Data sources and considerations

Q 6-02. What are the key characteristics of the data to choose?

The purpose of the forecast will determine many of the key characteristics of the forecast and these will drive the requirements for the data. For example, a forecast of daily cash balances will require higher frequency (that is, daily rather than monthly) data than will a macroeconomic forecast of annual GDP growth. Among the characteristics that must be selected in assembling a forecast are the frequency of observations or readings, the fineness or granularity of the data, and the time horizon.

Q 6-02.01. How does frequency limit a forecast?

Frequency refers to the time lapse between observations. It is not necessary for the observations to cover a period shorter than the observation frequency. One might have, for example, monthly forecasts of the quarterly or annual GDP figure. In fact, to the extent that changing perceptions affect behavior, such a statistic could be useful for some forecasts. Frequency will affect how often a forecast can be updated.

Q 6-02.02. How does granularity limit a forecast?

Granularity refers to the level of detail either in the period covered by an observation or the disaggregation of composite funds flows. For example, a report on closing cash balances for each day of the week that is released in a weekly bundle has a daily granularity, if a weekly frequency. From a disaggregation viewpoint, granularity refers to what level in the rolling up of data flows are the data reported. It might be possible, for example, to gather customs duties information for each entry point in the country, but this level of granularity may not make forecasting customs duties easier. Granularity will affect how low in detail a forecast can go. The challenge is to choose a level at which the differences among the elements at a particular level have significant differences that if captured will improve the forecast.

Q 6-02.03. How does the selection of a time horizon affect a forecast?

The selection of a time horizon will be driven by the purpose of the forecast and will be limited by the granularity and frequency of the data available. Together, these factors, along with the rate at which forecast error grows, will determine how far forward a useful forecast should project. Generating a forecast entails costs in resources and time. There is little point in investing what is needed to produce a forecast that provides detail at too fine a level to grasp its meaning efficiently , that is at too gross a level to make decisions, or that extends so far into the future that

forecast error overwhelms the data. In the last circumstance, the range about the point estimate that is attributable to error probabilities is greater than the point estimate itself. The forecast horizon should be set such that data (or projections) exist for the data flows that feed into the forecast for the same horizon, the time period is far enough out to use the forecast to take actions that are needed, and the forecast is not overwhelmed by forecast noise.

The definition of any time horizon will entail some arbitrary decision. To ease discussions in this chapter, however, the following rough categories are suggested:

- Short term – Up to three months forward from forecast release date
- Medium term – Three months to a year forward
- Long term – one year or longer forward.

Q 6-02.04. What tasks are needed to prepare the data for building forecasts?

Given that the object of the forecast has been carefully selected and the time horizon has been based on this, the first step is to select the variables that will be forecast and the data that will be used to generate the projection. Selecting the output variable is the step of answering the question of what is going to be measured.

Only after deciding on the output variable should one choose the input data. The data selected to build the forecast should have an arguable relevance or functional relationship to the predicted variable.[1] Describe clearly the relationship or linkage between the input and the output before making a selection. This avoids rationalizing a choice made with no real reason. Do not select variables just because they are easily available for use. Finally, be sure the data exist to cover the period being forecast. An accurate model constructed using contemporaneous or slightly leading variables is of little forecast value if the same data or good forecasts of them are not available for the period being estimated. For example, same-quarter GDP values may provide a very accurate forecast of customs revenue. This relationship is of little value for forecasting customs duties six months in advance unless a good forecast of GDP is available and has been tested to demonstrate its usefulness.

The next task is to locate the offices that have the input data and to secure their cooperation in providing the data timely and regularly.

[1] This assertion may seem obvious. Within the philosophy of statistics, however, there is a view that requiring independent variables to have an a priori relationship to the dependent variable in a regression model may actually lead to an unwarranted bias in constructing a model. Still, the method suggested here is recommended as a good starting point.

Q 6-02.05. Who has a role in developing funds forecasts?

Funds forecasts draw data from individual cash flows. In many governments, responsibility for managing these funds flows fall under various offices or ministries. The team responsible for creating the cash forecast must secure the assistance of teams responsible for:

- treasury management
- budget
- external and internal debt management
- macroeconomic policy and analysis
- tax and customs
- other revenues
- outlays, including all major spending units
- the central bank (or serving commercial banks)

The information technology office has to be a part of the design team. An integrated financial system should be able to generate the reports or data needed to develop and monitor the cash flow forecasts.

Q 6-02.06. What, then, is the forecaster's role?

The forecast's primary nature is that of an informational tool rather than as an accounting or mathematical result. Thus, the principal role of the forecaster is to compile and interpret information from many sources.

The forecaster must determine the scope of the forecast and identify sources of information. Since some, if not most, of the critical information may reside outside the treasury, the forecaster must ask other agencies for it. In soliciting the data, the forecaster must shape into the data into usable format and assemble pieces into a unified product. If the data are drawn from various sources, it often is necessary to transform the data to a common frequency, scale, and format for the uses of the forecast. This may be done either by the forecaster or by the source. For example, one may ask the revenue or customs authorities for a monthly estimate and break it down to daily by oneself or ask them to do this.

The forecasting process, then, is to gather information from reliable sources and apply non-political statistical methods to highlight critical features of the data and to develop a forecast. The resulting information must be disseminated to all interested parties with an interpretation of the findings.