CS-BIGS 1(1): 1-14 http://www.bentley.edu/csbigs/vol1-1/lan.pdf

Bright Lights, Big Dreams - A Case Study of Factors Relating to the Success of Broadway Shows

Lan Ma Nygren
Rider University, USA

Jeffrey S. Simonoff
New York University, USA

The Cox proportional hazards model technique is demonstrated in an analysis of the longevity of Broadway shows. The necessity of using the Cox proportional hazards model to identify the factors associated with show longevity becomes clear by comparison to an ordinary linear regression analysis. Emphases are given to the model interpretation, including residual analysis, tests for the proportional hazards assumption, and fitting the stratified proportional hazards model. The intended audience consists of all applied statisticians who may consider using the Cox proportional hazards model. A basic to intermediate level of statistical analysis background is needed to understand the case.

Background

Located in and around Times Square in New York City, the 39 theaters called "Broadway" draw millions of visitors each year from all over the world. From June 2005 through May 2006 (the 2005-2006 season), theatergoers bought 12 million tickets to the Great White Way's musicals and plays, yielding gross revenues of more than \$860 million. According to The League of American Theatres and Producers, 57% of the Broadway audience in the 2005-2006 season were tourists, with foreign tourists a vital portion. The Broadway theater, known especially for the musical (one of the few purely American art forms), has become a global phenomenon in the live entertainment industry.

The business of the Broadway theater industry is as exciting as what appears on the stage. Failure in the theater has always been much more common than success. For instance, more than half of the Broadway shows that opened in the three seasons from 1996-97 through 1998-99 closed after 100 or fewer performances (roughly three months), and only six shows, all of them musicals, ran for more than 800 performances. Large musicals can bring in as much as \$50,000 in revenues per performance, but investors in a loser can see their investment disappear before their eyes, to the tune of as much as \$10 million for a musical.

In the Broadway theater business, where attracting the largest possible theatergoing audience is critical (and can be thought of as the cornerstone of Broadway show management), it is crucial for theater owners, managers, and stage producers to understand what factors are associated with the success of their offerings. To this end, we conduct an empirical study to investigate the factors that relate to the success of Broadway shows.

Related Work

Although there has been an increasing amount of empirical research addressing the determinants of success in the film industry (De Vany, 2004), relatively little work has focused on investigating the drivers of Broadway show success. Reddy, Swaminathan, and Motley (1998) employed an ordinary linear regression to model the types of information that affect the longevity and success of Broadway shows based on data from the 1980-1982 seasons. They found that critic reviews (particularly, those in the *New York Times*), preopening advertising in the *New York Times*, show type, and timing of the opening of the show were significant predictors of the longevity (number of performances) of the show. In various empirical investigations of the motion picture industry, the following factors have been found to (sometimes) be associated with the success of a motion picture: the genre (action, comedy, etc.) of the film, the Motion Picture Association of America rating of the film, critical reviews, Academy Award nominations and wins, measures of "star power," whether or not the movie was a sequel, first week's revenue, and the budget for the film. See Simonoff and Ma (2003) for more complete references. A different emphasis is given in this study (compared to Simonoff and Ma, 2003) by focusing on demonstrating the application of the Cox proportional hazards model technique. New material includes comparison to an ordinary linear regression analysis, assessment of model adequacy both graphically and through the use of hypothesis tests, and fitting the stratified proportional hazards model.

Data

The following discussion is based on Simonoff and Ma (2003). The data constitute a census of shows eligible for the Tony Awards opening on Broadway for a 3-year period (the 1996-1997 through 1998-99 seasons), which are then followed until the end of the 1999-2000 season on May 3, 2000 (the traditional end of the season corresponds to the last day of eligibility of a show for the Tony Awards). The response variable of interest is the longevity of the show. Show longevity is chosen as the dependent variable because high production costs, combined with the limited seating capacity of a theater, suggest that the commercial success of a show depends on the length of time the show runs, which is measured as the total number of performances of the show. Indeed, in these data the correlations between the (logged) total revenue of the show and the (logged) total number of performances is very high (.943), so examination of the length of the show's run is effectively equivalent to examining revenues. Having said that, analyzing longevity rather than revenues has the advantage of avoiding issues related to the very different costs associated with different types of shows. According to The League of American Theatres and Producers, in the 2004-2005 season (for example), the average paid admission for musicals was $67.92, while that for plays was $60.79, and that for other types of shows was $41.26. Given the high production costs for a musical, higher ticket prices have to be charged to generate more revenue, but this higher revenue does not necessarily translate into profitability (which would allow the show to remain open); it is the latter measure that is of interest, rather than the former.

An important feature of the response variable is that some of the observations are right-censored. To be more specific, by the end of the 1999-2000 season, seven of the shows had not closed. All that is known for these shows is that the number of performances is at least the observed value. Figure 1 gives a histogram of the total number of performances for the shows. The value for the seven shows that had not closed by May 3, 2000, are marked with an ×, so the total number of performances for those shows until closing is, in fact, larger than the recorded values. The long right tail in the histogram reflects the fact that roughly half of the shows closed after 100 or fewer performances (roughly three months), while six had more than 800 performances. If we think of the longevity of a show as its "survival time," the right-skewedness shown in Figure 1 is a typical pattern for lifetime data. There were a total of 95 shows opening from the 1996-1997 through 1998-1999 seasons. Among them, four shows ("Eugene Onegin," "Into the Whirlwind," "The Cherry Orchard," and one production of "The Three Sisters") were contractually limited to very few performances, and these are not included in the data analyses, resulting in data for 91 shows.

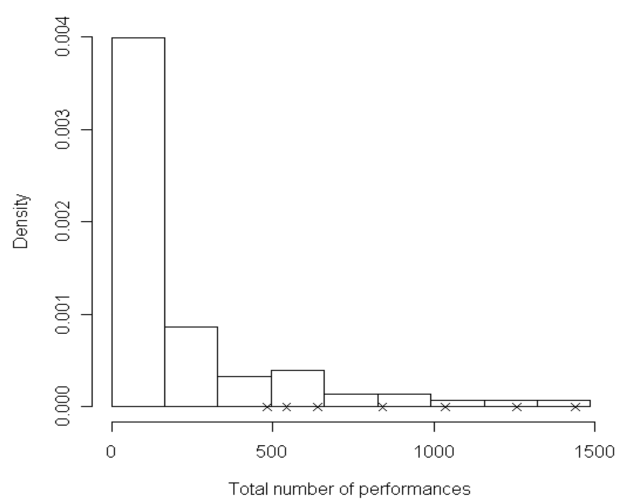


Figure 1. Histogram of observed total number of performances for each show. Numbers for shows that had not closed by May 3, 2000 are marked with the symbol ×.

To identify the possible predictors of a show's longevity, it is important to understand the nature of the product and the risks associated with paying to attend a Broadway show. The theater (together with movies and art) is an *experiential* good (one that people choose and use solely for the experience of pleasure). Since such a good would typically be limited to one (or few) experiences, and the monetary risk involved in the choice of a theatrical show is relatively high (the ticket price of a show on Broadway is $25 to $110), consumers seek to minimize risk by obtaining information from external sources. A potentially powerful source of such information is a critic's review. Critic reviews provide information about theatrical productions both experientially (expressing the feel of the theater-going experience) and objectively (information about the cast, plot, and genre of the show), and if the critic is viewed as unbiased, his or her opinions can be persuasive. Further, a critic with tastes similar to those of the potential audience would tend to predict success or failure well, even if the reviews did not have a direct causal effect on that success or failure. According to the *Demographics of the Broadway Audience 2005-2006* report prepared by the League of American Theatres and Producers, 28% of the Broadway show attendees identified critics' reviews as a major influence to see a show.

The predictors used to measure the impact of information sources are critic reviews from the *New York Times* and the *New York Daily News*. Since the reviews from these two newspapers do not give numerical ratings of the shows, three doctoral students interested in the Broadway theater served as judges to rate each review on a scale from 1 (poor evaluation) to 5 (high evaluation). To avoid bias, each judge read and rated each review independently and instructions on the criteria for each scale were provided to the judges. The correlations between the ratings of the three judges ranged from .829 to .872 for the *Times* and from .853 to .887 for the *Daily News*. The high interrater correlations suggest strong consensus among the judges. The ratings for each newspaper are therefore taken to be the average of the ratings for the three judges.

Besides information sources such as critic's reviews, objective features of a show (i.e. show type, timing of the opening, initial audience reaction) are also key factors that influence the success of a theatrical product. Show type is included as a potential predictor of show longevity for several reasons. Many studies have found genre to be predictive of success of movies. Musicals are considerably more expensive to produce than other shows, and such increased production values might appeal to consumers. Musicals also typically charge higher ticket prices than other types of shows, so a producer might keep a musical open longer to try to recover more of the initial fixed cost. Whether or not a show is a revival is also a potential predictor, since a show being revived has appeared on Broadway before, a potentially positive signal to consumers.

We anticipate that the timing of a theatrical opening could affect its success, since research on movies indicates that timing is a significant predictor of movie success (see, for example, Radas and Shugan 1998, Ravid 1999, and Simonoff and Sparrow 2000). The opening month of a show is included as a possible predictor to investigate the seasonal effects on show longevity.

Initial audience reaction to the show is measured using the percentage of available seats sold during the first week of the run, as reported by the League of American Theatres and Producers. First week's attendance is a good indicator of the effectiveness of preopening publicity, since most Broadway show tickets are bought in advance, and also provides a large pool of potential (amateur) critics to spread the (hopefully) good word about the show (the *Demographics of the Broadway Audience 2005-2006* report mentioned earlier noted that personal recommendation was the single strongest reported influence in choosing a show to attend).

The fact that Academy Award nominations and wins provide a boost to demand of movies is supported by both conventional wisdom and empirical evidence (see, for example, Simonoff and Sparrow 2000). The predictors attempting to capture the effect of awards in the context of Broadway shows are Antoinette Perry (Tony) Award nominations and wins. Here, we restrict ourselves to the major categories of best musical and best play (revival and nonrevival), best director (musical and play), leading actor and actress (musical and play), and featured actor and actress (musical and play), as these are the categories of greatest interest to the general public. Winning or not winning the most important of Broadway theater awards is a powerful information source that would be expected to play a significant role in generating awareness and influencing choice of theatrical shows, in the same way that critic reviews do this. Nominations and awards are also likely to be indicative of the overall talent level of the people involved in the show and, of course, of the quality of the show itself.

The nature of the Tony Awards—and their impact on Broadway productions—affects the way show success should be measured. Shows will often be kept open until Tony nominations are announced; troubled shows that are nominated benefit from related advertising, while those that are not nominated close quickly. Thus, the total number of performances of a show can be related to its opening date in a way that has nothing to do with audience approval of the show. A troubled show that opens in February or March stays open several weeks (and dozens of performances) longer than one that opens in April, simply because it has longer to wait before Tony nominations or awards are announced. For this reason, longevity will be defined here in three distinct ways: total number of performances from opening night, total number of performances after the announcement of Tony Award nominations, and total number of performances after the announcement of Tony Award winners. For each of these targets, the connection with seasonal effects will also be investigated.

Methods

To motivate the proportional hazards model approach to the analysis of the data, we first pursue a linear regression model to see what problems arise if we ignore the censoring in the target variable. Since the total number of performances is right skewed, as indicated in Figure 1, the natural logarithm of total number of performances will be used as the dependent variable. Table 1 summarizes the results of the least squares linear regression model fit based on the subset of the available predictors that are numerical, for illustrative purposes.

Table 1. Results of Least Squares Linear Regression Model Fit for Total Number of Performances

Variable	Coefficient	t	p
Constant	3.3405	7.40	<0.001
First-week attendance	0.0032	0.36	0.719
Second-week attendance	0.0104	1.17	0.246
Tony nominations	0.0690	1.00	0.319
Tony awards	0.3369	2.97	0.004
New York Times review	-0.1272	-1.38	0.171
Daily News review	0.1804	2.11	0.038

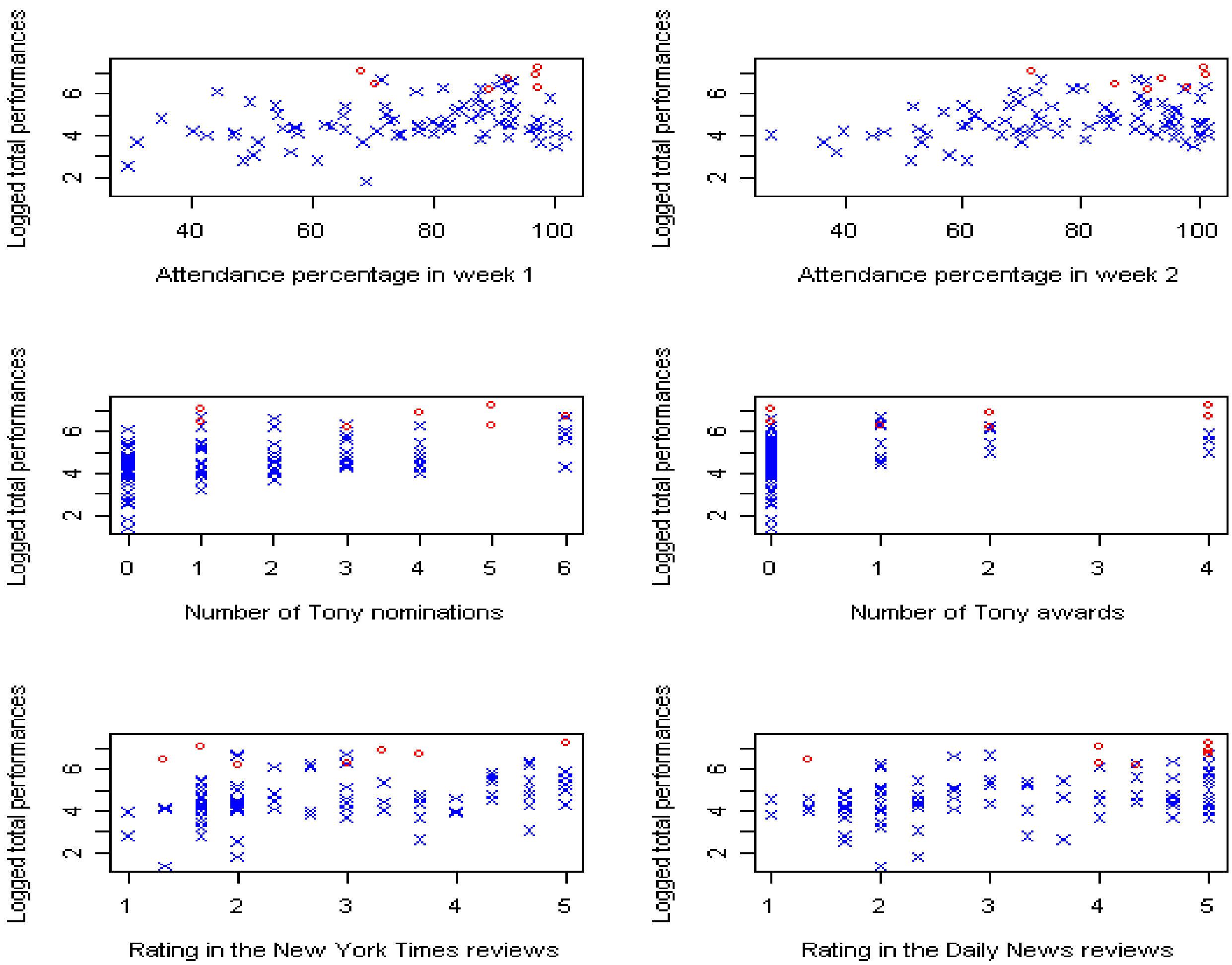
Note. Target variable is the logged total number of performances. Predictors include percentages of seats sold in the first week and the second week, review ratings in the *New York Daily News* and the *New York Times*, the number of Tony nominations in the major categories, and the number of Tony Awards in the major categories. Residual standard error: 0.8364 on 79 degrees of freedom; F-statistic: 7.59 on 6 and 79 degrees of freedom (p-value = 2×10^{-6}); R-squared: 0.3658, adjusted R-squared: 0.3176.

The only explanatory variables that are significantly related to the show longevity are the number of Tony Awards in the major categories and the reviews in the *Daily News*. The coefficients imply that each additional Tony Award is associated with multiplying the estimated total number of performances by roughly 1.4 ($e^{0.3369}=1.40$), and a rating one point higher in the *Daily News* review is associated with multiplying it by almost 1.2 (holding all else fixed). In contrast to the positive effect of the *Daily News* reviews on show longevity, the rating in the *Times* reviews has a negative coefficient, which indicates that better ratings in the *Times* reviews hurt longevity of a Broadway show (although this coefficient is not statistically significantly different from zero).

To take into account the censoring in the target variable, we use a censoring indicator variable c to keep track of the different types of observations in the data. Specifically, we use $c = 1$ to denote that the observed total number of performances is an actual survival time (i.e., the show has closed by the end of the 1999-2000 season), and we use $c = 0$ to denote that the observed total number of performances is a censored value (i.e., the show is still open at the end of the study). This variable is not used in the least squares regression fit, but is crucial for the survival analysis models to be described shortly. Figure 2 presents the scatter plots of the logged total number of performances versus predictors, where different plotting symbols are used for the censored and non-censored data points.

Several interesting findings emerge in Figure 2. The dependent variable, show longevity, must take on positive values. This discourages use of a strictly linear model in favor of a loglinear model, as fitted values from a linear model could be negative, especially for shows closed with very few performances. In addition, we note that the censored points (i.e., the shows still open at the end of the study) are all distributed systematically towards the top of the plots. This trend is anticipated since the shows that are still open at the end of the study have more performances, while the shows that have short runs had already closed by the end of the final season examined. To see how this pattern can lead to biases in the analyses, we now repeat the linear regression fitting omitting the censored observations. The results are presented in Table 2.

The values of the coefficients are noticeably different from those in Table 1, with most changing in the direction of a weaker relationship with logged total number of performances. Now the only significant predictor is Tony Awards. The R-squared is lower, even though the standard error of the estimate is also smaller, which is not surprising given the existence of less variability in the logged total number of performances from omitting several large values.



o = censored observation × = non-censored observation

Figure 2. Scatter plots of logged total number of performances versus predictors. The plotting symbols represent whether or not the observation is censored.

Table 2. Results of Least Squares Linear Regression Model Fit For Total Number of Performances Omitting the Censored Observations

Variable	Coefficient	*t*	*p*
Constant	3.3276	7.76	<0.001
First-week attendance	0.0056	0.66	0.513
Second-week attendance	0.0057	0.68	0.498
Tony nominations	0.0717	1.09	0.280
Tony awards	0.2714	2.21	0.030
New York Times review	-0.0304	-0.33	0.741
Daily News review	0.1305	1.56	0.124

Note. Residual standard error: 0.78 on 72 degrees of freedom; F-statistic: 4.66 on 6 and 72 degrees of freedom (p-value = 0.0005); R-squared: 0.2797, adjusted R-squared: 0.2197.

Table 3 summarizes the observed total number of performances and the predicted total number of performances using the model presented in Table 2 for the seven shows that were still open at the end of the study period (i.e., the censored observations). It is clear that the model in Table 2 seriously underestimates the total number of performances of the censored observations.

As was noted earlier, show longevities can be viewed as "survival times." Censored data are common in survival data, as it is often the case that a subject is still alive at the end of the study period. For such data (including the data here), statistical models and methods designed for modeling survival time with censored observations should be used to analyze the data. Many statistical packages will perform these analyses; the results given here were obtained using the package R (R Development Core Team, 2006).

A useful descriptive statistic to begin with in any analysis of survival time is the Kaplan-Meier estimator of the survival function, also known as the *product limit* estimator (Hosmer and Lemeshow, 1999, provide a thorough discussion of this and the other survival analysis methods discussed here). It provides a description of the overall pattern of survival times. The survival function in the current context is the probability of observing the total number of performances of a show greater than or equal to some stated value, denoted as *t*. The survival function $S(t)$ satisfies $S(t) = P(T \geq t) = \int_t^\infty f(x)dx$, where $f(\cdot)$ is the density function of the survival time.

TABLE 3. Observed and Predicted Total Number of Performances for Censored Observations

Show	Total Performances through May 3, 2000	Prediction
Annie Get Your Gun	482	273
Cabaret	841	621
Chicago	1437	594
Footloose	639	82
Fosse!	543	242
Jekyll and Hyde	1257	105
The Lion King	1036	338

A graph is an effective way to display an estimate of a survival function. Figure 3 presents the plot of the Kaplan-Meier estimate of the survival function for the total number of performances.

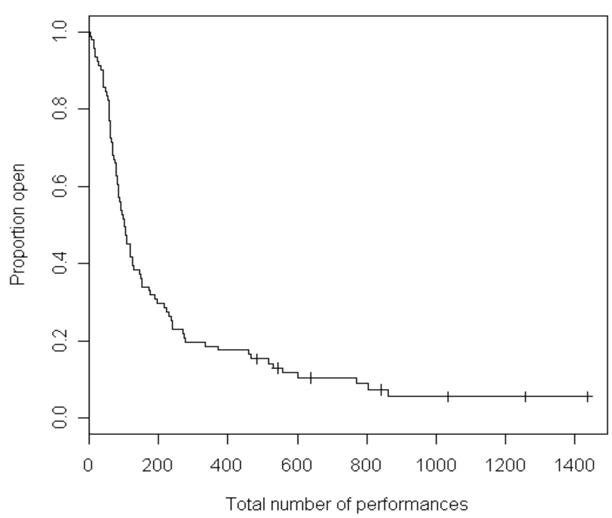


Figure 3. Kaplan-Meier estimate of the survival function for the total number of performances.

The Kaplan-Meier curve shows the decreasing step function defined by the estimated survival function. It drops at a value of the total number of performances if a show closed at that number, and is constant otherwise. An important advantage of the Kaplan-Meier curve is that the method takes into account censored data. In Figure 3, small vertical tick-marks indicate censored observations (i.e., shows still open by the end of the study). Note that the survival function estimate descends sharply at first and then tails off gradually. The initial steep descent demonstrates that there were many shows that closed shortly after opening. The relative long right tail represents the few shows that had many performances. The minimum value of the survival function is not zero since the largest number of performances was a censored observation, and hence there is positive estimated probability of survival past that number of performances.

A measure that directly captures the essence of the fundamental underlying survival process is the hazard function $h(t)$. The hazard function is the instantaneous risk of failure (i.e., the show closing), given that it has survived to time *t*, and satisfies $h(t) = f(t)/S(t)$. As was noted earlier, $f(t)$ is the density function of the survival time and $S(t)$ is the survival function. Since the goal here is to identify the covariates that are related to the show longevity, it is more important to characterize how the distribution of the survival time changes as a function of the covariates than to specify the basic underlying distribution of survival time. In this setting, a regression model for the hazard function that describes survival time in a comparative sense provides more flexibility than one that requires estimation of the underlying survival distribution.

The most frequently used regression model of this type is the *proportional hazards* model, first proposed by Cox (1972). The data for each show are denoted by the triple, (t, c, $\mathbf{x}$), where t is the number of performances until either the show closes or the end of the 1999-2000 season, c is the censoring indictor variable as specified earlier, and $\mathbf{x}$ represents values of the predictor variables for the show. In the Cox model, the hazard function $h(t, \mathbf{x}, \boldsymbol{\beta})$ is modeled as satisfying

$$h(t, \mathbf{x}, \boldsymbol{\beta}) = h_0(t)e^{\mathbf{x}'\boldsymbol{\beta}}. \quad (1)$$

This is a semiparametric model in that the baseline hazard function $h_0(t)$ is not specified, which allows for a wide variety of possible survival distributions. This baseline hazard corresponds to that when each of the covariates equals zero, or equivalently (if the predictors are treated as centered in the analysis, as is typical) when they each equal their mean value. The model implies that hazard functions for different $\mathbf{x}$ values are multiplicatively related (hence the name proportional hazards), or equivalently that $S(t, \mathbf{x}, \boldsymbol{\beta}) = [S_0(t)]^{\exp(\mathbf{x}'\boldsymbol{\beta})}$, where $S_0(\cdot)$ is the baseline survival function. This model postulates an exponential effect of a covariate on the per show closing rate, holding all else in the model fixed (or, equivalently, a linear effect of a covariate on the logged closing rate). This exponential relationship is more reasonable than a linear one, given the nonnegativity of survival times and the typically observed right-tailedness of survival time as shown in Figure 1.

Parameter estimates in the proportional hazards model are estimated by maximizing the partial likelihood function. Hypotheses regarding individual parameters are tested using Wald tests, which are defined in the same way as t-tests in least squares regression as the ratio of the estimated parameter to its estimated standard error. The overall significance of the regression relationship is tested using a (partial) likelihood ratio test, which corresponds to the F-test in least squares regression. Although this is not pursued here, the baseline survival function $S_0(t)$ can be estimated using the partial likelihood, resulting in graphical representations of the regression relationship and estimated survival times for specific shows.

Results

Total Number of Performances

We first describe the analysis of the total number of performances of the Broadway shows. Table 4 presents the results of the model fitting. First-week attendance was not available for two shows ("More to Love" and "Rollin on the T.O.B.A."), and one show ("Summer and Smoke") was not reviewed in the *Daily News*, leaving 88 shows in the sample. Note that since it is hazard that is being modeled, a positive coefficient implies an increased risk of a show closing and, hence, a shorter expected survival time.

Table 4. Results of Proportional Hazards Model Fit for Total Number of Performances

Variable	Coefficient	exp(Coefficient)	z	p
Type of show:				
Musical	-0.9747	0.377	-3.12	0.002
Musical revue	-0.8568	0.425	-1.72	0.086
Revival	0.2591	1.296	0.98	0.330
Opens in July	2.7218	15.208	3.15	0.002
First-week attendance	-0.0174	0.983	-2.39	0.017
Daily News review	-0.2642	0.768	-2.25	0.025
New York Times review	0.0482	1.049	0.38	0.700
Tony nominations	0.0438	1.045	0.47	0.640
Tony Awards	-0.5295	0.589	-2.78	0.005

Note. Target variable is the total number of performances. Predictors include two indicator variables identifying the type of show (musical and musical revue), an indicator variable identifying if the show was a revival, a seasonality effect represented by an indicator variable identifying if the show opened in July (the only monthly indicator variable that was statistically significant), the percentage of seats sold in the first week, review ratings in the *New York Daily News* and the *New York Times*, the number of Tony nominations in the major categories, and the number of Tony Awards in the major categories. The entries under "exp(Coefficient)" are the multiplicative effect of a one-unit increase of the predictor on the hazard function given the other variables are held fixed. z refers to the Wald test of significance of the coefficient, with p the associated two-tailed significance level. The partial likelihood ratio test LR is compared with a χ^2 random variable on the appropriate number of degrees of freedom and tests the overall significance of the predictors. Overall significance: $LR = 54.1$ on 9 degrees of freedom ($p = 2\times 10^{-8}$).

The table shows that the model provides significant predictive power for the risk of a show closing, as the partial likelihood ratio test of overall significance is highly significant. The type of show is significantly related to show survival, with (as expected) musical shows surviving longer than plays. The coefficients indicate that given the other variables, the hazard for a musical is 62.3% lower than for a play, and is 57.5% lower for a musical revue than for a play. In contrast, whether or not a show is a revival is not a significant factor in survival, suggesting that this form of information does not influence consumer behavior.

The only component of the seasonal effect of opening month that is significantly associated with survival is whether or not the show opened in July, with a July opening increasing the hazard by a factor of more than 15. This result, although not necessarily unexpected (shows with a high profile are unlikely to open in July since the New York social scene moves to eastern Long Island during the summer), should be viewed with caution, since it is based on only two shows ("A Thousand Clowns" and "Twelfth Night"), both of which closed quickly. Only nine shows opened on Broadway during the months May through September of the study period, suggesting that producers consider seasonal patterns when deciding when to introduce a show. This would naturally decrease the chances of observing of a strong seasonal effect on longevity.

Initial customer reaction to a show is directly related to show success, as anticipated. Each additional percentage point of attendance (as a percent of total available seats) during the first week of opening is associated with a 1.7% decrease in hazard, holding all other variables in the model fixed.

As expected from the discussions given earlier, critic reviews can be important in predicting longevity of a show, but in ways that are different from those observed by Reddy et al. (1998) in an important way. A positive review in the *Daily News* is associated with a significant more successful show, as a rating one point higher is associated with a 23.2% drop in the hazard, given the other variables. On the other hand, reviews in the *Times* are not at all related to show longevity. Simonoff and Ma (2003) gives an extensive discussion of this interesting finding. The essence seems to be that while a review in the *Times* might once have had a strong effect on the success of a show, that is no longer the case because of the way that the audience for and marketing of Broadway shows has changed in the last 20 years.

Winning awards is also apparently associated with show longevity. While the number of Tony Award nominations in major categories is not predictive for the risk of a show closing, the actual number of awards is, with each additional award associated with a 41.1% decrease in the hazard holding all else in the model fixed. There are many reasons why this pattern might occur: inherently higher quality of the show, greater opportunities for positive advertising, and the presence of a positive information source for potential attendees.

Number of Performances after Tony Nominations

As was noted earlier, Tony nominations and Awards are very important for the survival of Broadway shows. Given this, some shows might stay open and survive longer than expected just to try to make it to the awards announcements, when they would otherwise close. To remove this potential source of bias, in this section we examine show longevity as measured by the number of performances after the announcement of Tony award nominations.

Table 5 summarizes the results of a proportional hazards model fit to the 57 shows in the sample that were open at the time of the Tony nominations. There is no seasonal (opening month) effect included in this case, since none of the underlying indicator variables is close to statistically significant. In addition, the show type effect is represented only by an indicator variable for musicals, which implies that musical revues and plays are pooled together. Audience reaction to the show and the effect of Tony nominations on attendance are captured using the percentage of seats sold in the week after the announcement of the nominations. The model with first-week attendance included does not provide additional predictive power, suggesting that attendance the first week after the nominations captures the useful information related to attendance.

Table 5 shares quite a bit of similarity with Table 4. Musicals are more successful than other shows, having a 79.3% lower risk of closing given the other variables. Attendance in the week after nominations are announced is predictive for survival, with an additional percentage point of seats filled associated with a 1.8% decrease in the hazard of closing holding all else fixed. As was true when modeling total performances, positive reviews in the *Daily News* are associated with increased longevity, while reviews in the *Times* are not at all related to show success. Although the number of Tony nominations in major categories is unrelated to hazard, each additional Tony Award is associated with a 39.7% decrease in the risk of the show closing, given the other predictors.

There are two differences between tables 4 and 5 that are worth pointing out. First, there is no evidence of a seasonal (opening month) effect for postnomination longevity. This is actually not surprising, given that the only effect for total longevity was related to opening in July, and neither of the two shows that opened in July was still open at the time of Tony Award nominations, more than 8 months later. The other difference is that whether or not a show is a revival is a statistically significant predictor for hazard, with revivals having more than twice the risk of closing, given the other variables. The reason that

Table 5. Results of Proportional Hazards Model Fit for Number of Performances After Tony Nomination Announcement

Variable	Coefficient	exp(Coefficient)	z	p
Musical	-1.5750	0.207	-3.85	<0.001
Revival	0.7106	2.035	2.09	0.037
Attendance after nominations	-0.0178	0.982	-1.92	0.055
Daily News review	-0.3242	0.723	-2.00	0.046
New York Times review	0.0018	1.002	0.01	0.990
Tony nominations	0.1492	1.161	1.24	0.210
Tony Awards	-0.5052	0.603	-2.41	0.016

Note. Target variable is the number of performances after the announcement of Tony Award nominations. Predictors include a show-type effect represented by an indicator variable identifying if the show was a musical (the only indicator needed to define the statistical extent of the effect), an indicator variable identifying if the show was a revival, the percentage of seats sold in the week after the announcement of nominations, review ratings in the *New York Daily News* and the *New York Times*, the number of Tony nominations in the major categories, and the number of Tony Awards in the major categories. Overall significance: LR = 42.1 on 7 degrees of freedom ($p = 5 \times 10^{-7}$).

the revival effect is not significant for total number of performances is that shows that were no longer open at the time of the Tony nominations (which are not used in the analysis in this section) had very short run times (an average of approximate 66 performances), with relatively little difference between revivals and nonrevivals (revivals averaging 78 performances and nonrevivals 59 performances). Shows that were still open when the Tony nominations were announced were generally longer-lived as anticipated, but, among this group, revivals were less successful (revivals averaging 178 performances while nonrevivals 267 performances). One reason for this pattern could be that the familiarity to the potential audience of the material in a revival is viewed as a lack of novelty, keeping audiences away.

Number of Performances after Tony Awards

In this section show longevity is measured as the total number of performances after the announcement of Tony Award winners. Table 6 summarizes the results of a proportional hazards model fit to the 50 shows in the sample that were open when the Tony Award winners were announced. Just as was true when modeling performances after the Tony nomination announcement, there is no seasonal effect included, and the show type effect is represented only by an indicator variable for musicals. Audience reaction to the show, and the effect of Tony nominations and Awards on attendance, is captured using the percentage of seats sold in the week after the announcement of the nominations and percentage of seats sold in the week after the announcement of the awards.

Table 6. Results of Proportional Hazards Model Fit for Number of Performances After Tony Awards Announcement

Variable	Coefficient	exp(Coefficient)	z	p
Musical	-1.8229	0.162	-3.94	<0.001
Revival	1.2492	3.487	2.86	0.004
Attendance after nominations	-0.0413	0.959	-2.50	0.012
Attendance after awards	0.0254	1.026	1.40	0.160
Daily News review	-0.3169	0.728	-1.74	0.083
New York Times review	-0.0131	0.987	-0.07	0.950
Losing Tony nominations	0.2587	1.295	1.81	0.070
Tony Awards	-0.3873	0.679	-1.98	0.048

Note. Target variable is the number of performances after the announcement of Tony Award winners. Predictors include a show-type effect represented by an indicator variable identifying if the show was a musical (the only indicator needed to define the statistical extent of the effect), an indicator variable identifying if the show was a revival, the percentage of seats sold in the week after the announcement of nominations, the percentage of seats sold in the week after the announcement of winners, review ratings in the *New York Daily News* and the *New York Times*, the number of losing Tony nominations in the major categories, and the number of Tony Awards in the major categories. Overall significance: $LR = 43.4$ on 8 degrees of freedom ($p = 7 \times 10^{-7}$).

Several covariate relationships are similar to earlier results. Once again musicals run longer than other types of shows, having 83.8% lower risk of closing given the other variables. The relative cost of a show being a revival is even stronger here, with revivals having 3.5 times the risk of closing, holding all else fixed. Once again, reviews in the *Times* are unrelated to show longevity, but now the favorable nature of a positive review in the *Daily News* is only marginally significant. This is actually not surprising, since all of the shows had been open at least 4 weeks by the time of the Tony Awards ceremony, (so a lessening of the effect of an opening-night review would be expected), and one would expect that Tony Awards and nominations would provide important "official" information for potential customers replacing that of reviews. One somewhat puzzling result is that while attendance in the week after announcement of nominations is significantly associated with postaward longevity (one additional percentage point of attendance reducing the risk of closing by 4% holding all else fixed), attendance in the week after announcement of the awards is not.

The relationship of postaward longevity to Tony nominations and Awards is particularly interesting. Unlike in the earlier situations, the number of Tony nominations is (weakly) associated with survival. In order to make clearer what the model is saying, we divide Tony nominations into two components: winning nominations (Tony Awards) and losing nominations (i.e., the number of Tony Awards subtracted from the number of nominations). Consistent with the earlier analyses, receiving more Tony Awards is associated with longer survival, with an additional award associated with a 32.1% decrease in the risk of closing given the other variables[1].

[1] Changing the predictors in the proportional hazard model from total nominations and Tony Awards to losing nominations and Tony Awards does not affect the fit of the model in any way. The only change is to the coefficient for Tony Awards, since it now represents something slightly different than before; it is now an estimate of the exponential effect on survival of one additional award, given that the total number of losing nominations (and other predictors) is kept fixed, rather than an estimate of the exponential effect on survival of one additional award, given that the total number of nominations (and other predictors) is kept fixed.

The number of losing Tony nominations is inversely related to postaward longevity, which might seem counterintuitive but is actually reasonable. Each additional losing Tony Award nomination is associated with a 29.5% increase in the risk of closing, but this is given that all else is held fixed, including the number of Tony Awards. Thus, losing nominations (additional nominations without additional wins) are apparently viewed as negative information by the public and are associated with increased risk of the show closing.

Assessment of Model Adequacy

Residuals

As is true for any statistical model, inferences from the proportional hazards model are valid only if the assumptions being made hold, at least approximately. In this section, we check if there are any outliers (shows whose longevity is strongly out of line with what is expected) and in the next section, we discuss whether the proportional hazards assumption holds.

Since the response variable in these analyses includes censored observations, and the proportional hazards model attempts to fit the hazard function, rather than show longevity directly, there is no obvious definition of residuals for the model. One commonly used set of residuals is the martingale residuals,

$$e_i = c_i - \hat{H}(t_i, \mathbf{x}, \hat{\beta}),$$

where $\hat{H}(t_i, \mathbf{x}, \hat{\beta})$ is the estimated cumulative hazard at time t_i (that is, the integral of the estimated hazard function through t_i). The martingale residual estimates a value that has mean zero if the model is correct, with positive values corresponding to shows that closed earlier than expected and negative values corresponding to shows that closed later than expected. Figure 4 plots the martingale residuals for the proportional hazards model in Table 4.

As shown in the graph, three shows had notable (negative) martingale residuals: "Footloose" (639+ performances for a musical with no Tony Awards, 70% first-week attendance, and a poor review in the *Daily News*, implying an expected longevity of 175 performances), "Jackie: An American Life" (128 performances for a play with no Tony Awards, 35% first-week attendance, and a poor review in the *Daily News*, which would imply an expected longevity of 59 performances), and "Jekyll and Hyde" (1,257+ performances for a musical with no Tony Awards, 68% first-week attendance, and a good review in the *Daily News*, implying an expected longevity of 310 performances). The martingale residuals do not exhibit any autocorrelation, supporting the assumption of independence of the survival observations in the sample (the observations in the sample are ordered based on the date the show opened).

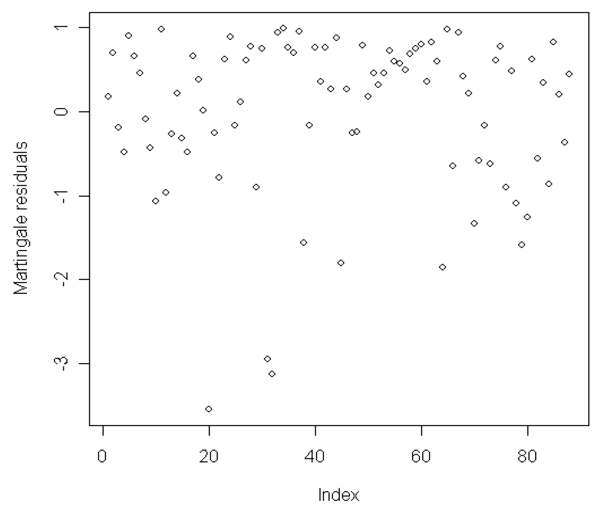


Figure 4. Martingale residuals for the proportional hazards model in Table 4.

Figure 5 plots the martingale residuals for the proportional hazards model for postnomination performances corresponding to Table 5. Note that none of the shows are outlying according to the martingale residuals. The martingale residuals also do not exhibit any autocorrelation. Figure 6 gives the martingale residual plot for the proportional hazards model for postaward performances corresponding to Table 6. None of the shows are outlying in this plot as well. The plot also provides no evidence for autocorrelation in the martingale residuals.

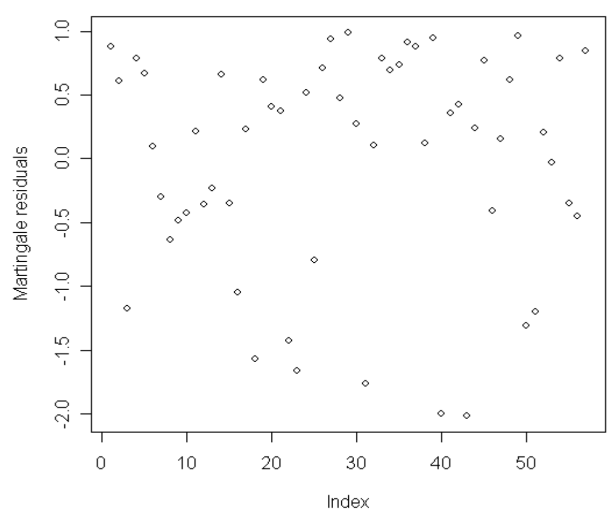


Figure 5. Martingale residuals for the proportional hazards model in Table 5.

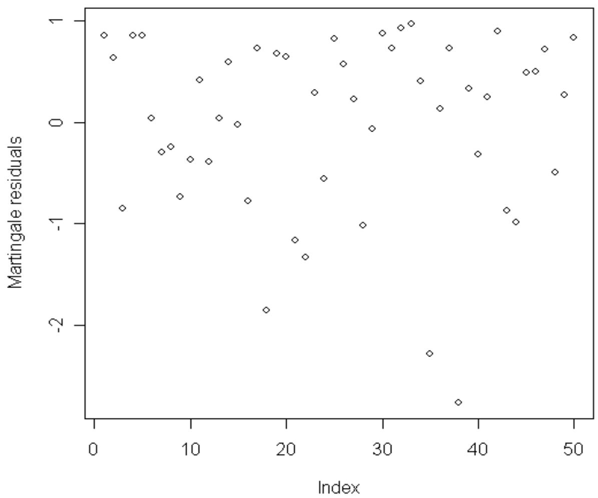


Figure 6. Martingale residuals for the proportional hazards model in Table 6.

Table 7 presents the results of the proportional hazards model fit when these three shows are omitted from the sample. Comparing the results in Table 7 with those in Table 4, we see that the effects that were statistically significant remain so (and become slightly stronger), whiles those that were not statistically significant remain insignificant. Thus, the three outliers do not substantially affect the implications of the analysis.

Assessing the Proportional Hazards Assumption

The key assumption in the proportional hazard model (1) is that the effect of a covariate on the hazard function is the same at all times. Grambsch and Therneau (1994) proposed testing this assumption via the specific form of time-varying coefficient

$$\beta_j(t) = \beta_j + \gamma_j g_j(t),$$

where $g_j(t)$ is a specified function of time such as the identity, logs, or ranks. Proportional hazards correspond to $\gamma_j = 0$ for all j, and one way to test the hypothesis that $\gamma_j = 0$ for each coefficient is via a score test. Hosmer and Lemeshow (1999, page 207) recommend using the log function in these tests, and that is what is done here.

Table 8 presents the p-values for score tests for the proportional hazards model shown in Table 4. Small p-values provide evidence of nonproportional hazards related to that variable. The results in Table 8 imply nonproportionality of hazards linked to the type of show. Violation of the proportionality assumption in this model can be accounted for by fitting a model stratified on the type of show, which will be discussed further in the next section.

TABLE 7. Results of Proportional Hazards Model Fit For Total Number of Performances Omitting Three Outliers

Variable	Coefficient	exp(Coefficient)	*z*	*p*
Type of show				
Musical	-0.8085	0.446	-2.64	0.008
Musical revue	-1.2185	0.296	-2.41	0.016
Revival	0.0361	1.037	0.14	0.890
Opens in July	3.0521	21.159	3.50	<0.001
First-week attendance	-0.0283	0.972	-3.64	<0.001
Daily News review	-0.2885	0.749	-2.35	0.019
New York Times review	-0.0010	0.999	-0.01	0.990
Tony nominations	-0.0299	0.970	-0.33	0.740
Tony Awards	-0.6352	0.530	-3.02	0.003

Note. Target variable is the total number of performances. Predictors are the same as those in Table 4. The sample excludes the three outliers. Overall significance: *LR* = 68.7 on 9 degrees of freedom ($p = 3 \times 10^{-11}$).

Table 8. p-Values of Score Tests for Proportional Hazards Assumption For Model Summarized in Table 4

Variable	p
Type of show	
Musical	0.003
Musical revue	0.747
Revival	0.070
Opens in July	0.514
First-week attendance	0.327
Daily News review	0.361
New York Times review	0.411
Tony nominations	0.146
Tony Awards	0.718

Table 9 gives corresponding p-values for tests based on the model summarized in Table 5. Since there is not strong evidence of nonproportionality of hazards here, we conclude that the nonstratified proportional hazards model in Table 5 seems appropriate for modeling post-nomination longevity.

Table 9. p-Values of Score Tests for Proportional Hazards Assumption For Model Summarized in Table 5

Variable	*p*
Musical	0.086
Revival	0.235
Attendance after nominations	0.632
Daily News review	0.450
New York Times review	0.766
Tony nominations	0.500
Tony Awards	0.647

Table 10 is used to check the proportionality assumption for the proportional hazards model in Table 6. The p-values for score tests in the table indicate nonproportionality linked to several variables, suggesting a stratified model fitting may be appropriate.

Table 10. p-Values of Score Tests for Proportional Hazards Assumption For Model Summarized in Table 6

Variable	p
Musical	0.081
Revival	0.517
Attendance after nominations	0.008
Attendance after Awards	0.033
Daily News review	0.134
New York Times review	0.419
Losing Tony nominations	0.055
Tony Awards	0.456

Model Extension for Nonproportionality of Hazards

When the proportional hazards assumption is violated, a useful generalization of the model that can treat the problem is the stratified proportional hazards model. Consider a show characteristic defined by a nominal variable, such as show type (i.e., musical, musical revue, or play). The model (1) assumes that, given the other covariates, the hazard function for, say, musicals is a constant multiple of that for, say, musical revues for all times t. That is, the relative hazard of a musical with specified characteristics closing at time t compared to that of a musical revue with the same characteristics closing at time t is the same for all values of t. This might not be the case, as different types of shows might "age" differently, and the resultant nonproportionality of hazards means that the proportional hazards model is no longer appropriate.

A simple way to address this nonproportionality is the *stratified proportional hazards model*, which assumes different baseline hazard functions for the different levels of the nominal variable (i.e., for each type of show). For this model, the hazard function satisfies

$$h_s(t, \mathbf{x}, \boldsymbol{\beta}) = h_{s0}(t) e^{\mathbf{x}'\boldsymbol{\beta}},$$

where $h_{s0}(\cdot)$ is the baseline hazard for level s. Under this model, the effect of being in level s can be summarized using, for example, the median baseline survival time for shows at that level, while the regression coefficients are interpreted as multiplicative effects on the hazard as always.

Table 11 summarizes the results of the stratified proportional hazards model fit for the total number of performances of the Broadway shows. Table 8 identified that the type of show should be a stratification variable, so the indicator variables identifying the type of show (musical and musical revue) are excluded from the stratified model. Thus, while in the earlier analysis each different type of show had a different multiplicative effect on the baseline hazard function, in this analysis each different type of show has a different baseline hazard function. Further, while in the nonstratified model for the total number of performances an indicator variable identifying if the show opened in July was used to capture the seasonal effect, this indicator variable is no longer appropriate for the stratified model since (as was noted earlier) only two shows in the sample opened in July, and both are plays. Instead, in this model an indicator variable identifying if the show opened in summer months (May through September) is used to incorporate the seasonal effect into the stratified model.

Table 11. Results of Stratified Proportional Hazards Model Fit For Total Number of Performances

Variable	Coefficient	exp(Coefficient)	z	p
Revival	0.4107	1.508	1.52	0.130
Opens in summer	0.9989	2.715	2.19	0.028
First-week attendance	-0.0206	0.980	-2.61	0.009
Daily News review	-0.3064	0.736	-2.45	0.014
New York Times review	-0.0390	0.962	-0.30	0.760
Tony nominations	0.0652	1.067	0.69	0.490
Tony Awards	-0.6388	0.528	-3.08	0.002

Note. Target variable is the total number of performances. The stratification variable is the type of show. Predictors include an indicator variable identifying if the show was a revival, a seasonality effect represented by an indicator variable identifying if the show opened in summer months (May through September), the percentage of seats sold in the first week, review ratings in the *New York Daily News* and the *New York Times*, the number of Tony nominations in the major categories, and the number of Tony Awards in the major categories. Overall significance: LR = 45.3 on 7 degrees of freedom ($p = 1.2 \times 10^{-7}$).

The results in Table 11 are nearly identical to those in Table 4. The interpretation of the estimates would not differ substantially from those discussed earlier and thus are not repeated here. The score tests of nonproportionality of hazards as shown in Table 12 show no problems with model assumptions based on the stratified model.

Table 13 presents the results of a stratified proportional hazards model fit for the number of postaward performances. The stratification variables used are the number of losing Tony nominations and whether or not a show is a musical. The coefficients of the covariates in the stratified model are similar to those for the nonstratified model (as shown in Table 6), and the resultant inferences are also very similar. The p-values in Table 14 do not suggest significant nonproportionality of hazards.

TABLE 12. p-Values of Score Tests for Proportional Hazards Assumption For Model Summarized in Table 11

Variable	*p*
Revival	0.105
Opens in summer	0.159
First-week attendance	0.461
Daily News review	0.332
New York Times review	0.242
Tony nominations	0.113
Tony Awards	0.598

Table 13. Results of Stratified Proportional Hazards Model Fit For Number of Performances After Tony Awards Announcement

Variable	Coefficient	exp(Coefficient)	*z*	*p*
Revival	2.2931	9.905	3.29	0.001
Attendance after nominations	-0.0512	0.950	-2.59	0.010
Attendance after awards	0.0361	1.037	1.89	0.059
Daily News review	-0.2087	0.812	-0.82	0.410
New York Times review	-0.4481	0.639	-1.51	0.130
Tony Awards	-0.3780	0.685	-1.43	0.150

NOTE. Target variable is the number of performances after the announcement of Tony Award winner. The stratification variables are the number of losing Tony nominations and whether or not a show is a musical. Predictors include an indicator variable identifying if the show was a revival, the percentage of seats sold in the week after the announcement of nominations, the percentage of seats sold in the week after the announcement of winners, review ratings in the *New York Daily News* and the *New York Times*, and the number of Tony Awards in the major categories. Overall significance: $LR = 28.2$ on 6 degrees of freedom ($p = 9 \times 10^{-5}$).

Table 14. p-Values of Score Tests for Proportional Hazards Assumption For Model Summarized in Table 13

Variable	*p*
Revival	0.263
Attendance after nominations	0.701
Attendance after awards	0.660
Daily News review	0.567
New York Times review	0.775
Tony Awards	0.662

Conclusion

In this study, we have used the proportional hazards model to investigate the factors relating to the longevity of Broadway shows. We find that the type of show is an important predictor for show longevity, with musicals having longer run times than other shows. Critic reviews in the *Daily News* are related to longevity, as would be expected, but, in contrast to earlier investigations, reviews in the *New York Times* are unrelated to the success of a show. Winning major Tony Awards is associated with greater success, but being nominated and then losing is negatively related to postaward longevity. Rather than being a positively viewed stamp of approval, the status of a show as a revival is inversely related to show success, at least after the Tony Award nominations have been announced. As anticipated, increased early attendance is associated with greater success of a show.

REFERENCES

Cox, D. R. 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* 34:187-220.

De Vany, A. 2004. *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry.* London: Routledge.

Grambsch, P. M., and Therneau, T. M. 1994. Proportional hazards tests in diagnostics based on weighted residuals. *Biometrika* 81:515-26.

Hosmer, D. W., Jr., and Lemeshow, S. 1999. *Applied Survival Analysis.* New York: Wiley & Sons.

R Development Core Team 2006. *R: A Language and Environment for Statistical Computing.* Vienna, Austria.

Radas, S., and Shugan, S. M. 1998. Seasonal marketing and timing new product introductions. *Journal of Marketing Science* 35:296-315.

Ravid, S. A. 1999. Information, blockbusters and stars—a study of the film industry. *Journal of Business* 72:463-92.

Reddy, S. K., Swaminathan, V., and Motley, C. M. 1998. Exploring the determinants of Broadway show success. *Journal of Marketing Research* 35:370-83.

Simonoff, J. S., and Ma, L. 2003. An empirical study of factors relating to the success of Broadway shows. *Journal of Business* 76:135-50.

Simonoff, J. S., and Sparrow, I. R. 2000. Predicting movie grosses: Winners and losers, blockbusters and sleepers. *Chance* 13, no. 3:15-24.

Appendix: Illustrative R code for censored data model fitting

```
# See http://www.r-project.org/ for information on the free package R

library(survival)

# Kaplan-Meier estimated survival function

plot(survfit(Surv(Total.performances, Closed) ~ 1), conf.int=F, ylab="Proportion open", xlab="Total num-
ber of performances"))

# Proportional hazards model in Table 4

ph1 <- coxph(Surv(Total.performances, Closed) ~ Att.week.1 +
  Tony.nominations + Tony.awards + Musical + Musical.revue + Reviv +
  NYT.rating + DN.rating + July, iter.max=100, na.action=na.omit)

# Proportional hazards model in Table 5

ph2 <- coxph(Surv(Post.nomination.performances,Closed) ~ Att.post.nom +
  Tony.nominations + Tony.awards + Musical + Reviv + NYT.rating +
  DN.rating, iter.max=100, na.action=na.omit)

# Proportional hazards model in Table 6

ph3 <- coxph(Surv(Post.award.performances,Closed) ~ Att.post.nom +
  Att.post.award + Losing.nominations + Tony.awards + Musical + Reviv +
  NYT.rating + DN.rating, iter.max=100, na.action=na.omit)

# Tests of proportionality of hazards in Tables 8 - 10

cox.zph(ph1, transform=”log”)
cox.zph(ph2, transform=”log”)
cox.zph(ph3, transform=”log”)

# Stratified proportional hazards model in Table 11

ph4 <- coxph(Surv(Total.performances,Closed) ~ Att.week.1 +
  Tony.nominations + Tony.awards + Reviv + NYT.rating + DN.rating +
  Summer + strata(Type), iter.max=100, na.action=na.omit)

# Test of proportionality of hazards in Table 12

cox.zph(ph4, transform=”log”)

# Stratified proportional hazards model in Table 13

ph5 <- coxph(Surv(Post.award.performances,Closed) ~ Att.post.nom +
  Att.post.award + strata(Losing.nominations) + Tony.awards +
  strata(Type) + Reviv + NYT.rating + DN.rating, iter.max=100,
  na.action=na.omit)

# Test of proportionality of hazards in Table 14

cox.zph(ph5, transform=”log”)
```

Correspondence: lnygren@rider.edu