THE DANGERS OF DATA FITTING

The success of leading indicators in market economies encouraged many countries to try to adapt them to fit local conditions. The problem is that many economists tend to create models or simplified representations of the economy that assume recent trends will continue into the future—a surefire recipe for missing a turning

point. In some countries, economists took better-known leading indicators of the business cycle and made simplifying assumptions to argue that their models now incorporated cyclical risks. In other countries, economists insisted that because the structure of their economy differed from that of the United States, different indicators would be needed to predict their business cycles.

There is some truth to the latter idea. But the tendency to use simplified models based on past data inevitably led to glaring failures in recession forecasting. The Organization for Economic Cooperation and Development (OECD) helps each member country select indicators for its own economy, based in part on what local economists feel would work best given the unique circumstances of their own country. Yet these indicators seldom work (actually they seem to work in retrospect, but mostly because the OECD revises the trends in hindsight). A South Korean delegate, for example, reported at an OECD meeting in 1996 that the South Korean government's leading index, developed based on its ability to "predict" past turns in the economic cycle, failed to anticipate future turns. To correct this, the indicators were repeatedly refitted to include the new turning point data, only to fail again at a subsequent turning point. The delegate asked, "What good are leading indicators if they predict only past turning points?"

It was exactly the outcome that Moore had warned of. Some indicators might seem to provide a good fit for specific countries for a period of time, but the question was whether those indicators could accurately predict future turning points. He believed that the drivers of the business cycle should be the fundamental guideline in the selection of indicators, no matter what economy was being examined. Our own South Korean leading indexes have

predicted turning points there well, precisely because they do not reflect the limited circumstances of a particular time and place.

But while the principal drivers of the business cycle—like profits and inventories—remain the basis for accurate indicators in every country we examined, it was sometimes difficult to find suitable measures of those drivers. Frequently, the data in a country would be unreliable or biased in some manner.

For example, when we began working with Mexico's central bank in 1998 to develop a cyclical monitoring system, we used the jobless rate as a coincident indicator to help establish the business cycle chronology. Going back to the 1980s, the Mexican jobless rate occasionally fell as low as 2 percent. This ran against our understanding of the extent of joblessness in the Mexican economy. When we investigated further, we found that the jobless data were strongly biased by local definitions of joblessness. In Mexico, if a person had done any kind of work, loosely defined, in the last twelve months, he or she was considered employed. Clearly, many "unemployed" people fit this classification.

Although this made the Mexican jobless data unsuitable for many forecasting models, we were able to use the data anyway. The number itself was biased, and the actual level of unemployment was undoubtedly much higher. But the timing of the cyclical turns in the data confirmed the timing found in other coincident measures of the business cycle, such as production and sales. Although biased, the relative rise and fall of the unemployment figures worked just fine for our purposes in identifying and predicting the timing of economic turning points.

The United States, relatively speaking, has some of the world's best data, yet economists here, too, sometimes attribute

their forecasting failures to data inaccuracies. For example, until revisions in late July 2002 showed that U.S. GDP had declined in the first three quarters of 2001 rather than in just the third quarter, many prominent economists who had failed to predict the 2001 recession continued to claim there had been no recession at all. When they finally faced the truth that they had been wrong, they blamed it on bad data.

In developing countries where data are often sparse and flawed, the problem is worse. This makes it difficult to build forecasting models that rely on high data quality in order to work properly. Quite often, these are also the economies that can least afford the time, money, and effort needed to build a world-class system of statistical data collection. The lack of high-quality data in developing countries poses major problems for standard model-based approaches to forecasting. But, as the example of Mexico illustrates, when the focus is primarily on identifying turning points in the cycle, the knowledge of indicators that turn in durable sequences in country after country is invaluable in identifying the right selection of leading indicators.