ECON 546: Themes in Econometrics

Lab Exercises #8
(10 March, 2010)

In this lab. class we are going to get some more practice in estimating "count" data models, using MLE. *It would be advisable to read the first three sections of the attached paper, "Survival of the Hippest: Life at the Top of the Hot 100", prior to the lab. class*. That paper uses survival analysis to analyze the same data-set you will be working with. These data are in the Excel workbook, **S:\Social Sciences\Economics\Econ546\Hot100.xls**. A description of the variables is given below the data records in the Excel file.

(a)   Load the data into an EViews workfile. Create a new variable, EXTRA = WEEKS – 1. This variable represents the number of additional weeks that a number one hit stays at the top of the charts, beyond the minimum of one week that such a hit must achieve, by definition. View the "Histogram and Stats" for the variable "EXTRA". Do you think it is reasonable to model this variable using a Poisson process?

(b)   Estimate a negative binomial regression model that explains "EXTRA" as a function of variables that you think may be important. Include a constant in the model, and be careful in your use of the various dummy variables so as to avoid "perfect multicollinearity". Experiment with the model specification, using the Huber/White standard errors, and try and find a version that includes only significant explanatory variables. [**Note:** You can use the SIC to discriminate between alternative non-nested models, and you might consider examining "interactions" between the various dummy variables and other variables, such as "FEMALE*YEAR".]

(c)   For use below, estimate the Poisson version of your preferred model.

(d)   Was being Elvis or The Beatles a significant factor in determining the number of extra weeks at the top for a Hot 100 number one hit? What about gender, year of the hit, or being a purely instrumental number?

(e)   Construct an "Actual & Fitted" plot for your preferred specification.

(f)   Using this last model, use the Wald test to test if the Poisson model is preferred to the Negative Binomial model. What do you conclude?

(g)   Use the Cameron-Trivedi LM test for over-dispersion. Is the outcome of the test consistent with your conclusion in part (f)?

(h)   Can you think of a further way of testing the null of a Poisson model against the alternative of a Negative Binomial model?

(i)   Using either the Poisson or Negative Binomial model (as determined by your result in parts (f) and (g) above), calculate the marginal effect (on the number of extra weeks at the top of the Hot 100) of the number of number one hits that were achieved by all artists in that year.

Survival of the hippest: life at the top of the hot 100

David E. Giles

Department of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., V8W 2Y2 Canada
E-mail: dgiles@uvic.ca

We analyse the survival characteristics of recordings that reached the number one spot on the US popular music charts over the period 1955 to 2003. Our results show that there has been a statistically significant change in the time spent at number one since 'album cuts' were included in the compilation of *Billboard*'s Hot 100. Survival time is significantly improved if the recording is by a female solo artist, or if it is an instrumental tune. We also find a significant 'Elvis effect'.

'I'll never be a saint, it's true. I'm too busy surviving!' (Madonna, 1994)

I. Introduction

In this article we undertake a survival analysis of musical recordings that attained 'number one hit' status in the US music industry over the period 1955 to 2003. In economic terms, the article deals with product survival. The product is a number one hit recording, and its lifespan is measured in terms of the (possibly nonconsecutive) weeks that the recording remains at the number one spot on the *Billboard* Hot 100 chart. The survival of a recording on the charts, and particularly at the premier position, is of great importance to the artist(s), their agent(s) and the company producing the recording. Survival has obvious implications for immediate revenue generation and also subsequent revenue from future releases, concert attendances, product endorsements, etc. Consequently, an understanding of survival patterns, and of the factors that are significant determinants of survival, is central to our appreciation of revenue patterns in this industry.

Several other authors have studied the survival of other forms of artistic performance. For example, Simonoff and Ma (2003) and Maddison (2005) have

considered Broadway theatre productions, and earlier work by de Vany and Walls (1997) and Walls (1998) analysed the life-span of first-run motion pictures in different countries and languages. However, there appears to be only one study that addresses survival in the popular music industry. Strobl and Tucker (2000) consider various aspects of the market for pre-recorded popular music in the United Kingdom, and part of their analysis involves modelling the durations of albums on the British charts between January 1991 and January 1993.

There are surprisingly few other related studies for the popular music industry. Hamlen (1991) undertook an empirical test of the 'superstardom' hypothesis using US recording sales; Chung and Cox (1994) analysed the underlying probability distribution for the generation of 'gold record' awards in the USA; and Burke (1996) investigated the dynamics of product differentiation in the British recording industry. Recently, Connolly and Krueger (2005) have provided a wide-ranging economic analysis of the rock and roll industry, with a particular emphasis on concert revenues. However, they do not discuss the survival of artists or their recordings.

Here, we consider all 'singles' recordings that reached the 'number one' position on the 'Hot 100' chart (and its predecessors back to 1955) in the USA, up to the end of 2003. This provides a sample of nearly 1000 'lifetimes', where a life (or spell) is measured in terms of the number of weeks that a hit single stays at the top spot on the charts. These spells are modelled, using both nonparametric and parametric procedures with the objective of isolating some of the significant determinants of a long-lived number one hit. Although this modelling abstracts from certain dynamic effects that may be important, we are able to control for a wide range of different factors and obtain quite robust models.

In the next Section we provide a brief background to the product and market that we are considering, with particular emphasis on the charts that are compiled by *Billboard* magazine. The data in this study are described in Section III. The basic concepts associated with modelling survival data are presented in Section IV, together with a preliminary nonparametric analysis. Our main parametric survival modelling, and the associated results, are discussed in Section V, and our conclusions and thoughts for extending the research in various directions are provided in the last section.

II. Historical Background

Prior to the first publication of *Billboard* magazine's 'Hot 100' chart on 4 August 1958, various charts tracked the fortunes of popular music recordings. For example, between 1955 and 1958 *Billboard* published the 'Top 100' chart (from 12 November 1955 to 28 July 1958), as well as the 'Best Sellers in Stores', 'Most Played by Jockeys' and 'Most Played in Jukeboxes' charts. The 'Most Played In Jukeboxes' was a twenty-spot chart that continued until 17 June 1957; the 'Most Played By Jockeys' chart was discontinued on 28 July 1958 and had 20–25 positions at different times; while the 'Best Sellers In Stores' was a top 25 or top 50 list that ended on 13 October 1958. By the end of 1958 all of these charts had been merged into the 'Hot 100'.

Billboard magazine itself was founded in 1894, with a focus on carnival entertainment, it published its first music 'hit parade' on 4 January 1936, and the first *Music Popularity Chart* was calculated on 27 July 1940. The first three singles to reach the top of this chart were *I'll Never Smile Again* and *The Breeze and I*, both by Tommy Dorsey and his Orchestra, followed by Glenn Miller's *Imagination*. Today, *Billboard* publishes a range of music charts but the Hot 100 remains the premier singles chart for North American popular music. Currently, a new chart goes into effect every Saturday, but the information is posted on the magazine's website on the previous Thursday *Billboard Magazine*, 2005).

Originally, only singles that were available for purchase were considered for inclusion in the Hot 100. However, with the decline in importance of singles sales over time, a major modification of the Hot 100 took place on 5 December, 1998. Since then, so-called 'album cuts' have been eligible for the chart. Album cuts are variety singles that receive air time, and hence attract 'airplay points' from Nielsen Broadcast Data Systems (2005), but cannot be purchased as traditional singles. Sales performances of the latter have been tracked by Nielsen SoundScan (2005) since 1991. The first album cut single to reach number one (for one week) on the Hot 100 was Aaliyah's *Try Again*, on 17 June 2000. So, since 1998 both sales and airtime have entered the formula that determines the Hot 100 chart, and we test for the significance of this structural break in the data in our subsequent analysis.

Despite the pre-eminent role of the *Billboard* Hot 1000 chart, it has not been without its critics. One criticizm has been that it has been biased against so-called 'nonRhythmic' songs. *Billboard*'s response was to introduce the 'Pop 100' and 'Pop 100 Airplay' charts on 12 February 2005. At this time, the *Billboard* Hot 100 responded to the impact of the internet on the recorded music industry, and began to track *paid* digital downloads from such websites as Napster (2005), Musicmatch (2005) and Rhapsody (2005). These changes in early 2005 comprised the first major revision of the Hot 100's formula in more than sixteen years.[1]

One of the most successful musical recordings since World War II was the soundtrack from the musical *South Pacific*, which was rated number one by various polls for a total of 61 nonconsecutive weeks, beginning in 1947. Immediately prior to the Hot 100 era, Elvis Presley's *Hound Dog/Don't Be Cruel* stayed at number one for 11 consecutive weeks in 1956. Interestingly, this benchmark figure was not surpassed until 1992, when Boyz II Men took *End of the Road* to the top for 13 weeks on 15 August, and then Whitney Houston's *I Will Always Love You* held the number one slot for 14 weeks from 28 November

---

[1] Billboard magazine keeps up with the times – it now provides a 'Hot Digital Tracks' chart, and even a ten-item 'Hot Ringtones' chart for hard-core cellular telephone users.

of that year. Although this record was equalled by Boyz II Men in 1994, the next (and current) record duration was the 16 week spell set by Mariah Carey and Boyz II Men with *One Sweet Day*, beginning on 2 December 1995. Our interest is in modelling the determinants of these spells. Other issues that will be addressed include changes in the number of number one hits per year; and the impact of major changes in the way in which the *Billboard* Hot 100 chart is compiled.

III. Data

The data that we use have been constructed from information made available on the internet by de Haan (2005), *Billboard Magazine* (2005) and *Answers.com* (2005). The variables are defined in the Appendix.[2] Our primary data-set relates to recordings that reached number one during the period August 1958 to December 2003. The first of these hits was *Poor Little Fool*, by Ricky Nelson, which stayed in the number one spot for two weeks and the last one was *Hey Ya*!, by OutKast, which reached the number one spot on 13 December 2003, and stayed in that position for nine consecutive weeks. It is important to note that a 'recording' is defined here in terms of *both* the artist and the song (tune). So, a song that reaches the top of the charts more than once, because a 'cover version' is recorded subsequently by a different artist, is counted as two separate recordings for our purposes. In fact, there are nine examples of this in our sample, and in addition 'The Twist', by Chubby Checker, made two separate ascents to number one, first in 1960, and again in 1962.

The start of our sample period is determined by the creation of the *Billboard* Hot 100 chart on 4 August 1958, and its finish date avoids much of the recent impact of downloading digital music on the internet. We have also considered an extended sample, beginning in January 1955, using the data from De Haan (2005), and the dummy variable D1958 is included only in this part of the analysis, of course. The dummy variable D1998 accounts for the introduction of album cuts into the compilation of the Hot 100 chart in December 1998. Digital

downloads on the internet started to become an important issue in 2000, so this dummy variable may also be controlling for some of this effect.

This larger sample has the disadvantage of including a period of multiple charts, and slightly ambiguous measures of the WEEKS variable, but it does have the merit of capturing the commencement of the rock and roll era – *Rock Around the Clock*, by Bill Haley and His Comets, became the first rock and roll single to top the charts, on 9 July 1955. Based on different numbers of charts, De Hann (2005) reports that it stayed at number one for nine weeks while the *Answers.com* site reports only eight weeks. In the case of these pre-Hot 100 ambiguities, we have used de Haan's figures in this study. The extended period has the additional advantage of including ten of Elvis Presley's sixteen number one hits.[3] One interesting feature of the data is that several number one hits lost their premier position to other recordings, but subsequently returned to the top spot. The first such example was Bobby Darin's recording of *Mack the Knife*, which initially reached the top of the *Billboard* Hot 100 on 5 October 1959, and ultimately held this position for a total of nine nonconsecutive weeks. We allow for such effects through the NONCON dummy variable. Finally, the YEAR variable controls for the possibility of systematic changes in the life-length of top hits over the sample. While the covariates that we have considered are not totally exhaustive, they capture a wide range of important characteristics.

The characteristics of the data are summarized in Table 1 and Fig. 1. The primary and extended samples comprise 901 and 965 observations respectively. None of the observations for the WEEKS variable are subject to censoring, and all of the covariates are time-independent.[4]

IV. Modelling Survival Data

We begin with some basic definitions.[5] Let T be a continuous, nonnegative, random variable that measures the passage of time, and let t denote a particular realization (duration) of this random variable. Then, the distribution function for the

---

[2] The data are available from the author, in an Excel worksheet, on request.

[3] This tally is actually eleven out of seventeen when the eleven-week double-sided hit record, *Hound Dog/Don't be Cruel* is counted as two separate hits.

[4] With regard to time-invariance, recall that YEAR measures the year in which the recording *first* reached number one. With regard to the absence of censoring, we are making the reasonable assumption that none of the number one recordings prior to December 2003 will again attain the number one spot after June 2005.

[5] Kiefer (1988) provides a very readable introduction to the application of duration analysis to economic data, and we adopt his fairly standard notation in this section.

Table 1. Time at the top – summary statistics (weeks)

Variable	Mean	Median	Min.	Max.	SD	Skew
Hot 100 sample: 4 August 1958 – 13 December 2003 ($N = 901$)						
WEEKS	2.643	2	1	16	2.164	2.362
NONCON	0.040	0	0	1	0.196	4.697
GROUP	0.505	0	0	1	0.500	−0.020
MALE	0.292	0	0	1	0.455	0.915
FEMALE	0.179	0	0	1	1.677	3.814
INST	0.027	0	0	1	0.161	5.880
ELVIS	0.007	0	0	1	0.081	12.132
BEATLES	0.020	0	0	1	0.140	6.861
D1958	n.a.	n.a.	n.a.	n.a	n.a.	n.a.
D1998	0.071	0	0	1	0.257	3.486
Extended sample: 1 January 1955 – 13 December 2003 ($N = 965$)						
WEEKS	2.754	2	1	16	2.242	2.141
NONCON	0.037	0	0	1	0.190	4.883
GROUP	0.485	0	0	1	0.500	0.060
MALE	0.317	0	0	1	0.466	0.786
FEMALE	0.172	0	0	1	0.378	1.738
INST	0.028	0	0	1	0.165	5.724
ELVIS	0.017	0	0	1	0.128	7.572
BEATLES	0.019	0	0	1	0.135	7.115
D1958	0.066	0	0	1	0.249	n.a.
D1998	0.066	0	0	1	0.249	3.486

Notes: The last number one recording for 2003 topped the charts on 13 December and stayed in that position for nine consecutive weeks, and hence into 2004.
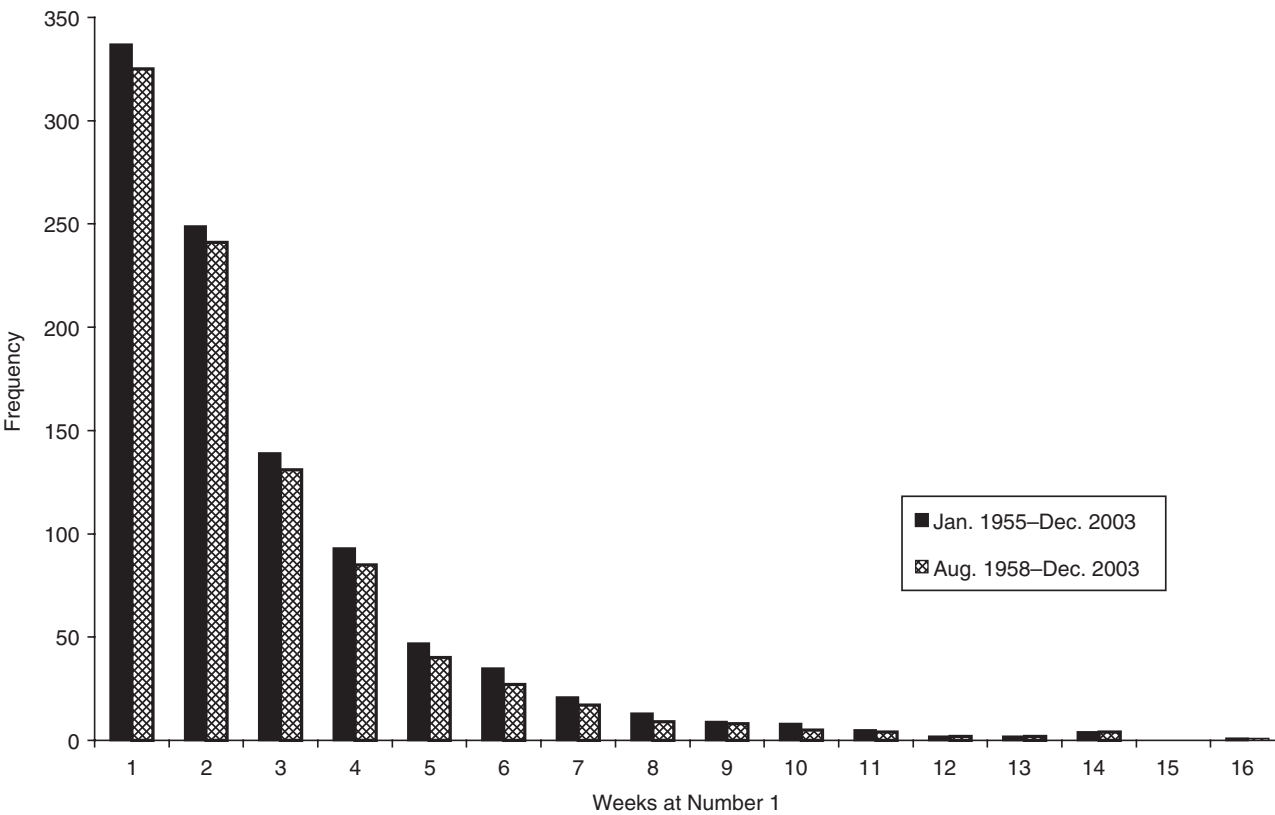

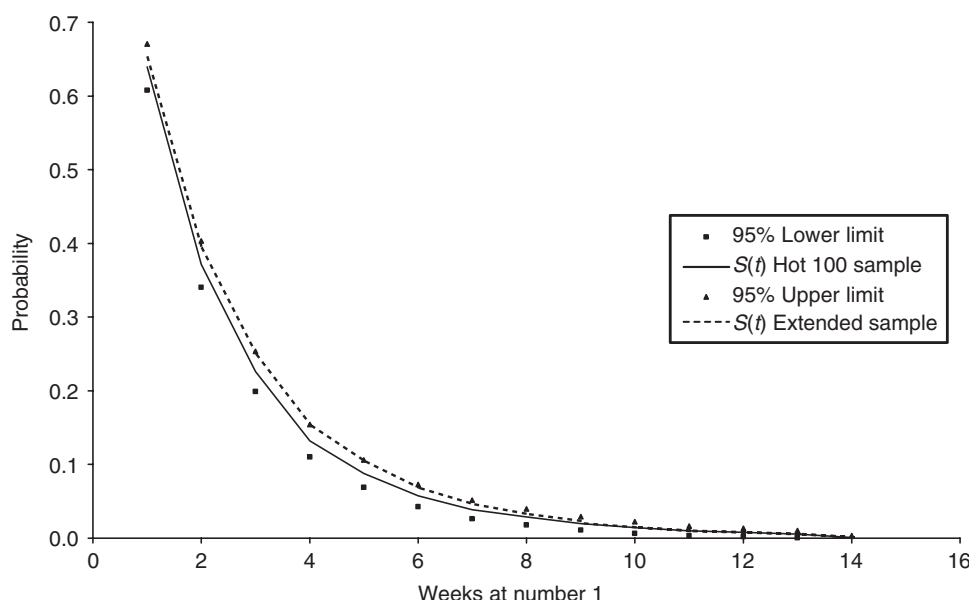**Fig. 1.   Distribution of duration of spells at number one spot**

Fig. 2. Kaplan–Meier survival functions

duration is $F(t) = \Pr[T < t]$, and the corresponding density function (assuming that it exists), is defined as $f(t) = dF(t)/dt$. We will be especially interested in the 'survival function', $S(t) = [1 - F(t)] = \Pr[T \geq t]$, and the 'hazard function', $\lambda(t) = f(t)/S(t)$. The latter is essentially the rate at which spells will be completed at duration t, conditional upon having lasted that long. The functions F, f, S and λ simply provide alternative ways of characterizing the distribution of T.

It is easy to show that $\lambda(t) = -[d\log_e S(t)/dt]$, and one important role of the hazard function is that it provides a basis for defining 'duration dependence'. The underlying random variable is said to exhibit positive (negative) duration dependence at some time, t^*, if $[d\lambda(t)/dt]|_{t=t^*} > 0 \ (< 0)$. Positive (negative) duration dependence implies that the probability that a spell is about to end increases (decreases) with an increase in the spell length. Finally, the so-called 'integrated hazard function', defined as $\Lambda(t) = \int_0^t \lambda(s)\,ds = -\log_e S(t)$, can be useful for specification testing in any parametric survival analysis.

We begin by constructing the well-known Kaplan and Meier (1958) product-limit survival and hazard functions. These provide a valuable graphical aid to model specification, and their construction is especially simple as we have no censoring issues to

deal with. Censoring is therefore ignored in the following discussion. Suppose that we have N observations on T, and these are ordered as $t_1 < t_2 < t_3 < \cdots < t_K$. If there are 'ties' in the data, then $K < N$. In our primary data-set, $N = 965$ and $K = 15$, for example.[6] Let h_j denote the number of completed spells of duration $t_j \ (j = 1, 2, \ldots, K)$. Let n_j be the number of spells that are not completed before duration t_j, so that $n_j = \sum_{i \geq j}^{K} h_i$. Then, a natural estimator of the hazard function is $\hat{\lambda}(t_j) = (h_j/n_j)$, and the corresponding estimator of the survival function is

$$\hat{S}(t_j) = \prod_{i=1}^{j} \left[\frac{n_i - h_i}{n_i}\right] = \prod_{i=1}^{j} (1 - \hat{\lambda}_j), \quad j = 1, 2, \ldots, K$$

This is the Kaplan–Meier product-limit estimator of the survival function, and it will be noted that it is fully nonparametric as no assumptions have been made about the underlying distribution of T.[7] In what follows we have used Greenwood's (1926) formula to estimate the variance of the survival function at each data point, and to construct an asymptotically valid 95% confidence interval.

The estimated Kaplan–Meier survival functions for our primary and extended samples appear in Fig. 2, together with the 95% confidence points for the former function.[8] The associated hazard

[6] The spells are of length one to 16 weeks, but there are no spells of length 15 weeks.

[7] Johansen (1978) shows that the product-limit estimator is also a maximum likelihood estimator.

[8] For an informative recent discussion of the relative merits of the Kaplan–Meier and various parametric estimators, see Meier *et al.* (2004).
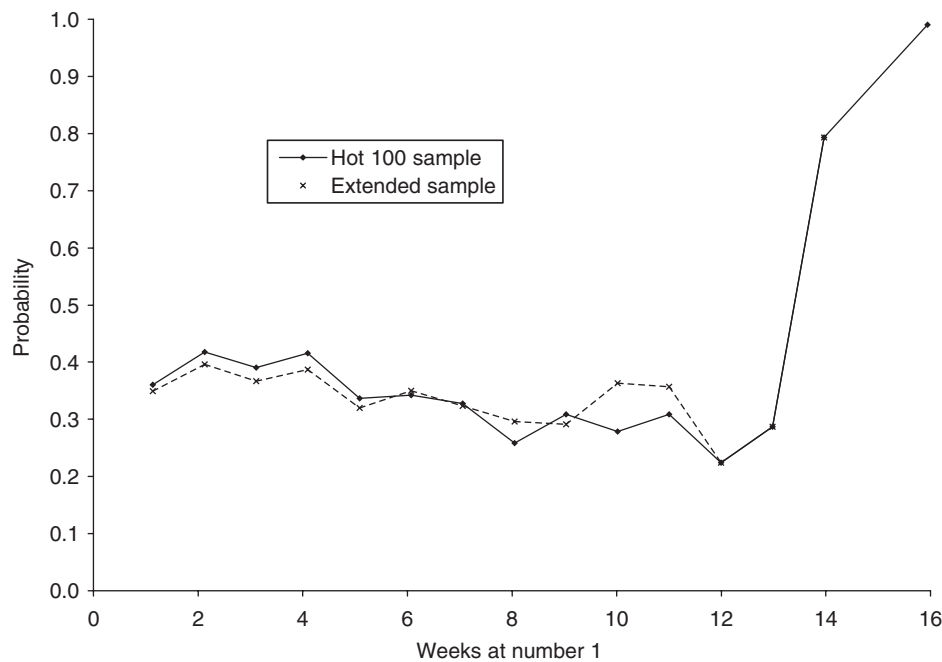
Fig. 3   Kaplan–Meier hazard functions

functions are shown in Fig. 3. As can be seen, the results are quite robust to the choice of sample period. The survival functions have the anticipated general shape, while Fig. 3 suggests that the hazard varies with time. Specifically, there appears to be a slight rise in the hazard over the first two weeks of life at the top of the charts, and then a gradual decline for the next ten weeks. The increase in the Kaplan–Meier hazard for lives in excess of twelve weeks may be an artefact of the small number of observations in this range, but interestingly the shape of the hazard functions in Fig. 3 is remarkably similar to their counterpart in Walls' (1998) study of the Hong Kong cinema industry. One implication of these results is that as we turn to parametric survival models in order to examine the significance of covariates, we will need to consider distributions that allow for a time-varying hazard.

V. Parametric Survival Models With Covariates

Our parametric modelling uses the well-known class of 'accelerated failure time' models (Kalbfleisch and Prentice, 1980), as these avoid certain restrictive features of the 'proportional hazard' model. The term 'accelerated' refers to the fact that an increase in a linear predictor 'accelerates' one's position along the hazard curve. Specifically, we present results based on three underlying parametric distributions – Exponential, Weibull and Log-logistic. The Exponential distribution has a hazard function that is constant over time, but we include it as a basic comparator. We also explored the Generalized Gamma distribution, which nests these (and certain other) distributions, but were unable to maximize the associated likelihood function. Details of the various survival, hazard and integrated hazard functions are given in Table 2. Explanatory variables are incorporated into the accelerated failure time model by setting $\theta = \exp(-\beta' x_i)$, where x_i is a vector of values for the covariates at observation 'i', and β is the corresponding vector of coefficients. In these models the coefficients are the partial derivatives of $\log_e(t)$ with respect to the covariates, so their signs are readily interpreted.[9]

Estimation of the models was undertaken with the LIMDEP econometrics package (Greene, 2002). A general-to-specific modelling strategy was adopted, and the preferred results for each parametric model appear in Table 3. These specifications are based on the significance of the covariates, and results are presented for both the primary and extended samples. The Exponential model is nested within the

[9] This is a feature of the accelerated failure time and the proportional hazard models. In general, the interpretation of the coefficients in survival models may be complicated.

Table 2. Density, survival and hazard functions

	Log-logistic	Weibull	Exponential
$f(t)$	$\theta p (\theta t)^{p-1} / [1 + (\theta t)^p]^2$	$\theta p (\theta t)^{p-1} \exp[-(\theta t)^p]$	$\theta \exp(-\theta t)$
$S(t)$	$1/[1 + (\theta t)^p]$	$\exp[-(\theta t)^p]$	$\exp(-\theta t)$
$\lambda(t)$	$\theta p (\theta t)^{p-1} / [1 + (\theta t)^p]$	$\theta p (\theta t)^{p-1}$	θ
$\Lambda(t)$	$\log[1 + (\theta t)^p]$	$(\theta t)^p$	θt

Notes: p and θ are (positive) shape and scale parameters. The Weibull distribution collapses to the Exponential distribution when $p = 1$.

Weibull model (by setting $p = 1$), so we can readily test the former specification against the latter. The (asymptotically standard normal) 't-statistics' for testing the restriction that $p = 1$ in the Weibull models take the values 10.348 and 10.953 for the samples beginning in 1958 and 1955 respectively, so we strongly reject the Exponential model. Applying likelihood ratio tests (and allowing for the different numbers of covariates in the models), the test statistics have values of 214.862 and 242.696 for the two samples. Asymptotically these statistics are chi-square with 3 and 4 degrees of freedom respectively, so again the Exponential model is clearly rejected. The Log-logistic and Weibull models are nonnested, but the values of Akaike's information criterion (AIC) suggest a preference for the Log-logistic model. The survival and hazard functions for the preferred log-logistic model appear in Fig. 4.

The fact that the integrated hazard function in Fig. 5 is close to a straight line through the origin over much of its range is also supportive of the chosen Log-logistic specification. This model also has the advantage over the Weibull model that its hazard function may be either monotonic decreasing, or it may rise and then fall as the survival time increases. In our estimated Weibull models the shape parameter (p) exceeds unity, implying that the hazard functions increase monotonically over time, and the same result was found by de Vany and Walls (1997) and Walls (1998) for first-run movies in the United states of America and Hong Kong; and by Strobl and Tucker (2000) for record albums on the UK charts. However, those authors did not consider more flexible distributions that allow for nonmonotonic hazards, as we have done here. The hazard functions in Fig. 5 are not significantly different across the two samples, on the basis of the log-rank and generalized Wilcoxon tests.[10]

The results in Table 3 are very robust to the inclusion of the early sample period prior to

the inception of the *Billboard* Hot 100. The only real exception is that the 'Elvis factor' is (essentially) significant in the Weibull model when data prior to 1958 are included. This is consistent with our previous observation that the majority of this artist's number one hits were released in the period from 1955 to 1958. Focusing on the preferred Log-logistic specification, we see that the model predicts a median 'life at the top' of just two weeks – a figure that exactly matches the sample information in Table 1. In this respect, the Log-logistic model is again superior to the other specifications. The shape of the hazard function in Fig. 5 indicates that there is positive duration dependence up to 2.5 weeks, and negative duration dependence thereafter. This implies that the probability that a number one hit is about to decline in the charts increases weekly during its first two or three weeks at the top, but if the recording can survive in the premier spot for that long, its chances of surviving another week improve.

Apparently, a number one hit's 'life-at-the-top' is enhanced significantly if it was recorded by a female solo artist, if it is an instrumental piece, or if it is able to 'bounce back' for a second spell. There have been very few number one hits in the latter two categories in recent years, so this aspect of the results may be of essentially historical interest. The durability of instrumental hits is particularly interesting. There have been 27 instrumental chart-toppers in the USA since 1955, with 23 of these since the advent of the *Billboard* Hot 100. The last such hit was the 'Miami Vice Theme', which reached number one for just one week in November 1985. The average duration for an instrumental Hot 100 chart-topper was 3.13 weeks, compared with 2.76 weeks for other types of number one recordings.

Of course, being Elvis Presley was also a significant advantage, but interestingly The Beatles did not have the same experience according to our preferred model – at least in the USA! These last results are

---

[10] The values for the log-rank and generalized Wilcoxon test statistics are 1.189 and 1.009, with p-values of 0.266 and 0.355 respectively. Each statistic is distributed as chi-square with one degree of freedom under the null hypothesis of equality in survivals.

1884

D. E. Giles

Downloaded By: [University of Victoria] At: 17:25 5 March 2010

Table 3. Maximum likelihood estimation of parametric models

	Exponential				Weibull				Log-logistic			
Start Year:	1958		1955		1958		1955		1958		1955	
Intercept	−12.177	(−1.728)	−11.959	(−1.698)	−18.842	(−5.619)	−19.930	(−6.031)	0.603	(23.771)	0.600	(23.507)
NONCON	0.4340	(1.307)	0.440	(1.306)	0.369	(2.526)	0.371	(2.610)	0.574	(4.560)	0.572	(4.497)
YEAR	0.007	(1.861)	0.007	(1.830)	0.010	(5.920)	0.011	(6.337)				
FEMALE									0.125	(2.204)	0.137	(2.417)
INST					0.267	(2.350)	0.295	(2.654)	0.185	(1.465)	0.240	(1.957)
ELVIS							0.412	(1.644)	0.532	(1.831)	0.496	(2.792)
BEATLES					0.237	(1.763)	0.249	(1.885)				
D1958	n.a.		0.700	(3.113)	n.a.		0.680	(6.137)	n.a.		0.588	(6.956)
D1998	0.342	(1.777)	0.344	(1.791)	0.272	(2.977)	0.264	(2.949)	0.491	(6.296)	0.490	(6.220)
p	1.000	(n.a.)	1.000	(n.a.)	1.479	(31.952)	1.493	(33.170)	2.634	(26.855)	2.614	(27.812)
θ	0.385	(21.776)	0.372	(22.269)	0.345	(35.536)	0.334	(37.731)	0.500	(45.897)	0.481	(47.646)
Median	1.800	(21.780)	1.863	(22.273)	2.264	(35.519)	2.345	(37.744)	1.997	(45.899)	2.078	(47.629)
AIC	1.233		1.234		1.117		1.113		1.014		1.021	

Notes: Asymptotic *t*-ratios appear in parentheses. 'Median' is the estimated median survival time. 'Median' and 'θ' are calculated using the sample means of the data.
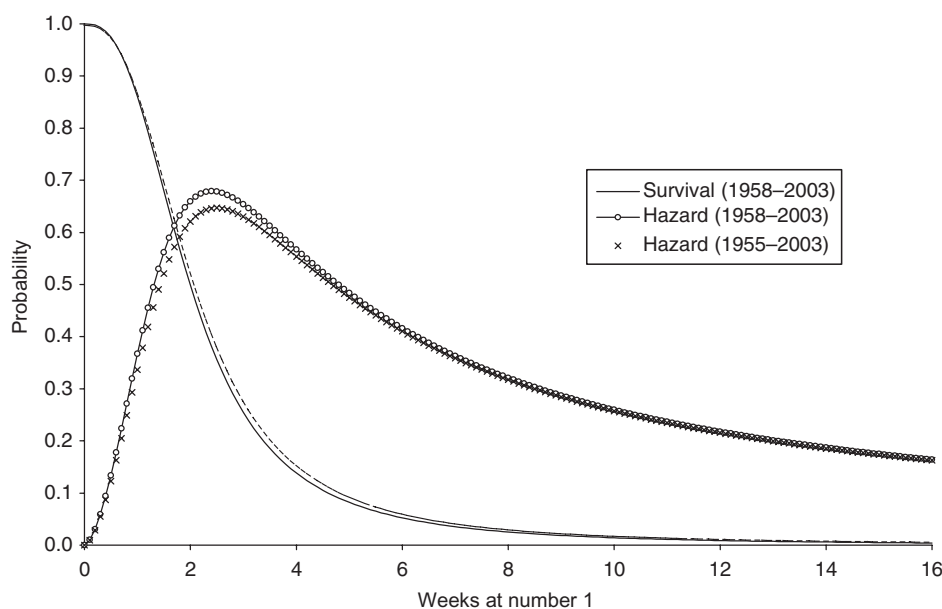
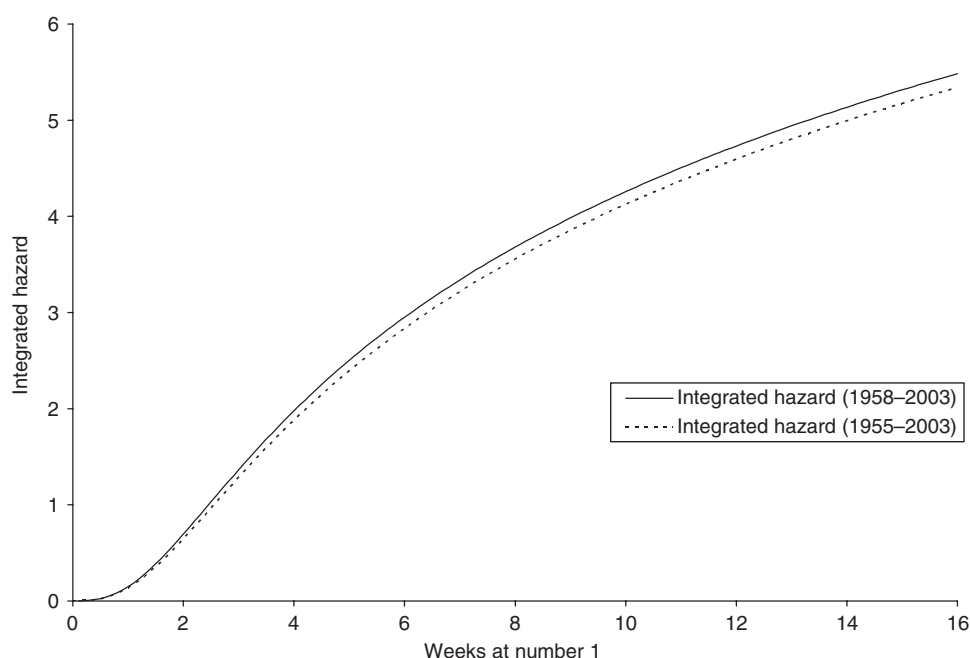**Fig. 4.   Log-logistic survival & hazard functions**

Fig. 5.   Log-logistic integrated hazard functions

supported by a simple inspection of the data. The Beatles' 20 number one hits averaged 2.95 weeks at the top of the Hot 100, while Elvis Presley's 17 chart-toppers averaged 5.29 weeks. The latter figure breaks down to 6.27 weeks for the period prior to the emergence of the Hot 100, and 3.5 weeks thereafter.

All of the covariates in the model are dummy variables, so the estimated coefficients measure the (additive) differential impacts on the logarithm of the durations. Therefore, the estimated marginal effect of being Elvis was to add nearly two weeks to the life of a number one hit on the Hot 100 chart.[11]

---

[11] Note that $\exp(0.532) = 1.7$ weeks. In spite of reported Elvis sightings in recent times, and the predilection of some of his colleagues, the author prefers to refer to 'The King' in the past tense. Contrarian readers should feel free to hum or whistle The Monkees' seven-week number one hit song of December 1966.
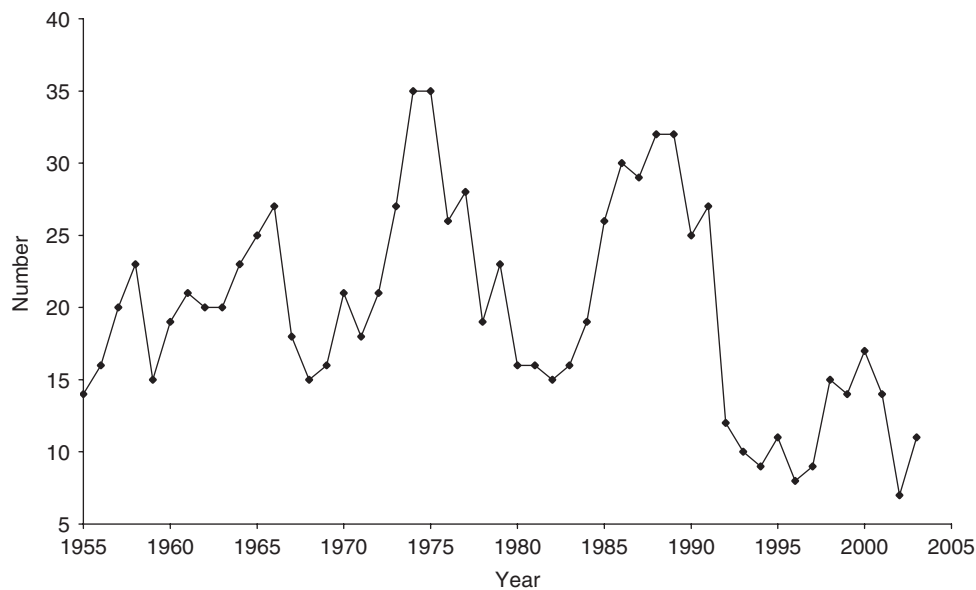
Fig. 6.   Annual number of number one hits

This 'Elvis effect' may reflect a form of market dominance that relates to the degree of heterogeneity of the reputation of the artists at any point in time. One might anticipate that the more homogeneous is the group of leading recording artists, the shorter the likely duration of a number one hit. Star performers such as Elvis Presley and The Beatles increase the heterogeneity of the market. A new release by these artists has a natural advantage over competing releases. The most extreme example of superstar dominance of the Hot 100 chart occurred in 1964. For the week of 4 April, The Beatles had 12 singles in the Hot 100, including the five top spots. This rose to a record 14 singles on this chart in the following week. Judging by the figures in the previous paragraph, this saturation of the market apparently had a bearing on the average life of a Beatles' number one hit.

The important change to the compilation of the *Billboard* Hot 100 that took effect in December 1998 had a significant and positive impact on the duration of number one hits, adding an estimated 1.6 weeks to a spell in the top position, *ceteris paribus*. More durable number one hits imply, of course, that there will be fewer such recordings per year. In fact, the actual average number of number one hits per year from 1955 to 1998 was 20.5, while it was only 12.6 per year between 1999 and 2003. This feature of the data is consistent with the significance of the D1998 dummy variable in all of our models, and that of the YEAR variable in the Exponential and Weibull models. The full distribution of number of chart-toppers per year is shown in Fig. 6, and the declining

trend in annual number of number one hits since 1975 may indicate that there has *not* been an increase in competition in this industry. This is an issue that deserves further investigation. Finally, we note that, although there is no significant difference between the survival functions for our primary and extended samples, the merging of the various charts that existed prior to August 1958 into the Hot 100 chart, did have a significant and positive impact on the duration of number one hits. Apparently, there is a significant difference between the pre-1958 and post-1958 survival functions. We have not computed the former separately due to the relatively short time-span involved.

VI. Conclusions

In this article we have explored some of the characteristics and determinants of the survival of popular music recordings at the number one spot on the *Billboard* Hot 100. It is found that a log-logistic model provides a good account of these survival characteristics, and there is evidence of positive duration dependence for between two and three weeks at the top of the chart, followed by negative duration dependence. Historically, instrumental recordings and recordings by solo female artists have enjoyed significantly longer lives at number one. Not surprisingly, Elvis Presley had a dominant impact on survival times, as did important changes in the formula used to compile the Hot 100 in 1998.

A number of interesting issues remain to be explored. These include dynamic effects such as those associated with an artist's past performance, and the influence of the strategic timing of new releases. Work in progress considers such matters, and also addresses tests for returns to information in this market.

Acknowledgements

References

Answers.com (2005) Available at: http://www.answers.com/topic/billboard-hot-100

Billboard Magazine. (2005) Available at: http://www.billboard.com

Burke, A. E. (1996) The dynamics of product differentiation in the British record industry, *Journal of Cultural Economics*, **20**, 145–64.

Chung, K. H. and Cox, A. K. (1994) A stochastic model of superstardom: an application of the yule distribution, *Review of Economics and Statistics*, **76**, 771–5.

Connolly. M. and Krueger, A. B. (2005) Rockonomics: the economics of popular music, National Bureau of Economic Research Working Paper 11282.

De Haan, D. (2005) *Unofficial Billboard Hot 100 Hitarchiv*. Available at: http://home.planet.nl/~haan0654/hot100.htm

De Vany, A. and Walls, W. (1997) The market for motion pictures: rank, revenue and survival, *Economic Inquiry*, **35**, 783–97.

Greene, W. H. (2002) *LIMDEP Version 8.0: Econometric Modelling Guide*, Vol. 2, Econometric Software, Plainview, NY.

Greenwood, M. (1926) The Natural Duration of Cancer, Reports on Public Health and Medical Subjects, His Majesty's Stationery Office, London Vol. 33, pp. 1–26.

Hamlen, A. W. (1991) Superstardom in popular music: empirical evidence, *Review of Economics and Statistics*, **73**, 729–33.

Johansen, S. (1978) The product limit estimator as product limit estimator, *Scandinavian Journal of Statistics*, **5**, 195–9.

Kalbeisch, J. and Prentice, R. (1980) *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Kaplan, E. L. and Meier, P. (1958) Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457–81.

Kiefer, N. M. (1988) Economic duration data and hazard function, *Journal of Economic Literature*, **XXVI**, 646–79.

Maddison, D. (2005) Increasing returns to information and the survival of broadway theatre productions, *Applied Economics Letters*, **11**, 639–43.

Madonna, (1994) Survival, *Bedtime Stories Album*, Warner Brothers Records, Hollywood, CA.

Meier, P., Karrison, T., Chappell, R. and Xie, H. (2004) The price of Kaplan-Meier, *Journal of the American Statistical Association*, **99**, 890–6.

Musicmatch (2005) Available at: http://www.musicmatch.com

Napster. (2005) Available at: http://www.napster.com.

Nielsen Broadcast Data Systems. (2005) Available at: http://www.bdsonline.com/

Nielsen SoundScan. (2005) Available at: http://www.soundscan.com/

Rhapsody. (2005) Available at: http://www.rhapsody.com/.

Simonoff, J. and Ma, L. (2003) An empirical study of factors relating to the success of broadway shows, *Journal of Business*, **76**, 135–50.

Strobl, E. A. and Tucker, C. (2000) The dynamics of chart successes in the UK. pre-recorded popular music industry, *Journal of Cultural Economics*, **24**, 113–34.

Walls, W. (1998) Product survival at the cinema: evidence from Hong Kong,", *Applied Economics Letters*, **5**, 215–19.

Appendix: Data Definitions

WEEKS	Total number of weeks at 'number one'
YEAR	Year in which number one spot was *first* achieved, 1955 to 2003
NONCON	Dummy variable: $=1$ if weeks were nonconsecutive; $=0$ otherwise
GROUP	Dummy variable: $=1$ if performer was group/duo; $=0$ otherwise
MALE	Dummy variable: $=1$ if performer was male solo singer; $=0$ otherwise
FEMALE	Dummy variable: $=1$ if performer was female solo singer; $=0$ otherwise
INST	Dummy variable: $=1$ if recording was strictly instrumental; $=0$ otherwise
ELVIS	Dummy variable: $=1$ if artist was Elvis Presley; $=0$ otherwise
BEATLES	Dummy variable: $=1$ if recording was by The Beatles; $=0$ otherwise
D1958	Dummy variable: $=1$ prior to 4 August 1958; $=0$ otherwise
D1998	Dummy variable; $=1$ since 5 December 1998; $=0$ otherwise