Improving Economic Forecasting in Minnesota

by Vladimir Cherkassky and Filip Mulier

The decisions of the people and institutions that make up our economy rely upon the accurate analysis of economic data and forecasting of future economic activity. Yet the effective use of economic data is often impeded by a lack of flexible tools for data analysis. Standard statistical methods often rely upon rigid assumptions about the relationships and dependencies among economic variables, and it is assumed that changes in certain variables will consistently correspond with changes in other variables. Unfortunately, the interdependencies among economic variables are often inconsistent or unknown and standard methods may produce very inaccurate results. This report describes the use of new nonlinear modeling techniques to better forecast economic behavior in Minnesota where more traditional modeling methods have proved ineffective.

The specific problems discussed in this report are those faced by two of Minnesota's state economic agencies. First, the Economic Analysis Office of the Minnesota Department of Trade and Economic Development was interested in identifying relationships between investments in education and the productivity of manufacturing. Second, the Research and Statistics Office of the Minnesota Department of Economic Security wanted better predictions for employment in several industries. Existing predictions, based on standard linear-regression models, had seemed unreasonable in light of national predictions for the same industries.

In both cases, we applied the new nonlinear modeling techniques to economic data supplied by the agencies to see, in the first case, if meaningful relationships could be detected between manufacturing productivity and education indices, and in the second, if nonlinear modeling could provide better estimates of future employment in certain Minnesota industries. Following a brief discussion of economic data and the distinction between linear and nonlinear dependencies and modeling techniques, we will present our findings and some of the policy implications they suggest for economic planning in Minnesota.

One area where manufacturing productivity is especially high is in firms that use unskilled workers and machines that are responsible for most of the production.

Economic Data and Forecasting
The goal of economic forecasting is to develop a model useful for predicting economic behavior in the future, based on available historical data. This is an inherently difficult problem, made more difficult by certain characteristics of economic data.

First, the data samples available for economic modeling are often small. For example, an analysis of economic data for Minnesota counties is limited to eighty-seven data sets, and as some counties do not record economic data, the number of data sets is usually smaller. And in contrast to the experimental sciences, where data samples can be generated in a controlled environment, economic data cannot be generated at will. Economic data must be taken as is reported.

Second, economic data is very noisy—that is, it is often poorly defined and subject to measurement error. For example, data on personal income will not include unreported cash payments. Moreover, many economic variables are estimated by sampling techniques rather than measured directly. Large samples typically cannot be obtained due to cost constraints, and small
samples are inherently less reliable. Finally, most economic indices change with time due to political, technological, or structural changes affecting the economy. Hence, economic forecasting typically uses only recent, more reliable observations. This can make forecasting based on time series models impractical because changes in an input variable over time might alter or negate its ability to predict the output variable.

Linear and Nonlinear Relationships
When we refer to dependencies between economic variables, we are describing a relationship in which changes in certain input or predictor variables correspond to changes in other output or response variables. Such dependencies are typically expressed as an equation, where the value of the "y" output variable is a function of the "x" input variables. For example, an unknown dependency between Minnesota employment in the general building contractors industry (the y output variable) and two x input variables—the anticipated interest rate (x₁) and the anticipated increase in Minnesota population (x₂)—would be mathematically described as

\[ y = \text{function}(x₁, x₂) + \text{error} \]

where error denotes error due to unknown factors other than x₁ or x₂ that could possibly affect y.

We can try to estimate this unknown dependency using known data—that is, the values of y, x₁, and x₂ in past years. Past values of these variables and apparent dependencies among them enable us to assume future values and dependencies. Depending on the nature of our assumptions, we can use one of two types of model—linear or nonlinear—to forecast future relationships between these variables.

A prime example of a linear model is linear regression. In a perfect linear regression, a change in an x input variable yields a proportionate change in the y output variable. Unfortunately, if what appears to be a linear dependency proves to be nonlinear, a linear model will be highly inaccurate in predicting future economic behavior. Such models, in other words, work well only when the underlying assumption of a linear dependency is strong. In addition, dependencies among variables can change, or may prove over a period of time to be cyclical or aperiodic rather than linear. The flexibility of nonlinear methods—methods that avoid rigid assumptions about underlying dependencies among variables—enables us to model nonlinear dependencies more accurately.

The two nonlinear methods used in this study were, first, Multivariate Adaptive Regression Splines (MARS), a method derived from statistics, and second, Constrained Topological Mapping (CTM), a method derived from neural network research. Both the MARS and CTM methods rely on partitioning—that is, dividing the domain of x input variables into separate regions, and estimating responses to these variables separately for each region. The main difference between MARS and CTM is the way in which the domain of input variables is partitioned.

Manufacturing Productivity and Work Force Education
Minnesota’s Economic Analysis Office (part of the state’s Department of Trade and Economic Development) develops strategies to promote economic growth in Minnesota, and disseminates information on how such growth might be pursued. The office had a data set of county-level economic indicators relating to educational investment and manufacturing productivity for 1990, and wanted to identify possible meaningful
Figure 1. Minnesota Manufacturing Productivity as a Function of Labor Cost and Proportion of Professional/Technical Labor (analysis by counties)

relationships between these indices. The response variable ($y$) was value added in manufacturing per production hour (a quantitative measure of manufacturing productivity), and the eight possible predictor variables included such factors as the proportion of the work force with a post-secondary education, per capita tuition paid by post-secondary students, the proportion of the work force which could be classified as professional or technical, and the labor cost in manufacturing per hour. Although the original data set included eighty-seven samples (one for each county in the state), missing data for many counties limited our modeling of the response variable to fifty-four counties. The data came as county level summaries, not data from individual firms within each county.

We modeled the response ($y$) to the eight input variables using three different methods: one linear (multiple linear regression) and two nonlinear (MARS and CTM). Our goal was first to estimate which of the three models had the smallest prediction error. The prediction error is estimated through resampling.

We found that there was no pronounced difference between the linear and nonlinear models, but the nonlinear model based on the CTM method had the lowest estimated prediction error. The CTM model showed a dependency between $y$ and two predictor variables—$x_1$ (percent of labor force in professional/technical work) and $x_2$ (labor cost in manufacturing per hour). This dependency can be charted on a three-dimensional graph (Figure 1). We see that as labor cost per hour increases along the first horizontal axis, value added in manufacturing per hour also increases along the vertical axis, especially after labor costs increase above a certain threshold. Along the second horizontal axis, we see a slightly different pattern. Value added in manufacturing initially decreases as the percent of the labor force in professional/technical work increases, and then begins to increase after this proportion passes a threshold between 35 and 40 percent.

Combining the effect of the two input variables on the response variable of value added in manufacturing per hour, we see two areas where the response to both input variables is particularly high: first, where the percent of the professional/technical work force is low and labor cost in manufacturing is high, and second, where the percent of the professional/technical work force is high and labor cost in manufacturing is high (Figure 1). The first of these areas corresponds to industries using unskilled workers in a factory setting where machines are responsible for most of the production. The second area corresponds to industries where production depends on the skill and education of the work force.

The Economic Analysis Office sees several policy implications in these results. First, it appears that increased wage rates above a certain threshold accompany increased productivity, while wage rates paid below this threshold do not increase productivity. This may be because laborers commanding wages below the threshold do not possess useful skills for production (an education and training policy issue) or are not able to locate positions where they can use their skills productively (a labor mobility and information policy issue). On the other hand, firms may be unwilling to pay wages above the threshold level (a market information failure).

Second, it appears that firms employing high levels of professional and technical personnel are highly productive, as are firms at the other end of the scale employing low levels of such skilled personnel. Firms employing more balanced proportions of technical and nontechnical personnel are less productive. These results suggest two strategies for improving productivity:

- Firms that choose to employ low levels of professional and technical labor should invest heavily in automated capital inputs.
- Firms employing high levels of professional and technical labor need not invest as heavily in automated capital inputs, but may instead rely on their investments in intellectual capital to increase productivity.

It is interesting to note that most changes in productivity are gained by adjusting the proportion of professional and technical labor in the production process. It appears that firms that do not choose to move toward one or the other of these production strategies incur increased labor costs with little increased productivity.

Forecasting Minnesota Employment

The Research and Statistics Office of the Minnesota Department of Economic Security projects future employment demand for various industries. Specifically, the office tries to forecast significantly growing and declining industries, to compare state and national trends, and to identify major deviations from historic trends. This information is used by educational program planners, employment counselors, public policymakers, businesses, job seekers, and the news media.

The Research and Statistics Office's projections have typically been based on linear modeling methods following the guidelines of various national agencies (for example, the Bureau of Labor Statistics) and they use national projections of economic indices such as employment, interest rates, and population growth, as well as projected state economic indices. The office found that standard linear modeling approaches were not producing reasonable projections for two sectors of the state's economy: the general building contractors sector and the electronic equipment manufacturing sector. Using nonlinear modeling methods, we were able to successfully analyze data sets for these two sectors.

The Minnesota Department of Economic Security supplied data sets for both sectors, including data on employment and various economic indicators for the past twenty years and the next decade.
years (1972-1992). Estimates of variables were given for 1993 to 2002. The goal in each case was to predict future employment levels (the response variable) based on past observations and on future estimates for the predictor variables. Our first step was to identify the most important predictor variables—that is, those variables having the closest correlation to the response variable. Once significant dependencies between predictor and response variables were established, we forecast future employment based on the projected estimates of the predictor variables available from state and national agencies.

Employment for General Building Contractors. The response variable we sought to predict in this instance was employment in Minnesota’s general building contractors industry. The predictor variables we examined were: national employment in sectors that include the general contractors industry, the U.S. conventional mortgage rate, the total Minnesota population, Minnesota’s population of twenty-five through forty-four year olds, Minnesota’s population of forty-five through sixty-four year olds, the ratio of Minnesota’s total population to the United States’ total population, and U.S. construction employment. Our analysis determined that the most important of these predictor variables were the total Minnesota population and the U.S. conventional mortgage rate. Specifically, the greatest correlation was observed with the anticipated interest rate two years ahead of any given year and with the anticipated change in the Minnesota population for the year following any given year.

The best model, then, for predicting employment among general building contractors used these two variables—the anticipated interest rate in two years and the anticipated change in Minnesota’s population for the next year—in a CTM-based nonlinear model. Figure 2 charts the results of using both the linear modeling approach and the CTM-based nonlinear approach in predicting employment for general building contractors. The CTM-based model more closely corresponds to the true values of employment for general building contractors in 1972-1992 than does the linear model. It also projects a higher value for employment among building contractors in future years.

Looking more specifically at the modeled relationship between the two predictor variables and the response variable on a three-dimensional graph (Figure 3), we see along the first horizontal axis that increasing growth in Minnesota’s population generally corresponds to increasing construction employment along the vertical axis, as might be expected. Less intuitively, however, rising interest rates along the second horizontal axis also correspond to increasing construction employment along the vertical axis. Common sense knowledge would predict the opposite result—that construction activity would slow with rising interest rates. The results shown on the model may have been skewed by a few decisions to start large construction projects immediately, based on the expectation of rising interest rates. It is also possible that this model may be unreliable, being derived from too few data points.

Employment in Electronic Equipment Manufacturing. As with our analysis of general building contractor employment, we sought to predict employment in the electronic equipment manufacturing sector. Here, employment might be dependent on such predictor variables as Minnesota’s total personal income, national and Minnesota unemployment rates, national and Minnesota employment in computer and office equipment manufacturing, and national or Minnesota employment in industrial machinery manufacturing. In this instance, there were many high correlations between pre-

Figure 3. Minnesota Employment of General Building Contractors as a Function of Anticipated Mortgage Rates and Population Growth (analysis for entire state)
dictor variables and the response variable, but the highest correlation (for which there was also a projected value available) was the state’s total personal income in 1982 dollars. The high correlation has a simple interpretation: personal income drives spending on electronic goods, which in turn affects employment in the electronic equipment manufacturing industry.

The high correlation between Minnesota’s total personal income and employment in the manufacturing of electronic equipment suggests that standard linear regression modeling should also work well in this situation. However, we applied nonlinear modeling, again using the CTM method, to achieve still better prediction accuracy (Figure 4).

And looking specifically at employment in this sector as a function of Minnesota’s total personal income, we gain some interesting insights (Figure 5). The general trend has already been noted—more income drives more spending on electronic goods, which in turn increases employment in the electronic industry. But there are also two distinct regions of dependency, corresponding to lower and higher total personal income in the state, with the break-point occurring between $55 and $60 billion on the horizontal axis. At this point, employment in the electronic equipment manufacturing sector, charted on the vertical axis, sharply increases. This suggests a possible qualitative difference in the type of electronic goods purchased by consumers at different income levels, with more expensive goods, such as computers, being purchased only when personal income has reached a certain level.

The Minnesota Department of Economic Security is constantly reviewing methods of forecasting employment for the state so that the forecasts can be as accurate as possible. They liked the improved accuracy resulting from the use of nonlinear modeling. Though the application of MARS and CTM techniques is fairly new to the analysis of economic data, they saw that these techniques deserve further review. A large part of the improvement in our analysis of employment predictions for building contractors and electronic equipment manufacturing came from the process of selecting meaningful predictor variables before the nonlinear modeling methods were applied. In each instance we selected one or two variables, as compared with the total of ten or sixteen used in predictions by the Economic Security Office. This selection process is an important component of the nonlinear modeling technique.

The use of nonlinear modeling to help predict future employment in Minnesota industries has implications for businesses, educational institutions, and individuals. Such projections may affect a business’s decision to open, expand, or close. They suggest how resources for vocational and technical education programs and internship programs should be allocated. And they may influence an individual’s decision to pursue a particular career. To the extent that nonlinear modeling can provide more accurate predictions of employment trends, these decisions can be made more intelligently.

Vladimir Cherkassky is associate professor of electrical and computer engineering at the University of Minnesota. His current research is on theory...
and methods for learning dependencies from data. He was director of the NATO Advanced Study Institute (ASI) From Statistics to Neural Networks: Theory and Pattern Recognition Applications held in France in 1993 and has been a member of the program committee at several major conferences on neural networks. He has taught several short courses and tutorials on neural network and statistical methods for industry and academia.

Filip Mulev was a doctoral candidate in electrical engineering at the University of Minnesota at the time he worked on this study with Cerkassky. He received his degree in 1994 and presently works at a large multinational corporation based in St. Paul in the area of neural networks and statistical data analysis. He has written a book (co-authored with Cerkassky) Learning From Data: Concepts, Theory and Method that will be published by Wiley in 1998.

This study was supported by an interactive research grant from CURA and the Office of the Vice President for Research, University of Minnesota. Interactive research grants have been created to encourage University faculty to carry out research projects that involve significant issues of public policy for the state and that include interaction with community groups, agencies, or organizations in Minnesota. These grants are available to regular faculty members at the University of Minnesota and are awarded annually on a competitive basis.

Credits:
Cover photo and photos on pages 3, 5, and 11 by Robert Friedman.
Photos on pages 8, 9, and 12 by Nancy Conroy.
Maps and figures prepared for the authors by the Cartography Laboratory, Department of Geography, University of Minnesota.