Jointly published by Akadémiai Kiadó, Budapest *Scientometrics,*
and Springer, Dordrecht *Vol. 62, No. 3 (2005) 329–342*

Modelling citation age data with right censoring

JESÚS BASULTO SANTOS, FRANCISCO JAVIER ORTEGA IRIZO

Department of Applied Economics, University of Sevilla (Spain)

In order to model the variable T (the age of citations received by scientific works) with data elaborated by the Institute of Scientific Information, we have used some of the instruments already developed in the survival models to this type of retrospective analyses in the presence of censored data. This analysis is used because, usually, the citations of ages greater than or equal to 10 years appear added together. For a set of journals related to the field of Applied Economics, we have explored which models fit better among those commonly used. Two different approaches to assess the goodness-of-fit for each selected model have been suggested: an analysis through graphical methods and a formal analysis to estimate the parameters of each model by the method of maximum likelihood estimation with data censored to the right.

Introduction

In bibliometric studies, it is interesting to determine a probabilistic model for a nonnegative random variable T, age of citations received by a journal, where age refers to the time elapsing from the publication of the cited journal until the publication of the citing collection under study, the "citing" source. The data, corresponding to this variable, are usually obtained in retrospective studies (BURRELL, 2001). For example, if we want to obtain the corresponding observations to a certain journal A, we consider a set of journals at a particular date t, and then we look back to the age distribution of citations to the journal A at t, t-1, t-2 etc. (which correspond to the date of publication of the journal A). In this way, we will obtain the distribution of the age of citations T=0,1,2 etc., received by the journal A at the date t of the citing source.

This type of data shows similarity with survival, duration or failure time models (BURRELL, 2002; ORTEGA, 2003). For example, manufactured items, such as mechanical or electronic components, are often subjected to life test in order to obtain information on their endurance. This involves putting items in operation, often in a laboratory setting, and observing them until they fail. It is common here to refer to lifetimes as "failure times", since when an item cease operating satisfactory, it said to have "failed". Similarity, when we observe in a year t_1 that a citation of an article

Received June 1, 2004

Address for correspondence:
JESÚS BASULTO SANTOS
Dpto. de Economía Aplicada I., Facultad de Ciencias Económicas y Empresariales, Universidad de Sevilla
Avda. Ramón y Cajal, nº 1., 41018 Sevilla, Spain
E-mail: basulto@us.es

0138–9130/2005/US $ 20.00

published in the year t_0 has taken place, we can interpret the age of the citation (t_1-t_0) as the time elapsing from the publication of the cited article until the event "to be mentioned" happens, which allows you to use all the tools already developed by the survival models to this type of retrospective analyses.

One of the most important databases about age of citations is elaborated by the Institute of Scientific Information (ISI); but, in this database, all citations greater than or equal to 10 years appear added together, which it supposes a serious limitation for obtaining a model that describes the behaviour of the variable T. This limitation is usually termed "censoring". Essentially, data are said to be "censored" when there are individuals in the sample for which only a lower (or upper) bound on lifetime is available. For example, if after observing a specific electronic component during a determined time L, it still continues operative, we will only know that the time of failure is superior or equal to L, then this data is said to be right censored (LAWLESS, 1982). Similarly, some citations provided by the ISI database have ages greater than or equal to 10, so we can consider them as data censored to the right. We can use the full WOK of the JCR data in order to avoid the problem of the censoring data, although, this requires more time.

Our main purpose is to explore the applicability of different models that best describe the age data of citations with data censored to the right. We have considered the models Log-normal, Weibull and Log-logistic, since they have been used in previous studies (EGGHE & RAO, 1992). We have selected a group of 10 journals, all of them belonging to Applied Economics research field. We have used graphical methods to assess the quality of the fitting Log-normal, Weibull and Log-logistic distributions, as in BURRELL, (2002), and formal statistical methods to compare the empirical and theoretical distributions once the parameters have been estimated by the method of maximum likelihood estimation.

Journals and theirs age of citations

In the *Journal Citation Reports* (JCR) of ISI, we have selected 10 journals related to the field of Applied Economics corresponding to year 2001. The selected journals are given in Table 1.

We have obtained for each journal the age of the received citations in the ISI database (2001). Although the variable T is continuous, we only have annual data since one value of T is evaluated from a citation produced in year t_1 from an article published in the year t_0. Being coherent with the method used by ISI to calculate the median of any distribution, we will assign the value (t_1-t_0) +0.5 as the age of citations because the observed years of citations must fall into $[t_1 - t_0 , t_1 - t_0 + 1]$ and we assume that the distribution of the age of citations into this interval is uniform (BASULTO & ORTEGA, 2002).

Table 1. Journals selected

	Abbreviated journal title	Full journal title
1	ECONOMET THEOR	*Econometric Theory*
2	ECONOMETRICA	*Econometrica*
3	INSUR MATH ECON	*Insurance Mathematics & Economics*
4	J APPL ECONOM	*Journal of Applied Econometrics*
5	J BUS ECON STAT	*Journal of Business & Economic Statistics*
6	J ECONOMETRICS	*Journal of Econometrics*
7	J MATH ECON	*Journal of Mathematical Economics*
8	J ROY STAT SOC A STA	*Journal of the Royal Statistical Society Series A – Statistics in Society*
9	OXFORD B ECON STAT	*Oxford Bulletin of Economics and Statistics*
10	REV ECON STAT	*Review of Economics and Statistics*

Thus, the age of citations of the articles published in 2001 falls into the interval [0,1) where the class mark is 0.5; for the year 2000, the interval is [1,2) with a class mark equal to 1.5, etc. The last year, where the citations are observed is 1992, the interval is [9, 10) with a class mark equal to 9.5. The rest of citations are grouped into the interval $[10,\infty)$. For non censored data, the upper and lower limits of class intervals and the class marks are $[L_{j-1},L_j)$ and t_j respectively. The data set collected of the ten journals is given in Table 2.

Table 2. Distributions of citations for each journal selected

	Age of citation	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	Rest	Total
	Class Interval	[0,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10)	>10	
	Class Mark	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5		
1	ECONOMET THEOR	7	30	36	33	42	51	76	49	28	40	156	548
2	ECONOMETRICA	20	73	127	162	181	244	178	252	329	261	7080	8907
3	INSUR MATH ECON	3	16	23	27	28	11	14	13	15	12	45	207
4	J APPL ECONOM	1	26	33	27	53	74	77	26	73	49	161	600
5	J BUS ECON STAT	6	14	43	92	66	61	139	115	75	127	376	1114
6	J ECONO-METRICS	9	72	142	145	182	295	214	198	138	261	1560	3216
7	J MATH ECON	7	20	22	20	20	40	16	21	10	18	325	519
8	J ROY STAT SOC A STA	10	27	45	29	28	74	64	46	24	30	597	974
9	OXFORD B ECON STAT	0	11	65	19	26	49	50	15	43	114	259	651
10	REV ECON STAT	12	51	116	152	137	186	108	112	111	111	1780	2876

Methodology

Two distinct approaches to assess the goodness-of-fit for each selected model have been suggested: A) an analysis through graphical methods and B) a formal analysis to estimate the parameters of each model by the method of maximum likelihood estimation with data censored to the right.

A) We have used a graphical method similar to the one followed in BURRELL (2002). For each model, we have looked for a transformation of the survivor function $S(t) = P[T>t]$ that has a linear behaviour against the logarithm of variable T. Later, we have represented the nonparametric estimation of the survivor function $\hat{S}(t_j)$ against $\log(t_j)$, being an appropriate model if the points, $(\log(t_j), \hat{S}(t_j)$, (throughout log will denote natural logarithm), come near to a straight line. The main difference with BURRELL (2002) is that we have used the Kaplan-Meier estimator (LAWLESS, 1982).This last estimator is an adaptation of the empirical survivor function for data censored to the right. The Kaplan-Meier estimator is provided in Appendix 2.

The transformation corresponding to the Weibull Model is: $\log[-\log(S(t))] = \beta \log t + \beta \log \lambda$, where $S(t) = e^{-(t\lambda)^{\beta}}$, $\lambda, \beta > 0$ is the survivor function. Therefore, if $\log\left[-\log\left(\hat{S}\left(t_j\right)\right)\right]$ is plotted versus $\log(t_j)$, the resultant graph should be approximately linear if a Weibull model is appropriate.

For Log-normal model, the transformation is:

$$\Phi^{-1}(1 - S(t)) = (1/\sigma)\log t - \mu/\sigma\ \Phi^{-1}(1 - S(t)),$$

where $\Phi(.)$ is the standard Normal distribution function (Φ^{-1} is the inverse of Φ) and $S(t)=1-\Phi((\log t-\mu)/\sigma)$, $\mu \in \mathbb{R}$, $\sigma>0$ is the survivor function. Thus, if $\Phi^{-1}(1 - S(t_j))$ is plotted versus $\log(t_j)$, the resultant graph should be roughly linear if a Log-normal model is appropriate.

Last, the transformation corresponding to the Log-logistic model is: $\log\left((1 - S(t))/S(t)\right) = \gamma \log t + \gamma \log \rho$, where $S(t) = 1/\left[1 + (\rho t)^{\gamma}\right]$, $\rho, \gamma > 0$ is the survivor function. Thus, if $\log\left((1 - \hat{S}(t_j))/\hat{S}(t_j)\right)$ is plotted versus $\log t_j$, the resultant graph should be roughly linear if a Log-logistic model is appropriate.

B) First, we have calculated the likelihood function with data censored to the right (LAWLESS, 1982) for each selected model. These functions are provided in Appendix 3. Second, we have estimated the parameters of each one of the selected models by the method of maximum likelihood estimation with data censored to the right. These estimations are used in section Results, where we compare the empirical and theoretical distribution function in the points L_j. Third, for each model and journal, we have defined errors as differences between the empirical and estimated distribution functions evaluated in the points L_j, j=1,2,...,10. We have calculated the root of the mean squared errors (RMSE's) as the measure of goodness-of-fit. More specific, for each journal, we have defined $e_{ij} = \hat{F}(L_j) - F_i(L_j),\ i = 1,2,3,\ \ j = 1,...,10$, where we will call the cases Log-normal, Weibull and Log-logistic as models 1, 2 and 3 respectively, and we have calculated, $RMSE_i = \sqrt{\sum_{j=1}^{10} e_{ij}^2 n_j \Big/ \sum_{j=1}^{10} n_j}$, which will measure the degree of goodness-of-fit of the i-th model, where n_j is the number of citations of age t_j received by the considered journal.

Results

Graphical analysis

In Figures 1–10 we have given separate plots of $\hat{S}(t_j)$ against $\log(t_j)$, j=1, 2,…,10, for the three models and each journal collected. These figures are provided in Appendix 1. A least squares linear fit is superimposed in each case as well as the R^2-values.

In Figures 1–10 we can appreciate that the R^2-values are very similar for the three models (with a small disadvantage for the Log-normal model). We can also observe how the R^2-values seem to depend more on the journal than on the model selected. For example, the journal OXFORD B ECON STAT presents an anomalous data that causes that the R^2-value diminishes significantly for the three models; journals J APPL ECONOM and J BUS ECON STAT also presents the R^2-values somewhat inferior to the rest of journals. A journal with R^2-value greater than all the other journals is superior in the three models (for example, it is the case of ECONOMETRICA and J ECONOMETRICS). The greatest difference between the models occurs in journals ECONOMET THEOR and J BUS ECON STAT, in which the R^2-values of the Log-normal model are sensibly inferior to Weibull and Log-logistic models.

In summary, from this graphical analysis, it can be deduced that the behaviour of the three models is very similar, but it seems to have a slight difference in favour of the Weibull and Log-logistic models, that could perhaps be due to the different transformations that they are used to make the graphs.

Comparison of empirical and theoretical distribution functions

As we have already indicated in the previous section, we will compare the empirical and theoretical distribution functions of each model and each journal, where we have estimated the parameters of the different models by the method of maximum likelihood estimation. In the case of the Log-normal model, instead of obtaining the estimations of

μ and σ, we have estimated the transformations $\lambda = e^{-\mu}$ and $p = \dfrac{1}{\sigma}$. The estimated parameter as well as their standard errors are provided in Table 3.

Table 3. Parameter estimates and standard errors

	Log-normal		Weibull		Log-logistic	
	λ	p	λ	β	ρ	γ
ECONOMET THEOR	0.1443 (0.0055)	1.2843 (0.0391)	0.1134 (0.0031)	1.8730 (0.0835)	0.1414 (0.0047)	2.3578 (0.0926)
ECONOMETRICA	0.0414 (0.0010)	0.9423 (0.0158)	0.0450 (0.0010)	1.8406 (0.0436)	0.0498 (0.0011)	1.9451 (0.0438)
INSUR MATH ECON	0.1790 (0.0104)	1.2598 (0.0669)	0.1325 (0.0062)	1.6245 (0.1300)	0.1774 (0.0098)	2.1790 (0.1518)
J APPL ECONOM	0.1399 (0.0042)	1.5429 (0.0474)	0.1134 (0.0025)	2.2039 (0.0949)	0.1373 (0.0036)	2.7925 (0.1065)
J BUS ECON STAT	0.1235 (0.0028)	1.5552 (0.0308)	0.1033 (0.0017)	2.2039 (0.0949)	0.1224 (2.8703)	0.0024 (0.0837)
J ECONOMETRICS	0.1012 (0.0018)	1.2605 (0.0212)	0.0851 (0.0012)	1.9364 (0.0487)	0.1026 (0.0015)	2.2746 (0.0506)
J MATH ECON	0.0671 (0.0054)	0.8077 (0.0454)	0.0569 (0.0042)	1.3225 (.1056)	0.0717 (0.0050)	1.4865 (0.1070)
J ROY STAT SOC A STA	0.0723 (0.0037)	0.9063 (0.0336)	0.0626 (0.0028)	1.4954 (0.0817)	0.0771 (0.0033)	1.6891 (0.0824)
OXFORD B ECON STAT	0.1134 (0.0039)	1.3867 (0.0561)	0.0943 (0.0025)	2.1298 (0.0995)	0.1116 (0.0035)	2.4703 (0.1109)
REV ECON STAT	0.0742 (0.0020)	0.9985 (0.0237)	0.0639 (0.0017)	1.6015 (0.0563)	0.0778 (0.0019)	1.8083 (0.0574)

The theoretical distribution functions are $F_i(t) = 1 - S_i(t),\ i = 1,2,3$, where we will call the cases Log-normal, Weibull and Log-logistic as models 1, 2 and 3 respectively. For the Log-normal model, the estimated parameter have been $\lambda = e^{-\mu}$ and $p = 1/\sigma$, therefore we have to consider that new distribution function is:

$$F_1(t) = 1 - S_1(t) = \Phi\left((\log t - \mu)/\sigma\right) = \Phi\left((1/\sigma)\log\left(te^{-\mu}\right)\right) = \Phi\left(p\log(\lambda t)\right)\cdot$$

The empirical distribution function for each journal is obtained from the empirical survivor function, which is calculated using the formula of Kaplan-Meier as indicated in the previous section.

As an example, we offer in Table 4 the complete results for the journal ECONOMET THEOR, where it can be appreciated that the model with a better goodness-of-fit is the Weibull, followed very close by the Log-logistic and Log-normal models.

Table 4. Empirical and theoretical distribution function of ECONOMET THEOR

$L_{j-1} - L_j$	t_j	n_j	$\hat{F}(L_j)$	Log-normal $F_1(L_j)$	Weibull $F_2(L_j)$	Log-logistic $F_3(L_j)$
[0-1)	0.5	7	0.0128	0.0065	0.0168	0.0098
[1-2)	1.5	30	0.0675	0.0553	0.0602	0.0484
[2-3)	2.5	36	0.1332	0.1412	0.1243	0.1169
[3-4)	3.5	33	0.1934	0.2402	0.2035	0.2069
[4-5)	4.5	42	0.2701	0.3376	0.2921	0.3063
[5-6)	5.5	51	0.3631	0.4267	0.3850	0.4042
[6-7)	6.5	76	0.5018	0.5053	0.4774	0.4939
[7-8)	7.5	49	0.5912	0.5733	0.5654	0.5721
[8-9)	8.5	28	0.6423	0.6315	0.6462	0.6384
[9-10)	9.5	40	0.7153	0.6813	0.7179	0.6935
= or > 10		156				
Total		548				
RMSE				0.0372	0.0183	0.0231

For each model and journal, we have calculated the root of the mean squared error as the measure of goodness-of-fit.

The values of the $RMSE_i$ coefficients, i=1,2,3, for each selected journal as well as the average of them are provided in Table 5, where the minimal values for each journal has been emphasized with shaded cells. The last row shows the RMSE average of the journals corresponding to each model.

Table 5. RMSE for models and journals

	Log-normal $RMSE_1$	Weibull $RMSE_2$	Log-logistic $RMSE_3$
ECONOMET THEOR	0.0372	0.0183	0.0231
ECONOMETRICA	0.0056	0.0026	0.0025
INSUR MATH ECON	0.0157	0.0331	0.0194
J APPL ECONOM	0.0318	0.0174	0.0187
J BUS ECON STAT	0.0298	0.0132	0.0189
J ECONOMETRICS	0.0113	0.0148	0.0096
J MATH ECON	0.0114	0.0185	0.0141
J ROY STAT SOC A STA	0.0150	0.0193	0.0153
OXFORD B ECON STAT	0.0465	0.0371	0.0415
REV ECON STAT	0.0082	0.0180	0.0131
Average	**0.0213**	**0.0192**	**0.0176**

As it can be appreciated in Table 5 and already showed in the graphical analysis, the behaviour of the three models is very similar, and it is practically impossible to say that one of them obtains the best goodness-of-fit. In addition, we can see how the disadvantage observed for the Log-normal model in the graphical analysis is not appreciated clearly with the RMSE criterion, since it obtains the best goodness-of-fit in four of the ten journals. Although it is also certain that the Log-normal model presents superior average RMSE for the set, for example in the journals ECONOMET THEOR, J APPL ECONOM and J BUS ECON STAT the values of $RMSE_1$ are superior to $RMSE_2$ and $RMSE_3$. The contrary happens in the model Log-logistic; it has the best goodness-of-fit in two cases and presents the smaller average, since in no journal $RMSE_3$ is sensibly greater than $RMSE_1$ and $RMSE_2$.

Discussion and conclusions

The analysis of survival data is an interesting tool for bibliometric studies. This analysis is used to model the variable T; the age of citations received by scientific works. Unlike survival data, in retrospective citations studies you first observe the citation and thereafter you can look for the date of the cited work. The age of citations is interpretable as the elapsed time since a certain experiment begins (date which the citing source is published) until a certain event happens (the work is cited).

In situations where we are looking for the best model from a set of models, it is better to have a data set in which censored observations do not exist. However, this is not possible if we want to use the databases of ISI, where citations of age which are greater than or equal to 10 years are added together. Consequently we must take advantage of some instruments previously developed for the duration models.

The graphical analysis is a simple method to select the most appropriate model from a set of models. This procedure shows an acceptable and a very similar goodness-of-fit for the three models (Log-normal, Weibull and Log-logistic), with a slight disadvantage to the Log-normal model.

The comparison of empirical and theoretical distribution functions practically obtains the same conclusions as in the graphical analysis, since the RMSE's values are very similar for all journals.

If we rely on the criterion of the average RMSE for the set of journals, our conclusion is that the Log-logistic model is the best fit for this particular set of data and that the Log-normal model shows the worst behaviour. In addition, the Log-logistic model does not present a value of RMSE sensibly greater than other models in any journals. The exceptions are the RMSE's values of the Log-normal model in journals ECONOMET THEOR, J APPL ECONOM and J BUS ECON STAT. On the other hand, the Weibull model presents a value of RMSE sensibly greater than the other models for the journal INSUR MATH ECON. We must consider small differences amongst the three models when reading our conclusions.

Finally, we emphasize the importance of models studies in the present paper, mainly in citation studies of obsolescence. For example, in the ISI databases, the calculation of the median (or period of time during which a half of citations was received by the cited journal) is not possible without any models when the percentage of censored observations is greater than 50% (BASULTO & ORTEGA, 2002; ORTEGA, 2003).

References

BASULTO, J., ORTEGA, F. J. (2002), Modelización de la antigüedad de las citas en la literatura científica con datos censurados a la derecha, *Revista Española de Documentación Científica,* 25 : 141–150.

BURREL, Q. L. (2001), Stochastic modelling of the first-citation distribution, *Scientometrics,* 52 : 3–12.

BURREL, Q. L. (2002), Modelling citation age data: Simple graphical methods from reliability theory, *Scientometrics*, 55 : 273–285.

EGGHE, L., RAO, I. K. RAVICHANDRA (1992), Citation age data and the obsolescence functions: Fits explanations, *Information Processing & Management,* 28 : 210–217.

LAWLESS, J. F. (1982), *Statistical Models and Methods for Lifetime Data,* New York: John Wiley & Sons.

ORTEGA, F. J. (2003), Propuesta de Mejora del Índice Agregado de Impacto. Aplicación a la Valoración de Sexenios de Investigación, *Revista Española de Documentación Científica,* 26 : 418–432.

Appendix 1
Figures

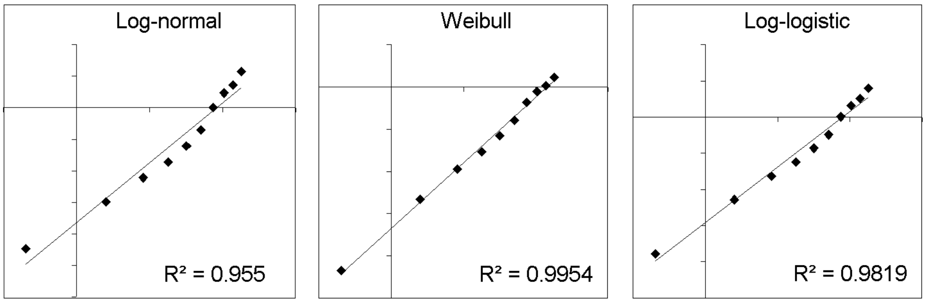

Figure 1. Plots for ECONOMET THEOR with Log-normal, Weibull and Log-logistic models

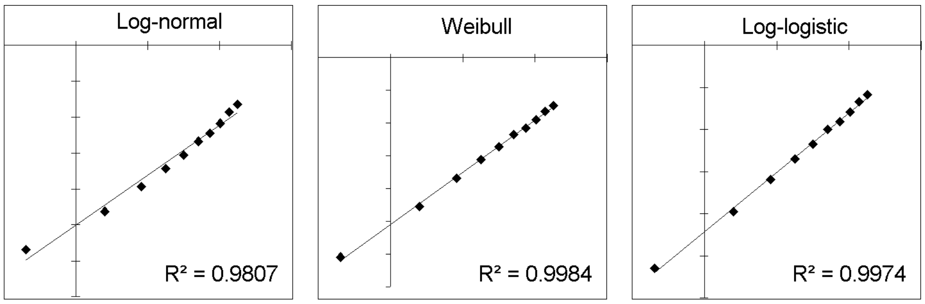

Figure 2. Plots for ECONOMETRICA with Log-normal, Weibull and Log-logistic models

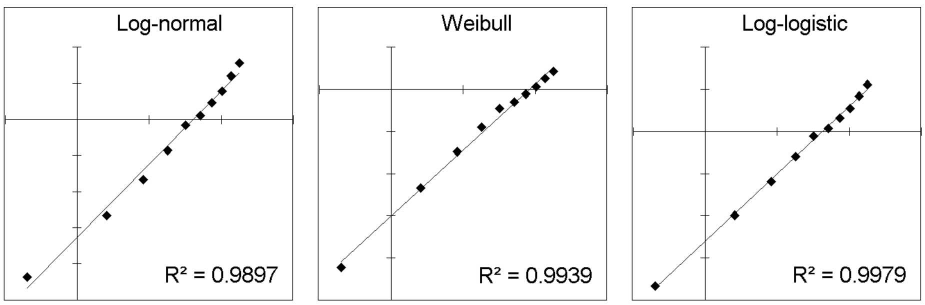

Figure 3. Plots for INSUR MATH ECON with Log-normal, Weibull and Log-logistic models

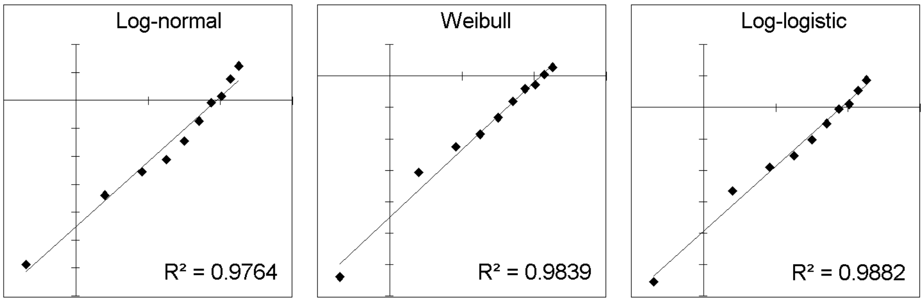

Figure 4. Plots for J APPL ECONOM with Log-normal, Weibull and Log-logistic models

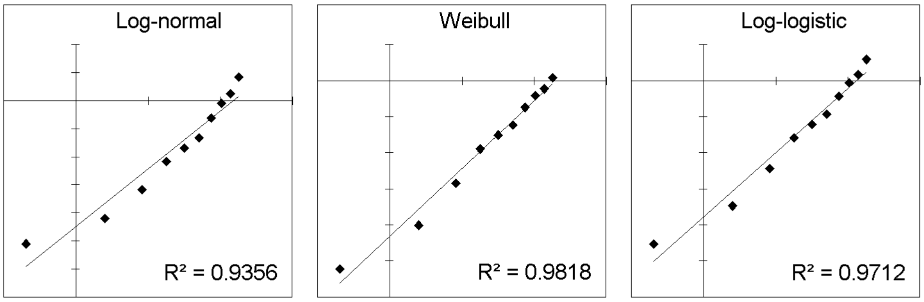

Figure 5. Plots for J BUS ECON STAT with Log-normal, Weibull and Log-logistic models

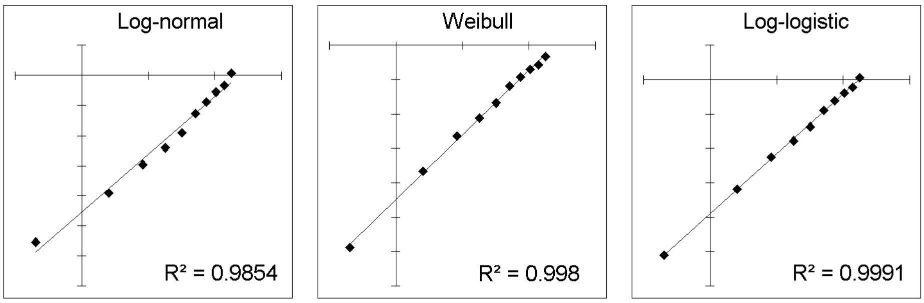

Figure 6. Plots for J ECONOMETRICS with Log-normal, Weibull and Log-logistic models

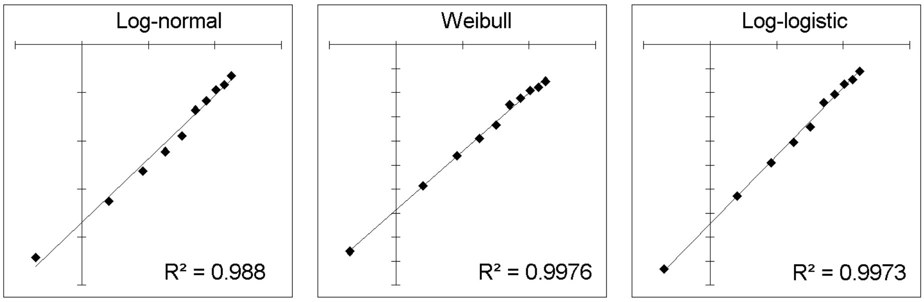

Figure 7. Plots for J MATH ECON with Log-normal, Weibull and Log-logistic models

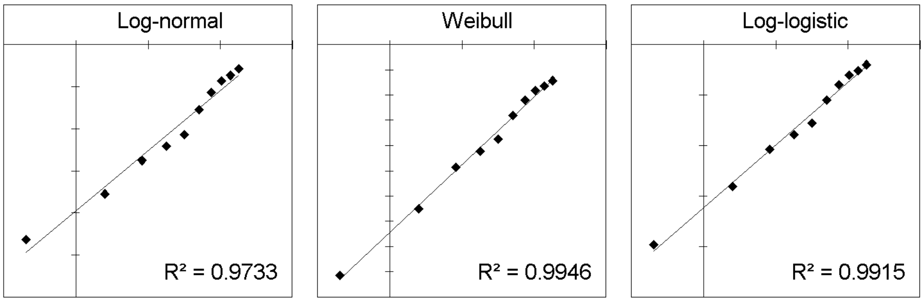

Figure 8. Plots for J ROY STAT SOC A STA with Log-normal, Weibull and Log-logistic models

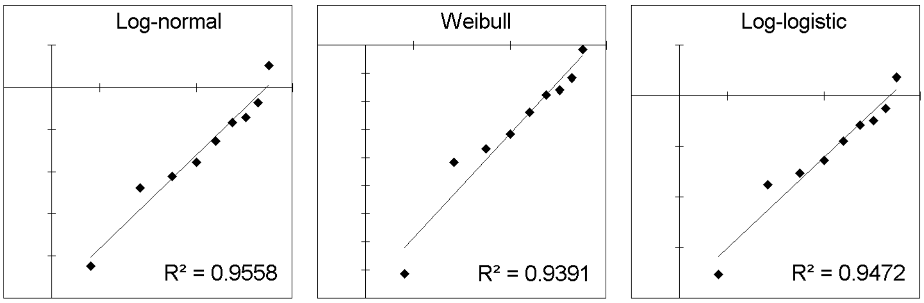

Figure 9. Plots for OXFORD B ECON STAT with Log-normal, Weibull and Log-logistic models

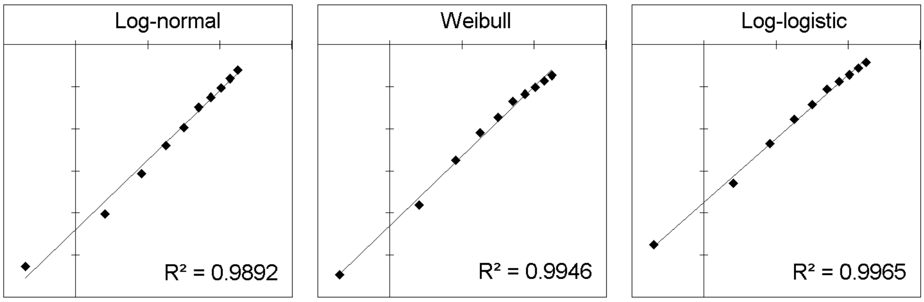

Figure 10. Plots for REV ECON STAT with Log-normal, Weibull and Log-logistic models

Appendix 2
Kaplan-Meier estimator

Suppose that there are observations on n citations and that there are k ($k \leq n$) different dates $t_1 < t_2 < ... < t_k$ where the citations occur. We let n_j the number of citations at t_j. In addition, there are also censoring data C_i for individuals whose citations are not observed. The Kaplan-Meier estimate of S(t) is defined as

$$\hat{S}(t) = \prod_{j:t_j<t} \frac{N_j - n_j}{N_j},$$

where N_j is the number of citations uncensored just prior to t_j. (LAWLESS, 1982, p. 72).

Appendix 3
Log-likelihood function with right censoring

Suppose that the variable T has pdf $f(t,\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}^k$ and Θ is an open set. We let $t_1, t_2, ..., t_k$ the (n-r) citations observed and $C_{n-r+1}, ..., C_n$ the r citations censored. The likelihood function is:

$$L = \prod_{i=1}^{n-r} f(t_i;\theta) \prod_{i=n-r+1}^{n} S(C_i;\theta),$$

where S(t) is the survivor function of T. (LAWLESS, 1982, p. 36).

For the data under study, where $C_i = C = 10,\ \forall i = n-r+1,...,n$, the likelihood function is:

$$L = S(C;\theta)^r \prod_{i=1}^{n-r} f(t_i;\theta),$$

and thus the log-likelihood function is:

$$\ell = r\log\left(S(C;\theta)\right) + \sum_{i=1}^{n-r}\log\left(f(t_i;\theta)\right).$$

We will call the cases Log-normal, Weibull and Log-logistic models as models 1, 2 and 3, respectively. The pdf's and survivor functions are:

$$f_1(t,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,t\sigma} e^{-\frac{1}{2\sigma^2}(\log t-\mu)^2} \qquad ,S_1(t,\mu,\sigma) = 1-\Phi\left((\log t-\mu)/\sigma\right) \qquad ,\ \mu\in\mathbb{R},\ \sigma>0$$

$$f_2(t,\lambda,\beta) = \lambda\beta\left(\lambda t\right)^{\beta-1} e^{-(\lambda t)^\beta} \qquad ,S_2(t,\lambda,\beta) = e^{-(\lambda t)^\beta} \qquad ,\ \lambda,\beta>0$$

$$f_3(t,\rho,\gamma) = \gamma\rho\frac{\left(\rho t\right)^{-1+\gamma}}{\left(1+(\rho t)^\gamma\right)^2} \qquad ,S_3(t,\rho,\gamma) = \frac{1}{1+\left(\rho t\right)^\gamma} \qquad ,\ \rho,\gamma>0$$

where $\Phi(\cdot)$ is the standard Normal distribution function.

We observe that the pdf and survivor functions of the Log-normal model for the parameters $\lambda = e^{-\mu}$ and $p = 1/\sigma$ are respectively:

$$f_1(t,\lambda,p) = \frac{1}{\sqrt{2\pi}\,t} pe^{-p^2\log(\lambda t)} \qquad ,S_1(t,\lambda,p) = 1-\Phi\left(p\log(\lambda t)\right) \qquad ,\lambda,p>0.$$

Thus, the log-likelihood for the three models are respectively:

$$\ell_1 = r\log\left(1-\Phi\left(p\log(\lambda C)\right)\right) + (n-r)\log p - \sum_{i=1}^{n-r}\log t_i - \sum_{i=1}^{n-r}\log(\lambda t_i).$$

$$\ell_2 = -r(\lambda C)^\beta + (n-r)\log(\lambda\beta) + (\beta-1)\sum_{i=1}^{n-r}\log\left(\lambda t_i\right) - \lambda^\beta\sum_{i=1}^{n-r} t_i^\beta.$$

$$\ell_3 = -r\log\left(1+(\rho C)^\gamma\right) + (n-r)\log(\gamma\rho) + (\gamma-1)\log\gamma + (\gamma-1)\sum_{i=1}^{n-r}\log t_i - 2\sum_{i=1}^{n-r}\log\left((1+\rho t_i)^\gamma\right).$$