PROFILING WORKERS FOR UNEMPLOYMENT INSURANCE: VARIABLE SELECTION IN THE KENTUCKY MODEL

Mark C. Berger, Dan A. Black and Amitabh Chandra
Department of Economics and
Center for Business and Economic Research
Gatton College of Business and Economics
University of Kentucky

Abstract

We discuss the choice of variables used in Kentucky's model for worker profiling. These variables covered several categories, including: monetary, economic, industry, occupation, job-characteristics and prior spells of unemployment. We also derived industry-specific variables to reflect one year and two year changing economic patterns, using commuting data to weigh the significance of these changes in other counties for each county. After removing a 10% sample for validation study, we estimated several models, designing a profiling system for each model. The inclusion of regional economic variables significantly improved the explanatory power of our model.

Introduction

In June 1994, through a proposal initiated by the Kentucky Department for Employment Services (DES), the Commonwealth of Kentucky was selected as a prototype state for the development of a system for profiling unemployment insurance claimants and providing them with reemployment services. The Center for Business and Economic Research (CBER) at the University of Kentucky developed a profiling system that generated a single score for each claimant that measured the probability of exhausting his or her benefits. In this report we present our research methodology in developing that model and focus on the process of variable selection for the Kentucky model.

This report will discuss methodological issues in variable selection for the Kentucky model at three levels: First, we discuss our motivations and criteria for variable selection. We then list the variables that were included in the final Kentucky model and comment on their explanatory power. Finally, we outline our thoughts and recommendations for future research in this area.

Criteria for Variable Selection

Because Kentucky was a prototype state for the development of a profiling system, we were unable to incorporate the experiences of other states in the development of our model. We therefore relied on a rigorous survey of the existing academic literature on unemployment spells to identify a core group of variables that could be incorporated into our model. To capture the extent to which there is regional variation in economic activity we placed special emphasis on

including variables that could discern such differences. While estimating our models we found that the inclusion of such variables significantly increased the explanatory power of our models.

The methodology in estimating our model was as follows. First, we divided our data by Area Development District (ADD) aggregates. The 15 ADDs divide the state into regions which comprise similar counties in terms of type of industry, demographics, and geography. Initially, we were inclined to estimate 15 separate models—one for each ADD. A closer examination of population levels and industry structure in each ADD, however, discouraged us from using this method. Several ADDs contained too few UI claimants (leading to potential problems with cell size), while others had economic conditions that were similar to adjacent ADDs. After close examination of ADDs, we combined them into eight groups after comparing population, employment earnings, income, industrial structure, welfare payments, UI benefits, and the growth rates of these variables across ADDs. We then constructed variables from a number of sources that are discussed below. After removing a ten percent sample for validation study, we estimated several models, designing a profiling system for each model. To chose between competing models, we compared correlation coefficients between the predicted percentage of benefits that were exhausted and the actual percentage of UI benefits that were exhausted. For the double-limit tobit model that was finally chosen, we correctly identified 81.57% of claimants in the ten percent validation sample who actually exhausted their UI benefits.

Selection of Variables

On the basis of a literature survey, we identified the following variables as being most significant in the previous literature: Weeks of full and regular benefits, high and base quarter

earnings, industry, schooling, number of dependents, tenure, plant closure, residence in a MSA, state unemployment rate, initial length of benefits, UI replacement rate, UI benefits and duration of potential benefits.

Additionally, all researchers use various controls for age, ethnicity, and gender; variables that the Department of Justice forbids us from using in our profiling models. It is however, important to recognize that the omission of these variables will potentially result in the econometric problem of "omitted variable bias." The extent of this bias will be determined by the size and signs of the correlation coefficients between the omitted and included variables. The omission of a variable such as age will cause the coefficient of a highly correlated variable such as tenure to increase to absorb the effect of excluded age. Omitting gender and ethnicity, will tend to lead biased coefficients on variables such as education, because of their correlation with this variable.

After extensive consultations with DES and DIS staff we were able to identify a final set of variables for inclusion into the Kentucky model. They are derived from five years of claimant data which were obtained from the Kentucky UI mainframe computer databases. In addition, we used variables from the Enhanced National Data System (ENDS), the U.S. Bureau of Economic Analysis' Regional Economic Information System (REIS), the 1990 U.S. Census and the ES-202 database.

Monetary Variables: annual wage, benefit amount, reservation wage

Economic Variables: Does the worker have a phone? Is the worker economically disadvantaged? Is the worker on welfare? Does the worker hold a driver's license?

Education Variables: level of education, including type of degree (ex. GED)

Industry Variables: Standard Industrial Classification (SIC) codes or industry codes were available from the claimant's wage record database and were interfaced with information from the ES-202 database. The ES202 database is a UI report of workers and wages by industry. Its data provides information on the economic growth and decline of a claimant's industrial area on a quarterly basis, thus giving another measure of a claimant's reemployment probability.

Occupation Variables: These data were accessed from the ENDS database, and the codes were interfaced with the Occupational Employment Statistics (OES) database and converted to OES codes. The OES codes reveal the claimant's occupational history and the type of employment that he or she is likely to be seeking. The OES database can provide data on occupations that are in demand and those that are not. With this information, an assessment of the claimant's likelihood of finding employment in his or her usual line of work can be made.

Characteristics of Job Variables: The tenure variable was derived from the ENDS database. The total unemployment rate by county was used for claimants who reside in Kentucky. If the claimant resided out of state, the unemployment rate in the county where the claimant filed a UI claim was used.

Prior Spells of Unemployment Insurance Recipiency Variables: Did the worker have a previous unemployment spell? Did the worker exhaust EUC benefits? Did the worker exhaust regular benefits?

Variables to capture Regional Economic Differences: We created four sets of variables, which measured changes in employment (by SIC) for each county, based on the weighted changes

in the employment of other counties in Kentucky. These variables were weighed using intrastate commuting patterns obtained from the U.S. 1990 Census. One set of variables measured a 1 yr. percentage change in these variables, the another a two year percentage change. For the *ith* county therefore, we used the following formula to derive an *other* employment rate: $\text{(Average Other Employment)}_i = \sum_{j \neq i} \frac{(\#\text{ who live in } i \text{ but work in county } j) \text{ Employment in county } j}{\#\text{ who live in } i \text{ but work elsewhere in KY}}$

Similarly, an *other* unemployment rate for each county was also calculated, using the same commuting patterns as weights. The formula for the other unemployment rate for the *ith* county is given by: $\text{(Average Other Employment)}_i = \sum_{j \neq i} \frac{(\#\text{ who live in } i \text{ but work in county } j) \text{ Unemployment rate in county } j}{\#\text{ who live in } i \text{ but work elsewhere in KY}}$

In addition, we included variables measuring a one year and two year percentage change in the employment (by SIC) of a particular county. In our estimations we found that while none of these variables were individually significant, collectively, they were highly significant.

Thoughts and Recommendations

The census commuting patterns are easily available and we encourage other states to experiment with this approach as an additional means of controlling for regional economies. Further research needs to be conducted on the extent to which omitted variable bias may be confounding the predictive power of our models. For larger states that do not face the problems of small cell size that were endemic to the Kentucky data, we recommend a full set of interactions between the variables that we have discussed. For example, we found that plant closures had different effects across occupations and were particularly detrimental for managers who were unemployed as a result of a plant shutdown.

Amitabh Chandra

Research Associate

Center for Business and Economic Research

Gatton College of Business and Economics

University of Kentucky

Lexington, KY 40506

(606) 257-4566 [Voice] (606) 257-7671 [Fax]

chandra@pop.uky.edu [e-mail]