Appendix B

This appendix describes the concept and construction procedure of the lead profile to evaluate leading indicators.

Evaluation of a Leading Indicator: The Lead Profile[6]

Testing for Cyclical Leads

It has long been recognized that leading indicators can be a valuable forecasting tool for forecasting cyclical turning points. They have, however, not always been properly evaluated. One method of evaluating leading indicators that has gained some popularity in recent years is the Granger causality test. It is thus interesting to note what Granger and Newbold (1986) have to say about the difficulty of evaluating the index of leading indicators:

"The index of leading indicators has become a widely quoted and generally trusted forecasting tool. However, it has been rather misinterpreted. The index is intended only to forecast the timing of turning points and not the size of the forthcoming downswing or upswing nor to be a general indicator of the economy at times other than near turning points. Because of this, evaluation of the index of leading indicators by standard statistical techniques is not easy."

This difficulty in evaluation has often led to flawed assessments of the performance of leading indicators, not necessarily based on their ability to anticipate turning points. Part of the problem has been a lack of familiarity with the standard methods of identifying turning points. Yet, since leading indicators are meant primarily to forecast business cycle turning points, the identification of turning points in time series is a sine qua non for an appropriate evaluation of their forecasting performance. In fact, an objective algorithm for turning point identification, based on a systematic codification of the judgmental procedures used for decades at the NBER, was devised almost three decades ago (Bry and Boschan, 1971), shortly after the creation of the index of leading indicators. The Bry-Boschan procedure has certainly stood the test of time.

Geoffrey Moore, who helped create the index of leading indicators (Moore and Shiskin, 1967), used the Bry-Boschan procedure extensively in the decades following its creation (e.g., Klein and Moore, 1985). Other users have included King and Plosser (1989), who provide a description of the procedure. Watson (1994) points out that the Bry-Boschan procedure provides a good way to define turning points since it is based on objective criteria for determining cyclical peaks and troughs.

The objective (though not mathematically simple) definition of turning points given by Bry and Boschan's algorithmic formulation of the classical NBER procedure makes it possible to evaluate the performance of leading indicators in terms of an objective measure of the leads of leading indicators at turning points. In that sense, the Bry-Boschan procedure permits a more appropriate evaluation of the performance of leading indicators.

Given the cyclical turning points of a potential leading indicator, it is possible to measure the lead of that indicator at each business cycle turning point. However, many leading indicators cover only a small number of cycles. Thus the evaluation of leading indicators by parametric statistical methods is usually constrained by the limited number of cyclical turning points covered by the data. In addition, the need to make a heroic assumption that the probability distribution of the leads has a standard functional form also precludes the use of parametric tests of statistical significance.

This appendix suggests a simple nonparametric test to evaluate the cyclical leads of leading indicators, and describes lead profile charts that graphically depict these leads in probabilistic terms, to aid in the selection and evaluation of leading indicators.

The Problem

A number of considerations go into the evaluation of any time series as a cyclical leading indicator. The main issue is the evaluation of the magnitude of the leads of a leading indicator compared with a reference cycle (such as the business cycle) at cyclical turns as well as their leads compared with one another when two or more series are being compared. In all of these cases, the magnitude (and even the direction) of the lead may vary from one turn to the next. The problem, then, is the statistical significance of the leads, or of the difference in leads, as the case may be.

We have cited Granger and Newbold (1986) who suggest, in effect, that standard statistical approaches to the evaluation of leading indicators may be fraught with problems. The simpler classical approach of just measuring the mean and standard deviation of the leads does not result in tests of statistical significance without an assumption that the probability distribution of the leads has a standard functional form. Thus, no tests of significance can usually be performed. Under such circumstances, simple nonparametric tests may be the most appropriate solution.

Appropriate Nonparametric Tests

Nonparametric tests are often called “distribution-free” because they do not assume that the observations were drawn from a population distributed in a certain way, e.g., from a normally distributed population. These tests also do not require the large samples needed to reliably estimate parameters of distributions assumed in parametric tests. Such tests should therefore be uniquely suited to testing the significance of leads, which may be small in number, and for which the probability distribution function is quite unknown.

Since the leads in question are differences in timing at cyclical turns (between a pair of indicators, for example), the appropriate nonparametric tests are those applicable to matched pairs of samples. The most powerful tests in this class assume interval scaled data (like temperature in degrees Celsius) where equal intervals at any point in the scale imply equal differences. Leads measured in months or quarters are at least interval scaled, so such tests can be used with data on leads.

The most appropriate test to assess the significance of leads within this class is the Randomization test for matched pairs. This test has a power-efficiency of 100%, because it uses all the information in the sample (Siegel, 1956), but it does not lend itself to manual computation for sample sizes greater than about nine pairs. In such cases, a simple computer program can be used.

The Randomization Test for Matched Pairs

The Randomization test (Fisher, 1935) is a simple and elegant way to test the significance of leads. The first step is to calculate the difference in timing at turns, that is, the leads of one indicator over another, or over the business cycle turning points. The null hypothesis, that these differences are not statistically significant, is to be tested against the alternative hypothesis that the leads are significant.

Now, some of the differences calculated in the first step may be positive, others negative. If the null hypothesis is true, the positive differences are just as likely to have been negative, and vice versa. So if there are N differences (from N pairs of observations), each difference is as likely to be positive as negative. Thus, the observed set of differences would be just one of 2^N equally likely outcomes under the null hypothesis.

Also, under the null hypothesis, the sum of the positive differences would, on average, equal the sum of the negative differences, so the expected sum of the positive and negative differences would be zero. If the alternative hypothesis was true, and the leads were positive and significant, the sum would very likely be positive.

The second step, therefore, is to sum the differences, assigning positive signs to each difference; then to switch the signs systematically, one by one, to generate all the outcomes which result in sums as high or higher than that observed. If there are R such outcomes, then the probability of the observed outcome (or a more extreme outcome) under the null hypothesis is $(R/2^N)$. In other words, the null hypothesis can be rejected at the $100(1-(R/2^N))$% confidence level.

An example of the manual computation involved is provided below.

Leads of a hypothetical leading indicator over business cycle troughs

The leads at troughs of this indicator compared to the business cycle troughs are 12, 4, 1, 0 and -27 months. The last figure represents a lag of 27 months. Although the convention is to use negative numbers for leads, and positive numbers for lags, it is simpler for the purpose of this exposition to think of leads as being positive, because we are, in general, concerned with the significance of leads, not lags.

The first step is to drop the zero-month lead from the analysis; keeping this observation would make no difference to the results, as is evident from the procedure for the Randomization test. Then N = 4, and the 4 observations are (12, 4, 1, -27), which add up to a sum of S = -10.

This sum S is now compared with the sums computed by starting with all positive numbers, and switching signs one by one so that the sums are in descending order until our sum of S = 10 is reached:

12	4	1	27	Sum =	44
12	4	-1	27	Sum =	42
12	-4	1	27	Sum =	36
12	-4	-1	27	Sum =	34
-12	4	1	27	Sum =	20
-12	4	-1	27	Sum =	18
-12	-4	1	27	Sum =	12
-12	-4	-1	27	Sum =	10
12	4	1	-27	Sum =	-10 = S

Since R = 9 sums out of 2^4 (i.e., 16) possible combinations are greater than or equal to -10, the probability of such an outcome under the null hypothesis ("leads not significant") is 9/16 = 0.5625, so that the null hypothesis can be rejected only at the 100 (1-0.5625)% = 43.75% level of confidence. Hence, the null hypothesis is not rejected for leads at troughs.

Lead Profiles

So far, the discussion has focused on the confidence level at which the null hypothesis ("leads not significantly different from zero") can be rejected in favor of the alternative hypothesis ("leads significantly greater than zero months"). Now, even if it is established that the leads are significantly greater than zero months, it might be interesting to know how much greater than zero months the leads are likely to be – for example, whether the leads are also significantly greater than one month.

This is easy to determine. All one needs to do is to subtract one month from each of the differences in timing at turns (already calculated in the first step of the Randomization test). Then, as before, one finds the confidence level at which the null hypothesis is rejected in favor of the alternative hypothesis that the difference in timing at turns significantly exceeds one month.

In this way one can also determine the confidence levels for the hypotheses that the leads exceed 2,3,4, …. K months – simply by subtracting 2,3,4, …. K respectively from the original differences before performing the Randomization test. We call this full set of confidence levels a "lead profile".

The lead profile is a graphical depiction of the leads in strictly probabilistic terms, that aids meaningful comparisons between the indices. It can be graphically represented in bar charts or “lead profile charts”. The question answered by this chart is whether the difference between the leads of the two indices is statistically significant.

The advantage of lead profile charts is that these use as input just the information on the length of the leads at each turning point. However, by gleaning statistical inferences from the data rather than relying solely on averages, and by displaying the results graphically, they afford additional insights into the significance of leads.

Another major advantage of lead profiles lies in the explicit statistical inferences that can be made about the significance of leads without making any assumptions about the probability distribution of leads, or any restrictions on sample size. These inferences can be made about the leads of a given cyclical indicator over a reference cycle, such as a set of business cycle turning points. They can also be made about the leads of one cyclical indicator over another, to assess whether one has significantly longer leads than the other. Moreover, it is convenient to put lead profiles in the form of bar charts, for easy and effective visual appraisal of the significance of lengths of leads.

---

[6] This section is based on Banerji (2000).