References about leading, lagging, and coincident economic indicators

Question Link

Reading about macroeconomics, I've come across a great deal of information regarding leading, lagging, and coincident indicators. I'd like to learn more, specifically information related to measurements.

For the most part, in the books that I've read, indicators are graphed along side changes in the economy. I imagine that economists do not just graph items and visually inspect whether they are (1) indicators and (2) if they are leading, lagging, or coincident, etc.

What articles/texts are available that discuss the statistical methods behind identifying and measuring economic indicators?

Answer

The Wooldridge texts are part of most graduate level econometric studies in the US. They are:

- Introductory Econometrics: A Modern Approach, 6th Edition
- Econometric Analysis of Cross Section and Panel Data, 2nd Edition

I would start with the first, unless you have a lot of formal training in statistics.

Hope this helps,

Justin

[tag:econometrics] [tag:macroeconomics] [tag:statistics] [tag:r]

UPDATED EXAMPLES: from *INTRODUCTORY ECONOMETRICS by Jeffrey M. Wooldridge*

Below is proof of how the text book referenced illustrates how Economists can use statistical methods to **identify and measure** leading, lagging, and coincident relationships among economic indicators.

The first examples are from chapter 16 on Simultaneous Equations. Examples `16.4` and `16.6` illustrate how to use **Instrumental-variable regression by two-stage least squares** for identifying and measuring **coincident** economic indicators. In this case, how annual inflation drops by about 1/3 of a percent for every 1 percent increase in the import share of GDP.

The second is an example `18.8` of the text and illustrate how **autoregresive** $AR(1)$ **and vector autoregressive** $VAR(1)$ **time series methods** can be used to identify and measure, intertemporal relationships between economic indicators, in an attempt to forecast the future unemployment rate using today's unemployment and inflation rates.

`Example 16.4`: INFLATION AND OPENNESS

> "Romer (1993) proposes theoretical models of inflation that imply that more"open" countries should have lower inflation rates. His empirical analysis explains average annual inflation rates (since 1973) in terms of the average share of imports in gross domestic product since 1973 - which is his measure of openness. In addition to estimating the key equation by OLS, he uses instrumental variables. While Romer does not specify both equations in a simultaneous system, he has in mind a two-equation system:"

$$inf = \beta_{10} + \alpha_1 open + \beta_{11} log(pcinc) + \mu_1$$

$$open = \beta_{20} + \alpha_2 inf + \beta_{21} log(pcinc) + \beta_{22} log(land) + \mu_2$$

"where *pcinc* is 1980 per capita income, in U.S. dollars, assumed to be exogenous, and *land* is the land area of the country in square miles, also assumed to be exogenous. The first equation is the one of interest, with the hypothesis that $\alpha < 0$. More open economies have lower inflation rates."

"The second equation reflects the fact that the degree of openness might depend on the average inflation rate, as well as other factors. The variable $log(pcinc)$ appears in both equations, but $log(land)$ is assumed to appear only in the second equation. The idea is that, ceteris paribus, a smaller country is likely to be more open, so $\beta_{22} < 0$."

"Using the identification rule that was stated earlier, the first equation is identified, provided $\beta_{22} \neq 0$. The second equation is *not* identified because it contains both exogenous variables. Be we are interested in the first equation.

`Example 16.6`: INFLATION AND OPENNESS

"Before we estimate the first equation in 16.4 using the data in *openness*, we check to see whether *open* has sufficient partial correlation with the proposed IV, $log(land)$. The reduced form regression is:"

$$\widehat{open} = \beta_0 + \beta_1 log(pcinc) + \beta_2 log(land)$$

```
open_model <-lm(open ~ lpcinc + lland, data = openness)
summary(open_model)
Coefficients:
        Estimate Std. Error t value Pr(>|t|)

(Intercept) 117.0845    15.8483   7.388 2.97e-11

lpcinc        0.5465     1.4932   0.366    0.715
lland        -7.5671     0.8142  -9.294 1.51e-15
```

"The t statistic on $log(land)$ is over nine in absolute value, which verifies Romer's assertion that smaller countries are more open. The fact that $log(pcinc)$ is so insignificant in this regression is irrelevant."

"Estimating the first equation using $log(land)$ as an IV for *open* gives:"

$$\widehat{inf} = \beta_0 + \beta_1 open + \beta_2 log(pcinc)$$

```
    library(AER)
    inflation_IV <- ivreg(inf ~ open + lpcinc | lpcinc + lland, data = openness)
    summary(inflation_IV)

    Coefficients:
            Estimate Std. Error t value Pr(>|t|)
`(Intercept)  26.8993    15.4012   1.747   0.0835 .`

    open         -0.3375     0.1441  -2.342   0.0210
    lpcinc        0.3758     2.0151   0.187   0.8524
```

"The coefficient on open is statistically significant at about the 1% level against a one sided alternative of $\alpha_1 < 0$. The effect is economically important as well: for every percentage point increase in the import share of GDP, annual inflation is about 1/3 of a percentage point lower. For comparison, the OLS estimate is -0.215, $se = 0.095$."

`Example 18.8`: FORECASTING THE U.S. UNEMPLOYMENT RATE

"We use the $PHILLIPS$ DATA, but only for the years 1948 through 1996, to forecast the U.S. civilian unemployment rate for 1997. We use two models. The first is a simple AR(1) model for $unem$:"

$$\widehat{unemp}_t = \beta_0 + \beta_1 unem_{t-1}$$

> "In a second model, we add inflation with a lag of one year:"

$$\widehat{unemp}_t = \beta_0 + \beta_1 unem_{t-1} + \beta_2 inf_{t-1}$$

```
library(dynlm)
phillips <- ts(phillips, start = 1948)
unem_AR1 <- dynlm(unem ~ unem_1, data = phillips, end = 1996)
unem_inf_VAR1 <- dynlm(unem ~ unem_1 + inf_1, data = phillips, end = 1996)

stargazer(unem_AR1, unem_inf_VAR1, keep.stat=c("n", "adj.rsq", "ser")
=================================================================
                                  Dependent variable:
                    ---------------------------------------------
                                         unem
                             (1)                     (2)
-----------------------------------------------------------------
unem_1                      0.732                   0.647
                           (0.097)                 (0.084)

inf_1                                               0.184
                                                   (0.041)

Constant                    1.572                   1.304
                           (0.577)                 (0.490)

-----------------------------------------------------------------
Observations                  48                      48
R2                          0.554                   0.691
Adjusted R2                 0.544                   0.677
Residual Std. Error    1.049 (df = 46)         0.883 (df = 45)
F Statistic         57.132   (df = 1; 46)  50.219   (df = 2; 45)
```

"The lagged inflation rate is very significant in the second model ($t \approx 4.5$), and the adjusted R-squared much higher than that from the first. Nevertheless, this does not necessarily mean that the second equation will produce a better forecast for 1997. All we can say so far is that, using the data up through 1996, a lag of inflation helps to explain variation in the unemployment rate."

"To obtain the forecasts for 1997, we need to know $unemployment$ and $inflation$ in 1996. These are 5.4 and 3.0, respectively. Therefore, the forecast of $unem_{1997}$ from the first equation is $1.572 + .732(5.4)$, or about 5.52. The forecast from the second equation is $1.304 + 0.647(5.4) + 0.184(3.0)$, or about 5.35. The actual civilian unemployment rate for 1997 was 4.9, so both equations over-predict the actual rate. The second equation does provide a somewhat better forecast."