2.0 Demand Forecasting for Existing Facilities

2.0 Demand Forecasting for Existing Facilities

A significant issue faced by public-sector transportation planners is determining the appropriate capabilities for various modal and intermodal facilities. Ideally, these facilities should be able to accommodate the demands placed on them with an appropriate amount of spare capacity. A basic set of information required by these planners is the expected demand for use of these facilities.

This chapter presents an introduction to forecasting freight demand for existing transportation facilities; and the next chapter presents an introduction to the more complex subject of forecasting freight demand for new transportation facilities. Additional information relating to the procedures described in this chapter is presented in Appendices C-E.

Sources of information on historic and current transport activity and facility usage are discussed in the first section of this chapter, and sources of economic forecasts are discussed in the second section. Section 2.3 presents a simple procedure for combining an economic forecast with historic data on transport activity to produce a forecast of transport demand, and several options for improving the quality of these forecasts are also described. Section 2.4 discusses several regression and extrapolation procedures that can produce forecasts of transport demand from time-series data. Finally, Section 2.5 discusses the identification and evaluation of alternative futures that should be considered by planners before determining any course of action.

■ 2.1 Current and Historic Data on Facility Usage and Transport Activity

The most readily available information about demand for an existing facility is information about past and/or current usage of the facility – i.e., about past and/or current demand for the facility under certain price and service conditions. If these price and service conditions are reasonably "normal" (e.g., there are no unusual supply constraints), this demand information can be used as the basis for forecasting demand under similar "normal" conditions. The procedures presented in this chapter use data on (or estimates of) past or current usage or transport activity as the basis for generating forecasts of demand.

There are three types of sources of data on facility usage and related transport activity:

- Data compiled by the facility operator;
- Data collected and published by Federal agencies and other public and private entities that monitor or analyze transport activity on a regional, state, national or international level; and
- Data collected as part of a special survey designed to supplement data available from other sources.

These three types of data sources are discussed in the following subsections.

Facility Data

Facilities that charge users directly for their services invariably collect usage data that is related to the fees collected, and they may collect additional data for their own planning purposes or because the data is required by a governmental agency. Usage data directly related to fee collections (e.g., facility usage by vehicles, vessels, containers, etc.) is likely to be quite accurate. However, detail is likely to be lacking on actual freight volume, commodities, origins and destinations, and, in some cases, whether freight is even being carried; and even when such detail is collected, the data may be of lower quality. Special surveys may be necessary if this detail is desired for forecasting or planning purposes.

Published and Proprietary Data

Appendix C contains information on approximately 35 compilations of data that are available from public or private-sector sources in printed or magnetic form. These sources vary in: their level of detail; the modes, commodities and types of movements covered; whether they incorporate information on all movements of a given type or on just a sample of such movements; and, in the case of sample data, the size and structure of the sample. Some of the more significant of these sources are:

- The ICC Carload Waybill sample public use file – contains tons, carloads, trailers, containers, revenue, commodity, and origin and destination BEA region[1] for a sample of rail shipments.
- Waterborne Commerce and Vessel Statistics – contains annual data on tons by commodity, harbor, waterway segment, direction, and type of movement (internal, coastwise, export or import) for all movements using domestic waterways.
- U.S. Air-Freight Origin Traffic Statistics – contains estimates developed by the Colography Group of annual weight, value and number of air-freight shipments for 73 industries by "market area" of origin.
- The 1993 Commodity Flow Survey – expected to contain estimates of tons and value of shipments by commodity, mode, and origin and destination state or NTAR;[2] eight modes will be distinguished (including private truck, for-hire truck, and air/surface parcel transport).

Special Surveys

Data from the above sources may be supplemented by information collected from special surveys conducted partly or primarily to contribute to the forecasting process. The type of survey to be used for this purpose depends upon whether or not the firms using the facility in question are known.

When the set of firms using a facility is known (e.g., from information maintained by the facility operator), a survey can be conducted of all or a sample of these firms. Such a survey can be designed to collect data on annual volume of usage by shipment characteristics of interest (e.g., shipment size, commodities, origins and destinations, etc.), as well as information on expected (near-term) changes in these volumes, use of competing facilities, and factors affecting the choice of facilities. When designing such a survey, it is important to limit the amount of information requested so that respondents do not find the survey to be burdensome.

When the set of firms using a facility is unknown or only partially known, an unbiased sample of firms using the facility cannot be constructed. Instead, the survey generally is designed to collect information on a sam-

[1] The Bureau of Economic Analysis (BEA) has divided the United States into 183 economic areas (or regions), each of which contains one or more cities and the surrounding hinterland.

[2] The National Transportation Analysis Regions (NTARs) are a set of 89 regions, obtained by aggregating the 183 BEA economic areas into larger units.

ple of movements by interviewing employees who are bringing shipments to or from an intermodal facility or who are transporting shipments over a modal facility (e.g., a road). This approach is commonly used for obtaining information on truck transport, the mode for which published data are sparsest. The North American Truck Survey (NATS), described in Appendix C, is performed by interviewing truck drivers at truck stops; and a special survey conducted on behalf of the State of Washington[3] was performed by interviewing truck drivers at weigh stations.

Movement-oriented surveys generally are limited to collecting data on a single movement (though the NATS survey collects data on both the truck's current movement, whether empty or loaded, and its preceding movement). Data collected can correspond to the annual shipment data collected from a firm (shipment size, commodity, vehicle type, origin and destination, etc.), though truck drivers and other carrier employees may have somewhat less complete information about these shipments than does the carrier. Also, the data are limited to individual shipments. Accordingly, a much larger sample is required to obtain a reliable indication of the overall use of a facility.

Information on facility usage obtained from movement-oriented surveys is most accurate when the facility is geographically confined (e.g., it is an intermodal facility or a relatively short road segment) and the survey is conducted at the facility. For geographically dispersed facilities, the sampling procedure may not be capable of surveying certain types of movement (e.g., short hauls that do not pass any survey locations, or overweight trucks that use bypasses to avoid weigh stations). Surveys conducted at truck stops are likely to pick up very few short movements (since, on short trips, drivers are relatively unlikely to stop at truck stops). Also, if multiple survey locations are used, movements on routes that pass more than one survey location are more likely to be sampled than movements on routes that pass only one such location (a type of sampling bias for which corrections can be readily developed).

Additional information on the design and use of special surveys is presented in Appendix D.

[3] William R. Gillis, Kenneth L. Casavant and Charles Howard, Jr., "Survey Methodology for Collecting Freight Truck Origin and Destination Data," The Gillis Group, Pitzville, Washington, July 1994.

■ 2.2 Sources of Economic Forecasts

In general terms, the freight transport system moves goods from the point where they are produced to a user or consumer of these goods, frequently via one or more intermediate storage facilities. The location and operation of the intermediate facilities have a significant influence on the character of transport demand (origins and destinations, shipment sizes, shipment frequency, delivery-time constraints, etc.). However, the most important determinants of transport demand are: the volume of goods that are produced and consumed; and the locations of production and consumption. For this reason, forecasts of freight demand most frequently are derived, at least in part, from forecasts of overall economic output or, better still, from forecasts of production and consumption.

Since economic forecasts have many applications aside from their use in forecasting transport demand, such forecasts are available from several sources. Accordingly, most forecasts of demand for freight transport are based to some extent on exogenous forecasts of changes in the economy. Potential sources of these forecasts are described below.

Several states fund research groups that monitor the state's economy and produce forecasts of changes in the economy. For example, the Center for the Continuing Study of the California Economy develops 20-year forecasts of the value of California products by two-digit Standard Industrial Classification (SIC) code. Similarly, the Texas Comptroller of Public Accounts develops 20-year forecasts of population for ten substate regions and 20-year forecasts of output and employment by one-digit SIC code and substate region; and a private firm produces 20-year forecasts of output and employment in Texas by three-digit SIC code.

Long-term economic forecasts also are available from two federal agencies. At 2½-year intervals, the Bureau of Labor Statistics (BLS) publishes low, medium and high 12 to 15-year forecasts of several economic variables, including real domestic output, real exports and imports, and employment, for each of 226 sectors (generally corresponding to groups of three-digit SIC industries).[4] Also, at five-year intervals, the Bureau of Economic Analysis (BEA) develops 50-year regional projections of population and personal income as well as employment and earnings by industry sector.[5] The BEA forecasts are published by state for 57 industries, and

[4] The most recent BLS forecasts are contained in U.S. Department of Labor, Bureau of Labor Statistics, *American Work Force 1992-2005*, Bulletin 2452, April 1994.

[5] See U.S. Department of Commerce, Bureau of Economic Analysis, *BEA Regional Projections to 2040*, Three Volumes, U.S. Government Printing Office, October 1990.

by metropolitan statistical area and BEA economic area for 14 industry groups.

In addition to the state and federal agencies, short and long-term economic forecasts are also available from several private sources. The private firms use government and industry data to develop their own models and analyses. Two of the better known private sources are DRI/McGraw Hill and the WEFA Group.

DRI provides national, regional, state, Metropolitan Statistical Area (MSA), and county-level macroeconomic forecasts on a contract or subscription basis. Variables forecast include gross domestic product, employment, imports, exports, and interest rates. DRI also produces short-term (2½ to 3 year) and long-term (20 to 25 year) industrial input and output forecasts for 250 industries (2, 3, or 4-digit SIC code). Industrial inputs include employment, energy, and materials used in production. These input/output forecasts are updated semiannually. Price and wage indices are also forecast for 650 different industries.

WEFA produces quarterly short (2½ to 3½ year) and long-term (10 and 25 year) and annual long-term (25 year) U.S. macroeconomic forecasts. Variables forecast include gross domestic product, employment, price indices, financial indicators, and foreign exchange rates. WEFA also produces short-term (3 year) output forecasts for 537 industries (at the 4-digit SIC level) on a quarterly basis and long-term (10 year) input and output forecasts for 480 industries semiannually.

■ 2.3 Economic Indicator Variables

A relatively simple procedure for deriving forecasts of transport demand from economic forecasts is to assume that demand for transport of various commodity groups is directly related to variations in corresponding economic *indicator variables*. These indicator variables can be used either to derive *annual growth rates* or to derive *growth factors* representing the ratios of forecast-year values to base-year values. The procedure requires data or estimates of transport activity or facility usage, by commodity group, for a reasonably "normal" base year as well as forecasts of growth in the corresponding indicator variables. The basic version of this procedure is:

1. Divide base-year transport activity or facility usage by commodity group.

2. Associate each commodity group with an economic indicator variable that is related to production or demand for that commodity group and for which forecasts are available from some exogenous source (e.g.,

transport of food products might be associated with production of food products.)

3. For each indicator variable, obtain either a *growth factor* by dividing its forecast-year value by its base-year value, or obtain a forecast *annual growth rate* (e.g., by determining the average annual growth rate implied by the variable's base-year value and its value in any forecast year).

4. For each commodity group, estimate forecast-year demand either by multiplying base-year activity by the corresponding growth factor or by applying the indicator variable's annual growth rate to base-year activity.

5. Aggregate the forecasts across commodity groups to produce forecasts of total transport demand and forecasts of transport demand for any set of commodity groups of interest.

The most desirable indicator variables are those that measure goods output or demand in physical units (tons, cubic feet, etc.). However, forecasts of such variables frequently are not available. More commonly available indicator variables are constant-dollar measures of output or demand, employment, or, for certain commodity groups, population or real personal income.

Some Examples

The Vessel Traffic Services Study

One example of the use of economic indicator variables is a set of forecasts of waterway freight traffic and freight-vessel traffic developed by Jack Faucett Associates for the Volpe National Transportation Systems Center (VNTSC) and for the U.S. Coast Guard.[6] Forecasts of traffic were required for study zones surrounding 24 major ports in order to estimate the value of Vessel Traffic Service (VTS) systems being considered to enhance the safety of vessels traveling to and from these ports.

For the VTS study, base-year data on freight (and vessel) traffic were obtained primarily from the U.S. Army Corps of Engineers (COE) commodity and vessel traffic files for 1987. For all but one of the study zones of interest, this file provided estimates of import, export, and domestic

[6] Jack Faucett Associates, *Commodity and Vessel Traffic Forecasts*, Task Report, prepared for the Transportation Systems Center, Cambridge, Massachusetts, March 1991.

freight traffic, in tons, by commodity and direction, for several waterway segments, for each of 159 commodity groups. Movements of four of these commodity groups were dropped from consideration because forecasts were not needed. On the other hand, a separate commodity code was created for liquefied natural gas (LNG), a commodity of particular concern for the VTS study; and information from the LNG import terminals was used to separate out base-year LNG movements from other movements of "petroleum and coal products, not elsewhere classified." For the one study zone for which COE data were not available (the Santa Barbara Channel), base-year estimates of freight traffic by commodity were derived from VNTSC estimates of vessel traffic through the channel and from commodity data for Los Angeles/Long Beach.

Forecasts of commodity traffic for four forecast years (1995, 2000, 2005, and 2010) were developed using annual forecasts for the 1986-2000 time period developed by the Bureau of Labor Statistics (BLS) in 1988.[7] The forecasts used were the moderate-growth forecasts of real domestic output, exports and imports, by industrial sector. For these purposes, a correspondence was developed between 127 of the BLS's 226 sectors and the 155 commodity groups for which forecasts were required. (The BLS sectors used were the 126 goods-producing sectors plus the scrap sector.)

For each commodity group, the average annual growth rate in real output of the corresponding BLS sector or sectors was determined, as were the corresponding growth rates in real exports and real imports. For each commodity group, these growth rates were then applied to the base-year estimates of domestic movements, exports and imports, respectively, to produce forecasts for each of the forecast years of interest.

For three commodity groups of special interest to the study, the above forecasts were modified on the basis of additional data; and for a fourth commodity group, a separate forecast was used. Forecasts of coastwise shipments of petroleum products for several ports were modified to reflect Bureau of Economic Analysis employment forecasts[8] for oil and gas extraction in Alaska and for petroleum refining in Texas and Louisiana. Forecasts of crude-oil imports entering three Texas port areas were adjusted to reflect the effect of a planned offshore petroleum terminal, using information from persons involved in the planning effort. Also, relatively conjectural forecasts of LNG imports were developed from data on 1990 LNG imports at two terminals and from information about capacity at these two terminals and at a third that was expected to resume

[7] U.S. Department of Labor, Bureau of Labor Statistics, *Projections 2000*, U.S. Government Printing Office, Washington, D.C., March 1988.

[8] U.S. Department of Commerce, Bureau of Economic Analysis, *BEA Regional Projections to 2040*, U.S. Government Printing Office, Washington, D.C., June 1990.

operation at the time of the study; and the forecasts for imports of all other petroleum products were reduced to be consistent with the forecasts of LNG imports.

The California Freight Energy Demand Model

The California Freight Energy Demand (CALFED) Model was developed by Jack Faucett Associates for the California Energy Commission in 1983.[9] This model is used by the Commission and by the California Air Resources Board for forecasting truck and rail-freight activity and energy consumption. These agencies expect to update and expand the model within the next two or three years.

The CALFED Model develops forecasts of truck and rail-freight traffic for eleven commodity groups for five regions of the state and additional forecasts of overall truck (freight and non-freight) activity by vehicle type and region. Forecasts of truck and rail freight activity are developed by applying growth factors to base-year estimates of activity by commodity, region, and vehicle type or railroad-car type.

Exhibit 2.1 lists the eleven commodity groups distinguished by the model and the corresponding economic indicators used for deriving the growth factors. California forecasts of all indicators shown in the exhibit are produced regularly by the Center for the Continuing Study of the California Economy (CCSCE). The CALFED Model uses forecasts expressed in physical units, where available, and forecasts of value of output or employment in most other cases; forecasts of population are used to derive growth factors to be applied to household-goods transport. The model uses exponential interpolation and extrapolation to derive forecasts for years in which CCSCE forecasts are not available.

The following is a somewhat simplified description of the development of base-year (1977) estimates of truck and rail traffic.

Base-year estimates of truck transport of manufactured goods were developed using 1977 Commodity Transportation Survey data[10] on movements between eight BEA economic areas in California and between these areas

---

[9] Jack Faucett Associates, *California Freight Energy Demand Model*, prepared for the California Energy Commission, Sacramento, California, June 1983.

[10] U.S. Bureau of the Census, 1977 Commodity Transportation Survey, special computer tabulations prepared for the Transportation Systems Center, Cambridge, Massachusetts. (This Census survey was last conducted in 1977 and has since been replaced by the Commodity Flow Survey discussed in Section 2.1 and in Appendix C.)

Exhibit 2.1 Economic Indicators Used by the CALFED Model

Commodity Groups	Economic Indicators
1. Fruits and Vegetables	Food Products (tonnes)
2. Other Agricultural	Food Products (tonnes)
3. Construction and Mining	Employment in construction
4. Timber and Lumber	Lumber, plywood, etc. (board feet)
5. Food Products	Food products (tonnes)
6. Paper Products	Paper products (tonnes)
7. Chemicals	Chemicals (1972 dollars)
8. Primary Metals	Primary metals and transport equipment (1972 dollars)
9. Machinery	Machinery (1972 dollars)
10. Other Manufacturing	Cement and glass (tonnes); output of SIC codes 22, 23, 25, 27, 30, 31, 34, 38, and 39 (1972 dollars)
11. Household Goods	Population

and 165 economic areas in the rest of the country. Estimates of ton-milesin each of the model's five substate regions were derived using likely mileages within each of these regions for movements for each of the origin/destination (O/D) pairs of BEA economic areas. For inter-state movements, separate mileages were assigned to each of several entry/exit routes (shown in Exhibit 2.2), and all O/D pairs were associated with one of these routes. Traffic through California originating and terminating in other states was assumed to be negligible (a simplifying assumption that could not be made for most other states).

Base-year estimates of truck ton-miles of nonmanufactured commodities were derived from 1977 Truck Inventory and Use Survey data on the VMT of heavy vehicles serving the corresponding sectors, estimates of effective average payload by commodity group, and, in the case of agricultural products, additional data from other sources.[11]

Base-year estimates of rail ton-miles by commodity group and California region were derived from 1977 railroad waybill data[12] using a variant of the procedure used for truck transport of manufactured goods.

Improving the Forecasts

The basic economic-indicator procedure, presented above, makes the simplifying assumption that, for any transport facility, the percentage change in demand for transport of each commodity group will be identical to the percentage change in the corresponding indicator variable. However, for various reasons, the two percentage changes are likely to be somewhat different from each other. These reasons include changes over time in:

- value of output per ton;
- output per employee;
- transportation requirements per ton; and

[11] Jack Faucett Associates, *The Multiregional Input-Output Accounts, 1977: Interregional Commodity Flows*, Volume VI, prepared for the U.S. Department of Health and Human Services, August 1982; U.S. Department of Agriculture, *Agricultural Statistics, 1980*, U.S. Government Printing Office, 1980; and U.S. Department of Agriculture, *Fresh Fruit and Vegetable Shipments*, Calendar Year 1978, FVUS-7, Washington, D.C., July 1979.

[12] U.S. Department of Transportation, Federal Railroad Administration, 1977 Waybill File, Washington, D.C.

Exhibit 2.2 Average Mileage in California for Interstate Truck Movements

		Mileage by Freight Model Region				
BEA Economic Area	**Entry/Exit Route**	**(1) San Francisco**	**(2) Los Angeles**	**(3) San Diego**	**(4) Sacramento**	**(5) Rest of State**
164. San Diego	I-5		138	60	71	520
	I-15		276	63		98
	I-8			70		
165. Los Angeles	I-5		68		71	581
	I-80		68		208	332
	I-40		244			5
	I-10		242			5
166. Fresno	I-5				71	403
	I-80				182	208
	I-15		150			159
	I-40		198			159
	I-10		256			151
167. Stockton	I-5				71	268
	I-80				198	73
	I-15		150			297
	I-40		198			297
	I-10		256			289
168. Sacramento	I-5				36	235
	I-80				146	55
	I-15		150		32	330
	I-40		198		32	330
	I-10		256		32	322
169. Redding	I-5					117
	I-80					196
	I-15		150		71	445
	I-40		198		71	445
	I-10		256		71	437
170. Eureka	I-5					164
	I-80					356
	I-15	129	150			450
	I-40	129	198			450
	I-10	129	256			442
171. San Francisco	I-5	80			28	235
	I-80	74			199	42
	I-15	55	150			313
	I-40	55	198			313
	I-10	55	255			296

- competition from other facilities and modes.

To the extent that the likely effects of these changes are understood and can be estimated at reasonable cost, the basic procedure should be modified to reflect these effects. These effects are discussed below.

Value Per Ton

For most commodity groups, the relationship between value of output (measured in constant dollars) and volume shipped (measured in pounds, tons, cubic feet, etc.) may change over time. These changes may be due to a change in the mix of commodities being produced within a given commodity group (e.g., more aluminum and less steel) or a change in the average real value per ton of major products within the group. These changes may result in changing value per ton in either direction. (The shift to personal computers from mainframes provides an important example of a product category, computers, in which the value per ton, or per pound, has decreased appreciably.)

When transport demand is being forecast for several different commodity groups, adjustments for expected changes in value per ton for all commodity groups will be relatively expensive to make and may not have a very significant effect on the overall forecast of transport demand. However, when there are one or two commodity groups that are of particular interest, some consideration should be given, at least in an informal way, to determining how real value per ton for these groups has been changing and how it is likely to change over the forecast period.

Output Per Employee

Employment is related to transport demand less closely than is real output. Hence, employment is a less desirable indicator variable. However, long-term forecasts of employment are more available than forecasts of output, so that, for some purposes, employment forecasts must be used.

As a result of improvements in labor productivity, real dollar-valued output per employee increases over time, and physical output (in tons or cubic feet) tends to increase as well. Forecasts of the overall increase in real dollar-valued output per employee for goods-producing industries (agriculture, mining, construction, and manufacturing) can be obtained from DRI/McGraw-Hill. In order to avoid a downward bias in the forecasts of transport demand, forecasts of percentage change in employment should be converted to forecasts of percentage change in (real dollar-valued) output by multiplying by estimated compound growth in labor productivity over the forecast period.

Transportation Requirements Per Ton of Output

Decreases in the real cost of transportation that have occurred over time have resulted in a general tendency for industry to increase its consumption of transport services in order to economize on other factors of production. This tendency has resulted in trends toward decreased shipment sizes and increases in both lengths of haul and standards of service, with the last effect resulting both in a demand for premium quality services (e.g., just-in-time delivery) provided by traditional modes and in diversion to more expensive modes that offer faster, more reliable service.

Statistical analyses, using procedures such as those presented subsequently in Section 2.4, should provide useful data for forecasting the extent to which these trends are likely to increase the overall demand for freight transport. However, similar analyses of the secular shift toward higher quality modes are unlikely to produce reliable results because of the difficulty in controlling for temporal changes in modal service quality.

Competitive Factors

Whenever relevant, forecasts of demand for a facility or mode should be adjusted to reflect expected changes in degree of competition from other facilities or modes. These changes may result from:

- expected changes in relative costs;
- the elimination of base-year supply constraints at the facility in question or at competing facilities; or
- the development of future supply constraints at the facility in question or at competing facilities; or
- the development of new competing facilities.

The forecasting problems posed by base-year supply constraints frequently can be avoided by choosing a base year when no significant supply constraints existed. When this is not practical, a combination of historic data and judgment may be used to adjust the estimates of base-year facility usage to eliminate the effects of the supply constraints, thus producing estimates of base-year demand in the absence of supply con straints; annual growth rates or growth factors can then be applied to these estimates of base-year demand to produce the forecast demand.

2.4 Statistical Techniques

Regression Analysis

One alternative to the use of economic indicator variables is regression analysis, an alternative that has a stronger theoretical underpinning. Regression analysis involves identifying one or more independent variables (the explanatory variables) which are believed to influence or determine the value of the dependent variable (the variable to be explained), and then calculating a set of parameters which characterize the relationship between the independent and dependent variables. For our purposes, the dependent variable normally would be some measure of freight activity (e.g., ton-miles) and the independent variables usually would include one or more measures of economic activity. For forecasting purposes, forecasts must be available for all independent variables. These forecasts may be obtained from exogenous sources or from other regression equations (provided that the system of equations is not circular), or they may be developed by the forecaster using other appropriate techniques.

For forecasting purposes, regressions normally use historic *time-series* data[13] obtained for both the dependent and independent variables over the course of several time periods (e.g., years). Regression techniques are applied to the historic data to estimate a relationship between the independent variables and the dependent variable; and this relationship is applied to forecasts of the independent variables for one or more future time periods to produce forecasts of the dependent variable for the corresponding time periods.

Software to estimate the coefficients of the independent variables is widely available, easy to use, and, once the data are assembled, very inexpensive to run. Software ranging from spreadsheets to advanced statistical packages such as SAS, SPSS, and TSP provide capabilities for running regressions. The researcher need only enter data for the independent and dependent variables and invoke the proper command to produce the parameter estimates. The packages also present the researcher with some statistical measures, discussed below, which can be used to assess the appropriateness of the model.

Appendix E contains an introduction to regression analysis along with references to several textbooks. Some of the basic requirements for using

[13] An alternative to time-series regression is cross-sectional regression which uses observations of the dependent and independent variables across a set of similar entities (e.g., states, industries or firms).

regression techniques for forecasting transport demand are discussed below.

Some Basic Requirements

The use of time-series regression analysis requires the availability of historic time-series data for the dependent variable and also for all independent variables that have a significant influence on the value of the dependent variable or for proxies for these independent variables. A frequently used proxy variable is "time" which can be used as a proxy for any influences (e.g., value per ton or output per employee) that tend to have had a uniformly increasing or decreasing effect over the historic time period and are expected to have a similar effect over the forecast period. However, care is required in using time as a proxy to make sure that it does not capture historic trends (e.g., modal diversion) that may not be expected to persist into the future.

A related issue is the use of transport activity as the dependent variable. As observed in Section 2.1., transport activity actually represents transport demand under certain price and service conditions. If these conditions were reasonably constant over the historic period being used for the regression and they are expected to remain constant over the forecast period, they need not be explicitly represented in the regression. However, any price and service conditions that have varied significantly or are expected to vary should be represented by the independent variables or otherwise be given special treatment.

Of particular concern in the use of transport activity as the dependent variable is the effect of any supply constraints at the facility of interest or at a competing facility. Such constraints have no effect on transport demand; however, they can have a significant, but temporary, effect on transport activity. If such constraints only affect transport activity in a single historic time period (e.g., a single year), it is appropriate to exclude data for that period from the regression. However, if activity is affected in several time periods, it may be preferable to represent the effect of the constraint in the regression, possibly by using a *dummy* variable (i.e., a variable that has a value of one in time periods when the effect is present and is zero in other time periods).

Other Time-Series Techniques

Unlike regression analysis, which is based on a presumed theoretical relationship between dependent and independent variables, classic time-series methods[14] are not based on economic theory or interaction. Time-series models do not explain behavior and consequently do not provide a basis for policy analysis. Instead, the basic time-series methods are sophisticated tools of extrapolation which allow the past behavior of a variable to be characterized and projected into the future.

Time-series models require less data than regression models. Historic data are required only for the variable to be forecast (e.g., ton-miles). The models implicitly presume that: the effects of all of the variable's significant influences (e.g., growth and cyclical variation in economic activity, changes in competition from competing facilities/modes) can be adequately captured by an analysis of the historic changes in the variable itself; and that, during the forecast period, these influences will not change in character (e.g., the character of the business cycle will not change and overall economic growth will not significantly accelerate or decelerate). The requirement that the influences on the variable to be forecast not change makes these techniques more appropriate for short-term forecasting than for developing the long-term forecasts that usually are required for facility planning. Also, it should be noted that, like all other techniques presented in this chapter, the use of historic data on transport *activity* to represent transport *demand* presumes that the activity data used do not reflect the effects of any significant supply constraints.

Time-series models require that the variable to be forecast be "stationary," a situation in which its random or stochastic properties do not vary with respect to time (i.e., its mean value, its variance, and its covariance with other observations of the variable are independent of time). A time series, Y that contains a trend can often be made stationary (detrended) by differencing. The resulting new series, Y^*, is then used as the input for the time-series analysis.

The most common nonregression time-series model is the ARIMA model, an acronym for AutoRegressive Integrated Moving Average. ARIMA tools are widely available in statistical software packages and spreadsheets.

An ARIMA model requires the analyst to specify three parameters, p, d, and q:

> p is the order of the autoregressive dimension of the model, i.e., the number of lagged values of Y^* in the model;

[14] The seminal work on time series methods is G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco, 1970 (revised edition, 1976).

d is the number of times Y is differenced to achieve the stationary form Y*;

q is the number of lagged values of the error term which represents the moving average component of the model.

To develop and use an ARIMA model, an analyst follows three steps:

1. model identification, in which the values of *p*, *d*, and *q* are determined;
2. estimation of other model parameters;
3. verification that the model is satisfactory.

Of these steps, model identification is most critical and most challenging. The analyst must interpret several statistics, including a *correlogram,*[15] to determine which model specification is best for the data series in question. ARIMA is often considered a partial art form because there is much room for interpretation.

Some more easily used time-series methods are exponential smoothing techniques and curve fitting.

Exponential smoothing involves removing the random fluctuations in a data series to establish its underlying pattern and using that pattern to develop forecasts. Forecasts developed through smoothing are most appropriate for a short horizon in which the underlying trends of the past are expected to continue to be the primary determinant of the variable's value.

Curve fitting estimates how well a time series "fits" or can be described by a standard mathematical function ("curve"). Some of these functional forms are very simple (such as a straight line) while others are more complex (such as a logistic curve). Most software packages provide a variety of functional forms to use for evaluating the data series and allow the analyst to project the curve beyond the estimation period. Forecasts developed in this way also are most appropriate for short-term use.

Time-series methods are discussed further in Appendix E.

---

[15] A *correlogram* is a plot of the autocorrelation coefficient, r_k. Its pattern can often reveal the particular form of the ARIMA model to an experienced analyst. For a good discussion of correlogram patterns and the specifications they suggest see Peter Kennedy, *A Guide to Econometrics*, Third Edition, MIT Press, Cambridge, Massachusetts, pp. 260-261.

The Structural Econometric Time-Series Approach

One of the limitations of the ARIMA model and other time-series methods is that the analyses lack any explanatory power. There is no underlying theoretical relationship specified between the dependent variable and those factors which might influence or determine its value as there is in a regression model. The dependent variable itself contains all information needed to estimate its own future values. That specification is unsatisfying to analysts who are interested in estimating how changes in other variables affect the dependent variable. However, time-series models often provide more accurate forecasts than the structural models, so they are not devoid of value.

Econometricians who were dissatisfied with the lack of a theoretical basis for time-series methods eventually developed a synthesis which combines the structural and time-series models. An approach known as the structural econometric time-series approach (SEMTSA) was one of the results of this effort. As Kennedy explains:

> SEMTSA is based on the observation that dynamic structural equation econometric models are special cases of multivariate time-series (Box-Jenkins) processes in which *a priori* restrictions suggested by economic theory have been imposed on the parameters. Furthermore, if the exogenous variables in the econometric model can be viewed as being generated by a multiple time-series (ARIMA) process, then each of the individual endogenous variables in the econometric model can be expressed as a univariate Box-Jenkins ARIMA process.[16]

SEMTSA develops a traditional, theoretically grounded, structural model, derives the properties of corresponding ARIMA equations, and uses time-series methods to estimate the ARIMA equations. The results are checked for consistency with the structural model. If inconsistencies are noted, the proposed structural model is re-examined to try to identify its probable flaws. In this way, the time-series method is used to improve the structural model.

■ 2.5 Alternative Futures

The two preceding sections have presented procedures for producing a single forecast of freight demand. The goal of these procedures is to produce as good a forecast as is practical with available resources. Planning

[16] *Ibid.*, p. 249.

decisions can then be made on the basis of this forecast. However, planners are cautioned that the forecast is likely not to be completely accurate – either because some of the assumptions (e.g., those relating to economic growth) prove to be inaccurate, or because of deficiencies in the procedure itself.

Because no forecast can be guaranteed to be perfectly accurate, effective planning requires that planning decisions be reasonably tolerant of inaccuracies in the forecast. For some capacity-expansion projects, the cost of not being able to accommodate demand may be much greater than the cost of over expansion. For such projects, planners may wish to develop, or even to focus on, a "high likely" forecast of demand, and to use this forecast as the basis of expansion plans. On the other hand, for bond-financed projects, the greater concern might be that capacity utilization be high enough to generate revenue that is adequate to pay off the bonds. For these projects, a greater concern may be to identify the lowest level of future demand that is likely to occur.

The conventional approach to analyzing the effects of alternative futures is to subject a forecast to some form of *sensitivity analysis*. This approach is discussed in the first subsection below, but with an emphasis on forecasting those alternative futures that are of greatest concern.

A different approach to sensitivity analysis involves *starting* by identifying the alternative futures of concern and *then* identifying the conditions under which these futures could occur. This alternative, which we shall call *futures analysis*, is discussed in the second subsection.

Sensitivity Analysis

The development of any forecast requires a number of assumptions to be made, either explicitly or implicitly. Some of the types of assumptions that may be incorporated into forecasts of demand for a transportation facility relate to:

- Economic growth – both nationally and locally;
- Growth in the economic sectors that generate significant volumes of freight handled by the facility;
- Transport requirements of these sectors (which may be affected by increased imports or exports, or by changes in production processes);
- Modal choice (which may be affected by changing transport requirements or changing cost and service characteristics of competing modes);

- Facility usage per unit of freight volume (which may be affected by changes in shipment size or container size);
- The availability and competitiveness of alternative facilities;
- Value per ton of output; and
- Output per employee (if employment is used as an indicator variable).

Sensitivity analysis consists of varying one or more of these assumptions in order to produce alternative forecasts. The most common alternative assumptions to be considered are those related to economic growth; and, indeed, economic forecasters (including BLS) frequently provide high and low forecasts of growth in addition to a medium (or most likely) forecast. These alternative forecasts of economic growth can be used to generate alternative forecasts of transport demand, and additional alternative forecasts of exogenous variables (e.g., trade) can be used to produce an even larger set of forecasts of transport demand (e.g., high growth, high trade; high growth, low trade; etc.) However, simply varying these exogenous forecasts generally will not produce a set of transport-demand forecasts that represents the full range of demand that might exist in future years of interest. To produce a better understanding of the range of demand that might exist in the future, a more thorough sensitivity analysis should be conducted.

One approach to conducting a thorough sensitivity analysis consists of reviewing each of the assumptions explicit or implicit in the analysis and, for each assumption, generating a pair of reasonably likely alternative assumptions, one that would increase the forecast of demand and one that would decrease it. A high forecast of demand can then be generated by using all the alternative assumptions that would tend to increase the forecast (or at least all those that are logically compatible with each other); and a low forecast can be generated by using all the alternative assumptions that would tend to decrease the forecast. These high and low forecasts should provide planners with appropriate information about the range of transport demand that could exist in the future. Planning decisions can then be made that are designed to produce acceptable results for any changes in transport demand within the forecast range.

A somewhat more systematic type of sensitivity analysis consists of making small changes in the analytic assumption, one at a time, and determining the effect of each change on forecast demand. The results of this effort are a set of estimates of the *sensitivity* of the forecast to each of the assumptions. These results permit identification of the assumptions to which the forecast is most sensitive. These assumptions can then be reviewed and, if appropriate, improved; and a subjective determination can be made about the degree of confidence one has in the accuracy of the assumptions. Assumptions that are not deemed to be highly accurate can be varied and the implications of such variation can be determined – either by repeating the forecasting process using appropriate sets of alternative assumptions; or by making the simplifying (and not necessarily accurate) assumption that the effect of changing each of the analytic assumptions is linear and deriving alternative forecasts from the original forecast and the previously estimated sensitivities.

The second type of sensitivity analysis can provide more insight into the relationships between the various analytic assumptions and the forecasts produced. However, this approach requires a greater expenditure of resources. Furthermore, the most important sensitivity results – high and low forecasts of demand – can be generated using either approach, though these forecasts will be affected by the alternative analytic assumptions used to generate them and the care with which the high and low forecasts are then generated.

Futures Analysis

The preceding subsection discussed the use of alternative assumptions about the future and possibly about economic relationships as the basis for generating alternative forecasts of transport demand. In futures analysis, this process is essentially reversed. More specifically, futures analysis may be viewed as consisting of two steps.

1. Identity those alternative futures (e.g., levels of future demand) that would warrant a different planning decision than the one indicated by the original forecasts; and

2. For each such alternative future, identify the circumstances under which it might occur.

For some capacity-expansion projects, an alternative future of concern might be inability to meet future demand without further capacity expansion that could most efficiently be accomplished as part of the current project. For such projects, the second step of the futures analysis would include a determination of the conditions under which future demand might exceed planned capacity. Some of the possible contributing causes to be considered would include:

- Higher-than-expected economic growth;

- Higher-than-expected growth in the mode(s) served (due to changes in transport requirements of shippers or in cost and service characteristics of competing modes);

- Higher-than-expected growth in transport demand in the region served by the facility (due to unusual growth in production and/or consumption in the region); and

- A temporary or permanent loss of capacity at a competing facility.

If there appears to be a significant probability that future demand would indeed exceed planned capacity, further analysis would then be performed to obtain a better understanding of this probability and the expected costs and benefits of expanding the extent of the planned expansion.

For a bond-financed project, an alternative future of concern would be one in which demand for the facility would not generate the revenue required to operate the facility and to pay off the bonds. We assume that the total cost of financing the facility is reasonably well known and that the cost of operating the facility (as a function of usage) is also understood. Then, for a given user-fee schedule, the minimum level of usage that will pay the financing and operating costs of the facility can be estimated.

The second step of a futures analysis for this type of project would include a determination of the conditions under which future usage would fall short of the required minimum level. Possible contributing causes would include lower-than-expected growth in the overall economy, in the mode(s) served, or in the region served, as well as an unanticipated increase in competition from other facilities (including potential new facilities). If it is determined that usage may not be adequate under the assumed user-fee schedule, other user-fee schedules should be considered, incorporating appropriate adjustments in usage forecasts to reflect the effects of the alternative fees.