in survival analysis contain censored data. A subject is considered to be censored if its exact survival time is unknown. Censoring usually occurs at the right side of a lifetime interval, but we can also have interval or left censoring.

Typically, we see right censoring for one of three reasons:

1. The subject does not experience the event of interest before the study ends. This is referred to as administrative censoring. For example, in a five year test of reliability, a machine is yet to fail by the end of the five year period.

2. The subject is lost to follow up during the study.

3. The subject withdraws from the study.

A standard assumption used in survival analysis is that of independent censoring. This means that the censoring mechanism is independent of (or non-informative for) failure. Intuitively, this assumption implies that upon observing a censoring at time $t_c$, we can infer no more about the true lifetime, $X$, of the subject beyond the obvious fact that $X > t_c$.

### 3.1.2 The Kaplan-Meier estimator

A standard non-parametric estimator of the survival function (or curve) was developed by Kaplan and Meier (1958). To calculate the Kaplan-Meier estimator, let
\( t_1, t_2, \ldots, t_n \) be the ordered failure times of the sample. Set \( t_0 = 0 \). Let \( D_i \) be the number of observations that fail at time \( t_i \), and let \( N_i \) be the number of observations in the risk set at time \( t_i \). The risk set contains the observations that are alive and under observation just prior to time \( t_i \). Then,

\[
\hat{S}_{KM}(t) = \prod_{i: \, t_i \leq t} \frac{N_i - D_i}{N_i}.
\]

(3.8)

The Kaplan-Meier estimator is a right-continuous step function with jumps at the failure times of the sample. In the absence of censoring, the estimator reduces to the empirical survival function. The Kaplan-Meier estimator is the non-parametric maximum likelihood estimator of the survival function (Kaplan and Meier, 1958, p. 475). Using Greenwood's formula, the variance of the estimator can be approximated:

\[
Var\left(\hat{S}_{KM}(t)\right) \approx (\hat{S}_{KM}(t))^2 \sum_{i: \, t_i \leq t} \frac{D_i}{N_i(N_i - D_i)}.
\]

(3.9)

3.1.3 The Cox Proportional Hazards model

The Cox Proportional Hazards (PH) model, introduced by Cox (1972), is a semi-parametric method for modeling the hazard function. Let \( x \) be a \( k \times 1 \) vector of time-fixed covariates. The Cox PH model represents the hazard of an event at time \( t \) as the product of two factors

\[
\lambda(t) = \lambda_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k\}.
\]

(3.10)

The first factor, \( \lambda_0(t) \), is referred to as the baseline hazard and is a function of time, but not of the covariates. The exact functional form of this term is left unspecified
in the model. For this reason the Cox PH model is characterized as semi-parametric.

The second factor, \( \exp\{\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k\} \), is a function of the covariates, but not of time.

For subject \( m \) with covariate vector \( x_m \) and subject \( n \) with covariate vector \( x_n \), the ratio of the hazards is

\[
\frac{\lambda_m(t)}{\lambda_n(t)} = \frac{\lambda_0(t) \exp\{\beta' x_m\}}{\lambda_0(t) \exp\{\beta' x_n\}} = \frac{\exp\{\beta' x_m\}}{\exp\{\beta' x_n\}}
\]

which is independent of time. Since the ratio of the hazards is constant in time, the hazards are proportional. It is from this property that the model derives its name.

The form of the Cox PH model allows not only extreme flexibility, but also easy interpretation of the covariate effects. The term \( \exp\{\beta_i\} \) gives the marginal multiplicative effect on the hazard resulting from a one unit change in covariate \( x_i \).

The model can be extended to allow for time-varying covariates. Let \( X \) denote a matrix of covariates indexed by time. Then Eqn. (3.10) can be rewritten as

\[
\lambda(t) = \lambda_0(t) \exp\{\beta_1 x_1(t) + \beta_2 x_2(t) + \ldots + \beta_k x_k(t)\}. \tag{3.12}
\]

Note that the incorporation of time-dependent covariates yields a ratio of hazards that is dependent on time. Some researchers, such as Kalbfleisch and Prentice (2002, p. 43) prefer to call the model with time-dependent covariates the relative risk model or simply the Cox model. We will not make such a distinction in our writing.
3.1.4 Remarks

Survival analysis is a natural fit to the structure of the dividend initiation data. As introduced above, a survival analysis variable of interest is the time until an event occurs. Our variable of interest is the time until dividend initiation. We can measure time relative to the incorporation of a firm or relative to the listing of a firm. There is censoring present in the data:

- At the end of the study, some firms are yet to initiate dividends.
- Some firms are lost during the course of the study. They may have merged with other firms, reprivatized, or gone bankrupt.

The logistic regression approach of Fama and French (2001) treats the data in a somewhat artificial manner. It does not take into account the time-ordered structure of the data. If given a data matrix of thirty observations, the logistic regression approach fails to take into consideration whether the data represents six firms tracked over a five year period or thirty firms tracked over a one year period. Therefore, this approach ignores the firm-level dynamic nature of dividend policy. Furthermore, the logistic regression approach does not cleanly deal with censorings.

However, the dividend initiation data does present some challenges to the standard survival analysis framework. As stated previously, a usual assumption employed in survival analysis is that of independent censoring. This is probably not a valid assumption to place upon the firms that dropped out of the sample due to bankruptcy.
A firm that was censored due to poor performance was probably not considering initiating dividends at any time in its foreseeable future. We do, however, feel that the independent censoring assumption is valid for firms that were censored due to mergers and reprivatizations. Therefore, we assert that our dataset contains both dependent and independent censoring.

3.2 Introduction to copulas

A copula is a multivariate distribution function whose marginal distributions are uniform on the interval [0, 1]. By making use of the probability integral transformation, which allows us to transform a random variable (by way of its cumulative distribution function) into a Uniform[0, 1] random variable, we see that copulas provide a method for modeling the dependence structure among random variables without having to first make assumptions regarding their marginal distributions.

3.2.1 Definition

For a mathematical definition of the bivariate copula, we give the following excerpts from Nelsen (1999). Nelsen begins by defining a subcopula and then introduces a copula as a subcopula with domain \( I^2 = [0, 1] \times [0, 1] \). First, preliminary definitions must be given.

- Let \( \mathbb{R} \) denote the real line \((-\infty, +\infty)\) and \( \overline{\mathbb{R}} \) denote the extended real line \([-\infty, +\infty]\). Let \( I = [0, 1] \).

- "A 2-place real function \( H \) is a function whose domain, \( \text{Dom}H \), is a subset of
\( \mathbb{R}^2 \) and whose range, \( \text{Ran}H \), is a subset of \( \mathbb{R}^2 \) (Nelsen, 1999, p. 6).

- "Let \( S_1 \) and \( S_2 \) be nonempty subsets of \( \mathbb{R} \), and let \( H \) be a function such that \( \text{Dom}H = S_1 \times S_2 \). Let \( B = [x_1, x_2] \times [y_1, y_2] \) be a rectangle all of whose vertices are in \( \text{Dom}H \). Then the \( H \)-volume of \( B \) is given by

\[
V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)
\]

(Nelsen, 1999, p. 6).

- "A 2-place real function \( H \) is 2-increasing if \( V_H(B) \geq 0 \) for all rectangles \( B \) whose vertices lie in \( \text{Dom}H \)" Nelsen (1999, p. 6).

- "Suppose \( S_1 \) has a least argument \( a_1 \) and that \( S_2 \) has a least argument \( a_2 \). We say that a function \( H \) from \( S_1 \times S_2 \) into \( \mathbb{R} \) is grounded if \( H(x, a_2) = 0 = H(a_1, y) \) for all \( (x, y) \) in \( S_1 \times S_2 \)" (Nelsen, 1999, p. 7).

**Definition 3.1.** "A two-dimensional subcopula (or 2-subcopula, or briefly, a subcopula) is a function \( C' \) with the following properties:

1. \( \text{Dom}C' = S_1 \times S_2 \), where \( S_1 \) and \( S_2 \) are subsets of \( I \) containing 0 and 1;
2. \( C' \) is grounded and 2-increasing;
3. For every \( u \) in \( S_1 \) and every \( v \) in \( S_2 \),

\[
C'(u, 1) = u \quad \text{and} \quad C'(1, v) = v
\]

(Nelsen, 1999, p. 8).

**Definition 3.2.** "A two-dimensional copula (or 2-copula, or briefly, a copula) is a 2-subcopula \( C \) whose domain is \( I^2 \).
Equivalently, a copula is a function \( C \) from \( I^2 \) to \( I \) with the following properties:
1. For every $u, v$ in $I$, 
\[ C(u, 0) = 0 = C(0, v) \]
and
\[ C(u, 1) = u \quad \text{and} \quad C(1, v) = v; \]
2. For every $u_1, u_2, v_1, v_2$ in $I$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, 
\[ C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0 \]
(Nelsen, 1999, p. 8).

Although we have provided the definition of the bivariate copula, it is worth noting that copulas can be further extended into higher dimensions. In this dissertation we shall limit our discussion to the two-dimensional case.

### 3.2.2 Archimedean copulas

Genest and MacKay (1986) provide a nice introduction to a special group of two-dimensional copulas, denoted $H(x, y)$, known as the Archimedean subclass.

**Definition 3.3.** "Consider a class $\Phi$ of functions $\phi : [0, 1] \rightarrow [0, \infty]$ that have two continuous derivatives on $(0,1)$ and satisfy:
\[ \phi(1) = 0, \quad \phi'(t) < 0, \quad \phi''(t) > 0 \]
for all $0 \leq t \leq 1$...[These conditions ensure that $\phi$ has an inverse with two derivatives]...Every member $\phi$ of the class $\Phi$ generates a bivariate distribution function for the pair $(X, Y)$ as follows:
\[ H(x, y) = \phi^{-1}[\phi(x) + \phi(y)] \quad \text{if} \ \phi(x) + \phi(y) \leq \phi(0) \]
\[ = 0 \quad \text{otherwise} \]


The authors provide a short list of elementary properties of $H(x, y)$. We briefly discuss three of these properties.
• The marginal distributions of $X$ and $Y$ are uniform on the interval $[0, 1]$.

proof: Let $F$ be the marginal distribution of $X$ and let $G$ be the marginal distribution of $Y$.

$$F_X(x) = \Pr(X \leq x, Y \leq 1) = H(x, 1) = \phi^{-1}[\phi(x) + \phi(1)] = \phi^{-1}[\phi(x) + 0] = x$$

(3.14)

By the same logic, $G_Y(y) = y$.

• $X$ and $Y$ are independent $\Leftrightarrow \phi(t) = -c \log(t)$, where $c > 0$ is arbitrary.

proof ($\Rightarrow$): Show that if $X$ and $Y$ are independent, then $\phi(t) = -c \log(t)$.

We have already shown that $F_X(x) = x$ and $G_Y(y) = y$

Under independence, $H(x, y) = F_X(x) \cdot G_Y(y) = xy$. To find the independence copula, we need to solve the equation

$$\phi(xy) = \phi[H(xy)] = \phi[\phi^{-1}[\phi(x) + \phi(y)]] = \phi(x) + \phi(y).$$

(3.15)

Differentiate Eqn. (3.15) with respect to $x$:

$$\phi'(xy) \cdot y = \phi'(x).$$

(3.16)

Differentiate Eqn. (3.16) with respect to $y$:

$$\phi''(xy) \cdot y \cdot x + \phi'(xy) = 0.$$

(3.17)

Rearrange terms:

$$\phi''(xy) \cdot y \cdot x = -\phi'(xy).$$

(3.18)
Perform change of variables and notation. Let $xy = t$ and $\phi'(xy) = g(t)$.

Rewriting Eqn. (3.18), we have:

$$g'(t) \cdot t = -g(t) \quad \text{or} \quad \frac{g'(t)}{g(t)} = \frac{-1}{t}.$$ (3.19)

Integrating Eqn. (3.19):

$$\log(g(t)) = -\log(t) + k = \log\left(\frac{k'}{t}\right).$$ (3.20)

$$\Rightarrow g(t) = \frac{k'}{t}.$$ (3.21)

As defined above, $g(t) = \phi'(t)$. Since $\phi'(t)$ is negative and $t = xy$ is positive, $k'$ is negative.

Integrate Eqn. (3.21) to find $\phi(t)$:

$$\int \phi'(t)dt = \int \frac{k'}{t} dt.$$ (3.22)

$$\phi(t) = k' \log(t) + r.$$ (3.23)

Since $\phi(1) = 0, r = 0$. Let $k' = -c$. Then we have, as desired,

$$\phi(t) = -c \log(t), \quad \text{where} \quad c > 0.$$ (3.24)

proof ($\Leftarrow$): Show that if $\phi(t) = -c \log(t)$, then $X$ and $Y$ are independent.

To show that $X$ and $Y$ are independent, we need to show that $H(x, y) = F_X(x) \cdot G_Y(y) = xy$. 
Find $\phi^{-1}(t)$:

\[
x = -c \log(y) \quad \Rightarrow \quad \frac{x}{-c} = \log(y) \quad \Rightarrow \quad y = \exp\left\{\frac{x}{-c}\right\}
\]

(3.25)

\[
\Rightarrow \phi^{-1}(t) = \exp\left\{\frac{-t}{c}\right\}.
\]

(3.26)

Plug in expressions for $\phi(t)$ and $\phi^{-1}(t)$ into the definition of $H(x, y)$:

\[
H(x, y) = \phi^{-1}[\phi(x) + \phi(y)]
\]

\[
= \exp\left\{\frac{-[-c \log(x) - c \log(y)]}{c}\right\}
\]

\[
= \exp\{\log(x) + \log(y)\}
\]

\[
= xy.
\]

• $\Pr(X > x, Y > y) = H(x, y) - x - y + 1$.

proof: Use the elementary probability property

\[
\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).
\]

(3.27)

Let $A = \{X \leq x\}$ and $B = \{Y \leq y\}$. Then $A \cap B = \{X \leq x, Y \leq y\}$ and $\Pr(A \cap B) = H(x, y)$. Since $X \sim \text{Uniform}[0, 1]$ and $Y \sim \text{Uniform}[0, 1]$, $\Pr(A) = x$ and $\Pr(B) = y$. Plugging into Eqn. (3.27), we get:

\[
H(x, y) = \Pr(A \cap B)
\]

\[
= \Pr(A) + \Pr(B) - \Pr(A \cup B)
\]

\[
= x + y - [1 - \Pr(X > x, Y > y)].
\]

(3.28)
Rearranging Eqn. (3.28) yields

\[ \Pr(X > x, Y > y) = H(x, y) - x - y + 1, \text{ as desired.} \]  

(3.29)

3.2.3 Remarks

By definition, a copula is a joint distribution function of random variables that are Uniform[0, 1] distributed. However, at times we will refer to the copula of the random variables \(X_1, X_2, \ldots, X_N\) with distribution functions \(F_{X_1}, F_{X_2}, \ldots, F_{X_N}\) that are not necessarily uniform. A straightforward application of the probability integral transformation, \(U_1 = F_{X_1}(X_1), U_2 = F_{X_2}(X_2), \ldots, U_N = F_{X_N}(X_N)\), yields random variables that are uniformly distributed. The copula of \(X_1, X_2, \ldots, X_N\) is given by the copula of the transformed \(U_i\) variables.
Chapter 4

Exploring Dividend Initiation Policy: Tracking the Disappearing Dividends

This chapter is stand-alone and can be read independently of the rest of the thesis. We detail our initial efforts at modeling the decreased propensity to pay effect using survival analysis techniques. The body of this chapter is excerpted from a paper written for the finance literature.

Chapter Abstract:

Recent papers in the finance literature report a marked decline in the proportion of dividend-paying firms over the period 1978 to 1999. Some researchers attribute this decline, in part, to a decrease in the propensity of firms to pay dividends. We test the hypothesis that firm age, as measured from incorporation, is a relevant input to dividend initiation decisions, and that its inclusion in a model of dividend initiation helps to account for the decreasing proportion of dividend-paying firms. Our results show that shifting the time origin of the analysis from listing to incorporation reduces the magnitude of the decrease in the propensity to pay. Furthermore, for the NYSE firms in our sample, shifting the time origin leaves the decreased propensity to pay effect statistically insignificant.

KEY WORDS: Life cycle theory, Propensity to pay
4.1 Introduction

The proportion of dividend-paying industrial (i.e. non-financial and non-utility) firms has been declining since the late 1970s. In their study of “disappearing dividends,” Fama and French (2001, p. 3) cite the proportion of publicly traded, industrial, dividend-paying firms to be 66.5% in 1978 and 20.8% in 1999. They hypothesize that this decline is due, in part, to a decrease in the propensity of firms to pay dividends. Given the recent changes in the tax code that reduced the double-taxation burden on dividends, the study of dividend initiation and prevalence is now more timely than ever.

In this chapter, we model the dividend initiation decisions of newly listed firms. Our goals are two fold. First, drawing from life cycle theory, we test the hypothesis that firm age, as measured from incorporation, represents a changing firm characteristic that helps to explain the decreased proportion of dividend-paying firms. If firm age proves to be a significant factor in our model, then its inclusion may reduce the amount of the dividend decline attributed to the decreased propensity to pay effect. Secondly, we investigate whether the decrease in the propensity to pay has affected the two major exchanges, the New York Stock Exchange (NYSE) and the Nasdaq, to the same degree.

We use survival analysis methodology to model the first transition of a firm from non-dividend-payer to dividend-payer. Our results consistently show that the decreased propensity to pay effect is mitigated when the time origin of the analysis is
moved from listing to incorporation. For our sample of NYSE firms, the decreased propensity to pay effect is reduced to the point of becoming statistically insignificant.

A growing body of recent work in the finance literature focuses not only on the modeling of dividend policy, but also on the prevalence of dividends. For example, Grullon and Michaely (2002) explore the substitution effect between dividends and share repurchases. Their findings support the hypothesis that firms are increasingly turning toward share repurchase activity along with or instead of dividends when setting payout policy. DeAngelo et al. (2003) investigate the real and nominal levels of dividend payout and their relationship to the distribution of earnings. They find that the decline in the proportion of dividend-paying firms comes primarily from the loss of firms which pay small dividends. Therefore, despite the decline in the proportion (and number) of dividend-paying firms, the real and nominal levels of dividend payout increased over the period 1978 to 2000.

Fama and French (2001) consider the population of publicly traded, industrial firms and report a marked decline in the proportion of dividend-payers over the period 1978 to 1999. They cite two causes for this reduction: 1) a shift in the characteristics of the population of firms toward the characteristics of firms which, historically, have never paid dividends, and 2) a reduced propensity to pay dividends. Fama and French also identify three firm characteristics that are relevant to the decision to pay dividends: profitability, investment (growth) opportunity, and size. Expanding upon their work and basing our model on life cycle theory, we test the hypothesis that
a fourth characteristic, firm age, should also be considered. We measure age from incorporation of the firm.

The life cycle model of dividend initiation that we develop in this chapter models the first transition of a firm from non-dividend-payer to dividend-payer. We do not seek to explain why dividend-paying firms increase or decrease the level of their payout. Nor do we explain why a firm that stopped paying dividends would choose to restart payments at a future date. In other words, our analysis tracks a firm until payment of its first dividend, and no further. Note, however, that the analysis of Fama and French (2001) is not limited to the decision of non-paying firms to initiate dividends. They study the behavior of current payers (to continue dividends), never payers (to initiate dividends), and former payers (to renew dividends).

The primary tools used in the finance literature for analyzing dividend policy have been ordinary regression and its simple variants, such as logistic regression. However, we use a different methodology to model dividend policy. As our data structure has a time-to-event flavor (with the event of interest being dividend initiation), it is a natural fit to the survival analysis framework. In the context of our analysis, we model the survival times of firms as non-dividend-payers.

In comparison to traditional regression models, the survival analysis approach better allows us to account for censored (i.e. incomplete) observations. There are two sources of censoring in our data: 1) at the end of the study (31 December 2002), some firms are yet to initiate dividends, and 2) some firms are lost during the course
of the study, most likely due to bankruptcy or merger.

Furthermore, the survival analysis approach better allows us to capture the temporal dynamics of dividend policy. The standard logistic regression approach treats the data in a somewhat artificial manner. It does not take into account the time-ordered structure of the data. For example, if given a data matrix of thirty firm-year observations, the logistic regression approach does not take into consideration whether the data represents six firms over a five year period or thirty firms over a one year period. Consequently, the effect of firm age and maturation is not captured by the logistic regression model.

In our analysis, we employ a methodology that preserves the temporal structure of the data by allowing the hazard (or, informally, chance) of dividend initiation to be a function of time. Hence, the effect of age is captured in the model not through a covariate, but by setting the time origin.

Previous studies of dividend policy have begun tracking firms at listing, implicitly setting the time origin of their models at listing. Testing our hypothesis that incorporation-age is a relevant input to dividend policy is done by designating incorporation as the time origin of our model.

This chapter is organized as follows. Section 4.2 describes our proposed model of dividend initiation and discusses the hypotheses we test. Section 4.3 gives a description and preliminary analysis of the data. In Section 4.4, we implement the model on our entire sample of firms in order to examine the effect of considering age. In Sec-
tion 4.5, we separate our sample by exchange and repeat the analysis to explore the existence of differing behaviors on the NYSE and the Nasdaq. Section 4.6 concludes.

4.2 Proposed model and statement of hypotheses

4.2.1 Consideration of age in a model of dividend initiation

Recent empirical research has identified a common set of covariates used to model dividend policy decisions. For example, Fama and French (2001) use profitability, investment opportunity, and size. Grullon and Michaely (2002) use earnings and size in their study of the substitution effect. In a recent study of dividend initiation, Deshmukh (2003) uses cash flow, investment opportunity, size, and the number of shareholders. Based on the literature, a model of dividend initiation should account for profitability (or earnings or cash flow), investment opportunity, size, and, perhaps, repurchase activity.

We test the hypothesis that firm age should also be accounted for in a model of dividend initiation. We measure age from the incorporation year of the firm. If age is a relevant factor, then changes in its distribution among new lists may help explain the decreased proportion of dividend-paying firms.

Our interest in the consideration of age in a model of dividend initiation is motivated by firm life cycle theory. A standard discussion of dividend policy (for example, Ross et al. (2002, p. 521)) states that tax laws dictate that it is generally more advantageous for firms to reinvest earnings in the firm than to use them to pay dividends. This implies that, ideally, firms delay dividend initiation until they run out of invest-
ment opportunities. Young firms with abundant investment opportunities will refrain from paying dividends. Mature firms will structure their payout policy around the fact that investment opportunities are dwindling. We refer to this as a life cycle model for dividend initiation. In short, we hypothesize that as firms age, they naturally develop a higher tendency to utilize dividends as a method for disbursing earnings.

The assumed relationship between investment (growth) opportunities and age underlying our model is supported by both theoretical and empirical research. Jovanovic (1982) offers a theoretical model for the relationship between size and growth. His model draws upon the evidence that smaller firms grow faster than larger firms. Jovanovic theorizes that differences in efficiency are the cause, and that small efficient firms grow and survive, while inefficient firms fail. Thus, Jovanovic's theory is consistent with our assumption that growth patterns are a life cycle characteristic of firms.


We note that the market conditions of the 1980s and early 1990s encouraged firms to go public relatively early in their life cycles. Furthermore, a sharp increase in the annual number of new lists began in 1979 (Fama and French, 2003). We wish to ascertain whether the decline in the proportion of dividend-paying firms over
the period 1978 to 1999 is partially attributable to the fact that the market was flooded with many new lists that were extraordinarily young by incorporation-age standards. In this manner, part of the dividend decline that was previously attributed to a decrease in the propensity to pay dividends may be, in fact, another example of shifting firm characteristics. In this instance, the shift is toward younger, more immature firms which have not yet begun to exhaust their growth opportunities and turn to dividends as an outlet for earnings.

4.2.2 Stratification by exchange

The second topic we explore is the presence of differing dividend initiation behavior across exchanges, namely the NYSE and the Nasdaq. The most important difference between the NYSE and the Nasdaq is their market structure. The NYSE has a specialist auction structure, while the Nasdaq has a dealer market structure. The exchanges also have different listing requirements, with the NYSE requirements generally being more restrictive.

It may be that by their structure and requirements, these two exchanges attract fundamentally different populations of firms. Firms may be forced to list on the Nasdaq because they are unable to meet the NYSE's listing requirements. There is evidence that firms are more likely to list on the same exchange as their competitors; technology stocks tend to gravitate toward the Nasdaq. A certain type of firm may value the prestige associated with listing on the NYSE.

Several researchers have found evidence of higher transactions costs on the Nasdaq
exchange. The papers of Bessembinder (1999), Weston (2000), and He and Wu (2003) compare execution costs on the Nasdaq relative to the NYSE before and after the 1997 Nasdaq order-handling reforms. Although the papers give differing opinions as to the extent and manner in which the reforms reduced trade execution costs on the Nasdaq, all agree that Nasdaq costs were significantly higher prior to 1997.

Relatedly, Freund and Webb (1999) analyze correlations between trading volume and measures of market volatility. They find Nasdaq volume to be highly correlated with stock-specific variance measures, whereas NYSE and AMEX volume is highly correlated with market variance measures. They interpret this as evidence that the type and quantity of information driving the Nasdaq exchange differs from that driving the NYSE and the AMEX. Bessembinder and Kaufman (1997) find that the returns of Nasdaq stocks are more volatile than the returns of NYSE stocks, even when firm size is accounted for.

Because of the differing market mechanisms and information driving these exchanges (and also due to differences in efficiency and listing populations), we hypothesize that NYSE and Nasdaq firms weight the inputs to a model of dividend initiation differently. For example, since firms have a preference for smoothing dividends and strongly avoid establishing a payout level they will be unable to sustain, we hypothesize that firms on the more volatile Nasdaq exchange will be less sensitive to profitability figures when deciding whether or not to initiate dividends. We also hypothesize that the rise of the high-tech industry (which is perceived to be highly
resistant to dividends) has led to a greater decreased propensity to pay effect on the tech-heavy Nasdaq.

4.3 Data

Our datasets were constructed using three sources. We began with a dataset compiled by Jovanovic and Rousseau (2001) which supplied the years of listing, incorporation, and founding for a sample of 7,732 firms. For these firms, dividend and covariate information were obtained using the CRSP and Compustat databases, respectively. Accordingly, we created a CRSP dataset and a Compustat dataset. To appear in the CRSP dataset, a firm must have sufficient coverage in CRSP that we can determine its year of dividend initiation. To appear in the Compustat dataset, a firm must appear in the CRSP dataset and have sufficient covariate coverage over its lifetime as a non-dividend-paying firm. The CRSP dataset is used to estimate the survival curves. The Compustat dataset is used to estimate the hazard function. Firms may appear in the CRSP dataset without appearing in the Compustat dataset (due to insufficient covariate coverage). However, all firms in the Compustat dataset also appear in the CRSP dataset.

In order to analyze the decreasing proportion of dividend-paying firms, we restrict our analysis to firms which list in two time intervals, 1965 to 1975 and 1985 to 1995. These intervals were chosen to exploit two important events documented by Fama and French: the 1978 peak in the proportion of dividend-paying firms and the 1979 start of the surge in the annual number of new lists. In this manner, the listing years
of 1965 to 1975 are set as the base period, and relative to it we measure the decreased propensity to pay effect present in the firms which listed from 1985 to 1995.

After filtering the data (e.g. removing financials and utilities, AMEX firms, and firms with insufficient CRSP and/or Compustat coverage) and narrowing it down to the designated listing intervals, our final CRSP dataset contains 1,207 firms, 249 from 1965-75 and 958 from 1985-95. Our final Compustat dataset contains 1,008 firms, 209 from 1965-75 and 799 from 1985-95. Further details on data filtering and formatting can be found in Appendix A.

4.3.1 Simple statistics of the CRSP sample

<table>
<thead>
<tr>
<th>Listing Group</th>
<th>n</th>
<th>Mean</th>
<th>S.D.</th>
<th>10th Pctl.</th>
<th>Median Pctl.</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965-75</td>
<td>249</td>
<td>30.62</td>
<td>25.77</td>
<td>3</td>
<td>22</td>
<td>114</td>
</tr>
<tr>
<td>1985-95</td>
<td>958</td>
<td>9.83</td>
<td>15.98</td>
<td>0</td>
<td>22</td>
<td>130</td>
</tr>
</tbody>
</table>

In Table 4.1, we analyze the age at listing of firms in our two groups. Age at listing is defined as the difference between the listing year and the incorporation year. This table establishes that the 1985-95 group tends to be younger at listing than the 1965-75 group. The median age at listing for the 1965-75 group is 22 years (with an average age of 30.62 years), while the median age at listing for the 1985-95 group is 5 years (with an average age of 9.83 years).

The sample contains firms from two exchanges, the NYSE and the Nasdaq. The
distribution of firms by exchange is shown in Table 4.2. The earlier listing group contains 174 NYSE firms and 75 Nasdaq firms. (The Nasdaq exchange was founded in 1971, so there are only five years of Nasdaq coverage falling within the listing interval of the 1965-75 group.) The later listing group contains 203 NYSE firms and 755 Nasdaq firms.

The percentage of payers and non-payers in each listing group is shown in Table 4.3. As expected, a higher proportion of the 1965-75 listing group initiates dividends. Over 88% (221 firms) of the 1965-75 listing group are classified as dividend-payers, while only 27% (260 firms) of the 1985-95 listing group are classified as such.

We have verified that the characteristics of the Compustat sample are comparable.
4.3.2 Simple statistics of the covariates

Table 4.4 gives summary statistics for our measures of profitability, investment opportunity, and size. These measures are used as covariates in our hazard models of dividend initiation. We use the same covariate definitions as Fama and French (2001, p. 19, 41) in their study of disappearing dividends. Profitability is calculated as \( \frac{E_t}{A_t} \), where \( E_t \) is earnings before interest but after taxes and \( A_t \) is assets. Investment opportunity is measured by the growth rate of assets, \( \frac{(A_t - A_{t-1})}{A_t} \). Size is measured by the percentage of NYSE firms with the same or lower market capitalization, referred to as \( NY P_t \). Each firm in the Compustat sample is tracked from the year following listing (due to the lagged structure of the investment opportunity covariate) up to and including the year of dividend initiation. Summary statistics for each covariate are displayed by decade and listing group.

Looking at profitability and investment opportunity (Panels A and B), a general theme is higher variability in the covariates among the 1985-95 listing group. This is consistent with previous research on new list profitability and survivorship. Fama and French (2003) hypothesize that reductions in the cost of equity capital have made public equity financing available to a wider population of firms. Many of these new inclusions are growth firms who were previously unable to secure public equity financing due to poor short-term profitability outlooks. We also note the presence of extreme observations, especially in the minimum values.
Table 4.4: Summary Statistics for Profitability, Investment Opportunity, and Size

Panel A: Summary Statistics for Profitability, $E_t/A_t$

<table>
<thead>
<tr>
<th>Listing Group</th>
<th>Time Interval</th>
<th>Firm-Yrs</th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965-75</td>
<td>1966-1969</td>
<td>98</td>
<td>0.094</td>
<td>0.091</td>
<td>0.030</td>
<td>0.033</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td>1970-1979</td>
<td>382</td>
<td>0.092</td>
<td>0.089</td>
<td>0.076</td>
<td>-0.646</td>
<td>0.335</td>
</tr>
<tr>
<td></td>
<td>1980-1989</td>
<td>123</td>
<td>0.060</td>
<td>0.076</td>
<td>0.089</td>
<td>-0.320</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td>1990-2001</td>
<td>66</td>
<td>-0.014</td>
<td>0.043</td>
<td>0.244</td>
<td>-1.095</td>
<td>0.227</td>
</tr>
<tr>
<td>1985-95</td>
<td>1985-1989</td>
<td>463</td>
<td>0.007</td>
<td>0.083</td>
<td>0.355</td>
<td>-4.768</td>
<td>0.443</td>
</tr>
<tr>
<td></td>
<td>1990-2001</td>
<td>5788</td>
<td>-0.027</td>
<td>0.056</td>
<td>0.328</td>
<td>-7.544</td>
<td>2.999</td>
</tr>
</tbody>
</table>

Panel B: Summary Statistics for Investment Opportunity, $dA_t/A_t$

<table>
<thead>
<tr>
<th>Listing Group</th>
<th>Time Interval</th>
<th>Firm-Yrs</th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965-75</td>
<td>1966-1969</td>
<td>98</td>
<td>0.165</td>
<td>0.133</td>
<td>0.122</td>
<td>-0.057</td>
<td>0.620</td>
</tr>
<tr>
<td></td>
<td>1970-1979</td>
<td>382</td>
<td>0.094</td>
<td>0.138</td>
<td>1.019</td>
<td>-19.486</td>
<td>0.656</td>
</tr>
<tr>
<td></td>
<td>1980-1989</td>
<td>123</td>
<td>0.020</td>
<td>0.055</td>
<td>0.309</td>
<td>-1.521</td>
<td>0.797</td>
</tr>
<tr>
<td></td>
<td>1990-2001</td>
<td>66</td>
<td>0.049</td>
<td>0.057</td>
<td>0.260</td>
<td>-1.106</td>
<td>0.647</td>
</tr>
<tr>
<td>1985-95</td>
<td>1985-1989</td>
<td>463</td>
<td>0.125</td>
<td>0.148</td>
<td>0.625</td>
<td>-11.620</td>
<td>0.818</td>
</tr>
<tr>
<td></td>
<td>1990-2001</td>
<td>5788</td>
<td>0.088</td>
<td>0.107</td>
<td>0.390</td>
<td>-8.547</td>
<td>0.928</td>
</tr>
</tbody>
</table>

Panel C: Summary Statistics for Size, $NYP_t$

<table>
<thead>
<tr>
<th>Listing Group</th>
<th>Time Interval</th>
<th>Firm-Yrs</th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965-75</td>
<td>1966-1969</td>
<td>98</td>
<td>0.505</td>
<td>0.471</td>
<td>0.209</td>
<td>0.064</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td>1970-1979</td>
<td>382</td>
<td>0.367</td>
<td>0.331</td>
<td>0.269</td>
<td>0.004</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>1980-1989</td>
<td>123</td>
<td>0.324</td>
<td>0.234</td>
<td>0.277</td>
<td>0.003</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>1990-2001</td>
<td>66</td>
<td>0.356</td>
<td>0.425</td>
<td>0.262</td>
<td>0.005</td>
<td>0.757</td>
</tr>
<tr>
<td>1985-95</td>
<td>1985-1989</td>
<td>463</td>
<td>0.200</td>
<td>0.104</td>
<td>0.217</td>
<td>0.001</td>
<td>0.910</td>
</tr>
<tr>
<td></td>
<td>1990-2001</td>
<td>5788</td>
<td>0.292</td>
<td>0.226</td>
<td>0.270</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
4.4 Consideration of age in a model of dividend initiation

In this section, we apply our life cycle model of dividend initiation to our sample of firms. Our primary objective is to explore the effect of considering firm age in a model of dividend initiation. We focus on the estimation of two functions, the survival function and the hazard function. We use two common survival analysis tools, the Kaplan-Meier estimator and the Cox Proportional Hazards (PH) model (see Kalbfleisch and Prentice (2002)), to model these functions. In our analysis, we are measuring survival with respect to a firm's lifetime as a non-dividend-payer.

4.4.1 Survival curve estimation

Before imposing a parametric model structure on dividend initiation and taking into account covariate effects, we first estimate Kaplan-Meier survival curves for our sample.

In Figure 4.1, we present a pair of graphs, with survival measured from (A) listing and (B) incorporation. Within each graph, separate survival curves are calculated for the two listing groups. The thick, solid curve represents the 1965-75 listing group, and the thin, dashed curve represents the 1985-95 listing group. Recall that we are measuring survival in terms of existence as a non-dividend-paying firm. The survival curve corresponding to the 1965-75 group falls below the survival curve of the 1985-95 group, indicating that the earlier group has shorter survival times as non-dividend-paying firms.
Figure 4.1: All Firms: Kaplan-Meier Survival Curves for Dividend Initiation with Time Origin (A) Listing and (B) Incorporation. The thick, solid line corresponds to the 1965-75 listing group. The thin, dashed line corresponds to the 1985-95 listing group.

Table 4.5: Tests of Equality of Survival Curves; All Firms

<table>
<thead>
<tr>
<th>Time Origin</th>
<th>Log-Rank</th>
<th>Wilcoxon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listing</td>
<td>396.3***</td>
<td>370.6***</td>
</tr>
<tr>
<td>Incorporation</td>
<td>25.5***</td>
<td>6.9**</td>
</tr>
</tbody>
</table>

NOTE: *** denotes significance at the 0.001 level; ** denotes significance at the 0.01 level; * denotes significance at the 0.05 level.
When we measure survival as a non-dividend-paying firm using listing as the time origin, we see a large separation between the group curves. However, when we measure survival using incorporation as the time origin, we see a smaller separation between the curves and correspondingly smaller values of the Log-Rank and Wilcoxon statistics in Table 4.5. (The Log-Rank and Wilcoxon statistics test the null hypothesis of equality of the curves.) This indicates that (in the absence of controlling for covariates) the dividend initiation behaviors of the two groups are more similar when we reference survival time from incorporation. Note, however, that for both time origins we strongly reject the null hypothesis of equality of the curves. The narrowing of the separation between the curves supports our hypothesis that changes in the distribution of age at listing among new lists help to explain the decreased proportion of dividend-paying firms.

4.4.2 Hazard modeling

The Kaplan-Meier approach for estimating the survival function does not take into account covariates. Fama and French's analysis suggests that the proportion of dividend-payers is declining partially in response to changes in firm characteristics. We cannot capture this effect by analyzing survival curves. Therefore, the next step of our analysis involves modeling the hazard of dividend initiation using the Cox PH model.

As before, we investigate two models, first using listing the time origin and then using incorporation as the time origin. For covariates, we use the proxies for prof-
itability, investment opportunity, and size. Our covariates change on an annual basis and are, thus, time-varying. One difficulty in this analysis is that the covariates are made publicly available only after a firm lists. In order to fit a model from incorporation, we must use a method of data imputation for the time period between incorporation and listing. We outline our methods for data imputation in Appendix A. We use the straight-line method for profitability and size and the backfill method for investment opportunity.

As previously noted, our covariates contain many extreme observations. We attempted to control for outliers by winsorizing and power transforming the data. Neither of these methods were sufficient and the models failed to satisfy the proportionality assumption of the Cox PH model. Therefore, we converted our time-varying numerical covariates into time-varying categorical covariates. Each covariate was transformed into dummy variables representing four levels, where level one corresponds to values at or below the 25th percentile and level four corresponds to values above the 75th percentile.

In addition to the time-varying categorical covariates for profitability, investment opportunity, and size, we include an additional fixed covariate. $GRPIND$ is an indicator of listing group affiliation. $GRPIND$ takes the value 0 for firms in the 1965-75 listing group and 1 for firms in the 1985-95 listing group. In this manner, the earlier listing group is set as the baseline and the value of $e^{GRPIND}$ gives the multiplicative effect of later listing group affiliation on the hazard of dividend initiation. This co-
efficient isolates the effect of group affiliation while controlling for covariate effects. Therefore, the exponentiated coefficient of $GRPIND$ expresses the propensity to pay dividends of the 1985-95 listing group relative to that of the 1965-75 listing group.

The exact form of our hazard model of dividend initiation is

$$
\lambda(t) = \lambda_0(t) \exp\{\beta_1 P2(t) + \beta_2 P3(t) + \beta_3 P4(t) + \\
\beta_4 I2(t) + \beta_5 I3(t) + \beta_6 I4(t) + \\
\beta_7 S2(t) + \beta_8 S3(t) + \beta_9 S4(t) + \beta_{10} GRPIND\} \tag{4.1}
$$

where $\lambda_0(t)$ is the baseline hazard function and $P2$, $P3$, and $P4$ denote the second, third, and fourth categorical levels of profitability (with similar conventions for investment opportunity and size). Depending upon the time origin, $t$ references time elapsed from either listing or incorporation.

Table 4.6 and Figure 4.2 display the results of fitting the PH model to the 1,008 firms in our Compustat dataset. Panels A show the model that uses listing as the time origin. Panels B show the model that uses incorporation as the time origin.

The third column of the table reports the exponentiated value of the coefficient. Referring back to our discussion of the $GRPIND$ covariate, this column gives the marginal effect of the covariate on the hazard of dividend initiation, relative to its baseline level. For the profitability, investment, and size covariates, the baseline level corresponds to values at or below the 25th percentile. For the $GRPIND$ covariate, the baseline level corresponds to the 1965-75 listing group.

For example, the value of 1.542 in Panel A for covariate $S3$ indicates that, all
Table 4.6: Cox PH Model for Dividend Initiation; All Firms

Panel A: All Firms; From Listing

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coef</th>
<th>Odds</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>Chisq</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2</td>
<td>0.721</td>
<td>2.055</td>
<td>(1.446, 2.921)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>0.845</td>
<td>2.327</td>
<td>(1.640, 3.303)</td>
<td>28.335***</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>0.933</td>
<td>2.543</td>
<td>(1.789, 3.615)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I2</td>
<td>0.080</td>
<td>1.083</td>
<td>(0.823, 1.426)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I3</td>
<td>-0.237</td>
<td>0.789</td>
<td>(0.589, 1.057)</td>
<td>67.318***</td>
<td></td>
</tr>
<tr>
<td>I4</td>
<td>-1.064</td>
<td>0.345</td>
<td>(0.250, 0.476)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.051</td>
<td>1.052</td>
<td>(0.760, 1.458)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0.433</td>
<td>1.542</td>
<td>(1.132, 2.101)</td>
<td>68.833***</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>0.986</td>
<td>2.681</td>
<td>(2.004, 3.586)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRPIND</td>
<td>-1.952</td>
<td>0.142</td>
<td>(0.116, 0.173)</td>
<td>369.165***</td>
<td></td>
</tr>
</tbody>
</table>

Panel B: All Firms; From Incorporation

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coef</th>
<th>Odds</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>Chisq</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2</td>
<td>0.997</td>
<td>2.711</td>
<td>(1.883, 3.904)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>1.581</td>
<td>4.862</td>
<td>(3.365, 7.026)</td>
<td>315.643***</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>2.117</td>
<td>8.308</td>
<td>(5.693, 12.125)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I2</td>
<td>-0.337</td>
<td>0.714</td>
<td>(0.545, 0.936)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I3</td>
<td>-0.638</td>
<td>0.528</td>
<td>(0.397, 0.703)</td>
<td>43.004***</td>
<td></td>
</tr>
<tr>
<td>I4</td>
<td>-1.011</td>
<td>0.364</td>
<td>(0.264, 0.501)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.286</td>
<td>1.330</td>
<td>(0.952, 1.859)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0.830</td>
<td>2.293</td>
<td>(1.652, 3.183)</td>
<td>136.218***</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>1.633</td>
<td>5.118</td>
<td>(3.714, 7.052)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRPIND</td>
<td>-0.884</td>
<td>0.413</td>
<td>(0.333, 0.513)</td>
<td>63.867***</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: *** denotes significance at the 0.001 level; ** denotes significance at the 0.01 level; * denotes significance at the 0.05 level.
other covariate values being equal, the hazard of dividend initiation for a firm with size between the 50th and 75th percentiles is 1.5 times the hazard for a firm with size at or below the 25th percentile. While the value 1.542 technically corresponds to the hazard ratio, we shall keep with more commonly used terminology and refer to it as the odds ratio.

For the GRPIND covariate, the Chisq column gives a Wald chi-squared test statistic for significance. For the categorical covariates, the Chisq column gives a joint test of significance for all levels of the covariate.
Figure 4.2 provides a graphical view of the categorical coefficients. Note that level one, the baseline level in the model, is represented on the graph with a coefficient value of zero. Based on life cycle theory, we hypothesize that the hazard of dividend initiation should increase as a firm becomes more profitable or grows larger. Likewise, the hazard of dividend initiation should decrease as investment opportunities become greater. Therefore, we expect the coefficients for profitability and size to be positive and the coefficients for investment to be negative. We also expect the coefficients for profitability and size to be non-decreasing in level and the coefficients for investment to be non-increasing in level.

We see that the signs and orderings of the coefficients are in agreement with the life cycle hypothesis of dividend initiation. (Although in Table 4.6, Panel A, the coefficient on $I2$ is positive, it is not significantly different from zero.) For both time origins, all four covariates are significant, indicating that the hazard of dividend initiation is a function of profitability, investment opportunity, size, and listing group affiliation.

For the model fit from listing, the coefficient on $GRPIND$ is $-1.952$. This implies that membership in the later listing group reduces the hazard of dividend initiation by a factor of 7 ($\exp\{-1.952\} \approx 0.142 \approx \frac{1}{7}$). This appears to be strong evidence of a decreased propensity to pay effect.

For the model fit from incorporation, the coefficient on $GRPIND$ is $-0.884$, indicating that membership in the later listing group reduces the hazard of dividend
initiation by a factor of 2.5. Although group affiliation is a significant covariate under both time origins, the decrease in the propensity to pay is smaller when the model is constructed using incorporation as the time origin. This is consistent with the conclusions drawn from the survival curves. The dividend initiation behaviors of the two listing groups are significantly different under either choice of time origin. However, the difference is more extreme when survival as a non-dividend-paying firm is measured from listing.

In order to provide a more concrete interpretation of the hazard model and the effect of the GRPIND covariate, Figure 4.3 displays the projected survival curves derived from the PH model fit to our Compustat sample using the listing time origin.

In Panel A, we chart the projected survival curves of a hypothetical growth firm. For years one to five, this firm's covariates are set to level one for profitability and size and level four for investment. For years six to fifteen, the firm's covariates are set to level two for profitability and size and level three for investment. Past year fifteen, the firm has matured and growth has levelled off. The covariates are set to level three for profitability and size and level two for investment.

The solid curve plots the survival prospects of this firm in the 1965-75 listing group, while the dashed curve plots survival for this firm in the 1985-95 listing group. The decreased propensity to pay effect can clearly be seen. Ten years after listing, the projected probability that the growth firm in the 1965-75 listing group remains a non-dividend-payer is 52.5%. If the firm is a member of the 1985-95 listing group,
Figure 4.3: *Projected Survival Curves Obtained from the Cox PH Model for a (A) Hypothetical Growth Firm and (B) Hypothetical Established Firm.* The thick, solid line corresponds to the 1965-75 listing group. The thin, dashed line corresponds to the 1985-95 listing group.

the probability is 91.2%.

In Figure 4.3, Panel B, we show the projected survival curves of a hypothetical established firm in each listing group. Our hypothetical established firm has reached a steady state of level three profitability and size and level one investment. Ten years after listing, the projected probability that the established firm in the 1965-75 group is yet to initiate dividends is 2.3%. For the established firm in the 1985-95 listing group, the probability of waiting at least ten years to initiate dividends is 58.6%.
4.5 Analysis of exchange-specific dividend initiation behavior

In this section, we explore the exchange-specific dividend initiation behavior of the NYSE and the Nasdaq. We repeat the analysis performed in the previous section, but model the exchanges separately. Our primary topic of investigation is whether the magnitude of the decreased propensity to pay effect differs between the exchanges. Secondly, we investigate whether the exchanges weight the inputs to the model of dividend initiation differently.

4.5.1 Exchange-specific survival curve estimation

The survival curves fit to the 377 NYSE firms in the CRSP sample are shown in Figure 4.4. For the 1965-75 listing group, 166 firms initiate dividends and 8 are censored. For the 1985-95 listing group, 114 firms initiate dividends and 89 are censored. The survival curves fit to the 830 Nasdaq firms are shown in Figure 4.5. For the 1965-75 listing group, 55 firms initiate dividends and 20 are censored. For the 1985-95 listing group, 146 firms initiate dividends and 609 are censored.

In comparison to the survival curves generated for the entire sample of firms in Figure 4.1, we see that the curves generated from listing (Panels A) all have the same basic shape. The NYSE curves are lower than the corresponding Nasdaq curves, indicating a higher prevalence of dividends among NYSE firms. Overall, the curves generated from listing consistently point to a significant difference in dividend initiation behavior between the groups. For all pairs of curves generated from listing,
Figure 4.4: NYSE Firms: Kaplan-Meier Survival Curves for Dividend Initiation with Time Origin (A) Listing and (B) Incorporation. The thick, solid line corresponds to the 1965-75 listing group. The thin, dashed line corresponds to the 1985-95 listing group.

Table 4.7: Tests of Equality of Survival Curves; NYSE Firms

<table>
<thead>
<tr>
<th>Time Origin</th>
<th>Log-Rank</th>
<th>Wilcoxon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listing</td>
<td>70.7***</td>
<td>66.3***</td>
</tr>
<tr>
<td>Incorporation</td>
<td>1.1</td>
<td>15.9***</td>
</tr>
</tbody>
</table>

NOTE: *** denotes significance at the 0.001 level; ** denotes significance at the 0.01 level; * denotes significance at the 0.05 level.
Figure 4.5: Nasdaq Firms: Kaplan-Meier Survival Curves for Dividend Initiation with Time Origin (A) Listing and (B) Incorporation. The thick, solid line corresponds to the 1965-75 listing group. The thin, dashed line corresponds to the 1985-95 listing group.

### Table 4.8: Tests of Equality of Survival Curves; Nasdaq Firms

<table>
<thead>
<tr>
<th>Time Origin</th>
<th>Log-Rank</th>
<th>Wilcoxon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listing</td>
<td>149.5***</td>
<td>151.9***</td>
</tr>
<tr>
<td>Incorporation</td>
<td>38.3***</td>
<td>38.8***</td>
</tr>
</tbody>
</table>

NOTE: *** denotes significance at the 0.001 level; ** denotes significance at the 0.01 level; * denotes significance at the 0.05 level.
the Log-Rank and Wilcoxon statistics given in Tables 4.5, 4.7, and 4.8 reject the null hypothesis of equality of the curves.

Focusing our attention on the curves generated from incorporation, we see something different. For the Nasdaq sample (Figure 4.5) we see a plot much like Figure 4.1. In comparison to the listing curves, the incorporation curves are closer together, but still significantly different. However, for the NYSE sample (Figure 4.4) the curves generated from incorporation look much different. The curve for the 1985-95 group lies below that of the 1965-75 group for the first fifty years post-incorporation. The Log-Rank statistic in Table 4.7 fails to reject the null hypothesis of equality of the curves.

In summary, the survival curves fit to the Nasdaq sample tell the same story as the survival curves fit to the entire sample. The dividend initiation behaviors of the two listing groups differ significantly under either choice of time origin, but the difference is less extreme when the curves are generated from incorporation. For the NYSE sample under the incorporation time origin, we see evidence that dividend initiation behaviors do not differ between the listing groups.

4.5.2 Exchange-specific hazard modeling

Table 4.9 displays the results of fitting the Cox PH model for dividend initiation to the 334 NYSE firms in the Compustat sample. Table 4.10 displays the results of fitting the model to the 674 Nasdaq firms in the Compustat sample. The signs and orderings of the coefficients for all four models are consistent with life cycle theory.
Table 4.9: Cox PH Model for Dividend Initiation; NYSE Firms

Panel A: NYSE Firms; From Listing

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coef</th>
<th>Odds</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>Chisq</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2</td>
<td>0.005</td>
<td>1.005</td>
<td>1.005</td>
<td>(0.688, 1.467)</td>
<td>3.468</td>
</tr>
<tr>
<td>P3</td>
<td>0.232</td>
<td>1.261</td>
<td>1.261</td>
<td>(0.881, 1.807)</td>
<td>30.221***</td>
</tr>
<tr>
<td>P4</td>
<td>0.267</td>
<td>1.306</td>
<td>1.306</td>
<td>(0.903, 1.888)</td>
<td>30.221***</td>
</tr>
<tr>
<td>I2</td>
<td>-0.195</td>
<td>0.823</td>
<td>0.823</td>
<td>(0.578, 1.172)</td>
<td>30.221***</td>
</tr>
<tr>
<td>I3</td>
<td>0.015</td>
<td>1.015</td>
<td>1.015</td>
<td>(0.722, 1.426)</td>
<td>30.221***</td>
</tr>
<tr>
<td>I4</td>
<td>-1.015</td>
<td>0.362</td>
<td>0.362</td>
<td>(0.238, 0.551)</td>
<td>30.221***</td>
</tr>
<tr>
<td>S2</td>
<td>-0.012</td>
<td>0.988</td>
<td>0.988</td>
<td>(0.672, 1.452)</td>
<td>30.221***</td>
</tr>
<tr>
<td>S3</td>
<td>0.436</td>
<td>1.547</td>
<td>1.547</td>
<td>(1.087, 2.203)</td>
<td>30.221***</td>
</tr>
<tr>
<td>S4</td>
<td>0.299</td>
<td>1.348</td>
<td>1.348</td>
<td>(0.928, 1.959)</td>
<td>30.221***</td>
</tr>
<tr>
<td>GRPIND</td>
<td>-1.047</td>
<td>0.351</td>
<td>0.351</td>
<td>(0.270, 0.456)</td>
<td>30.221***</td>
</tr>
</tbody>
</table>

Panel B: NYSE Firms; From Incorporation

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coef</th>
<th>Odds</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>Chisq</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2</td>
<td>0.362</td>
<td>1.436</td>
<td>1.436</td>
<td>(0.973, 2.118)</td>
<td>66.410***</td>
</tr>
<tr>
<td>P3</td>
<td>1.003</td>
<td>2.727</td>
<td>2.727</td>
<td>(1.864, 3.990)</td>
<td>66.410***</td>
</tr>
<tr>
<td>P4</td>
<td>1.547</td>
<td>4.697</td>
<td>4.697</td>
<td>(3.140, 7.027)</td>
<td>66.410***</td>
</tr>
<tr>
<td>I2</td>
<td>-0.606</td>
<td>0.545</td>
<td>0.545</td>
<td>(0.385, 0.773)</td>
<td>66.410***</td>
</tr>
<tr>
<td>I3</td>
<td>-0.466</td>
<td>0.627</td>
<td>0.627</td>
<td>(0.446, 0.882)</td>
<td>66.410***</td>
</tr>
<tr>
<td>I4</td>
<td>-1.000</td>
<td>0.368</td>
<td>0.368</td>
<td>(0.242, 0.558)</td>
<td>66.410***</td>
</tr>
<tr>
<td>S2</td>
<td>0.265</td>
<td>1.304</td>
<td>1.304</td>
<td>(0.880, 1.931)</td>
<td>66.410***</td>
</tr>
<tr>
<td>S3</td>
<td>1.002</td>
<td>2.724</td>
<td>2.724</td>
<td>(1.853, 4.003)</td>
<td>49.217***</td>
</tr>
<tr>
<td>S4</td>
<td>1.276</td>
<td>3.582</td>
<td>3.582</td>
<td>(2.378, 5.394)</td>
<td>49.217***</td>
</tr>
<tr>
<td>GRPIND</td>
<td>-0.081</td>
<td>0.922</td>
<td>0.922</td>
<td>(0.700, 1.215)</td>
<td>0.329</td>
</tr>
</tbody>
</table>

NOTE: *** denotes significance at the 0.001 level; ** denotes significance at the 0.01 level; * denotes significance at the 0.05 level.
### Table 4.10: Cox PH Model for Dividend Initiation; Nasdaq Firms

**Panel A: Nasdaq Firms; From Listing**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coef</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>Chisq</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2</td>
<td>1.189</td>
<td>3.283</td>
<td>(1.665, 6.474)</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>1.549</td>
<td>4.706</td>
<td>(2.408, 9.198)</td>
<td>24.194***</td>
</tr>
<tr>
<td>P4</td>
<td>1.616</td>
<td>5.033</td>
<td>(2.561, 9.888)</td>
<td></td>
</tr>
<tr>
<td>I2</td>
<td>0.160</td>
<td>1.173</td>
<td>(0.746, 1.845)</td>
<td></td>
</tr>
<tr>
<td>I3</td>
<td>-0.450</td>
<td>0.638</td>
<td>(0.390, 1.042)</td>
<td>20.736***</td>
</tr>
<tr>
<td>I4</td>
<td>-0.717</td>
<td>0.488</td>
<td>(0.297, 0.802)</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.404</td>
<td>1.497</td>
<td>(0.872, 2.570)</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0.602</td>
<td>1.825</td>
<td>(1.077, 3.092)</td>
<td>23.220***</td>
</tr>
<tr>
<td>S4</td>
<td>1.118</td>
<td>3.059</td>
<td>(1.830, 5.113)</td>
<td></td>
</tr>
<tr>
<td>GRPIND</td>
<td>-1.955</td>
<td>0.142</td>
<td>(0.100, 0.200)</td>
<td>121.902***</td>
</tr>
</tbody>
</table>

**Panel B: Nasdaq Firms; From Incorporation**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coef</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>Chisq</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2</td>
<td>1.608</td>
<td>4.991</td>
<td>(2.456, 10.142)</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>2.321</td>
<td>10.183</td>
<td>(5.016, 20.672)</td>
<td>70.155***</td>
</tr>
<tr>
<td>P4</td>
<td>2.840</td>
<td>17.114</td>
<td>(8.310, 35.245)</td>
<td></td>
</tr>
<tr>
<td>I2</td>
<td>-0.343</td>
<td>0.710</td>
<td>(0.448, 1.124)</td>
<td></td>
</tr>
<tr>
<td>I3</td>
<td>-0.873</td>
<td>0.418</td>
<td>(0.256, 0.683)</td>
<td>20.620***</td>
</tr>
<tr>
<td>I4</td>
<td>-0.974</td>
<td>0.378</td>
<td>(0.229, 0.622)</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.512</td>
<td>1.668</td>
<td>(0.959, 2.902)</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0.846</td>
<td>2.330</td>
<td>(1.338, 4.060)</td>
<td>34.029***</td>
</tr>
<tr>
<td>S4</td>
<td>1.466</td>
<td>4.333</td>
<td>(2.493, 7.530)</td>
<td></td>
</tr>
<tr>
<td>GRPIND</td>
<td>-1.251</td>
<td>0.286</td>
<td>(0.197, 0.416)</td>
<td>43.189***</td>
</tr>
</tbody>
</table>

**NOTE:** *** denotes significance at the 0.001 level; ** denotes significance at the 0.01 level; * denotes significance at the 0.05 level.
All coefficients with the opposite sign as hypothesized (e.g. the coefficient on $I3$ in Table 4.9, Panel A) are not significantly different from zero. For all pairs where the ordering is not as hypothesized (e.g. the coefficients on $I2$ and $I3$ in Table 4.9, Panel B), the coefficients do not significantly differ from one another.

For the models constructed from listing (Panel A), the hazard of dividend initiation for the NYSE sample is dependent on investment opportunity, size, and group affiliation. Profitability is not significant. For the Nasdaq sample, the hazard of dividend initiation is dependent upon all four covariates. We also find that under both time origins, the Nasdaq sample is more sensitive to profitability than the NYSE sample. Although we hypothesized that because of the Nasdaq's higher return volatility, its firms would be less sensitive to profitability, we observe completely the opposite relationship in the fitted models.

For the Nasdaq firms, the coefficient on $GRPIND$ is $-1.955$ when the time origin is listing and $-1.251$ when the time origin is incorporation. Thus, when using listing as the time origin, membership in the later listing group decreases the hazard of dividend initiation by a factor of seven. When using incorporation as the time origin, membership in the later listing group decreases the hazard of dividend initiation by a factor of 3.5. Although group affiliation is a significant covariate under both time origins, the decreased propensity to pay effect is smaller in magnitude when incorporation is used.

For the NYSE firms, the coefficient on $GRPIND$ is $-1.047$ when the time origin is
listing and $-0.081$ when the time origin is incorporation. The latter is not significantly different from zero. When using listing as the time origin, membership in the later listing group decreases the hazard of dividend initiation by a factor of three. However, under the incorporation time origin, the \textit{GRPIND} covariate is not significant. In other words, when the model is applied to the NYSE sample and incorporation is set as the time origin, we see no statistically significant evidence of a decrease in the propensity to pay dividends.

4.6 Conclusions

Drawing from life cycle theory, we have tested the hypothesis that a previously unconsidered factor, firm age, should be accounted for in a model of dividend initiation. Our results support the hypothesis that changes in the distribution of incorporation-age among newly listed firms partially explain the decrease in the proportion of dividend-paying firms. In this manner, part of the decrease that was previously attributed to the decreased propensity to pay effect is yet another example of changing firm characteristics. In this instance, the shift is toward younger, more immature firms who have not yet reached their natural age for dividend initiation.

We have also tested the hypothesis that the magnitude of the decrease in the propensity to pay differs across exchanges. In a comparison of a sample of NYSE firms to a sample of Nasdaq firms, our results show a much stronger decrease in the propensity to pay on the Nasdaq. In fact, our PH model finds no statistically significant evidence of a decrease in the propensity to pay for our NYSE sample.
when incorporation is used as the time origin. We hypothesize that differing market structures and listing populations have led to differing dividend initiation behaviors on these two exchanges.

One area for future research is to draw from the work of Grullon and Michaely (2002) and incorporate the effect of share repurchases into our model of dividend initiation. This could be done by including repurchase activity as a covariate or by redefining our event to include the initiation of repurchase activity.

A second line of future investigation regards the bankruptcy censorings in our data. In the analysis presented in this chapter, we have assumed independent censoring. This means that observing a censoring gives us no additional information regarding the firm's underlying lifetime as a non-dividend-paying firm. However, it is plausible to assume that a firm who was censored due to poor performance would not have been considering initiating dividends at any time in its foreseeable future. Under this assumption, censoring by bankruptcy gives us information regarding the firm's underlying lifetime as a non-dividend-payer—namely, that it is quite long.

We shall refer to the bankruptcy censorings in our data as dependent censorings. Of the 1,207 firms in our CRSP sample, 119 are dependent censorings. These 119 firms represent 9.9% of our sample. However, the dependent censorings are not divided equally between the exchanges. 12.3% (102 firms) of our CRSP Nasdaq sample are dependent censorings, while 4.5% (17 firms) of our CRSP NYSE sample are dependent censorings.
It is clear that our analysis of the Nasdaq sample is most strongly affected by the possible biases that are introduced by ignoring the dependent censorings. Likewise, the NYSE sample, with its small percentage of dependent censorings, is least affected.

When estimating the survival curves, a crude method of addressing the dependent censorings is to recode them as censorings at infinity. Under this convention, the survival curves for the NYSE sample are nearly indistinguishable from those shown in Figure 4.4. The Log-Rank and Wilcoxon statistics are virtually unchanged from the values reported in Table 4.7.

Therefore, we have reason to believe that the biases introduced into the analysis of the NYSE sample are negligible and that our conclusions regarding the absence of a decrease in the propensity to pay are valid. In the following chapters, we present a copula-based method for recovering marginal survival that accommodates the censoring mechanism present in the dividend initiation data.
Chapter 5

Data Structure and Dependent Censoring

The purpose of this chapter is to discuss our data structure and to survey previous copula-based efforts in addressing dependent censoring when estimating marginal survival.

In the traditional survival analysis framework, the observable data consist of the pairs \((T_i, \delta_i), i = 1, \ldots, n\), where

\[ T_i = \min(X_i, C_i) \quad \text{and} \quad \delta_i = I_{\{X_i \leq C_i\}}. \]  

(5.1)

In this manner, \(T_i\) is the minimum of the (potential) lifetime random variable \(X_i\) and the censoring random variable \(C_i\). \(\delta_i\) is the indicator of a failure event. The usual assumption made in survival analysis is that of independent censoring—within each pair \((X_i, C_i)\) the lifetime and censoring random variables are independent. Note that this assumption, by the incomplete nature of survival data, is untestable. Intuitively, independent censoring implies that knowing an observation is censored does not give us any additional information about its lifetime, other than the obvious fact that its...
lifetime exceeds its observed time in the study.

Although commonly made, the assumption of independent censoring may not be valid in certain instances. For example, if we are conducting a study of cancer patients and treatment response, independent censoring implies that patients who censor themselves by withdrawing from the study are no more (or less) terminal than patients who remain in the study. However, it is plausible that many patients who decide to leave the study do so because of poor response to the treatment. At the time of removing themselves from the study, these subjects face relatively poor survival prospects. Under this scenario, censoring is informative for survival and dependent censoring is present in the data.

Under independent censoring, it has been well established that the observable data $(T_i, \delta_i)$ are sufficient to uniquely determine the marginal distribution of $X$ and hence, its survival function, $S(t) = \Pr(X > t)$. In this case, the Kaplan-Meier estimator is an appropriate non-parametric estimator of $S$. However, if the independent censoring assumption is violated, the Kaplan-Meier estimator can produce misleading results. For further information, the reader is directed to a paper by Link (1989) in which he conducts a simulation study that illustrates the sensitivity of the Kaplan-Meier estimator to violations of the independent censoring assumption.

A scan of the recent literature reveals many methods for dealing with dependent censoring. We focus on the copula-based methods developed by Zheng and Klein (1994, 1995, 1996). These authors have developed a series of survival estimators that
use copulas to model the dependence structure between the lifetime and censoring random variables. Two of their estimators, the "self-consistent" estimator and the "copula-graphic" estimator are discussed. Rivest and Wells (2001) have furthered the development of the copula-graphic estimator and determined the closed form expression and limiting behavior of the estimator when the copula is a member of the Archimedean subclass.

5.1 Data structure and key result

The Zheng and Klein papers view the data in the dependent competing risks framework. We shall adopt their notation. Let $X$ be the lifetime variable of interest and let $Y$ be the lifetime of another "competing" event. The occurrence of $Y$ precludes us from observing $X$; $Y$ censors the event of interest. The observable data consists of $T_i = \min(X_i, Y_i)$ and $\delta_i = I\{X_i \leq Y_i\}, \ i = 1, \ldots, n$. Note that this interpretation is almost identical to the standard definition given in Eqn (5.1). We have simply replaced $C_i$ with $Y_i$. In the dependent competing risks framework, we place extra emphasis on the fact that $Y$ itself is an event and this event is informative for $X$.

The common assumption made in survival analysis is that $X$ and $Y$ are independent. This is untestable by the very nature of the data. It has been shown that if $X$ and $Y$ are independent, then the observed data are sufficient to uniquely determine the marginal distribution of $X$. Otherwise, some assumptions must be made about the dependence structure between $X$ and $Y$ in order to recover the marginals. Zheng and Klein (1995) show that if the copula of $X$ and $Y$ is known, then the observable
data are sufficient to uniquely determine the marginal distributions of $X$ and $Y$. This is the key result underlying their copula-based survival estimators.

5.2 The self-consistent estimator

Before focusing our attention on the copula-graphic estimator, we first discuss an earlier copula-based estimator devised by Zheng and Klein (1994). Loosely, the idea behind the self-consistent estimator is an attempt at reconstructing the complete bivariate sample $(X_i, Y_i), i = 1, \ldots, n$ from the observable competing risks data. If we had a complete bivariate sample, the obvious non-parametric estimators for marginal survival would be the empirical survival functions

$$\hat{P}_r(X > x) = \hat{S}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(X_i > x)}$$

(5.2)

$$\hat{P}_r(Y > y) = \hat{R}(y) = \frac{1}{n} \sum_{i=1}^{n} I_{(Y_i > y)}.$$  

(5.3)

However, within the competing risks framework, we can only observe points along the ray $X_i = Y_i = T_i$ and the indicator of $T_i = X_i$ or $T_i = Y_i$. Zheng and Klein (1994, p. 2303) explain that "this corresponds to observing directed line segments, starting from $T_i$ in the upper right hand quadrant of the plane."

In order to estimate $S(t)$ from the competing risks data, we need to determine the number of observations with an $X$ coordinate greater than or equal to $t$.

- **Case A:** If $\delta_i = 1$, the relation of $X_i$ to $t$ is clear. For this observation, we have observed an $X$ lifetime, so there is no ambiguity.
• **Case B**: If $\delta_i = 0$, we observe a $Y$ lifetime and the relationship of $X_i$ to $t$ may be uncertain. If $Y_i > t$, then we know $X_i > Y_i > t$ and the relationship is clear. However, if $Y_i < t$, then we do not know whether the underlying $X_i$ exceeds $t$.

Therefore, to estimate the number of observations with $X_i$ greater than or equal to $t$, we must estimate

$$\Pr(X > t | T_i = t_i, \delta_i = 0) = \Pr(X > t | X > t_i, Y = t_i). \quad (5.4)$$

This probability depends on the joint distribution function of $X$ and $Y$, which itself is a function of the copula, $C$, of $X$ and $Y$ and the unknown survival functions $S$ and $R$. If the probability in Eqn (5.4) could be estimated, then $S$ and $R$ would follow from

\begin{align*}
\hat{S}(t) &= \frac{1}{n} \sum_{i=1}^{n} I_{\{t_i > t\}} + \sum_{t_i < t} (1 - \delta_i) \hat{P}r(X > t | X > t_i, Y = t_i) \quad (5.5) \\
\hat{R}(t) &= \frac{1}{n} \sum_{i=1}^{n} I_{\{t_i > t\}} + \sum_{t_i < t} \delta_i \hat{P}r(Y > t | Y > t_i, X = t_i). \quad (5.6)
\end{align*}

The first term on the right hand side of Eqn (5.5) takes care of Case A and the first scenario under Case B. The second term gives the estimated probability necessitated by the second scenario of Case B. According to Zheng and Klein (1994, p. 2303), “Any estimators $\hat{S}(t)$ and $\hat{R}(t)$ for which substitution into the right hand side of Eqns (5.5) and (5.7) yields a fixed point are self-consistent estimators.”

If $X$ and $Y$ are assumed to be independent, $\Pr(X > t | X > t_i) = \frac{S(t)}{S(t_i)}$ and the self-consistent estimator resulting from Eqn (5.5) is the Kaplan-Meier estimator. This result is attributed to Efron (Zheng and Klein, 1994, p. 2304).
If $X$ and $Y$ are assumed to be dependent with known copula $C(u, v)$, then

$$
\hat{P}_r(X > t, \mid X > t_i, Y = t_i) = \frac{1 - C_v(1 - \hat{S}(t), 1 - \hat{R}(t_i))}{1 - C_v(1 - \hat{S}(t_i), 1 - \hat{R}(t_i))}
$$

(5.7)

where $C_v(a, b) = \frac{\partial C(u, v)}{\partial v}$ evaluated at the point $(u, v) = (a, b)$. A corresponding expression exists for $\hat{P}_r(Y > t, \mid Y > t_i, X = t_i)$ (Zheng and Klein, 1994, p. 2304).

Zheng and Klein (1994, p. 2304–2305) give a concise algorithm for calculating the self-consistent estimators:

1. Begin with initial guesses $\hat{S}_0(t)$ and $\hat{R}_0(t)$ and an assumed copula.

2. Using the assumed copula, evaluate Eqn (5.7) and the corresponding equation for $Y$ using $\hat{S}_0(t)$ and $\hat{R}_0(t)$ and plug into the right hand side of Eqns (5.5) and (5.7) to obtain $\hat{S}_1(t)$ and $\hat{R}_1(t)$.

3. Iterate until convergence.

Note that each iteration requires a pass through the data.

### 5.3 The copula-graphic estimator

Zheng and Klein (1995, p. 129) begin their development of the copula-graphic estimator by assuming that $\Pr(X_i = Y_i) = 0$ and noting that the following three quantities are estimable from the data:

$$
k(t) = \Pr(X > t, Y > t)
$$

$$
p_1(t) = \Pr(X \leq t, X < Y)
$$

$$
p_2(t) = \Pr(Y \leq t, Y < X).
$$

(5.8)
Let $F$ and $G$ denote the marginal distribution functions of $X$ and $Y$, respectively. Let $F$ be strictly increasing on $(0, t_1)$ with $F(t_1) = 1$, and let $G$ be strictly increasing on $(0, t_2)$ with $G(t_2) = 1$. Set $t^* = \min(t_1, t_2)$. Zheng and Klein prove that if the copula, $C$, of $X$ and $Y$ is known, then $F$ and $G$ are uniquely determined by \{\{k(t), p_1(t), p_2(t)\}$ on the interval $(0, t^*)$.

The idea behind the copula-graphic estimator is to find estimators of $F$ and $G$ that are consistent with the empirical estimates of $k(t)$ and $p_1(t)$ defined in Eqn (5.8). (Note that $p_2(t)$ need not be considered since $k(t) + p_1(t) + p_2(t) = 1$.)

The copula-graphic estimator is based on a mapping from $x - y$ space to $F - G$ space. If $F$ and $G$ are the marginal distributions of $X$ and $Y$, then for any $t$,

$$k(t) = \Pr(X > t, Y > t) = \int I_{A_t} \, d\mu_c = \mu_c(A_t) \tag{5.9}$$

$$p_1(t) = \Pr(X \leq t, X < Y) = \int I_{B_t} \, d\mu_c = \mu_c(B_t) \tag{5.10}$$

where

$$A_t = \{(u, v) : F(t) < u \leq 1, \ G(t) < v \leq 1\} \tag{5.11}$$

$$B_t = \{(u, v) : 0 \leq u \leq F(t), \ G(F^{-1}(u)) \leq v \leq 1\} \tag{5.12}$$

and $\mu_c$ is the probability measure corresponding to the copula, $C$, of $X$ and $Y$. In the words of Zheng and Klein (1995, p. 130), the goal is to "find estimators $\hat{F}$ and $\hat{G}$ of $F$ and $G$ which preserve these properties [Eqns (5.9) and (5.10)] on a selected grid of $m$ points $0 < t_1 < t_2 < \ldots < t_m < \max\{T_i, i = 1, \ldots, n\}".

To construct the estimators $\hat{F}$ and $\hat{G}$ of the marginals, begin by defining the
following estimates:

$$\hat{A}_t = \{(u,v) : \hat{F}(t) < u \leq 1, \hat{G}(t) < v \leq 1\} \quad (5.13)$$

$$\hat{B}_t = \{(u,v) : 0 \leq u \leq \hat{F}(t), \hat{G}(\hat{F}^{-1}(u)) \leq v \leq 1\} \quad (5.14)$$

$$\hat{k}(t) = \hat{P}_r(X > t, Y > t) = \frac{1}{n} \sum I_{\{T_i > t\}} \quad (5.15)$$

$$\hat{p}_i(t) = \hat{P}_r(X \leq t, X < Y) = \frac{1}{n} \sum I_{\{T_i \leq t, \delta_i = 1\}}. \quad (5.16)$$

Upon plugging these estimates into Eqns (5.9) and (5.10), \(\hat{F}\) and \(\hat{G}\) can be obtained by solving

$$\mu_c(\hat{A}_t) - \hat{k}(t_i) = 0 \quad (5.17)$$

$$\mu_c(\hat{B}_t) - \hat{p}_i(t_i) = 0 \quad (5.18)$$

and defining \(\hat{F}\) and \(\hat{G}\) to be straight lines in the intervals \((t_i, t_{i+1})\). Eqns (5.17) and (5.18) can be solved using a bisection root-finding algorithm.

However, if the grid points for the estimator are chosen to be the unique (i.e. \(X\) only or \(Y\) only) event times, then a simpler algorithm can be used. Using this grid allows us to exploit the fact that for each \(t_i\) representing an \(X\) event (\(\delta_i = 1\)), the estimate for \(F\) should jump, while the estimate for \(G\) should remain unchanged. Similarly, for each \(t_i\) representing a \(Y\) event (\(\delta_i = 0\)), the estimate for \(G\) should jump, while the estimate for \(F\) should remain unchanged. Letting \(t_0 = 0\) and setting \(\hat{F}(t_0) = \hat{G}(t_0) = 0\) we can solve for our estimators using the following two equations.

For \(\delta_i = 1\):

$$\hat{k}(t_i) = 1 - \hat{F}(t_i) - \hat{G}(t_{i-1}) + G\{\hat{F}(t_i), \hat{G}(t_{i-1})\}. \quad (5.19)$$
For $\delta_i = 0$:

$$\hat{k}(t_i) = 1 - \hat{F}(t_{i-1}) - \hat{G}(t_i) + C\{\hat{F}(t_{i-1}), \hat{G}(t_i)\}. \quad (5.20)$$

Note that Eqns (5.19) and (5.20) are direct results of the identity

$$\Pr(X \leq t_i) + \Pr(Y \leq t_i) - \Pr(X \leq t_i, Y \leq t_i) + \Pr(X > t_i, Y > t_i) = 1. \quad (5.21)$$

For example, when $\delta_i = 1$, Eqn (5.19) follows directly from Eqn (5.21) after the following substitutions are made:

$$\Pr(X \leq t_i) = F(t_i)$$
$$\Pr(Y \leq t_i) = G(t_{i-1})$$
$$\Pr(X \leq t_i, Y \leq t_i) = C\{F(t_i), G(t_{i-1})\}$$
$$\Pr(X > t_i, Y > t_i) = k(t_i). \quad (5.22)$$

It is to the estimator defined on this grid of unique event times that Zheng and Klein give the name “copula-graphic” estimator.

Zheng and Klein prove strong consistency and assert that the copula-graphic estimator is a maximum likelihood estimator. (For a proof, they refer the reader to Chapter 1 of Robertson and Uppuluri (1984).) Lastly, under the independence copula, the copula-graphic estimator reduces to the Kaplan-Meier estimator. This result is also given by Rivest and Wells (2001) and will be shown in the next section.

Zheng and Klein (1994, p. 406) report the results of a small scale simulation ($n = 20$; 10,000 trials) comparing the performance of the copula-graphic estimator, the self-consistent estimator, and a third copula-based estimator. They look at the
relative bias \( \{E[\hat{S}(t_p) - p]/p \} \) and relative mean squared error \( E[(\hat{S}(t_p) - p)^2] (p(1-p)) \) of the three estimators. The main conclusion of this simulation is that the self-consistent estimator has significantly larger bias than the other two estimators.

Zheng and Klein (1995, p. 132-133) also investigate the robustness of the copula-graphic estimator to misspecification of the copula. The robustness study suggests that the estimator is more sensitive to the implied strength of association between \( X \) and \( Y \) (commonly measured by Kendall's tau or Spearman's rho) than to the parametric form of the copula.

In real life, the true copula linking \( X \) and \( Y \) is never known. Therefore, the most promising use of a copula-based estimator is to establish bounds on the true survival function. In order to use two copulas, say \( C_1 \) and \( C_2 \), to establish bounds, we must ensure that \( C_1 \) and \( C_2 \) have an ordering, i.e. \( C_1(x,y) \leq C_2(x,y) \) for all \( x,y \), and that the relation somehow relates to the strength of dependence between \( X \) and \( Y \). There exist several common families of copulas that are "monotone in a single parameter \( \alpha \), so that specifying a range of values for this parameter is equivalent to specifying a range of values for Kendall's tau. This gives us a technique for specifying bounds on \( F \) and \( G \), based on a range of associations for an assumed family of copulas" (Zheng and Klein, 1995, p. 134).

5.4 Martingale extension of the copula-graphic estimator

Rivest and Wells (2001) expand upon the work of Zheng and Klein by deriving a
closed form expression for the copula-graphic estimator when the assumed copula is a member of the Archimedean subclass. Rivest and Wells begin by assuming that \( \{T_i, U_i\}_{i=1}^n \) are independent finite failure and censoring times, with observable data \( X_i = \min(T_i, U_i) \) and \( \delta_i = I_{\{T_i \leq U_i\}} \). They also assume that the distribution of \( X \) is absolutely continuous and there are no tied event times.

Let the joint distribution of \( T \) and \( U \) be given by

\[
H(t, u) = \phi^{-1}[\phi(S(t)) + \phi(C(u))] \tag{5.23}
\]

where \( S \) and \( C \) are the marginal survival functions of \( T \) and \( U \), respectively. Eqn (5.23) is familiar, as it is the definition of the Archimedean copula given in Eqn (3.13).

"Under Eqn (5.23), the copula graphic [sic] estimator for the survivor functions \( S(t) \) (resp. \( C(u) \)) is a right continuous decreasing step function \( \hat{S}(t) \) (resp. \( \hat{C}(u) \)) satisfying \( \hat{S}(0) = 1 \) (resp. \( \hat{C}(0) = 1 \)) with jumps at the points \( X_i \) where \( \delta_i = 1 \) (resp. \( \delta_i = 0 \)) such that

\[
\phi^{-1}[\phi(\hat{S}(X_i)) + \phi(\hat{C}(X_i))] = \hat{\pi}(X_i), \quad \text{for } i = 1, \ldots, n, \tag{5.24}
\]

where \( \hat{\pi} \) is the standard estimate of the survival function of \( X \), \( \hat{\pi}(x) = \sum I_{\{X_i \geq x\}}/n \)" (Rivest and Wells, 2001, p. 141).

To derive the closed form expression for \( \hat{S} \), consider the case where \( \delta_i = 1 \). We use the fact that \( \hat{C} \) does not jump at \( X_i \), coupled with the left continuity of \( \hat{\pi} \) (i.e. \( \hat{\pi}(X_i^-) = \hat{\pi}(X_i) \)) and right continuity of \( \hat{S} \) (i.e. \( \hat{S}(X_i^+) = \hat{S}(X_i) \)) to establish the
jump size of $\hat{S}$ at $X_i$:

$$\phi(\hat{S}(X_i^-)) = \phi(\hat{\pi}(X_i^-)) - \phi(\hat{C}(X_i^-)) = \phi(\hat{\pi}(X_i)) - \phi(\hat{C}(X_i^-))$$  \hspace{1cm} (5.25)

$$\phi(\hat{S}(X_i^+)) = \phi(\hat{\pi}(X_i^+)) - \phi(\hat{C}(X_i^+)) = \phi(\hat{\pi}(X_{i+1})) - \phi(\hat{C}(X_i^+))$$  \hspace{1cm} (5.26)

$$\Rightarrow$$

$$\phi(\hat{S}(X_i^-)) - \phi(\hat{S}(X_i)) = \phi(\hat{\pi}(X_i)) - \phi(\hat{\pi}(X_i) - 1/n).$$  \hspace{1cm} (5.27)

If we sum both sides of Eqn (5.27) over the $X_i$ that are less than or equal to $t$ and where $\delta_i = 1$, we get the closed form expression of the copula-graphic estimator of $S$ for Archimedean copulas

$$\hat{S}(t) = \phi^{-1} \left[ - \sum_{X_i \leq t, \delta_i = 1} \phi(\hat{\pi}(X_i)) - \phi(\hat{\pi}(X_i) - 1/n) \right]$$  \hspace{1cm} (5.28)

(Rivest and Wells, 2001, p. 141).

Under the independence copula, $\phi(t) = -\log(t)$, it is easy to see that Eqn (5.28)
reduces to the Kaplan-Meier estimator:

\[
\hat{S}(t) = \phi^{-1} \left[ - \sum_{X_i \leq t, \delta_i = 1} \phi(\hat{X}_i) - \phi(\hat{X}_i - 1/n) \right]
\]

\[
= \exp \left\{ \sum_{X_i \leq t, \delta_i = 1} - \log(\hat{X}_i) + \log(X_i) - 1/n \right\}
\]

\[
= \exp \left\{ \sum_{X_i \leq t, \delta_i = 1} - \log(n^{-1} \sum I_{(X_i \geq t)}) + \log(n^{-1} \sum I_{(X_i \geq t) - n^{-1}}) \right\}
\]

\[
= \prod_{X_i \leq t, \delta_i = 1} \frac{n^{-1} \sum I_{(X_i \geq t)} - n^{-1}}{n^{-1} \sum I_{(X_i \geq t)}}
\]

\[
= \prod_{X_i \leq t, \delta_i = 1} \left( 1 - \frac{1}{\sum I_{(X_i \geq t)}} \right)
\]

\[
= \prod_{X_i \leq t, \delta_i = 1} \left( 1 - \frac{D_i}{N_i} \right).
\]

Eqn (5.28) can be rewritten using counting process notation. Let

\[
N_i(t) = I_{(X_i \leq t, \delta_i = 1)} \quad Y_i(t) = I_{(X_i \geq t)}
\]

\[
\bar{N}(t) = \sum_{i=1}^{n} N_i(t) \quad \bar{Y}(t) = \sum_{i=1}^{n} Y_i(t).
\]

(\bar{N}(t) gives the number of deaths accumulated by time \( t \) and \( \bar{Y}(t) \) gives the number of subjects in the risk set at time \( t \).) Then the counting process form of the copulographic estimator under an assumed Archimedean copula is

\[
\hat{S}(t) = \phi^{-1} \left[ \int_{\bar{Y}(u) > 0} \left\{ \phi \left( \frac{\bar{Y}(u) - 1}{n} \right) - \phi \left( \frac{\bar{Y}(u)}{n} \right) \right\} d\bar{N}(u) \right].
\]

Rivest and Wells (2001, p. 143) note that since \( \phi((\bar{Y}(u) - 1)/n) - \phi(\bar{Y}(u)/n) \approx -\phi'(\bar{Y}(u)/n)/n \),

\[
\hat{S}(t) \approx \phi^{-1} \left[ \frac{1}{n} \int_{\bar{Y}(u) > 0} \phi' \left( \frac{\bar{Y}(u)}{n} \right) d\bar{N}(u) \right].
\]
The right hand side of Eqn (5.37) is equivalent to the differential equation-based estimator of Zheng and Klein. (This estimator was briefly discussed in the simulation study of Zheng and Klein that we mentioned earlier. However, we were unable to find a published paper detailing its derivation.) Therefore, the copula-graphic estimator under an assumed Archimedean copula and the differential equation-based estimator have the same asymptotic behavior.

Rivest and Wells (2001) also show uniform consistency and asymptotic normality of the copula-graphic estimator when the assumed copula is a member of the Archimedean subclass.

5.5 Remarks on the copula approach

Using copulas to model the dependence structure between random variables avoids many of the problems associated with the practice of modeling joint distributions using marginals and a correlation coefficient. Embrechts et al. (2000) identify two common misconceptions regarding the modeling of dependence.

**Fallacy 1.** “Marginal distributions and correlation determine the joint distribution” (Embrechts et al., 2000, p. 73).

**Fallacy 2.** “Given marginal distributions $F_1$ and $F_2$ for $X_1$ and $X_2$, all linear correlations between -1 and 1 can be attained through suitable specification of the joint distribution” (Embrechts et al., 2000, p. 73).

In order to disprove Fallacy 1, Embrechts et al. (2000, p. 73) simulate two sets of bivariate data with identical marginal distributions and correlations, but differing tail dependence. When plotted, it is obvious that these two datasets do not share
the same joint distribution. However, specification of \( \{k(t), p_1(t), p_2(t), t > 0\} \) and a copula guarantees uniqueness of the marginals (Zheng and Klein, 1995, p. 129).

It is important to remember that the Pearson correlation measure captures linear dependence. In many real world applications, notably in the area of finance, measuring the linear dependence present in the data is not appropriate. In cases such as these, capturing the dependence in the tails of the data is critical. Copulas provide an excellent tool for modeling tail dependence.

With regard to Fallacy 2, Embrechts et al. (2000, p. 74) introduce a scenario in which \( X_1 \) is assumed to be lognormally distributed with mean 0 and standard deviation 1 and \( X_2 \) is assumed to be lognormally distributed with mean 0 and standard deviation 2. Their correlation is said to be 0.7. This must be false; no joint distribution for the prescribed marginals and correlation exists. The attainable interval of correlation coefficients for variables with the specified joint and marginal distributions is \([-0.090, 0.666]\). The copula-based approach avoids this problem. There will always exist a set of marginal distributions that are consistent with the assumed copula and any possible \( \{k(t), p_1(t), p_2(t), t > 0\} \) (Zheng and Klein, 1995, p. 131).
Chapter 6

The Extended Copula-Graphic Estimator

The methods outlined in the previous chapter are applicable to data with a bivariate competing risks structure. In this chapter, we extend the copula-graphic estimator to data with a slightly more complicated structure.

6.1 Trivariate partial competing risks data structure

The dependent competing risks framework outlined by Zheng and Klein (1994, 1995, 1996) accommodates two event types. We desire to estimate marginal survival functions for data with a slightly more complicated structure. Our intent is to examine the decision of firms to initiate dividends. Although our data is financial in nature, it has a "time-to-event" flavor and censoring is present. Therefore, it is well suited to survival analysis techniques.

Let $X$, our lifetime of interest, correspond to the lifetime of a firm as a non-dividend-payer. Let $Y$ represent a firm’s lifetime until bankruptcy. We have strong
reason to believe that $X$ and $Y$ are not independent. If a non-dividend-paying firm enters bankruptcy, it is unlikely that the firm was considering initiating a dividend in its near future. Dividends are paid to shareholders out of a firm's profits. If a firm reaches insolvency, then it presumably has negative profitability and poor prospects for the long run.

Bankruptcy is not the only source of censoring in our data. We have firms that leave the sample for other reasons, most notably mergers. Recall that our event of interest is dividend initiation. Therefore, we need only consider mergers as censorings if they involve non-dividend-paying firms. Since large, profitable, established firms tend to be dividend-payers, the majority of mergers that we consider censorings involve small, immature firms valued by investors for their growth prospects and future profitability.

Generally, firms in our sample are regarded as merger targets for two reasons. First, they may be smaller, well-performing firms that are attractive to larger firms due to their success. Secondly, they may be smaller, under-performing firms that are attractive to larger firms due to their unreached potential. Therefore, we regard mergers as an ambiguous signal for dividend initiation and firm sustainability. In this manner, we have identified a second source of censoring in our data that is independent of dividend initiation and bankruptcy.

Another important source of independent censoring in our data is the administrative censoring brought about by the end of the study. Firms remaining in the sample
that were yet to initiate a dividend by 31 December 2002 are marked as censored.

We shall extend the copula-graphic estimator to accommodate independent censoring. Our data structure accounts for three lifetime variables, $X$, $Y$, and $Z$. As before, the events corresponding to the $X$ and $Y$ lifetimes (dividend initiation and bankruptcy, respectively) exhibit dependence. $Z$ represents the lifetime to an event which is independent of (i.e. non-informative for) dividend initiation and bankruptcy.

Formally, our observable data is of the form

$$T_i = \min(X_i, Y_i, Z_i)$$

(6.1)

$$\delta_i = \begin{cases} 
\delta_{i,X} = I_{\{T_i = X_i\}} \\
\delta_{i,Y} = I_{\{T_i = Y_i\}} \\
\delta_{i,Z} = I_{\{T_i = Z_i\}} 
\end{cases}$$

(6.2)

for $i = 1, \ldots, n$. We shall call this trivariate partial competing risks (TPCR) data.

6.2 Recovery of quantities from TPCR data

As outlined in the previous chapter, the copula-graphic estimator hinges upon the estimation of two quantities from the observable bivariate competing risks data: $k(t)$ and $p_1(t)$. We extend the copula-graphic estimator to data of the TPCR form by showing that these quantities are still recoverable from the observable data.

6.2.1 Estimation of $k(t)$

The introduction of $Z$ does not present much of an obstacle to the estimation of $k(t) = \Pr(X > t, Y > t)$. In this case, we define $W = \min(X,Y)$. Since $Z$ is
independent of $W$, we can estimate $Pr(W > t) = Pr(X > t, Y > t)$ by using the Kaplan-Meier estimator for $W$ in the presence of independent censoring from $Z$.

### 6.2.2 Estimation of $p_1(t)$

In order to recover estimates of $p_1(t) = Pr(X \leq t, X < Y)$ in the presence of independent censoring from $Z$, we use the inverse-probability-of-censoring-weighting (IPCW) approach. This method was introduced by Robins and Rotnitzky (1992). Satten and Datta (2001) provide a nice explanation of the Kaplan-Meier estimator in the IPCW framework. We present an introduction to the IPCW approach using the notation of Satten and Datta with the observable data as defined in Eqn (5.1).

Denote the ordered failure or censoring times by $\tau_j, j = 1, \ldots, J$. Let $n_j$ be the number of observations that fail at time $\tau_j$ and let $m_j$ be the number of observations that are censored at time $\tau_j$. Then the risk set at time $t$, denoted $R(t)$, can be expressed as

$$R(t) = \sum_{j=1}^{J} (n_j + m_j) I_{(\tau_j \geq t)}$$  \hspace{1cm} (6.3)

and the Kaplan-Meier estimator $\hat{S}_{KM}(t)$ for the marginal survival of $X$ is defined as

$$\hat{S}_{KM}(t) = \prod_{j : \tau_j \leq t} \left( 1 - \frac{n_j}{R(\tau_j)} \right).$$  \hspace{1cm} (6.4)

Likewise, the Kaplan-Meier estimator of the censoring distribution, $K(t) = Pr(C > t)$, can also be calculated using the above definitions. In this case,

$$\hat{K}(t) = \prod_{j : \tau_j \leq t} \left( 1 - \frac{m_j}{R(\tau_j)} \right).$$  \hspace{1cm} (6.5)
Note that \( \hat{K}(t) \) estimates the probability that a censored observation survives past time \( t \).

In the absence of censoring, the marginal distribution of failure for \( X \) could be estimated by the empirical CDF

\[
\hat{F}(t) = \frac{1}{n} \sum_{i=1}^{n} I(t_i \leq t).
\]  

(6.6)

When censoring is present, the estimator defined in Eqn (6.6) can be adjusted by reweighting each term (generated by an observed death at time \( t_i \)) by the inverse of the probability that a censored observation survives to or past time \( t_i \). This probability corresponds to \( \Pr(C \geq t_i) = \Pr(C > t_i-) \), and is estimated by \( \hat{K}(t_i-) \). Thus, the IPCW estimator for the failure distribution of \( X \) is given by

\[
\hat{F}_{rr}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(t_i \leq t) \delta_i}{\hat{K}(t_i-)}.
\]  

(6.7)

The survival function generated by the IPCW approach is found by taking \( \hat{S}_{rr}(t) = 1 - \hat{F}_{rr}(t) \).

Intuitively, the estimator defined in Eqn (6.7) adjusts the cumulative number of observed failures at each event time to account for the censored observations present in the data. For example, suppose that the probability that a censored observation survives to or past time \( t_k \) is 1/3 and the number of failures observed by time \( t_k \) is 5. Then the number of "effective" failures observed by time \( t_k \) is set equal to 15 to account for:

- the observed failures. (This gives us 5 of the 15.)
• the 2/3 probability that a censored observation did not survive past time $t_k$.

(This gives us 10 of the 15. $10/15 = 2/3$.)

In other words, each failure observed prior to or at time $t_k$ is now taken to correspond to three failures.

Satten and Datta (2001) show that the IPCW estimator $\hat{F}^*_\text{rr}(t)$ and the standard Kaplan-Meier estimator $\hat{F}_KM(t) = 1 - \hat{S}_KM(t)$ are equivalent under the convention that ties between failures and censorings are broken by assuming that the failures occur instantaneously before the censorings.

With observable data of the form given in Eqn (6.1) and Eqn (6.2), the IPCW approach can be used to obtain an estimator for $p_1(t)$. In this case, $p_1(t) = \Pr(X \leq t, X < Y)$ can be estimated by

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{\hat{K}_Z(t_i^-)} \times I(t_i \leq t \delta_{i,X=1})$$

(6.8)

where $\hat{K}_Z(\cdot)$ is the Kaplan-Meier estimator for $Z$.

### 6.3 Implementation

Equipped with estimates of $k(t)$ and $p_1(t)$ for TPCR data, we now implement the extended copula-graphic estimator (ECGE). We use S-Plus (InsightfulCorp., 1988, 2002) to program the estimator in a series of functions. Descriptions of the functions are given in the following subsections. The actual code is given in Appendix B. In this section, we shall refer to our events as failures and censorings. $X$ and $Y$ events represent failures, while $Z$ events represent censorings.
We must first address the main issue to be dealt with when applying the estimator to discrete data—tied lifetimes. We can define ties to be of two types, based on whether or not they involve the censoring event. For example, if $T_i = T_j$ and $\delta_{i,Y} = \delta_{j,Z} = 1$, we have a tie involving the censoring event. In this case, observation $i$ and observation $j$ both have the same observed lifetime, and the lifetime of $i$ corresponds to a $Y$ event and the lifetime of $j$ corresponds to a censoring. If $T_i = T_j$ and $\delta_{i,X} = \delta_{j,Y} = 1$, then we have a tie between the failures. Here, observation $i$ and observation $j$ both have the same observed lifetime, and the lifetime of $i$ corresponds to an $X$ event and the lifetime of $j$ corresponds to a $Y$ event.

Ties that involve the censoring lifetime must be broken to maintain a consistent ordering between failures and censorings. Recall that when constructing the Kaplan-Meier estimator, the standard convention used for ties is to assume that the failures occur instantaneously before the censorings.

As outlined in the previous section, the IPCW approach requires us to calculate the survival function for the censorings. If we do this by simply recoding the censorings as failures and the failures as censorings and applying the Kaplan-Meier estimator to the recoded data, we will fail to maintain a consistent ordering between failures and censorings throughout the estimation process. Therefore, we add a slight amount, 0.01 years, to each $Z$ lifetime. This ensures that failures are consistently treated as occurring just before censorings.

Ties involving $X$ and $Y$ must also be broken so we can make use of the simplified
grid on which to compute our estimator. Recall that Zheng and Klein (1995) first define their estimator as the solution to

\[ \mu_c(\hat{A}_t) - \hat{k}(t_i) = 0 \]  

(6.9) and

\[ \mu_c(\hat{B}_t) - \hat{p}_1(t_i) = 0. \]  

(6.10)

They provide an outline of a bisection root-finding algorithm for constructing the estimator. Although the algorithm is simple to follow, the quantities involved in Eqn (6.10) are not easy to compute. \( \hat{B}_t \) has no closed form solution and would have to be evaluated numerically.

If the grid is chosen to be the unique failure times, the estimator is greatly simplified. Since we desire \( \hat{F} \) to jump only at the times of \( X \) failures (and likewise for \( \hat{G} \) and \( Y \) failures), we can construct the estimator based only on Eqn (6.9).

However, in the data we wish to analyze, we observe both an \( X \) failure and a \( Y \) failure at many of the event times in the dataset. Nearly 31% of the sample experiences a failure at a non-unique failure time. Therefore, there is concern that we may lack a sufficient number of unique failure times with which to construct a grid.

In order to maintain the ease of calculating the estimator, we shall break ties between failure times. This will be done by subtracting a small amount of time, 0.01 years, from either the \( X \) or \( Y \) lifetimes in the dataset. If the amount is subtracted from the \( X \) lifetimes, we are putting dividend initiations before tied bankruptcies. If the amount is subtracted from the \( Y \) lifetimes, we are putting bankruptcies before
tied dividend initiations. Tie breaking will be further discussed in Section 8.2.

6.3.1 Function 1: g.km.xy

This function takes in a vector of lifetimes, a vector of event types (coded as 1, 2, or 3), and a bound. Event types 1 and 2 correspond to failures (i.e. $X$ and $Y$) and event type 3 corresponds to censoring (i.e. $Z$). Note that this vector of event types provides the same information as $\delta_i$ in Eqn (6.2). The function then finds the right-continuous Kaplan-Meier estimator of $W = \min(X, Y)$ up to the specified bound. It outputs a plot of the Kaplan-Meier estimator of $W$ and a five column matrix

\[
\text{left.int} | \text{right.int} | \text{n.risk} | \text{n.failure} | \text{km.xy}
\]

where

- \text{left.int} corresponds to the left endpoint of the step function.
- \text{right.int} corresponds to the right endpoint of the step function.
- \text{n.risk} gives the number of observations in the risk set at the time given by \text{left.int}.
- \text{n.failure} gives the number of failures at the time given by \text{left.int}.
- \text{km.xy} gives the Kaplan-Meier estimator for $W = \min(X, Y)$.

Sample output from \text{g.km.xy} is given below and in Figure 6.1.

\begin{verbatim}
> t.test.vec <- c(1,1,2,3,1,3, 2,2,3,1,3,2, 2,3,2,3,2,2)
> t.time.vec <- c(3,3,2,9,4,3,12,9,1,5,3,9,8.5,17,7,15,3,4)
\end{verbatim}
The last column of this output gives our estimator of \( k(t) \).

![Kaplan-Meier estimator for \( W = \min(X,Y) \) in the presence of \( Z \)](image)

\[ \Pr(W > t) = \Pr(X > t, Y > t) \]

Figure 6.1: Sample Output from \( g.km.xy \).
6.3.2 Function 2: g.ipcw

As in the previous function, g.ipcw takes in a vector of lifetimes, a vector of event types (coded as 1, 2, or 3), and a bound. The function first finds the Kaplan-Meier for the censoring variable up to the specified bound. Here, as previously discussed, we must maintain consistency in the ordering of tied failures and censorings. Therefore, we add a small amount to each censoring lifetime when calculating the Kaplan-Meier estimator of $Z$. Next, the function finds the IPCW estimator of $p_1(t)$ up to the specified bound. To provide a quick check that the quantities are being correctly derived, the function also calculates the IPCW estimator of $p_2(t)$.

The function outputs three matrices and plots of the Kaplan-Meier estimator of the censorings and $\hat{p}_1(t)$. The first matrix corresponds to the Kaplan-Meier estimator of the censorings and is of the same form as the output of g.km.xy. The second matrix gives the estimator of $p_1(t)$, and the third matrix gives the estimator of $p_2(t)$.

Sample output from g.ipcw is given below and in Figure 6.2.

```r
> g.ipcw(t.time.vec, t.event.vec, 12)
$rat.mat:
  left.int right.int n.risk n.failure km.z
[1,] 0.00   1.01   18       0 1.0000000
[2,] 1.01   3.01   18  12 0.94444444
[3,] 3.01   9.01   18  13 0.7991453
[4,] 9.01  12.00   10  12 0.5993590

$ret.mat2:
  left.int.x right.int.x n.failure.x weight.z pi.hat
[1,]  0     3               18 0.0000000 0.0000000
[2,]  3     4               18 0.9444444 0.1176471
[3,]  4     5               12 0.7991453 0.1871658
[4,]  5    12               12 0.7991453 0.2566845
```
Function 3: \texttt{g.emp.prob.maker}

This function takes in vectors of lifetimes and event types. First, it calculates an upper bound, $R$, equal to the maximum of the $X$ and $Y$ event times. Our estimators
will be defined on the interval \([0, R]\). Next, it breaks the ties between \(X\) events and \(Y\) events, thereby enabling us to use the simpler algorithm for calculating the estimator. The function then calls \(g.\text{km.xy}\) and \(g.\text{ipcw}\) and returns estimates of \(k(t)\), \(p_1(t)\), and \(p_2(t)\) on the grid defined by the unique failure times. (The points of this grid are referred to in the output as \text{event.times.xy}.) It also adds these three estimates to ensure that they sum to one.

Sample output from \(g.\text{emp.prob.maker}\) is given below. For the sample code, we have broken ties by subtracting 0.01 from all \(Y\) event times.

\[
> g.\text{emp.prob.maker}(t.\text{time.vec}, t.\text{event.vec})
\]

\[
[1] "The bound is" "11.99"
\]

<table>
<thead>
<tr>
<th>event.times.xy</th>
<th>k.hat</th>
<th>p1.hat</th>
<th>p2.hat</th>
<th>sum.check</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.99</td>
<td>0.9411765</td>
<td>0.0000000</td>
<td>0.05882353</td>
<td>1</td>
</tr>
<tr>
<td>2.99</td>
<td>0.8823529</td>
<td>0.0000000</td>
<td>0.11764706</td>
<td>1</td>
</tr>
<tr>
<td>3.00</td>
<td>0.7647059</td>
<td>0.1176471</td>
<td>0.11764706</td>
<td>1</td>
</tr>
<tr>
<td>3.99</td>
<td>0.6951872</td>
<td>0.1176471</td>
<td>0.18716578</td>
<td>1</td>
</tr>
<tr>
<td>4.00</td>
<td>0.6256684</td>
<td>0.1871658</td>
<td>0.18716578</td>
<td>1</td>
</tr>
<tr>
<td>5.00</td>
<td>0.5561497</td>
<td>0.2566845</td>
<td>0.25668449</td>
<td>1</td>
</tr>
<tr>
<td>6.00</td>
<td>0.4866310</td>
<td>0.2566845</td>
<td>0.32620321</td>
<td>1</td>
</tr>
<tr>
<td>8.49</td>
<td>0.2780749</td>
<td>0.2566845</td>
<td>0.46524064</td>
<td>1</td>
</tr>
<tr>
<td>8.99</td>
<td>0.1853832</td>
<td>0.2566845</td>
<td>0.55793226</td>
<td>1</td>
</tr>
</tbody>
</table>

6.3.4 Function 4: \(g.\text{cg.est}\)

This is the function that produces the extended copula-graphic estimator of the marginal survival functions of \(X\) and \(Y\). It takes in an object created by \(g.\text{emp.prob.maker}\) and a copula specification (family and parameter). The available copula families are:
• Independence (independence):

\[ C_I(x, y) = x \times y \]  \hspace{1cm} (6.11)

When using the independence copula, which involves no parameter, a dummy parameter must be given to the program.

• Gamma Frailty (gamma), \( \alpha > 0 \):

\[ C_\alpha(x, y) = x + y - 1 + \left\{ \left( \frac{1}{1 - x} \right)^{1/\alpha} + \left( \frac{1}{1 - y} \right)^{1/\alpha} - 1 \right\}^{-\alpha} \]  \hspace{1cm} (6.12)

See Oakes (1989, 1982) for details on the Gamma Frailty copula.

• Frank (frank), \( \alpha > 0, \alpha \neq 1 \):

\[ C_\alpha(x, y) = \log_\alpha \left\{ 1 + \frac{(\alpha x - 1)(\alpha y - 1)}{\alpha - 1} \right\} \]  \hspace{1cm} (6.13)

See Genest (1987) for details on the Frank family of copulas.

• Gumbel-Hougaard (gumbel.hougaard), \( \alpha \geq 1 \):

\[ C_\alpha(x, y) = \exp \left[ - \{(- \log x)^\alpha + (- \log y)^\alpha\}^{1/\alpha} \right] \]  \hspace{1cm} (6.14)

See Hutchinson and Lai (1990) for details on the Gumbel-Hougaard copula.

The function calls two helper functions `copula.eqns.for.x` and `copula.eqns.for.y` which solve the following equations for the specified copula:

for \( \delta_i = 1 \):

\[ \hat{k}(t_i) = 1 - \hat{F}(t_i) - \hat{G}(t_{i-1}) + C\{\hat{F}(t_i), \hat{G}(t_{i-1})\}. \]  \hspace{1cm} (6.15)
For some of the available copulas, namely the Frank and Gumbel-Hougaard copulas, the right hand sides of Eqn (6.15) and Eqn (6.16) cannot easily be manipulated algebraically to solve for $\hat{F}(t_i)$ and $\hat{G}(t_i)$. In this case, a second set of helper functions, `nm.copula.for.x` and `nm.copula.for.y` recover $\hat{F}(t_i)$ and $\hat{G}(t_i)$ from Eqn (6.15) and Eqn (6.16) using Newton's method.

Sample output of `g.cg.est` using a Gamma Frailty copula with parameter 1.5 is given below and in Figure 6.3.

```r
> cg.test <- g.emp.prob.maker(t.time.vec, t.event.vec)
[1] "The bound is" "11.99"
> g.cg.est(cg.test, "gamma", 1.5)
        xy.points F.est G.est F.bar.est G.bar.est
[1,] 0.00 0.0000000 0.00000000 1.0000000 1.0000000
[2,] 1.99 0.0000000 0.05882353 1.0000000 0.9411765
[3,] 2.99 0.0000000 0.11764706 1.0000000 0.8823529
[4,] 3.00 0.1410115 0.11764706 0.8589885 0.8823529
[5,] 3.99 0.1410115 0.20337005 0.8589885 0.7966300
[6,] 4.00 0.2356886 0.20337005 0.7643114 0.7966300
[7,] 5.00 0.3287183 0.20337005 0.6712817 0.7966300
[8,] 6.99 0.3287183 0.32033887 0.6712817 0.6796611
[9,] 8.49 0.3287183 0.43283449 0.6712817 0.5671655
[10,] 8.99 0.3287183 0.64392412 0.6712817 0.3560759
[11,] 11.99 0.3287183 0.77367841 0.6712817 0.2263216
```
Estimated Survival Functions for X and Y

Estimated Survival for X

Estimated Survival for Y

Copula Type: gamma, with parameter = 1.5

Figure 6.3: Sample Output from g.cg.est.
Chapter 7

Simulation Study of the ECGE

In this chapter, we explore the implementation of the extended copula-graphic estimator (ECGE) on simulated data. The first section provides background information on two commonly used measures of association, Spearman’s rho and Kendall’s tau. In the second section, we discuss our efforts in developing an S-Plus function to simulate copula data. The third section details three simulation studies.

7.1 Measures of association

The key component to estimation of the marginal survival functions is not the parametric form of the assumed copula, but rather an understanding of the degree of association between $X$ and $Y$ (Zheng and Klein, 1995). Two common non-parametric measures of association are Spearman’s rho and Kendall’s tau.

7.1.1 Spearman’s rho

Spearman’s rho, $\rho_s$, is a simple rank-based variant of the standard Pearson product-moment correlation. To calculate Spearman’s rho, the data vectors $X$ and $Y$ are
converted into their rank vectors, \( R_x \) and \( R_y \). Then the Pearson correlation statistic is calculated using the rank vectors. However, Spearman's rho is commonly estimated by applying the Spearman rank correlation formula

\[
rs = 1 - \frac{6 \sum d_i^2}{n^3 - n} \tag{7.1}
\]

where \( d_i = R_{x_i} - R_{y_i} \). It can be shown that the Pearson product-moment correlation formula applied to the rank vectors simplifies to Eqn (7.1) in the absence of ties (Fisher and van Belle, 1993, p. 386).

A formal mathematical definition of Spearman's rho emphasizes the link to copulas:

**Definition 7.1.** (Joe, 1991, p.32) “Let \( F \) be a continuous bivariate cdf with univariate margins \( F_1, F_2 \) and let \( (X_1, X_2) \sim F \); then Spearman's rho is the correlation of \( F_1(X_1) \) and \( F_2(X_2) \). Since \( F_1(X_1) \) and \( F_2(X_2) \) are \( U(0,1) \) rvs (under the assumption of continuity), their expectations are 1/2, their variances are 1/12, and Spearman's rho is

\[
\rho_S = 12 \int \int F_1(x_1)F_2(x_2)dF(x_1,x_2) - 3. \tag{7.2}
\]

Since \( \rho_S \) is invariant under monotone transformations, Spearman's rho can also be given in terms of the copula of \( X \) and \( Y \):

\[
\rho_S = 12 \int \int uv \ dC(u,v) - 3. \tag{7.3}
\]

### 7.1.2 Kendall's tau

Kendall's tau, \( \tau \), is a measure of association based on the difference in the probabilities of concordant and discordant pairs. The following definition links copulas to Kendall's tau:

\[
\rho_S = 12 \int \int uv \ dC(u,v) - 3. \tag{7.3}
\]
Definition 7.2. (Joe, 1997, p.32). "Let $F$ be a continuous bivariate cdf and let $(X_1, X_2), (X'_1, X'_2)$ be independent pairs with distribution $F$. Then Kendall's tau is

$$
\tau = \Pr[(X_1 - X'_1)(X_2 - X'_2) > 0] - \Pr[(X_1 - X'_1)(X_2 - X'_2) < 0]
= 2 \Pr[(X_1 - X'_1)(X_2 - X'_2) > 0] - 1
= 4 \int F dF - 1.
$$

(7.4)

Since $\tau$ is invariant under monotone transformations, Kendall’s tau can also be defined in terms of the copula:

$$
\tau = 4 \int \int C \, dC - 1.
$$

(7.5)

To estimate Kendall’s tau in practice, consider the $n(n - 1)/2$ pairings $(X_i, Y_i)$ and $(X_j, Y_j), j \neq i$. Count 1 if $(X_i - X_j)(Y_i - Y_j) > 0$ (concordant pair), count $-1$ if $(X_i - X_j)(Y_i - Y_j) < 0$ (discordant pair), and count 0 if $(X_i - X_j)(Y_i - Y_j) = 0$. Let $\kappa$ be the sum of these counts. Kendall’s tau is calculated as

$$
\tau = \frac{\kappa}{n(n - 1)/2}.
$$

(7.6)

For many copulas, Spearman’s rho and Kendall’s tau can be expressed as a simple function of the copula parameter $\alpha$. The following table gives expressions for Spearman’s rho and Kendall’s tau for the copulas we have implemented for the ECGE.
Table 7.1: Spearman's Rho and Kendall's Tau for Implemented Copulas

<table>
<thead>
<tr>
<th>Parametric Copula</th>
<th>Spearman's Rho</th>
<th>Kendall's Tau</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Gamma Frailty</td>
<td>$1/(1 + 2\alpha)$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Frank</td>
<td>$1 + 12\gamma^{-1}{D_2(\gamma) - D_1(\gamma)}$</td>
<td>$1 + 4\gamma^{-1}{D_1(\gamma) - 1}$</td>
<td>2, 3</td>
</tr>
<tr>
<td>Gumbel-Hougaard</td>
<td>$(\alpha - 1)/\alpha$</td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

Notes for Table 7.1

1. Oakes (1982), Hutchinson and Lai (1990), and Cook and Johnson (1981) have investigated the expression of Spearman’s rho for the Gamma Frailty copula. However, no simple closed form has been found.

2. $\gamma = -\log \alpha$.

3. $D_k$, the Debye function, is defined as

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} \, dt. \quad (7.7)$$

4. We have found no references regarding the expression of Spearman’s rho for the Gumbel-Hougaard copula.

7.2 Simulating copula data

In order to run simulation studies of our copula-based estimator, we must first develop an *S-Plus* function to simulate copula data. We have implemented four copulas for use with the ECGE: the Gumbel-Hougaard copula, the Gamma Frailty copula, the Frank copula, and, by default, the independence copula. The simulation algorithms are described in the following subsections and the code is provided in Appendix C.
7.2.1 Simulating from the Gumbel-Hougaard copula

\[ C_\alpha(x, y) = \exp \left[ - \left\{ (-\log x)^\alpha + (-\log y)^\alpha \right\}^{1/\alpha} \right], \quad \alpha \geq 1 \quad (7.8) \]


Marshall and Olkin (1988) discuss the generation of multivariate distribution functions with marginals as parameters. They focus on a technique using mixture models. In an example, they show that the Gumbel-Hougaard copula can be obtained using this mixture method.

They begin their discussion with the following motivation. Let \( F \) and \( G \) be univariate distribution functions. Let \( \bar{G} \) denote the survival function corresponding to the distribution function \( G \) and suppose that \( \bar{G}(0) = 1 \). Then, since \( F^\theta \) is also a distribution function for all \( \theta > 0 \), the mixture

\[ H(x) = \int F^\theta(x) \, dG(\theta) \quad (7.9) \]

is a distribution function. Marshall and Olkin (1988, p. 839) state that their "results are possible only because this kind of mixture has the following property: For any specified pair of distribution functions \( G \) and \( H \) such that \( \bar{G}(0) = 1 \), there exists a distribution function \( F \) for which (7.9) holds. To see this, observe that \( H(x) = \phi[\log F(x)] \), where \( \phi \) is the Laplace transform of \( G \)."

Next, they extend their argument to the bivariate dimension. In this case, \( G \) is a bivariate distribution function such that \( \bar{G}(0, 0) = 1 \) with marginals \( G_1 \) and \( G_2 \). For
the mixture defined by

\[ H(x_1, x_2) = \int \int F_{\theta_1}^\theta_1(x_1) F_{\theta_2}^\theta_2(x_2) \, dG(\theta_1, \theta_2) \]  
(7.10)

the marginal distributions of \( H \) are of the form given in Eqn (7.9):

\[ H_1(x) = \int F_{\theta_1}^\theta_1(x) \, dG_1(\theta_1) \]
\[ H_2(x) = \int F_{\theta_2}^\theta_2(x) \, dG_2(\theta_2). \]

Extending from the univariate case, it follows that \( H \) is a bivariate distribution function with marginals \( H_1 \) and \( H_2 \) if \( F_i(x) = \exp\left\{-\phi_i^{-1}(H_i(x))\right\} \), where \( \phi_i \) denotes the Laplace transform of \( C_i \), for \( i = 1, 2 \).

Marshall and Olkin (1988, p. 834) give the following theorem:

**Theorem 7.1.** "Let \( H_1, \ldots, H_n \) be univariate distribution functions, and let \( G \) be an \( n \)-variate distribution function such that \( G(0, \ldots, 0) = 1 \), with univariate marginals \( G_i \) (\( i = 1, \ldots, n \)). Denote the Laplace transform of \( G \) and \( G_i \), respectively, by \( \phi \) and \( \phi_i \) (\( i = 1, \ldots, n \)). Let \( K \) be an \( n \)-variate distribution function with all univariate marginals uniform on \([0, 1] \). If \( F_i(x) = \exp\left[-\phi_i^{-1}H_i(x)\right] \) (\( i = 1, \ldots, n \)), then

\[ H(x_1, \ldots, x_n) = \int \cdots \int K(F_{\theta_1}^\theta_1(x_1), \ldots, F_{\theta_n}^\theta_n(x_n)) \, dG(\theta_1, \ldots, \theta_n) \]  
(7.11)

is an \( n \)-variate distribution function with marginals \( H_1, \ldots, H_n \).

They show that the Gumbel-Hougaard copula results from the above theorem when \( K \) is the case of independence, the univariate marginal distributions of \( G \) are all equal and \( G \) is the upper Frechet bound, and \( \phi(s) = \exp(-s^{1/\alpha}). \)

Based on the constructions given in Eqn (7.10) and Eqn (7.11), Marshall and Olkin (1988, p. 840) give the following algorithm for simulating from the Gumbel-Hougaard copula.
• Step 1: Generate an observation, $\gamma$, from $G$. This is done by noting that $\phi(s) = \exp(-s^{1/\alpha})$ corresponds to the Laplace transform of a positive stable random variable. A method of simulating positive stable random variables is given in Chambers et al. (1976).

• Step 2: Generate an observation $(y_1, y_2)$ from $K$. For the Gumbel-Hougaard copula this is simple: $K$ has independent marginals that are uniform on $[0, 1]$.

• Step 3: Let $x_i = F_i^{-1}(y_i^{1/\gamma})$, where $F_i(u) = \exp\{-\phi^{-1}(H_i(u))\}$.

This results in a simulated pair $(x_1, x_2)$ from the Gumbel-Hougaard copula with parameter $\alpha$.

For Step 3, we note that $H_i(u) = u$ for $i = 1, 2$. (This follows since $H$ is a copula and has uniform marginals on $[0, 1]$.) Therefore, $F_i(u) = \exp\{-\phi^{-1}(u)\}$. Carrying out the algebra, we have

$$F_i(x_i) = y_i^{1/\gamma}$$  \hspace{1cm} (7.12)

$$\exp\{-\phi^{-1}(x_i)\} = y_i^{1/\gamma}$$  \hspace{1cm} (7.13)

$$\phi^{-1}(x_i) = -\log(y_i^{1/\gamma}).$$  \hspace{1cm} (7.14)

Now, substituting in for $\phi$, we can solve for $x_i$:

$$x_i = \phi(-\log(y_i^{1/\gamma}))$$  \hspace{1cm} (7.15)

$$= \exp\{-(-\log(y_i^{1/\gamma}))^{1/\alpha}\}$$  \hspace{1cm} (7.16)

$$= \exp\{-(-\gamma^{-1}\log(y_i))^{1/\alpha}\}.$$  \hspace{1cm} (7.17)
For the Gumbel-Hougaard copula, Kendall’s tau is equal to \((\alpha - 1)/\alpha\). Since \(\alpha \geq 1\), this indicates that the Gumbel-Hougaard copula is only able to capture positive association.

7.2.2 Simulating from the Gamma Frailty copula

\[
C_\alpha(x, y) = x + y - 1 + \left\{ \left( \frac{1}{1 - x} \right)^{1/\alpha} + \left( \frac{1}{1 - y} \right)^{1/\alpha} - 1 \right\}^{-\alpha}, \quad \alpha > 0 \tag{7.18}
\]

As an exercise, we derived the Gamma Frailty copula from first principles. This undertaking made clear the origin of the copula’s name. This derivation is given in Appendix D.

Given the derivation, the simulation algorithm is simple.

1. Generate a gamma distributed random variable with shape parameter \(\alpha\) and rate parameter \(\alpha\), i.e. a gamma distributed random variable with mean 1.

2. Generate \(y_1\) and \(y_2\), independently distributed exponential random variables with mean equal to the inverse of the gamma random variable generated in Step 1.

3. Transform \(y_1\) and \(y_2\) into Uniform(0,1) random variables using the probability integral transform \(x_i = F(y_i)\). We know that \(y_1\) and \(y_2\) are marginally Pareto distributed, and under the current parametrization, \(F(t) = 1 - (1 + t/\alpha)^{-\alpha}\).

This algorithm yields the simulated pair \((x_1, x_2)\) from the Gamma Frailty copula with parameter \(\alpha\). For the Gamma Frailty copula, \(\tau = (1 + 2\alpha)^{-1}\). Again, since \(\alpha\) is
constrained to be positive, this copula can only capture positive association.

### 7.2.3 Simulating from the Frank copula

\[
C_\alpha(x, y) = \log_\alpha \left\{ 1 + \frac{(\alpha^x - 1)(\alpha^y - 1)}{\alpha - 1} \right\}, \quad \alpha > 0, \quad \alpha \neq 1 \tag{7.19}
\]

Observations from the Frank copula can be simulated using a straightforward method. Since the desired bivariate distribution is a copula, the marginal distributions are uniform. Therefore, for a copula pairing \((x_1, x_2)\), \(x_1\) can readily be generated as a Uniform(0,1) random variable. This suggests an obvious method for simulating \(x_2\) if the conditional distribution of \(x_2\) given \(x_1\) is known. For the Frank copula, this conditional distribution has a closed form and is given in Nelsen (1986).

To simulate from the Frank copula:

1. Generate two independent Uniform(0,1) random variables \(x_1\) and \(y\).

2. Making use of the conditional distribution of \(x_2\) given \(x_1\) and the probability integral transform, let

\[
x_2 = \log_\alpha \left\{ 1 + \frac{y(\alpha - 1)}{\alpha^x(1 - y) + y} \right\}. \tag{7.20}
\]

This yields the simulated pair \((x_1, x_2)\) from the Frank copula with parameter \(\alpha\).

For the Frank copula,

\[
\tau = 1 + \frac{4}{-\log \alpha} [D_1(-\log \alpha) - 1] \tag{7.21}
\]
where $D_1(x)$ is the Debye function defined by

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} \, dt. \quad (7.22)$$

Furthermore, since $D_k(-x) = D_k(x) + kx/(k + 1)$, $\tau(1/\alpha) = -\tau(\alpha)$ (Nelsen, 1986, p. 3280). Therefore, unlike the Gumbel-Hougaard and Gamma Frailty copulas, the Frank copula is able to capture negative association.

### 7.2.4 Simulating from the independence copula

$$C_I(x, y) = xy \quad (7.23)$$

Simulating from the independence copula is trivial, but is mentioned for the sake of completeness. To simulate a pair from the independence copula, generate two (independent) Uniform(0,1) random variables.

### 7.3 Simulations

We now discuss the performance of our estimator in three simulation studies. In the first simulation, we assess the performance of the ECGE and its variance under the assumption of the correct dependence structure. In the other simulations, we focus on the performance of the estimator in situations where the assumed dependence is not the true dependence that generated the data.
Simulation 1 Set up

<table>
<thead>
<tr>
<th>Size</th>
<th>2,000 trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100 simulated data points in each trial</td>
</tr>
<tr>
<td></td>
<td>300 bootstrap replications per trial</td>
</tr>
</tbody>
</table>

**Distributions**

- $X \sim \text{Weibull}(2.55, 5)$
- $Y \sim \text{Weibull}(2.35, 4.8)$
- $Z \sim \text{Uniform}(0, 20)$

**Dependence Structure**

- Truth: Gumbel-Hougaard with $\tau = 0.8$
- Assumed: Gumbel-Hougaard with $\tau = 0.8$

### 7.3.1 Simulation 1: Assessing performance and variance

Our simulation study is structured as follows. For each trial, we generate 100 pairs $(X_i, Y_i)$, where $X_i$ and $Y_i$ are Weibull distributed with parameters $(2.55, 5)$ and $(2.35, 4.8)$, respectively, and linked by a Gumbel-Hougaard copula with $\alpha = 5$ (corresponding to $\tau = 0.8$). Next, we generate 100 realizations of our $Z$ random variable, which is Uniform$(0, 20)$ distributed and independent of $X$ and $Y$. The “observed” data consists of two vectors, a vector $T_i = \min(X_i, Y_i, Z_i)$ indicating event times and a vector indicating event types. The ECGE is calculated using the observed data and an assumed Gumbel-Hougaard copula with $\alpha = 5$. Estimates of the survival functions for $X$ and $Y$ are recovered.

This process is repeated 2,000 times. Our final outputs are pointwise averages of the 2,000 estimated survival curves. By pointwise, we mean that the x-axis (time) is gridded, and the average values of the survival estimates are determined at these grid points. In our simulations, the grid is set over the interval from 0 to the 99.5$^{th}$
percentile of the underlying distribution in increments of 0.1. For this simulation, the 99.5\textsuperscript{th} percentile of the Weibull(2.55, 5) distribution is 9.61492, resulting in a grid of 97 points. The 99.5\textsuperscript{th} percentile of the Weibull(2.35, 4.8) distribution is 9.758545, resulting in a grid of 98 points.

Note that in this simulation, we assume the correct dependence structure when calculating the ECGE. In the simulations that follow, this will not necessarily be the case.

We do not have a closed form expression for the variance of the ECGE. Therefore, we construct a bootstrapped variance estimate. For each of the 2,000 trials, the observed data is bootstrapped (sampled with replacement) 300 times, and for each of these bootstrap replication datasets, the ECGE is calculated. Then the (pointwise) variance of the 300 bootstrapped survival curves is calculated. At the end of the simulation, we obtain the average bootstrapped variance from the 2,000 trials. This bootstrapped variance estimate is then compared to the empirical sample variance. The empirical sample variance is simply the (pointwise) variance of the 2,000 survival curve estimates derived from the original (i.e. non-bootstrapped) data.

Figure 7.1 shows the output of the ECGE (pink line) in the simulation described above. The true underlying distribution that generated the data is shown in black. For comparison, the Kaplan-Meier estimator is also shown in blue. (Recall that the Kaplan-Meier estimator assumes independence between \(X\) and \(Y\).) The curves for \(X\) are shown on the left, and the curves for \(Y\) are shown on the right.
Figure 7.1: Simulation 1: Performance of the ECGE under Assumption of the Correct Dependence Structure. Black line, true underlying distribution that generated the data; Pink line, ECGE; Blue line, K-M estimator. Dependence structure: Gumbel-Hougaard copula with $\tau = 0.8$. Estimates for $X$ on left and $Y$ on right. Average over 2000 trials.

The ECGE clearly outperforms the Kaplan-Meier estimator. Along much of the time interval, the ECGE nearly perfectly captures the true distribution, while the Kaplan-Meier estimator lies considerably above it. This makes intuitive sense; when positive dependence between $X$ and $Y$ is ignored, survival will be overstated. The ECGE is not quite as accurate in the tail of the survival distribution. This is not surprising—by the construction of survival data, it is often difficult to observe events
in the right tail. However, even in the tail of the distributions (roughly $T > 7$), the ECGE only slightly overestimates survival, and its performance is far superior to the Kaplan-Meier estimator.

![Figure 7.2: Simulation 1: Comparison of the Empirical Sample Variance of the ECGE to the Bootstrapped Variance Estimate. Black line, bootstrapped variance estimate; Pink line, empirical sample variance. Estimates for $X$ on left and $Y$ on right. Average over 2,000 trials with 300 bootstrap replications per trial.](image)

In Figure 7.2 we compare the empirical sample variance of the ECGE to the bootstrapped variance estimate. The bootstrapped variance estimate is very close to the empirical sample variance, although we see some divergence in the tails. Again,
we attribute this behavior to a lack of data in the tails.

### 7.3.2 Simulation 2: Investigating sensitivity to the choice of copula or implied strength of association

<table>
<thead>
<tr>
<th>Simulation 2 Set up</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Distributions</strong></td>
</tr>
<tr>
<td>$X \sim \text{Weibull}(2.5, 5)$</td>
</tr>
<tr>
<td>$Y \sim \text{Gamma}(4, 1)$</td>
</tr>
<tr>
<td>$Z \sim \text{Lognormal}(1.75, 0.25)$</td>
</tr>
<tr>
<td><strong>Dependence Structure</strong></td>
</tr>
<tr>
<td>Truth: Gamma Frailty with $\tau = 0.8$</td>
</tr>
<tr>
<td>Assumed: Gamma Frailty with $\tau = 0.8$</td>
</tr>
<tr>
<td>Gumbel-Hougaard with $\tau = 0.2$</td>
</tr>
<tr>
<td>Gumbel-Hougaard with $\tau = 0.5$</td>
</tr>
<tr>
<td>Gumbel-Hougaard with $\tau = 0.8$</td>
</tr>
</tbody>
</table>

In this simulation, $X$ is generated by the Weibull$(2.5, 5)$ distribution, and $Y$ is generated by the Gamma$(4, 1)$ distribution. They are linked by a Gamma Frailty copula with $\alpha = 0.125$, corresponding to $\tau = 0.8$. Again, we have strong, positive dependence between $X$ and $Y$. $Z$ is generated by the Lognormal$(1.75, 0.25)$ distribution. 1,000 data points are simulated for each trial.

In Figure 7.3 we show the output of the ECGE (pink line) when the correct dependence structure is assumed. For $X$, the fit is very good and clearly superior to the Kaplan-Meier estimator (blue line). For $Y$, the ECGE outperforms the Kaplan-Meier estimator, but it is obvious that there is a lack of data past $T = 8$.

In Figure 7.4 we show the results of the ECGE under the assumption of the incor-
Figure 7.3: *Simulation 2: Performance of the ECGE under Assumption of the Correct Dependence Structure.* Black line, true underlying distribution that generated the data; Pink line, ECGE; Blue line, K-M estimator. Dependence structure: Gamma Frailty copula with $\tau = 0.8$. Estimates for $X$ on left and $Y$ on right. Average over 2000 trials.

rect parametric copula. While the true dependence underlying the data is generated by a Gamma Frailty copula, we assume a Gumbel-Hougaard copula when calculating the ECGE. Furthermore, we assume a variety of dependence strengths:

- $\tau = 0$, corresponding to independence and the Kaplan-Meier estimator (blue line).
Figure 7.4: Simulation 2: Performance of the ECGE under Assumption of the Incorrect Copula and Varying Strengths of Dependence. Black line, true underlying distribution that generated the data; Blue line, K-M estimator; Green line, ECGE under assumption of G-H copula with $\tau = 0.2$; Orange line, ECGE under assumption of G-H copula with $\tau = 0.5$; Pink line, ECGE under assumption of G-H copula with $\tau = 0.8$. True dependence structure: Gamma Frailty copula with $\tau = 0.8$. Estimates for $X$ on left and $Y$ on right. Average over 2000 trials.

- $\tau = 0.2$, corresponding to the Gumbel-Hougaard copula with $\alpha = 1.25$ (green line).

- $\tau = 0.5$, corresponding to the Gumbel-Hougaard copula with $\alpha = 2$ (orange line).
• \( \tau = 0.8 \), corresponding to the Gumbel-Hougaard copula with \( \alpha = 5 \) (pink line).

As expected, as the value of \( \tau \) nears the true value of 0.8, the ECGE output moves further from the Kaplan-Meier estimator and closer to the true underlying distribution. Furthermore, the ECGE output under the assumption of \( \tau = 0.8 \) and a Gamma Frailty copula (Figure 7.3) and the ECGE output under the assumption of \( \tau = 0.8 \) and a Gumbel-Hougaard copula (Figure 7.4) are very similar. Comparable results are also obtained under the assumption of the Frank copula with \( \tau = 0.8 \). This supports the conclusions drawn by Zheng and Klein in their simulation study. The estimator is not particularly sensitive to the parametric choice of copula. Rather, the quality of the results are more dependent on the implied strength of association.

### 7.3.3 Simulation 3: Investigating the Frank copula and negative dependence

<table>
<thead>
<tr>
<th>Simulation 3 Set up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Distributions</td>
</tr>
<tr>
<td>( X \sim \text{Gamma}(5, 1/6) )</td>
</tr>
<tr>
<td>( Y \sim \text{Exponential}(10/19) )</td>
</tr>
<tr>
<td>( Z \sim \text{Uniform}(0.35, 10) )</td>
</tr>
<tr>
<td>Dependence Structure</td>
</tr>
<tr>
<td>Truth: Frank with ( \tau = -0.5 )</td>
</tr>
<tr>
<td>Assumed: Frank with ( \tau = -0.5 )</td>
</tr>
<tr>
<td>Frank with ( \tau = -0.2 )</td>
</tr>
<tr>
<td>Frank with ( \tau = -0.8 )</td>
</tr>
</tbody>
</table>

In our final simulation, we test our estimator on data where \( X \) and \( Y \) exhibit negative dependence. (This is the type of dependence we hypothesize is present in the
dividend initiation data.) In this simulation, \( X \) is generated by the Gamma(5, 1/6) distribution and \( Y \) is generated by the Exponential(10/19) distribution. They are linked by a Frank copula with \( \alpha = 1/\exp(-5.736283) \), corresponding to \( \tau = -0.5 \). This indicates moderate negative association between \( X \) and \( Y \). \( Z \) is generated by the Uniform(0.35, 10) distribution, independently of \( X \) and \( Y \). 1,000 data points are simulated for each trial.

Figure 7.5: Simulation 3: Performance of the ECGE under Assumption of the Correct Dependence Structure. Black line, true underlying distribution that generated the data; Pink line, ECGE; Blue line, K-M estimator. Dependence structure: Frank copula with \( \tau = -0.5 \). Estimates for \( X \) on left and \( Y \) on right. Average over 2000 trials.
Figure 7.5 shows the results of the ECGE (pink line) under the correctly specified copula and dependence structure. Again, the Kaplan-Meier estimator (blue line) is shown for reference. For both $X$ and $Y$, there is a severe lack of data past $T = 1.25$. Up to this point, the ECGE performs well.

Figure 7.6: Simulation 3: Performance of the ECGE under Assumption of the Correct Copula and Varying Strengths of Dependence. Black line, true underlying distribution that generated the data; Blue line, K-M estimator; Green line, ECGE under assumption of Frank copula with $\tau = -0.2$; Orange line, ECGE under assumption of Frank copula with $\tau = -0.8$. True dependence structure: Frank copula with $\tau = -0.5$. Estimates for $X$ on left and $Y$ on right. Average over 2000 trials.

In Figure 7.6 we show the ECGE under the assumption of the correct parametric
copula and varying strengths of dependence, with $\tau = 0$ (blue line), $\tau = -0.2$ (green line), and $\tau = -0.8$ (orange line). Comparing $\tau = 0$ to $\tau = -0.2$ (blue to green) shows the gains in accuracy that are achieved by recognizing a small amount of negative dependence. The orange line corresponding to $\tau = -0.8$ reflects the danger of overestimating the strength of negative association.
Chapter 8

Application: Revisiting Dividend Initiation

In this chapter we return to the analysis of dividend initiation and revisit the survival curves presented in Chapter 4. We focus our efforts on the curves generated from incorporation. Whereas the curves generated in Chapter 4 are Kaplan-Meier estimators and built under the assumption of independent censoring, we now account for the dependence between $X$ and $Y$ using the ECGE approach.

Recall that in the dividend data, our event of interest is dividend initiation. However, several firms drop out of the sample due to bankruptcy. Common sense tells us that there is a negative association between dividend initiation, $X$, and bankruptcy, $Y$. Furthermore, there are other sources of censoring in the data, most notably mergers, that we consider to be independent of (or at least ambiguous signals for) dividend initiation. These non-bankruptcy censorings are treated as a third class of event referred to as $Z$. Censoring by end of the study is also included in this third group.
Acknowledging the presence of dependent censoring in our data is especially critical because the bankruptcies are not distributed evenly throughout the sample. For example, we see a much higher proportion of bankruptcy censorings in the 1985-95 listing group than in the 1965-75 listing group (11.6% vs. 3.2%). We also see a much higher proportion in the Nasdaq sample than in the NYSE sample (12.4% vs. 4.5%). This is illustrated in Table 8.1.

Table 8.1: Distribution of Event Types

<table>
<thead>
<tr>
<th>Listing Grp</th>
<th>n</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dividend Initiation</td>
<td>Bankruptcy</td>
<td>Independent Censoring</td>
</tr>
<tr>
<td>1965-75</td>
<td>249</td>
<td>221 (88.8%)</td>
<td>8 (3.2%)</td>
<td>20 (8.0%)</td>
</tr>
<tr>
<td>1985-95</td>
<td>958</td>
<td>260 (27.1%)</td>
<td>111 (11.6%)</td>
<td>587 (61.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>1207</td>
<td>481 (39.9%)</td>
<td>119 (9.9%)</td>
<td>607 (50.3%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Listing Grp</th>
<th>n</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dividend Initiation</td>
<td>Bankruptcy</td>
<td>Independent Censoring</td>
</tr>
<tr>
<td>1965-75</td>
<td>174</td>
<td>166 (95.4%)</td>
<td>3 (1.7%)</td>
<td>5 (2.9%)</td>
</tr>
<tr>
<td>1985-95</td>
<td>203</td>
<td>114 (56.2%)</td>
<td>14 (6.9%)</td>
<td>75 (36.9%)</td>
</tr>
<tr>
<td>Total</td>
<td>377</td>
<td>280 (74.3%)</td>
<td>17 (4.5%)</td>
<td>80 (21.2%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Listing Grp</th>
<th>n</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dividend Initiation</td>
<td>Bankruptcy</td>
<td>Independent Censoring</td>
</tr>
<tr>
<td>1965-75</td>
<td>75</td>
<td>55 (73.3%)</td>
<td>5 (6.7%)</td>
<td>15 (20.0%)</td>
</tr>
<tr>
<td>1985-95</td>
<td>755</td>
<td>146 (19.3%)</td>
<td>97 (12.8%)</td>
<td>512 (67.8%)</td>
</tr>
<tr>
<td>Total</td>
<td>830</td>
<td>201 (24.2%)</td>
<td>103 (12.4%)</td>
<td>527 (63.5%)</td>
</tr>
</tbody>
</table>
8.1 Re-estimating the survival curves

In order to account for the dependence between dividend initiation and bankruptcy using the ECGE, we must specify a parametric form for the dependence and the strength of association. In other words, we must select a copula and a copula parameter. Since we believe that dividend initiation and bankruptcy are negatively associated, the Frank copula is an obvious choice for the parametric form. We also believe that the association between dividend initiation and bankruptcy is substantial. Following the suggestion of Zheng and Klein, we will use the ECGE to establish bounds on the true survival function by assuming that the value of Kendall’s tau lies between -0.6 and -0.9.

We first return to the survival curves generated for the entire sample of firms. These curves were presented in Figure 4.1 on page 52. There are a total of 1,207 firms, with 249 in the 1965-75 listing group and 958 in the 1985-95 listing group. We observe 119 bankruptcy censorings, representing 3.2% of the 1965-75 group and 11.6% of the 1985-95 group.

The adjusted survival curves are shown in Figure 8.1. The blue curves correspond to the 1965-75 listing group and the pink curves correspond to the 1985-95 listing group. The dotted lines show the Kaplan-Meier survival estimate for each group and correspond to the survival estimates given in Figure 4.1. They are included to provide a reference for gauging the effect of accounting for dependent censoring. The solid curves correspond to the ECGE generated under the assumption of a Frank
Figure 8.1: *All Firms: ECGE Survival Curves for Dividend Initiation with Time Origin Incorporation*. Blue curves correspond to the 1965-75 listing group; Pink curves correspond to the 1985-95 listing group. Solid curves show ECGE output under the assumption of a Frank copula with \( \tau = -0.6 \) (lower curve) and \( \tau = -0.9 \) (upper curve). Dotted curves show K-M survival estimates.

Because we are unable to determine the actual strength of association between dividend initiation and bankruptcy, the solid curves are meant to represent bounds on the true survival function.

For the 1965-75 listing group, the ECGE survival curves do not vary much from their Kaplan-Meier counterpart. This is not surprising; there are very few dependent
Figure 8.2: All Firms: Confidence Limits for ECGE Survival Curves. Blue curves correspond to the 1965-75 listing group; Pink curves correspond to the 1985-95 listing group. Dotted curves show ECGE output under the assumption of a Frank copula with $\tau = -0.6$ (lower curve) and $\tau = -0.9$ (upper curve). Lower solid curve shows the lower 95% confidence limit for $\tau = -0.6$ and upper solid curve shows the upper 95% confidence limit for $\tau = -0.9$.

censorings in this group. For the 1985-95 listing group, the ECGE survival curves are much higher than the Kaplan-Meier survival curve. The overall effect is a widening of the separation between the groups. When dependent censoring is taken into account, the dividend initiation behaviors of the groups are more dissimilar than the Kaplan-Meier estimators would lead us to believe.

Bootstrapped confidence limits established by the “percentile” method are shown
by the solid lines in Figure 8.2. For an explanation and discussion of this method, see Chernick (1999, p. 53). The upper solid curve gives the upper 95% confidence limit for \( \tau = -0.9 \), and the lower solid curve gives the lower 95% confidence limit for \( \tau = -0.6 \). The dotted curves show the ECGE for \( \tau = -0.6 \) and \( \tau = -0.9 \). The clear separation of the confidence limits over much of the time interval supports the hypothesis of differing dividend initiation behaviors for the two groups.

Figure 8.3: NYSE Firms: ECGE Survival Curves for Dividend Initiation with Time Origin Incorporation. Blue curves correspond to the 1965-75 listing group; Pink curves correspond to the 1985-95 listing group. Solid curves show ECGE output under the assumption of a Frank copula with \( \tau = -0.6 \) (lower curve) and \( \tau = -0.9 \) (upper curve). Dotted curves show K-M survival estimates.
Next we address the survival curves generated for the NYSE sample. These curves were first presented in Figure 4.4 on page 62. There are a total of 377 firms, with 174 from the 1965-75 listing group and 203 from the 1985-95 listing group. There are only 17 bankruptcy censorings in the NYSE sample, corresponding to 1.7% of the 1965-75 group and 6.9% of the 1985-95 group.

Because of the low percentage of bankruptcy censorings in the NYSE sample, the ECGE survival curves shown in Figure 8.3 are not dramatically different from the Kaplan-Meier survival curves. For the 1965-75 group, the ECGE and Kaplan-Meier curves are nearly indistinguishable. For the 1985-95 group, the difference between the ECGE and Kaplan-Meier curves is somewhat greater.

However, the overall narrowing of the separation between the groups is slight and not substantial enough to overturn our previous conclusions. Based on the curves in Figure 8.3 it still appears that the 1985-95 group initiates dividends somewhat faster than the 1965-75 group in the early years following incorporation and that this ordering reverses later on. Despite the fact that the separation past (approximately) \( T = 50 \) is widened, the ECGE curves still do not suggest that the dividend initiation behaviors of the NYSE listing groups are vastly different.

Figure 8.4 shows bootstrapped confidence limits established by the percentile method for the NYSE sample. There is separation between the confidence limits up to approximately twenty years post-incorporation and considerable overlap thereafter. Again, this supports our previous conclusion that the 1985-95 listing group
initiates dividends somewhat faster initially, but thereafter the dividend initiation behaviors of the groups become similar.

Lastly we return to the survival curves generated for the Nasdaq sample of firms. These curves were presented in Figure 4.5 on page 63. There are a total of 830 firms, 75 from the 1965-75 listing group and 755 from the 1985-95 listing group. Bankruptcy censorings account for 6.7% of the 1965-75 group and 12.4% of the 1985-95 group.
Figure 8.5: Nasdaq Firms: ECGE Survival Curves for Dividend Initiation with Time Origin Incorporation. Blue curves correspond to the 1965-75 listing group; Pink curves correspond to the 1985-95 listing group. Solid curves show ECGE output under the assumption of a Frank copula with $\tau = -0.6$ (lower curve) and $\tau = -0.9$ (upper curve). Dotted curves show K-M survival estimates.

The adjusted survival curves for the Nasdaq sample are shown in Figure 8.5. Because of the high proportion of bankruptcy censorings in the Nasdaq sample, the ECGE curves are noticeably higher than the Kaplan-Meier curves, especially for the 1985-95 listing group. This widens the separation between the groups. After taking dependent censoring into account, the dividend initiation behaviors of the listing groups are more dissimilar than indicated in the analysis presented in Chapter 4.
Figure 8.6: *Nasdaq Firms: Confidence Limits for ECGE Survival Curves.* Blue curves correspond to the 1965-75 listing group; Pink curves correspond to the 1985-95 listing group. Dotted curves show ECGE output under the assumption of a Frank copula with $\tau = -0.6$ (lower curve) and $\tau = -0.9$ (upper curve). Lower solid curve shows the lower 95% confidence limit for $\tau = -0.6$ and upper solid curve shows the upper 95% confidence limit for $\tau = -0.9$.

The bootstrapped confidence limits for the Nasdaq sample shown in Figure 8.6 reinforce the clear separation between the listing groups.

Adjusting the survival curves to account for the dependence between dividend initiation and bankruptcy has the general effect of widening the separation between the groups. This is not surprising, given the greater proportion of bankruptcy censorings in the later listing group. However, for our sample of primary interest, the NYSE
sample, we argue that accounting for dependent censoring does not produce a strong enough effect to reverse our previous conclusion. The dividend initiation behavior of the NYSE sample does not seem to vary significantly between the listing groups.

8.2 Sensitivity to tie breaking

Next we investigate the sensitivity of our estimator to our tie breaking scheme. We have identified the need to break ties between $X$ and $Y$ events in order to maximize the grid on which the estimator is calculated. Up to this point, we have avoided ties by subtracting a small amount, 0.01 years, from each observed $Y$ lifetime in our dataset. In this manner, our convention dictates that if Firm A experiences dividend initiation at $T = 5$ years and Firm B experiences bankruptcy at $T = 5$ years, Firm B's event time is recoded to 4.99.

We could have just as easily avoided ties by shortening the observed $X$ lifetime. For the sake of convenience, we shall refer to the method of subtracting from the $Y$ lifetimes as the $Y$-Minus technique and the method of subtracting from the $X$ lifetimes as the $X$-Minus technique.

We return to the NYSE sample and the survival curves plotted in Figure 8.3. These curves were calculated using the $Y$-Minus technique. We recalculate the six curves using the $X$-Minus technique and analyze the differences. An overlay plot of the survival estimates generated by the two techniques would yield six pairs of nearly indistinguishable curves. Therefore, in Figure 8.7 we present plots of the differences in the estimators generated by the two tie breaking approaches. The difference value
is calculated as the \( Y{\text{-}}\text{Minus} \) estimate minus the \( X{\text{-}}\text{Minus} \) estimate. Table 8.2 gives the values of the maximum difference between each pair.

Obviously, the estimates will differ over the intervals \([t-0.01, t)\), where \( t \) represents a tied lifetime. Note, however, that these intervals can be made as short as we desire. The selection of 0.01 as the value for subtraction was purely arbitrary. Therefore, we ignore these intervals of known discrepancy when analyzing the differences in the estimators resulting from the two tie breaking approaches.

<table>
<thead>
<tr>
<th>Copula Specification</th>
<th>1965-75 Listing Group</th>
<th>1985-95 Listing Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>( 4.591314 \times 10^{-4} )</td>
<td>( 1.661852 \times 10^{-3} )</td>
</tr>
<tr>
<td>Frank ((\tau = -0.6))</td>
<td>( 8.193980 \times 10^{-5} )</td>
<td>( 1.047021 \times 10^{-4} )</td>
</tr>
<tr>
<td>Frank ((\tau = -0.9))</td>
<td>( 1.978474 \times 10^{-8} )</td>
<td>( 6.978255 \times 10^{-10} )</td>
</tr>
</tbody>
</table>

For the 1965-75 listing group, there are 74 total event times. At three of these times, we observe both \( X \) and \( Y \) failures. For the 1985-95 listing group, there are 54 total event times. At eight of these times, we observe both \( X \) and \( Y \) failures. These non-unique event times involve 17% of the 1985-95 listing group. For this group, it is clear that a sizable proportion of the observations experience an event at a non-unique event time.

Based on Table 8.2 and Figure 8.7, we assert that our estimator is not overly sensitive to the tie breaking approach. The maximum difference in the survival estimates is \( 1.661852 \times 10^{-3} \) for the independence copula and even smaller for the Frank
Figure 8.7: Differences in Survival Estimates from Two Tie Breaking Approaches. Results for the 1965-75 group shown in Panels A, C, and E. Results for the 1985-95 group shown in Panels B, D, and F. First row shows differences in survival estimates under the independence copula; second row under the Frank copula with \( \tau = -0.6 \); third row under the Frank copula with \( \tau = -0.9 \). For the dividend initiation data, \( \tau \) likely between -0.6 and -0.9.
copulas. For the Frank copula with $\tau = -0.9$, the differences are so small that the scale of the $y$-axis would have to be drastically altered in order to see them.

For the NYSE sample, tie breaking does not seem to be an issue. For the entire sample, where nearly 31% of the firms experience an event at a non-unique time, tie breaking also does not seem to be problematic. For the six pairs of survival curves, the maximum difference between the tie breaking approaches is $2.565086 \times 10^{-3}$, occurring for the independence copula on the 1985-95 listing group.
Chapter 9

Conclusion

Recent papers in the finance literature report a marked decline in the proportion of dividend-paying industrial firms over the period 1978 to 1999. Fama and French (2001) attribute this decline to a decrease in the propensity of firms to pay dividends and to changing firm characteristics. Drawing from life cycle theory, we tested the hypothesis that a previously unconsidered characteristic, firm age, should be accounted for in a model of dividend initiation. Our analysis supports the hypothesis that changes in the distribution of incorporation-age among newly listed firms partially explain the decline in the proportion of dividend-paying firms. In this manner, part of the decline that was previously attributed to the decreased propensity to pay effect is yet another example of changing firm characteristics. By taking incorporation-age into account, our model reduces the magnitude of the decreased propensity to pay effect.

We also tested the hypothesis that the magnitude of the decreased propensity to pay effect differs across exchanges. In a comparison of a sample of NYSE firms to a
sample of Nasdaq firms, our analysis shows a much stronger decrease in the propensity to pay on the Nasdaq. In fact, we find no statistically significant evidence of a decrease in the propensity to pay dividends for our NYSE sample when incorporation-age is taken into account.

We have analyzed the dividend initiation data using survival analysis techniques. Because the data has a time-to-event structure and censoring is present, survival analysis (rather than logistic regression) is the most appropriate method for analyzing the data.

The conclusions described above were obtained using standard survival analysis tools, namely the Kaplan-Meier estimator for marginal survival and the Cox Proportional Hazards model. However, the dividend initiation data does not satisfy one of the usual assumptions of survival analysis—that of independent censoring. The most prevalent source of non-administrative censoring in the data, bankruptcy, should be considered informative for dividend initiation. At the same time, however, there remain sources of censoring other than bankruptcy that can be treated as non-informative for the event of interest.

The challenge of dealing with the dependent and independent censoring present in the dividend initiation data led to the development of the Extended Copula-Graphic Estimator (ECGE). The ECGE is an extension of Zheng and Klein's copula-graphic estimator and allows for the estimation of marginal survival functions in the presence of both dependent and independent censoring.
Using the ECGE, we are able to refine and improve the marginal survival estimates presented in our analysis of dividend initiation. Adjusting the survival curves to account for dependent censoring has the general effect of increasing the magnitude of the decreased propensity to pay effect (relative to the levels established in our initial analysis). However, for the NYSE sample, accounting for dependent censoring still does not produce a strong enough effect to indicate a significant decrease in the propensity to pay dividends among newly listed firms.

We have identified two areas for future work. First, it would be helpful to develop a Log-Rank-type test statistic for curves generated by the ECGE. Currently, we have no formal statistic for testing equality of survival curves generated by the ECGE. Such a statistic would clearly strengthen our discussion of the survival estimates for the dividend initiation data.

Secondly, it would be useful to extend the ECGE to multiple sources of dependent censoring. Currently, the estimator accommodates only two dependent competing risks. However, potential applications may have richer data structures that go beyond a dependence between $X$ and $Y$. Given the ease with which copulas can be extended beyond the bivariate case, a copula-based approach should be able to handle multiple dependent competing risks. In the dividend initiation data, this would give us the opportunity to explore the effect of mergers being treated as informative for dividend initiation.

Although outside the scope of our current research focusing on marginal survival,
it would be a significant contribution to the field of survival analysis to modify the Cox PH model to account for dependent censoring. Along these lines, Huang and Wolfe (2002) have developed a Cox-based frailty model for dependent censoring. Their model is applicable to clustered data.

The analysis of dividend initiation is but one of many possible applications of survival analysis techniques to financial data. A quick scan of the finance literature reveals recent papers using survival analysis in a "conventional" application—the study of firm survivorship. See, for example, Manigart et al. (2002) and Turetsky and McEwen (2001). In a novel application of survival analysis techniques, Lo et al. (2002) model the execution of limit orders. (A limit order is a directive to buy a specified amount of stock at a specified price. Because the order may be cancelled at any time or may fail to be executed if the specified price is not met before expiration, the study of limit orders involves censored data.)

Beyond these applications, we foresee much potential for the use of survival analysis in financial modeling. One of the timeliest topics in finance, credit risk, is a natural fit to the time-to-event structure of survival analysis. For example, survival analysis could be used to model time-to-prepayment or time-to-default of consumer mortgages or could be used on an institutional level to model credit downgrades.
Appendix A

Filtering and Formatting the Data

A.1 Creating the CRSP sample

To create the CRSP sample, we follow the steps outlined by Fama and French (2001, pp. 40–42). We pull monthly data from January 1925 to December 2002 by PERMNO for the 7,732 firms in the Jovanovic and Rousseau (2001) dataset. (Recall that this dataset provides the years of founding (FND), incorporation (INC), and listing (LIST).) The following variables are pulled from the CRSP database: PERMNO, DATE, SHRCID, EXCHCD, SICCD, SHROUT, PRC, RET, RETX, and DLSTCD. 3,352 firms are eliminated from our sample due to at least one of the following: they are not new lists, they are a utility or financial firm, they have an invalid value for SHRCID, or they list before 1925 (the start of CRSP coverage). Details on the first three exclusion criteria can be found in Fama and French.

This leaves us with a CRSP sample of 4,380 firms. 1,309 of our 4,380 firms are missing INC. Ten firms have CRSP coverage which predates their INC value. These firms are removed. This reduces our CRSP sample to 3,061 firms. After removing
the AMEX firms and restricting our analysis to firms which listed in the intervals 1965-75 and 1985-95, our final CRSP sample consists of 1,207 firms.

A.2 Creating the Compustat sample

To create the Compustat sample, we again follow the steps outlined by Fama and French (2001, pp. 40-42). We pull annual data from 1965 to 2002 by PERMNO for the 1,207 firms in the CRSP sample. The following data items are pulled: 6, 15, 18, 25, 50, and 199. The following quantities are calculated as outlined by Fama and French:

- Market Equity \( (M_E_t) = DATA199 \times DATA25. \)
- Earnings Before Interest \( (E_t) = DATA18 + DATA15 + DATA50 \) (if available).
- Assets \( (A_t) = DATA6. \)

In their analysis, Fama and French use three explanatory variables which are derived from the quantities given above: profitability = \( E_t/A_t \), the growth rate of assets = \( dA_t/A_t = (A_t - A_{t-1})/A_t \), and the percentage of NYSE firms with the same or lower market capitalization = \( NYP_t \). (Fama and French also consider a fourth explanatory variable, the market-to-book ratio, as an additional proxy for investment opportunity. However, one key assumption of their method for establishing the decreased propensity to pay effect is that the proxies have constant meaning throughout the sample period. They state that this assumption is especially suspect for the
market-to-book ratio (Fama and French, 2001, p. 24). Therefore, all their analyses are run twice—with and without the market-to-book ratio included as a covariate. Either way, the same general results are obtained.) 1,008 firms (of the 1,207 CRSP firms) are sufficiently covered in the Compustat database.

Missing values are an issue with the Compustat data. Five firms are eliminated from the sample because they are completely lacking Compustat coverage over a span of several years. Other firms are missing Compustat coverage for a single year or are missing values of certain covariates in certain years. In order to maintain the size of our sample, we replace a missing covariate value in year $t$ with the average of the covariate values in years $t-1$ and $t+1$. Using this method, a firm is excluded from the Compustat sample due to missing data only if it is lacking a covariate value for three consecutive years.

In order to fit the Cox PH model with incorporation as the time origin, we need covariate values for all years between incorporation and dividend/censoring. However, this information is only publicly available for the time period between listing and dividend/censoring. Therefore, we need a method to impute covariate values for the time period between incorporation and listing. We propose two such methods, the straight-line method and the backfill method.

For the straight-line method, assume that the covariate values are zero at founding and identify the covariate values in the first year after listing. Fill in the values for the years between using a straight-line imputation. If a firm is missing the value of $FND$,
use the value suggested by the median length of the \([FND, INC]\) interval among other firms with the same listing year. For the categorical models, determine the percentiles of the covariate values in the first year after listing. Assume the covariate percentiles are zero at founding and fill in the percentile values for the years between using a straight-line imputation. Assign the imputed percentile values to their proper categorical level.

For the backfill method, fill in the years between \(INC\) and \(LIST\) using the covariate values from the first year after listing. For the categorical models, fill in using the categorial level from the first year after listing.

### A.3 Formatting the data for survival analysis

To get the data into the form for survival analysis, we must define the event time and set up the status variable. The purpose of the status variable is to indicate if the observation experiences the event of interest or is censored. Our definition of the event corresponds to that of Fama and French (2001). A firm is classified as a dividend-payer in a given year if at least one monthly with-dividend-return \((RET)\) exceeds its without-dividend-return \((RETX)\). For each firm, we identify the year in which this relation first occurs and designate it \(MINDIVYR\). For firms that are censored, \(MINDIVYR\) is set equal to their censoring year.

Our analysis contains two different definitions of event time. \(FROMLAST\) is the number of years from listing to the initiation of dividends. \(FROMINC\) is the number of years from incorporation to the initiation of dividends.
In survival analysis, immediate failures are not allowed. Despite the fact that some firms list and initiate dividends within a twelve month period, we cannot assign them a FROMLIST value of zero. Furthermore, our values of INC and LIST are rounded. We know only the year. Therefore, we must make the following clarification and adjustment to the way event times are defined in our sample.

FROMINC is recorded as MINDIVYR - INC. If a firm incorporates in January of year \( t \) and initiates its first dividend in year \( t + 1 \), FROMINC = 1. If a firm incorporates in December of year \( t \) and initiates its first dividend in year \( t + 1 \), FROMINC is also coded as 1. As imprecise as this may be, we are limited by our inexact knowledge of the incorporation date. For the sake of consistency, FROMLIST is calculated in the same manner.

If a firm incorporates in year \( t \) and pays its first dividend in year \( t \), FROMINC is coded as 1. The same holds for FROMLIST. To be consistent when setting up our Compustat sample, we ignore the covariate values in the actual year of listing, \( t \), and begin tracking the covariate values in year \( t + 1 \). This convention is also necessitated by the construction of the investment covariate.

The status variable, PAYER, equals 1 if the firm initiates dividends and 0 if the firm is censored. Firms are marked as censored if they are still actively traded at the end of the study (12/31/2002) but are yet to initiate dividends or if they delist between LIST and 12/31/2002 without ever having paid dividends.
Appendix B

_S-Plus_ Code for the ECGE

B.1  _g.km.xy_

g.km.xy <- function(time.vec, event.vec, bound){
  ## Recode X's and Y's to events and Z's to censorings
  event.vec2 <- rep(0, length(event.vec))
onestwos <- event.vec < 3
  event.vec2 <- 1*onestwos

  ## Identify the (unique) event and censoring times and order them
  # Remove multiplicities
  censor.times <- sort(unique(time.vec[event.vec2 == 0]))

  ## Want to add leading zero to vector of event times
  # All event times
  event.times <- time.vec[event.vec < 3]
  # Sorted (unique) event times w/ leading 0
  event.times2 <- c(0, sort(unique(event.times)))

  ## For each event time, identify number and risk and number of failures
  n.risk <- rep(0, length(event.times2))
n.failure <- rep(0, length(event.times2))
  for (i in 1:length(event.times2)){
    n.risk[i] <- sum(time.vec >= event.times2[i])
    n.failure[i] <- sum(event.times == event.times2[i])
  }

  ## Calculate K-M estimator
  km.contrib <- rep(0, length(event.times2))
  km.contrib <- (n.risk - n.failure)/n.risk
  km.xy <- cumprod(km.contrib)
## Restrict output to prespecified bound

1. Make intervals of form \([\text{left.int}, \text{right.int})\)

\[
\text{left.int} \leftarrow \text{event.times2}
\]

\[
\text{right.int} \leftarrow \text{c(event.times2[-1],1000)}
\]

\[
\text{intervals} \leftarrow \text{cbind(left.int, right.int)}
\]

2. Identify last row for intervals; this is where \(\text{left.int} \leq \text{bound}\)
   and \(\text{right.int} > \text{bound}\)

\[
\text{gt.rows} \leftarrow \text{intervals[\text{right.int} > \text{bound},]}
\]

\[
\text{gt.rows} \leftarrow \text{as.matrix(gt.rows, nrow=1)}
\]

3. S-Plus is being difficult here. If there is only one row in \(\text{gt.rows}\),
   it sets the dimensions as 2x1; we want a 1x2 so we can \text{rbind} it; must
   transpose this special case

\[
\text{if (nrow(gt.rows)==2 & ncol(gt.rows)==1)}{
    \text{intervals2} \leftarrow \text{rbind(intervals[\text{right.int} \leq \text{bound},], t(gt.rows))}
}\]

4. If number of rows in \(\text{gt.rows} > 1\), then we are OK and can just
   \text{rbind} first row

\[\text{else } \{
    \text{intervals2} \leftarrow \text{rbind(intervals[\text{right.int} \leq \text{bound},], \text{gt.rows}[1,])}
\}\]

5. Set final \text{right.int} equal to the bound

\[
\text{intervals2[nrow(intervals2),2]} \leftarrow \text{bound}
\]

6. Format output

\[
\text{length.out} \leftarrow \text{nrow(intervals2)}
\]

\[
\text{n.risk} \leftarrow \text{n.risk[1:length.out]}
\]

\[
\text{n.failure} \leftarrow \text{n.failure[1:length.out]}
\]

\[
\text{km.xy} \leftarrow \text{km.xy[1:length.out]}
\]

\[
\text{ret.mat} \leftarrow \text{cbind(intervals2, n.risk, n.failure, km.xy)}
\]

7. Plot K-M estimator

\[
\text{event.mark} \leftarrow \text{intervals2[,1]}
\]

\[
\text{plot(event.mark, km.xy, pch=18, xlab="Time", ylab="Survival",}
\]

\[
\text{ylim=c(0,1), xlim=c(0, bound))}
\]

\[
\text{title(\"Kaplan-Meier estimator for W=\min(X,Y) in the presence of Z \ Pr(W>t) = Pr(X>t, Y>t)\")}
\]

\[
\text{event.markb} \leftarrow \text{event.mark[-1]}
\]

\[
\text{segments(event.mark[-length(event.mark)], km.xy[-length(km.xy)],}
\]

\[
\text{event.markb, km.xy[-length(km.xy)]}
\]

\[
\text{segments(event.markb[length(event.markb)], km.xy[length(km.xy)],}
\]

\[
\text{km.xy[length(km.xy)]}
\]
g.ipcw <- function(time.vec, event.vec, bound){

    # First, compute K-M for censorings
    
    ## Recode the Z's to events and the X's and Y's to failures
    event.vec.z <- rep(0, length(event.vec))
    threes <- event.vec == 3
    event.vec.z <- 1*threes

    ## Identify the (unique) z event and xy censoring times and order them
    # Remove multiplicities
    censor.times.xy <- sort(unique(time.vec[event.vec.z == 0]))

    ## Want to add leading zero to vector of z event times
    event.times.z <- time.vec[event.vec == 3] # All event times
    ## Add small amount of time to each Z
    event.times.z <- event.times.z + 0.01
    event.times2.z <- c(0, sort(unique(event.times.z))) # Sorted (unique)
    # event times w/
    # leading 0

    ## Create new time.vec with 0.01 added to Z times
    new.time.vec <- c(time.vec[event.vec<3], event.times.z)

    ## For each z event time, identify number and risk and
    ## number of failures
    n.risk <- rep(0, length(event.times2.z))
    n.failure <- rep(0, length(event.times2.z))
    for (i in 1:length(event.times2.z)){
        n.risk[i] <- sum(new.time.vec >= event.times2.z[i])
        n.failure[i] <- sum(event.times.z == event.times2.z[i])
    }

    ## Calculate K-M estimator
    km.contrib <- rep(0, length(event.times2.z))

}
km.contrib <- (n.risk - n.failure)/n.risk
km.z <- cumprod(km.contrib)

## Restrict output to prespecified bound
# Make intervals of form [left.int,right.int)
left.int <- event.times2.z
right.int <- c(event.times2.z[-1],1000)
intervals <- cbind(left.int,right.int)

# Identify last row for intervals; this is where
# left.int <= bound and right.int > bound
gt.rows <- intervals[right.int > bound,]
if (nrow(gt.rows)==2 & ncol(gt.rows)==1){
    intervals2 <- rbind(intervals[right.int <= bound,],t(gt.rows))
}
else {
    intervals2 <- rbind(intervals[right.int <= bound,],gt.rows[1,])
}

# Set final right.int equal to the bound
intervals2[nrow(intervals2),2] <- bound

## Format output
length.out <- nrow(intervals2)
n.risk <- n.risk[1:length.out]
n.failure <- n.failure[1:length.out]
km.z <- km.z[1:length.out]
ret.mat <- cbind(intervals2,n.risk,n.failure,km.z)

## Plot K-M estimator
event.mark <- intervals2[,1]
plot(event.mark,km.z,pch=18,xlab="Time",ylab="Survival",
     ylim=c(0,1),xlim=c(0,bound))
title("Kaplan-Meier estimator for Z in the presence of (X,Y)
Pr(Z>t)")
event.markb <- event.mark[-1]
segments(event.mark[-length(event.mark)],km.z[-length(km.z)],
## Calculate estimator of $\hat{P}(t)$ up to bound

# Identify event times (X's)
```r
event.vec.x <- rep(0, length(event.vec))
one.s <- event.vec == 1
event.vec.x <- 1*one.s
event.times.x <- time.vec[event.vec.x == 1]
event.times2.x <- c(0, sort(unique(event.times.x)))
```

# Identify number of events at each X event time
```r
n.failure.x <- rep(0, length(event.times2.x))
for (i in 1:length(n.failure.x)) {
  n.failure.x[i] = sum(event.times.x == event.times2.x[i])
}
```

# Identify weight (from km.z) for each X event time
# Don't worry about weight at time zero; this will be
# concatenated in manually
```r
weight.z <- rep(0, length(event.times2.x)-1)
for (i in 1:length(weight.z)) {
  in.right <- event.times2.x[i+1] > intervals[,1]
  in.left <- event.times2.x[i+1] <= intervals[,2]
  in.bounds = in.right*in.left
  weight.z[i] <- km.z[in.bounds == 1][1]
}
```

# Concatenate in weight for time zero
```r
weight.z <- c(1, weight.z)
```

## Calculate $\hat{P}(t)$
```r
p1.contrib <- n.failure.x/weight.z/length(time.vec)
p1.hat <- cumsum(p1.contrib)
```

## Plot estimator of $\hat{P}(t)$
```r
plot(event.times2.x, p1.hat, pch=18, xlab="Time", ylab="p1.hat",
     main="Estimate of $\hat{P}(t) = \Pr(X \leq t, X < Y)$",
     xlim=c(0,bound), ylim=c(0,1))
segments(event.times2.x, p1.hat, c(event.times2.x[-1], bound), p1.hat)
segments(event.times2.x[-1], p1.hat[-length(p1.hat)],
         event.times2.x[-1], p1.hat[-1], lty=2)
```
## Make intervals

```r
left.int.x <- event.times2.x
right.int.x <- c(event.times2.x[-1], bound)
intervals.x <- cbind(left.int.x, right.int.x)
```

```r
ret.mat2 <- cbind(intervals.x, n.failure.x, weight.z, p1.hat)
```

## For the fun of it, calculate estimator of p2.hat

```r
## Identify event times (Y's)
event.vec.y <- rep(0, length(event.vec))
twos <- event.vec == 2
event.vec.y <- 1*twos
event.times.y <- time.vec[event.vec.y == 1]
event.times2.y <- c(0, sort(unique(event.times.y)))

## Identify number of events at each Y event time
n.failure.y <- rep(0, length(event.times2.y))
for (i in 1:length(n.failure.y)){
    n.failure.y[i] = sum(event.times.y == event.times2.y[i])
}

## Identify weight (from km.z) for each Y event time
# Don't worry about weight at time zero; this will be
# concatenated in manually
weight.z <- rep(0, length(event.times2.y)-1)
for (i in 1:length(weight.z)){
    in.right <- event.times2.y[i+1] > intervals[,1]
    in.left <- event.times2.y[i+1] <= intervals[,2]
    in.bounds = in.right*in.left
    weight.z[i] <- km.z[in.bounds == 1][1]
}

## Concatenate in weight for time zero
weight.z <- c(1, weight.z)
```

```r
## Calculate p2.hat
p2.contrib <- n.failure.y/weight.z/length(time.vec)
p2.hat <- cumsum(p2.contrib)
```

## Make more intervals

```r
left.int.y <- event.times2.y
right.int.y <- c(event.times2.y[-1], bound)
intervals.y <- cbind(left.int.y, right.int.y)
```
ret.mat3 <- cbind(intervals.y,n.failure.y,weight.z,p2.hat)

return(ret.mat,ret.mat2,ret.mat3)
}

B.3 g.emp.prob.maker

g.emp.prob.maker. <- function(time.vec,event.vec){

## Address ties between X and Y
time.vec <- time.vec - .01*(event.vec==2)
#time.vec <- time.vec - .01*(event.vec==1)

par(mfrow=c(3,1))
# Find bound; bound will be max(X,Y)
l.bound <- max(time.vec[event.vec<3])
print(c("The bound is", l.bound))

# Find estimator of P(X>t,Y>t) (k.hat) up to bound
step.one <- g.km.xy(time.vec,event.vec,l.bound)

# Find Pi.hat up to bound
step.two <- g.ipcw(time.vec,event.vec,l.bound)

# Identify relevant X, Y times
event.times.xy <- time.vec[event.vec < 3]
event.types.xy <- event.vec[event.vec < 3]
event.times.xy <- sort(unique(event.times.xy))

## For each X, Y time, find k.hat, this estimator is
## RIGHT CONTINUOUS
# Get intervals out of step.one
k.hat.left.int <- step.one[,1]
k.hat.right.int <- step.one[,2]
k.hat.vals <- step.one[,5]

k.hat <- rep(0,length(event.times.xy))
for (i in 1:length(k.hat)){
  left.bound <- event.times.xy[i] >= k.hat.left.int
  right.bound <- event.times.xy[i] < k.hat.right.int
  in.bound <- left.bound*right.bound
  k.hat[i] <- k.hat.vals[in.bound==1][1]
# May need to do last interval by hand
# (last interval contains both left and right endpts)
last.left.int <- k.hat.left.int[length(k.hat.left.int)]
last.right.int <- k.hat.right.int[length(k.hat.right.int)]
last.event <- event.times.xy[length(event.times.xy)]

if (last.left.int == last.right.int & last.event == last.left.int)
  k.hat[length(k.hat)] <- k.hat.vals[length(k.hat.vals)]

## For each $X, Y$ time, find $\hat{\pi}$, this estimator is
## RIGHT CONTINUOUS
# Get intervals out of step.two
pi.hat.left.int <- step.two$ret.mat2[,1]
pi.hat.right.int <- step.two$ret.mat2[,2]
pi.hat.vals <- step.two$ret.mat2[,5]

pi.hat <- rep(0,length(event.times.xy))
for (i in 1:length(pi.hat)){
  left.bound <- event.times.xy[i] >= pi.hat.left.int
  right.bound <- event.times.xy[i] < pi.hat.right.int
  in.bound <- left.bound*right.bound
  pi.hat[i] <- pi.hat.vals[in.bound==i][1]
}

## May need to do last interval by hand
## (last interval contains both left and right endpts)
last.right.int <- pi.hat.right.int[length(pi.hat.right.int)]
if (last.event == last.right.int)
  pi.hat[length(pi.hat)] <- pi.hat.vals[length(pi.hat.vals)]

## For each $X, Y$ time, find $\hat{p_2}$, this estimator is
## RIGHT CONTINUOUS
# Get intervals out of step.two
p2.hat.left.int <- step.two$ret.mat3[,1]
p2.hat.right.int <- step.two$ret.mat3[,2]
p2.hat.vals <- step.two$ret.mat3[,5]

p2.hat <- rep(0,length(event.times.xy))
for (i in 1:length(p2.hat)){
  left.bound <- event.times.xy[i] >= p2.hat.left.int
  right.bound <- event.times.xy[i] < p2.hat.right.int
  in.bound <- left.bound*right.bound
  p2.hat[i] <- p2.hat.vals[in.bound==i][1]
## May need to do last interval by hand
## (last interval contains both left and right endpts)
last.right.int <- p2.hat.right.int[length(p2.hat.right.int)]
if (last.event == last.right.int)
  p2.hat[length(p2.hat)] <- p2.hat.vals[length(p2.hat.vals)]

## Find relevant event times
#relevant.x <- sort(unique(time.vec[event.vec == 1]))
#relevant.y <- sort(unique(time.vec[event.vec == 2]))
##relevant.xy <- cbind(relevant.x, relevant.y)

## Format output
emp.probs <- cbind(event.times.xy, k.hat, pi.hat, p2.hat)
sum.check <- k.hat + pi.hat + p2.hat
emp.probs <- cbind(emp.probs, sum.check)
return(emp.probs)

}  

B.4 g.cg.est

```r

# g.cg.est

g.cg.est <- function(g.emp.prob.maker.obj, copula, alpha.param) {

  xy.points <- g.emp.prob.maker.obj[, 1]
  k.hat <- g.emp.prob.maker.obj[, 2]
  p1.hat <- g.emp.prob.maker.obj[, 3]
  p2.hat <- g.emp.prob.maker.obj[, 4]

  ## Take xy.points and recover whether the observation is an X or a Y
  ## Do so by looking at the jumps in p1.hat and p2.hat

  # Assign first point in xy.points
  is.x.a <- 1 * (p1.hat[1] > 0)
  is.y.a <- 1 * (p2.hat[1] > 0)

  # Assign points 2 - n in xy.points
  is.x.b <- 1 * (p1.hat[-1] > p1.hat[-length(p1.hat)])
  is.y.b <- 1 * (p2.hat[-1] > p2.hat[-length(p2.hat)])

  is.x <- c(is.x.a, is.x.b)
  is.y <- c(is.y.a, is.y.b)

```
# Identify relevant X points and relevant Y points
rel.x <- xy.points[is.x == 1]
rel.y <- xy.points[is.y == 1]

# For each xy.point, follow the algorithm

# Need leading row for time = 0
xy.points <- c(0,xy.points)
G.est <- rep(0,length(xy.points))
F.est <- rep(0,length(xy.points))
k.hat <- c(1,k.hat)
is.x <- c(0,is.x)
is.y <- c(0,is.y)

for (i in 2:length(xy.points)){
    # If point is an X event, use eqn (4.5)
    if (is.x[i] == 1) {
        F.est[i] <- copula.eqns.for.x(k.hat[i],F.est[i-1],
        G.est[i-1],copula,alpha.param)
        G.est[i] <- G.est[i-1]
    }

    # If point is a Y event, use eqn (4.6)
    if (is.y[i] == 1) {
        G.est[i] <- copula.eqns.for.y(k.hat[i],F.est[i-1],
        G.est[i-1],copula,alpha.param)
        F.est[i] <- F.est[i-1]
    }
}
F.bar.est <- 1 - F.est
G.bar.est <- 1 - G.est

# Get vectors of X events and Y events and the survival function
# at these points
x.vec <- xy.points[is.x == 1]
F.bar.est.x <- F.bar.est[is.x == 1]
y.vec <- xy.points[is.y == 1]
G.bar.est.y <- G.bar.est[is.y == 1]

# Plot survival functions for F and G
par(oma=c(2,2,2,2))
par(mfrow=c(1,2))
plot(xy.points, F.bar.est, xlab="Time", ylab="Survival", ylim=c(0,1),
type="n", main="Estimated Survival for X")
segments(xy.points[-length(xy.points)], F.bar.est[-length(F.bar.est)],
```r
xy.points[-1], F.bar.est[-length(F.bar.est)])
segments(x.vec, c(1,F.bar.est.x[-length(F.bar.est.x)]),
x.vec, F.bar.est.x)

plot(xy.points, G.bar.est, xlab="Time", ylab="Survival", ylim=c(0,1),
type="n", main="Estimated Survival for Y")
segments(xy.points[-length(xy.points)], G.bar.est[-length(G.bar.est)],
xy.points[-1], G.bar.est[-length(G.bar.est)])
segments(y.vec, c(1,G.bar.est.y[-length(G.bar.est.y)]),
y.vec, G.bar.est.y)

mtext("Estimated Survival Functions for X and Y", outer=T,
side=3, cex=2)
mtext("Copula Type: ", side=1, outer=T, cex=1.5, line=-1, adj=1, at=0)
mtext(copula, outer=T, side=!, cex=1.5, line=-1, adj=1,
at=sort(xy.points)[length(xy.points-2)])

if (copula != "independence")
  param.text <- as.character(alpha.param)
  mtext(param.text, outer=T, side=!, cex=1.5, line=-1, adj=1,
at=sort(xy.points)[length(xy.points-2)])

est.mat<-cbind(xy.points,F.est,G.est,F.bar.est,G.bar.est)

return(est.mat)
}

B.5 copula.eqns.for.x
copula.eqns.for.x <- function(k.val, F.val, G.val, copula.type,
 alpha){
  F.val.cp <- 0
  if (copula.type == "independence") {
    F.val.cp <- (k.val - 1 + G.val)/(G.val - 1);
  }
  if (copula.type == "gamma"){
    F.val.cp <- 1 - (k.val-(1-alpha) + 1 -
    (1/(1-G.val))"(alpha-1))"(-1/(alpha-1))
  }
  if (copula.type == "frank" | copula.type == "gumbel.hougaard"){
    F.val.cp <- nm.copula.for.x(k.val, F.val, G.val,
```
alpha, copula.type)
}
return(F.val.cp)
}

The function copula.eqns.for.y is similar to the above.

B.6 nm.copula.for.x

nm.copula.for.x <- function(k.val, F.valg, G.valg, alpha.val, copula.name){
  # G.val, k.val, and alpha.val are known. We are trying to solve for F.val.
  if (copula.name == "frank") {
    copula.eqn <- function(F.v, G.v, alpha.v){
      ce <- logb((1 + (alpha.v^-F.v - 1)*(alpha.v^-G.v - 1))/(alpha.v - 1)), alpha.v)
      return(ce)
    }
  }
  if (copula.name == "gumbel.hougaard"){
    copula.eqn <- function(F.v, G.v, alpha.v){
      ce <- exp(-((-log(F.v))-alpha.v+(-log(G.v))-alpha.v) - 1/alpha.v))
      return(ce)
    }
  }
  # Initialize the difference tolerance and the maximum number of
  # iterations
  F.diff.tol <- 0.0001
  max.iter <- 1000
  # Initialize difference value and iteration counter
  F.diff <- 99
  iter <- 0
  # Start with a previous guess for F.val; make it the value from the
  # previous iteration of g.cg.est
  F.minus1 <- F.valg
  # Evaluate the function at the previous guess for G.val
  fn.F.minus1 <- 1 - F.minus1 - G.valg +
    copula.eqn(F.minus1, G.valg, alpha.val) - k.val
  # Come up with an initial guess for F.val
  F.guess <- F.minus1 + 0.01
# Enter loop; set loop to break if F.diff is less than the
tolerance or iterations greater than max number allowable
while (F.diff > F.diff.tol & iter < max.iter){

    # Evaluate function at current guess for F.val
    fn.F.guess <- 1 - F.guess - G.val +
               copula.eqn(F.guess, G.valg, alpha.val) - k.val

    # Evaluate derivative of function at current guess for F.val
    deriv.F.guess <- (fn.F.guess - fn.F.minus1)/(F.guess - F.minus1)

    # Calculate next guess for F.val
    F.plus1 <- F.guess - fn.F.guess/deriv.F.guess

    # Evaluate difference between current guess and next guess
    F.diff <- abs(F.plus1 - F.guess)

    # Update values for next iteration
    F.minus1 <- F.guess                   # Current guess becomes previous guess
    F.guess <- F.plus1                   # Next guess becomes current guess
    fn.F.minus1 <- fn.F.guess            # Function evaluated at current guess
                                           # becomes function evaluated at
                                           # previous guess

    # Count iterations
    iter <- iter + 1
}

# If algorithm converged, output number of iterations and new F value.
ret.vec <- c(F.plus1)
if (iter < max.iter & F.plus1 > F.valg) return(ret.vec)

# If algorithm converged, but new F value less than previous F value,
# something is wrong.
if (iter < max.iter & F.plus1 <= F.valg)
    return("Algorithm converged to non-valid F.
            Check F,G, and K vectors.")

# If algorithm did not converge, notify user.
if (iter == max.iter) return("Algorithm did not converge.")

The function nm.copula.for.y similar to the above.
Appendix C

S-Plus Code for Copula Simulation

copula.sim <- function(n,copula.type,param){

# For independence copula
if (copula.type == "independence"){
  u1 <- runif(n,0,1)
  u2 <- runif(n,0,1)
}

# For gamma frailty copula (also see Devroye (1986) p. 600)
if (copula.type == "gamma"){
  # Generate gamma r.v. w/ shape parameter alpha and rate
  # parameter alpha
  frailties <- rgamma(n, shape=param, rate=param)

  # Now the marginal distributions of x1 and x2 conditional
  # on W are exponential with mean equal to the inverse of
  # the gamma variable generated above
  x1 <- rexp(n, rate=frailties)
  x2 <- rexp(n, rate=frailties)

  # Transform x1 and x2 into uniform (0,1) random variables
  # using probability integral transform U = F(X). We know
  # that x1 and x2 are marginally Pareto distributed; their
  # cdf is given in Eqn (14)
  u1 <- 1 - (1+x1/param)^(-param)
  u2 <- 1 - (1+x2/param)^(-param)
}

# For Gumbel-Houggaard copula (Frees and Valdez (1997) p. 12)
if (copula.type == "gumbel.hougaard"){

159
# Step 1: Generate a latent random variable, gamma, having Laplace transform \( \tau = \exp(-s^{-1}(1/a)) \). This is the Laplace transform corresponding to the positive stable distribution. Use Chambers et al. (1976) p. 341 to simulate from the positive stable distribution.

```r
alph <- 1/param

theta <- runif(n, 0, pi)
a.top <- sin((1-alph)*theta) * ((sin(alph*theta))^(alph/(1-alph)))
a.bottom <- (sin(theta))^(1/(1-alph))
a.exp <- a.top/a.bottom

W <- rexp(n,1)
gam <- (a.exp/W)^(((1-alph)/alph))
```

# Step 2: Independently of Step 1, generate an \( n \times 2 \) matrix of uniform r.v.

```r
U.mat <- matrix(runif(2*n,0,1), ncol=2)
```

# Step 3: For \( k = 1,2 \), calculate \( u_1 \) and \( u_2 \) vectors

```r
u1 <- exp(-(-(1/gam)*log(U.mat[,1]))^((1/param)))
u2 <- exp(-(-(1/gam)*log(U.mat[,2]))^((1/param)))
```}

# For Frank's copula (from Nelsen (1986) p. 3282)

```r
if (copula.type == "frank"){
  u1 <- runif(n,0,1)
  v <- runif(n,0,1)
  top.fun <- v*(param - 1)
  bottom.fun <- (param~u1)*(1 - v) + v
  u2 <- logb(1+top.fun/bottom.fun, param)
}
```

```r
par(mfrow=c(1,1))
plot(u1, u2, xlim=c(0,1), ylim=c(0,1), main="Simulated Copula Data")
```

```r
par(mfrow=c(1,2))
hist(u1)
hist(u2)
```

```r
u.mat <- cbind(u1,u2)
```

```r
return(u.mat)
```
Appendix D

Deriving the Gamma Frailty Copula

This derivation was aided by an example found in Frees and Valdez (1998, p. 3-4).

Start with $X_1$ and $X_2$, defined to be independently distributed exponential random variables with mean 1. For exponential random variables, the hazard is constant and equal to the inverse of the mean. Therefore, for $X_1$ and $X_2$,

$$\lambda_i(t) = 1 \quad i = 1, 2 \quad \text{(D.1)}$$

Now, introduce a frailty parameter, $W$, that is gamma($\alpha, \beta$) distributed and acts multiplicatively upon the hazards. The conditional hazard functions have the following form:

$$\lambda_i(t|w) = 1 \times w \quad i = 1, 2 \quad \text{(D.2)}$$

Derive the conditional survival functions from the conditional hazard functions.

$$S_i(t|w) = \exp\{-\int_0^t \lambda_i(t|w) \, dt\}$$

$$= \exp\{-\int_0^t w \, dt\}$$

$$= e^{-wt} \quad i = 1, 2 \quad \text{(D.3)}$$
Since $F = 1 - S$, the conditional distribution functions are given by

$$F_i(t|w) = 1 - e^{-wt} \quad i = 1, 2. \quad (D.4)$$

This indicates the conditional distributions of $X_1$ and $X_2$ are exponential with mean $1/w$. Integrate out the $w$ in Eqn. (D.4) to get the unconditional distribution functions.

$$F_i(t) = \int_0^\infty F_i(t|w)f_W(w)dw \quad (D.5)$$

$$= \int_0^\infty (1 - e^{-wt}) \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} w^{\alpha-1} e^{-w/\beta} dw \quad (D.6)$$

$$= 1 - \int_0^\infty e^{-wt} \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} w^{\alpha-1} e^{-w/\beta} dw \quad (D.7)$$

$$= 1 - \int_0^\infty \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} w^{\alpha-1} e^{-w(1/\beta)} dw \quad (D.8)$$

$$= 1 - \int_0^\infty \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} w^{\alpha-1} e^{-w(t+1/\beta)} dw \quad (D.9)$$

$$= 1 - \int_0^\infty \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} w^{\alpha-1} e^{-w/\beta(t+1)} dw \quad (D.10)$$

To complete the integration, note that the integrand looks very close to a gamma($\alpha, \frac{\beta}{\beta(t+1)}$). Multiply and divide by Eqn (D.11) by $(\beta t + 1)^\alpha$.

$$F_i(t) = 1 - \frac{1}{(\beta t + 1)^\alpha} \int_0^\infty \frac{1}{\Gamma(\alpha)} \left( \frac{\beta t + 1}{\beta} \right)^\alpha w^{\alpha-1} e^{-w/\beta(t+1)} dw \quad (D.12)$$

$$= 1 - \frac{1}{(\beta t + 1)^\alpha} \quad (D.13)$$

$$= 1 - (1 + \beta t)^{-\alpha} \quad (D.14)$$

This is the cdf of the Pareto distribution.

Conditional on $w$, $X_1$ and $X_2$ are i.i.d. Now we will derive their joint distribution, $F(x_1, x_2)$. First, let us recall a basic result:

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) \quad (D.15)$$

$$= 1 - P(X_1 > x_1) - P(X_2 > x_2) + P(X_1 > x_1, X_2 > x_2) \quad (D.16)$$

$$= 1 - [1 - F_1(t)] - [1 - F_2(t)] + P(X_1 > x_1, X_2 > x_2) \quad (D.17)$$
We have derived all the pieces of Eqn (D.17) except for the last term, $P(X_1 > x_1, X_2 > x_2)$.

\[
P(X_1 > x_1, X_2 > x_2) = \int_{0}^{\infty} P(X_1 > x_1|w)P(X_2 > x_2|w)f_w(w)dw \tag{D.18}
\]

\[
= \int_{0}^{\infty} e^{-wx_1}e^{-wx_2} \frac{1}{\Gamma(\alpha)\beta^\alpha} w^{\alpha-1}e^{-w/\beta}dw \tag{D.19}
\]

\[
= \int_{0}^{\infty} \frac{1}{\Gamma(\alpha)\beta^\alpha} w^{\alpha-1}e^{-w(x_1+x_2+1/\beta)}dw \tag{D.20}
\]

Use the same trick as before. This time, the integrand is close to that of a gamma($\alpha, \frac{\beta}{\beta x_1 + \beta x_2 + 1}$). Multiply and divide by Eqn (D.20) by $(\beta x_1 + \beta x_2 + 1)^\alpha$.

\[
P(X_1 > x_1, X_2 > x_2) = \frac{1}{(\beta x_1 + \beta x_2 + 1)^\alpha} \times \int_{0}^{\infty} \frac{1}{\Gamma(\alpha)} \left(\frac{\beta x_1 + \beta x_2 + 1}{\beta}\right)^\alpha w^{\alpha-1} \times e^{-w(\beta x_1 + \beta x_2 + 1)}dw \tag{D.21}
\]

\[
= (1 + \beta x_1 + \beta x_2)^{-\alpha} \tag{D.22}
\]

Now we can plug into Eqn (D.17) and solve for the joint distribution function.

\[
F(x_1, x_2) = 1 - (1 + \beta x_1)^{-\alpha} - (1 + \beta x_2)^{-\alpha} + (1 + \beta x_1 + \beta x_2)^{-\alpha} \tag{D.23}
\]

We can simplify Eqn (D.25) by making some substitutions. Recall that $F_1(x_1) = 1 - (1 + \beta x_1)^{-\alpha}$ and $-(1 + \beta x_2)^{-\alpha} = F_2(x_2) - 1$. We have

\[
F(x_1, x_2) = F_1(x_1) + F_2(x_2) - 1 + [(1 - F_1(x_1))^{-1/\alpha} + (1 - F_2(x_2))^{-1/\alpha} - 1]^{-\alpha}. \tag{D.24}
\]

This yields the copula function

\[
C(u_1, u_2) = u_1 + u_2 - 1 + \left[\left(\frac{1}{1-u_1}\right)^{1/\alpha} + \left(\frac{1}{1-u_2}\right)^{1/\alpha} - 1\right]^{-\alpha}. \tag{D.25}
\]
Bibliography


