ECON 8602: Problem Set #1
(due Tue., 11/13)

I have downloaded data for 1997 from the Small Business Administration (http://www.sba.gov/advo/stats/data.html) on all US firms classified by their estimated receipts. I have put it into an excel spreadsheet which you can download from my web page.

1. Read the article by Axtell in Science magazine (you can download the pdf file from my web page). Using the data I have given you, see if you replicate Axtell's result (Fig. 2) using all firms. Explain exactly your procedure in going from the raw data to the figure (I did not find his explanation totally clear). You can try to follow what Axtell did, or proceed using what you think is a good logical approach, but be precise.

2. Do the same as above for manufacturing and separately for one (or all) 2-digit industry within manufacturing (i.e. the 4-digit industries whose SIC begins with 2 or 3 and ends with 00). Beware! The industry disaggregation is tricky. They include the firm in an industry if it has any establishment primarily in that industry. But the size of the firm remains the total size of the firm (including its establishments in other industries). (You can read about this issue in the documentation provided at the SBA web site.)

3. Do you find your evidence convincing in favor of the Pareto distribution rather than the log normal distribution?

4. Write a computer program that simulates a growth process for a population of firms. Try to get it to replicate the Pareto distribution claimed by Axtell (and perhaps corroborated by you). For hints, see the Axtell paper, and the paper by Gabaix. Check that the sizes of your largest simulated firms follow Zipf's law. (Stanley et. al. show that the largest firms do not conform to this law, and Axtell admits that the Zipf distribution seems off for the very largest firms).

5. Using your computer program, generate a balanced sample of 1000 firms over a 6-year period. Can you produce a scatter plot that matches Figure 1 in Bronwyn Hall's paper?