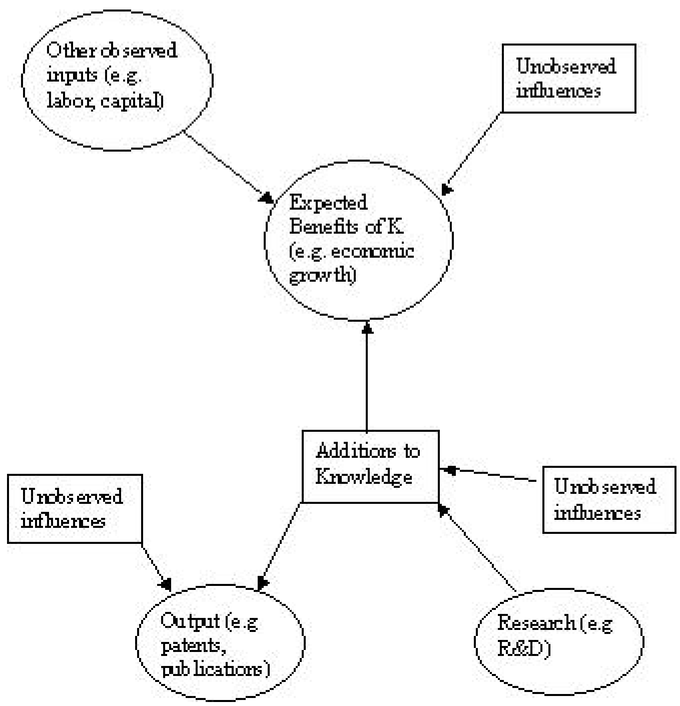
Lecture # 4 – Measuring Knowledge

I. A Model of Knowledge and the Economy

- Our goal in this class is to discuss various types of data that can be used to measure the effects of knowledge on the economy.
- To begin, we consider a model of how knowledge affects the economy.
    - This is important because we can measure inputs and outputs of the process, but we rarely measure output directly.
    - Thus, different measures may be right for different tasks.
    - Note: in the diagram below, circles represent observed variables, and boxes represent unobserved variables.



- Although we would like to observe knowledge directly, we cannot.
- We can observe:
    - Inputs into the creation of knowledge
        - For example, we can look at R&D data, spending on higher education, etc.
    - Outputs of the creation of knowledge
        - We can look at patents & publications
        - We can measure the effect of knowledge on economic growth (e.g. TFP).

- Unfortunately, there are unobserved effects between any of the observed variables and the effect of knowledge.
    - For example, not all R&D is created equal.
    - Not all inventions are patented.
    - Changes in quality might be mismeasured because of price changes.
- In today's class, we look in more detail at each of the observed variables in this model.

II. Measuring the Effect of Knowledge

- Since any discussion of different measures of research effort ultimately want to tell us about the effect of knowledge on well-being, we begin by discussing how economists measure the effect of knowledge on the economy.
- Possible measures of the effect of knowledge on output:
    - One "measure", introduced in the first week of class, is simply the portion of growth that cannot be explained by changes in the inputs of the economy. We called this <u>total factor productivity (TFP)</u>.
        - Using g to represent growth rates, this can be defined as
            - $g_{TFP} = g_Q - s_K g_K - s_L g_L$
    - Economists can treat knowledge as an input to production, and use the measures discussed later in class to find the effect of these inputs on output.
    - Economists can calculate the rate of return on an investment, as in the Mansfield paper.
        - More appropriate for micro-level studies of a single technology.
        - One drawback of case studies is that they tend to focus on successful inventions (or large failures).
            - Thus, they might not be typical of innovation as a whole.
- Complications
    - A difficulty with interpreting output data is that changes in prices may hide technological change – changes in quality.
    - One reason is that the Consumer Price Index (CPI) (and the Producer Price Index) do not accurately measure changes in quality.
    - For example, if we look at health care costs, we would see large increases in spending. However, the quality of care received has improved. Computers are similar.
    - A 1998 government study estimated that omitting technological change overstated inflation by 0.6% -- that is, inflation is actually 0.6% lower.

- Example of effect of higher quality: *Economist* article on measuring benefits of the Internet
    - GDP only measures economic transactions.
    - Doesn't consider consumer surplus – the extra welfare generated by a new innovation
        - Suppose, in 1999, broadband cost $20/month
        - In 2006, suppose broadband costs $17/month
            - Consumers who were willing to buy broadband in 1999 now receive a $3 surplus
            - Using calculations like this, Greenstein/McDevitt estimate the consumer surplus from broadband to be $5-7 billion per year
        - Note that this is conservative
            - Assumes consumers get the same benefit (e.g. have the same demand curve) for broadband in 1999 and 2006
        - In reality, if demand is higher (e.g. because of more available on Internet today), surplus is also higher
    - Other possible measures:
        - Survey: ask people willingness to pay
        - Value of time
            - Can look at value of time saved by using Internet to find information
            - Must assume a value of time to do this
            - Also could look at increased leisure time spent on Internet
- How does this affect our measure of technological change?
    - Recall that $g_{TFP} = g_Q - s_K g_K - s_L g_L$
    - These calculations must be done using real dollars.
    - If inflation is overstated, adjustments to g_Q are too high.
    - This suggests that output actually grows faster than shown in data that ignores technological progress.
    - Assuming that the quality of K and L is measured correctly (it may not be), TFP would be underestimated.
- Other potential problems
    - Knowledge assets are hard to measure
        - Can count stocks of accumulated R&D, but don't know future value of what it will create.
    - Capital stocks hard to measure
        - Consider differences in labor productivity between the US and Europe
            - Non-farm labor productivity in the US grew 3%/year from 1996-2002, and has been growing
            - In contrast, European growth rates were around 2%/year, and falling.
    - Why are these different?
        - Increased capital stocks in the US (e.g. IT) made labor more productive.

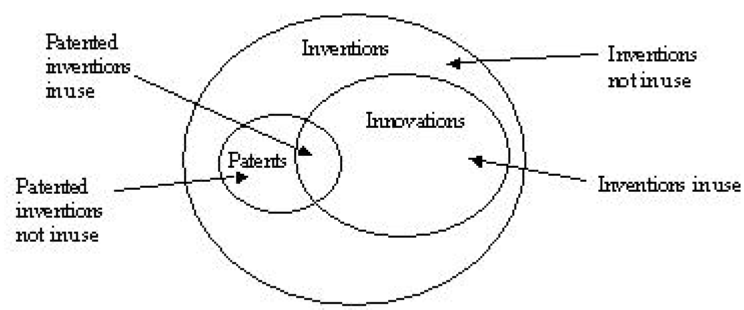
        - TFP controls for this:
            - For GDP/Inputs, both numerator and denominator up.
            - For GDP/L, only numerator up.
        - Quality adjustments are also important.
            - The US corrects prices to control for the increased quality of computers.
            - Makes GDP growth bigger, because acknowledges we get more power for the same cost.
        - In the European data, for the TFP calculation, both the numerator and denominator are incorrect. Thus, some of the error cancels out.
        - In contrast, for GDP/L, only the numerator is incorrect. Growth is measured too slowly, so labor productivity suffers.

III. Measures of Research Inputs

- The most common measure of input into the scientific process is measures of research and development (R&D) spending.
- Advantages of R&D data
    - R&D offer direct measures of the inputs into the research process
        - For example, if you want to know the opportunity cost of increasing government funding for R&D, what matters is how much money is spent.
    - Data is available at a reasonable level of detail
- Disadvantages/complications
    - The return on a dollar of R&D may vary.
        - That is, the level of inputs is certainly correlated with, but not a direct measure of, the output of research.
    - Time lags
        - A research project may take a couple of years. Thus, the effects of R&D might not be immediate.
    - Much government R&D in US goes to defense or space exploration.
        - The benefits of these projects will not traditionally show up in output measures, since not typically sold in markets.

IV. Measures of Research Output

- Measures of output can help to measure the unobserved success or failure of research inputs. However, there are other complications with the data.
- Two widely used measures of the research patents are:
    - Patents
    - Scientific publications
- Patents
    - Patents provide monopoly rights to an invention in return for publicly disclosing the invention.
    - Advantages of patent data
        - Long time series of historical data available
            - US patent system began in 1793
        - Greater detail available on patents than with R&D statistics
        - All patents must meet at least some minimum quality standard. To receive a patent, an invention must be:
            - Novel
            - Useful
            - Non-obvious
    - What do patents measure?
        - A patent represents the output of a research process.
        - However, not all inventions are patented, and not all patents are equal.
            - Only a subset of inventions are brought to market.
            - Furthermore, only a subset of inventions is patented.



- Implications
  - Patents proxy for the knowledge created in the research process, but they are not a complete count of innovation.
    - How problematic these variations are depends on how the data is used.
  - To look at innovation across time, we need to assume that the likelihood of patenting an invention stays the same across time.
    - Not always true. Patent policy changes do affect patenting rates. However, one can control for these if necessary.
  - To compare data across industries, we need to assume that the likelihood of patenting is the same across industries.
    - We will see evidence in the patents section of the course that this is not true.
  - When comparing patents across countries, we need to be aware of differences in national patent systems.
  - Patent application data is more reliable than patent grant data
    - Grants depend on patent office behavior. For example, fewer patents were granted in 1979, due to budget cuts.
- Validation of patent statistics
  - Despite these problems, much research has been done using patent statistics.
  - Typical findings:
    - There is a strong correlation between patents and R&D within an industry.
    - The ratio of patents to R&D has, until recently been falling. As a result, the correlation across time is not as strong.
      - Is this due to changes in patenting behavior, or changes in research productivity?
    - Variations in quality
      - Most patents have little value. However, some are extremely valuable.
      - Possible controls for quality:
        - Patent citations
          - Patent citations are references to earlier patents.
          - Weighted patent counts tend to correlate stronger with other measures of technological change.
          - Citations are also useful to tell us about knowledge flows.
        - The number of countries patent protection is granted in.
          - Because you have to pay an application fee in each country, inventors only file abroad if the invention is good.

- Research on scientific publications is known as bibliometrics.
  - Alternative methods:
    - Publication counts
    - Citation counts
      - Again, controlling for quality is an important issue.
        - Counting citations made to each publication is a way to control for quality.
        - For example, in the article by May, note that the top publishing countries do even better when citations are used to control for quality.
  - See today's slides for examples of data from King's article

V. R&D and Productivity

- Once we have measures of productivity and R&D, we can use the data to assess the effectiveness of the inputs (e.g. R&D) on the desired output (e.g. productivity).
- As noted before, methods include case studies, such as Mansfield, and econometric studies. Here we look at some of the econometric evidence.
- Methodology
  - Two related approaches:
    1. Production function studies estimate the effect of R&D on output or TFP
    2. Cost function studies estimate the effect of R&D on costs
  - Both are related by theory, but have different data requirements
  - Data can be cross-section, time series, or both
- Theory
  - Begin with a production function that includes R&D as an input:
  - This leads to the following estimating equation, which uses growth rates:
    - $g_Q = A + ag_K + bg_L + cg_R + \varepsilon_t$
    - Here, c is the elasticity of output with respect to R&D – the percentage increase in output from a one percent increase in R&D.

  - Problems
    - Endogeneity due to omitted variables (e.g. if firms productive for reasons unrelated to inputs, such as managerial skill)
    - Measurement issues
      - Quality changes might not be captured in measures of output
      - Aggregating past research activities into one measure of knowledge is difficult.
        - For example, there may be time lags between R&D and the observed effect.
        - Does R&D depreciate?
      - Potential double counting – are labor and K used for R&D also included in L & K?
      - If R is confined to a specific industry or firm, spillovers from other industries are not measured
  - Cost function models are similar, but relate R&D to reductions in production cost
    - Require more data, as need to know the costs of inputs
    - A typical cost function would be:
      - cost = f(Q, w, r, R&D)
      - Input prices likely exogenous, since determined for the whole industry
      - However, other endogeneity issues remain
- Results of private returns to R&D
  - Elasticity of R&D on output positive and significant, typically around 0.1 – 0.2
    - Findings are stronger in micro-level cross sections than macro-level time series studies.
  - Elasticity of R&D on TFP positive and significant, typically around 0.2-0.3
  - Elasticity using cost functions typically around 0.2-0.25
  - In all studies, there is greater variation across industries than within industries.
    - That is, R&D has different effects in different industries.
- Results on social returns to R&D
  - Two ways to study spillovers
    0. Compare estimated private returns to R&D to estimates of returns to other types of investments. Typically, the returns to R&D are higher.
    1. Include variables to capture outside R&D.
    - For example, look at R&D done by industries that supply a firm
  - These studies support the idea of spillovers
  - An often-cited ratio is that social returns are 4x that of private returns.