The Size Distribution of Innovations Revisited: an Application of Extreme Value Statistics to Citation and Value Measures of Patent significance

Gerald Silverberg and Bart Verspagen
MERIT, Maastricht University
gerald.silverberg@merit.unimaas.nl.

This paper focuses on the analysis of size distributions of innovations, which are known to be highly skewed. We use patent citations as one indicator of innovation significance, constructing two large datasets from the European and US Patent Offices at a high level of aggregation, and the Trajtenberg (1990) dataset on CT scanners at a very low one. We also study self-assessed reports of patented innovation values using two very recent patent valuation datasets from the Netherlands and the UK, as well as a small dataset of patent license revenues of Harvard University. Statistical methods are applied to analyse the properties of the empirical size distributions, where we put special emphasis on testing for the existence of 'heavy tails', i.e., whether or not the probability of very large innovations declines more slowly than exponentially. While overall the distributions appear to resemble a lognormal, we argue that the tails are indeed fat. We invoke some recent results from extreme value statistics and apply the Hill (1975) estimator with data-driven cut-offs to determine the tail index for the right tails of all datasets except the NL and UK patent valuations. On these latter datasets we use a maximum likelihood estimator for grouped data to estimate the tail index for varying definitions of the right tail. We find significantly and consistently lower tail estimates for the returns data than the citation data (around 0.6-1 vs. 3-5). The EPO and US patent citation tail indices are roughly constant over time, but the latter estimates are significantly lower than the former. The heaviness of the tails, particularly as measured by value indicators, we argue, has significant implications for technology policy and growth theory, since the second and possibly even the first moments of these distributions may not exist.

JEL Codes: C16, O31, O33

Keywords: returns to invention, patent citations, extreme-value statistics, skewed distributions, heavy tails.