The Data

The analysis in the paper utilizes productivity statistics, labor market variables, and business failure data in addition to the technology series. The sources for each are discussed below.

A.1. Annual Employment and Productivity Statistics 1909-49

We rely on data from Kendrick (1961) for output per worker in the private non-farm economy as well as for the manufacturing and transportation sectors. Data on aggregate employment and the aggregate unemployment rate come from the *Historical Statistics of the United States Millennial Edition* (Table Ba 470-477) which are, in turn, taken from Weir (1992) whose numbers represent a re-working of those originally developed by Lebergott (1964). GNP per person (1947 constant dollars) comes from the Economic Almanac of the National Conference Board while Solow (1957) and Goldsmith (1956) are the sources for data required to compute TFP for the private non-farm economy. Our TFP series are calculated using the equation:

$$\frac{\Delta TFP}{TFP} = \frac{\Delta\left(\frac{GNP}{L}\right)}{\left(\frac{GNP}{L}\right)} - \omega_K \frac{\Delta\left(\frac{K}{L}\right)}{\left(\frac{K}{L}\right)},$$

where GNP is private non-farm GNP, L is the number of hours worked, K is the capital stock, and ω_K is Solow's (1957) share of property in income.

A.2. Monthly Productivity measures and Labor Market variables

We make use of two labor productivity measures. At the aggregate level, we define productivity in a given month to be the Federal Reserve Board's index of industrial production divided by the number of employees working in nonagricultural establishments (available from the NBER http://www.nber.org/databases/macrohistory/). For manufacturing, we use the ratio of industrial production to employment in the sector.

We use 10 series taken from the NBER's Macrohistory Database to estimate the relationship between detailed labor market variables and technical change at monthly frequencies. The series spanning January 1929 – December 1949 include Employees in Nonagricultural Establishments and Total Production Worker Employment in Manufacturing

produced by the Bureau of Labor Statistics (BLS) and the unemployment rate as reported by the National Conference Board. For the period June 1920 to December 1949, we have data on the index of total production worker-hours in manufacturing drawn from the National Industrial Conference board, the NBER and the Bureau of the Census, vacancies estimated from Help-Wanted Ads in newspapers produced by W.A. Berridge of Metropolitan Life Insurance Company, and, finally, labor turnover in manufacturing – the quit rate, layoff rate, discharge rate, total separation rate, and gross accession rate –collected by the BLS. The turnover data represent unweighted medians of rates derived from the reports of samples of companies in non- seasonal industries and are based on the average number of wage earning employees on the payroll.

A.3. The Book Based indicators

The annual book based indicators are created from information contained in the Machine Readable Cataloguing records (MARC records) from the Library of Congress (LOC). Each MARC record contains information on the type of publication (for example, a new edition or a new title), place, language, library classification code, and a list of subjects treated in the volume. The data is the same as that used in the paper Measuring our Ignorance, one book at a time, by Alexopoulos and Cohen (JME, 2009), and focuses on new titles in different fields of technology (T and subgroups of T- TK, TJ, TS, and the automotive titles (primarily in a subgroup of TL)) published in English in the US between 1909 and 1949.[1] We then exclude all books that include history as a descriptor (either in the title or in subject keywords).

We also created a series of indicators based on the following data drawn from *PW:* date, place of publication, language, and type (new titles excluding new editions). Included in these indicators are all technical publications encompassed by the Dewey classification numbers 600-609, 620-629, and 660-699. These classifications were chosen so that the materials closely resemble those found in the LOC's T classification used for the annual indices.

[1] For more detailed descriptions of the Library of Congress' classification, see https://www.loc.gov/catdir/cpso/lcco/.