Projecting State and Area Industry Employment

Prepared for the
Projections Workgroup
Employment and Training Administration
U.S. Department of Labor

Harvey Goldstein, Ph.D.

The University of North Carolina at Chapel Hill
Department of City and Regional Planning
New East Building, Campus Box 3140
Chapel Hill, NC 27599-3140
www.planning.unc.edu/

Email: hgold@email.unc.edu

Seventh Edition
August 2005

Prepared for the Projections Workgroup through a grant from the
Employment and Training Administration
U.S. Department of Labor

Preface and Acknowledgements

This most recent edition of *Projecting State and Area Industry Employment* updates sections on data, including the conversion from SIC to NAICS, and adds new material on the use of dummy variables in regression models. It also includes a new section on the use of "new economy" indicators as independent variables in regression model specification. Other changes or additions reflect the organization of the Projections Workgroup and some modifications in the LTP software system developed and maintained by the Utah staff.

Larry Less of Ohio and Anthony Hayden of New York carefully read an earlier draft. They each provided a number of highly useful suggestions for improvement and identified errors I had missed. Dorothy Gattis and Steve Brock also provided helpful information on recent changes in data and the LTP software, respectively.

Earlier editions have benefited from a number of individuals involved in the projections process, including the many state analysts who have attended training sessions over the last 25 years and have provided valuable feedback and ideas for presentation of the material. They are too numerous to list, but some individuals deserve special mention: Neal Rosenthal, formerly Associate Commissioner of BLS; Dick Dempsey, formerly of NOICC; Alan Eck of BLS; and Stuart Sweeney of the University of California at Santa Barbara; in addition to Larry Less of Ohio.

Chapter 1

Introduction and Overview

This monograph has been written to provide technical guidance to the staff of labor market information (LMI) divisions of state employment security agencies or other public sector organizations that develop or use state and substate industry employment projections. The material is intended to cover the data, techniques, models, and procedures recommended for the development of the industry employment projections, that, in turn, are input to the development of occupational employment projections within the context of the Occupational Employment Statistics (OES) projections program of the U.S. Bureau of Labor Statistics (BLS) and the Projections Workgroup initiative of the Employment and Training Administration of the U.S. Department of Labor.

Technical support for the state and substate projections program has existed for many years. In 1981 BLS sponsored a set of week-long training sessions that NOICC continued through the early 1990s. Several editions of this guidebook were written during this period to serve as a technical reference for analysts embarking on this training or as a refresher for those who completed the training. Yet because the states varied in their data, techniques, and software for developing their projections, it was difficult to design the training material and the guidebook to maximally meet the needs of the analysts from all the states.

With the development of the Long-Term Industry Projection (LTP) System in the early 1990s, state LMI analysts finally had a common, comprehensive tool to develop industry projections. Throughout the 1990s and up until the present, the LTP system has gone through numerous changes and improvements. A major change in the classification of industry sectors – from the SIC to NAICS – has, of necessity, led to some changes in the procedures to developing the data base for the industry employment projections. Also, several of the projection models have been "tweaked" and the set of explanatory variables for specifying regression models have been rethought.

There have also been some notable changes in the larger context of state and substate employment projections, including how the Bureau of Labor Statistics develops its national projections, how state replacement rates are estimated, the incorporation of wage data in the OES program, the development of the short-term projection system and the integration of the long-term projection system, short-term projection system, and the micromatrix system into a "suite" of programs. Some of these changes and innovations have direct implications for the approach taken toward developing long-term employment projections. This edition of the monograph is written to reflect the most recent data, techniques, and software incorporated into the LTP system, as well as to introduce some new ideas and approaches to industry employment projections more generally. While it is not intended to serve as the technical documentation for the LTP system, it is meant to provide guidance about the analytic process that forms the basis for it.

1.1 Themes of the Approach

There are three closely related themes which underlie the approach toward developing state and substate employment projections described in this monograph and embodied in the LTP system.

The Projections Process Should be Analytical, Rather than Mechanical.

A strictly mechanical approach to developing projections most likely will lead to less than acceptable levels of projection accuracy. The low accuracy eventually will lead to the loss of credibility by projection users.

Also, (i) the greater uncertainty in national and international economies and markets, (ii) the greater "openness" of state and substate economies to the global economy, (iii) the more rapid rate of technological change, (iv) and the wide variation in rates of economic growth and performance between states and substate regions, have necessitated a more analytical approach to developing projections. The exercise of *analytical judgment* would include, for instance, identifying special local factors or conditions which might require adjustment of rates or ratios derived from technically accepted national data, and choosing the most appropriate projection models based upon the validity of their underlying economic assumptions.

While the projections process should not utilize a mechanical approach, it still should be a highly systematic one rather than a series of ad hoc procedures. A systematic process means following a logically connected set of steps, or elements. The projection process can and should be analytical *and* systematic at the same time. This can be achieved by considering that at each step there are choices among alternative procedures, models, or data. Analytical judgment is exercised in choosing the most appropriate option such that the *validity* and *usefulness* of the projections will be maximized within available time and resource constraints.

It follows that in adopting a more analytical approach to developing projections, the insight and experience of the LMI analysts themselves become the agency's most valued resource and most critical input in the entire projections process. Insight and experience can be gained in many ways, including attending training sessions and conferences sponsored by the Projections Managing Partnership (PMP), BLS, ETA, or other organizations. But perhaps analysts can learn the most by actually going through the complete projections process and then subsequently *evaluating* those projections. For instance, addressing such questions as: which industries were projected accurately, and which ones were not? Is there a pattern to those industries which had relatively large projection errors? Which assumptions proved to be incorrect? What relevant factors should have been taken into account? In summary, LMI staffs should adopt an ongoing projections evaluation process and strive to continue to climb the learning curve in developing projections by retaining good analysts and keeping them assigned to the projections program.

In adopting a less mechanical and more analytical approach, *pre-projection analysis* and *post-projection review* and adjustment play prominent roles in the process. Pre-projection analysis refers to basic investigations of the economic behavior of the state or substate area as a whole, and of the region's major or key industries, before actually starting the development of the projections. The results of these analyses should help the staff make better choices, decisions, and judgments at each subsequent step of the projections process. These may include choices of: economic assumptions about the projection period, whether to use national trends as proxies for state and substate trends, and which projection model specifications to use. Post-projection adjustment refers to using additional information and knowledge for adjusting the numbers produced by the quantitative models and techniques before final acceptance, publication, and dissemination of the projections. It is felt that a well-designed adjustment process which takes into account the experience and insight of both the LMI in-house staff and outside experts might be the single most critical element for improving the accuracy of state and substate area projections in an environment of rapid economic change. Yet, before the final projection review adjustment, intermediate review, evaluation, and adjustments should be undertaken within each major step in the process. This investment in effort at these intermediate points will lead to significant payoffs later in terms of increased accuracy and credibility of the projections.

The Projections Process Should be Cost-Effective

In facing the reality of budget constraints for the projections program, LMI divisions must make hard choices about how they can best use the limited resources available in developing the most accurate and useful projections as possible. This means, for example, *setting priorities* between industry and occupational groups in the development of projections. It would not be efficient to spend an equal amount of time developing projections for each detailed industry and occupation when some are more important and larger than others. In addition, choices among alternative techniques at particular elements in the projections process should take into account the differences in costs of implementing them. The analyst should consider whether the expected gain in accuracy or validity by using a more sophisticated technique is justified by the additional cost required to implement the technique. The maxim here is to use the most simple, least costly technique if "it works.” At the same time, it is hoped that continued research and evaluation on state and area projections will lead to further innovations that will enhance cost-effectiveness and that will be shared among the states.

The Projections Process Should be Pragmatic

It should be remembered that the ultimate purpose of the projections is to make more rational programmatic and planning decisions regarding job training, education, and career guidance. This implies that the projections should be developed, published, and disseminated to maximize the usefulness of the projections to a variety of decision-makers, while retaining their technical validity. To accomplish this, LMI divisions must understand the (often conflicting) information needs of various types of users before deciding such important technical issues as length of projection period, geographical coverage, and industry and occupational detail. They must understand how the validity of their projections might be affected by meeting various user requests or needs. And they should be able to explain to users those relationships in an understandable way. Finally, state LMI staff should understand and appreciate the importance of having a common set of procedures and standards among the states as envisioned by the PMP.

BLS-sponsored research on the usefulness of state and substate projections has suggested that the quality of the published reports is an important determinant of the extent to which the projections are actually used in decision-making (Cruze, Goldstein, et al. 1985). Presentation formats which emphasize communicability, and clearly written documentation of underlying assumptions and methods used should lead to increased credibility of the projections. The essential point here is that the most accurate projections will not be useful if they are not also understandable and credible.

1.2 The Audience for this Monograph

This monograph is designed primarily to serve as a reference manual for analysts who are active in the development of state and area industry employment projections. In conjunction with other material and organized training sessions, it also should serve as a training manual for those without any prior experience in the projections process. Since not all topics will be covered in maximal detail, a list of references is provided for the reader wishing for a more thorough or more rigorous coverage of particular topics. But with any science or art, however (developing projections is a combination), mastery comes with hands-on practice and experience. This guidebook will be most useful both to the highly experienced and relatively inexperienced analysts when it is read and studied while actually developing a set of projections using the LTP system.

1.3 Overview of the Industry Employment Projections Process

The process can be conceived as a system of steps that are logically linked together (i.e., the outputs of one step are required as one of the inputs at the next step). The subsequent chapters in this guidebook are organized around *four* major steps: (1) pre-projection decisions and issues; (2) developing the base-year industry employment estimates (and relevant historical time-series); (3) performing pre-projection analyses, and (4) developing the set of industry employment projections. The accuracy and usefulness of the end product, or output, of this system depends upon the cumulative accuracy of the steps along the way.

1.3.1 **Pre-Projection Decisions and Issues (Chapter 2)**

The starting point in the projections process is making some basic decisions about the data and methods that will be used. These decisions should begin with an assessment of the *users' information needs***,** and an understanding of the importance of all states using a common framework for the development of their projections. This assessment will help analysts decide (1) for which *geographical areas (substate***)** the projections will be developed, (2) for *which target years* the projections will be developed, (3) for what *levels of industry and occupational detail* the projections will be developed, and (4) what *employment concept* (e.g., wage and salary employment vs. total employment) will be consistently used in the process. In addition, pre-projection analyses should be undertaken to identify:

(i) the availability of data, especially for smaller substate areas;

(ii) any special factors affecting the relevant geographic areas that might influence the type of data (variables) and qualitative information that should be collected, the choice of industry employment projection models to be used, and the length of industry employment time-series to be collected;

(iii) the key industry sectors in each geographical area for which disproportionate resources would be allocated in the projections process.

In general these decisions will influence the choices of data, methods, models, and procedures, and the intermediate projection results in all the remaining steps of the process. Thus, they should be carefully considered before plunging into the more technical steps.

1.3.2 **Preparing the Base-Year and Historical Data (Chapter 3)**

The base-year industry employment estimates are required both for developing the set of industry employment projections and for developing the base-year staffing pattern matrix. Topics covered include: how to break out the Current Employment Survey (CES) data with Quarterly Census of Employment and Wages (QCEW) data for most industries, how to use the decennial Census and Current Population Survey (CPS) data for estimating base-year employment in non-CES or QCEW covered industries, importing data into the LTIP software system, and modifying and editing data within the LTIP system.

1.3.3 **Performing Pre-Projection Analyses (Chapter 4)**

Techniques covered in this chapter help the analyst decide what modeling strategies and approaches will most likely yield the most valid projections for each industry. The results of relatively simple techniques from regional economics that are incorporated into the LTIP system, including the location quotient, time-series plots, measures of the temporal stability of employment, comparative employment growth rates, and similarity/dissimilarity of industry composition, will enable the analyst to become more time-efficient when developing projections for the full set of state industries as well as for substate areas.

1.3.4 **Developing Industry Employment Projections (Chapter 5)**

Three broad classes of projection techniques are covered in this chapter for both state and substate areas: (1) share and shift-share models; (2) simple time-series (or trend) models; and (3) single-equation regression models. Within each of these classes a number of more specific models are described in terms of their assumptions and the conditions under which each type of model may yield valid projections. Regression models are given most emphasis, including guidelines for the analyst to choose the set of independent variables for a given industry sector that will produce a model that fits the historical data well and also performs well in a projection mode. Criteria and measures for evaluating projection models are presented. Finally, suggested guidelines for the important review and adjustment of the projections before dissemination to users is provided, including a discussion of taking into account labor supply considerations.

Chapter 2

Pre-Projection Decisions and Issues

There are several basic decisions that analysts will need to make before beginning the actual technical steps in the development of the projections. Some of these decisions will be based on the assessment of user information needs, while others will be based on considerations of data availability, data reliability, and the likely validity of the projections. Still others should be based on a preliminary analysis of the factors which are shaping the general economic and employment trends in the particular state or substate areas.

2.1 Geography: Projections for Which Areas?

All state LMI divisions develop *statewide* projections. Moreover, ETA now requires statewide occupational employment projections every two years as a deliverable from the One-Stop/LMI Core Product Funds ETA provides to each state's LMI office. Substate area projections are optional, but are highly recommended, particularly in the larger states with a large number of diverse labor market areas.

A relevant issue, then, is for which *substate areas* projections should be developed. In general, states should develop projections for substate areas that come as close as possible to labor market areas (LMAs). LMAs are designated by the Bureau of Labor Statistics. An LMA is defined as a geographic area consisting of a central community and contiguous areas which are economically integrated. The chief operating attribute of an LMA, is that within a given LMA, workers can change jobs without having to change their residential location. There are large and small LMAs. Large MSAs are economically integrated set of counties with a central community of at least 50,000 population. These are MSAs designated by the Office of Management and Budget and by the Census Bureau. Small LMAs are single counties or a set of contiguous counties economically integrated with a central community of at least 5,000 population.

If state policy or law mandates projections for administrative regions that are not approximately LMAs, then it is recommended that such projections be developed by adjusting projections for the LMA that is closest to the administrative region, or by "sharing" state projections, rather than by independent models. At the same time the LMI division should include caveats or warnings about the validity of such projections in projection publications or other dissemination media, so that the reputation and credibility of the LMI division is not damaged.

Beyond this, the decision involves considerations of *user needs*, *data availability and reliability*, and *projection validity*. The problem is that these considerations often conflict. The following are relevant concerns:

- Historical time-series employment data (QCEW, formerly ES 202) exist for individual counties, but there are no suitable data for generating projections for *portions* of counties, such as municipalities.
- Many states will not have converted SIC-based historical employment data to NAICS. This will be particularly true for non MSAs. Insufficiently long time-series may preclude the ability to analyze long-term trends or the use of regression models for projections.
- Projections for small areas tend to have low validity because of the volatility of employment levels due to only a few establishments in a given industry. Moreover, confidentiality restrictions on publishing projections in cases of a small number of establishments is a further reason not to develop projections for substate areas of low employment.

- Projections for portions of metropolitan or labor market areas can be very misleading since the labor supply can be met for any job opening from generally anywhere within the labor market area.[1] Also there are rather weak theoretical bases for projecting the location of many types of industry employment *within* labor market areas.

The types of substate areas for which projections *might* be developed, and the issues which may arise in each, are shown in Exhibit 2.1. The table indicates that consistent and detailed historical employment time-series data often are not available for geographic units less than counties. It should be noted that a number of states are now using GIS to spatially disaggregate ES 202 data below the county level. While this tool may allow the construction of historical employment time-series data for micro-spatial units, the issue of low theoretical validity for projecting employment at a spatial level of detail less than the LMA still remains. The validity of industry employment projections for non-labor market areas is low for at least three reasons.

- First, one is restricted to using the simplest, most naive projection techniques because the data requirements of more sophisticated techniques can not be met.
- Second, the "behavior" of industry employment time-series for small, non-metropolitan areas tends to be more volatile because of fewer establishments.
- Third, there is less theoretical guidance for understanding or explaining labor demand and supply in non-labor market areas.

There are also additional problems beyond those that arise in developing industry projections. Many non-labor market areas are not sampled for OES data. This may mean using statewide OES staffing pattern data instead, which may introduce significant error in the development of occupational employment projections. Many states use the Estimation Delivery System (EDS) (developed by the North Carolina LMI Division) to generate substate occupational estimates for non-sampled areas. Although these EDS estimates may contain anomalies; they are more accurate when the non-sampled area is "closer" to a sampled area. In general, while the use of statewide staffing patterns, national change factors, and national separation rates are often necessary because the data for estimating these at the substate area level may not exist, they become questionable for application to substate areas that are not representative, or typical of statewide or national patterns.

In a number of states LMI divisions have been instructed or mandated to develop projections for WIAs and economic development regions that do not conform to labor market areas.

It is recommended that LMI divisions place highest priority on developing projections for *metropolitan areas* and other *large labor market areas*. It is possible to develop valid projections for other areas which approximately conform to labor market areas *and* are either single counties or county aggregates. Yet because industry employment projection techniques available for these areas will be less analytically powerful (because of data constraints), the resulting projections should be interpreted cautiously. On the other hand, sub-county and sub-metropolitan area projections should not be attempted for the reasons discussed above. By

[1]A labor market area, for purposes of this discussion, is an area in which the very large majority of the labor demand and supply is located within this area. In principle, anyone living within the area can take a job located anywhere within the area.

EXHIBIT 2.1

Projections for Types of Substate Areas
Reliability and Validity Issues

Type of Substate Area	Principal Sources of Industry Employment Data	Data Reliability Problems	Projection Validity Issues	Other Issues
Major metropolitan MSAs		ES-202; CES 790	--	*Some economic variables not available in consistent annual time-series;
Small metropolitan MSAs		ES 202	--	*Insufficient length of time-series
Non-metropolitan LMAs	ES-202 (QCEW)	--	*Possible volatility due to small number of establishments	*Suppression (confidentiality) *Paucity of explanatory variables *Non-sampled OES
Multicounty regions (e.g.,WIAs, econ dev't areas)	ES-202 (QCEW)	--	*Volatility	*Suppression *Not suitable for demand/supply analysis *Paucity of explanatory variables if not an LMA *Non-sampled OES
Submetropolitan area (whole county)	ES-202 (QCEW)	--	*Not suitable for demand/supply analysis *Labor market theory not valid	*Non-sampled OES
Subcounty	Census of Population/ CPS; Censuses of Business	These data not suitable; Employment data from CPS is by place of residence; Frequency (every 5 years)	*Labor market theory not valid *Not suitable unit for demand/supply analysis	*Non-sampled OES

attempting. to develop these, credibility for *all* of the LMI division's employment projections may be harmed.

2.2 Choice of Time Periods: The Projection Year(s) and Data Requirements

Consistency in the base and projection years among the states is important for many data users. ETA now expects all states to develop long-term employment projections to the target year that conforms to the BLS' national employment projections. Yet because different users and clients have needs for different length of projection periods, some states will choose to develop projections for several different target years.

There is a definite relationship between length of the projection period and accuracy of the projections. The length of the projection period also has important implications for the choice of industry employment projection model, or technique. There are also data availability constraints to be considered. For example, the BLS national industry employment projections (and the national change factor matrix) are developed for specific projection periods and target years.

For states that are considering developing projections for additional target years, the following issues might be considered:

- Projection accuracy tends to decrease the longer the projection period beyond about 5 years, since uncertainty about target-year macroeconomic and sectoral conditions increases.

- The techniques available for developing detailed industry employment projections (except fully-specified econometric models) are not suitable for developing short-term projections of less than, say, two years. For these short-term projections *cyclical* factors play more important roles than secular or structural factors in affecting employment levels. These cyclical factors are difficult to take into account in most of the projection techniques. However, analysts now have access to the Short-Term Industry Employment Projections Module which is based upon analysis of the employment time-series rather than causal explanations of changes in industry employment.

- The longer the length of projection period, the longer the length of the historical employment time-series needed as a basis for extrapolating trends.

- The decision to develop state or substate projections for a target year which is not one for the BLS national projections will create substantially more work in data development and in projecting many national economic variables exogenously. For instance, national industry employment projections will have to be interpolated or extrapolated, national change factors for the matrix will have to be adjusted; and some national data used to develop the base-year matrix in some non-OES covered industries may not be available.

No matter how long the length of projection period, the analyst should utilize the longest possible historical time-series of industry employment. Even if regression models will not be used for every industry sector, a long historical time-series when plotted on a graph against time will be an important aid in the pre-projection analysis. Issues such as relative employment stability, cyclical turning points, structural change, and whether the state or area trend mirrors the national trend can be inferred from the examination of the historical time-series employment data of each industry sector. These inferences, in turn, can lead to better choices about which types of industry employment projection models and their particular specifications to use.

Decisions about collecting historical data for other variables should wait until decisions about specific industry employment projection model specifications are made (see chapter 5).

2.3 Choice of Employment Concept

In the past, states had to choose between developing projections for only *non-agricultural wage and salary employment* or *total employment*. The latter includes agricultural workers, the self-employed (SE) and unpaid family workers (UFW), in addition to non-agricultural wage and salary employment. In order to be consistent with the BLS national employment projections, ETA now expects all states to use the total employment concept. It is also preferred because it is more comprehensive: leaving out the self-employed from the projections, for example, will distort the actual picture of future labor supply and demand conditions in many occupations.

Having said that the total employment concept is preferred, some states develop projections for only "covered" employment because of the difficulty of putting together a consistent and accurate time-series of non-covered employment categories such as agricultural employment and railroad employment. Sources of employment data in the non-covered sectors are discussed in chapter 3.
The remainder of this monograph assumes the use of a total employment concept.

2.4 Decisions on Level of Industry Detail and Ownership

States will choose between developing projections at the 3-digit level of NAICS detail, 4-digit NAICS, or a combination of some 3- and some 4-digit NAICS. There are several considerations in making this choice. Perhaps the most important is the level of detail in the OES staffing pattern matrix. By developing industry projections at the same level of detail as the OES data, state analysts will be able to make maximal use of the information in the staffing pattern matrix and will avoid introducing additional occupational employment projection error as a result of heterogeneity of staffing patterns within more aggregate NAICS categories. The following issues should be considered when making decisions on the appropriate level of industry detail:

- OES staffing pattern matrix generally assumes industry employment at the 4-digit NAICS level (although some OES staffing pattern data are at the 3-digit and 5-digit NAICS level).

- In general, the higher the industry detail, the larger the effort needed to develop the industry employment projections, particularly when there are a large number of substate areas.

- Employment in large and key industry sectors should be projected at the most disaggregated level.

- Large dissimilarity of employment distribution between the nation and the state (or substate area) among the 4-digit NAICS industries within a given 3-digit NAICS category will often mean choosing the 4-digit NAICS detail. This is particularly true if the national industry employment trend will be used as a predictor of state or area industry employment.

- Projection evaluations generally show that the higher the degree of industry detail, the higher the projection error, on average.

- Evaluation studies also suggest that occupational employment projection error does not increase significantly when more aggregated industry data are substituted for more detailed industry categories in the staffing pattern matrix.

The degree of *variation* in staffing patterns among industries (from the most recent state and current national staffing pattern matrices) should be an important consideration in deciding where to use more aggregate industry employment categories.

In addition to the decision of appropriate level of industry detail is the decision of definition of industry in terms of ownership. For most industry sectors this is not an issue. But for some, the OES definition for purposes of estimating staffing patterns differs from that used by other BLS employment data. Because industry employment projections are used as input to the development of occupational employment projections, there is a good case to develop industry projections to match the OES industry definition for education, hospitals, postal service, and perhaps several others. Here, for example, employment in education within the public sector would normally be included under government in the ES 202 data, not under educational services. But because the staffing patterns of employment in public education would be much closer to the staffing pattern for educational services than for local government, the OES definition of education includes public schools and colleges and universities, in addition to private sector educational institutions.

2.5 Identifying Key Industries

LMI staff and budget constraints, as well as administrators' concern for cost-effective use of staff resources, dictate that in the projections process disproportionate attention be placed on the most important, critical, or key industry sectors, rather than treating each industry equally. There is no single set of specific criteria for designating the key industries, but some general guidance can be provided.

The identification of the key industries should be based, in part, on discussions with appropriate outside experts about current and prospective trends in the state (area) economy, about emerging sectors of the economy, and about sectors that are deemed strategic by elected and policy officials in state economic development plans.

As a starting point, analysts can start to compile rank-order lists of industry sectors (at a 4-digit NAICS level) in terms of:

- The largest in the state or substate area (based on most recent current employment data)
- The fastest growing in the state or substate area (based upon amount of employment growth over say, the last five years)
- The fastest *rate of growth* in the state or substate area
- The largest declines in employment in the state or substate area (based upon say, the last five years)
- Largest percentage of scientists, engineers, and technicians nationally (from recent OES Survey)
- Fastest growth rates nationally (over recent period or expected over projection period)
- Sectors deemed to be strategic in state or substate economic development policies.

One possible scheme for classifying industries for purposes of allocating staff time in the projections process is to create a three-level classification: (1*) key* industries, based upon indicators similar to those just described (2) the *least important* industries, and, (3) an intermediate group.

The "*least important*" category of industries might include those which meet some combination of the following criteria: small, neither significantly growing nor declining, not undergoing technological change, and having relatively large concentrations of workers in occupations not normally targets of job training programs and vocational education. The intermediate group would consist of the remaining industry sectors.

2.6 Identifying Broad Economic Factors and Trends in the State/Area Economies

This is a general research activity to be undertaken by the LMI staff prior to actually starting the projections. Its purposes are to gather relevant information that will help guide some technical decisions later, as well as to provide background information for the projection review and adjustment process. The following items represent examples of the types of questions that might be researched.

- What has been the degree of *conformity* between recent economic and key industry employment trends at the national level and those for the state or substate area? To what degree have they mirrored each other, or conversely, have followed independent paths?

- How sensitive is the state or substate area economy to national business cycles? Are the amplitudes similar or not? Has the timing been similar, in terms of duration of expansions, duration of recessions, turning points.

- What industry sectors comprise the state's or substate area’s export base?

- What is the geographic scope of the markets for the state's or substate area’s principal export sectors? Mainly regional? National? International?

- How has the state’s industry composition changed in, say, the last ten years, and how does that compare to changes in the nation’s industry composition?

- How *diversified or specialized* is the state or substate area economy?

- To what extent are the state's (area's) key industries part of industrial clusters versus independent of other industrial sectors?

- How much variation in economic and employment trends among the substate areas of the state?

- Are there peculiar characteristics or attributes of the state or substate area economy which may make it respond differently from the national average to exogenous economic or technological changes? For example, does a [particular substate area have a state capital, a concentration of universities, a concentration of military bases or defense spending, a concentration of oil/gas production-related industries, etc.?

- Is there a notable *functional specialization* of any the state's or substate area’s key industries: e.g., headquarters functions, production, research and development, distribution?

- Are there prospective changes in state or area public policies including legislation, regulation, or broad public attitudes on the horizon which might substantially affect the future rate of economic growth and development?

The methods for generating answers to these questions are not sophisticated and, for the most part, may be described as "quick and dirty." These include:

- Time-series trend analysis
- Location quotients
- Shift and share analyses
- Brief telephone interviews with experts on the state/area economy and on principal industries.

Some of these simple techniques are included in the LTIP system software and are described below in chapter 4.

Chapter 3

Preparing the Base-Year and Historical Data

In this chapter we review the preparation of the base-year industry employment estimates, the annual historical time-series of industry employment, and other economic data that will be used as predictor variables in the projection models. The preparation of the base-year industry employment estimates is a prerequisite for developing the staffing pattern matrix, while the historical time-series of employment and other variables for the state, substate areas, and the nation is a necessary step before conducting pre-projection analyses described in the next chapter. The analyst may also wish to modify data or transform variables, covered here, after developing an initial set of projection models.

3.1 Base-Year Industry Employment Estimates

There are two principal data sources--the Current Employment Survey (CES, or 790 data) and the quarterly UI Covered Workers (QCEW) reports—plus several supplementary sources for industries not covered by the above.

3.1.1 **CES-Covered Industries**

The CES data, available from state LMI divisions and then the Bureau of labor Statistics, generally are regarded as the best and most accurate estimates of current industry employment. The principal drawback of the CES data is that the level of industry and geographic detail is limited at state and substate levels due to reliability constraints imposed by the size of the samples of establishments. Some unpublished CES data for states and substate areas are available at more industry detail, but this is not sufficient to provide the desired industry detail (mostly 4-digit NAICS) for the staffing pattern matrix at accepted levels of reliability.

QCEW data, which are highly detailed, can be used to disaggregate, or "break out", the CES data. Yet because the coverage of QCEWdata is not as broad as the CES data in a few industries, it is better not to use it alone to provide current or historical industry employment estimates in these cases. When QCEWdata are used with the available CES data, however, the advantages of each source can be maximized while minimizing their respective disadvantages. In all cases, the CES and ES-202 should reflect the most recent benchmark adjustments. A simple example helps to illustrate the procedure for breaking out the CES data using QCEWdata (see Exhibit 3.1).

In the example, annual employment in a 3-digit NAICS industry for a given year is available from the CES but the 4-digit employment data are not. The annual average QCEW employment is calculated for each 4-digit NAICS industry, and then the proportionality factor for each is calculated (Column 5). Each proportionality factor then is multiplied by the given CES 3-digit NAICS employment estimate to obtain the final 4-digit annual employment estimate (Column 6). The CES 3-digit employment estimate thus serves as a control total for the constituent 4-digit QCEW employment estimates. If annual time-series are to be constructed, this procedure is repeated with similar employment data for each observation year to be included in the time-series. The calculations represented in Exhibit 3.1 can be programmed into the LTP system, or can be programmed and performed easily with a microcomputer-based spreadsheet template.

EXHIBIT 3.1

CES Disaggregation Using ES-202 Data

(1) Industry	(2) NAICS	(3) CES 790	(4) ES-202	(5) Proportion	(6) Final Estimate
Transp equip mfg	**336**	**32,700**	**32,966***	**1.0000**	**32,700**
Motor vehicle mfg.	3361	--	5,193	0.1575	5,150
Motor vehicle body.	3362	--	3,858	0.1170	3,826
Motor vehicle parts	3363	--	17,638	0.5350	17,495
Aerospace products	3364	--	2,062	0.0625	2,044
Railroad rolling stock	3365	--	**	0.0012	**
Ship/boat building	3366	--	3,896	0.1182	3,865
Other transp equip	3369	--	284	0.0086	281

Calculation:

STEP 1. Divide col. 4 row entry by col. 4, row 1 entry
= col. 5 row entry

STEP 2. Multiply col. 5 row entry by col. 3, row 1 entry
= col. 6 row entry

** Non-disclosed

3.1.2 **Non-CES Covered Industries**

The CES data cover only non-agricultural wage and salary employment. Assuming the use of a *total employment concept* in developing the set of occupational employment projections, then supplementary data must be used for obtaining estimates of non-covered workers. Non-covered workers include railroad workers, workers in many religious organizations, work-study students, employees of some non-profit organizations, commissioned sales workers, as well as agricultural workers, the self-employed (SE), private household workers, unpaid family workers (UFW), and federal government workers. Agriculture workers and private household workers must be estimated from Census and CPS data or from other supplemental information within the state. Federal employment estimates are provided by the CES and the QCEW data (the occupational distribution of federal employment is provided by the BLS OES survey staff to the Utah MicroMatrix Service Center). The self-employed (SE) and unpaid family workers (UFW) are estimated for all industries combined from the CPS and decennial Census. This information may also be produced in the MicroMatrix system by applying national ratios.

3.1.2.1 Agricultural Wage and Salary Workers

There are two principal data sources for agricultural employment estimates. These are the annual state CPS for wage and salary agricultural employment (total), and the decennial Census for employment in each of the 3-digit NAICS industries, 111 and 112, that are considered agricultural production. However, the CPS agricultural employment total should be used as the control total because the more detailed decennial census agricultural employment estimates are generally out-of-date. Before using the most recent annual CPS agricultural estimate, however, government workers in agriculture should be subtracted out using current CPS employment data by class of workers, if available. Otherwise they would be double-counted.[1] Detailed agricultural employment for NAICS codes 111 and 112 may be estimated by "breaking out" this control total using the latest decennial census data to calculating proportionality factors. This procedure is similar to that used when breaking out the CES data using QCEW proportionality factors (shown in Exhibit 3.1). However, this should be done only if CPS data are not available for a given state. An additional source of data for agricultural workers is the pentennial Census of Agriculture.

There are generally less current agricultural employment data available for substate areas than for states. Unless either unpublished census data or other sources can be made available, the analyst will have to make some assumptions about substate agricultural employment shares of state agricultural employment. One of the simplest assumptions to make is that a given substate share of state total agricultural employment remains the same from the decennial census year to the base-year. The current state CPS estimate then is broken out geographically, based on the latest decennial census' substate shares. Also, the full level of industry detail (3-digit NAICS) at the substate level may not be available. If this is the case, the analyst must use judgment in deciding how to disaggregate. For example, the analyst can assume statewide proportions of agricultural employment for all substate areas. Alternatively, the analyst could avoid this problem by "rolling up" the detailed agricultural employment industries in the staffing pattern matrix used for substate areas.
Finally, if supplementary agricultural employment surveys are used in a given state, then these should be considered for supplementing or even replacing the CPS and Census data for estimating current agricultural employment. The analyst should exercise his/her best judgment in deciding how the supplementary survey data may be used most appropriately, and what are the limits to their reliability and accuracy.

---

[1] Failure to subtract the government workers in agriculture because of data unavailability is not a serious problem, however, since the number is typically so small. See information provided by the LAUS program for identifying employment by class of worker to use in subtracting out government workers.

Data for forestry (NAICS 113) and fishing, hunting, and trapping (NAICS 114) should be handled in the same way as NAICS codes 111 and 112, but many states may opt to leave out these industries because of their small sizes.

3.1.2.2 Private Household (Wage and Salary) Workers

Analysts should use the most recent state CPS annual average to obtain estimates of wage and salary private household workers at the state level. For substate estimates, the analyst will need to make assumptions about the substate share of state employment. If one assumed constant substate shares of state private household employment, then one could calculate these percentages for the latest decennial census year. The same percentages then would be applied to the current state CPS estimate to obtain substate private household employment estimates.

3.1.2.3 Non-Farm Self-Employed

The annual state CPS, the decennial Census, and the BLS ratios file allow the analyst to estimate total self-employed workers (all industries combined).

3.1.2.4 Unpaid family Workers

The BLS ratios file can be used to estimate the number of the state's unpaid family workers.

3.1.2.5 Federal Government Employment

Current estimates of federal government employment for state and substate areas are obtained from the CES and QCEW. The estimates of federal government occupational employment by OES classification are sent by BLS to the States. They are compiled from data supplied by the U.S. Office of Personnel Management and the U.S. Postal Service.

3.2 Constructing Historical Annual Industry Employment Time-Series for Industry

The same procedures discussed above for obtaining base-year industry employment estimates should be used for preparing consistent annual time-series of wage and salary employment in all CES and non-CES- covered industries. The preparation of historical time-series employment is recommended for use in pre-projection analysis even if regression models are not used for the actual projections. As a practical matter, the analyst should construct annual time-series for all industries at the state level, and for at least all the key industries at the substate level.

Beginning in 2000, all LMI divisions had to deal with the change from the SIC to the NAICS-based employment data for developing consistent historical time-series as input to the industry employment projections and for maintaining compatibility with the industry sectors in the OES staffing pattern matrix.

The change from the SIC code to NAICS represents a discontinuity in the historical employment time-series, starting in 2001. The year 2000 was the last year employment data were collected and made available on an SIC basis. The Bureau of Labor Statistics, anticipating many of the problems the conversion causes for the states' projection process, converted annual SIC employment data for the nation and for states back to 1990.

These conversions were based on dual-coded SIC/NAICS data available from the Longitudinal Data Base (LDB) file. But in order to infer trends and to utilize regression or time-series models for projections, state analysts require a longer historical time-series than a starting year of 1990 permits. A ‘rough’ minimum number of annual observations for obtaining reliable estimates from OLS regression models is 20-25 years of historical data. This implies going back to *at least* 1980.

Many states have already reconstructed historical employment time-series prior to 1990. For those that have not, several resources are available on the Projections Workgroup web site (or contact Karen F. Duffy: duffyk@odjfs.state.oh.us) to provide guidance. An illustration of how to use ratios obtained from the LDB file to convert pre-1990 SIC annual employment data into NAICS (developed by Sandy Newman of Ohio) is shown in Exhibit 3.2 below.

A study to empirically test several options for extending the historical time-series prior to 1990 was conducted recently by this author. The conclusions from this study were that states should first use ratios obtained from each state’s dual-coded LDB file (and provided by BLS to all the states) to convert pre-1990 data on an SIC-basis to NAICS. But because these conversion ratios will all be based upon “snapshots” of the relationship between SIC and NAICS employment in 2002, errors will be introduced for estimates in earlier years. In general, the further one goes back in applying the ratios, the greater the magnitude of error. For this reason, the study also recommended the use of a dummy variable in OLS regression models for taking into account possible measurement error for years when annual employment estimates were converted from SIC to NAICS prior to 1990. The use of dummy variables in regression models will be described in more detail in chapter 5 below.

An additional recommendation is for state analysts to carefully examine the “new” historical employment time-series after applying the dual-coded ratios to the pre-1990 data. Results from the above-mentioned study found that for some industries, the application of the ratios from the LDB file by BLS may produce contaminated estimates even for the years after 1990. The best means for diagnosing this is for analysts to routinely produce a plot of the annual time-series (employment plotted against time) for each time-series in which SIC data were converted to NAICS, and then to visually examine the plot to see if there is an obvious discontinuity or outliers. This can be easily done with the LTP system and is discussed in section 3.6 below.

3.3 Time-Series for Non-Employment Variables

When regression models are to be used for developing industry employment projections, annual time-series must be constructed for all variables which are to be used as independent, or predictor variables. The selection of these will be discussed in Chapter 5 below. A number of national-level economic variables are maintained in historical time-series (as well as projected values) and provided by the BLS to the Projections Workgroup. These are pre-loaded onto the LTIP system. The variables include:

- Population
- Personal income; total, real, disposable, per capita
- Gross domestic product
- Consumption expenditures and investment
- Earnings, productivity, unemployment rate

Exhibit 3.2
Converting SIC to NAICS

Steel product manufacturing

Ratio	SIC	NAICS	1980	1981	1982	1983	1984	1985	1986	1987	1988	1990
		3312										11,822
	331		73,100	72,600	58,000	51,700	52,100	45,800	42,400	43,400	46,100	
0.4176			30,527	30,318	24,221	21,590	21,757	19,126	17,706	18,124	19,251	
	339		2,200	2,200	1,900	1,800	2,200	2,400	2,300	2,300	2,500	
0.0645			142	142	123	116	142	155	148	148	161	
Backcast NAICS 3312			30,668	30,460	24,343	21,706	21,899	19,281	17,855	18,272	19,413	
Time-series for 3312			30,668	30,460	24,343	21,706	21,899	19,281	17,855	18,272	19,413	11,822

SICs with ratios of less than 0.001 should be ignored, as they add little and take up time.To construct NAICS data for 3312 prior to 1990, the March 2002 ratios of SIC employment that were in the dual-coded cell for NAICS 3312 were used.

The ratios were generated by RATIOSxx programs, posted on the Projections Workgroup we site.

The two primary SICs comprising NAICS 3312 were 331 and 339. SICs with ratios of less than 0.001 should be ignored, as they add little and take up time.

As an example, in 1982, NAICS employment = (0.4176 X 58,000) + (0.0645 X 1,900) = 24,343.

The 1990 NAICS employment estimate is from the LDB file.

Note the discontinuity between 1989 and 1990.

Data for this exhibit provided by Sandy Newman, Ohio LMI.

- industrial production index
- Capacity utilization rate
- Interest rates, commercial and residential construction
- Exchange rates, trade-weighted value of the dollar
- Foreign trade: exports, imports, trade balance
- Price deflators: GDP and CPI

Time-series data for many economic variables at the state and substate levels are available from federal government sources through the internet (www.bls.gov; www.bea.doc.gov; www.census.gov; www.fed.stats.gov; www.icesa.org) and hardcopy publications. Other state and substate data sources will include state government offices of demography, state planning, and revenue/budget. University-based economic and business research centers, research units of major financial institutions, and proprietory research and consulting firms often can provide data on additional relevant variables.

Generally, there are fewer economic variables available in annual time-series for substate areas, and particularly non-metropolitan areas, than for states. Also when constructing time-series for substate areas, changes in geographic area definitions of MSAs based upon the decennial Census must be carefully noted in order to maintain data consistency.

If any of the non-employment variables are measured on an industry basis, such as output, or earnings derived from a particular sector, then the SIC to NAICS conversion will pose a problem for obtaining a consistent historical time-series. Here the availability of conversion ratios estimated from dual-coded data may not exist. If so, this will limit or prohibit the use of such variables.

3.4 Importing Time-Series Data to the Long-Term Projection System

Time-series employment data, as well as other data that can serve as independent variables in regression models, can be imported into the LTIP from several alternative formats. Once imported, the data can be adjusted or edited within the LTIP system to allow the analyst maximum flexibility in specifying a range of alternative models.

Employment data can be imported from spreadsheets or from text files. Data for constructing other variables can be imported from spreadsheets.

Depending upon whether the source data is ES 202, QCEW, or CES (790), and whether industry employment projections will be developed for NAICS or for OES industry definitions, the analyst can choose to (1) adjust the imported data into OES definitions; (2) "roll up" all government employment (such as employment in public hospitals and public schools) into the government NAICS; or (3) make no adjustments.

Likewise, the analyst is given the option of "rolling up" industries by (i) creating 3-digit NAICS industries from 4-digit industries; (ii) creating major industry divisions from 3-digit NAICS codes; or (iii) not rolling up the imported industry employment data.

3.5 Editing and Modifying Data in the LTP System

The LTP system under the main menu item EDIT, allows the analyst to add, delete, or modify observations or data values, as well as create new variables from data already entered in the system. The analyst can also edit the pre-loaded geographical (substate area) definitions and aggregate across industries to form new industry categories.

By choosing MODIFY DATA, and then MODIFY EMPLOYMENT DATA, the analyst can select a particular ownership class: federal government, state government, local government, private, or mixed, and in essence, treat each ownership class from a single NAICS industry as a separate industry with its own employment time-series.

One can add variables, delete variables, or edit the particular values of existing data (for instance, to correct an error, or to change the exogenously provided projected value of an independent variable) by choosing MODIFY VARIABLE DATA from the menu.

Ratio variables, such as state/national population, are particularly useful in specifying regression models, or to analyze whether the state has gained or lost its national share of some economic activity over time. These new variables can be created under the RATIO VARIABLES menu item, where the analyst can choose ratios of state to national, substate to national, or substate to state for any original economic variable already entered in the system.

Finally, any variable entered into the system can be transformed into the same variable lagged (choice of 1 or 2 years), the same variable led 1 year, or into its logarithmic, exponential, or inverse quadratic forms, using the VARIABLE TRANSFORMATION menu item. This option gives the analyst much added flexibility in specifying regression models when a nonlinear form of a variable is more theoretically valid or helps the model fit the data better.

3.6. Screening Time-Series for Errors and Outliers

The employment time-series should be screened and adjusted, if necessary, for any data recording errors or outliers prior to using for pre-projection analyses. Perhaps the easiest way to do this is a visual examination of a graphical plot of each historical time-series using the main menu item PREPROJECTION ANALYSIS, and then TIME SERIES GRAPH in the LTP system. The visual examination should detect most serious errors by noting discontinuities or obvious outliers. Discontinuities in a series can occur for many reasons (see Exhibit 3.3). These might include an uncorrected change in area definition, a change in data sources or employment concepts, use of different benchmarked employment data, or possible miscalculations. Perhaps the most prominent reason for a discontinuity is the change in industry classification from SIC to NAICS. A state that does not convert time-series data measured in SIC to NAICS will end up with significant discontinuities in its historical employment time-series. It will most likely appear as an additive discontinuity as shown in Exhibit 3.3. Yet even when states apply the dual coded ratios to pre-1990 annual employment data, there is a good chance an additive discontinuity will occur between 1989 and 1990. Often the analyst will not be able to correct this problem since it is caused by a ratio estimated with 2002 data that is not valid for earlier years. The antidote is to use a dummy variable in a regression model (see chapter 5) to take into account the discontinuity caused by the industry classification change.

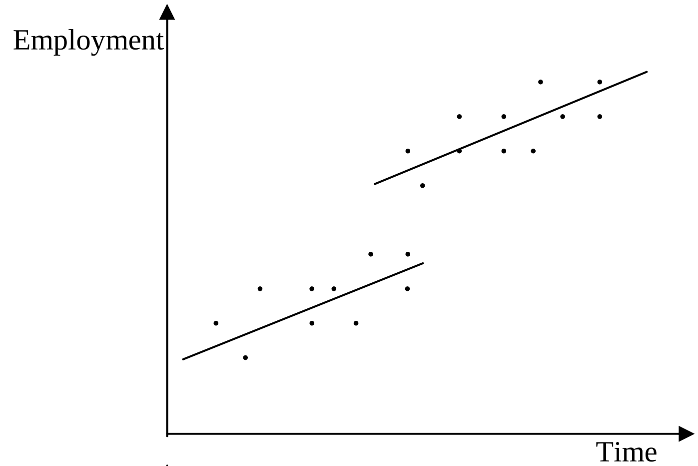
The analyst should investigate the reason(s) for the discontinuity in the time-series and attempt to correct it at the source of the error. If the source of the error can not be diagnosed easily then the analyst can choose to

EXHIBIT 3.3

Examples of Discontinuities and Outliers in Annual Employment Time-Series Data

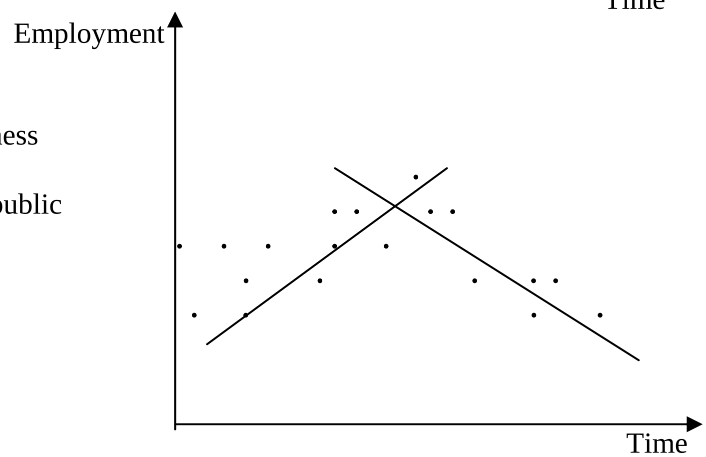
1. Discontinuity, Additive

Plausible Diagnoses:
*Change from SIC to NAICS
*Change in substate area boundaries
*New plant opening



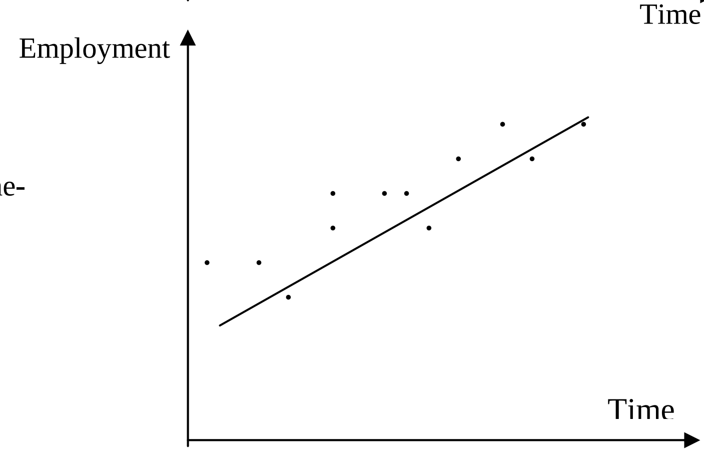
2. Discontinuity, Non-Additive

Plausible Diagnoses:
*Decline of state or area competitiveness
*State/area industry disinvestments
*Technological change or change in public policies affecting the industry



3. Outlier, Singular

Plausible Diagnoses:
*Error in reporting on ES-202 form
*Error in transcribing data to machine-readable files
*Strike, layoff



mechanically adjust the series. Mechanical adjustment is simplest when there is an additive discontinuity. The analyst should adjust the earlier portion of the time series. If the source of error can not be detected, if the discontinuity is not a simple additive one, and the industry is not a key, or critical industry, then it may not be worth the effort to correct and use the time-series. But for these cases regression analysis should *not* be used as the projection technique.

Although *discontinuities* in the time-series plots most often reflect errors or changes in the industry or geographic definitions, *outliers* may represent real events. If investigation fails to turn up any of the types of errors mentioned above, then the analyst should consider outliers as real. The explanations for these might include strikes, large plant closings or openings, a one-time large contract to one establishment (e.g., a defense contractor), or a highly volatile industry sector (e.g., some of the mining industries). In general, the analyst should not adjust outliers at this stage of the projections process if they represent real events. The analyst might consider later the use of a dummy variable or a data smoothing procedure at the stage of model specification if a regression analysis approach is used.

Chapter 4

Pre-Projection Analysis

4.1 Introduction

The exercise of analytical insight and judgment can make the greatest difference in the accuracy of the final set of industry employment projections. Experience has shown that investing some time in the analysis of the state (substate) economy and key industries prior to developing projections will significantly improve the usefulness of the projections as well as make the entire projections process more time-efficient. Below we describe several relatively simple analytical techniques that are included in the LTP system. The information gained from the application of these techniques can help the analyst choose the most appropriate projection models.

4.2 Classifying Industry Sectors: the Location Quotient

The location quotient (LQ) is useful for identifying many of the state's (or substate's) key industry sectors. The LQ is a measure of the relative concentration of an industry sector (in terms of employment) in a state or substate area compared to a reference region. The reference region is normally the nation. Exhibit 4.1 shows how the LQ is calculated for each industry sector. The minimum value of an LQ is 0, when there is no employment at all in a given industry sector in the region. There is no upward limit on the magnitude of an LQ. The "normal" value of an LQ is 1.0. An LQ = 1.0 means that an industry represents the same proportion of employment in the region as it does in the reference region. Thus any industry sector with an LQ > 1.0 means that it is disproportionately concentrated in the region; those industry sectors with an LQ < 1.0 are under represented. As a first cut, all industries with an LQ >1.0, and at the same time, meet a threshold level of employment, might be considered "key" industries. Other criteria can be applied to the designation of key industries: rapidly growing, targets of regional economic development policy, or those undergoing significant technological change.

Beyond the designation of key industry sectors, the analyst also should attempt to classify industry sectors by their product market orientation. More specifically, is a state (substate) industry primarily export-oriented, or local-serving? Since this need be only a coarse designation, location quotients are suitable for identifying clear export-oriented industries. Many industries that are local-serving are self-evident (e.g., grocery stores) but some industry sectors vary considerably in their primarily export ***vis a vis*** local-serving orientation across regions. For instance, banking may be export-oriented in a few large centers; eating and drinking establishments maybe export-oriented in tourist areas. Here the analyst should utilize informal sources of information for correct industry designation.

Among export-oriented industries, it is useful to identify whether the products are shipped *directly* or *indirectly* (i.e., as an intermediate input to another locally-produced good or service which in turn is shipped outside the study area). In either case, classifying the industry by the principal geographic scope of the eventual export market also is useful: regional, national, international. Among local-serving industries, the analyst should attempt to identify whether the market is primarily other local businesses or local households. The analyst will need outside information to answer these questions. The complete topology for classifying industry sectors by their product market orientation is shown in Exhibit 4.2.

Within the LTP system the analyst can select the year for calculating the LQ, the reference area, and the LQ threshold value for designating export or local-serving industries. The year should be the most recent for which there are employment data. There is some value, however, in calculating LQ's for some earlier year and then examining the changes in the relative concentration of key industries over the historical

EXHIBIT 4.1

Location Quotients

LQ_i = $E_i / E_T \;/\; US_i / US_T$

where E_i = employment in industry i for the study region.

E_T = total (all industry) employment for the study region.

US_i = employment in industry i for the nation.

US_T = total (all industry) employment for the nation.

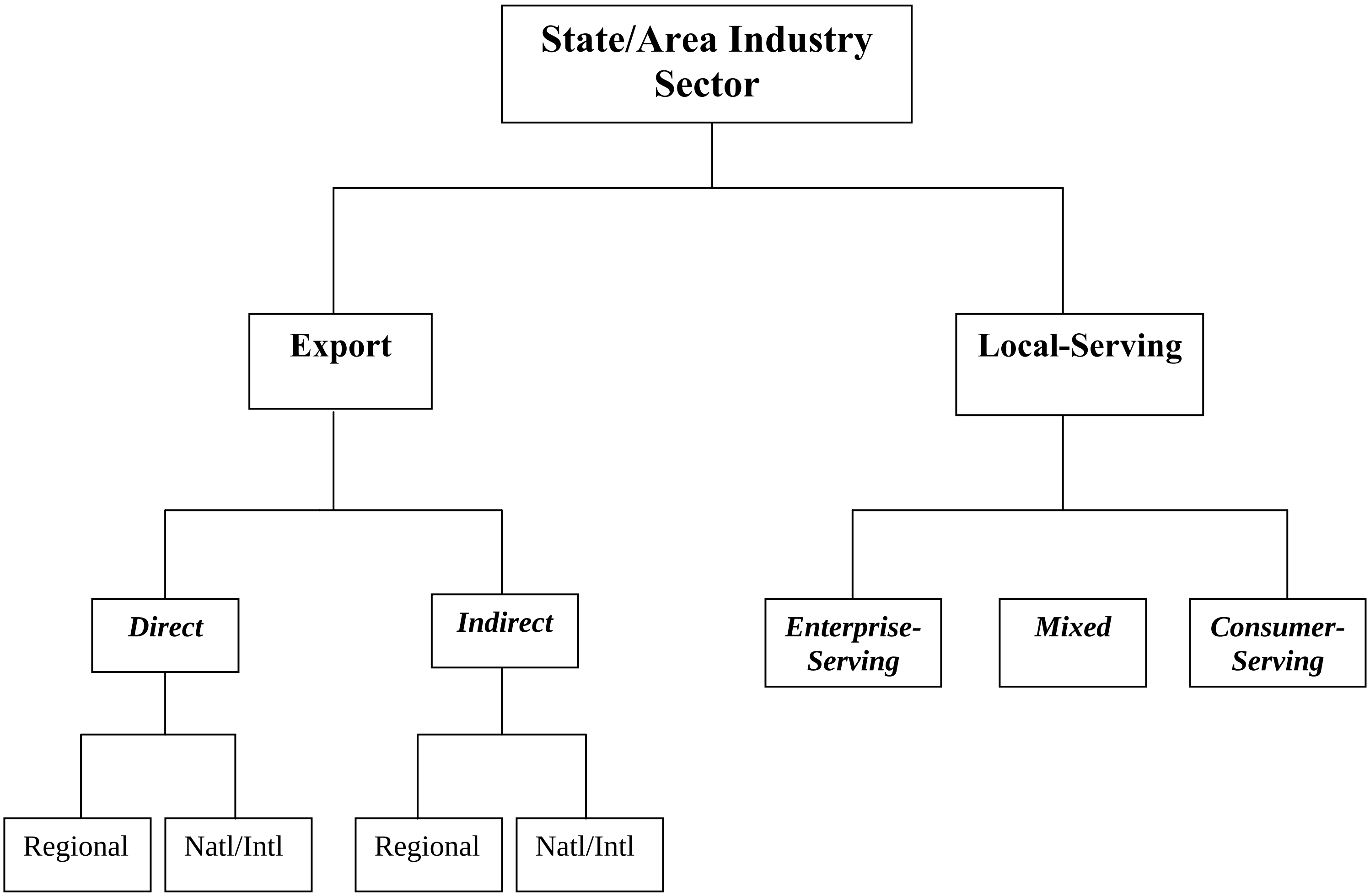
If $LQ_i >> 1$ the region is more specialized than the nation in industry i. Likely to be primarily an export industry.

If $LQ_i << 1$, the region is less specialized than the nation in industry i. Likely to be primarily local-serving.

Note: Employment data should all be for the same year (e.g. the base-year).

EXHIBIT 4.2

Typology of Industry Sectors Based on Product Market Orientation



period. The reference area should be the nation when calculating the LQs for the state industries. For substate analysis one can choose the state as the reference area to identify key industries, though the nation should still be used to identify export industries. The default threshold value of the LQs for identifying export sectors has been set at 1.2 in the LTP. The reasoning for setting it at 1.2 rather than 1.0 is to distinguish those sectors that are clearly over- represented in the region from those that are above, but still close to, the "norm" of 1.0. The analyst can select an alternative threshold value to be able to make "export sector" designation either more or less stringent. After making these choices, the LTP system will list those sectors in the state or substate area that are classified as "export" and those that are classified as local serving, based upon the magnitude of their respective LQ.

4.3 Analyzing Employment Trends: Time Series Graphs

An examination of the historical trend of employment in a given state or substate industry can reveal a number of important items about its behavior that might be expected to continue into the future. Trend analysis reveals the average rate of growth or decline; the extent to which the change in level of employment has been smooth, or linear; the extent to which it has followed economic cycles; and the extent to which the rate of employment has paralleled the nation's for the same industry. It also can reveal outliers and data input errors by visually examining the time-series plot.

Two types of time-series graphs can be produced in the LTP system under the "Time-Series Graph" menu item for pre-projection analysis. After selecting the industry and historical period, the analyst can select a graph of employment vs. time, or area/national employment ratio vs. time. The latter conveniently allows one to see if changes in the rate of employment in the state (substate) industry have mirrored the rate of change in employment in the national industry. The closer the two have moved together, the closer the time-series plot will be a straight horizontal line (see Exhibit 4.3 for examples). If the graph reveals strong divergence between the state (substate) industry employment and that of the nation, the analyst should explore, with outside information, the reasons for it. Some typical reasons for sharp divergences might include: differences in industry composition at a more detailed industry level between state (substate) and the nation; differences in functional orientation of the industry compared to the nation, e.g., concentration in R&D activity); productivity or production cost differences; or location in an above average or below average growth region of the nation.

In addition to showing the graph, the LTP system calculates the mean, the standard deviation, and the minimum and maximum values of employment during the designated historical time period. The best-fitting straight line through the time-series observations of employment is also drawn, whose slope indicates the average annual rate of industry employment growth for the period.

4.4 Employment Stability

The relative stability of an industry's employment level or rate of growth over the recent historical time period is often a reliable indicator of how accurately one can expect to project industry employment using different projection techniques. Knowledge of the relative stability, or volatility, also helps the analyst decide what factors will need to be taken into account in projecting industry employment.

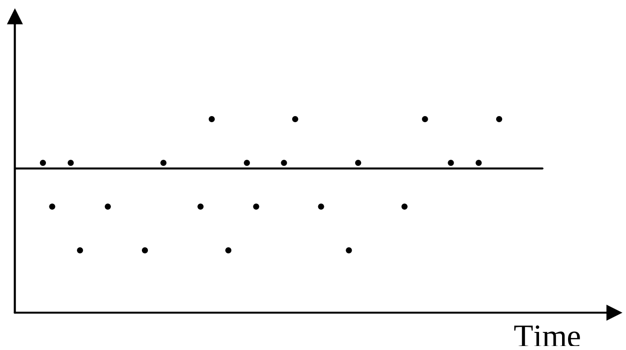
Employment stability can be measured in a number of alternative ways. A relatively simple, but useful, way is to fit the least squares regression line through the scatter plot of time-series observations. The extent to which the actual employment deviates from the linear trend is an indication of that industry's employment volatility. The coefficient of determination, or R-squared, of the least squares regression line, measures the overall extent of the deviation of the actual employment levels from the historical trend. The closer the R-squared is to 1.0, the greater the stability. In the LTP system, the analyst can examine the stability of industry

EXHIBIT 4.3

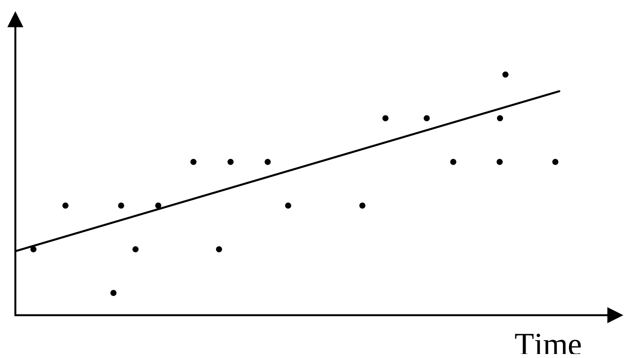
Hypothetical Time-Series Plots of the Ratio of State to National Industry Employment

* Constant state share of
national industry employment;

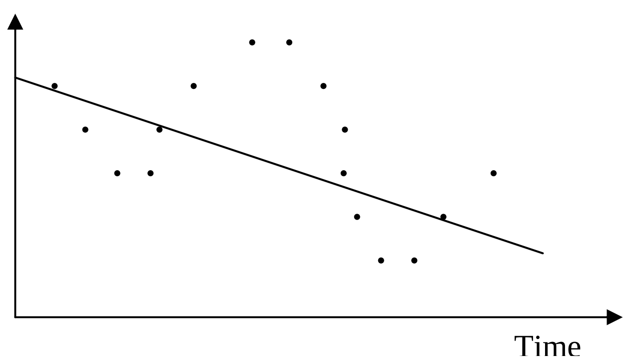
* Stable relationship



* Increasing state share of
national industry employment;
* Stable relationship

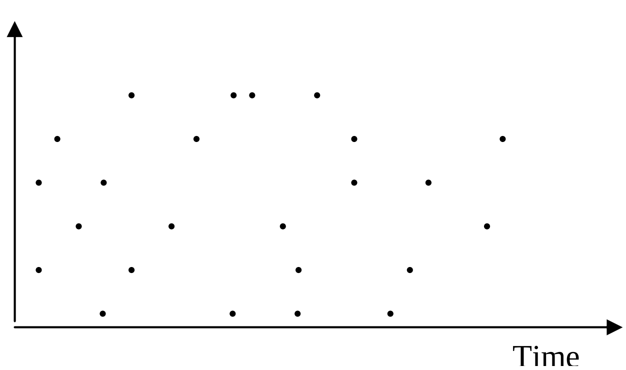


* Decreasing state share of
national industry employment;
* Strong state cyclical fluctuations



* No relationship between changes in
state and national industry employment

employment stability will be measured.



employment over the historical period, and also the stability of the ratio of state (or substate) to national industry employment over the historical period. The analyst chooses the beginning and end years of the historical period over which

4.5 Comparative Growth Rates

The historical growth rate of state (substate) industry employment, both in absolute terms, and relative to the historic and projected growth rate for the same industry sector for the nation, can be useful for selecting among alternative projection models. And after developing a tentative state (substate) projection, the comparison with the projected national industry employment growth rate can provide a valuable external validity measure.

After the industry, reference area, and time period have been chosen, the LTP system calculates the average annual (linear) industry employment growth rate for the state or substate area over the historical period, for the reference area over the same period, and the projected average annual employment growth rate for the reference area. The comparison of these growth rates will inform the analyst about the extent to which the state or substate industry employment "kept up" with the national over the same period. If the growth rates were comparable, then using projected national industry employment as a predictor variable in a regression model or in a shift-share model will likely lead to an accurate projection for the state or substate industry. If the historical growth rates diverge, then this becomes a signal to the analyst that other explanatory factors driving employment change will need to be identified for the particular industry sector.

The comparison of the projected national and historical industry employment growth rate with the growth rate corresponding to the preliminary state or substate industry employment projection provides the analyst with a reference, or benchmark for evaluating the state (substate) projection. For example, if the national historical growth rate for a given industry sector had been 2.0 percent annual, the state historical average annual growth rate equal to 1.0 percent, and a projected national industry employment growth rate from BLS equal to 1.5 percent annual, then a model that resulted in a state projection of 3.0 percent annual should be viewed skeptically by the analyst. One would want to be able to explain the plausibility of why the state industry would grow at twice the rate as the national industry when historically the state lagged the nation.

4.6 Industry Mix: the Index of Dissimilarity

Industry mix refers to the distribution of employment within a given industry sector among more detailed NAICS for the state or substate area, compared to a reference area. For example, in one state, NAICS 336 (transportation equipment manufacturing) may be concentrated much more in NAICS 3363 (motor vehicle parts), compared to the nation, and have a smaller proportion of employment in railroad equipment and ship building compared to the nation (see Exhibit 4.4). Information about the similarity of industry composition between the state (or substate area) and the nation can be useful for helping the analyst decide whether to develop projections at the 3- or 4-digit NAICS level of detail. Thus, if the state's distribution of employment among 4-digit NAICS industries differs significantly from the nation, then one it would *not* be wise to develop projections at the 3-digit level.

Analysis of industry mix also can inform the analyst about the extent to which national industry employment and other national industry indicators likely will be valid predictors of industry employment trends at the state or substate levels. If the industry composition markedly differs between the state or substate area and the nation for a given industry sector, then national industry variables may not prove to be good predictors in a projection model. One would either develop projections at a more detailed level of industry detail, or rely upon non-national indicators as predictor variables.

Differences in industry composition can be measured in several ways. One of the simplest, and incorporated in the LTP system, is the Index of Dissimilarity. This index is computed by taking two employment distributions – one for the state (substate area) and one for the nation – within a given NAICS, computing the proportion of employment in each detailed industry, and subtracting the difference in proportion for each detailed NAICS industry. These differences are summed and then divided by 2.0. The result is the value of the index. The minimum value of the index is 0.0, when the employment distributions for the state and nation are exactly the same. The maximum value is 1.0. Values greater than about 0.3 indicate a relatively high level of dissimilarity. The calculation of the Index of Dissimilarity is shown in Exhibit 4.4.

4.7 Summary

In this chapter we have reviewed several relatively simple analytic techniques -- the location quotient, employment time-series plots, trend lines to measure temporal employment, and industry composition analysis -- whose results are useful for identifying the factors that are ‘driving” growth or decline state or substate in industry employment. These same techniques also are helpful for selecting among a range of alternative projection models and model specifications to be described in the next chapter.

Exhibit 4.4

Index of Dissimilarity

NAICS	U.S. Industry (Thousands)	Empl Proportion US_i / US_T	Ohio Empl (Thousands)	Proportion ST_i / ST_T
3361	258.1	**0.1472**	31.2	**0.2031**
3362	152.8	**0.0871**	8.9	**0.0579**
3363	699.7	**0.3990**	96.2	**0.6263**
3364	438.1	**0.2498**	14.2	**0.0924**
3365	22.9	**0.0130**	0.5	**0.0033**
3366	143.8	**0.0820**	0.7	**0.0046**
3369	38.3	**0.0218**	1.9	**0.0124**
336 (Total)	1,753.7	**1.0000**	153.6	**1.0000**

Index = Sum {ABS [$(ST_i / ST_T) - (US_i / US_T)$] } / 2

= 0.2831

Note: ABS is absolute value

Chapter 5

Projection Models

5.1 Introduction

The analyst is faced with a large number of alternative projection models from which to choose for any given industry sector. This is both good news and bad. The good news is that with a large number of choices, the chances are higher that the analyst will find at least one that will produce accurate projections. The bad news is that the task of finding a 'good" model may seem rather complex and overly time-consuming for some analysts. The major purpose of this chapter is to provide both theoretical and practical guidance to the analyst for choosing among the wide array of models available, including a discussion of the strengths and weaknesses of different types of models for certain classes of industry sectors. The results of recent research to test alternative measures of state comparative advantage as predictor variables in regression models are discussed. And the use of dummy variables within the context of multiple regression models has been added to help the analyst deal with some of the modeling implications of the shift from SIC to NAICS employment data.

5.2 The Major Approaches to Developing Projections

There are at least five principal approaches that have been used for projecting state or substate industry employment: (1) employer surveys; (2) judgmental (expert) techniques; (3) extrapolation or allocation models (e.g., simple linear trend models, shift-share models); (4) single-equation regression models; and (5) fully-specified econometric models. Each of these approaches has some advantages and disadvantages compared to the others against such criteria as data requirements, cost, accuracy, versatility, etc. as shown in Exhibit 5.1. Yet not all criteria should be given the same weight. Because of likely *low accuracy* and *reliability*, employer surveys should not be chosen as an approach for developing long-term projections. Likewise, judgmental techniques should not be considered as a primary approach, although they should be relied upon heavily as a supplementary approach in the projection review and adjustment process. Assumption, a type of judgmental technique, alone may be appropriate for certain types of industries which are relatively small, non-key, and have exhibited historically stable behavior.

The analyst, in practice, is left with three approaches for developing projections for the majority of industry sectors: (1) extrapolation and allocation models, (2) single equation regression models, and (3) fully-specified econometric models. In general there is a role for each of these approaches within the context of the LMI projections program.

5.2.1 **The Appropriate Roles for Econometric Models**

The decision to use an econometric model should depend upon whether the expected improvements in accuracy, reliability, and credibility of the projections are sufficient to justify the greater expenses of building and maintaining such models. Available evidence, although limited, suggests that employment projections from econometric models are not significantly more accurate than those developed from simple single-equation regression models (see, for instance, Fishkind and Roberts, 1978; Texas Employment Commission, 1984; Michigan Employment Security Commission 1984). There also is evidence that the level of projection accuracy of econometric models decreases significantly as the level of industry detail increases from total employment to the major industry division level and then to 2-digit and 3-digit (in the

EXHIBIT 5.1

The Advantages and Disadvantages of Major Approaches for Projecting State and Substate Industry Employment

Approach	Data Requirements	Comprehensibility	*Industry detail*	Versatility *Length of Projection Period*	*Consideration of Alternative Futures*	Accuracy	Theoretical Validity
Employer Survey	High	High	Unconstrained	Valuable for short-range; Not suitable for long-range	Low	Low	Low
Judgmental (Expert)	Low	High	Constrained by knowledge of industry behavior & trends at detailed levels	Unconstrained. Some advantage for long-range	Moderate	Low-Moderate	Low
Extrapolation models	Low-moderate	Moderate-high	Not suitable for high levels disaggregation because of loss of accuracy	Suitable for medium or long-range projections only during stable periods	Low	Low-Moderate	Low
Single-equation regression models	Moderate	Moderate	Constrained only by availability of time-series data and projected values of independent variables	Not suitable for periods less than 2 years. Risky beyond 8-10 year period	Low-moderate	Low-high	Moderate
Fully-specified econometric models	High	Low	Not generally suitable to high levels of industry disaggrega-tion because of date requirements & model complexity	Generally one quarter to 2-3 years. No advantages beyond 3 years	High	Moderate-high	Moderate-high

old SIC) which were needed for LMI purposes (Bjornstad and Tepel, 1986).

The amount of extra costs of using an econometric model depend mostly on whether a suitable state The amount of extra costs of using an econometric model depend mostly on whether a suitable state model already exists in another public (or private) agency, and whether those forecasts would be available to the LMI unit. If such forecasts are available, the LMI staff should assess the performance, or track record, of the model's forecasts, and consider whether such attributes of the model as the length of the forecast period, the employment concept used, and level of industry detail of the employment forecasts are appropriate for LMI purposes. The performance of the model should be gauged by how accurate the model's forecasts have been over periods of different economic conditions, over various lengths of forecast periods, and for different industry sectors. One should bear in mind that most econometric models are designed for short-term forecasting purposes (from one quarter to two years), and that forecasting accuracy tends to decrease steadily with increasing length of forecast periods. If the model has produced reasonably accurate forecasts for projection periods comparable to those of the OES-based projections, and if there is comparability of employment categories between the model and the OES industry sector definitions, then the LMI staff might consider the use of a state econometric model in two alternative ways:

Option 1--Econometric Model Forecasts as Validity Checks

Forecasts of employment in industry sectors produced by the econometric model can be compared to the industry employment projections developed by LMI using either single-equation regression models or extrapolation/allocation techniques. In general, this can be done for the most aggregate industry sectors (e.g., total employment, major industry divisions, some 3-digit NAICS sectors) because most state econometric models are highly aggregated. The LMI analyst should "flag" the cases where there are relatively large discrepancies between the two projections, and particularly when the projected trends seem to be going in opposite directions. In these cases the analyst may want to compare the underlying assumptions of the technique or model used by LMI with those of the econometric model. One then has several choices: (1) stay with the LMI model projection; (2) re-specify the LMI unit’s projection model to yield a new projection; (3) adjust the projection from the LMI’s model based on judgment; (4) adopt the projection from the econometric model. The choice among these options should not be an arbitrary one, but based upon a careful review. The amount of effort devoted to the review in each case should depend upon whether the industry sector is a key one in the state or substate area.

Option 2--Econometric Model Forecasts as Control Totals

A more important active role for econometric models is the use of their more aggregate industry employment forecasts (i.e., total employment and major industry division) as control totals for the LMI unit's projections of employment at the 3- and 4-digit NAICS level. The latter would be developed by either single-equation regression models or extrapolation/allocation techniques. This option implicitly acknowledges that at the most aggregate levels, fully-specified econometric models are probably more accurate than the less sophisticated approaches. The LMI staff also may have the option of adjusting the state econometric model's employment projections that would serve as control totals after reviewing the underlying assumptions in the model.

5.2.2 **Single-Equation Regression Models vs. Extrapolation/Allocation Models**

There are at least two principal issues involved in choosing among these two types of models, accuracy and ease. Single-equation regression models, in general, should be more accurate than extrapolation or allocation techniques because they can make use of more information. This means a greater number of temporal observations, plus more flexibility in choice of factors, or predictor variables. On the other hand, the process of calibrating single-equation regression models is more time-consuming and requires more data than do shift-and-share and simple time-series models.

Given the tradeoff between accuracy and ease, it is recommended that single-equation regression models should be used for *at least* all the key industry sectors. Beyond that there are other rules of thumb based on our experience of developing and evaluating alternative industry employment projection techniques:

* Avoid the use of shift-and-share and related techniques which rely on only two historical observations for highly volatile and cyclically-sensitive industries.

* In cases where one 4-digit NAICS industry quantitatively dominates a given 3-digit industry, the analyst might use extrapolation techniques for the other 3-digit industries rather than developing independent projections.

* In cases where there is relatively little historical variation in level of industry employment, regression models may not perform well. A simple extrapolation or allocation technique, or even judgment, should be used when the industry sector is a non-key one.

* Multiple regression models produce unreliable and unstable estimates with fewer than 20 temporal annual observations in the historical time-series.

* For substate areas there are fewer potential predictor variables available in annual time series for use in regression models. Thus, shift-share and simple time-series models (extrapolation and allocation models) may be the only viable option for some substate areas, particularly if they are small or non-metropolitan

There is a great amount of flexibility in specifying alternative single-equation regression models, i.e., choosing among different combinations of independent variables and their mathematical form. Customizing the regression model for each individual industry sector is discussed later in this chapter. Variations among shift-share and simple time-series models in terms of their underlying assumptions and the conditions under which their projections would have higher validity are summarized in the following sections. More details can be obtained from several references listed at the end of this guidebook.

5.3. Shift-Share Models

The LTIP system includes seven different shift-share models: (i) constant share; (ii) constant regional rate; (iii) constant share of aggregate industry employment; (iv) implicit shift-share; (v) modified implicit shift-share; (vi) classical shift-share; and (vii) fixed employment to population ratio model.

What these models have in common in terms of strengths is that they require very little data, base their projections on either one or two temporal (historical) observations, are easy to understand, and are easy to calculate. A spreadsheet program is all that is needed if one is not using the LTIP software. Their common weaknesses are that the assumptions about industry employment change are very simplistic, the projections are produced mechanically, rather than analytically, and the projections are highly sensitive to the choice of the one or two historical years used to calibrate trends or shares. It is important that the analyst consider whether the assumptions underlying a given model “fit” the historical behavior of employment change in a given industry sector, and whether these assumptions are likely to continue to hold into the future, at least until the projection year (see Exhibit 5.2).

EXHIBIT 5.2

A Synopsis of Shift Share and Simple Time Series Models

1. Constant Share $E_i^{t+n} = US_i^{t+n} * E_i^t / US_i^t$

2. Constant Regional Rate $E_i^{t+n} = E_i^t + s * E_i^t [(E_i^t - E_i^{t-m}) / E_i^{t-m}]$

3. Constant Share of Aggregate Industry Employment $E_i^{t+n} = E_j^{t+n} * E_i^t / E_j^t$

4. Implicit Shift Share $E_i^{t+n} = E_i^t + E_i^t * \{(US_i^{t+n} / US_i^t) - 1\} + s E_i^t * \{E_i^t / US_i^t - E_i^{t-m} / US_i^{t-m} \}$

5. Modified Implicit Shift Share $E_i^{t+n} = E_i^t + E_i^t * \{(US_i^{t+n} / US_i^t) - 1\} + s E_i^t * \{E_i^{t+n} / US_i^{t+n} - E_i^t / US_i^t\}$

6. Classical Shift Share $E_i^{t+n} = E_i^t + E_i^t \{(US_T^{t+n} / US_T^t) - 1\} + E_i^t \{ US_i^{t+n} / US_i^t - US_T^{t+n} / US_T^t\} + s * E_i^t \{ E_i^t / E_i^{t-m} - US_i^t /US_i^{t-m} \}$

7. Linear Time Series $E_{i,t} = a + b$ TIME)

8. Log-Linear Time Series $E_{i,t} = a + b$ (log TIME)

9. Exponential Time Series $\log E_{i,t} = a + b$ TIME

where E is regional employment in industry i at time t,
US is national employment in industry i at time t,
t is the base year,
t + n is the projection year,
t - m is an historical year,
s is a scaling factor = (n / m), the ratio of length of the projection period to the length of the calibration period,
a, b are estimated regression coefficients
TIME is expressed as year: 1970, 1971, . . . 2000, . . .

5.3.1 **Constant Share Model**

This model projects industry employment on the basis of (i) an exogenous projection of national industry employment (from BLS in the LTIP system), and (ii) the share, or ratio, of state (substate) industry employment to national industry employment as calibrated for a chosen historical observation. The accuracy of the projection thus will depend upon the accuracy of the exogenous national industry employment projection and the historical share remaining the same between the calibration year and the projection year. If the time-series plot (discussed in chapter 4 above) for a given industry reveals that the share, or ratio, of employment has been unstable or has been systematically increasing or decreasing over time, then the analyst would be wise not to rely upon this particular model.

5.3.2 **Constant Regional Rate Model**

The Constant Regional Rate Model assumes regional industry employment will grow at the same rate during the projection period as it did in the historical period. The analyst selects the beginning year and the end year to define the historical period, and the model, in essence, extends the straight line connecting these two points out to the projection year. Clearly, the projection value will be sensitive to the choice of beginning and end years. If industry employment has been either highly cyclical or unstable, then the projection values could differ substantially depending upon the particular years chosen. In general, one should select the beginning year and the end year so that they are at comparable points of the business cycle. The model should only be used when the historical trend has been linear and stable, and when there is no reason to believe that the regional industry behavior in the projection period will be different from that in the historical period.

5.3.3 **Constant Share of Aggregate Industry Employment**

This model projects regional industry employment as a constant ratio of employment in another regional industry. It requires an exogenous projection of industry employment in this other sector. The model is often used to develop 4-digit NAICS industry projections after 3-digit NAICS industry employment projections have been developed. The analyst should first check to make sure that the ratio, or share, of the particular 4-digit industry employment to the 3-digit NAICS industry employment has been relatively constant over the historical period. This check is easily made using the time-series graph from the Pre-projection Analysis main menu option.

5.3.4 **Implicit Shift Share**

The Implicit Shift Share Model estimates two employment growth components and adds these together with base year employment to yield projected year employment. The first component is referred to as the national share. It calculates the amount of industry employment growth that would occur (hypothetically) if the state or substate area's industry employment grew at the same rate during the projection period as the national industry employment is projected to grow. Thus an exogenous national industry employment projection is required for this model.

The second component is the regional shift, or competitive, factor. It calculates the change, or shift, in the region's share of national industry employment between the base year (t) and some prior historical year (t-n), and assumes that the gain or loss of the region's share will continue at the same rate during the projection period. As with several of the previously described models, the projection will be quite sensitive to the analyst's selection of the base and the historical years. The two years should define a period that represents the long-term growth trend of the regional industry, and should be at comparable points on the business cycle. The shift component is especially sensitive to the choice of years.

5.3.5 **Modified Implicit Shift Share**

The "modification" to the implicit shift share model is that the regional shift component is calculated based on a projected change in regional share of industry employment, rather than the historical change in share that is assumed to continue at the same rate in the projection period. This model then requires an exogenous projection of regional share of national industry employment. In most cases the analyst will develop this by doing a trend analysis of the historical share, using time-series regression (to be discussed in section 5.4). The modified version of the model has the advantage of allowing the analyst to take into account that the regional share may not be changing linearly over time.

5.3.6 **Classical Shift Share**

This model is similar to the implicit shift share, except that it splits the national share component of the latter into two separable components: national total employment growth share, and national industry mix. An exogenous national industry employment projection is required. As with the implicit shift share model, the analyst selects the base year and a prior historical year. The change in regional share of national industry employment during this historical period is assumed by the model to continue at the same rate and in the same direction from the base year to the projection year. Thus the projection will be sensitive to the choice of the beginning and end years of the historical period.

5.3.7 **Fixed Employment-to-Population Ratio Model**

This model uses an exogenous population projection for the state or substate area. The model assumes that the ratio of industry employment to regional population will remain fixed between the base year and the projection year. Simple multiplication of this ratio by the exogenous population projection will yield the projected industry employment. Before using this model, the analyst is advised to examine the employment to population ratio over a few selected years in the historical period to verify that the ration is indeed relatively constant. If the ratio has had an increasing or decreasing trend, or if the ratio has been highly unstable, then the projection will have low validity. This model may be most appropriate for consumer services or retail sectors in which the technology of service provision, and consumer tastes are not expected to change significantly over the projection period.

.

5.3.8 **A Test Case of the Accuracy of Shift Share Models**

We will illustrate the relative accuracy of the particular types of extrapolation/allocation models using a test data set. The case is SIC 34 in Massachusetts. The base year is 1974 and the projection year is 1985. This case uses old data, but the lessons are still pertinent. In this test, because we know the *actual* 1985 employment level we can evaluate the accuracy of each model. Exhibit 5.3 lists the input data used and Exhibit 5.4 shows and evaluates the projection results. In the 1985 *unconditional* projections, the 1985 target-year national industry employment projections are those which were developed by BLS in 1978. In the *conditional* projections, the actual 1985 national industry employment is used as if they were known with certainty to the analysts when making the projections. Having the set of unconditional and conditional projections allows us to separate the effect of errors in exogenous projections of predictor variables from errors due to invalid assumptions in the models.

The results in this example show: (1) there tends to be a wide range of projections from which the analyst would have to choose the "best"; (2) the more complex shift-and-share models, which include a competitive, or regional shift component, were less optimistic than the simpler models, and had smaller projection errors; (3) certain knowledge of national industry employment did not necessarily lead to better projections--the shift-and-share models did worse with the actual values of national industry employment

EXHIBIT 5.3

Test Data for Shift-Share Models
Massachusetts and U.S. (in thousands)

Variable	t-1 = 1963	t = 1974	t+1 = 1985
State industry employment (SIC 34)	E_{ij}^{t-1} = 38.6	E_{ij}^{t} = 45.4	---
State total employment	E_{j}^{t-1} = 1,946.9	E_{j}^{t} = 2,374.9	---
State population	P_{j}^{t-1} = 5,344.0	P_{j}^{t} = 5,799.0	P_{j}^{t+1} = 6,208.*
National population	P_{k}^{t-1} = 189,240.0	P_{k}^{t} = 211,389	---
National industry employment (SIC 34)	E_{ik}^{t-1} = 1,150.1	E_{ik}^{t} = 1,505.3.0	E_{ik}^{t+1} = 1,900.**
National total employment	E_{k}^{t-1} = 56,707.0	E_{k}^{t} = 78,413.0	E_{k}^{t+1} = 98,810.**

Sources: State employment estimates are from Employment and Earnings, States and Areas 1939-1975 (BLS Bulletin 1310-12).

*Population estimates provided by the U.S. Bureau of the Census
**1985 National employment projections are the OES wage and salary employment projections

EXHIBIT 5.4

Projection Results of Shift-Share Models with Test Data

Model	1974 Employment	1985 Actual Employment	Unconditional Projection: 1985 Projected Employment	Unconditional Projection: Percent Projection Error	Conditional Projection: 1985 Projection Employment	Conditional Projection: Percent Projection Error
1. Constant Share of National Employment Model	45,400	47,500	57,304	20.6	44,405	-6.5
2. Constant Regional Rate Model	45,400	47,500	53,398	12.4	53,398*	12.4*
3. Employment/Population Fixed Ratio	45,400	47,500	48,602	2.3	48,602*	2.3
4. Implicit Shift and Share Model	45,400	47,500	50,840	7.0	39,396	-17.1
5. Modified Implicit Shift and Share Model	45,400	47,500	44,403	-6.5	34,408	-27.6
6. Classical Shift and Share Model	45,400	47,500	51,281	8.0	38,380	-19.2

Sources: U.S. Bureau of Labor Statistics, "OES Matrix/Projections Memorandum No.9," 1978; Massachusetts Division of Employment Security, Economic Analysis and Research; U.S. Bureau of Labor Statistics, Employment and Earnings, States and Areas 1939-1975;

than with the BLS projected value; (4) the model which performed the best, the employment/population fixed ratio model, would not ordinarily be chosen by the analyst for this export industry sector based on its implicit assumptions; and (5) that the unconditional projections using shift-and-share models were better than the conditional projections was fortuitous. The former badly *under-estimated* the future state share of national industry employment, but this was compensated in the opposite direction by the large exogenous BLS *over-estimated* national industry employment. In summary, the accuracy of this class of models is not predictable. Their performance is unstable because of their simple assumptions about the forces driving state industry employment and their reliance upon very limited historical data. Very good analyst judgment is required to decide if any of these models' assumptions will fit the particular case.

5.4 Simple Time-Series Models

An alternative approach to extrapolating historical trends into the future is to use simple time-series regression models. Here, state or substate industry employment is regressed against the variable time (TIME), or, alternatively, a *nonlinear function* of time. The assumed relationship between industry employment and time can be linear or nonlinear. In the LTP model, there is a linear, logarithmic, exponential, and polynomial function of time available.

In the linear time-series model, the best fitting straight line to the annual historical industry employment data is estimated (by ordinary least squares, OLS), and this straight line is simply ‘extended” to the projection year to estimate projected industry employment.

Often, however, a nonlinear function of time fits the annual historical time series better than the linear function. In the logarithmic time-series model, industry employment is regressed against the natural logarithm of TIME; the LTP model estimates the regression coefficients and then uses these coefficients to estimate industry employment for the projection year. In the exponential function, after mathematical manipulation, the log of industry employment becomes the dependent variable, and this is regressed against TIME. The estimated coefficients are then used to estimate employment for the projected year.

To use any of the described simple time-series models in the LTP model, the analyst selects the historical period (beginning year and end year) over which the model is to be estimated. The software will then find the best fitting straight line or curve) to the historical data, and indicate the “goodness of fit” by the R-squared, of each time series model estimated. Thus, the analyst could estimate the linear, logarithmic, exponential, and polynomial versions, and select the “best” of these on the basis of the respective R-squared.

5.5 Single Equation Regression Models

Single-equation regression models offer the advantages of a high degree of flexibility in their specification. This allows the analyst to take into account knowledge of a potentially wide variety of state or substate industry characteristics gained in the pre-projection analysis discussed in chapter 4. They require more data and staff effort than the extrapolation/allocation techniques but less than that which would be required by an econometric model. They also are more "accessible" and understandable for both LMI analysts and projection users than econometric models. Thus, the projections may be seen as more credible. Single-equation regression models are less equipped than econometric models to produce alternate sets of employment projections based on alternative future scenarios. Finally, the accuracy of single-equation regression models can be enhanced significantly, if a systematic projection *review* and *adjustment* process is conducted in conjunction with the models’ results.

5.5.1 **The Mathematical Form**

In single-equation regression models the dependent variable in each equation is the estimated or predicted level of employment in a given industry. Single-equation regression models have the property that each equation can be solved independently of all other equations.[1] Mathematically,

(Eq. 1) $E_{it} = a + b_1X_{1t} + b_2X_{2t} + ... + b_nX_{nt} + e_t$

where:

E_{it} is state or substate area employment in industry i at time t;

X_{1t} ... X_{nt} are values of independent or predictor variables at time t;

e_t is the error term;

a,b_1, ..., b_n are the regression model parameters.

Ordinary least squares (OLS) are used to calibrate (estimate) the model in Eq. 1. using annual historical time-series data. Once calibrated and tested, the OLS estimate of Eq. 1 is used to predict the level of state (substate) industry employment at time T+h where T is the base-year and h is the length of the projection period in years:

(Eq. 2) $E_{i,T+h} = a + b_1X_{1,T+h} + b_2X_{2,T+h} + ... + b_nX_{n,T+h}$

The values of the independent variables at time T+h must be supplied exogenously. These values may be known with certainty (e.g. for the variable TIME), may be obtained from BLS, or may be projected independently by the analyst.

The choice of independent variables is left to the analyst's knowledge of the factors which should help explain or predict changes in state or substate industry employment, and knowledge of the statistical assumptions underlying the linear regression model. It is in the flexibility of specifying the regression model that gives this approach an important advantage over others. This advantage can only be exploited, however, if the analyst uses sound judgment in model specification rather than adopting a purely mechanical approach.

The simplest regression model specifications are those in which the only independent variables are either: (1) time, (2) national employment, and (3) time and national employment. In picking the first model the analyst is assuming that a constant rate of employment change over the historical period will continue through the projection period. This is the same model as the simple time-series (linear) model described in section 5.3 above. In picking the second model the analyst is assuming that state (substate) industry employment as a share of national (or state) industry employment has been constant and this constant share will continue through the projection period. In the third model the analyst is assuming that some portion of the change in state (substate) industry employment is due to a change in national (or state) industry employment and that the other part is a linear function of time. The principal advantage of these relatively simple regression models is that the analyst can either determine the projection year value of time with certainty, and/or the projection of national industry employment is provided by BLS from their econometric model. The principal disadvantage is that they do not take into account other factors besides national industry employment and time which may

[1]This does not preclude recursively-related equations. Here, employment in industry i is projected in a first equation, and industry i's projected employment is used as a *predictor* variable in a second equation in which employment in industry j is projected.

be affecting the level of industry employment in a given state or substate area. If there are important *local* factors and these are not taken into account in the model, then the model's projections are likely to have large errors.

To exploit the analyst's knowledge of all the relevant factors that are affecting state or substate industry employment, a framework for specifying *customized* regression models is described below. The analyst then is in a position to use customized regression model specifications for many key industries when the data are available. Pre-defined (pre-specified) regression models may be appropriate for the non-key industries when development of customized models may be precluded because of limited staff resources.

5.5.2 **Regression Model Options in the LTP System**

The Long Term Industry Projection Model software provides the analyst with three options for developing single-equation regression models: pre-defined model specifications, customized model specifications, and a batch industry option.

The first option is to use a set of "pre-defined" model specifications already programmed into the model for individual industry sectors. These pre-defined specifications vary by type of industry, and specifically by whether the industry is primarily export-oriented or primarily local-serving. The market orientation of the industry can be determined automatically by the magnitude of the location quotient. The default value of the LQ threshold is 1.2, but the analyst can change this value. Alternatively the analyst can override the classification determined by the LQ for any given industry. For either an export or local serving industry, a set of alternative models automatically will be estimated, with the associated projection estimate and model diagnostics (to be described below). The analyst will still need to choose from among the set of model specifications/projections for each industry sector with the aid of the diagnostics. The principal advantage of using the pre-defined model option is the time savings of not having to design a set of candidate models for each industry sector from scratch. The principal disadvantage is that in some cases none of the pre-defined specifications may yield good calibration results and an accurate projection, because of particular behavioral attributes of the given industry sector in that state or area . Within the set of pre-defined models, the analyst has several additional options. These include re-specification with an automatic correction for serial correction (using the Hildreth-Lu transformation), and performing an ex-post projection test measured by the Theil-U statistic to gauge the performance of the model outside the calibration period. Both of these options are described in more detail below.

The second option is to develop customized model specifications for each individual industry sector. Here the analyst can select any combination of independent, or predictive, variables from the full set of data/variables that have been entered in the software. This option gives the analyst maximal flexibility for discovering the "best" model for each individual industry sector, taking into account the unique factors driving employment change in each sector for that state or area. The principal disadvantage of this option is that it can become quite time-consuming "thinking through" and iteratively testing and fine-tuning a series of models for each industry sector, when there are several hundred sectors for the state and every substate area. On the other hand, it is an excellent option to use with the most important, fastest growing, and largest industry sectors. As with the pre-defined model option, the analyst has the choice of automatically correcting for serial correlation and to calculate the Theil-U statistic.

The third option is to use the batch process for developing pre-defined models for a group of (or all) industry sectors, rather than one industry at a time. The analyst can designate groups of industries by NAICS detail (e.g., all 4-digit) or by range (e.g., NAICS 311-339). The advantage of this option is obviously the time savings. The analyst can push one or two buttons, go to lunch, and return to get the results for the full set of industry sectors designated. The principal disadvantage is that the analyst does not take into account the specific behavior, and different factors driving employment change in each sector.

5.5.3 **A Framework for Specifying Regression Models for Export Oriented Industries**

The recommended framework combines elements of export-base theory and the principle of comparative advantage from regional economic theory. The framework relies heavily on the availability of a reliable set of exogenous national industry employment projections. The latter allow the formulation of models which retain essential theoretical validity within the constraint of single-equation regression models. It also helps to avoid the collection and maintenance of a data base beyond the realistic capacity of state agencies faced with quite modest resources for research and analysis.

Employment demand in area export industries theoretically is assumed to be a function of output (product) demand, as well as the relative price of local labor.

(Eq. 3) $E_i = f\ (X_i,\ p_l/p_k)$

where:

E_i = employment demand by industry i
X_i = demand for industry i's product from all regions
p_l/p_k = the ratio of the unit price of labor to the unit price of capital.

National industry employment, whose projected values are available from BLS, serves as a proxy variable for projected demand for industry output at the national level. These national projections already take into account estimated changes in employment demand due to factor substitution between capital and labor because of changes in their relative prices. *With this factor included in the model, one only needs to account for why state or substate area industry employment may not change at the same rate as projected national industry employment.*

There are three general types of factors which may explain for a given industry the differential between state (or area) and national industry employment growth rates: (l) the state (area) has comparative advantages or disadvantages; (2) local cyclical effects -- the state (or area) economy cyclical behavior does not conform to the national economy in terms of severity, or amplitude; and (3) a given state (area) industry is *unusually* affected by the level of output in technologically-linked industries within the respective State(area). These factors and some suggested indicators of each are discussed below.

<u>Export Product Demand</u>. As stated, national industry employment serves as a proxy measure for level of product demand. Historical time-series data and projections for this variable are available from the U.S. Bureau of Labor Statistics. Occasionally national industry employment will not turn out to be a statistically significant predictor variable. The most likely reason for this is that the mix of state or area industries at more disaggregated industry levels is significantly different from the mix at the national level and there is large variation in employment growth trends among the industries at the more disaggregated level of detail. When projecting employment at the substate level, state or multi-state industry employment may be a better proxy than national industry employment for product demand in some cases.

<u>State (Area) Comparative Advantage.</u> Businesses in *a given industry* in a specific state or area may face advantages or disadvantages on either the demand side (their market) or the supply side (their production costs) compared to businesses in other states (or areas). If the state (area) is located in a region that is growing well-above the national average, then it is likely that the market for the state (area) industry's products is growing faster than the national average. Hence employment will expand at a faster rate than the national average for that given industry. On the supply side, the state (area) may have production costs (such as labor,

energy, transportation) that are well above or well below other producing regions. A state (area) may also have productivity advantages or disadvantages due to above or below average labor skills, physical infrastructure, or age of capital. Proxy measures might include the state's (area's) share of national population, national employment, or national personal income on the demand side. On the supply side the measures might be: ratios of state to national average wage rates and energy prices, and ratio of state and national average labor productivity.

<u>New Economy Variables.</u> Research was conducted recently by this author to investigate whether including variables that measure the competitiveness of a given state specifically within the “new economy” would increase the explanatory power of regression models.[2] Nine industry sectors in three states (Massachusetts, Minnesota, and Ohio) were selected as the test cases. Historical time-series data were assembled for about 25 “new economy” variables for the nation and for the three states. These variables included, for instance, annual industrial R&D investments, annual university R&D spending, bachelor degrees awarded in science and engineering, doctoral degrees awarded, patents granted, and so on. Regression models were calibrated using one or more of the new economy variables as predictor variables, and compared to the results for same model specification but without the new economy variable(s). Evaluation of the calibration results showed that including a new economy variable significantly increased the explanatory power of the regression model (adjusted R^2) and reduced serial correlation in a large percentage of the test cases. Such improvements were not limited to the models for just high tech industries but extended to traditional industries as well.

As a result of the tests, it was recommended that five “new economy” variables be added to the LTIP data set as alternative measures of state comparative advantage that analysts can consider when selecting among regression model specifications. These variables all at the state-level, are: (1) total industrial R&D investments as a percent of the U.S, contemporaneous and lagged (in constant $); (2) university R&D expenditures in science, engineering and health as a percent of the U.S., contemporaneous and lagged (in constant $); (3) Number of bachelor degrees in science, engineering and health awarded per capita, contemporaneous and lagged; (4) Percent of population with bachelor degrees; and (50 average wage/salary per job (in constant $).

There are a larger number of possible indicators of state comparative advantage focusing on the new economy that may be appropriate explanatory variables in regression models. A number are listed on the website for the report. In practice the analyst will be constrained by data availability in having to construct consistent annual historical time-series.

<u>State or Area Cycle Effect</u>. Each industry at the *national* level has its own characteristic business and growth cycles. These depend upon the industry's position in the various stages of processing (e.g. extraction, crude materials processing, intermediate goods processing, finished capital goods or business services, final consumer goods or consumer services), the stage in its product life-cycle, and the nature of foreign competition and the relative value of the dollar, and trade restrictions, to name a few. Yet, it is known that there is substantial variation in the cyclical performance of a given industry *between* states and areas. Here cyclical performance includes relative amplitudes of cyclical fluctuations, timing of turning points, and return-to-peak duration. The relative age mix (a proxy for productivity) of plant and equipment, the mix of functions of establishments (e.g. production, corporate headquarters, sales offices, or transportation and warehousing), and also peculiar inter-industry linkages in the state or area economy will affect how state or area industry employment will respond to national cyclical fluctuations. The state or area cycle effect measures the amplitude (and timing) of state or area industry employment cycles *relative* to the amplitude of U.S. economic cycles. Alternative *indirect* measures of national cyclical fluctuations include the U.S. annual

[2] See H. Goldstein, *New Economy Indicators for Long-Term Industry Employment Projections*, May 2005, on the Projections Workgroup website, for more details.

unemployment rate, and the annual rate of change in Gross National Product.[3] The estimated regression coefficient is interpreted as the sensitivity of state or area industry employment to U.S. economic cyclical fluctuations. The two alternative measures of U.S. economic cyclical fluctuations are mentioned to take into account that employment in a given state or area industry may be not equally sensitive to each measure.

Local Inter-Industry Demand. Employment in a state (area) industry that exports its products *indirectly* through other technologically-related state (area) industry generally is affected by the growth rate of these state or area purchasers. Since this agglomeration effect may be peculiar to the particular state (area), it can partially account for the differential between state (area) and national industry employment growth rates. If the local purchasing sectors of the products of a given industry i are expanding, everything else equal, we should expect higher than "expected" employment growth in industry i. On the other hand, if the state or area purchasing sectors are contracting, state or area industry employment would grow at a lower than "expected" rate. The proxy for state or area inter-industry demand is the level of state or area employment in the identified purchasing sectors.[4] Historical time-series data are available from the state LMI division or from the U.S. Bureau of Labor Statistics, *Employment, Hours, and Earnings*. Projections of employment in state or area purchasing sectors must be developed by the analyst before projecting employment in industry i, using a "recursive" strategy.

The types of predictor variables for export industries are summarized in Exhibits 5.5 and 5.6.

5.5.4 **Specifying Regression Models for Local-Serving Industries**

Employment in local-serving industries is primarily a function of output demand based on *state* or *area* conditions rather than national economic conditions. National industry employment projections from the BLS Economic Growth Model, however, take into account technological changes (i.e., factor input substitution), productivity changes and changes in consumer preferences at the national average. There also may be a state (area) cyclical factor which influences local-serving industry employment through a "roll-out" effect from the cyclical fluctuations of the state's (area's) export industries and from general macroeconomic conditions. Thus, the specification of projection models for local-serving industries draws on measures of: a) local demand, b) changes in technology, productivity or consumer tastes, and c) state or area business and growth cycles.

Local Product Demand. Local-serving industries may serve state or area businesses or the residential population, or both. Alternative indicators of local product demand include state (area) population, number of households, personal income, total state (area) employment, state (area) employment in a particular sector, or the number of state (area) enterprises. Which of these measures is best depends upon the particular product market orientation of the industry. Annual historical time-series for all these variables except number of households are available for all states and most MSAs from the BLS, the U.S. Bureau of Economic Analysis (BEA), or the Census Bureau. Projections of state and area population and personal income are available from BEA.

Changes in Technology, Productivity, or Consumer Tastes. The national industry employment projection takes into account the estimated impact of changes in technology, productivity, or consumer tastes in the industry at the national level. To obtain a standardized measure, we divide U.S. industry employment by either U.S. total employment, U.S. population, or U.S. households. Thus the time-series of the ratio of U.S.

---

[3] These measures may be lagged. Historical time-series and annual projections are developed and available on the Projections Workgroup website.

[4] The identification of purchasing sectors can be from first-hand knowledge of local inter-industry relationships or from "synthetic" state or substate input-output tables. These are now available from the U.S. Bureau of Economic Analysis, Regional Analysis Division, for a modest cost.

EXHIBIT 5.5

Examples of Predictor Variables For State Export Industry Models

Determinants of State <u>Industry Employment</u>	**<u>Predictor variables</u>**
1. *National Product Demand*	a. National Industry Employment
2. *State Comparative Advantage or Disadvantage*	a. State Share of National Population
	b. State Share of National Personal Income
	c. State Share of National Total Employment
	d. State Share of National Manufacturing Employment
	e. State/National Average Wage
	f. State/National Energy Price
	g. Time Trends
3. *State Cycle Effect*	a. National Unemployment Rate
	b. Percent Annual Change in GDP
4. *Inter-Industry Effect*	a. Employment in Related State Industry

EXHIBIT 5.6

Examples of Predictor Variables
For Substate Area Export Industry Models

Determinant of Area Industry Employment	**Predictor Variables**
1. *National Product Demand* ***OR*** *State Product Demand*	a. National Industry Employment or State Industry Employment
2. *Area Comparative Advantage or Disadvantage*	a. Area Share of National(State) Population
	b. Area Share of National(State) Personal Income
	c. Area Share of National(State) Total Employment
	d. Area Share of National(State) Manufacturing Employment
	e. Area/National Average Wage
	f. Area/National Energy Price
	g. Time Trends
3. Area Cycle Effect	a. National (State) Unemployment Rate
	b. Percent Annual Change in GDP
4. *Inter-Industry Effect*	a. Employment in Related Area Industry

industry employment to U.S. total employment takes into account the national average pattern of combined impacts of changes in the technology of purchasing sectors and productivity within the given industry. The time-series of the ratio of national industry employment to U.S. population or U.S. households takes into account the national average pattern of combined effects of changes in consumer preferences for the product of a given industry, and productivity within the given industry. Finally, in some cases, national industry employment by itself may be superior (in a statistical sense) to its ratio form, when the effects of other included predictor variables are considered.

State or Area Cycle Effect. The existence of state or area business cycles which differ in amplitude and/or timing from the national business cycle may affect employment in local-serving industries as well as export industries. The same measures are relevant here.

The classes of variables for local-serving industries are summarized in Exhibit 5.7 and 5.8.

5.5.5 **Procedures for Specifying Projection Models**

There are at least five valuable sources of information available to the analyst for choosing the correct model specification for a given industry sector. These are: (1) the results of the pre-projection analyses of the state (area) economy as a whole and the industry classification exercise; (2) time plots of state (area) industry employment, (state) area share of national (or state) industry employment, and industry share of state (area) total employment; (3) results from shift-and-share analyses of the components of area industry employment change using historical employment data; (4) informal and subjective knowledge obtained from state (area) economic and industry experts; and (5) results of model calibration tests. The first three information sources are most useful as guides for developing the initial model specification while the last two are most useful for systematically improving the initial model specifications.

5.5.5.1 Initial Model Specification

The industry classification results help the analyst identify for each industry the appropriate indicator of product demand, and to identify each industry as principally export-oriented (direct or indirect) or local-serving.[5] Predictor variables are selected from the various types illustrated in Exhibits 5.5 -- 5.8.

For export-oriented industries the time plot of state (area) share of national industry employment can inform the analyst about any evident trend in state (area) competitive advantage for the particular industry. This would be indicated graphically by a non-zero slope of the trend line. The plot also can help to identify the existence of an area or state cycle effect operating on that particular industry. A unique state (area) cycle effect is graphically evident by an oscillating pattern around the hypothetical linear trend line of the state (area)/national industry employment ratio (see Exhibit 4.3).

This time plot of state/national industry employment also can help the analyst judge how good national industry employment is as a proxy measure of product demand. If there is no evident trend line in the plot one may hypothesize that the state (area) industry might be serving a regional market which is more or less independent of the national market. The time plot of the industry share of total state (area) employment for local-serving industries may help the analyst indicate the extent to which the size of the state (area) economy, measured in total employment, can serve as a proxy for state (area) demand. If the plot is more or less a horizontal line, then it might be a very good indicator. Otherwise, the analyst may consider an alternative

---

[5] In some cases the analysis results suggest that the industry simultaneously has both orientations. The analyst may choose to combine variables from each specification framework or to develop separate models for constituent industries at a higher level of industry detail.

EXHIBIT 5.7

Examples of Predictor Variables For State Local-Serving Industry Models

Determinant of State Industry Employment	**Predictor Variables**
1. State Product Demand	a. State population
	b. State personal income (constant dollars)
	c. State total employment
	d. State manufacturing or export employment
	e. State share of national population
	f. State share of national personal income
	g. State share of national employment
2. Industry Productivity, Technology, or Consumer Preferences	a. National industry employment
	b. National industry share of total employment
	c. National industry employment per capita
3. State Cycle Effect	a. National or state unemployment rate
	b. Percent annual change in GDP

EXHIBIT 5.8

Examples of Predictor Variables
For Substate Area Local-Serving Industry Models

Determinant of State Industry Employment	Predictor Variables
1. *Area Product Demand*	a. Area Population
	b. Area Personal Income (constant dollars)
	c. Area Total Employment
	d. Area Manufacturing or Export Employment
	e. Area Share of National (State) Population
	f. Area Share of National (State) Personal Income
	g. Area Share of National (State) Employment
2. *Industry Productivity, Technology, or Consumer Preferences*	a. National(State) Industry Employment
	b. National(State) Industry Share of Total Employment
	c. National (State) Industry Employment Per Capita
3. *Area Cycle Effect*	a. National(State)Unemployment Rate
	b. Percent Annual Change in GDP

measure of state (area) demand or presume that some significant change in industry productivity, technology, consumer tastes or import-substitution may be occurring.

A shift-and-share analysis of state (area) industry employment change (particularly for export-oriented industries) can help the analyst identify which components of state (area) industry employment change have been most important in the recent historical period. A large regional shift component, for instance, would suggest that one or more variables which measure regional economic growth should be selected for inclusion in the initial model specification.

In general, not all of the factors will be judged relevant for representation in the initial model specification. Usually, only a particular variable from each relevant factor would be selected. The analyst should strive to include no more than three independent variables in any regression equation, to minimize the loss of precision in estimating regression parameters, and hence to minimize the loss of precision of the projected value of the dependent variable.

5.5.5.2 Improving the Initial Model Specification

The analyst should expect to be able to improve the initial specification of the projection model after examining the calibration results. Here the adjusted R^2 and the Durbin-Watson statistic (D-W) are the key indicators of possible misspecification. The presence of serial correlation indicates that the residual has a systematic component and a random component. This often implies there is an omitted, *relevant* independent (or predictor) variable(s) which, if included, could remove the systematic component from the error term. This will usually result in a better fitting model (higher adjusted R^2). The analyst's objective, then, is to make step-wise improvements in the model specification as indicated by a higher adjusted R^2 and a lower degree of serial correlation, subject to conceptual or theoretical validity of model specification candidates. Decisions about what changes in the model specification will result in the greatest statistical improvement at the next step are based upon a detailed examination of the calibration results at the previous step. The results of standard model specification (i.e., U.S. industry employment and time as the two predictor variables) can be used as a standard for judging the relative improvement in the calibration results for alternative model specifications. If the initial model specification produces poor calibration results, then information from "outside" industry experts may be useful for diagnosing the poor fit and for identifying any possible omitted relevant variable(s). Sometimes the analyst may decide that no regression model candidate is acceptable and may opt instead to use a simpler technique.

5.5.5.3 Including Dummy Variables

Including dummy variables in regression models is often useful for the model to be able to take into account one-time, or discrete, changes in the data. With the change from SIC-based employment and other economic data to NAICS-based data, the analyst must deal with “interrupted” historical time-series” caused by new industry definitions introduced in 2000. And even though BLS has converted national and state employment data from SIC to NAICS back to 1990, and many states went back even earlier, there are still interruptions in the time-series because the conversion factors are not error-free. Accordingly, it is recommended that all states use dummy variables when specifying regression models to project industry employment.

A dummy variable is one in which there are only two possible values for a given observation (year), 0 and 1. In this context we would designate a value of 0 for years in which the data were provided on an SIC-basis, e.g., pre-1990, and 1 for years in which the data were provided on a NAICS-basis (1990 and forward (and including the projection year). In some cases states may want to designate 0 for years prior to 2000 and 1 for 2000 and forward if there is good reason to believe that the conversion factors used by BLS are not valid for the particular state.

The dummy variable is added as an additional independent variable on the right-hand side. If it turns out that there was an additive "interruption" in the data coinciding with the coding for the dummy variable, then the estimation of the model will indicate that the dummy variable is statistically significant. We interpret the magnitude of the estimated coefficient for the dummy variable as the amount of difference in employment due to the change in industry definition, for instance. So, for example, if the estimated coefficient, b, is 200.0 (and statistically significant), then the effect of the change from SIC to NAICS resulted in, on average, an increase of 200 jobs to that particular industry sector. If the estimated coefficient were negative, say -200.0, then it would mean that the new NAICS-based industry became smaller by 200 jobs. See Exhibit 5.9 for a graphical representation of a model that includes a statistically significant dummy variable.

5.5.5.4 Transformations

Sometimes where a step-wise model specification improvement procedure fails to achieve acceptable calibration results, a transformation of the dependent variable or one or more of the predictor variables can significantly improve calibration results.

One common type of transformation is from a linear to a nonlinear relationship. Nonlinear transformations of one dependent variable or one or more predictor variables, however, should be conceptually based. The idea that there may be a nonlinear relationship between the dependent and the predictor variable(s), frequently comes from prior calibration results and earlier examination of time-plots. Logarithmic transformations, in practice, may work well since many nonlinear relationships can be approximated by logarithmic functions. Moreover, many statistical software packages have logarithmic transformation functions that can be easily used on specific variables.

A second type of transformation is used to correct for serial correlation if the latter cannot be eliminated by inclusion of the "correct" set of predictor variables. In this case the dependent variable, each independent variable, and the error term are transformed at follows:

Eq. 4 $$Y_t^* = a + b_1 X_{1t}^* + ... + b_n X_{nt}^* + v_t$$

where:

$$Y_t^* = Y_t - pY_{t-1}$$

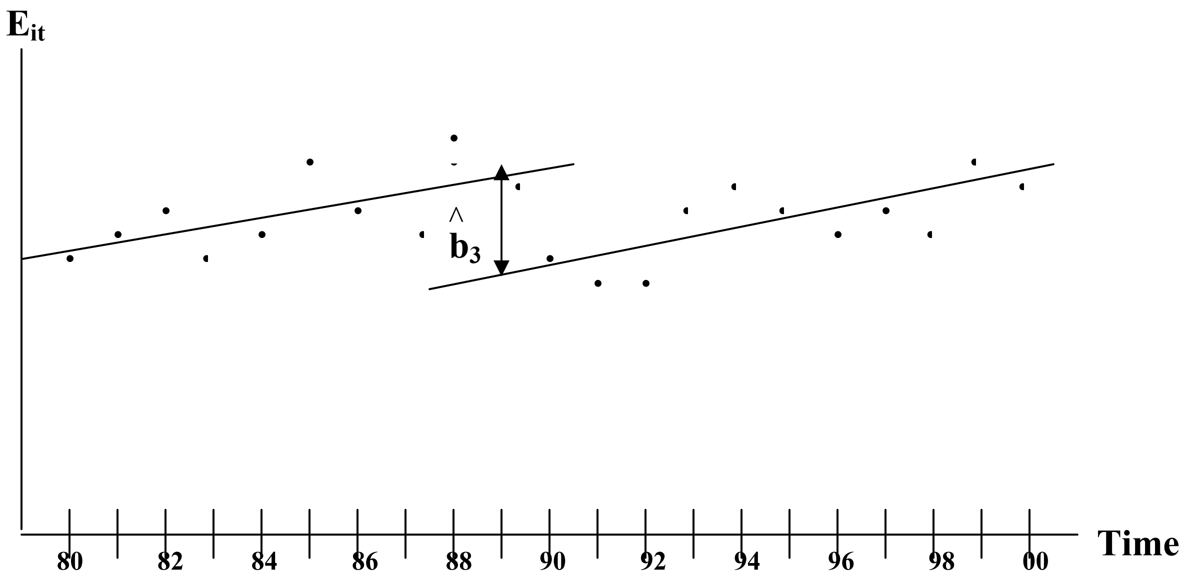
$$X_{1t}^* = X_{1t} - pX_{1t-1}$$

$$X_{nt}^* = X_{nt} - pX_{n\,t-1}$$

$$v_t = e_t - pe_{t-1}$$

In general the R-squared of the transformed model will be lower than in the original model, but the transformed model will be a better performing model. The LTP system includes this type of transformation as an option.

EXHIBIT 5.9
The Use of Dummy Variables in Regression Models



$$\hat{Y}_t = \hat{a} + \hat{b}_1 X_{1t} + \hat{b}_2 X_{2t} + \hat{b}_3 X_{3t}$$

where: Y_t is annual average employment in industry i
X_{1t} is national industry employment,
X_{2t} is state population
X_{3t} is a dummy variable = 0 for years before 1990
= 1 for 1990 and after

b_3 is the estimated magnitude of difference in industry employment due to conversion from SIC to NAICS.

5.5.6 **Special Considerations for Specifying Substate Area Models**

The logic for specifying regression models described above is applicable to substate areas as well as for states. There are, however, several special considerations when developing substate area models. First, some economic variables available at the state level may not be available for substate areas (e.g., annual population estimates). In some cases some economic variables may be available for state's major metropolitan areas only. The possible unavailability of some economic variables at the substate area level may preclude the analyst from
some of the pre-defined models in the LTP system, and will constrain specification options using the customized model option.. Second, the analyst will have to choose between the national economy or the state economy "driving" the substate area economy in specifying models. As a general "rule of thumb," national economic variables should be chosen when the substate area is a major metropolitan area, and/or when the particular industry sector exports primarily outside the state. The analyst should consider examining the time-series graphs of the ratio of substate area industry employment to state industry employment, and then the ratio of substate to national industry employment, to determine whether the area industry employment more closely mirrors state or national industry employment.

Third, the state's projections staff may not be sufficiently large to develop customized models for every sector in every substate area, and so "short cuts" may need to be taken. Several strategies exist, but good analyst judgment and a thorough review process will be necessary no matter what, to yield an accurate set of projections. One strategy is to just use the predefined specifications in the LTP system that can be estimated with the substate area data available, A second strategy is to use a constant share of state industry employment model for smaller industries and for those that have maintained a stable share of the state industry employment over the historical period. Finally, one can estimate all of the shift share and simple time-series models and choose the "best" from among the alternatives. Some criteria for judging the "best" model are discussed later in this chapter.

5.6 Evaluating the Projection Performance of Regression Models

In the previous section, a framework for developing customized regression models for particular state and substate industries was presented. This framework gives the analyst maximum flexibility and discretion in model building. But with such breadth of choice the analyst needs better evaluative tools for helping to decide which one of many alternative regression model specifications for a given state or substate industry will likely yield the most accurate projection. This section describes some available evaluative measures and tools. We can divide these into four types: (l) calibration tests, (2) theoretical validity criteria, (3) ex-post projection tests, and (4) external validity checks of actual projections.

5.6.1 **Calibration Results**

The calibration results provide information about the overall "goodness-of-fit" of the model during the historical period. They also can indicate possible violations of the assumptions of the classical linear regression (CLR) model. These violations can affect the desired unbiasedness and efficiency properties of the estimators, which may result in biased projection values and/or large uncertainty about the estimate of the true projection.

The most important evaluative, or diagnostic, measures from the calibration results are:

* Adjusted R^2

* The Durbin-Watson (DW) Statistic

* The signs and significance of the estimated regression coefficients

* The standard error of the forecast (SEF)

* The time plot of the residuals

The *adjusted* R^2 tells the analyst how well the model has "tracked" the dependent variable over the historical period. A model with a high adjusted R^2, however, does not mean necessarily that the model will continue to predict the dependent variable accurately during the projection period. For instance, structural relationships among the variables may change, or the high R^2 may have occurred because of spurious correlation. In addition, large errors in the projected values of the independent variables can cause large projection errors.

The *Durbin-Watson statistic* tells the analyst whether there is serial correlation (first-order) in the model. The presence of serial correlation is a violation of one of the principal assumptions of the CLR model. Yet the presence of serial correlation in single-equation, time-series regression models indicates likely model *misspecification.* More specifically, it is likely to mean that a relevant variable has been omitted from the model, causing the error term to have a systematic, rather than a random, pattern to it. Failing to correct for omission of a relevant variable can lead to biased estimators (regression coefficients) and thus biased projections of the dependent variable. It also can lead to inefficient estimators, which means that uncertainty about the projected value of the dependent variable grows larger.

The *statistical significance* of estimated regression coefficients can help the analyst determine whether there are irrelevant variables included in the model. Irrelevant variables can decrease the efficiency of the estimators and increase the uncertainty of the projected value. They can also, of course, tell the analyst which variables are associated most strongly with the dependent variable, and thus should be retained in the model. The *signs* of the coefficients provide the analyst a test of the theoretical validity of the model. These will be discussed below.

The *standard error of the forecast* (SEF) is used as an estimate of how much uncertainty in the projected value of the dependent variable the analyst should expect. The SEF does not take into account the effect of any error in the projection of the independent variables. The SEF/projection ratio often is used to standardize the SEF for scale differences in the magnitudes of the dependent variable across industry sectors.

The *time plot of the residuals* often is helpful in helping the analyst diagnose whether the model may be losing its ability to track the dependent variable in the most recent part of the calibration period due to changes in structural relationships. One also can identify what annual observations during the calibration period the model produced the most and least accurate predictions. For instance, the model may not be able to predict employment accurately in peak and trough years of business cycles for cyclically sensitive industries which do not conform to the national industry's cyclical behavior.

As a general guide, the analyst should aim to meet the following criteria for calibration results:

* Adjusted $R^2 \geq 0.90$;

* DW value that indicates no serial correlation *or* is in the indeterminate range;

* Ratio of SEF/Projection < 0.1;

* t-ratios of estimated regression coefficients $\geq$ 1.96;

* Signs of all coefficients are in the "correct" direction.

5.6.2 **Theoretical Validity**

Theoretical validity refers to whether the independent variables are causally and linearly related to the dependent variable, and whether there are no relevant variables omitted. In practice it means including variables which take into account the principal relevant factors which determine employment levels in a given state or substate industry. There are two general reasons why theoretical validity may not be met. The first is inadequate knowledge of the relevant determining factors of industry employment. The specification framework described above combined with analysts' own knowledge of peculiar state and substate economic and industrial conditions should help to minimize this. The second reason is data unavailability. Frequent lack of annual time-series data at the state and particularly substate levels for variables known to be causally related to the dependent variable lead to the specification of models which do not fully meet theoretical validity criteria.

The analyst should decide that there is a plausible linear, causal relationship between each independent variable included in a model and the dependent variable. The analyst should not include independent variables to "load up" the R^2 or otherwise improve the calibration results unless there is a theoretical justification for including each variable. After model calibration the analyst should examine the signs of the coefficients of each of the independent variables. If any of the signs are in the opposite direction of what theory would predict, then the analyst should reconsider the use of the particular model for projections purposes.[6]

5.6.3 **Ex-Post Projection Tests**

Ex-post projection tests allow the analyst to measure how well the model can "track" the dependent variable outside the calibration period by simulating a projection. The projection can be evaluated because the actual employment level in the target year is known (since the calibration period is shortened to allow an ex-post projection period). Exhibit 5.10 shows graphically how an ex-post projection test differs from a calibration test and from actual (or ex-ante) projections.

Ex-post projection tests may be conducted when the analyst uses the actual values of the independent variables during the projection period (sometimes called unconditional forecasting) or when the analyst uses projected or estimated values of the independent variables (conditional forecasting). In unconditional forecasting, the analyst has removed that portion of the forecasting error due to uncertainty in the projected year estimates of the independent variables in order to concentrate on the evaluation of the performance of the model alone. In conditional forecasting, the analyst jointly evaluates the performance of the model *and* the exogenous projections of the independent variables.

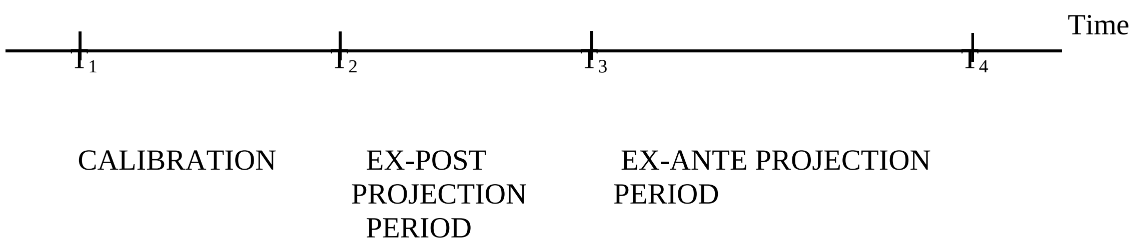
There are, however, several limitations to the ex-post projection test results as a valid basis for assessing likely model performance in an actual (ex-ante) projection mode.

First, because the length of the available calibration period for ex-ante projections is abridged to allow for an ex-post projection period, the regression coefficients estimated with the abridged time-series data may have significantly larger standard errors when the number of available time-series observations is already small

[6] An "incorrect" sign of a relevant variable can occur as a result of high mulitcollinearity. The analyst should attempt to reduce the multicollinearity by eliminating or substituting variables even at the expense of the R^2, so that the model is "credible".

EXHIBIT 5.10

Ex-Post Projection Tests



T_3 = Base Year (for Ex-Ante Projections)

T_4 = Projection Year (for Ex-Ante Projections

T_2 = Base Year (for Ex-Post Projection Tests)

T_3 = Projection Year (for Ex-Post Projection Tests)

T_1 = Beginning Year of Calibration Period (typically 20-30 annual observations for states and substate areas).

Second, the loss of the most recent observations may result in a biased set of regression coefficient estimates if there are significant structural changes occurring within the industry during the recent period from which observations are "withheld" (e.g. technological change, or market share adjustments due to increased international competition).

Third, the length of the ex-post projection period typically will be short to make maximum use of annual time-series observations for the calibration period. The ex-post projection tests the performance of projection models over a shorter projection period (typically only 2 to 4 years) compared to the typical length of the period (5-10 years) in developing actual projections. Thus, the ex-post projection test results may not necessarily be a valid indicator of the performance of the same models for longer projection periods. Despite these three qualifications, ex-post projection tests give the analyst evaluative information not available from just the calibration results.

There are several standard measures of projection error in ex-post projection tests. The mean absolute percent error (MAPE), the root mean squared (RMS) error, and the Theil U-statistic are those used most frequently. Each essentially measures accuracy as the difference between the predicted values of the model and the actual values, but they do differ in some important ways. Each is briefly discussed below.

The MAPE is defined as:

$$MAPE = [1 / T \sum^{T} | (P_t - A_t)/A_t |] \times 100$$

where:

P_t is the predicted value at time t

A_t is the actual value at time t

T is the number of years for which projections are developed.

The RMSE (absolute version) is defined as:

$$RMSE = 1 / T [\sum (P_t - A_t)^2]^{1/2}$$

The RMSE, by squaring each error, places disproportionately large penalties on large individual errors. Because each error is squared, the absolute value does not need to be taken as is done for the MAPE.

The RMSE (change version) is defined as:

$$\text{RMSE (change)} = 1/T \left[\sum (p_t - a_t)^2 \right]^{1/2}$$

where:

$p_t = (P_t - A_{t-1}) / A_{t-1}$

$a_t = (A_t - A_{t-1}) / A_{t-1}$

Here both the predicted and actual values are measured as percent changes over the previous time period. Thus, the measure places emphasis on being able to accurately forecast *changes* in employment from one period to another.

The Theil U-statistic is defined as:

$$U = \left[(1/T) \sum (p_t - a_t)^2 \right]^{1/2} / \left[(1/T) \sum (a_t^2) \right]^{1/2}$$

The U-statistic has some useful properties. U = O if and only if the predictions are all perfect ($p_t = a_t$ for every t). On the other hand, U = 1 when the model leads to the same RMS error (change version) as that which results when the analyst makes a naive, *no-change extrapolation* to obtain a set of predicted values. A model's projections are worse than what would be obtained by the no-change extrapolation if the value of U is calculated to be greater than 1.0. Thus the no-change extrapolation serves as a standard for evaluating a set of projections obtained from single-equation regression models. A case of U = 0.40, for example, is interpreted as the regression model eliminating 60 percent of the RMS error (change version) that would have occurred if the analyst had simply used the naive, no-change, extrapolation. The calculation of the Theil-U is an option built into the pre-defined and customized model specification choices of the LTP. The length of the ex-post projection period is three years.

5.6.4 **External Validity Checks of Actual Projections**

The fourth type of evaluation criterion for projections is to use outside information to assess the reasonableness of the projection values. The outside information might consist of: employment projections developed by another agency or private consultant; historical annual average industry employment growth rates, projected national industry employment growth rates, or projected industry employment growth rates developed by a similar state; and subjective or informal knowledge of the analysts or outside experts. Basically, these "outside" pieces of information are used to develop informal "standards" for the magnitude of what would constitute a reasonable projection for a given state or substate industry. This information would be useful for screening out models and projections which do not seem reasonable, rather than for a systematic projection review and adjustment process, which is described below.

5.7 Reviewing and Adjusting Industry Employment Projections

The need to apply systematic review and adjustment procedures to projections obtained from **any** statistically-based model is well understood by most forecasters. Lawrence Klein, based on his experience in developing forecasts for the national economy, and for state and metropolitan regions, concludes that even the most carefully constructed, theoretically valid, and statistically accurate models may not yield accurate projections if the models are used in a purely mechanical fashion (Klein and Young 1980). Judgment and intuition are critical elements in both the analysis of the statistical results and in the development of the final set of projections.

The review and adjustment process should accomplish at least four objectives: (1) achieve consistency of projections between industries and between the state and substate areas; (2) take into account the informal knowledge of in-house staff and possibly outside experts that is not reflected in the statistical data used to develop the projections; (3) help minimize the chances of large, embarrassing errors in any of the projections; and (4) achieve consistency with labor force and other economic projections for the state or area.

5.7.1 **Achieving Internal Consistency**

For the state and for each of the substate areas, there should be consistency among the set of industry employment projections such that the detailed industry sector projections sum to the projection for the next higher level of aggregation. For example, the projections for all the 4-digit NAICS industries within NAICS 36 should add up to the projection for NAICS 36. There are at least three alternative approaches that can be used to achieve this type of consistency: *top-down*, *modified top-down,* and *bottom-up*.

In a *pure top-down* approach, the employment projections of, for example, each of the 4-digit NAICS industries in the 3-digit NAICS36, no matter how developed, are adjusted proportionately so that their sum is forced to the given employment projection for NAICS 36 (which in turn may have been adjusted to meet the manufacturing employment total, and so on).

In a *modified top-down* approach, some projections for the set of 4-digit NAICS industries may be designated as *given* (or constrained) for the adjustment process. Such a designation might be based on having excellent calibration and/or ex-post projection tests for these industries. Each of the remaining 4-digit industries would be adjusted proportionally to add up to the respective 3-digit industry projection. This strategy is based on projection evaluation results that indicate accuracy tends to increase at higher levels of industry aggregation. But it also takes account that there can be large variation in the accuracy of projections among, say, the 4-digit NAICS industries *within* the given 3-digit industry.

As a further modification to the top-down approach, the analyst occasionally may want to evaluate the quality of the projection for the respective 3-digit NAICS industry before "automatically" using these as control totals for adjusting the more detailed industry projections. The modified top-down approach is illustrated in Exhibit 5.11.

The *bottom-up* approach is to take all the 4-digit NAICS industry projections from the model as given, sum them up, and adjust the respective 3-digit projections so that it equals the sum of the 4-digit projections. The top-down and modified top-down adjustments to achieve consistency are available as options in the LTIP system.

The second type of consistency constraint to be met is between the state projection and the set of substate projections for each industry. How this should be met depends upon whether the substate areas are collectively exhaustive (and mutually exclusive) for the state. There also is the issue of whether any of the substate areas cross state lines, which is the case for a handful of MSAs. If the substate areas are collectively exhaustive, mutually exclusive, and do not cross state boundaries, then the modified top-down procedure described above (but applied to geographic areas) is the recommended approach. Here, the projections for the major metropolitan areas would be the likely candidates for designated constrained projections (assuming the calibration/ex-post projection tests are "good"), and the projections for the remaining substate areas would be adjusted proportionally to meet the state projection for each industry. There, are, of course, cases where the analyst will want to "modify" the modified top-down procedure. These might be, for example, in cases where state employment is concentrated in non-metropolitan areas for a given industry, or when the metropolitan

EXHIBIT 5.11

Modified Top-Down Approach for Achieving Projection Consistency

SIC	Industry Title	Projection Designation[1]	Base-Year Employment	Unadjusted Projection	Proportionality Factor[2]	Adjusted Projection
37	*Transportation Equipment*	*Constrained*	*67,216*	*83,081*		*83,081*
371	Motor Vehicles	Constrained	22,305	26,000		26,000
372	Aircraft and Parts	Constrained	11,441	14,810		14,810
373	Ship and Boat Building	Adjustable	8,665	9,682	0.8686	8,410
374	Railroad Equipment	Constrained	17,127	19,600		19,600
375	Motorcycles, Bicycles	Adjustable	1,960	2,595	0.8686	2,254
376	Guided Missiles	Adjustable	6,668	7,821	0.8686	6,793
379	Misc. Transportation Equip.	Adjustable	5,050	6,004	0.8686	5,215

[1]Determined exogenously, e.g., based on model calibration performance.

[2]Calculated as:

$$\frac{\text{Projection SIC 37 – Constrained (3-digit) Projections}}{\sum \text{Adjustable (3-digit) Projections}} = \frac{22{,}671}{26{,}102} = 0.8686$$

area projections are considered weak. If the substate areas are collectively exhaustive of the state, but if some substate areas overlap or cross state boundaries, then the portion of each substate projection which overlaps or is included in another state must be estimated. If the overlapping or out-of-state portions are counties or county aggregates, then the simplest way to estimate these portions is to use the base-year QCEW county industry employment estimates to form a base-year proportion of the overlapping or out-of-state county to the substate area as a whole. This proportion would be used to estimate that portion of the projection which is in an overlapping county or is out-of-state for each industry. A slightly more sophisticated approach would be to estimate the trend of this proportion using historical ES-202 data. The latter might be justified when, say, the out-of-state county industry employment has been growing at a much higher (or lower) rate than the substate area as a whole. The trend could be estimated with a simple regression analysis or extrapolated based on two representative historical, annual observations (as in shift-and-share models). The estimates of the overlapping or out-of-state portions of substate areas would be subtracted from the respective substate area projections for each industry sector and the modified top-down approach discussed above could then be used.

In states where the substate areas are not collectively exhaustive, then the use of the state projections control totals for adjusting the substate area projections will not work. Instead, the analyst should form for each industry a geographic residual category by subtracting the sum of the projections for the substate areas from the state projection. Each of the residual projections should be reviewed to identify cases in which the residual projection is negative, and also "unrealistically" large or small. Using the ES-202 historical data, a rough "projection" can be made for the residual area for each industry using either a simple regression model (residual industry employment against time) or a shift-and-share model. This "rough" residual area projection can serve as a criterion for judging the realism of the difference between the state projection and the sum of the substate projections for each industry, as well as for adjusting the substate projections when the residual is not realistic. This adjustment can be made using the modified, top-down approach. Here the residual area is included as one of the exhaustive set of substate areas and the "rough" projection for the residual area is designated as constrained. This procedure is illustrated in Exhibit 5.12.

5.7.2 **Internal Review and Adjustment**

After the set of industry employment projections have been adjusted to be internally consistent, *systematic* in-house and outside reviews of the projections should be undertaken. There is no one, accepted procedure to organize the reviews, but there are some general guidelines and tasks which should help make the review process an effective one.

Before a staff meeting, the following information should be assembled for each industry to be reviewed (at least each *key* industry):

* Most recent monthly industry employment time-series starting from the base-year used in developing the projections.

* Calculations of the projected average annual industry employment growth rates along with recent historical average annual growth rates.

* Employment projections developed by other agencies or private consulting firms.

* Projections developed by the LMI staff in states with relatively similar economies and industrial structure; calculations of the respective projected average annual industry employment growth rates.

EXHIBIT 5.12

Achieving Consistency between State and Substate Industry Projections: The Case of Non-Exhaustive Substate Areas

Geographic Area	NAICS	Projection Designation	Unadjusted Projection	Proportionality Factor[1]	Adjusted Projection
State	336	Constrained	50,000		50,000
Substate area A	336	Constrained	25,000		25,000
Substate area B	336	Constrained	15,000		15,000
Substate area C	336	Adjustable	5,000	0.9375	4,688
Substate area D	336	Adjustable	3,000	0.9375	2,812
Imputed Residual Area	336		2,000		
Exogenous (rough) Residual area Projection	336	Constrained	2,500		2,500

[1]Calculated as: $\frac{\text{(State projection - Constrained substate projections)}}{\text{Adjustable substate projections}}$

* The calibration results, any ex-post projection tests, and any special assumptions underlying the projection models used.

* Any current announcements of prospective establishment openings, expansion, contractions or closings. Some of this information can be obtained from state commerce or economic development agencies, local industrial revenue bond authorities, local chambers of commerce, etc.

This information should be distributed in advance to each of the staff who will take part in the review and adjustment process. Each staff member would be asked to review all the information but to take special responsibility for a more intensive prior review of the projections of a group of related industries (e.g. durable good manufacturing industries). Discussion at the staff meeting would focus on (1) *whether* the projection should be adjusted and if so *why,* (i.e., the reasons) and then (2) *how much* of an adjustment, for each key industry. At this point various types of group decision making rules can be used for gaining either a consensus or group average projection.[7] Among the simplest and most reliable rule would be to ask each staff person, after the discussion, to write his or her "best" employment projection on a sheet of paper, and then calculate the staff average, for each industry being reviewed. It is important though that the "adjusted best projection" be based on the preliminary projection estimate(s) and the supplementary information presented, rather than be a number which "falls out of the sky" and could not be defended with "good reasons."

After the adjustments are made, the analyst then would need to go back and again achieve internal projection consistency between industries and between the state and substate areas. To avoid doing this again at this point, the internal review could take place before adjusting the projections to meet consistency requirements.

5.7.3 **External Review**

An external review of the industry employment projections might be conducted at this stage, followed by a more formal external review after the set of occupational employment projections are developed. Alternatively the LMI staff may wish to have only one external review (of both the set of industry and occupational employment projections) after the development of the occupational employment projections.

5.8 **Reconciling Employment Projections with Labor Force Projections**

The industry employment projections methodology described above is almost demand-side driven. That is, the factors predicting change in industry employment levels are demand for goods and services produced in a particular state or substate area. Implicitly it is assumed that there will not be a scarcity of any of the factors of production – land, labor, capital – that would constrain the ability to expand production and employment, should demand for product grow in the future.

In recent years this implicit assumption has become more questionable at the national level, but even more so in particular states and substate areas. Specifically, those underlying forces that determine the future size of the labor force – the age structure of the population, labor force participation rates, and levels of unemployment – have been shifting from their long-term historical trends, leading to increased uncertainty of future levels of labor supply, and distinct possibilities that future industry employment in some areas may be constrained by unavailability of labor (U.S. Department of Labor, 1998).

To increase the accuracy of the long-term industry employment projections, as well as to avoid confusion

---

[7] See Armstrong, 1978, chapter 6 for a description. Consult other chapters for more elaborate review and evaluation techniques.

by LMI users over seemingly inconsistent projections, state LMI analysts should take into account labor force projections at the review and adjustment stage. A recent survey of state LMI agencies (U.S. Department of Labor, 1998) has indicated that a small minority of states attempt to reconcile employment projections with labor force projections, though nearly all respondents recognize the importance of doing so. Not having access to software that is easy to use and based on theoretically valid procedures was cited as the major reason for not taking this step.

There are a number of methodological options for incorporating labor supply issues into the demand-driven employment projections process, some more theoretically valid, but also more complex, than others. But regardless of which option is chosen, it is important that LMI analysts understand the relationships among basic demographic and economic variables in order to be able to take labor supply issues into account to improve the accuracy of employment projections.

5.8.1 **Relationships among Demographic and Economic Variables**

The basic relationships between demographic and economic variables are shown in Exhibit 5.13 The demographic realm consists of total population, migration, and natural increase (births minus deaths). The economic realm consists of the labor force (labor supply), income and output (demand), and wages/unemployment/employment (supply and demand matching). The arrows indicate directional linkages. The dashed lines indicate the primary pairwise interaction within each realm. The reader should note that the output-employment relationship, the basis for the specification of the industry employment projection models in the L TP, is only one aspect of the total picture. To incorporate labor supply factors necessarily involves knowledge of population trends. In turn, population growth and decline is a function of both natural increase and net migration. The entire system is highly simultaneous, and a substantial body of economic and demographic theory is needed to fully specify the relationships among all of these variables.

Insight into the details of the labor supply and demographic components in Exhibit 5.13 can be gained with the help of three simple, but important, accounting identities. The first identity relates change in a region's population between two periods to the demographic components of change:

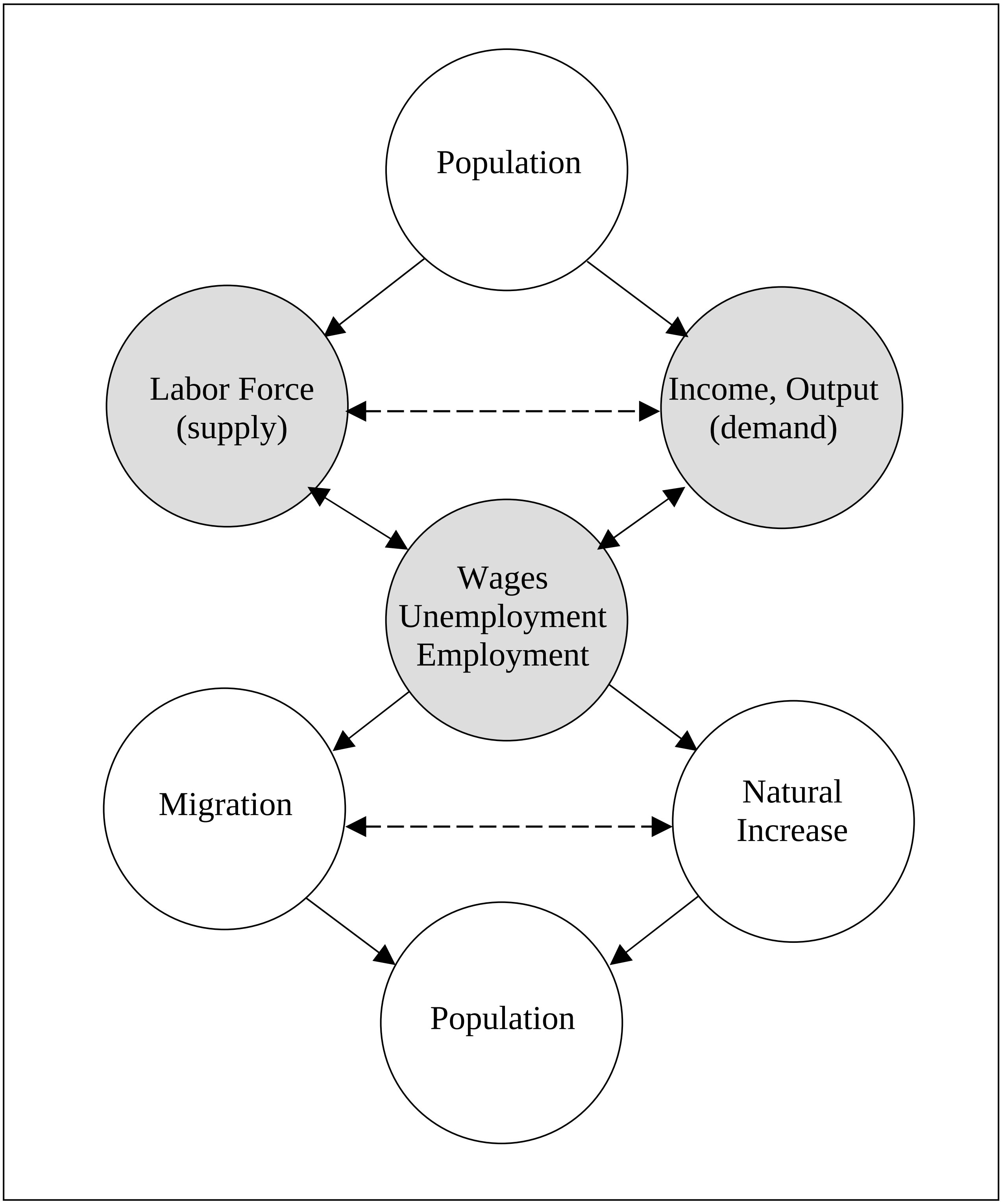
$$P_{t+1} = P_t + B_{t,t+t} - D_{t,t+t} + I_{t,t+t} - 0_{t,t+t} + IM_{t,t+t} - EM_{t,t+t}$$

This equation states that population at time *t+1* is equal to population at time t plus the births (B), in-migration (I), and immigration (1M) over the period, and less the deaths (D), out-migration (0), and emigration (EM) over the period. This basic identity could be further refined to isolate particular age or socio-economic groups. The net result of births minus deaths is termed natural increase or natural decrease. Migration flows are partitioned into international flows (immigration and emigration) and domestic interregional flows (in- and out-migration). Net migration is typically defined as the result of in-migration minus out-migration. Emigration data are unavailable in the U.S., so the representation of international flows is restricted to just immigration.

The decomposition of population change is an important conceptual step since different structural economic relationships are needed to explain variation in each component. For example, variables considered to influence fertility are different than those considered to influence out-migration. The importance of each component of demographic change will likely vary by region as well. So, as another example, states in the mid-West with elderly age structures may need to pay particular attention to modeling mortality and amenity-driven out-migration, while border states may need special attention to immigration. As a general rule, the dominant demographic forces at the regional level are births arid internal migration, while immigration is certainly an important labor supply factor in states like California, Florida, and Texas. Regional fertility rates are directly related to women's labor force participation rate. The general theoretical concept linking birthing

EXHIBIT 5.13

Relationships among Demographic and Economic Variables



decisions to economic factors is the life-time, or dynamic, labor supply. Women's employment histories, marital histories, and fertility histories are all interrelated. It is possible, therefore, to posit both income and price effects on fertility. So, as the price of children increases, the number demanded should fall. The price of children is tied closely to women's labor force participation, since female wages represent the opportunity cost of child bearing and child rearing. A pure increase in income should theoretically increase demand for children. The income effect is ambiguous, however, because higher income groups choose to invest more per child. Other factors that affect the price of children such as the price and availability of child care, or changes in social mores regarding day care, should also affect the female labor supply. On a regional scale migration is also linked to fertility since the regional stock of child-bearing aged women can increase or decrease depending upon the prevailing age-specific migration patterns. Factors typically taken into account to estimate regional fertility rates include wage and income measures, the female labor force participation rate, unemployment rate and, when available, the marriage rate.

There are several well-established economic theories of migration behavior. In neoclassical regional growth theory migration from low-wage to high-wage regions is a major part of the equilibration process. At the individual level, human capital and job-search models are often used to frame migration behavior. In both theoretical frameworks, individuals are assumed to choose among a set of locations (regions) based upon the discounted stream of costs and benefits associated with each alternative location. The job-search framework adds uncertainty about the future stream of costs and benefits to the utility maximization problem. Given this framework, regions can be viewed as possessing their own particular "basket" of attractive and repulsive characteristics. These include economic factors such as employment growth and wage and salary levels, and non-economic factors such as amenities and climate. Age, educational attainment, sex, race, and labor market status differences also explain the migration propensity of individuals.

Two additional accounting identities clarify the connections between demographic variables and the "standard" economic concepts of labor market analysis. Population in a single period can be decomposed into the employed (E), the unemployed (U), and those not in the labor force (N):

$$P_t = E_t + U_t + N_t$$

The above equation can be rearranged to isolate the labor force participation rate as,

$$LFPR = \frac{E_t + U_t}{P_{t(x>16)}}$$

where Pt (x >16) indicates the population age 16 and older. The importance of this decomposition, is that, once again, the components of population change in Eq. X differentially affect components of the labor market (E, U, and N), and therefore labor supply. For example, since births are linked to women's labor supply decisions, increasing regional births will generally be accompanied by declining labor supply. Employment growth in a region will be in conjunction with lagged in-migration response, while unemployment will be lagged by an increased out-migration response. As Treyz et al. (1992) note, "The indigenous labor force competes with potential migrants for jobs. Migrants in turn alter the long-term demographic structure of the population. The demographic-economic interaction becomes pronounced for rapidly growing or declining regions, and models that ignore this interaction may be misleading."

5.8.2 **Choosing Among Options for Reconciling Labor Force and Employment Projections**

Several options that attempt to include all or most of the relationships discussed above in section 5.8.1 are discussed in Goldstein and Sweeney (2000). However, those such as integrated demo-economic models that simultaneously project employment demand and supply will not become feasible for most states in the short- or medium-term, due to data constraints and that the present LTP system would need to be significantly altered. Yet in the long-term there is promise for linking a labor supply module directly into the LTP system.

In the meantime, prototype software has been developed by the ALMIS Long-Term Employment Projections and Census Tools Consortium that has the potential to meet the needs and analytical capability of LMI agencies in the short and medium terms. This software produces labor supply projections at the state level, and allows the LMI analyst to conveniently compare these projections to the demand-driven employment projections, at least at the aggregate level. The analyst is then in the position of knowing if the two sets of projections are consistent or not. If they are inconsistent, the analyst should consider which (or both) set of projections should be revised, based upon a critical examination of the assumptions underlying both sets, against an examination of selected demographic and economic variables that may reveal recent shifts. This may be best accomplished with the inclusion of a state demographer, and perhaps others knowledgeable of labor supply trends, within the review and adjustment process.

REFERENCES

Armstrong, J. Scott (1978). *Long Range Forecasting*. New York: John Wiley and Sons.

Bendavid-Val, Avrom (1991). Regional *and Local Economic Analysis for Practitioners*, 4th edition. New York: Praeger.

Bjornstad, David J. and Richard C. Tepel. "A Comparative Analysis of Errors in Long-Term Econometric Forecasts for Aggregate and Component Variables." Unpublished paper, Oak Ridge National Laboratory.

Cruze, Alvin M., Harvey A. Goldstein, John E.S. Lawrence, Edward M. Bergman, and Katherine A. Desmond (1984). *Evaluation of Industry and Occupational Employment Projections Made by State Employment Security Agencies*. Final report to the U.S. Bureau of Labor Statistics. Research Triangle Park NC: Research Triangle Institute.

Fishkind, Henry and R. Blaine Roberts (1978). "Two Methods of Projecting Occupational Employment," *Monthly Labor Review*, May, 57-58.

Goldstein, Harold (1965). "Methods of Projecting Supply and Demand in High Level Occupations," *Proceedings of the American Statistics Association, Social Statistics Section*, 89-93.

Goldstein, Harvey A. (2004). *The Conversion to NAICS: Some Implications for States to Develop Long-Term Industry Projections*. Final Report to the Projections Workgroup, U.S. Department of Labor, Employment and Training Administration.

Goldstein, Harvey A. (2005). *New Economy Indicators for Long-Term Industry Employment Projections.* Final Report to the Projections Workgroup, U.S. Department of Labor, Employment and Training Administration.

Goldstein, Harvey A. (1981). *Occupational Employment Projections for Labor Market Areas: An Analysis of Alternative Approaches*. U.S. Department of Labor, Employment and Training Administration, R&D Monograph 80.

Goldstein, Harvey A. and Karen L. Becker (1996). *Projections Utilization Study*. Final report prepared for the ALMIS Long-Term Employment and Census Tools Consortium, U.S. Department of Labor, Employment and Training Administration.

Goldstein, Harvey A. and Alvin M. Cruze (1987). "An Evaluation of State Projections for Industry and Occupational Employment," *Monthly Labor Review*, October, 29-38.

Goldstein, Harvey A. and Andrius Nemickas (1998). *Evaluation of States' Industry Employment Projections to 1995*. Final report to the ALMIS Long-Term Employment Projections and Census Tools Consortium, U.S. Department of Labor, Employment and Training Administration.

Goldstein, Harvey A. and Stuart Sweeney (2000). *Demo-Economic Specifications for State*

Industry Employment Projections. Final report to the ALMIS Long-Term Employment Projections and Census Tools Consortium, U.S. Department of Labor, Employment and Training Administration.

Isserman, Andrew (1986). *Population Change and the Economy: Social Science Theories and Models.* Boston: Kluwer-Nijhoff.

Klein, Lawrence R., and Richard M. Young (1980). *An Introduction to Econometric Forecasting Models*. Lexington, MA: D.C. Heath and Co.

Less, Lawrence J. (1992). "An Evaluation of Industry Projections: A Case Study of the Ohio Economy," *Economic Development Quarterly*, August.

Less, Lawrence J. (1997). "Ohio Industry Employment Projections for 1995: An Evaluation." ALMIS Long-Tem1lndustry Projections and Census Tools Consortium, U.S. Department of Labor, Employment and Training Administration.

Less, Lawrence J. (1999). "Ohio Industry Employment Projections for 1995: A Case Study Evaluation. Final report to the ALMIS Long-Tem1lndustry Projections, U.S. Department of Labor, Employment and Training Administration.

McLean, Mary L. and Kenneth P. Voytek (1992). *Understanding Your Economy: Using Analysis to Guide Local Strategic Planning*. Chicago, IL: Planners Press.

Michigan Employment Security Commission and the University of Michigan (1984). *Alternative Projections Project for Michigan, Final Report*. Lansing MI: Michigan Employment Security Commission.

Pindyck, Robert S. and David L. Rubinfeld (1997). *Econometric Models and Economic Forecasts*, 4th edition. New York: McGraw-Hill.

Stapleford, John E., William R. Latham, Omar J. Borla, and Young Doo Wang. *Alternative Projections of 1990 - Occupational Employment for Delaware, Final Report*. Newark DE: Bureau of Economic and Business Research.

Texas Employment Commission. *Industry Employment Projections. Evaluation of Alternative Methods, Final report*. Austin, TX: Texas Employment Commission.

Theil, Henri (1966). *Applied Economic Forecasting.* Amsterdam: North Holland.

Treyz, George, D. Rickman, and G. Shao (1992). "The REMI Economic-demographic Forecasting and Simulation Model." *International Regional Science Review*. 14(3): 221-253.

U.S. Department of Commerce (1992). *Regional Multipliers: A User Handbook for the Regional Input-Output Modeling System (RIMS II).* Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor, Bureau of Labor Statistics (1969). *Tomorrow's Manpower Needs,* Vols. I-IV. Washington, DC: U.S. Bureau of Labor Statistics.

U.S. Department of Labor, Employment and Training Administration, ALMIS Long- Term Employment and Census Tools Consortium (1996). *Guidelines for Long-Term Employment Projections.*

U.S. Department of Labor, Employment and Training Administration, ALMIS Long-Term Projections and Census Tools Consortium (1998). *Survey of Labor Supply Projections*.