Growth
Decline
High Demand
Moderate Demand
Low Demand
Good Bets
Best Bets

Occupational Descriptors

Enhancing Information For Customer Choice

Summer 2002

Illinois Department of Employment Security

Economic Information and Analysis Division

George W. Putnam
Waleed K. Almousa

Edited by:
Yolanda Y. Harris

AMERICA'S WORKFORCE NETWORK SM

U.S. Department of Labor
Employment and Training Administration

A Report for the Projections Managing Partnership

Table of Contents

Executive Summary

Occupational projections are one of the most popular types of data customers use to help them make informed decisions about their job prospects, career, or school curriculum. This is especially evident through the roughly one million hits the Occupational Outlook Handbook receives each week on the Internet. In response, states are seeking innovative ways to display occupational projections that are more informative to customers than in times past. More specifically, states are replacing their traditional numerical display of occupational projections with descriptors, such as “hot jobs,” or “very favorable.”

The descriptors tend to be more comprehensible to customers than mere numbers alone; and they convey a more realistic level of imprecision. However, as this report reveals, the methods and statistical rigor used to develop occupational descriptors vary from state to state. The impact on customers is twofold: (a) customers are misinformed about the outlook of occupations; and (b) customers are unable to appropriately compare occupational projections across state borders.

This report attributes such results to specific weaknesses inherent in some descriptor methodologies. For one, states should examine the statistical properties of data series before defining descriptors. Second, states should consider the use of multiple criteria, i.e., data series, to assign descriptors. Third, states should select a descriptor methodology that is best suited to their data constraints and meets the needs of customers.

The Workforce Investment Act of 1998 challenges states to improve their development and publication of labor market information. The data must be comprehensible and reliable to customers. To this end, this report focuses on improving and standardizing states’ occupational descriptor methodologies.

The purpose of this paper is to offer a systematic approach to developing descriptors for both short- and long-term projections. The paper proposes a methodology that involves four steps: (1) selecting the number of data series; (2) testing each series' statistical properties; (3) determining the number of descriptor categories; and (4) building the descriptor table.

Through one case study, the paper demonstrates a method for integrating descriptors into short- and long-term projections. Through a second case, this report investigates two different types of state descriptor methodologies. The results of the second study reveal a method that increases the accuracy and reliability of descriptors. It is a method that standardizes the framework for development of descriptors across all states and that improves the quality of projections information released to customers.

Introduction

This paper presents guidelines for developing occupational descriptors. It provides tips on selecting the number of data series to use, testing each series' statistical properties, and determining the number of categories to present. It also reviews the weakness inherent in many states' methodologies and underscores the need for standardization and increased accuracy. Two case studies provide step-by-step instructions for developing occupational descriptors across different time horizons (two-year and ten-year projections) and improving the accuracy of state-level descriptors.

The demand for timely and accurate labor market information, particularly on career opportunities, is growing. In a 1999 survey of labor market information users from across the Nation's public and private sectors, Keith, Wilson, and Ermatinger found that 93% had used one or more sources of occupational information within the past year. Although the information they needed varied greatly, more than one-half had requested both short-term (two-year) and long-term (ten-year) occupational projections to counsel students, help job seekers, conduct economic research and develop training programs.

The survey also asked these individuals to evaluate two report formats most commonly used to present projections results. The first format published projections on a numerical basis only. In this, the traditional format, key numbers such as the projected number of new jobs, the projected growth rate, the projected number of replacements, and the total projected number of job openings were listed for each occupation.

A second format, a synthetic report, provided both a numeric and a descriptive presentation of results. It put the numbers in context, such as describing different rates of annual growth as faster or slower than average. This hybrid approach represented a departure from the way projections have most often been published, and survey results indicated it was successful.

Most respondents (two-thirds) preferred the synthetic report format, regardless of whether they used the information for counseling, planning, or training. Not surprisingly, most found the traditional report difficult to read and understand and indicated that they would need additional information to interpret the projections.

In response to users' growing demand for projections, the number of state analysts who have begun or are considering producing occupational descriptors is high. Well over one-half of the states are either producing descriptors or considering the incorporation of descriptors in the dissemination of occupational projections (Wettemann, 1997). Some analysts utilize a single data series, and others use multiple data series, to allocate occupations among descriptor categories. Maine, for example, bases its descriptors on projected annual percentage growth. Washington, on the other hand, creates a more elaborate framework, using percentage growth, the number of job openings, the ratio of supply to demand, and the rate of placement to base its descriptors.

However well-intentioned states' responsiveness to users' needs, the lack of uniformity in descriptor methodology across states fosters the likelihood of confusing, and even misleading, users. A "hot job" occupation may be described in terms of a significant difference from the average occupation (standard deviation) or merely as a ranking in the top 10 percent of occupations (decile). Whether a series can be described using standardized statistics (standard deviation) or less rigorous measures, such as quartiles or deciles, depends on its statistical properties, e.g., normal distribution of data values. States must examine these properties to apportion correctly the data and assign cut-off levels. Analysts and users need to understand the implication of these statistical requirements for the presentation of occupational information.

Last, the number of descriptor categories states use to assign to occupations varies widely. Some states report projections results using six descriptors (ranging from excellent to declining), while others use only one label, most notably "hot jobs" or "fastest growing." The value-added of descriptor categories is, in part, its intuitive appeal to users. Occupational descriptors enhance the usability of projections results.

In contrast to numeric presentations, the implementation of descriptors communicates a more realistic level of imprecision in the projections than is the case with numeric presentations. As analysts are aware, both the input data and estimation methodology are subject to a measure of error. In fact, states routinely publish numeric results of occupational projections, e.g., 330 annual openings for management analysts, without an error estimate or confidence levels.

Evaluation of Descriptors

In recognition of the error embedded in the projections process, evaluations of the results typically are anchored in the accuracy of descriptors and not the numeric presentation of the data, i.e., projected annual total openings. Veneri (1997) measured the accuracy of the national 1984 to 1995 occupational projections by direction of change and the proximity of the actual category to the original designation. For the first measure, direction, an occupational descriptor was accurate when the occupation that was projected to decline or grow actually declined or increased. As for proximity, a projection was accurate if the actual descriptor was the same category as the projected descriptor or only one category off. Using this methodology Veneri found that the direction of employment change was correct for more than 70 percent of the occupations studied and 52 percent were on target or only one category away.

Do these percentages hold up at the state level? Barylo and Senf (1998) answered this question, duplicating Veneri's study with data from Minnesota. They found only 58% of Minnesota's projections scored correctly on direction and 35% on accuracy of proximity. These findings suggest important implications for the implementation of descriptor methodology at the state level.

First, states' input data and estimation methodology may produce projections that are less reliable than at the national level and, consequently, states should deliberate carefully the implication of national descriptors for the presentation of state projections. The national projections are reported in six groups: much faster than average, faster than average, about as fast as average, slower than average, little or no change, and decline. Fewer categories, e.g., four, may produce more of a convergence in the accuracy of presenting national and state projections to users. The boost in validity may outweigh the disadvantage in information loss to users inherent in the decline from six to four descriptor labels.

Second, Barylo and Senf reported that the accuracy of the descriptor was correlated with the employment size of an occupation. Occupations with larger-sized employment tend to produce a tighter fit between projected and actual descriptors. Third, both the national and state studies based descriptor categories on a measure of growth rate. The findings encourage state analysts to reconsider a reliance on only a single data series and to establish a framework for descriptors using multiple variables. This approach may lend greater stability to the robustness of descriptors over time.

Improving state and local labor market information is a priority of the Workforce Investment Act (WIA). Since employment projections are among the most frequently requested types of labor market information, they will be among the first to receive scrutiny under the requirements of this legislation. WIA challenges states to improve their projections. To this end, this report focuses on improving states' occupational descriptor methodologies.

Findings from Keith and others prod LMI producers to rethink traditional displays of occupational projections to include descriptors. Barylo and Senf offer compelling empirical evidence to support a synthetic and parsimonious approach to the presentation of state and local labor market information. WIA creates an accountability structure whereby producers must consult with local users in the improvement of local information. Implied in the linkage between producers and users is the standardization of information across local labor markets.

In the following section, we describe different solution scenarios for implementation of descriptors. We organize the material around the selection of data series (single or multiple variables), the testing of statistical properties of data series, the derivation of descriptors and the framing of a descriptor table. Finally, this report presents two case studies. The first case study explores the integration of short- and long-term projections into a descriptor methodology. The second case study proposes a state descriptor methodology and compares the accuracy results to the national descriptor methodology.

Developing a Descriptor Methodology: Different Solution Scenarios

The following sections explain in detail the series of steps that state analysts should take when selecting the number of data series to use, testing the statistical properties, apportioning the data to each descriptor category, and framing the descriptor table. Since analysts may choose any number of data series to develop their descriptors, this section reveals that in essence analysts must balance methodological issues and user needs. With this in mind, the section also discusses at length the pros and cons of using one or more data series.

Step 1: Select the number of data series

Single data series

Many of the previously mentioned studies based the descriptors on the projected employment growth rate, a single series. The appeal of this variable is intuitive: at a glance, users can identify the fastest growing occupations. Though this appeal is understandable, it presents a one-dimensional view.

It biases users' views against occupations that may not have fast rates of growth, but nonetheless will have a large number of projected job openings. Consequently, the growth rate identifies the fastest growing occupations, but not necessarily those with the largest number of job opportunities. The correlation between an occupation's employment size and accuracy of the descriptor is well known: the bigger the occupation's employment the more accurate the descriptor. Implementation of growth rate introduces a bias against large-sized occupations and, yet, the descriptors for these occupations are more accurate.

Multiple data series

When analysts use two or more data series to delineate descriptor categories, they provide users with a multidimensional view of occupational information. One of the recommendations arising from Minnesota's research is to include both the growth rate and employment size as a basis of each category. The outlook for a small-sized occupation with a fast growth rate may rate only "favorable" in contrast to a large-sized occupation with an average growth rate that may merit a "very favorable rating."

The kinds of variables states use varies widely. Florida, for example, computes a composite occupational index based on projected annual openings, projected annual growth rate, and expected entry-level wage. This index, coupled with expert review, is used to rank occupations. Utah takes a similar approach, although it grades occupations on a composite ranking. The top fifteen percent receive an "A" grade, the bottom fifty-five percent are designated with a "C" grade, and those that fall in the middle have a "B" grade. Pennsylvania expands beyond the standard labor demand variables to include measures of labor supply. Moreover, this state assigns different index weights to national, state, and local data series.

Step 2: Test the statistical properties of each data series

After selecting the number of variables to include, analysts need to examine the statistical properties of each data series to correctly apportion the number of occupations to each descriptor. In fact, the results of the investigation will narrow the range of statistics for establishing an empirical basis to allocate occupations to specific descriptor categories. The two key properties that state analysts should examine are (1) collinearity of data series and (2) the distribution of values within a data series. Collinearity assumes relevance only in the case of multiple data series, but the distribution of data values must be tested regardless of whether the descriptor is based on one or more data series.

Collinearity of data series

Two series are perfectly collinear when data values in series A and series B are a mathematical function of each other; or, stated differently, the function of one variable is a linear combination of the function of another variable (Pindyck and Rubinfeld, 1976). Projected annual growth rate and projected ten-year growth rate in an occupation are a linear combination. The latter is derived as a multiple of the former.

When perfect collinearity is ignored the designation of occupations to descriptors becomes biased. Since the projected ten-year rate is conditioned by the projected annual rate, a descriptor classification derived from these two series will reflect a duplicate weight on the projected rate. To most users this bias would be inconspicuous and, consequently, could flaw career choices.

Distribution of data values

The second statistical property analysts need to examine is the distribution of data values in a series. Since data values displayed in a frequency distribution can assume a variety of shapes, it is particularly important for the analyst to determine whether the distribution of data values approximates a normal curve. The normal curve exhibits several mathematical conditions (Blalock, 1972):

- Mean and median of the series are equal
- Variance around the mean is constant
- The curve is bell-shaped and symmetrical around the mean.

Tests for normality include plot graphs and statistical measures. The histogram, an often-used graph, segments data into equal intervals. A bar graph for each interval displays the count of observations and, taken as a single continuum, the height of the bars should approximate a normal curve. A more sophisticated graph is the normal probability plot in which each observed value is paired with its expected value from the normal distribution. Under the condition of normality, the pairing of observed and expected values will approach a straight line. A substantial departure from linearity suggests data abnormalities.

If analysts have trouble discerning a normal distribution based on plot graphs, statistical tests also offer clues as to the shape of the distribution. In a normally distributed series the mean and median converge, the skewness (symmetry) approaches zero and the peakedness (kurtosis) of the distribution approximates a value of three. In samples of fifty or more observations, the Kolmogorov-Smirnov (K-S) statistic tests whether the mean and variance of the data series equals the mean and variance of a normally distributed series (Blalock, 1972).

In a normally distributed series, tests for significance are appropriate to describe the data values. For example, projected annual growth rates beyond two standard deviations from the average projected rate can be described as significantly different at a certainty level of 99 percent. A series that fails to meet the conditions for normality can be represented only by descriptive statistics, such as proportions, percentages, quartiles, and deciles. In the case of the latter, users know only that “best bet” occupations, for example, are marginally better than the other. Diagnostic tests are available to analysts to guide them in choosing between statistics that measure significant differences and those that describe a segment of the distribution, e.g., top ten percent.

Step 3: Determine the number of descriptor categories

In evaluating the accuracy of its 1984-95 growth descriptors, the Bureau of Labor Statistics found that the "actual" category was some distance away from the "projected" category for almost one-half of the occupations studied. The projected growth descriptor for optometrists, for example, was "much faster than average" when in fact the descriptor that most accurately reflected the actual job growth in this profession was "average." Similarly childcare workers had a projected descriptor of "slower than average," but the descriptor that best conveyed their actual growth was "much faster than average."

Minnesota's accuracy rate of only 35% begs for alternative classifications to the national approach, an approach that uses six descriptor categories. The need to improve state-level accuracy is clear. The following paragraphs describe a few of these alternative classifications.

One- and two-descriptor categories

A one-category descriptor, "hot jobs," identifies some subpopulation of occupations based on a criterion or criteria. In Delaware, for example, the "best bet" occupations must satisfy three requirements as measured by total employment, projected annual openings and projected growth rate. The two-category descriptor differentiates growth from decline and increases the likelihood of a successful projection, but very often job seekers and program planners need more information than a simple dichotomy of positive and negative growth.

Three or more descriptor categories

Three descriptor categories capture direction, (growth/decline) coupled with a differentiation among growth rates: (1) unfavorable occupations- declining growth, (2) favorable occupations- ranging from some growth to average growth and (3) very favorable occupations-greater than average growth. This classification moves the data one step beyond the simplicity of direction and employs a more specific breakout of growth. The remaining classifications combine direction with a finer delineation of levels within direction. A common four-category descriptor is (1) very unfavorable occupations- declining growth, (2) unfavorable occupations- slow growth, (3) favorable occupations- average growth and (4) very favorable occupations- faster than average growth.

Whether analysts choose to capture a subpopulation (best bets), direction, or some combination of direction and magnitude of change, their choice should reflect a deliberate consideration of, and coordination between, critical methodological issues (data reliability) and sensitivity to user needs. In essence, state analysts must balance user needs while remaining committed to disseminating accurate labor market information.

Step 4: Build the descriptor table

The steps enumerated thus far include selecting the data series, investigating the statistical properties, and determining the descriptor categories. The following examples bring to the forefront a few of the methodological choices analysts have.

Table A:
Single Data Series and a Two-Category Descriptor

	Projected Growth Rate	
	Negative	Positive
Descriptor	unfavorable	favorable

In Table A, two descriptors ("unfavorable" and "favorable") describe a single data series, the projected growth rate. Those occupations assigned an unfavorable descriptor have negative projected rates while those assigned a favorable descriptor have positive rates of growth. The descriptor labels are, of course, a matter of choice. Since this table uses a single data series, there is no need to check for collinearity. Moreover, since this descriptor expresses the growth rate in terms of a dichotomy (positive and negative) and not statistical significance, there is no need to check for normality.

In Table B, two data series--projected growth rate and employment size--are juxtaposed to construct three descriptors: "unfavorable," "favorable," and "very favorable." The next step is to correctly apportion the occupations among the categories. Since this example includes multiple data series, we must test for collinearity of the data series and determine whether each data series is normally distributed.

Table B:
Two Data Series and a Three-Category Descriptor

	Projected Growth Rate		
Employment Size	Slow	Average	Fast
Small	unfavorable	favorable	very favorable
Moderate	unfavorable	favorable	very favorable
Large	favorable	very favorable	very favorable

Collinearity of data series. Projected growth rate and employment size are interval data and, therefore, the Pearson correlation is an appropriate measure of collinearity of these series.

(In the case of categorical data, non-parametric tests, such as the Spearman correlation or Kendall tau, should be implemented.) For the purpose of this example, let's assume that the data series are not collinear. Thus, the results of the analysis show that the null hypothesis, no perfect collinearity, cannot be rejected.

Distribution of data series. Distribution plots that display the data values in graphs provide an opportunity to observe the symmetry of the series. Statistical measures, both descriptive and analytical, test whether the series meets critical mathematical conditions. In this example, we assume that the projected growth rate series is normally distributed and, therefore, the rules of statistical significance can apply. "Average" growth rate is the mean rate plus or minus one standard deviation (see Table B). Occupations designated as "fast" have a growth rate in excess of one standard deviation from the mean while those considered "slow" have rates of growth that fall one standard deviation below the mean. Consequently, the rate of growth for occupations in the "fast" category is significantly higher than for those in either the "average" or "slow" categories.

Typically, occupations are not normally distributed by employment size because small-sized occupations tend to cluster at the tail of the distribution. Consequently, the median will invariably fall below a reasonable proximity of the mean. Although significance measures are not appropriate, analysts have several choices to differentiate employment levels. An arbitrary proportion designation could segment the distribution. As an illustration for the example in Table B, the smallest 50% fall to the small category, the top 20% to the large category, and the remainder to the moderate-sized category.

Build the descriptor table. The descriptor labels (unfavorable, favorable, and very favorable) communicate a level of complexity beyond a simple dichotomy and, indeed, reflect gradations of a combination between growth and employment size. Once the data series have been selected and analyzed and the number of descriptor labels decided, the final step is to build a descriptor table. The specific task is to assign descriptors to occupations with similar combinations of growth and employment size.

Continuing with the illustration in Table B, occupations with a significantly higher than average projected growth rate, regardless of size, certainly warrant the most positive outlook. In addition, since large occupations with an average projected rate represent a sizeable portion of labor demand they merit a very favorable label. The favorable designation is also reserved for moderate- or small-sized occupations exhibiting average growth or large-sized occupations with slow growth. Once again, the latter are a special case in which even slow growing occupations with large employment base represents substantial labor demand. The remaining occupations offer the least positive outlook, as indicated by an unfavorable grade.

Users of occupational information advocate for the incorporation of descriptors into the display of results from the short-term and long-term projections. The following case study investigates an approach to achieve that objective based on Illinois occupational projections covering two time horizons, 1997-1999 and 1996-2006. We discuss the benefits of this approach, propose a methodology, and analyze the results. In addition, we highlight the major variations in labor demand by explaining the assumptions on which two- and ten-year industry projections are usually based. In a subsequent case study, a comparison is made between national and state descriptor methodologies.

Case Study: A Comparison of Descriptors for Two Projection Horizons

Students, educators and job seekers benefit from the juxtaposition of short-term (two-year) and long-term (ten-year) projections to assess changes in the job market. Since short- and long-term job projections may differ markedly for some occupations, analysts must understand the forces that produce such discrepancies and present employment statistics in a way that helps users evaluate their career paths. This case study explains how Illinois developed its short- and long-term occupational descriptors.

Demand for workers varies widely by industry and occupation. Since some occupations are unique to particular industries while others are common to many, industry projections are the key to understanding occupational employment projections. Short-term industry projection models explain how employment fluctuates with the demand for goods and services over the business cycle. Long-term models, by way of contrast, explain the impact of demographic shifts and changes in workplace technology on the demand for workers in various industries. Consequently, the employment outlook may differ over the two horizons.

Health services and construction are two key industries whose demand for labor is projected to differ markedly over the short and long term. Within health care, for example, the long-term demographic impetus of an aging population will expand health care services and increase demand for all health care workers. In the short-run, however, opportunities for some health care workers, most notably health paraprofessionals, will diminish as health care organizations merge and consolidate departments to cut costs. The job outlook in construction is another case in point. Over the short-run, low interest rates combined with increased spending on the infrastructure will likely increase demand for construction workers; whereas over the long term, demand will probably slow and settle closer to the trend line of the last two decades.

Providing Context with Descriptors

When analysts express projected hiring as a function of business cycle movement and structural change, they present users with a more comprehensive profile of labor demand. Very often, however, when analysts provide this information, they list the projected number of job opening in both the short- and long-term scenarios and inevitably force users to discern the magnitude of these differences. In interpreting the projected job openings for loan officers for example, users face the challenge of assessing the significance of 580 average annual job openings in the short term compared to 420 openings in the long term. Had analysts used descriptors to present the results, users would know that loan officers had a very favorable outlook in both the short and long term, a context that is more intuitive.

Certainly cases exist when the same occupation may be assigned a different descriptor in the two projection horizons. Such situations arise when the business cycle diverges from long-term trends. Job opportunities in education provide one case in point. In the long term, as the baby boomlet swells and increases enrollments, education-related occupations should have a very favorable outlook. In the near term, however, the outlook is more constrained, held back in part by fiscal pressures on state and local budgets. While this technical explanation may not be obvious to users, the shift in outlook from one time horizon to the other can be more readily understood.

The following paragraphs reveal the series of steps Illinois used to develop occupational descriptors covering two time periods, the 1997-1999 short-term projections and the1996-2006 long-term projections. The four steps are as follows: selecting the data series, testing the statistical properties, determining the number of descriptors, and building the descriptor table.

Step 1: Select the data series

To develop these descriptors we selected two data series, the projected rate of annual growth and employment size. The former measures the change in projected labor demand over time. Although this series biases against occupations with a large employment base--the impact of 1,000 new jobs is less in an occupation with 10,000 workers than in an occupation employing 100 workers--it still provides critical information on projected demand as above or below past demand.

The second data series--employment size--offsets the weakness embedded with the growth rate by juxtaposing the projected pace of labor demand with the level of demand. Database administrators, for example, have one of the fastest projected growth rates of all workers with jobs increasing on average 7 percent, but even with this rapid rate of growth they should gain on average only 140 new jobs per year. In contrast, jobs for teacher aides are expected to grow at less than one-half the rate of database administrators, yet the number of job openings for teacher aides is more than four times the number projected for database administrators. Basing descriptors on both these series balances the shortcomings of each individual series. To compare employment levels in both periods, we expressed employment size as a percentage of total employment.

Step 2: Test the statistical properties

Next, we evaluated the statistical properties of each data series, first checking for collinearity then for normality. Perfect collinearity exists when one variable is a linear combination of another variable. As previously mentioned, the projected annual growth rate and projected ten-year growth rate violate this condition. The latter is a linear combination, a multiple, of the former.

Collinearity of data series

In this Illinois case study, the projected rate and percentage of employment are not, a priori, a linear combination. However, some attendant association should be expected. All things being equal, large-sized occupations will tend to have lower growth rates than small-sized occupations. As expected, the Pearson correlation test for collinearity produced a small measure of association-- only .10 for the short term and .09 for the long term, well below the 1.0 found when there is perfect collinearity.

Distribution of data series

Next, we inspected the distribution of values within the percentage employed data series. Histograms based on both the short- and long-term series showed an unambiguous clustering at the lowest end of the percentage scale, which violates the conditions of normality (*see Charts 1 and 2 in the Appendix*). Similarly, the data points in the probability plots deviated noticeably from the line (*see Charts 3 and 4 in the Appendix*). Under the condition of normality the points are linear.

The statistical tests validated the graphs that the percentage of total employment is not normally distributed (*see Table 1*). The mean and median did not converge and both the skewness and kurtosis substantially exceeded the threshold of normality. The K-S significance level was less than .01. Therefore, we rejected the null hypothesis, which assumed the presence of normality for both the short- and long-term series.

Table 1: Percentage Employment		
	Short-Term	Long-Term
Mean	0.131	0.131
Median	0.035	0.035
Skewness	5.583	5.531
Kurtosis	39.668	38.884
K-S (p-value)	0.000	0.000

Although not absolutely conclusive, the preponderance of evidence supports the claim that projected annual growth rate approaches a normal distribution. The histograms and probability plots exhibited properties associated with normality (*see Charts 5-8 in the Appendix*). From the histograms of both series, we discerned a bell-shaped curve, and from the probability plots we observed that points deviated only marginally from the straight line. Moreover, the mean and median converged and the skewness was near zero (*see Table 2*). The kurtosis statistic, however, captured an anomaly, a strong central tendency among the observations toward the mean of the growth rate continuum. The extent of the clustering around the mean exceeded what is expected under the condition of normality. The K-S statistic was significant and supported the claim of an aberration in the series.

Table 2: Projected Annual Growth Rate		
	Short-Term	Long-Term
Mean	0.664	1.013
Median	0.846	0.992
Skewness	-0.748	0.308
Kurtosis	6.607	8.750
K-S (p-value)	0.000	0.000

We must now determine whether this clustering around the mean seriously violates the conditions of normality. To investigate the viability of this assertion and to discern whether the abnormalities are localized or pervade the series, we partitioned the long-term projected annual growth rate into deciles and reported the findings in each decile. The middle section—third through eighth deciles-- exhibited common characteristics that are distinct from those deciles representing the tails of the series (*see Table 3*). The mean and median converged and the skewness and kurtosis met the conditions of normality. Overall, the K-S statistic reported a significance level greater than .01 and supported the null hypothesis of a normal distribution.

In those deciles representing the tails of the series, however, these conditions were not fulfilled. Since these data abnormalities were localized and not pervasive, we continued our analysis under the assumption that the overall series did not seriously breech the conditions of normality. The short-term projected rate series produced similar results.

Table 3: Projected Annual Growth Rate (Deciles)					
	1st	2nd	3rd	4th	5th
Mean	-3.282	-.413	.205	.586	.885
Median	-2.569	-.370	.217	.584	.881
Skewness	-.731	-.437	-.149	.048	-.016
Kurtosis	- .652	-.995	-1.170	-1.224	-.764
K-S (p-value)	.000	.006	.017	.069	.182
	6th	7th	8th	9th	10th
Mean	1.152	1.506	1.943	2.605	4.902
Median	1.150	1.502	1.941	2.520	4.203
Skewness	.162	.035	.155	.370	3.518
Kurtosis	-.989	-1.401	-1.037	-1.331	18.123
K-S (p-value)	.034	.029	.200	.001	.000

Step 3: Determine the number of descriptors

Having analyzed the statistical properties for clues on how to segment the values of each data series into descriptor categories, we can now determine the number of categories. Based on the statistical measures appropriate to the properties of each data series, we delineated the empirically-defined upper and lower bounds of each category and, thus, created the rows (percentage of total employment) and columns (annual growth rate) of the descriptor table.

We segmented the growth rate series by direction, differentiating between negative and positive values. Next, we allocated the positive values among the slow, moderate and fast growth categories. Since this series was normally distributed, we established categories based on significant differences from the mean (*see Table 4*). Moderate growth included values that were within plus or minus one-half standard deviation of the mean. Fast growth exceeded the mean by more than one-half standard deviation, and slow growth fell below the mean by more than one-half standard deviation.

Table 4: Prototype Descriptor Table

	Projected Annual Growth Rate			
% Employment	Negative ($x \leq 0$)	Slow ($0 < x \leq \mu - .5\sigma$)	Moderate ($\mu - .5\sigma < x \leq \mu + .5\sigma$)	Fast ($x > \mu + .5\sigma$)
Small ($0 \leq x \leq 25\%$)	VUF	UF	UF	F
Medium ($25\% < x \leq 50\%$)	VUF	UF	F	VF
Large ($50\% < x \leq 75\%$)	VUF	UF	F	VF
Very Large ($x > 75\%$)	VUF	F	VF	VF

Since the employment percentage series showed a discernible clustering of observations among small-sized occupations, it failed to meet the conditions of normality. Due to this bias, we segmented the series into quartiles (*see Table 4*), a less rigorous descriptive statistic. The lower and upper twenty-five percent of observations fell to the small and very large categories. We evenly allocated the remaining cases to the medium and large categories.

Step 4: Build the descriptor table

The final task in the construction of the table is to decide on the descriptor labels and the allocation of labels to the table cells. By using the approach proposed in this report, we effectively reduced the number of designations from more than 750 (one for each occupation) to

16 (a 4x4 table comprised of four employment categories and four growth-rate categories). To display this information, we selected four descriptor labels: very unfavorable (VUF), unfavorable (UF), favorable (F), and very favorable (VF) (*see Table 4*).

Assigning descriptor labels. Regardless of employment size, we assigned a very unfavorable outlook to occupations with a negative growth. Most fast growing occupations, on the other hand, were rated as very favorable. The only exceptions were fast growing occupations in the small-size category, which received only a favorable rating. Even though these occupations had rapid growth, the small number of new jobs did not warrant a very favorable rating. The opposite reasoning explained our distribution of descriptors under the slow growth category (*see Table 4*). In this category, all occupations had an unfavorable designation except those employing a very large number of workers (> 75th percentile). These occupations, we rated as favorable.

The multiplicative effect of size and growth also explained the very favorable designation for very large occupations with moderate growth. The unfavorable designation for small-sized occupations with moderate growth rectified the misleading notion of viewing occupational projections using only data from the growth rate series. If this series were the only criterion, these occupations would reside near the top in any fastest growing list for a job search or training program. In reality, these occupations offer few opportunities because the employment base is small. The two remaining cells were favorable for moderate growth rate for medium-sized and large-sized occupations.

Analyzing Short- and Long-Term Descriptors

Since each data series had identical statistical properties in both the short-term and the long-term horizons, the prototype descriptor table applies equally to the short- and long-term data and, thus, will facilitate a comparison of the data across time horizons. Tables 5 and 6 show the count of occupations in each cell based on the data from the short- and long-term. Each row total was nearly identical in these Tables. This finding was not too surprising given that the total number of occupations is the same for each table, and we used quartiles to segment the employment percentage series.

Table 5: Illinois Descriptors for Short-Term Projections

	Projected Annual Growth Rate				
% Employment	Negative	Slow	Moderate	Fast	Row Total
Small	100 VUF	35 UF	39 UF	17 F	191
Medium	53 VUF	53 UF	59 F	26 VF	191
Large	34 VUF	56 UF	64 F	37 VF	191
Very Large	26 VUF	42 F	90 VF	32 VF	190
Column Total	213	186	252	112	763

Table 6: Illinois Descriptors for Long-Term Projections					
	Projected Annual Growth Rate				
% Employment	Negative	Slow	Moderate	Fast	Row Total
Small	73 VUF	54 UF	47 UF	17 F	191
Medium	39 VUF	59 UF	74 F	19 VF	191
Large	25 VUF	54 UF	72 F	39 VF	190
Very Large	19 VUF	37 F	98 VF	37 VF	191
Column Total	156	204	291	112	763

Differences in labor demand are reflected in each Table's column totals. That they differed is reflective of how labor demand was projected to diverge over the two projection periods. Total annual growth averaged 1.46% in the long-term scenario, compared to 1.18% in the short term. This difference was not evenly distributed across all categories. The biggest difference occurred in the moderate category, where 38.1% of the occupations in the long-term scenario are located and 33.0% percent in the short-term scenario. In the fast growing category, there was no difference: the percentage of occupations was identical, 14.7%.

A more relevant comparison is the count of occupations by descriptor label for each projected period. Table 7 lists the total number of occupations within each category for both the short and long term. Although the proportions were not that far off in the favorable category: 24.2 % (short-term) vs. 25.3% (long-term), a more detailed look reveals that the occupations comprising these cells are different. Twenty-three occupations that had a very favorable outlook in the long term had only a favorable, or worse, outlook in the short-term run. That most of these occupations were health-related results from the two different time horizons projected for this important industry.

Table 7: Occupation Counts by Descriptor				
	Short-Term		Long-Term	
	Count	Percent	Count	Percent
Very Favorable	185	24.2	193	25.3
Favorable	182	23.8	200	26.2
Unfavorable	183	23.9	214	28.0
Very Unfavorable	213	27.9	156	20.4
Total	763	100.0	763	100.0

The long-term projections mirrored the aging of the population. That is, demand for all health care workers would increase rapidly as the population ages and increases the demand for additional health services. In the short-run however, as more and more health organizations merge and consolidate departments to cut costs, some health care workers were projected to have more pessimistic outlooks. The hiring expectation for paraprofessionals, for example, was much more pessimistic than that for professionals. Consequently, highly-paid professionals, such as doctors and nurses who provide key services, had a very favorable outlook in both the two- and ten-year scenarios while some technician occupations, such as laboratory and radiology, had a more pessimistic outlook in the short term than in the long term.

Case Study: A Comparison of National and State Descriptor Methodology

As a follow up to Minnesota's earlier study, this case study compares the accuracy of descriptors derived from the national methodology to those obtained from the state methodology as outlined in the previous chapter. Recall that Minnesota researchers found large discrepancies between the projected and actual descriptor designations when they applied the national methodology, i.e. the growth rate data series and six descriptors, to their 1986-1993 data. More specifically, Minnesota found that only 13 percent of the 448 occupations evaluated had identical descriptor labels based on projected and actual employment. Users desire, as a priority, the presentation of occupational projections that extend beyond the traditional numeric format. The challenge to LMI producers, therefore, is to select the most appropriate descriptor methodology given the constraint of data limitations.

The national descriptor methodology employs six descriptors, ranging from much faster to decline, derived from a single data series (growth rate). These two features restrict the methodology's suitability for presentation of state and local occupational projections. The six descriptors capture negative annual growth, no or little growth, and four gradations of annual growth. Most producers of labor market information concur that state and local occupational employment estimates are less statistically robust than national estimates due to less reliable sample estimation cells. This deficiency warrants fewer descriptor categories.

Moreover, the disadvantage of growth rate as the single data series for the estimation of descriptors is well known: a bias toward occupations with a smaller base employment. In fact, the Minnesota analysis reports an inverse correlation between the employment size of an occupation and the error in its projected growth rate. The error was greatest among the smallest occupations and diminished among the medium- and large-sized occupations. This finding encourages a consideration of multiple data series for inclusion in the descriptor methodology.

In the following sections, we re-analyze the Minnesota data using the construct proposed for state-level data, four descriptor labels (very unfavorable, unfavorable, favorable, very favorable) and two occupational employment series, growth rate and percentage of total employment. We explain how we implemented the state descriptor methodology, examine how well this descriptor methodology handled both projected and actual 1986-93 employment data, and compare these results to those obtained from the national methodology (single data series/six descriptors).

Re-estimating Minnesota descriptors using state methodology

The steps for implementing the proposed state descriptor methodology--four descriptor labels (very unfavorable, unfavorable, favorable, and very favorable) and two employment series (the growth rate and size of employment)--remain the same: examine each employment series statistical properties, determine the number of descriptors and build the table. Test results on each of Minnesota's data series were consistent with those obtained from Illinois. Only a modicum of collinearity existed between the 1986 percentage of total employment and either the 1993 projected annual growth rate (Pearson correlation .17) or the 1993 actual annual growth rate (Pearson correlation -.01). Therefore a descriptor table based on percentage of employment and either the 1993 projected or actual growth rate does not present redundant and potentially misleading information.

Test results for normality were also consistent with those obtained by Illinois. Like Illinois, Minnesota's percentage of total employment was not normally distributed as evidenced by a distinct skewness toward small-sized occupations. As a result, we segmented this data series using quartiles: 1. small-sized occupations (less than .031% of total employment); 2. medium-sized occupations (.031% to .079%); 3. large-sized occupations (.080% to .20%); and, 4. very large-sized occupations (greater than .20%).

As in Illinois, both growth rate series (1993 projected and 1993 actual) approximated a normal distribution. Therefore, we used standardized statistics to create each descriptor category's empirical upper and lower bound, i.e., numerical cut-off level. Specifically, we defined the negative rate category to include occupations with zero or less than zero growth. The mean and standard deviation defined the empirical boundaries of the slow, moderate, and fast categories. They are as follows: 1. in the projected growth rate series, the mean of the positive values was 1.9% and one-half of the standard deviation was .9%; and, 2. in the actual growth rate series, the mean of the positive values was 7.6% and one-half of the standard deviation was 4.4%.

Comparing projected and actual descriptors using the state methodology

Cross tabulations of the 1986 percentage employed and the 1993 projected rate (*see Table 8*) and the 1986 percentage employed and the 1993 actual rate (*see Table 9*) provide important information on the distribution of occupations across the aforementioned size and rate categories. Since both tables used the 1986 percentage employment data series, the total number of occupations in each size category is the same, i.e., each row contained a total of 112 occupations (*see Table 8 and Table 9*). Because the projected and actual growth rate differed, the column totals varied widely.

Table 8: Minnesota Descriptors for Projected Employment					
	Projected Annual Growth Rate				
% Employment	Negative	Slow	Moderate	Fast	Row Total
Small	29 VUF	19 UF	53 UF	11 F	112
Medium	6 VUF	30 UF	59 F	17 VF	112
Large	9 VUF	17 UF	67 F	19 VF	112
Very Large	8 VUF	15 F	79 VF	10 VF	112
Column Total	52	81	258	57	448

Table 9: Minnesota Descriptors for Actual Employment					
	Actual Annual Growth Rate				
% Employment	Negative	Slow	Moderate	Fast	Row Total
Small	58 VUF	13 UF	20 UF	21 F	112
Medium	54 VUF	14 UF	30 F	14 VF	112
Large	40 VUF	34 UF	29 F	9 VF	112
Very Large	44 VUF	30 F	32 VF	6 VF	112
Column Total	196	91	111	50	448

While the column totals in the slow and fast categories were fairly close in both tables (81 vs. 91 occupations and 57 vs. 50 occupations, respectively), the counts differed substantially in the negative and moderate categories (*see Table 8 and Table 9*). In Table 8, only 52 occupations were projected to decline while the actual number was 196 (Table 9). The gross underestimation of occupations in this category had multiple sources of error, including both the industry projection models and the occupational staffing patterns, but the implication for the assignment of descriptors is pronounced.

In the moderate-growth rate category, the discrepancy was also large, 258 projected (*see Table 8*) to 111 actual (*see Table 9*). Although the total number of occupations in the fastest category was similar in both tables (57 and 50 occupations), this category represented approximately twenty percent (19.8%) of all positive growth occupations in the actual series and less than fifteen percent (14.4%) in the projected series. Not surprisingly, the mean growth rate among occupations with positive growth was significantly higher in the actual series compared to the projected series, 7.6% to 1.9%.

Table 10 lists the total number of occupations that fell within each category under both scenarios. The row designations show the distribution of projected descriptors while the columns list the results for the actual descriptors. As expected, the largest discrepancy occurred for the very unfavorable (VUF) outlook (*see Table 10*).

Minnesota projected only 52 occupations (11.6% of the total) to have a very unfavorable (VUF) outlook. In reality almost four times as many occupations (196 or 43.8% of the total) should have received this designation (*see Table 10*). In brief, Minnesota over-projected employment growth. This optimism was also apparent in the very favorable (VF) category. Only 13.6% of Minnesota's occupations had met the criteria for very favorable, less than one-half the percentage projected (28%) (*see Table 10*).

Table 10: Comparison of Minnesota Projected/Actual Descriptors					
	Actual Descriptors				
Projected Descriptors	Very Favorable	Favorable	Unfavorable	Very Unfavorable	Row Total
Very Favorable Row % Column %	41 32.80% 67.21%	33 26.40% 30.00%	6 4.80% 7.41%	45 36.00% 22.96%	125 27.90%
Favorable Row % Column %	13 8.55% 21.31%	49 32.24% 44.55%	30 19.74% 37.04%	60 39.47% 30.61%	152 33.93%
Unfavorable Row % Column %	5 4.20% 8.20%	20 16.81% 18.18%	35 29.41% 43.21%	59 49.58% 30.10%	119 26.56%
Very Unfavorable Row % Column %	2 3.85% 3.28%	8 15.38% 7.27%	10 19.23% 12.35%	32 61.54% 16.33%	52 11.61%
Column Total Column %	61 13.62%	110 24.55%	81 18.08%	196 43.75%	448

Minnesota obtained better results in the favorable (F) and unfavorable (UF) categories. Using actual data, 110 occupations (24%) had a favorable designation compared to a projection of 152 occupations or 34%. For the unfavorable category, the results were 18% and 26%, respectively (*see Table 10*).

Comparing growth descriptors using state and national methodologies

Measuring Accuracy: on-target. How do these results compare to those derived from the national methodology? Using the national methodology (growth rate and six descriptors), only 13 percent of the occupations (58 occupations) had identical projected and actual descriptors, i.e., fell within the same range of percent growth. The on-target accuracy within the much faster growth and average growth descriptors was even lower. For example, Minnesota projected 13 occupations to grow at a much-faster rate while approximately 135 occupations actually fell into that category. In contrast, Minnesota projected 123 occupations to have average growth, but only 38 met that expectation.

Using the state descriptor methodology, Minnesota's on-target accuracy rate improved, increasing to 35 percent. In fact, approximately 30 percent of the occupations in the very favorable, favorable, and unfavorable categories had identical descriptors. In the very unfavorable category the matched percentage jumped to 60 percent.

Measuring Accuracy: good advice. In evaluating its growth descriptors BLS developed a measure of "good advice" accuracy. The logic of this measure is that the projected descriptor retains some validity even in instances where the actual descriptor, although not identical to the projected descriptor, is only one category removed. For example, a descriptor assigned to an occupation that was projected to grow at a "faster" rate still provides accurate information even though its actual rate growth was "average." The assumption underlying the measurement of "good advice" is that reasonable proximity does not grievously misdirect a user's job choice.

The state descriptor methodology employs a more rigorous measurement of one-step removed or "good advice." The delineation between favorable and unfavorable designations is a fault line. The mixing of occupations in these two descriptor categories, although one-step removed, constitutes misleading information for users in a way, and to a degree, that is not true for combining very favorable with favorable or unfavorable with very unfavorable. Even using this more conservative definition, 61% of all the Minnesota occupations met the requirement of "good advice."

Summary. These comparisons reveal that the state methodology provided far more accurate descriptors of Minnesota's job growth than those derived from the national methodology. Had Minnesota used the two series and four-descriptor approach proposed in this report, its "on-target" accuracy would have increased from 13 to 35 percent. Similarly, the percentage of occupational descriptors providing "good advice" would have increased from 35 to more than 60 percent. The inherent deficiencies of a single-series, six-category approach are magnified at the state level and may overwhelm any information gains from the presentation of subnational occupational employment projections. Under the Workforce Investment Act, state and local labor market information assumes a prominent role in the LMI infrastructure. Customers depend on information developed within this infrastructure to guide critical decisions, such as career decisions, job searches and curriculum development plans. It is incumbent upon producers within this infrastructure to employ sound statistical methodologies as a basis for communicating reliable information as input to user decisionmaking.

References

Barylo, Lusiena and Dave Senf (1998). Minnesota Occupation Employment Projections: An Evaluation. Report to ALMIS Long-Term Projections Consortium. U.S. Department of Labor: Washington, DC.

Blalock, Hubert M. (1972). Social Statistics (2nd ed.). New York: McGraw-Hill Book Company.

Keith, Brynn, JoAnn Wilson, and Kathy Ermatinger (1998). Occupational Information Users' Survey. Report to ALMIS Short-Term Forecasts Consortium. U.S. Department of Labor: Washington, DC.

Pindyck, Robert S. and Daniel L. Rubinfield. (1991). Econometric Models and Economic Forecasts (3rd ed.). New York: McGraw-Hill Book Company.

Veneri., Carolyn M (1997). "The 1995 Employment Projections: How Accurate Were They?" Occupational Outlook Quarterly (Fall): 35-52.

Wettemann, Martha (1997). Using Descriptors to Improve the Analysis of Labor Market Data. Report to ALMIS Short-Term Forecasts Consortium. U.S. Department of Labor: Washington, DC.

APPENDIX

Chart 1: Short-Term % Employment

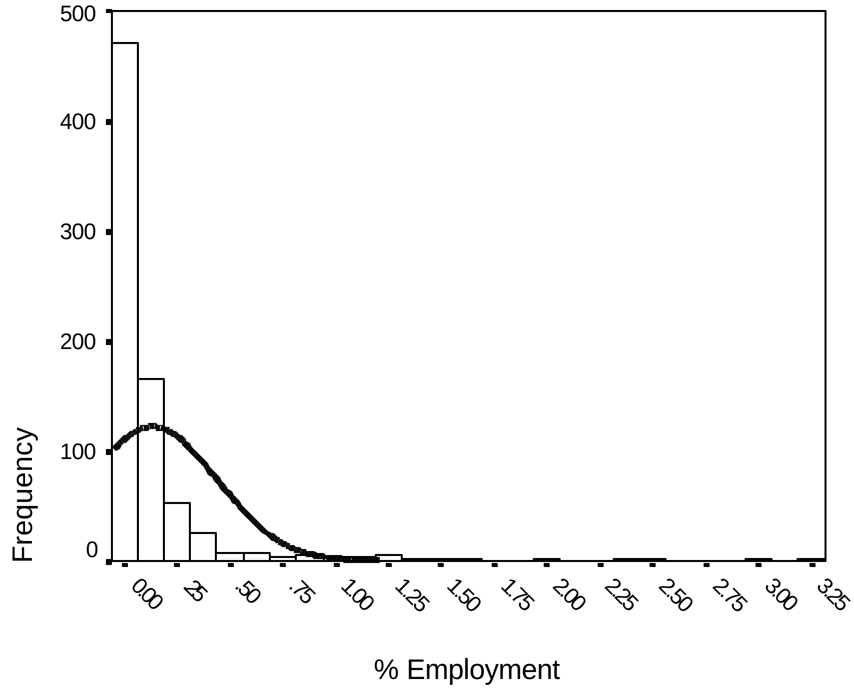

Chart 2: Long-Term % Employment

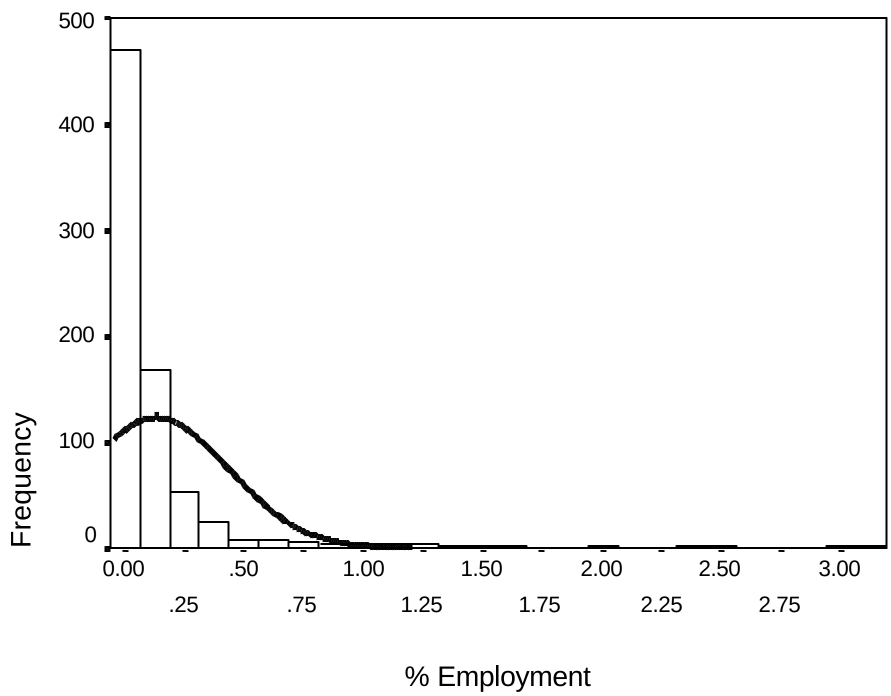

Chart 3: Short-Term % Employment

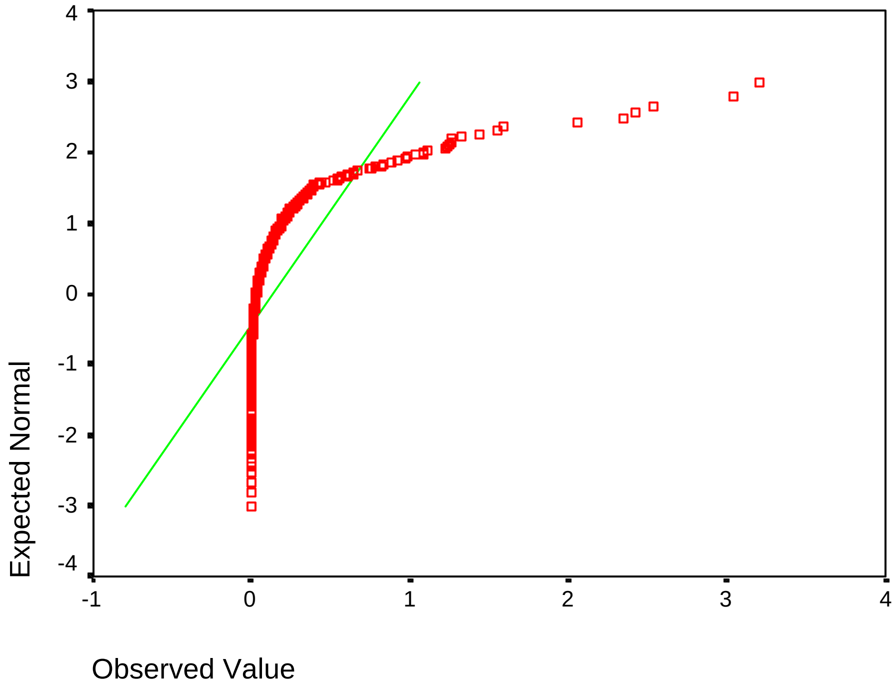

Chart 4: Long-Term % Employment

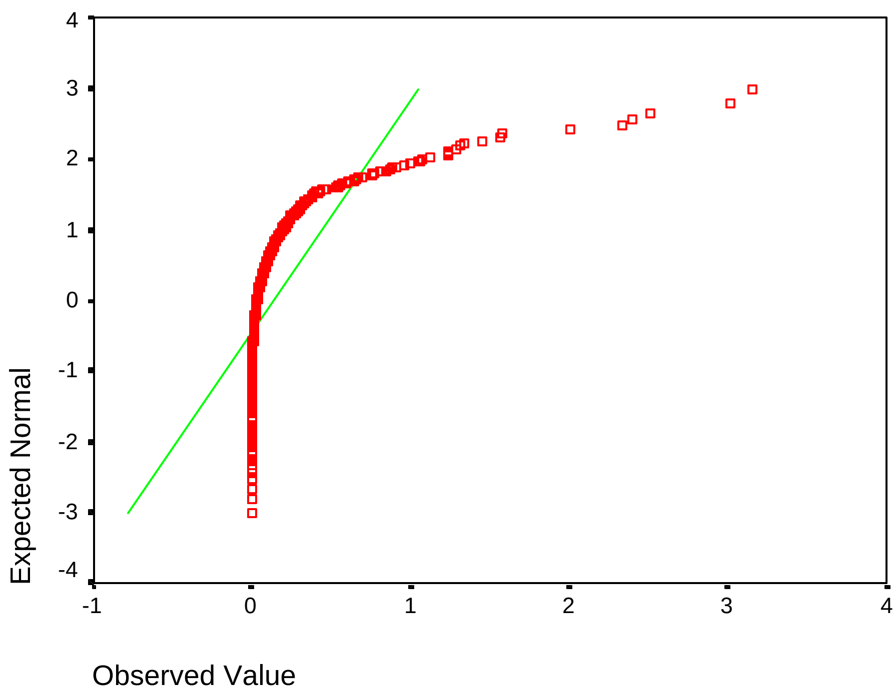

Chart 5: Short-Term Projected Annual Growth

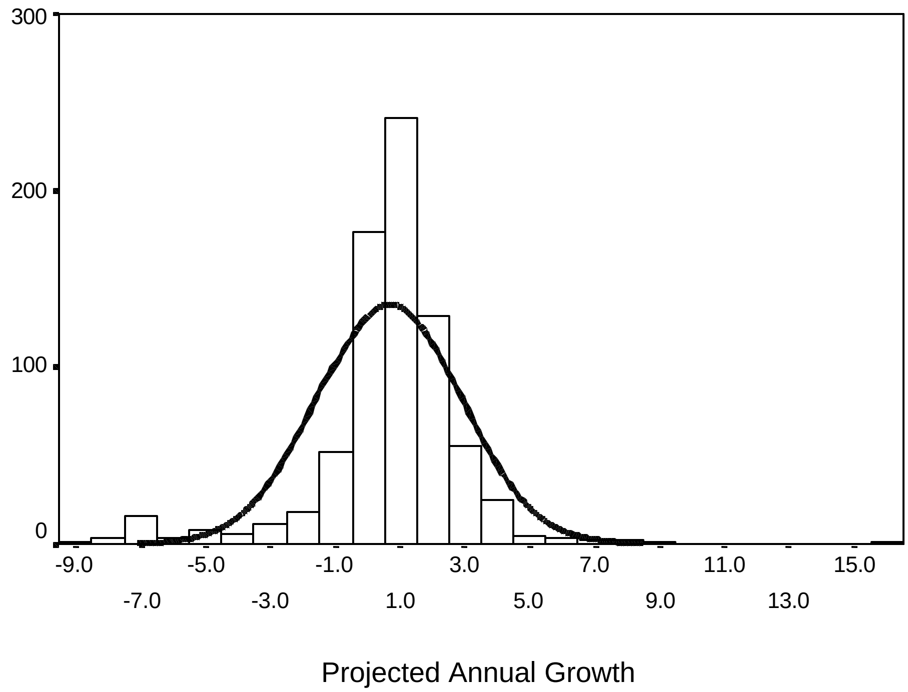

Chart 6: Long-Term Projected Annual Growth

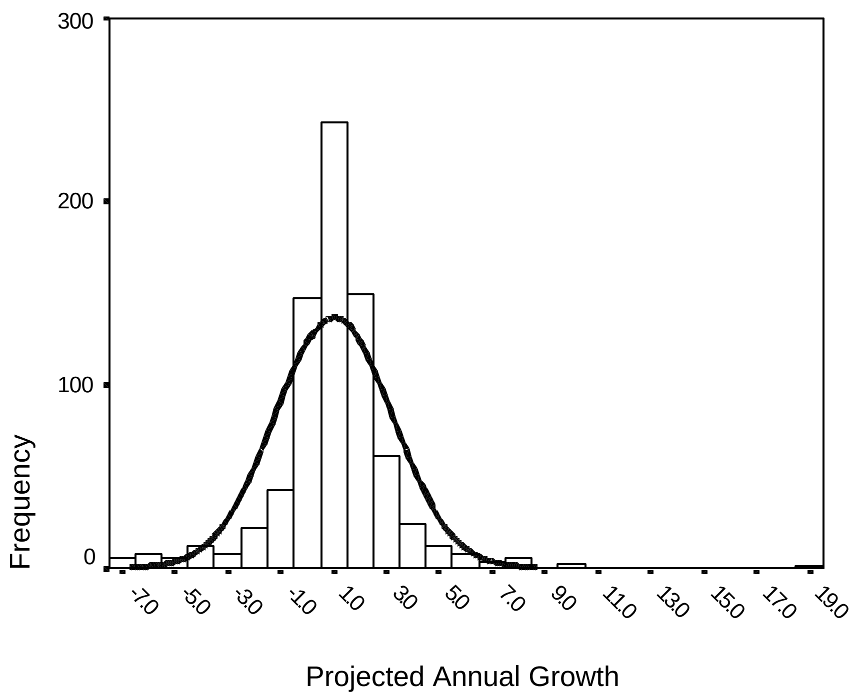

Chart 7: Short-Term Projected Annual Growth

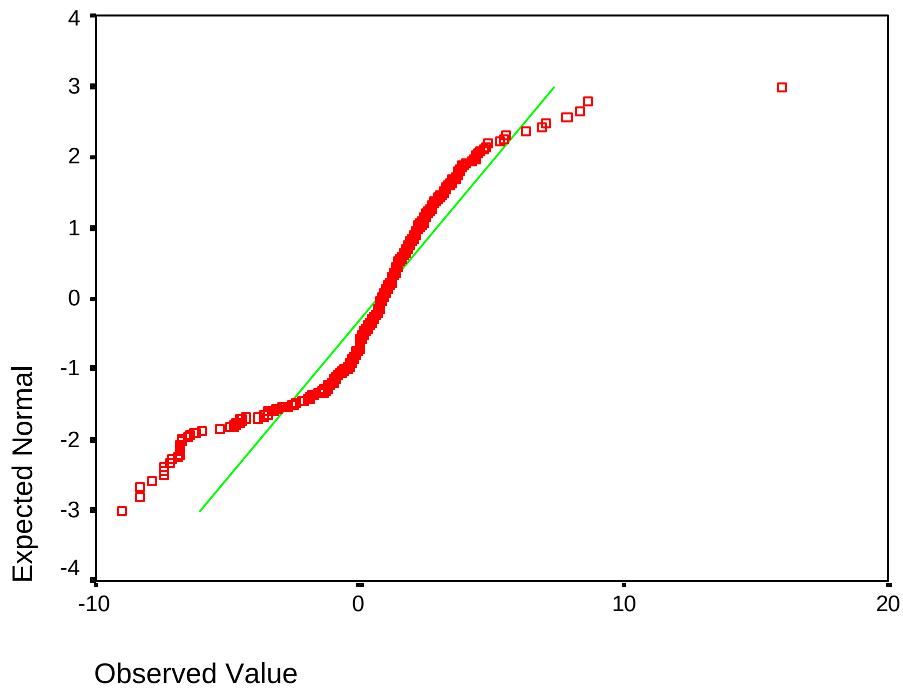

Chart 8: Long-Term Projected Annual Growth