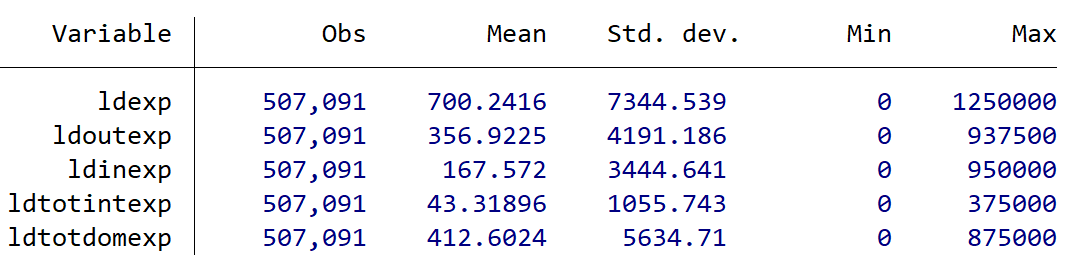
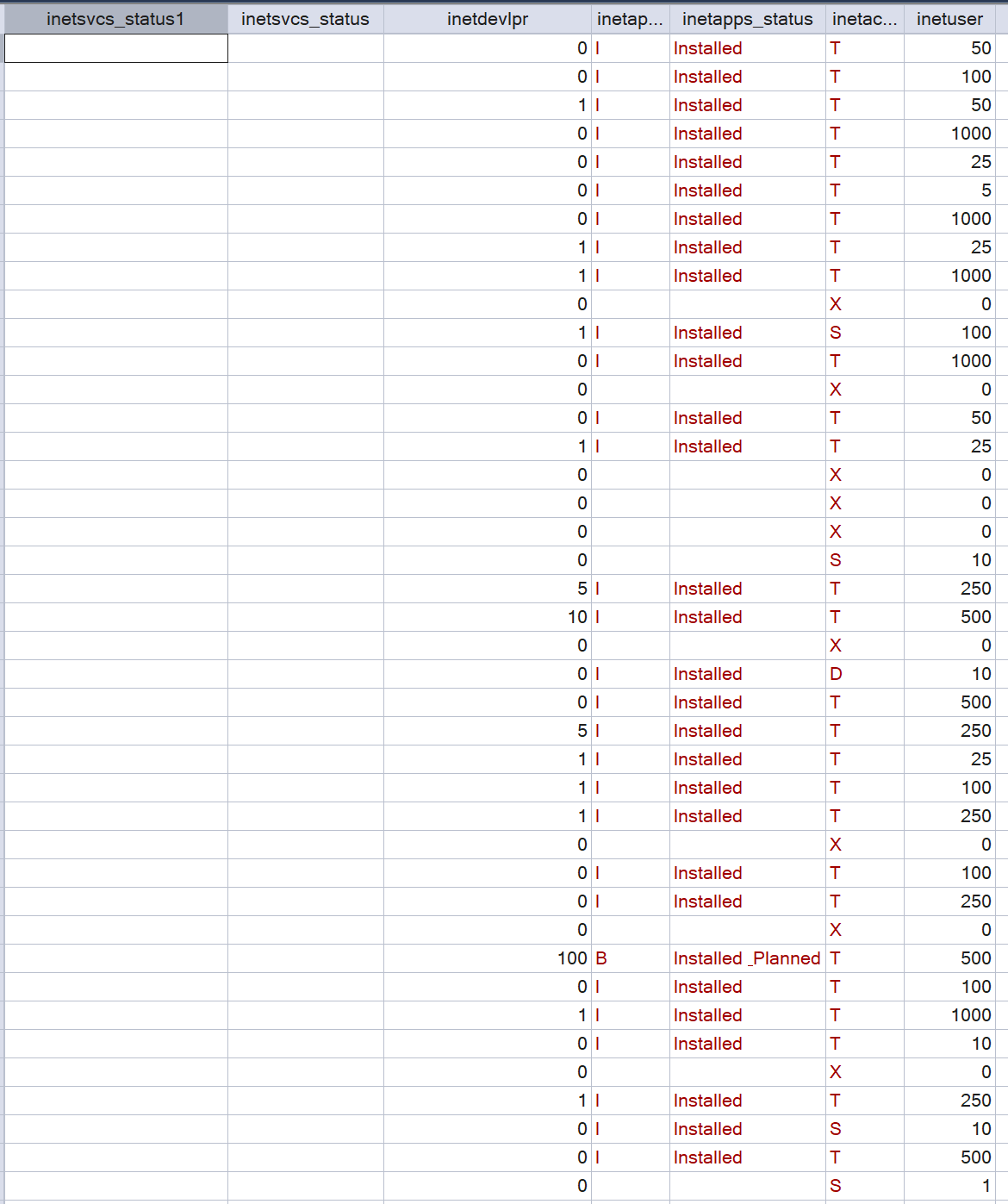
3.24:

1. Summarize information on the IT spending/other dollar amount related to the IT-investment in the HH dataset.
   1. Site Communications Summary Table: dollar amount variables are mostly long distance expenditure (categorized to outbound, inbound, international, domestic). But really not sure this is related to IT spending but does capture the cost of remote communication (but it could be external or internal)



* 1. IMO the number of pc, server, node etc (which were used & constructed in the do file) were the most useful variables in HH database.
  2. Related to Bloom et al., 2014: the access to intranet:



* 1. The number of white worker and desk worker might be useful to calculate the computer per worker?

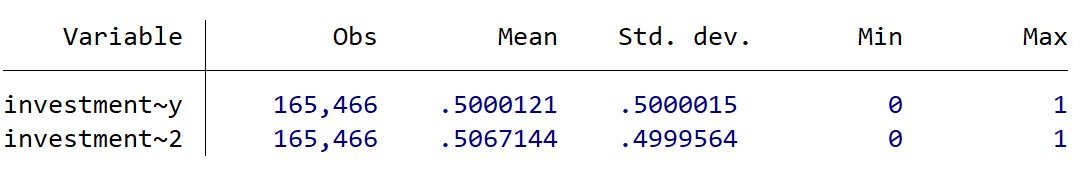
1. Classification of HH dataset: detailed data on firms’ investments in hardware, software, storage, networking, telecom IT investments, and ERP systems
2. Please follow the paper to construct the variables for all Compustat firms from 2000 to 2020: “Capital + R&D + Advertising Expenditure Efficiency”
   1. Paper: Goodman-ManagementForecastQuality-2014.pdf
      1. Location: Variable\_Construction/Management Forecast Accuracy
   2. Using STATA to complete and generate the dta file.
   3. Please also save the management forecast accuracy data you constructed before as the dta file.
3. When constructing these two variables for all Compustat firms, Please make sure firms in “Firm\_identifier.dta” are included – Added the CIK and checked: 3 firms are not matched.
4. Classification of industry competitiveness
   1. Literature review on the industry competitiveness – textual based (Hoberg-Phillips)
5. Construct competition related variables:
   1. Hoberg-Phillips textual measure
      1. HHI and total similarity data (they have the gvkey, year available for each firm). The data is downloaded from: <https://hobergphillips.tuck.dartmouth.edu/industryconcen.htm>
      2. Notes about the data from their technical details:
         * Each file contains a gvkey and year firm identifier.
         * The TNIC3HHI variable is a concentration measure and TNIC3TSIMM is a total similarity measure.
         * On date conventions, for convenience, the year field in this database is based on Compustat calendar years obtained as the first four digits of the YYYYMMDD datadate variable.
         * These HHI and total similarity data are computed using TNIC designations that include the firm itself in part of the HHI calculation. All HHIs are based on firm sales data from COMPUSTAT, and are computed using the Herfindahl-Hirschmann sum of squared market shares formulation.
   2. Industry concentration variables – HHI and four-firm ratio; those in Li, 2010
      1. (Draft) Steps to construct the 9 variables in Li, 2010:
         * Start with the Historical Segments data in Compustats
         * Merge with funda to get firm level financials ().
6. Check the matching between the segment data and sites in (Public\_hh\_dataset\_kept.dta)

Detailed steps to generate management forecast quality and investment efficiency in *Goodman et al., 2014: Management Forecast Quality and Capital Investment Decisions*

1. Main Results of the paper: Managers with higher quality external forecast make better investment decision (acquisition and capital expenditure decision)
2. Management forecast quality: earning forecast accuracy
   1. Definition: absolute value of the difference b/w management forecast EPS and the actual EPS, divided by stock price 3 days prior to the forecast; annual EPS.
   2. Y: Average of 3-year forecast accuracy prior to the unexpected investment
   3. I included 3 version of forecast accuracy:
      1. Abs(error)/actual
      2. Abs(error)/year-end price
      3. Abs(error)/3-day prior price
3. Investment efficiency
   1. Definition of unexpected investment: absolute value of the residual of the following industry-year regression (for industry with at least 30 observations):

Capex (sum of capital, R&D and AD expenditure) ~ Lag Tobin’s Q + CFFO + lag asset growth + lag investment

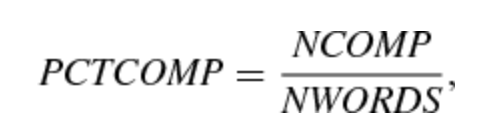
* 1. Definition of investment efficiency: indicator variable equals to 1 if the unexpected investment is below the median of the distribution of unexpected investment, otherwise 0.
  2. Steps to construct the efficiency variable involves two steps:
     1. Get firm-year level financials including in 2000-2020 for all compustats firms (in SAS)
        1. Capital (which item in compustats?)
        2. R&D expenditure (O)
        3. Advertising Expenditure (which item?)
        4. Lag Tobin’s Q
        5. CFFO
        6. Lag asset growth: start from asset
        7. Lag investment: start from investment
     2. Regress at industry-year (for industry with 30 or more observations) to get the residual value, median and the indicator variable (in stata)



The descriptive data is similar to Goodman et al., 2014.

Literature review on the industry competitiveness

1. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis (RFS, Hoberg and Phillips, 2010)
   1. Main results: Using textual analysis (similarity) of 10K product description, they find transaction stock returns, ex-post cashflows, and growth in product descriptions all increase for transactions for firms with similar product market language, especially in **competitive product markets.**
   2. Measurement of product market in their study:
      1. Textual-based product similarity (nearest 10 rivals & broad, firm-year level): Local and broad cosine similarity of pair wise product description. Higher similarity indicates higher competition
      2. % Neighbor Patent Words: For each firm i, this is the percentage of words in its product description having the word roots “patent,” “copyright,” and “trademark.”: High percentages indicate a firm’s emphasis on proprietary technology or products.
      3. Sales Herfindahl-Hirschman Index (HHI) (SIC-3): sales-based Herfindahl ratios for each three-digit SIC code
2. Text-Based Network Industries and Endogenous Product Differentiation (JPE, Hoberg and Phillips, 2016)
   1. Main results: created a time-varying industry classification and firm-specific competitor network
   2. The measurement of industry competition
      1. Similarity measure: similar to Hoberg and Phillips, 2010. Firms that have many peers with similar product description are considered to have high competition.
      2. High competition dummy: 1 if the manager cites “high competition”, or one of its synonyms, in MD&A
3. Product market competition, disclosure framing, and casting in earnings conference calls (JAE, Allee et al., 2021)
   1. Main story: the intensity of competition in the product market is associated with more negative and uncertain earnings conference calls
   2. Measure of intensity of competition: Ali et al. (2009) show that Compustat-based measures of competition using static industry classification are unreliable. Thus, we use Hoberg and Phillips’ (2016) text-based network industry classification (TNIC) measure as a proxy for product market competition.
4. Industry concentration and corporate disclosure policy (JAE, Ali et al., 2014)
   1. Main results: In more concentrated industries firms׳ management earnings forecasts are less frequent and have shorter horizons, their disclosure ratings by analysts are lower, and they have more opaque information environments.
   2. Measurement of industry concentration (as a measure of competition)
      1. HHI: is calculated for a 6-digit NAICS industry within the manufacturing sector by summing the squares of the individual company market shares of the 50 largest public and private companies in the industry or all the public and private companies in the industry, whichever is lower.
      2. Four-firm ratio: is the sum of the market shares of the four largest firms in the industry in terms of market share.
      3. Data source: Census of Manufactures publications for the 1997, 2002, and 2007 U.S. Census years
      4. a measure for competition from potential entrants from Li (2010) and Li et al. (2013). The measure for competition from potential entrants is based on proxies for entry costs and market size.
5. The impacts of product market competition on the quantity and quality of voluntary disclosures (RAS, Li, 2010)
   1. construct variables to separately measure competition from potential entrants and competition from existing rivals by conducting principal component analysis on commonly employed 9 proxies of competition:
      1. Industry concentration: 1): HHI and 2): four-firm ratio
      2. Karuna (2007) suggests three determinants of industry competition: 3): industry PP&E, 4): product market size, and 5): price-cost margin.
      3. He also included 6): industry research and development intensity, 7): industry capital expenditures, 8): total number of firms operating in an industry, 9): industry return on assets as additional measures for competition
6. A Measure of Competition Based on 10-K Filings (JAR, Li et al., 2012)
   1. This paper present a dictionary based measure of competition based a firm's 10-K filing.
   2. We measure management's perceptions of the intensity of the competition they face using word count of competition word list in the firm's 10-K filing, divided by the length of 10K



1. Spatial competition: Hotelling, 1929; Auditing literature; Distance b/w audit firms and client firms.