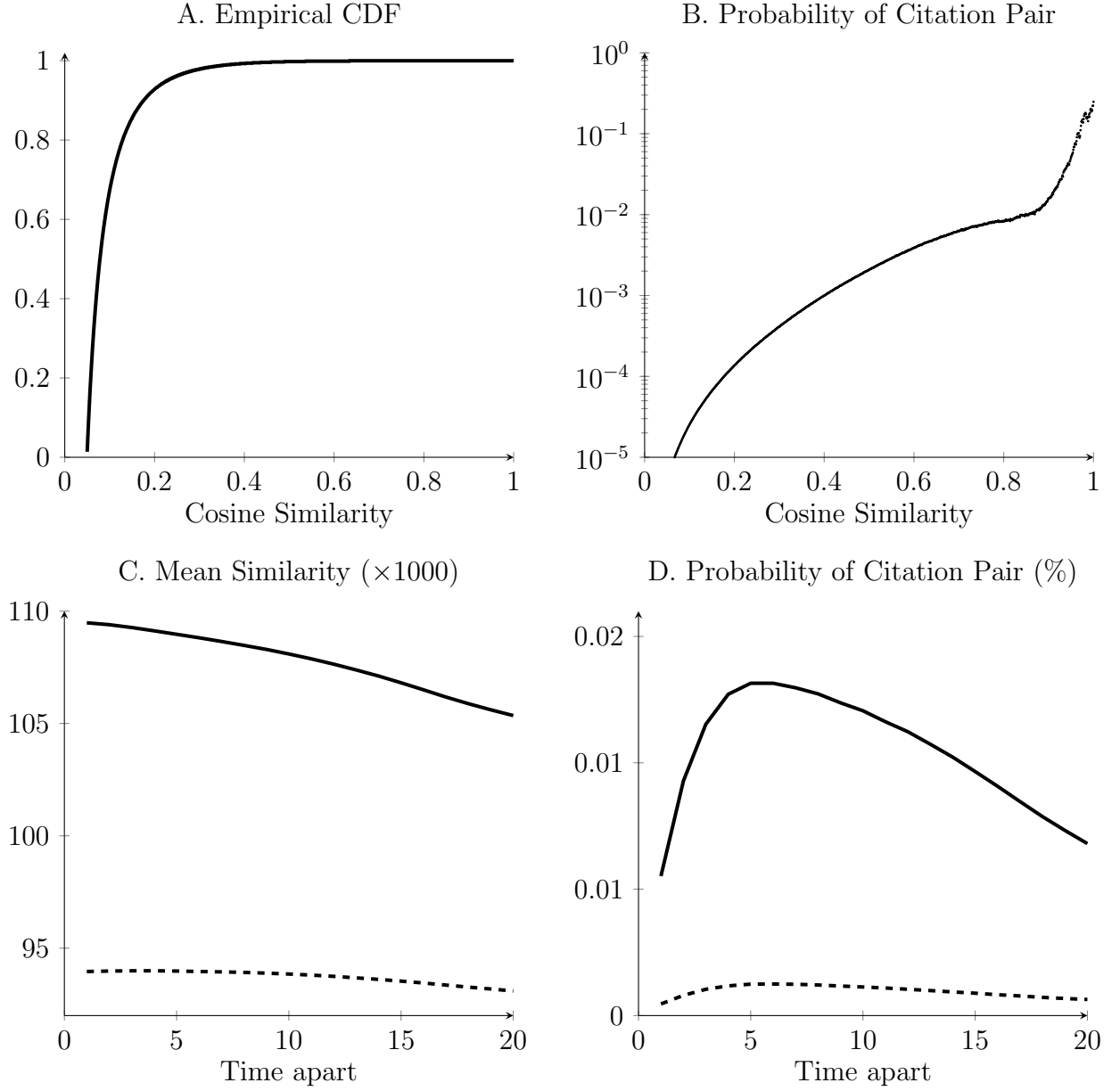


Figure 1: Pairwise similarity and citation linkages



Panel A plots the empirical CDF of our similarity measure $\rho_{i,j}$ across patent citation pairs. Panel B plots the conditional probability that patent j cites an earlier patent i as a function of the text-based similarity score between the two patents, $\rho_{i,j}$, computed in equation (??) in the main text. For computational reasons, we exclude similarity pairs with $\rho_{i,j} \leq 0.5\%$. Figure uses data only post 1945, since citations were not consistently recorded prior to that year. We use data only post 1945, since citations were not consistently recorded prior to that year. Panel C plots the mean similarity across patent pairs i and j as a function of the distance in filing years between the two patents, and whether the two patents belong in the same tech class or not. Panel D performs the same exercise for the mean number of citations across pairs. Similarity refers to the text-based similarity score between the two patents, $\rho_{i,j}$, computed in equation (??) in the main text. For computational reasons, we exclude similarity pairs with $\rho_{i,j} \leq 5\%$.

Figure 2: Pairwise similarity and citation linkages

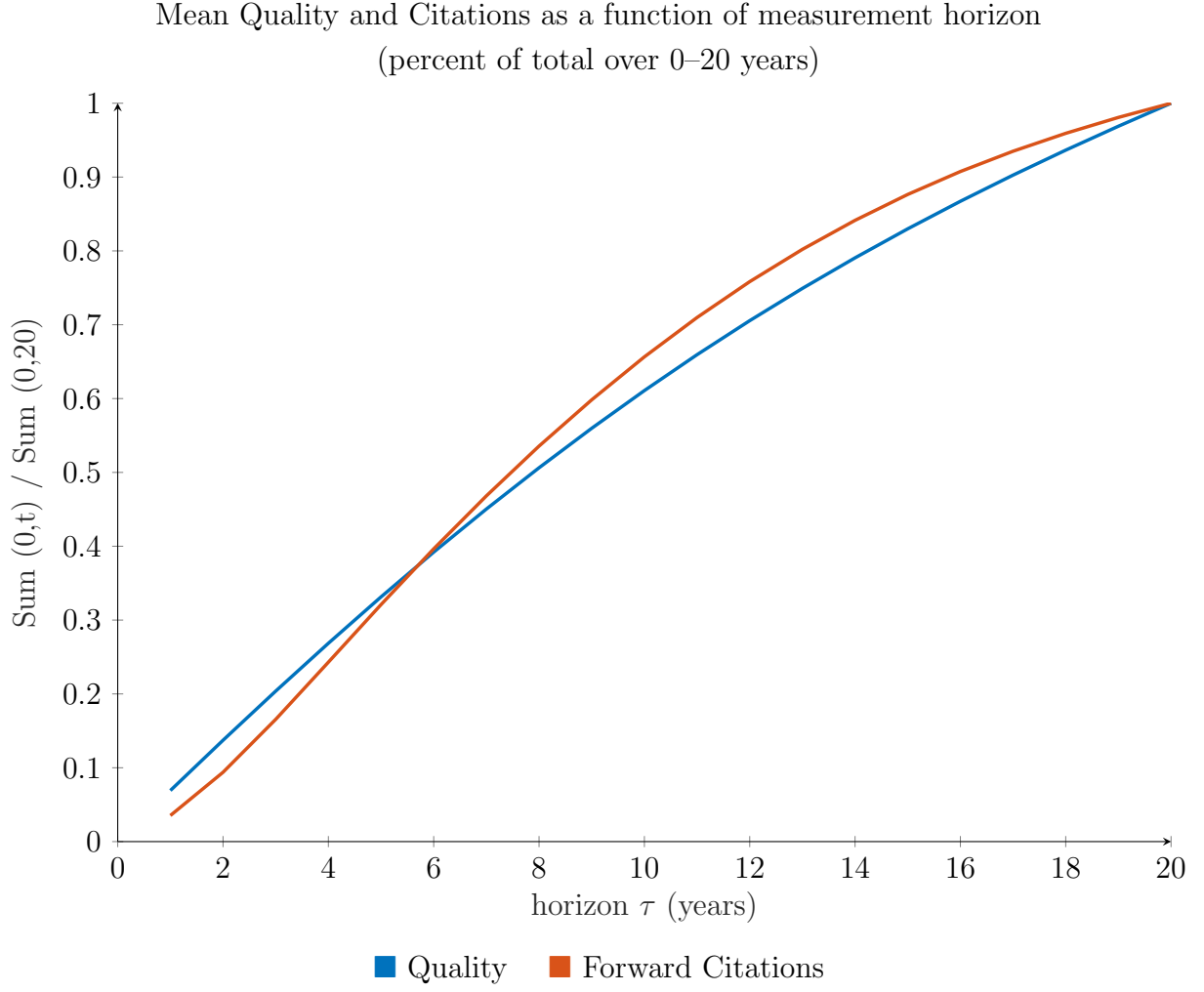


Figure examines the speed at which information about the quality of the patent is reflected in our quality measure and in forward citations. Specifically, we plot the mean across patent pairs of $x_{0,\tau}$ where x refers to either our quality indicator or forward citations measured over τ years subsequent to the patent, scaled by $x_{0,20}$.

Figure 3: Similarity Networks, Examples

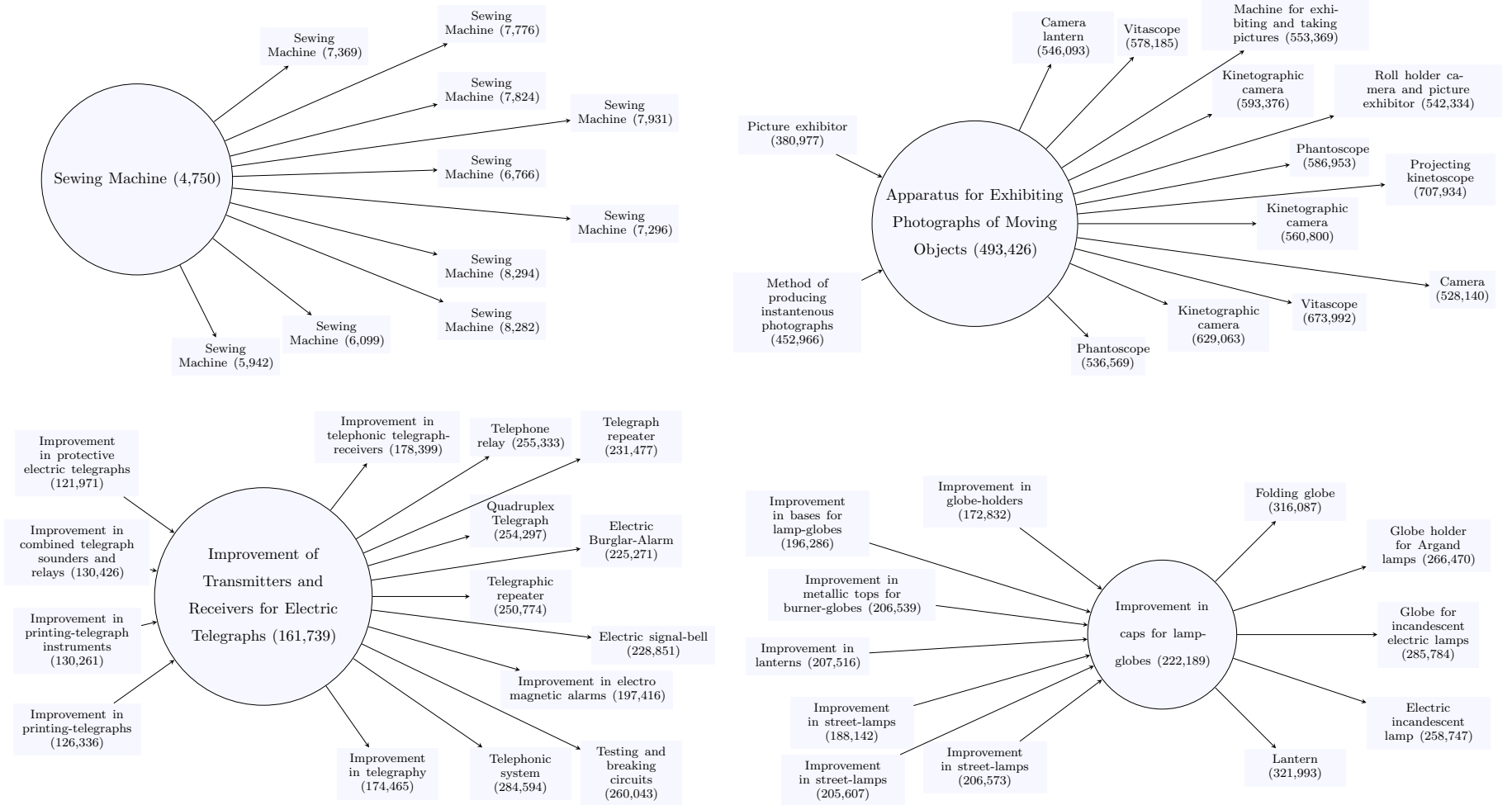


Figure displays the similarity network for four patents: the patent for the first sewing machine (top left); one of the earlier patents for moving pictures (top right); one of the early patents that led to the telephone (bottom left) and a randomly chosen patent from the 1800s (bottom right). In plotting the similarity links, we restrict attention to patents pairs filed at most five years apart and with a cosine similarity greater than 50%.

Figure 4: Measuring Patent Quality, Examples

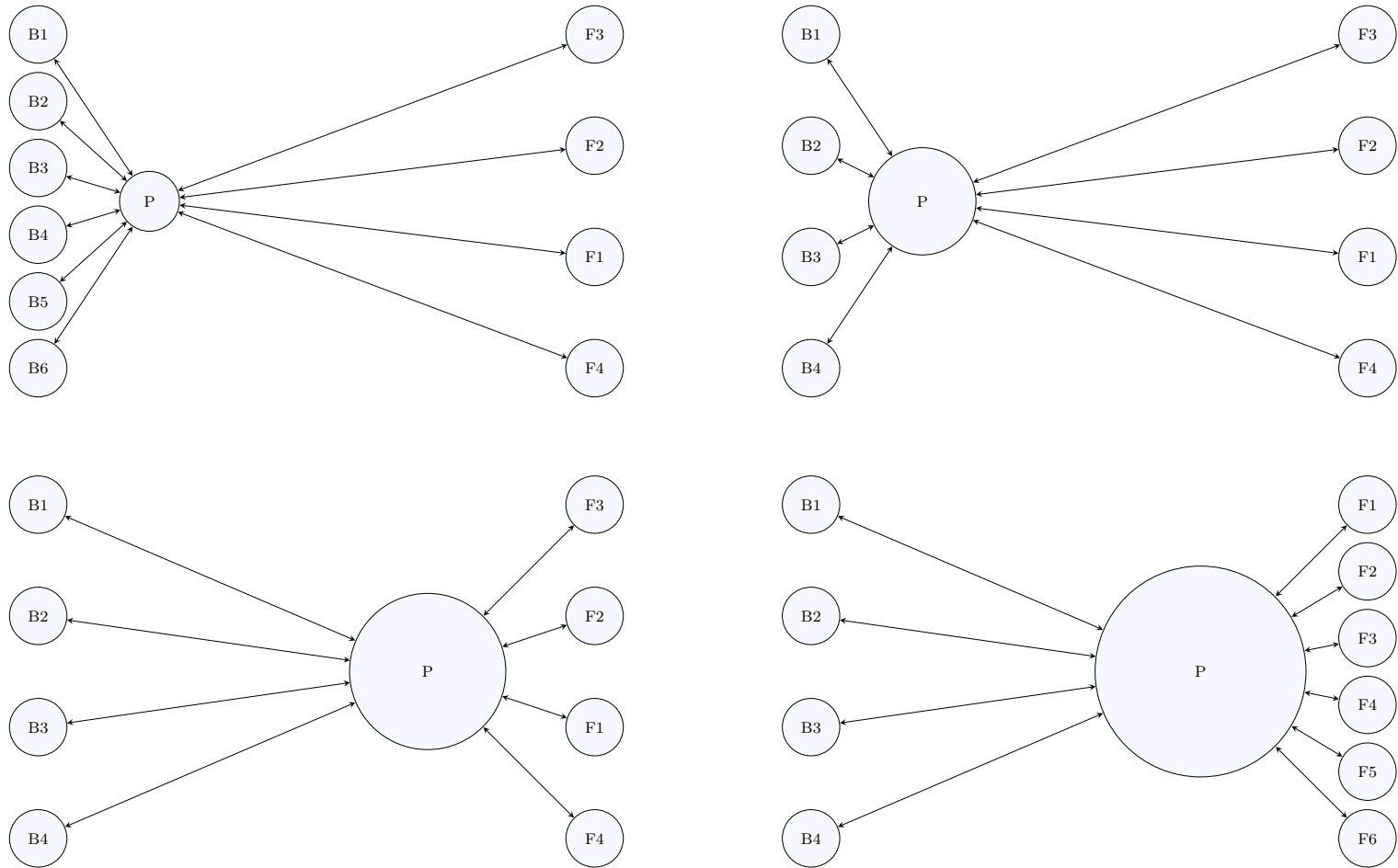


Figure displays four examples of measuring patent quality. In this figure, the length of the arrows indicate distance between patents, while the size of the node associated with the focal patent (P) indicates quality. Patents are more likely to be significant if they are further away from previous patents and/or closer to future patents.

Figure 5: Distribution of Quality and Citations over time

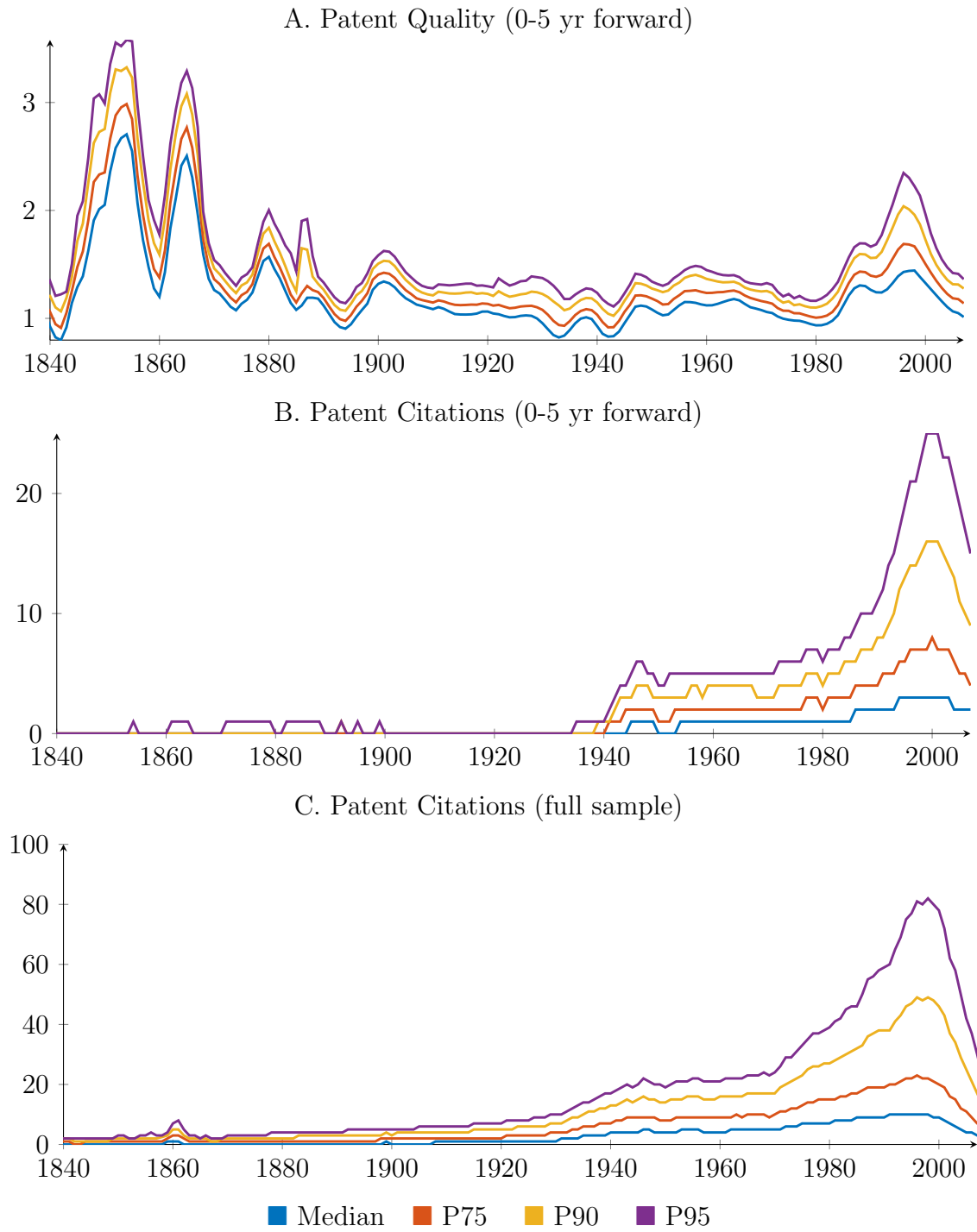
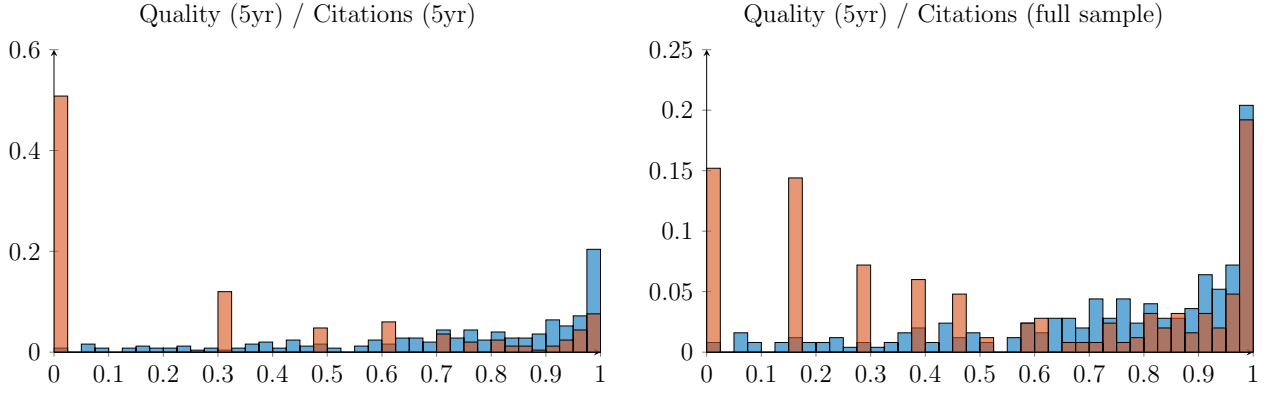


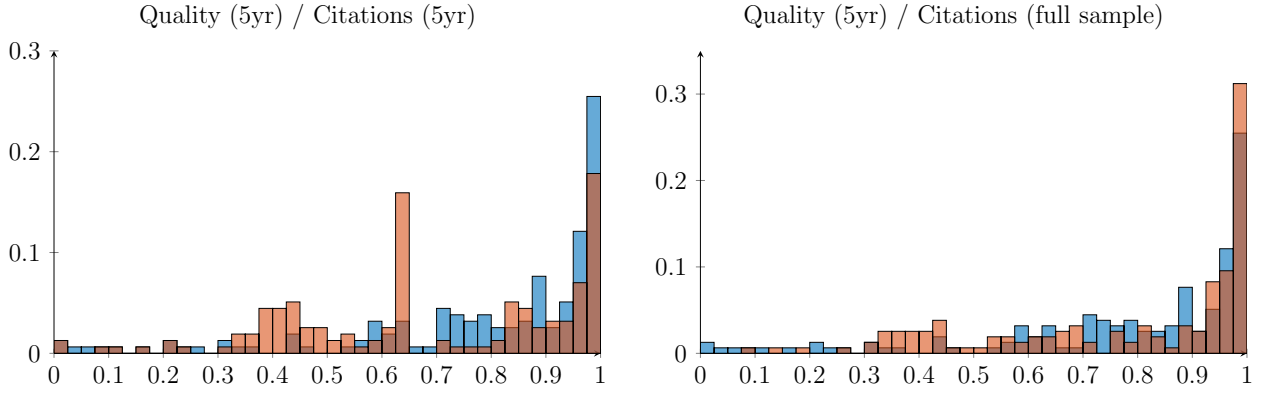
Figure plots the cross-sectional distribution of our quality measure (Panel A) and forward citations (Panels B and C) over time.

Figure 6: Important Patents: Quality vs Citations

Panel A. Comparison across cohorts: no adjustment



Panel B. Comparison across cohorts: remove year FE



Panel C. Comparison within cohorts

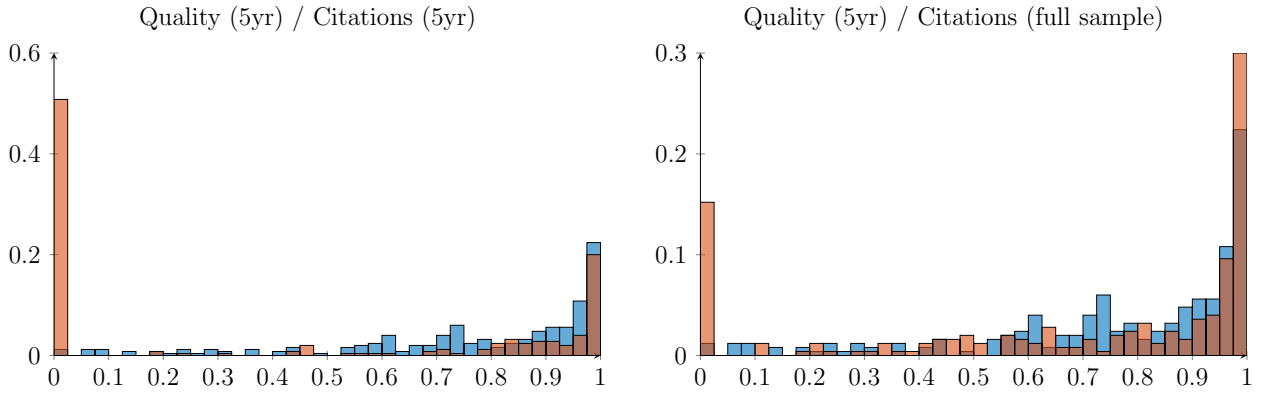


Figure compares the extent to which our quality indicator successfully identifies historically important patents, and compares with patent citations. The figure plots the distribution of patent percentile ranks based on our quality indicator (solid line) and forward citations (dashed line). A value of x% indicates that a given patent scores higher than x% of all other patents unconditionally (panel A); unconditionally, but adjust quality and citations by removing year-fixed effects (Panel B); or relative to patents that are issued in the same year (panel C). The list of patents, along with their source, appears in Appendix Table ??

Figure 7: Patent quality and citations

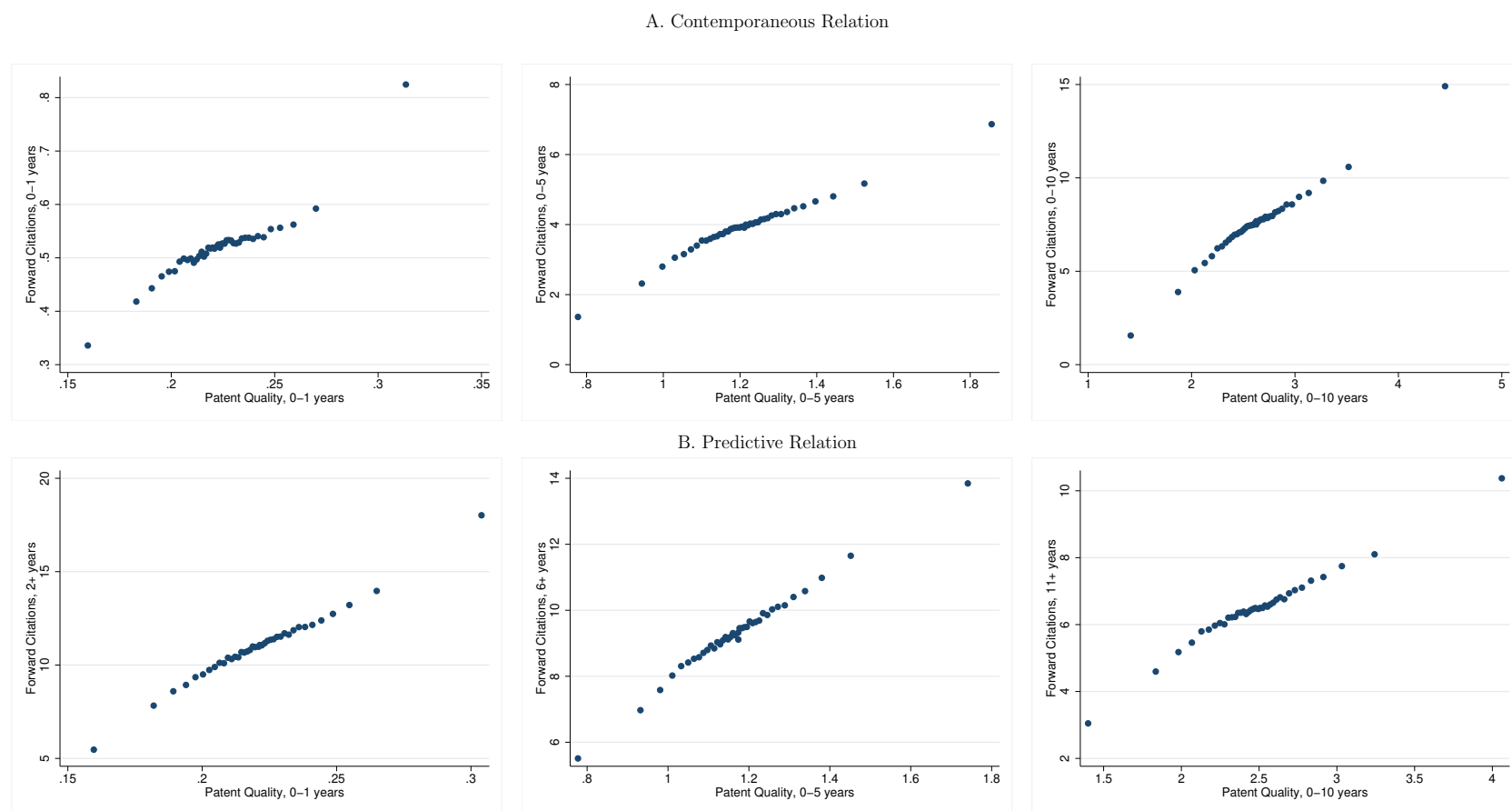


Figure plots the relation between the number of forward citations to our quality measure (both in levels). Panel A relates our quality measure to patent citations, when both are measured over the same horizon. The binned scatter plots control for fixed effects for technology class, and the interaction between assignee and patent grant year. Panel B plots the predictive relation between our quality measure and future citations; in addition to technology and assignee-issue year fixed effects, we also control for the number of citation the patent has received over the same horizon that our quality measure is computed.

Figure 8: Technological Innovation over the Long Run: Existing Indicators

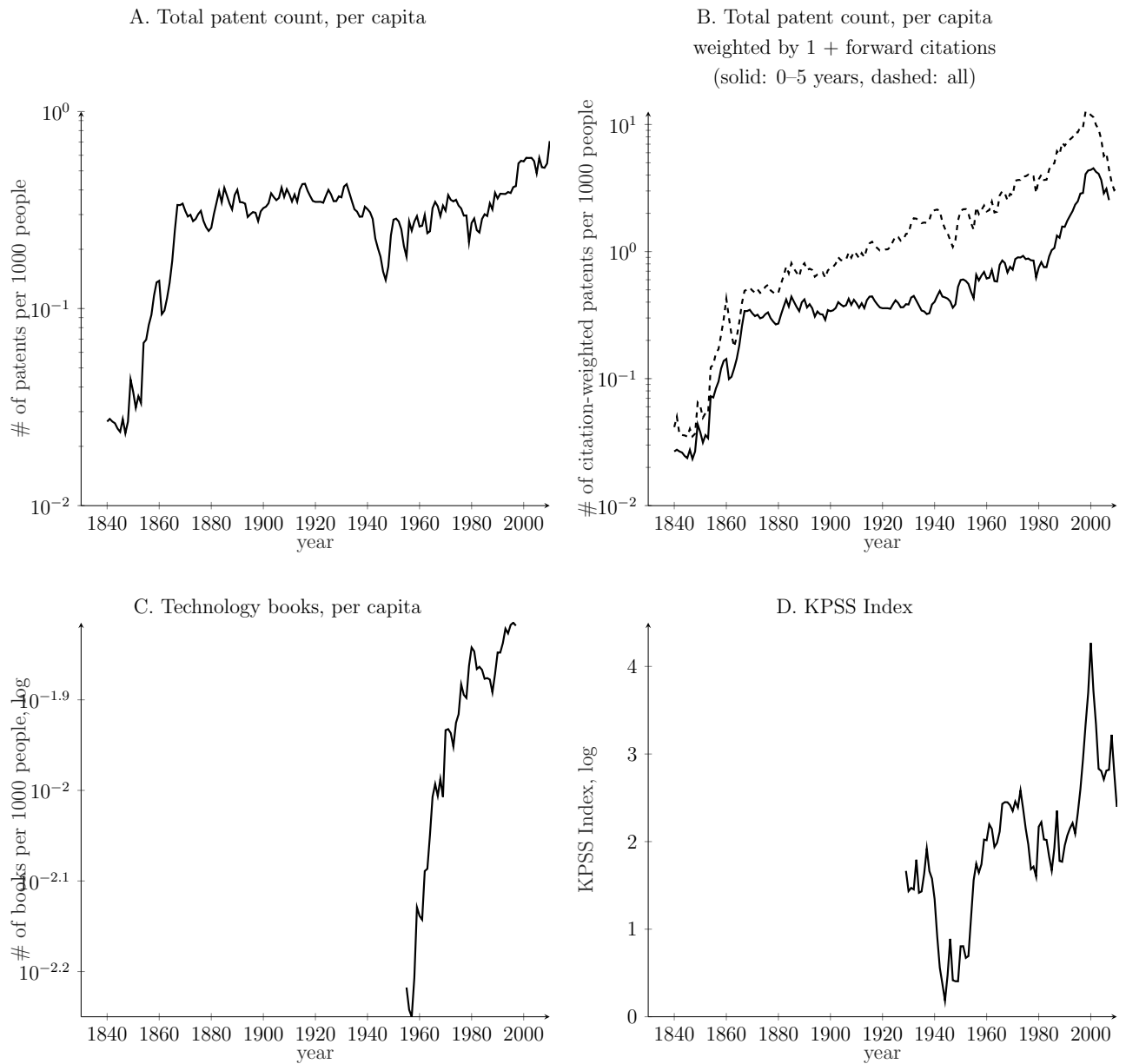
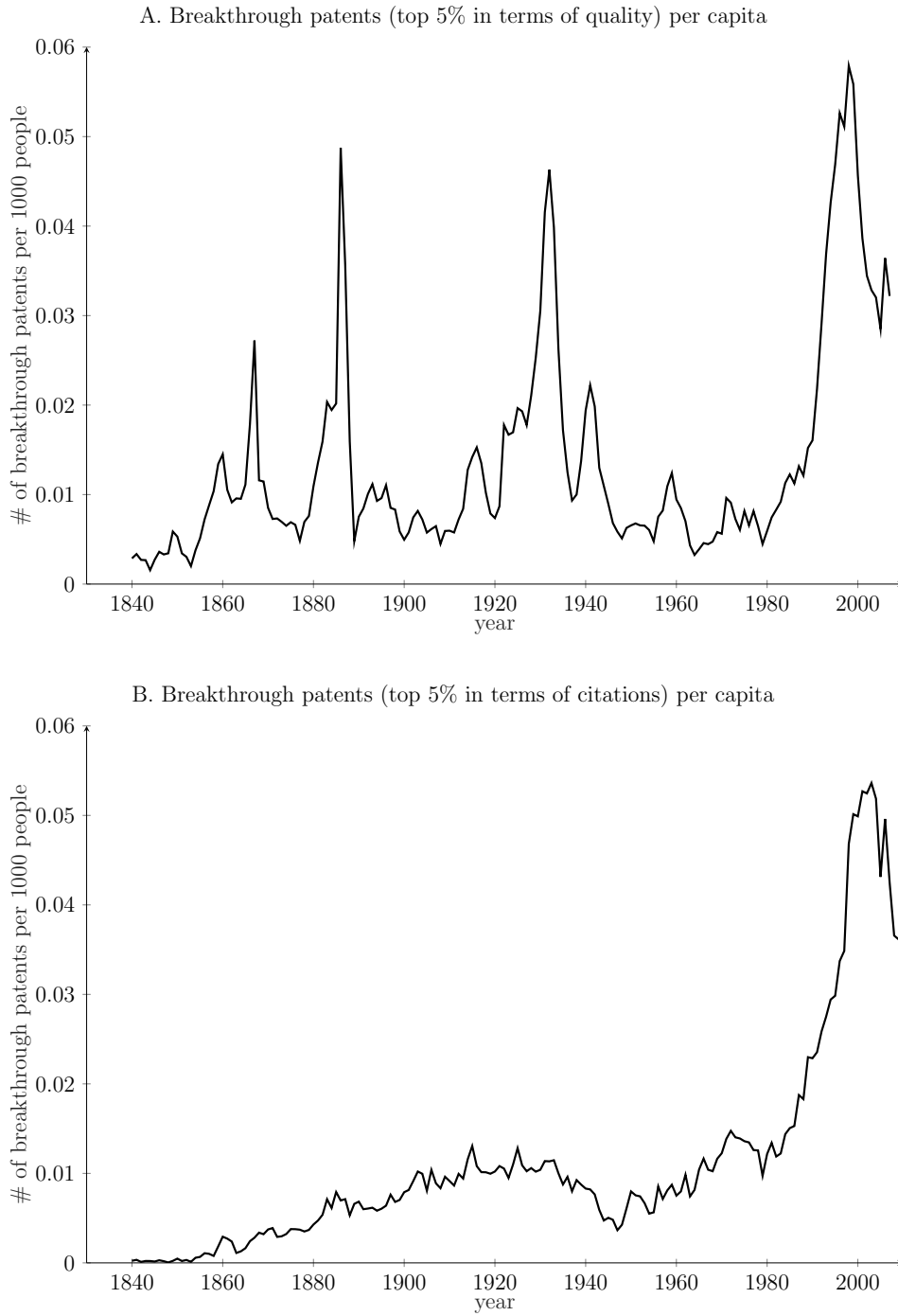


Figure plots existing indices of technological innovation.

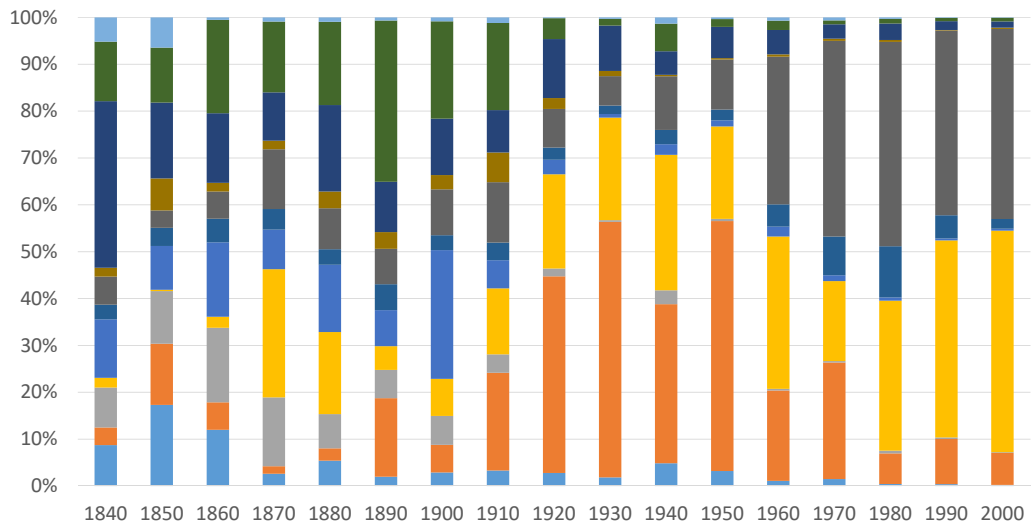
Figure 9: Technological Innovation over the Long Run: Breakthrough Patents



Panel A plots the number of breakthrough patents, defined as the number of patents per year that fall in the top 5% of the unconditional distribution of our baseline quality measure (defined as the ratio of the 5-yr forward to the 5-yr backward similarity) net of year fixed effects. We normalize by US population. In Panel B we plots the number of patents that fall in the top 5% of the unconditional distribution of forward citations (net of year fixed effects), again scaled by US population. The solid line denotes the index based on 5-year forward citations, the dotted line uses the total number of citations over the lifetime of the patent.

Figure 10: Breakdown by Technology Classes

Panel A: Breakthrough (Top 5%) Patents



Panel B: All patents

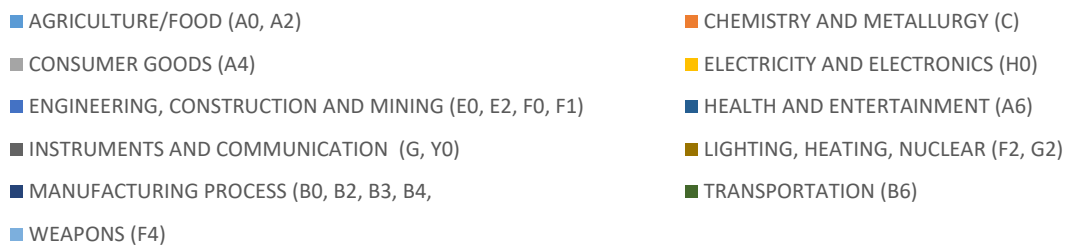
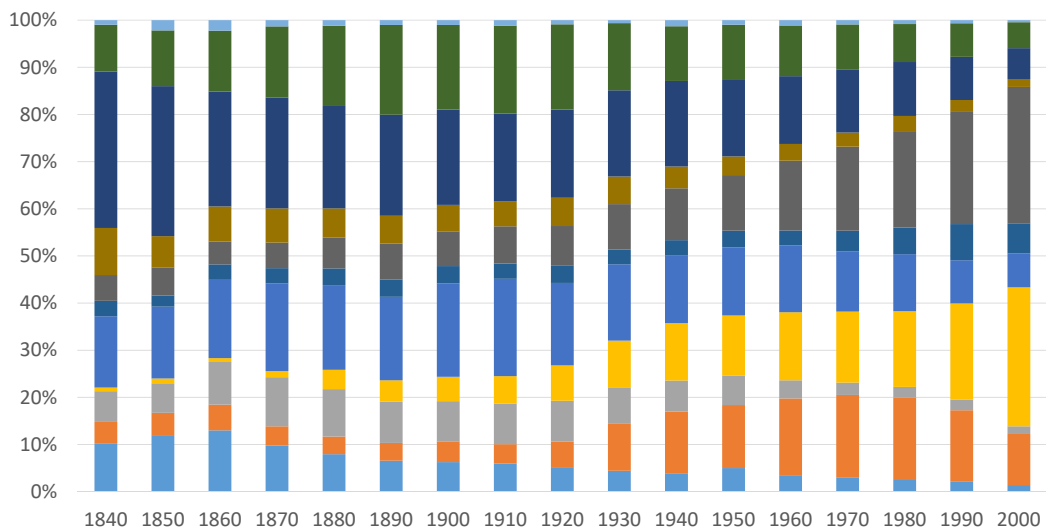
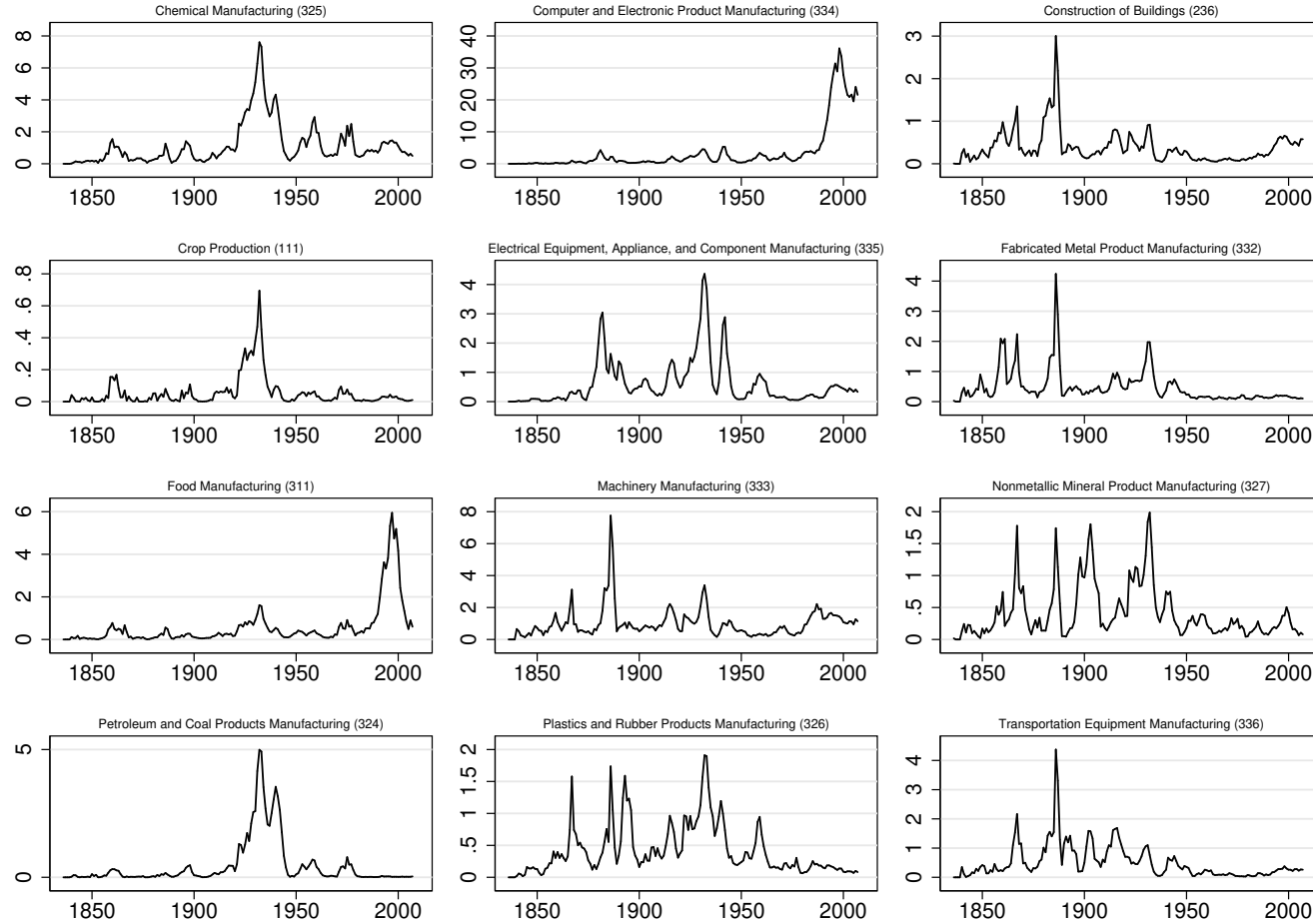


Figure 11: Innovation across industries: Breakthrough patents



Panel plots the number of breakthrough patents in each industry (NAICS 3-digit code), defined as the number of patents per year that fall in the top 5% of our baseline quality measure (defined as the ratio of the 5-yr forward to the 5-yr backward similarity) net of year fixed effects. We use the mapping from CPC4 codes to 3-digit NAICS codes provided by @@@. We restrict attention to the 12 most innovative industries (defined by the total number of breakthrough patents over that period).

Figure 12: Breakthrough patents and Aggregate TFP

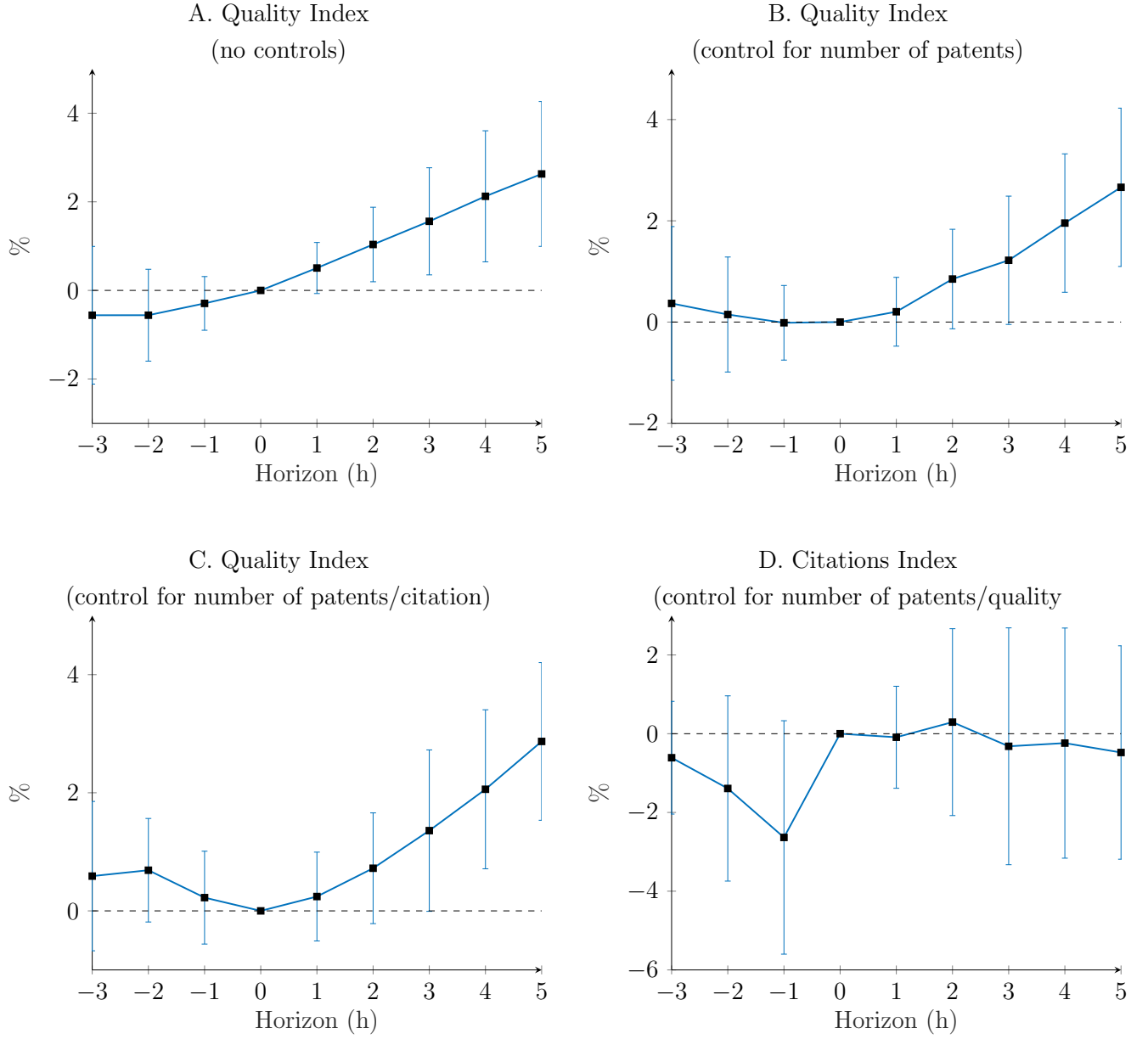


Figure plots the response of total factor productivity, adjusted for utilization, to a unit standard deviation shock to our technological innovation index (Panels A to C) and to the corresponding index based on citations (Panel D). Panels C and D plot the coefficients from a multi-variate regression. TFP is utilization-adjusted total factor productivity from ?. We include 95% confidence intervals, computed using ? standard errors.

Figure 13: Breakthrough patents and Industry TFP

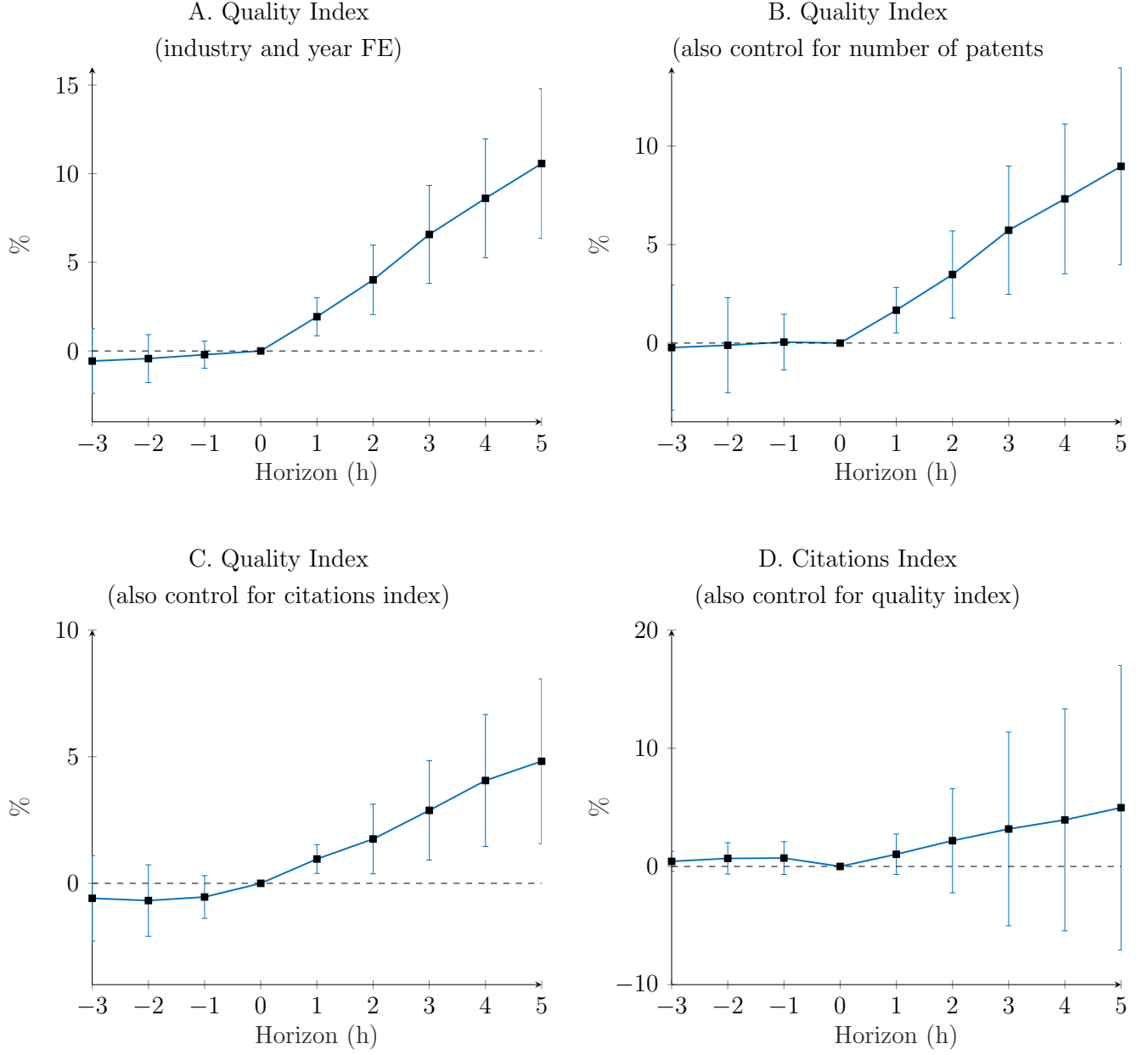


Figure plots the response of total factor productivity, adjusted for utilization, to a unit standard deviation shock to our technological innovation index (Panels A to C) and to a corresponding index by citations (Panels D). Panels C and D plot the coefficients from a multi-variate regression. Industry productivity data comes from the World KLEMS database (April 2013 release). Industry definitions are based on ISIC classification codes. We construct industry indices using the CPC4 to ISIC crosswalk constructed by ?. We only consider KLEMS sectors with non-zero patenting activity, which leaves us with 15 sectors covering the 1947–2010 period: basic metals; chemicals; petroleum and nuclear fission; electrical equipment; electricity, gas, and water supply; food; machinery; various manufacturing; mining and quarrying; non-metallic mining; paper; rubber and plastics; textiles; transport equipment; and wood. We include 95% confidence intervals, computed using standard errors clustered by industry and year.

Figure 14: Breakthrough patents and firm profitability

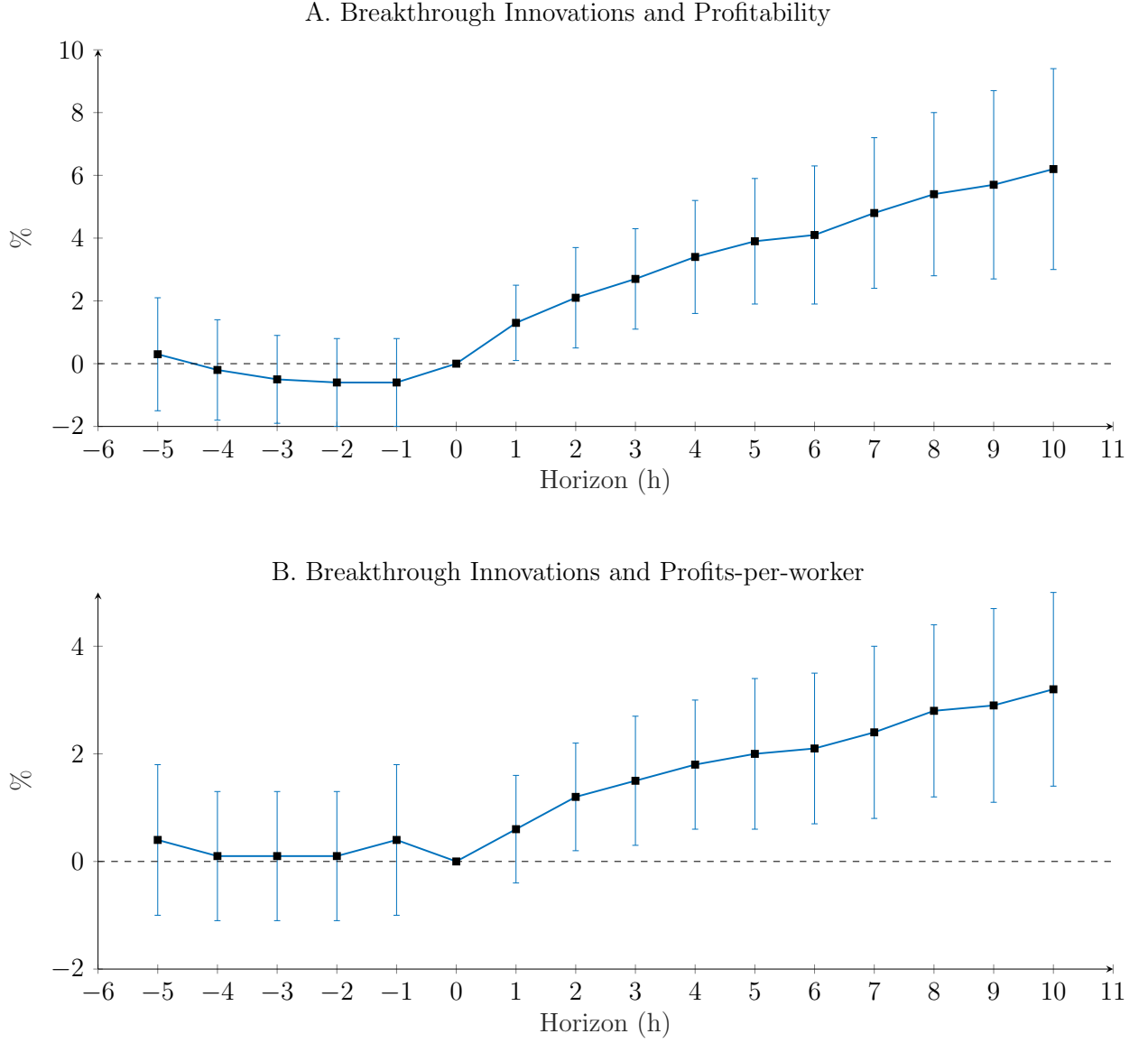


Figure plots the response of firm profits (panel A) and output per worker (panel B) to a dummy variable that takes the value of one if the firm has a breakthrough patent. The patents are dated as of the filing year ($t = 0$). Controls include a dummy variable for whether the firm has filed any patents during this period, the log number of patents, and industry-year fixed effects. Breakthrough patents are those that fall in the top 5% of our quality measure (net of year fixed effects, see text for details); patent quality is measured as the ratio of the 5-year forward similarity to the 5-year backward similarity. Profits are sales (Compustat: sale) minus costs of goods sold (Compustat: cogs); profits per worker is profits divided by the number of employees (Compustat: emp). We include 95% confidence intervals, computed using standard errors clustered by firm and year.