Survival Analysis with Artificially Constructed Events

by Sanjiv Jaggia, Professor, Suffolk University, Department of Economics, 8 Ash­
burton Place, Boston, MA 02108, USA and Satish Thosar, Associate Professor, Uni­
versity of Technology, School of Finance and Economics, PO Box 123, Sydney,
Broadway, NSW 2007, Australia

Abstract

One of the key elements of survival models is that they enable the researcher to de­
termine whether the length of time an individual (or economic entity) spends in a par­
ticular state affects the probability of exiting that state. Natural applications in
ecomics and finance include the analysis of unemployment spells, corporate
bankruptcies and mortgage pre-payments. The distinguishing feature of most appli­
cations is the definitive event that marks the transition from the origin to the transition
state. We believe that limiting the use of survival analysis to applications in which the
event duration appears to be 'naturally' available is an unnecessary constraint. For
example, the date of emergence from Chapter 11 bankruptcy protection is a subjec­
tive management decision and the true event duration, though treated as definitive,
is in reality quite ambiguous. We propose that survival models can and should be ex­
tended to analyze researcher-defined events such as the length of time a stock
takes to reach a pre-set price target. We illustrate our point with an examination of
IPO aftermarket behavior.

I. Introduction

The analysis of duration data via survival models was pioneered in the medical and
engineering fields. The variable of interest (dependent variable) is the duration of
time that precedes an event; in particular, the length of time an entity (say a medical
patient or machine part) spends in a given state. Survival models have been used to
explain the differences in event duration. The research interest lies in identifying the
covariates (independent variables) that explain the duration of time that elapses be­
tween the origin state and the transition state as well as the instantaneous probabil­
ity of exit over time. It is important to determine if this probability of exit increases or
decreases with the elapsed duration in the origin state. The distinguishing feature of
most survival model applications is the definitive event that marks the transition from
the origin to the transition state. In this paper we argue that this is largely a self­imposed constraint and that many interesting applications can and should be stud­
tied where the event is user-defined (or artificially constructed).

The extension of the survival model methodology to a number of applications
in the social sciences has been a natural progression. For instance, in sociology,
survival analysis has been used to examine duration of marriages. When a couple
marries, they enter the origin state (marriage) and the marriage proceeds along a
time path. Over time the marriage either survives or terminates with a divorce. The
data are censored when all we know is that the marriage survived up to a given point,
for example when one spouse dies. The divorce outcome is itself an event that signi­
fies the transition between the well-defined states of 'married' and 'not married'. An
early application in economics was in the study of unemployment spells. Using sur­vival models, researchers could establish the impact of elapsed time in the unem­
ployment spell on the probability of re-entering the workforce. Other applications include the analysis of strikes, bankruptcies, and mortgage pre-payments.\textsuperscript{2}

A key element seems to be that the choice of application is driven largely by the accessibility of measurement of the event duration. It is obviously easier, and perhaps expedient, to model events where the origin and exit dates can be identified definitively as in the case of marriage/divorce. In our opinion, this is an unnecessary and largely self-imposed constraint by researchers. We believe that there are many interesting economic phenomena which may not offer up a ready-made duration measure but which are still amenable to analysis using survival models. For instance, Jaggia and Thosar (2004), henceforth JT, employ a hazard model to study momentum and reversal effects in the IPO aftermarket. The event is artificially constructed by the researcher— in this case the amount of time the IPO stock takes to reach a pre-specified price target. The survival model is used to ascertain if the IPO aftermarket exhibits momentum and reversal patterns (duration dependence) and if the chosen covariate affects the probability of hitting the target.

This kind of artificial event construction could be criticized on the ground that it could be misused to implement a sort of data mining. Our view is that virtually all exploratory research (such as the capital market anomalies literature) is subject to the same potential criticism. Furthermore, there is already an element of artificiality in the measurement of duration in the more common finance applications. Unlike a patient’s death or a machine failure, the date on which a company files for or emerges from bankruptcy protection is a subjective management decision and the true event duration though treated as precise is in reality quite ambiguous. This is also true for the marriage example cited earlier. The couple is treated as having entered the ‘married’ state on the date that the marriage certificate is issued and the ‘divorced’ state on the date on which the legal papers are signed. In reality, the couple may have entered into a marriage-like arrangement long before becoming formally married and the marriage likely failed (or terminated) long before the legally determined divorce date. Therefore, the marriage duration as sampled by the researcher is also subject to potentially valid criticism.

In this paper, we use the JT study to illustrate our point and propose that survival models can and should be used to study phenomena whose durations are defined by the researcher. They may be especially useful in high uncertainty (volatile) environments such as market behavior following IPOs, takeover announcements, etc. In the next section, we describe the research objectives, the sample, the way in which the dependent variable is constructed and the explanatory variables used in the analysis. Section 3 contains a detailed methodological review of parametric duration models focusing on the Weibull and Weibull-Gamma specifications. Section 4 contains a brief discussion of our results. The final section contains concluding comments.

\textbf{II. Research Objectives and Data Description}

The motivation for the JT study was the general observation that in the late 1990s, high-tech IPOs appeared to be severely underpriced and experienced considerable positive momentum in the medium-term aftermarket. The research objectives were: (a) to capture the momentum and reversal patterns of the aftermarket in a rigorous
framework and (b) to ascertain whether ex-ante technical and fundamental variables had predictive power in terms of the probability of hitting researcher specified price targets in the IPO aftermarket. The duration event was defined when the cumulative market-adjusted return (CMAR) crossed the researcher determined target of 25, 50, or 100 percent.

The primary sample consists of 301 high-tech IPOs launched in the USA in the late 1990s. Aftermarket daily open price data is collected for each firm and the NASDAQ for 125 trading days post-IPO. Accordingly, the event is defined when the IPO stock reaches a pre-specified price target. More precisely, let $P_{i1}$ represent the Day 1 open price of the $i^{th}$ firm and let $P_{m1}$ be the corresponding level of the market (NASDAQ) index. Similarly, $P_{it}$ and $P_{mt}$ represent the open price at time $t$ of the $i^{th}$ firm and the market respectively. The CMAR of the firm at time $t$ is calculated as

$$CMAR_t = \frac{P_{i1}}{P_{m1}} - 1$$

(1)

The time in question does not refer to calendar time, but to the time from the IPO date. We use 25, 50, and 100 percent as CMAR thresholds; for instance, an event occurs when the CMAR exceeds 50 percent and we record the time in weeks from day 1 till the date on which the threshold is reached. This event duration becomes the dependent variable in the survival model.

The explanatory variables are described below:

**Technical variables**

- *Percentage Price Change*: Initial underpricing (or overpricing variable) that measures the change from the offer price to the open price. Expected sign: negative; the logic being that if there is a large opening positive spike, flipping activity will dampen the aftermarket.

- *Adjusted Return in Week 1*: Momentum (purely technical) variable representing the average market-adjusted return in the first week of trading. Expected sign: positive.

- *Market Return In Week 1*: Momentum (purely technical) variable representing the market return after the first week of trading. Expected sign: positive.

**Fundamental and IPO contract-specific variables**


- *Offer Size*: Offer price multiplied by the number of shares sold in the IPO. We use the log value of the offer size. Expected sign: No prior; included as a size proxy and control variable.

- *Age before the IPO*: Represents the number of years from the date the firm was incorporated to the IPO date. Expected sign: positive.
**Underwriter Reputation:** Lead underwriter’s reputation variable using the Carter-Manaster measure. (The dummy variable takes value 1 if the CM measure exceeds 7; 0 otherwise). Expected sign: negative; the reasoning is that highly ranked underwriters will be associated with less underpricing and lower aftermarket volatility.

**Green Shoe Dummy:** Dummy variable for a green shoe provision in the IPO contract. This provision gives the underwriter the option to purchase additional shares at the offer price to cover over allotments. Expected sign: negative; this option facilitates aftermarket stabilization by underwriters.

**Internet Services and Software Dummy:** Dummy variable takes value 1 if the firm belongs either to the Internet services or Internet software sectors. Expected sign: positive; the reasoning is that Internet firms have greater uncertainty about their fundamentals and are more susceptible to irrational trading.

Table 1 contains some descriptive statistics of the explanatory variables. Each variable is ordered and split up into four quartiles. Means are calculated for all observations and for each of the quartiles. It is worth noting that the initial underpricing and the momentum variables are remarkably different across quartile segments.

### III. Methodology

Survival models are often used to explain the duration of an event of interest. Their use is facilitated by the availability of various statistical packages that routinely estimate these models. The survival model methodology allows us to analyze the influence of individual characteristics as well as the elapsed duration on the probability of an event. In this paper we use continuous time parametric models to study how long it takes for a speculative investor to reach a desired threshold return. We study the influence of ‘fundamental’ as well as ‘technical’ characteristics that affect the duration. Further, we are able to document if the probability of reaching a desired threshold changes with the length of time the investor stays invested.

Let $t$ represent the observed duration of a completed event, which is the realization of a continuous random variable, $T$. In continuous time models, it is customary to specify the parametric form of the model in terms of its hazard function rather than the probability density function even though the two functions are mathematically equivalent. The following are the hazard, survivor and the density functions conditional on the vector of personal characteristics, $X$:

$$h(t;X) = \frac{f(t;X)}{S(t;X)}; S(t;X) = P(T > t); f(t;X) = \frac{-\partial S(t;X)}{\partial t}$$  \hspace{1cm} (2)

$f(t;X)dt$ is the unconditional probability that the event concludes in a small interval around $t$ and $S(t;X)$ is the probability that the duration exceeds $t$. The hazard represents the instantaneous probability that an event concludes at $t$, given that it had not already done so. In the context of this application, the hazard rate is the instantaneous probability that a given threshold is crossed.
For parametric estimation, we must specify \textit{a priori} the functional form of the hazard function. A commonly used duration distribution is Weibull with hazard, \( h(t; X) = \mu \alpha t^{\alpha - 1} \). The scale parameter is usually specified as \( \mu = \exp(X\beta) \) and the shape parameter \( \alpha \) allows for positive (\( \alpha > 1 \)), negative (\( \alpha < 1 \)) or the absence (\( \alpha = 1 \)) of duration dependence. The survivor function is \( S(t; X) = \exp(-\mu t^\alpha) \).

Duration data are often 'right' censored due to a finite data acquisition period. In our sample, we follow each IPO firm for 125 trading days (25 weeks) and the observation is censored if the desired threshold is not reached by the end of this period. The sample consists of N observed durations on \( t_i \) along with a censoring variable \( C_i \) that equals 1 if the threshold is reached and 0 if not. The following log-likelihood function incorporates all observations, where right censored spells contribute \( S(t; X) \) rather than \( f(t; X) \) to the function:

\[
L = \sum_{i=1}^{N} C_i \ln f(t_i; X_i) + (1 - C_i) \ln S(t_i; X_i) = \sum_{i=1}^{N} C_i \ln h(t_i; X_i) + \ln S(t_i; X_i) \tag{3}
\]

It is preferable to specify the likelihood function in terms of log durations since it greatly facilitates its maximum likelihood estimation. For a Weibull model, with \( S(t; X) = \exp(-\exp(X\beta + \alpha \ln t)) \) and \( h(t; X) = \alpha \exp(X\beta + \alpha \ln t) \), the likelihood is specified as:

\[
L = \sum_{i=1}^{N} C_i (\ln \alpha + X_i \beta + \alpha \ln t_i) - \exp(X_i \beta + \alpha \ln t_i) \tag{4}
\]

It is well documented that the maximum likelihood estimates may lead to spurious inferences if the model is incorrectly specified (see MacKinnon (1992)). It is, therefore, crucial to test the validity of the given parametric model before any meaningful inferences can be made. Two common sources of misspecification in the context of a survival model are (a) incorrect functional form of the hazard function and (b) neglected heterogeneity. Further, unlike linear regression models, there is no natural way of defining a residual for diagnostic testing of survival models. Tests in survival models are generally based on the generalized errors, \( \varepsilon = -\ln S(t; X) \). Under the null hypothesis of no model misspecification, \( \varepsilon \) has a unit exponential distribution with \( E(\varepsilon^1) = 1 \). The parameters are replaced by their maximum likelihood estimates to obtain generalized residuals, which are used for model testing. For a Weibull model,

\[
\hat{\varepsilon} = \exp(X\beta + \hat{\alpha} \ln t)
\tag{5}
\]

The conditional moment restriction tests based on \( E(\varepsilon^1) = 1 \) are easy to implement (see Jaggia (1991b)). For a Weibull model, a test of second moment restriction, \( E(\varepsilon^2) = 2 \), has traditionally been interpreted as a test for neglected heterogeneity, caused by the omission of relevant factors (see Lancaster (1985)). Jaggia (1991a) shows that the outcome of this test is valid only under a very restrictive alternative and leads to distorted results when heterogeneity is not the only source of misspecification. Therefore, as in Jaggia and Thosar (1995), we interpret the second moment restriction test as a general misspecification ("something is wrong") test rather than a test for heterogeneity per se.
Given that some of our observations are right censored, the above residuals are suitably modified as \( \hat{e} = \hat{e} + 1 - \hat{C} \). Although the modified generalized errors do not have a unit exponential distribution, they still have a unit mean with variance = \( \pi^* \) where \( \pi^* \) is the expected probability of censoring (see Lancaster and Chesher (1985)). We conduct a test that examines if the sample variance of the modified residuals equals the sample proportion of censored observations. The test statistic has an easily implementable form when the BHHH estimator is used for the information matrix. An OLS is run where the left hand side variable is unity and the right hand side consists of \((\hat{e}_i - 1)^2 - C_i\) and \([\partial I_i / \partial \theta_j]\) where \( L = \sum I_i \) and \( \theta_1, \theta_2, ..., \theta_k \) are the parameters of the model. The test is computed as \( \chi^2 \) (uncentered) and has an asymptotic \( \chi^2 \) distribution under the null.

If the Weibull specification is rejected, an appropriate formulation is \( \mu = \nu \exp(X\beta) \) where \( \nu \) is a positive continuous random variable that accounts for neglected heterogeneity or more generally some intrinsic randomness in the model. As \( \nu \) is not observable, the unconditional survivor function is:

\[
S(t;X) = \int_0^\infty e^{-\nu \exp(X\beta)} f(\nu) d\nu
\]

Once the mixing distribution \( f(\nu) \) is specified, the above function is computed to estimate the parameters of the model. Following the popular convention, we use the gamma distribution as a convenient mixing distribution. With a unit mean and variance equal to \( \sigma^2 \),

\[
S(t;X) = \left[ 1 + \sigma^2 \exp(X\beta)^{1/\sigma} \right]^{-\sigma^2}
\]

Note that the above function collapses to the Weibull survivor function for \( \sigma^2 \rightarrow 0 \). The corresponding hazard and the density functions are derived similarly. As in Jaggia and Thosar (1995), we suggest that the mixing distribution is used not only to compensate for omitted factors but also to correct for an overly restrictive Weibull hazard function. The log-likelihood function specified in terms of log-durations, with \( w_i = x_i \beta + \alpha \ln t_i \) is:

\[
L = \sum_{i=1}^N C_i \left( \ln \alpha + w_i - \ln(1 + \sigma^2 \exp(w_i)) - \frac{1}{\sigma^2} \ln(1 + \sigma^2 \exp(w_i)) \right)
\]

For the second moment restriction test, the generalized residuals for the Weibull-gamma model are:

\[
\hat{e} = \frac{1}{\sigma^2} \ln(1 + \sigma^2 \exp(\chi\beta + \alpha \ln t))
\]

IV. Results

We start with a brief analysis of the non-parametric estimate of raw hazard for different desired thresholds (see Figure 1). The Kaplan-Meier or product-limit estimator at a given time is calculated as the number of 'exits' divided by the 'number at risk' (see Lancaster 1990)). In other words, we calculate the number of firms for which the threshold target is reached at a given point in time divided by the total number of
firms for which it had not reached up to that point in time. It is worth mentioning that in
general, the accuracy of this estimator is better for shorter durations as inferences
about longer durations are based on fewer observations. In Figures 2 and 3, we
present the Weibull and Weibull-gamma (WG) hazards respectively for comparison.
It should be borne in mind that the raw hazard does not allow for the heterogeneous
nature of durations whereas the estimated parametric models are based on average
factor values. Although not directly comparable, it appears that the basic non-
monotonic pattern of the estimated WG model resembles the raw hazard fairly well.

The results of the estimated Weibull and WG models are presented in Tables 2
and 3 respectively. The $\chi^2$ test statistics for the second moment restriction test are
presented in the last row of the tables. The Weibull model is clearly inappropriate
since the null hypothesis of no misspecification is rejected for all models. On the
other hand the WG model passes the test for the 25 and 100 percent threshold mod-
els. Further, we also compute the likelihood ratio (LR) tests since the Weibull model
is nested in the WG model as $\sigma^2 \rightarrow 0$. The LR, computed as two times the difference
between the maximized log-likelihood values of the Weibull and the WG models,
has an asymptotic $\chi^2(1)$ distribution under the Weibull specification. The sample val-
ues of the test statistics are 103.91, 51.30, and 43.53 for the 25, 50, and 100 percent
threshold models respectively, which corroborate the finding that the WG model is
appropriate. Since both the visual and formal specification tests indicate that the WG
model represents the better fit, we will limit our discussion to the WG estimation re-
sults.

The most important finding in the context of this study is that both shape pa-
rameters ($\alpha$ and $\sigma^2$) are strongly significant for all three thresholds. The estimated
hazard functions, evaluated at mean explanatory variables, for the three thresholds
(Figure 3) are revealing. Their inverted U-shaped pattern indicates that the instanta-
neous probability of crossing the defined return threshold rises sharply at the outset,
peaks and then gradually wanes. This indicates the presence of duration depen-
dence (non-constant hazard) and points to initial momentum and gradual reversal in
the aftermarket CMAR. The probability of reaching the investor determined price tar-
gets increases for the first few weeks (momentum) before declining (reversal). It is
worth noting that the hazard for the lowest threshold (25 percent) peaks earlier and
at a higher level than the more aggressive 50 and 100 percent thresholds. This
makes intuitive sense because it is reasonable to assume that, from a speculative
investor's point of view, the probability of reaching a modest threshold (25 percent) is
higher (and more imminent) than a more demanding target.

From Table 3, we find that the strongest predictors of aftermarket momentum-
reversal are the purely technical variables. For instance, if the stock's CMAR in the
first week after going public is strongly positive this improves the probability (or
equivalently shortens the expected duration) of reaching the desired threshold. Fun-
damental variables such as pre-IPO profitability or age of the IPO firm have little or
no effect. The lead underwriter's reputation (contrary to our expectation) and design-
nation as an Internet firm have positive influences.

In linear regression models, the coefficients $\beta$ measure the marginal effect of
the explanatory variable on the dependent variable. The interpretation of $\beta$ in the
hazard (non-linear) model is not straightforward. Most empirical studies in survival models report only the significance of various explanatory variables, ignoring the more important part of interpreting the coefficients. We think that a good way to get a feel for the economic significance of individual regressors is to carry out a form of sensitivity analysis. For instance, we can assess the survival probability, $P(T > t)$, for the various values of a given variable. It should be mentioned that any factor that increases the hazard would decrease the survival probability.

In Table 4, we present the survival probabilities at week 5 for a given variable evaluated at the 1st and 4th quartile means respectively, while all other variables are at their overall means (see Table 1 for overall and quartile means). These probabilities are computed from (7) above with $\beta, a, \sigma^2$ replaced by their respective parameter estimates and with t set at five weeks. We find, for instance, if the firm’s CMAR in the first week post-IPO is -22.48 percent (1st quartile mean) the survival probability at week 5 is 0.501 for reaching the 25 percent threshold and 0.974 for the 100 percent threshold. In contrast if the first week’s CMAR is 34.85 percent (4th quartile mean) the survival probabilities for the 25 and 100 percent thresholds are 0.000 and 0.171 respectively. The impact of the other variables is analyzed similarly.

V. Conclusion

Our objective in this paper has been to make a case for extending a potentially powerful methodology (survival modeling) to applications in the fields of accounting, finance and economics. Two commonly used parametric specifications, namely the Weibull and the Weibull-gamma models, are described. We also highlight the consequences of a misspecified parametric model and provide detailed information to implement some important diagnostic tests.

Most survival model applications are constrained by the fact that they are based on the definitive event that marks the transition from the origin to the transition state. We argue that survival models can and should be extended to analyze researcher-defined events. We provide an example of a researcher (or practitioner) defined event like the amount of time a stock takes to reach a pre-set price target. The use of survival modeling using artificially defined events in this setting has yielded useful insights into the dynamics of IPO aftermarket behavior. The general results are of interest to academics and possibly regulators. Also, duration dependence implied by the shape of the estimated hazard function graphs (Figure 3) and the sensitivity analysis yielding survival probabilities (Table 4) could be useful analytical tools for investment practitioners.

Acknowledgement

We are grateful to Marc LeClere for his detailed comments. Any remaining errors are ours alone.
Table 1

Overall and Quartile Segment Means

<table>
<thead>
<tr>
<th>Variables</th>
<th>All</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Price Change</td>
<td>61.92</td>
<td>-1.75</td>
<td>20.04</td>
<td>51.01</td>
<td>167.66</td>
</tr>
<tr>
<td>Adjusted Return in Week 1</td>
<td>2.63</td>
<td>-22.48</td>
<td>-7.42</td>
<td>3.78</td>
<td>34.85</td>
</tr>
<tr>
<td>Market Return in Week 1</td>
<td>0.15</td>
<td>-2.86</td>
<td>-0.72</td>
<td>0.83</td>
<td>3.28</td>
</tr>
<tr>
<td>Net Income/Revenue</td>
<td>-1.77</td>
<td>-6.24</td>
<td>-1.10</td>
<td>-0.31</td>
<td>0.42</td>
</tr>
<tr>
<td>Offer Size</td>
<td>17.94</td>
<td>16.96</td>
<td>17.71</td>
<td>18.07</td>
<td>18.96</td>
</tr>
<tr>
<td>Age before the IPO (in years)</td>
<td>5.33</td>
<td>1.31</td>
<td>3.15</td>
<td>4.72</td>
<td>11.65</td>
</tr>
<tr>
<td>Underwriter Reputation</td>
<td>0.80</td>
<td>0.20</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Green Shoe Dummy</td>
<td>0.57</td>
<td>0.00</td>
<td>0.29</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Internet Dummy</td>
<td>0.60</td>
<td>0.00</td>
<td>0.41</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

For each variable, observations are ordered and split up into four equal quartile segments. Means are calculated for all observations and for each of the four segments, Q_i.
Table 2
Estimates of the Weibull hazard model for the 25, 50, and 100 percent thresholds

<table>
<thead>
<tr>
<th>Threshold</th>
<th>25%</th>
<th>50%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameters</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.839 (-0.532)</td>
<td>-1.777 (-0.994)</td>
<td>-1.899 (-0.804)</td>
</tr>
<tr>
<td>Percentage Price Change</td>
<td>-0.001 (-1.451)</td>
<td>-0.001 (-0.751)</td>
<td>-0.003* (-1.816)</td>
</tr>
<tr>
<td>Adjusted Return in Week 1</td>
<td>0.032** (13.281)</td>
<td>0.038** (12.892)</td>
<td>0.034** (9.522)</td>
</tr>
<tr>
<td>Market Return in Week 1</td>
<td>0.053* (1.698)</td>
<td>0.107** (3.011)</td>
<td>0.050 (1.213)</td>
</tr>
<tr>
<td>Net Income/Revenue</td>
<td>-0.003 (-0.127)</td>
<td>0.050* (1.811)</td>
<td>-0.013 (-0.409)</td>
</tr>
<tr>
<td>Offer Size</td>
<td>-0.107 (-1.172)</td>
<td>-0.106 (-1.033)</td>
<td>-0.155 (-1.145)</td>
</tr>
<tr>
<td>Age before the IPO (in years)</td>
<td>-0.006 (-0.425)</td>
<td>-0.006 (-0.334)</td>
<td>-0.006 (-0.246)</td>
</tr>
<tr>
<td>Underwriter Reputation</td>
<td>0.721** (3.318)</td>
<td>0.488** (2.045)</td>
<td>0.728** (2.253)</td>
</tr>
<tr>
<td>Green Shoe Dummy</td>
<td>0.469** (3.016)</td>
<td>0.323* (1.881)</td>
<td>-0.105 (-0.497)</td>
</tr>
<tr>
<td>Internet Dummy</td>
<td>0.251 (1.612)</td>
<td>0.596** (3.419)</td>
<td>0.505** (2.375)</td>
</tr>
<tr>
<td>$\alpha$ (shape parameter)</td>
<td>0.716** (-6.935)</td>
<td>0.866** (-2.383)</td>
<td>0.965 (-0.462)</td>
</tr>
<tr>
<td>$\chi^2$ Value (Second Moment Restriction Test)</td>
<td>56.168**</td>
<td>33.396**</td>
<td>22.664**</td>
</tr>
</tbody>
</table>

** and * denote significance at 5 and 10 percent respectively. For $\alpha$, the t-statistic is evaluated at 1.
<table>
<thead>
<tr>
<th>Threshold</th>
<th>25%</th>
<th>50%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameters</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>-2.879 (-0.533)</td>
<td>-4.261 (-0.762)</td>
<td>1.561 (0.320)</td>
</tr>
<tr>
<td><strong>Percentage Price Change</strong></td>
<td>-0.005* (-1.838)</td>
<td>-0.003 (-1.171)</td>
<td>-0.006* (-1.956)</td>
</tr>
<tr>
<td><strong>Adjusted Return in Week 1</strong></td>
<td>0.154** (6.420)</td>
<td>0.125** (4.524)</td>
<td>0.091** (6.276)</td>
</tr>
<tr>
<td><strong>Market Return in Week 1</strong></td>
<td>0.175** (2.006)</td>
<td>0.270** (2.812)</td>
<td>0.168** (2.081)</td>
</tr>
<tr>
<td><strong>Net Income/Revenue</strong></td>
<td>-0.073 (-1.058)</td>
<td>0.110 (1.603)</td>
<td>-0.034 (-0.568)</td>
</tr>
<tr>
<td><strong>Offer Size</strong></td>
<td>-0.120 (-0.388)</td>
<td>-0.127 (-0.415)</td>
<td>-0.609** (-2.072)</td>
</tr>
<tr>
<td><strong>Age before the IPO (in years)</strong></td>
<td>-0.023 (-0.611)</td>
<td>0.005 (0.137)</td>
<td>0.023 (0.470)</td>
</tr>
<tr>
<td><strong>Underwriter Reputation</strong></td>
<td>1.369** (2.279)</td>
<td>0.478 (0.794)</td>
<td>1.748** (2.934)</td>
</tr>
<tr>
<td><strong>Green Shoe Dummy</strong></td>
<td>1.723** (3.588)</td>
<td>0.897* (1.667)</td>
<td>0.453 (1.058)</td>
</tr>
<tr>
<td><strong>Internet Dummy</strong></td>
<td>1.099** (2.318)</td>
<td>2.149** (3.235)</td>
<td>2.064** (3.749)</td>
</tr>
<tr>
<td><strong>α (shape parameter)</strong></td>
<td>2.676** (4.239)</td>
<td>2.427** (2.570)</td>
<td>2.200** (3.583)</td>
</tr>
<tr>
<td><strong>σ² (shape parameter)</strong></td>
<td>5.514** (4.691)</td>
<td>5.188** (2.672)</td>
<td>4.434** (3.369)</td>
</tr>
<tr>
<td><strong>χ² Value (Second Moment Restriction Test)</strong></td>
<td>0.114</td>
<td>7.292**</td>
<td>2.195</td>
</tr>
</tbody>
</table>

** and * denote significance at 5 and 10 percent respectively. For α and σ², the t-statistics are evaluated at 1 and 0 respectively.
<table>
<thead>
<tr>
<th>Threshold</th>
<th>25%</th>
<th>50%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameters</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage Price Change</td>
<td>0.015</td>
<td>0.195</td>
<td>0.725</td>
</tr>
<tr>
<td>(0.034)</td>
<td></td>
<td>(0.287)</td>
<td>(0.879)</td>
</tr>
<tr>
<td>Adjusted Return in Week 1</td>
<td>0.501</td>
<td>0.871</td>
<td>0.974</td>
</tr>
<tr>
<td>(0.000)</td>
<td></td>
<td>(0.005)</td>
<td>(0.171)</td>
</tr>
<tr>
<td>Market Return in Week 1</td>
<td>0.034</td>
<td>0.398</td>
<td>0.865</td>
</tr>
<tr>
<td>(0.012)</td>
<td></td>
<td>(0.112)</td>
<td>(0.696)</td>
</tr>
<tr>
<td>Net Income/Revenue</td>
<td>0.015</td>
<td>0.324</td>
<td>0.769</td>
</tr>
<tr>
<td>(0.024)</td>
<td></td>
<td>(0.187)</td>
<td>(0.806)</td>
</tr>
<tr>
<td>Offer Size</td>
<td>0.018</td>
<td>0.206</td>
<td>0.681</td>
</tr>
<tr>
<td>(0.023)</td>
<td></td>
<td>(0.250)</td>
<td>(0.878)</td>
</tr>
<tr>
<td>Age before the IPO (in years)</td>
<td>0.019</td>
<td>0.230</td>
<td>0.809</td>
</tr>
<tr>
<td>(0.024)</td>
<td></td>
<td>(0.221)</td>
<td>(0.770)</td>
</tr>
<tr>
<td>Underwriter Reputation</td>
<td>0.046</td>
<td>0.281</td>
<td>0.917</td>
</tr>
<tr>
<td>(0.016)</td>
<td></td>
<td>(0.211)</td>
<td>(0.732)</td>
</tr>
<tr>
<td>Green Shoe Dummy</td>
<td>0.054</td>
<td>0.329</td>
<td>0.834</td>
</tr>
<tr>
<td>(0.010)</td>
<td></td>
<td>(0.167)</td>
<td>(0.761)</td>
</tr>
<tr>
<td>Internet Dummy</td>
<td>0.039</td>
<td>0.518</td>
<td>0.931</td>
</tr>
<tr>
<td>(0.013)</td>
<td></td>
<td>(0.111)</td>
<td>(0.631)</td>
</tr>
</tbody>
</table>

Survival probabilities at week 5 are calculated at the 1st and 4th quartile means (see Table 1) of the indicated variable, along with the overall means of all other variables. With all variables at their mean value, survival probabilities for the 25, 50, and 100 percent thresholds are 0.021, 0.227, and 0.795 respectively.
Figure 1. Estimated Raw Hazard (Kaplan Meier Estimator)

Figure 2. Estimated Hazard with Weibull Model
Figure 3. Estimated Hazard with Weibull-Gamma Model

Trading Weeks

Estimated Hazard
Endnotes

1. Survival models are also commonly referred to as *duration* or *hazard* models.


4. A similar analysis can be made on the downside. For instance, failure can be defined when the CMAR goes below, say 90 percent.

5. For instance, LIMDEP and STATA provide various estimation options. Other packages, especially Gauss, SAS, and Matlab, allow the user to specify the likelihood function for easy estimation.


7. Similar analysis can also be performed with expected durations.
References


