

Referee Report on “Investigating the Exponential Age Distribution of Firms”

The author’s main contribution is to assess whether the age distribution of firms is exponential using a variety of publicly available datasets. This is in contrast to the (large) literature which focuses on the size distribution of firms, and as such, is a potentially useful and interesting research question. Moreover, those studies that have looked at the related issue of survival rates have tended to be industry specific. Thus, the author is potentially able to complement existing research by analyzing aggregate level data.

Overall, I think that the paper could be a nice contribution but as it stands, I think there are various (and to my mind, significant) points that need addressing. Also, I find the paper difficult to follow in its current form because it jumps all over the place in terms of data analysis (young firms from the U.S.; the world’s oldest firms, and finally the airline industry), with no clear integration.

Section 1 (Introduction): Although potentially interesting, I think that the paper needs to be better motivated. You write that the reason for little prior research into this question is “presumably because of the difficulty in obtaining data”. I do not find this to be a motivating reason for why one should conduct research on the age distribution of firms. In other words, I would want you to better set out why one should investigate the age distribution of firms, especially as you do state it is a useful summary representation of the structure of industries.

Section 2 (Theoretical Modeling): My first question here is how does the analysis in this paper differ from Coad (2010)? You need to make it clearer what this paper adds or does differently. Secondly, you assume that the distribution of firm age is exponentially distributed. Given that your results in Section 2 are rather more qualitative (i.e. eyeballing a graph) than quantitative, I currently remain less than convinced. I think the paper would benefit from comparing the exponential with other distributions in order to assess which one fits the data best or at least discuss why you do not do this. This is especially relevant as you discuss in your introduction how various scholars have used the lognormal, the Pareto or the generalized beta as “best fits” for the observed empirical firm size distribution.

Section 3 (Young Firms): Given that your results appear to indicate that the very young and the very old firms do not appear to be exponentially distributed, you then attempt to assess why. I think the results on “exit hazards” for younger firms are interesting, and I am convinced this is true. Are exit hazard rates constant for middle-aged/older firms?

Section 4 (Old Firms): It is interesting that the data are nowhere close to what the exponential would predict, and I do like that you show the data, given that as you say, many studies might exclude these data points as “outliers”. However, I am not convinced by your argument that

these firms have survived this long because they are family-owned. You have around 1740 data points according to footnote 11 (how many are family owned?)

Section 5 (Sector Level Analysis): You state that you choose the airline industry because you “consider it to be a special case that is particularly likely to show a multi-modal age distribution”. It is not clear why you include this section at all. What does it add? It shows that in certain industries, the exponential does not do well, but your analysis does not follow smoothly from the previous section.

Minor Comments

1. At the end of your introduction, you state “the aim of this paper is to draw attention to the exponential age distribution”. You need to remove the word “exponential”. Surely, the aim is to assess whether the exponential is a good fit or not.
2. Page 3, 3rd paragraph: correct the typo “even in this situations....”
3. Page 5: you mean Figure 1, not Figure 2?
4. I find it extremely disconcerting to be faced with all these different datasets and charts that are using data from other studies. I would want to see a brief summary of the various datasets in an Appendix at the least that can help the reader understand the various data sources. E.g. Figure 2 (Coad and Tamvada, 2008); Figure 3 (Segarra et al, 2008); Figure 4 (Bottazzi et al, 2008); Figure 5 (many studies)