Profiling UI Claimants to Allocate Reemployment Services: Evidence and Recommendations for States

Final Report

March 2003

Principal Investigators

Dan A. Black
Center for Policy Research
Syracuse University

Jeffrey A. Smith
Department of Economics
University of Maryland

Miana Plesca
Department of Economics
University of Western Ontario

Suzanne Shannon
Center for Policy Research
Syracuse University

CONTENTS

LIST OF TABLES AND FIGURES

Executive Summary

This report does two important things. First, it develops and applies a state-of-the-art methodology for constructing or modifying statistical profiling models for the allocation of reemployment services that states can apply to their own data. Second, it provides substantive guidance on model development and modification to states based on our analysis of UI data from the Commonwealth of Kentucky. Our recommendations include ways of simplifying existing models without reducing their ability to predict which claimants will have long spells of unemployment as well as suggestions for improving predictive performance.

Simplifications of the Model:

We have four recommendations for making profiling models easier to estimate and implement. Our findings suggest that such simplifications may actually improve the predictive performance of the models as well.

Use Linear Models Estimated by Ordinary Least Squares:

Following the lead of the original Worker Profiling and Reemployment Services (WPRS) model, many states have relied on the use of discrete choice models such as logits and probits. While estimation of these discrete choice models is now feasible in standard statistical packages such as SAS, these models are difficult to interpret and are relatively computationally burdensome. Our results suggest that Ordinary Least Squares (OLS) estimation of the linear probability model generally outperforms the discrete choice models. For continuous dependent variables, we find that OLS estimation of simple linear models outperforms more sophisticated Tobit models. In all cases, linear models are easier to interpret and estimate (using OLS) than the corresponding non-linear logit, probit and

Tobit models. An additional advantage is that linear models allow researchers to look at an easily interpreted summary measure for goodness-of-fit (R^2) while the non-linear logit, probit and Tobit models require researchers to use summary statistics that are much harder to interpret.

Use Fraction of Benefits Exhausted as the Dependent Variable:

Again, following the lead of the WPRS model, many states use a binary variable for the dependent model of their profiling models: whether or not claimants exhaust their UI benefits. Our analysis suggests that there is a modest improvement in performance if the fraction of UI benefits exhausted is used as the dependent variable. Unlike the binary exhaustion variable, the fraction of benefits exhausted variable distinguishes claimants who use 22 weeks of UI benefits from claimants who use 2 weeks of UI benefits.

No Need to Use Local Unemployment Rates or Aggregate Industry Employment Growth Variables:

Our analysis suggests that the use of these two variables adds nothing to the predictive content of the model, which implies that they can be dropped from the model. The reason for this surprising finding is that all claimants who file a claim in a particular office in a particular week will have identical local unemployment rates. Many will also have the same industry employment growth rate variables. Thus, while including the local unemployment rate or industry employment variables improves the explanatory power of the model (e.g., the R^2 value), it does not affect the ordering of the claimants in terms of the likelihood that they will exhaust their benefits. Omitting these variables will ease implementation of the model as the remaining data needed for model estimation may simply be taken from the claimants' application forms for UI benefits.

Use of Regional Models:

Our analysis of the Kentucky data suggests that using separate models for regions within a state does not substantially improve the ability of the model to allocate reemployment services, relative to a model containing only regional dummy variables. Given the massive heterogeneity in the Kentucky economy, this suggests that other states with less heterogeneity may also not benefit from estimating regional models. Thus, a single state model (perhaps including regional dummy variables) will probably suffice. In addition, this result supports the external validity of our findings; that is, it suggests that they should apply to states other than Kentucky.

Improving the Predictions of the Model:

Richer Models Do Better:

While we spend considerable time trying to identify individual variables that substantially improve the ability of a profiling model to predict benefit exhaustion, we find that no single variable has a substantial impact. Collectively, however, richer models that control for a larger number of covariates outperform models with fewer covariates. Whether the increased predictive power is worth the added complexity depends in a large part on the expertise available to the states.

Model Performance Varies Over the Business Cycle:

We find that the predictive performance of the profiling models we examine varies substantially between the relatively high-unemployment period of the early 1990s and the relatively low-unemployment period of the middle to late 1990s. This finding suggests that occasional re-estimation may improve the performance of profiling models. We leave an exact answer to the question of how often to best re-estimate profiling models to future research.

1. Introduction

The Unemployment Compensation Amendments of 1993 (P.L. 103-152) amended the Social Security Act to require state agencies charged with administering state unemployment compensation laws to establish a Worker Profiling and Reemployment Services (WPRS) system. This system identifies Unemployment Insurance claimants "who will be likely to exhaust regular compensation and will need job search assistance services" and refers them to reemployment services to help them return to productive employment. The WPRS system consists of two major components: a profiling mechanism and a set of reemployment services. The U.S. Department of Labor (DOL) developed a "worker profiling" model that was first implemented at the state level in Maryland.

DOL guidelines originally recommended that states use the following variables in their profiling models: education, job tenure, aggregate industry-level employment changes, occupation, and local unemployment rate (Kelso 1998). We denote the model estimated using these five covariates as the "WPRS model" (Worden 1993). A Worker Profiling and Reemployment Services Policy Workgroup, however, recently recommended that "DOL should provide technical assistance to the States in improving their selection and referral processes" (Wandner and Messenger 1999).

More detailed descriptions of the U.S. WPRS appear in Balducchi (1996) and Wandner (1997). Additional information on the U.S. experience, as well as the story of an ambitious Canadian attempt to profile based on predicted impacts from services, appears in Eberts, O'Leary and Wandner (2002). OECD (1998) documents experiences with profiling in Australia, Canada, and the United Kingdom. Outside the U.S., the

Australians have had the most success with profiling. They have employed a sophisticated multivariate profiling system combining an automated model based on respondent characteristics with caseworker discretion to override the automated system based on factors not present in the administrative data, such as poor motivation.

This research project aims to help states improve their selection and referral procedures in two ways. First, we provide a general methodology for the evaluation, development and modification of profiling models. Second, we provide substantive guidance regarding types of changes to profiling models that should improve the assignment of claimants to reemployment services. Because only a limited number of UI claimants can be assigned to employment and training services in each time period, the selection model ought to perform the best job of selecting, out of the entire pool of UI claimants, those individuals who are most in need of receiving reemployment services.

Based on equity arguments, DOL has identified those UI claimants most likely to exhaust UI benefits as being the ones most in need of receiving employment and training services. Their argument is twofold. First, claimants who exhaust or nearly exhaust UI benefits face the strongest barriers to labor force reentry, and reemployment services should help them overcome such barriers. Second, if by participating in reemployment services UI claimants find a job and exit unemployment sooner, they will not collect the entire amount of entitled UI benefits and the states will economize on UI funds.

Because states may differ in their data collection and in their ability to implement various profiling models, we provide guidance across a range of models that vary in complexity. In order to guide states in making incremental improvements to their profiling models, according to their respective data collection and technical capacity, we systematically examine the determinants of benefit recipiency duration. We consider how different model specifications regarding the regressors included in estimation, the dependent variable, and the functional form that the dependent variable implies, affect the models' abilities to sort UI claimants.

To assess the predictive performance of each profiling model, we forecast the dependent variable (usually the probability of benefit exhaustion or the expected fraction of benefits exhausted) for each claimant. We then sort UI claimants by their forecasted values and compute the fraction of benefits actually exhausted within groups of claimants with varying forecasts. In particular, we compute the fraction of benefits exhausted by UI claimants in each of the five quintiles from the distribution of forecasts. If a model is accurate, the upper quintiles should show a high proportion of the average fraction of UI benefits exhausted, while the opposite should be true for the claimants in the lower forecast quintiles. We provide a methodology to determine the impact of each additional variable on the predictive performance of the profiling models and we recommend the specification we consider preferable given the tradeoff between model performance and ease of implementation.

In detail, our report carefully addresses the following questions:

- What functional form should the estimated model have? Would a sophisticated nonlinear specification like logit, probit, or tobit bring improvement over the standard linear model estimated by ordinary least squares (OLS)?
- What is the appropriate dependent variable? Is it the binary variable of whether or not the claimant exhausted UI benefits, or some other dichotomous

dependent variable, e.g. whether the claimants exhausted up to a large fraction of their maximum allowable UI benefits? What is the improvement from using – instead of a binary variable – a continuous dependent variable such as the ratio of benefits drawn to benefits entitled?

- Are all of the five explanatory variables used in the WPRS model, in particular the local unemployment rate and the change in aggregate employment in the claimants' industry of employment, relevant in improving the performance of the profiling model in allocating claimants to services?
- What additional variables might improve the assignment of claimants to reemployment services? Do the additional variables simply improve the statistical fit of the model, or do they alter the assignment of the claimants? Which are the variables that entail the largest improvement in predictive performance, and what is an optimal trade-off between model performance improvement and practical operational difficulty due to additional regressors?
- How important are regional differences in predicting the duration of recipiency?
- Is the assignment of claimants more accurate during recessions or during boom periods? Given existing capacity constraints, how do business cycles affect the performance of profiling models?

The analysis begins by estimating the WPRS model outlined in Kelso (1998) using the Kentucky data. We find that the linear model is the most versatile, and it has the added advantage of simple estimation by OLS. It also leads to the best results in most cases. We find that the local unemployment rate and the aggregate employment changes in the claimant's last industry variables included in the WPRS model do not improve the assignment of UI recipients. We find that a continuous dependent variable – fraction of benefits exhausted – does a better job at allocation than the dichotomous dependent variable – UI benefits exhausted or not – utilized in the WPRS model, because it incorporates the information contained in the durations of those claimants who do not exhaust their benefits. Adding explanatory variables improves the performance of the model, although at the cost of loss of simplicity. We examine how additional covariates, several of them currently used in some state models, improve the assignment of claimants to reemployment services. While we do not pinpoint one single best predictor, we make recommendations based on incremental improvements in the models' predictive power that result from adding various regressors. We find that there is no improvement to be gained from estimating separate regional models in the Kentucky data, relative to just including regional dummy variables. Making this change would greatly simplify some states' models. We also find that the predictive power of profiling models is sensitive to changes in the business cycle. The models predict best during periods of high unemployment, as there is greater heterogeneity among claimants.

While we argue that our substantive findings generalize to states other than those used in our empirical work, our report has great value even to readers who disagree with this assessment. Our methodology provides a template for states interested in investigating potential improvements in their UI profiling models using their own data, rather than relying on our findings obtained using data from Kentucky. Such states can simply repeat our analysis using their own data and then draw the appropriate conclusions.

2. Data

Our recommendations are based on the analysis of UI administrative data from the Commonwealth of Kentucky for the fiscal years 1989 to 1995. These seven years include a variety of different periods relative to the business cycle. The data also encompass a substantial amount of economic diversity within Kentucky. These facts, coupled with the finding (described in detail in Section 5.5) that estimating separate regional models does not have a large payoff in terms of predictive power, lead us to believe that our main substantive findings likely generalize to other states. The Kentucky data we utilize here are also very clean; alternative measures of the same concept (e.g., weeks of benefits received) constructed using different elements of the data give about the same answer.

The data we employ here are the same data utilized in Black, Smith, Berger and Noel (2002). They include all UI claims filed in the Commonwealth, except those on temporary layoff or hired from a union hall. UI claimants qualify for a maximum of 26 weeks of UI benefit recipiency. As the maximum amount of UI benefits to which a claimant is entitled is known, as well as the amount he or she actually collected, we can measure the extent to which claimants exhausted UI benefits as the ratio of benefits collected over benefits entitled.[1]

The data are quite rich and allow us not only to replicate the five-covariate WPRS model, but also to investigate how the profiling of UI clients is improved with additional regressors pertaining to the claimants' backgrounds, characteristics of last employment, social assistance recipiency, and UI claim histories.

The methodology for assessing the models' predictive performance, described in detail in the next section, requires that we withhold ten percent of the observations from the initial estimation; we use these observations for out-of-sample forecasts. Fortunately, the Kentucky sample is very large, about 330,000 observations, and therefore small sample sizes are never an issue, not even when the analysis is replicated on separate fiscal years. For some analyses, we report results using only data from fiscal year 1994, the most recent year for which we have data for the full fiscal year. In these cases, supplemental analyses not reported here indicate that our substantive conclusions hold in the other years of our data as well (which suggests that they also likely hold in years after our data run out in 1995).

3. Methodology

3.1. The WPRS Model

The profiling procedure involves two steps. In the first stage, all claimants are screened and some of them are a priori excluded from the whole process. These excluded claimants are either laid-off claimants with a known recall date, whose ties with their employers need not be severed, or union workers, who are referred to employment by their union. In the second stage, statistical models are estimated, and their predictive performance is tested, using a validation procedure that we describe below. The Kentucky data we use already excludes claimants from the first step.

Estimation of models to predict the length of spells of UI benefit receipt is a difficult task. Worden (1993) provided the baseline model for profiling UI claimants. She used a logit model with *UI benefit exhaustion* as the dependent variable and a parsimonious specification of the independent variables. Other researchers (e.g., Eberts and O'Leary 1996) have also used binary choice models in developing profiling systems.

---

[1] A ratio of one means the claimant exhausted his or her UI benefits.

There is also variation among states in the choice of dependent variable. For instance, Washington State uses an indicator variable for whether the claimant collected at least 90 percent of their benefit entitlement, while Idaho counts the number of weeks the claimant collected benefits (Kelso 1998).

The problem with using a dichotomous dependent variable is that all of the data variation among individuals who do not exhaust their UI benefits is ignored. Figure 1 presents a histogram of the fraction of benefits exhausted for claimants in Kentucky in 1994. It reveals that the data have a large mass of observations at one. In other words, almost forty percent of UI claimants exhaust their benefits. Thinking of the story depicted by Figure 1 alone, one would be inclined to advocate the use of a dichotomous model that splits claimants into exhausters and non-exhausters, as in the WPRS model specification.[2]

Nevertheless, as documented in Figure 2, when only non-exhausters are considered, there is substantial variation in the duration of benefits for the group that does not exhaust. Thus, a dichotomous model that treats all claimants who do not exhaust as identical ignores much useful information in the data. Indeed, one might expect non-exhauster claimants who use up almost all of their benefits to be more similar to claimants who exhaust their benefits than to those with very short spells of UI receipt. Model specifications that employ a continuous dependent variable exploit this variation in benefit duration among non-exhausters, and thus would be expected to yield better predictive results than the dichotomous models.

To document which specification of the dependent variable will result in the best assignment of claimants most in need of services, i.e., most likely to exhaust UI benefits,

[2] Exhausters comprise the group of claimants who exhaust their UI benefits. Non-exhausters comprise the group of claimants who do not exhaust their UI benefits.

we examine the following five specifications of the dependent variable as identified by Kelso (1998):

(1) a dichotomous variable for whether or not the claimant exhausts his or her benefit entitlement;

(2) a dichotomous variable for whether or not the claimant collects at least 90 percent of his or her benefit entitlement;

(3) the ratio of benefits drawn to benefit entitlement (i.e., fraction of benefits claimed);

(4) the number of weeks of benefits claimed; and

(5) a dichotomous variable for at least 26 weeks of benefits claimed.

For discrete variables we use the same logit model specification as the original WPRS model, that is, we use the dependent variables as described at (1), (2) and (5) above, and the five WPRS model regressors. Besides the logit model, we also estimate a similar probit specification on the same dependent and independent variables.[3]

For the continuous dependent variables – cases (3) and (4) above – we use OLS estimation of a simple linear model, along with a more sophisticated tobit model that accounts for masses of observations either at the lower or upper bounds of the support of the dependent variable.[4] For example, when the dependent variable is *fraction of benefits exhausted*, the observations are massed at one because many claimants exhaust their benefits.

---

[3] The difference between the logit and probit models comes from distributional assumptions regarding the error term. In the logit case the error term is assumed to come from a logistic distribution, while the error term is assumed to have a normal distribution in the probit specification.

[4] The support of a distribution of a random variable is the set of all possible values that the random variable takes on with positive (i.e., greater than zero) probability. In our case, the random variable

3.2. Additional Explanatory Variables

In addition to utilizing a binary dependent variable, the WPRS model includes only five covariates: education, job tenure, occupation, changes in aggregate employment in the industry of the claimant's last employment, and the local unemployment rate.[5] The WPRS model has a poor predictive performance in part because of the very small number of covariates it includes. As we show in detail in Section 3.4, increasing the number of covariates improves profiling performance. The problem with adding extra covariates is that too many covariates make the model difficult to implement in day-to-day operations by UI staff. In practice, states make different choices regarding this tradeoff. For example, the model for the state of Pennsylvania uses only eight covariates, while the model for Washington State, which is one of the larger state models, includes 26 covariates.[6] At the other extreme, the Kentucky model contains over 140 different covariates.[7]

Table 1 describes the five independent variables in the WPRS model as they are coded in the national, WPRS, and Kentucky applications. In the Kentucky data, the job tenure and the occupation variables refer to the last main job of the claimants. The industry variable gives the percentage change in employment in the industry of the claimant's last main job. The industry employment change is recorded as the percentage change from the previous month's employment figures. The unemployment rate variable is recorded monthly at the county level. Both the industry employment change and the

is the dependent variable, i.e. either the fraction of benefits exhausted or the number of weeks of benefits claimed.

[5] We call a specification employing the five covariates defined here the WPRS model. We refer to a reduced specification that only includes three covariates: education, tenure, and occupation, as the basic (or reduced) WPRS model.

[6] O'Leary, Decker, and Wandner (2002).

unemployment rate enter the estimation with a lag. The rationale behind using lagged values is that, at the moment of application, only last month's figures are known to the personnel who operate the profiling models.

Table 2 lists the variables in the WPRS model along with variables included in other state models. We compiled this list using Kelso (1998) and Berger, Black, Chandra, and Allen (1997). With three exceptions described in the notes to Table 2, the Kentucky data allow us to examine the relative importance of each variable considered by other states. The list in Table 2 also includes some variables that have not been used by any state, but which we thought might improve the allocation of claimants to services.

If some of the variables have missing values, rather than simply discarding those observations we use an imputation procedure. If the missing values come from a categorical variable, we create one more indicator variable for the category "missing." If the missing values come from a continuous variable, we add an indicator variable equal to one for the observations that had missing values and equal to zero otherwise, and we replace the missing values in the original variable with zero.

3.3. Assessing Predictive Performance

We use two different sets of measures to determine the performance and fit of our models. The first set of measures consists of the usual within-sample statistics reported for estimated models, such as R^2 for the linear model estimated by OLS or the log-likelihood value for models estimated by maximum likelihood, such as the logit or probit.

Unfortunately, statistics such as the R^2 or other within-sample forecast measures are often not compelling. First, within-sample statistics are not realistic tests given how

---

[7] Berger, Black, Chandra, and Allen (1997).

profiling models are actually implemented. Profiling is necessarily a forecast of the claimant's duration of recipiency, not a within-sample estimate. Indeed, this is the reason that Berger et al. (1997) exclude ten percent of the data from the estimation of their models and use these data to evaluate out-of-sample predictive performance.

Second, the use of R^2 and other summary measures of the goodness-of-fit of a model may overstate the impact of variables on the predictive power of the model. A simple example illustrates why. Suppose that a researcher includes a dummy variable for the month in which the worker's spell is initiated. Given the seasonal nature of unemployment, this might substantially improve the fit of the model. It will not necessarily improve, however, the assignment of workers to services because the variable will have the same effect on the predicted values of all workers beginning their spell of unemployment in the same month.[8] The inclusion of the dummy variable for the month will improve the statistical fit of the model, but may not alter the ranking of claimants for referral.

To avoid this so-called false accuracy problem, we rely on portions of the sample excluded from estimation to simulate the assignment of workers to treatment. By simulating the actual assignment, we may examine whether the addition of a variable, or the choice of a particular estimation methodology, improves the assignment mechanism. Our validation methodology is an improvement over much of the existing literature, in the sense that we measure how well the models predict for claimants not included in the estimation. Thus, we follow Berger et al. (1997) and randomly exclude from the estimation ten percent of the sample. We keep this ten percent sample for validation

[8] This is only approximately true in the case of non-linear models such as the logit or the probit.

purposes. The models are estimated on the remaining ninety percent of the sample. The forecasting performance is then tested on the ten percent validation sample.

Berger et al. (1997) create their validation sample by randomly excluding claimants *from all weeks* in which claims were filed. In contrast, we create our validation sample by excluding all claimants *in a random sample of weeks*. Claimants who file within the same week will face similar local economic conditions. Randomly selecting the validation sample on the basis of the weeks in which claims are filed limits the variation across these economic variables and thus avoids potentially inflating the predictive performance measures due to the variation in local economic conditions over time within the validation sample.

In creating the validation sample we randomly exclude four weeks out of fifty-two in each fiscal year. After discarding the observations in the validation sample, we estimate the model on the remaining forty-eight weeks of observations.[9] We use the coefficients obtained from the estimation sample to compute the predictions for the validation sample. In the case of dichotomous models, we predict the probability that each claimant will exhaust his or her UI benefits (or the probability of exhausting at least ninety percent of UI benefits when that is the dependent variable). The predicted probabilities lie within the interval from zero to one. In the case of models with a continuous dependent variable, we predict either the fraction of UI benefits exhausted – out of one hundred percent – or the duration of the claim in weeks – with a maximum of twenty-six weeks.

For assessing the predictive power of all of our models, we examine the actual UI benefit duration (as recorded in the data) at various percentiles of the distribution of predicted values of the dependent variable for the model under consideration. We use the term "predicted UI benefit duration/exhaustion" to denote, depending on the dependent variable in the model, the predicted probability of exhausting UI benefits, the predicted probability of collecting at least 90 percent of the benefit entitlement, the predicted fraction of benefits claimed, the predicted number of claimed weeks, or the predicted probability of consuming at least 26 weeks of benefits.

Claimants from the ten percent validation sample are sorted into quintiles based upon the distribution of predicted values of UI benefit duration/exhaustion. We look at quintiles rather than some other, finer partition such as deciles because examining the quintiles provides sufficient information to assess the predictive performance of our models.[10] Our measure of predictive performance is the average fraction of benefits exhausted for claimants in each quintile of the distribution of predicted values. If a model has good predictive performance, then the fraction of benefits exhausted by claimants from the top predicted quintiles should be large, and the fraction of benefits exhausted by claimants from the bottom predicted quintiles should be small. A useful performance benchmark is given by the average fractions of benefits exhausted by claimants in quintiles of the distribution of fraction of benefits received in the raw data (that is, quintiles of the distribution of realized values of fraction of benefits received). This comparison adjusts

[9] For the 1995 fiscal year we only have twenty-six weeks of data, so we exclude two random weeks and keep twenty-four weeks in estimation. Also, it may happen that a fiscal year has fifty-three weeks, so after excluding four weeks we keep forty-nine weeks in estimation.

[10] O'Leary, Decker, and Wandner (1998, 2002) assess predictive power in their application by inspecting deciles of predicted durations. Nevertheless, they only consider within-sample forecasts, rather than the more appropriate out-of-sample forecasts we examine here.

for differences across years (or across states, for cross-state comparisons) in the mean fraction of benefits exhausted.[11]

To evaluate how much better we can profile claimants by using sophisticated statistical techniques, we also consider a random assignment mechanism. Under random assignment, claimants are not profiled based upon their predicted UI benefit duration/exhaustion. Instead, the claimants are simply put in a random order. Comparing the average fraction of benefits received for quantiles of predicted values from a profiling model with the same averages for quantiles of the randomly ordered distribution quantifies the improvement generated by the statistical profiling model relative to random assignment to services. If no significant improvements in assignment arise from the models relative to random assignment, it is difficult to justify the use of statistical profiling.

The top panels in all of the tables presenting results on the predictive performance of profiling models provide the predictive performance measures computed using the raw data and using the random assignment mechanism. These measures serve as baselines for the performance of each estimated model, with the measures based on the realized duration and exhaustion data at one extreme (the best case) and those based on the random assignment mechanism at the other extreme (the worst case).

3.4. Comparing Profiling Models

The ability of a model to sort claimants by their predicted values of UI benefit duration/exhaustion is measured as the difference in average fraction of benefits exhausted

---

[11] We believe the results reported here to serve as broad guidelines for all states implementing profiling models. Our experience with data from Kentucky and other states indicates that different data sets may have substantially different sample averages for variables that encode the same basic

between the top and the bottom of the distribution of predicted values. Depending on the nature of the application at hand, we may be interested in knowing how the models perform at different points of the predicted distribution. Put differently, if approximately 80 percent of clients are assigned to reemployment services, then we want to make sure our models are able to predict correctly the top 80 percent of the distribution of actual spell lengths. If instead only 40 percent of clients are assigned to reemployment services, we prefer that our models identify accurately the top 40 percent of actual spell lengths. Nevertheless, because of capacity and budgetary constraints and because of fluctuations in the number of unemployment claimants, there is no clear-cut threshold to use.

Ideally, a model will perform well at all points of the distribution of predicted UI benefit duration/exhaustion values. To obtain an accurate picture of the predictive performance, we report results at different points in the distribution of predicted values. In particular, we compare the difference in the average fraction of benefits exhausted between the top 80 percent and bottom 20 percent of the distribution of predicted values, as well as the differences between the top 60 percent and bottom 40 percent, the top 40 percent and bottom 60 percent, and the top 20 percent and bottom 80 percent. We also report an average of the five differences. By reporting the results in this way we facilitate qualitative claims about the relative performance of various models.

To begin, we compare the performance of existing profiling models as reported in the literature. Given the limited number of covariates, and the inherent difficulties in predicting the exhaustion of UI benefits, it is not surprising that the explanatory power of the WPRS model is modest. Table 3 reports the differences in predictive power between

---

concept. This is why we recommend values relative to respective sample averages, rather than absolute values, for the predictive measures.

the Pennsylvania, Washington, and the 140-covariates Kentucky models. It also reports the predictive power results we obtained from a model with the five WPRS covariates and from our preferred model specification estimated on Kentucky data, both using fraction of benefits exhausted as the dependent variable. For the Pennsylvania and Washington models we report differences between the top 25 percent and bottom 75 percent of the predicted distribution, taken from O'Leary, Decker and Wandner (1998). For the other three models (all estimated on Kentucky data) we report differences between the top 60 percent and bottom 40 percent of the distribution of predicted UI benefit duration/exhaustion values, in order to be consistent with the way Berger et al. (1997) report the predictive results for the larger (140-covariates) Kentucky model.

The predictions for the larger (140-covariates) Kentucky specification are far superior to those of the model with the WPRS covariates, as well as the Washington and Pennsylvania models. Our preferred model specification, estimated on the Kentucky data, fares much better than the model with the WPRS covariates, but not as well as the more elaborate Kentucky model.[12]

The results in Table 3 actually overstate the performance of the WPRS model because we have used fraction of benefits exhausted as the dependent variable in order to emphasize the effects of the covariate set in comparing the five WPRS model covariates to our preferred specification and to the Kentucky model. It is the 140 different covariates (or some subset thereof) in the Kentucky model that account for its superior performance in Table 3. In what follows, we strive to find a balance between model simplicity (fewer covariates) and model performance.

4. Use Simple Linear Models Estimated by OLS

The original model developed by Worden (1993) consists of a logit model with UI benefit exhaustion as the dependent variable. Since the imprecision in logit models is larger at the tails, we expected a similar probit model to perform marginally better that the logit. It turns out that for all the models with a dichotomous dependent variable, the probit model does not yield any performance improvement over the logit specification.[13] Moreover, the linear probability model estimated by OLS performed at least as well as the nonlinear logit and probit estimated by maximum likelihood. For models with a continuous dependent variable, we estimated both simple linear models using OLS and more sophisticated tobit models that account for mass points at either end of the distribution by maximum likelihood.[14] Once again, none of the models outperformed the linear model estimated using OLS.

Table 4 shows comparative predictive results from linear probability and simple linear models estimated by OLS, as well as logit, probit, and tobit models estimated by maximum likelihood, using the Kentucky data. Although for simplicity we only report results from models including the five covariates from the WPRS model, our finding that all else equal, linear models estimated by OLS are not outperformed by any other specifications holds more generally. Moreover, in numerous instances linear models

---

[12] We use only the fiscal year 1994 in this set of estimation results for simplicity reasons. The detailed year-by-year analysis in Section 8 shows that 1994 is neither among the best years for prediction nor among the worst years for prediction. For the empirical results, see Table 11.

[13] The dichotomous dependent variables are UI benefits exhausted, at least 90 percent of UI benefits exhausted, and at least 26 weeks of claimed benefits. The continuous dependent variables are fraction of benefits exhausted and number of weeks of claimed benefits.

estimated by OLS actually yield better predictive results. Given the simplicity of the linear model and OLS estimation, and given the fact that more sophisticated models such as the double limit tobit or probit do not bring any improvement to the predictive power of the models, we recommend the use of the linear regression model in all profiling model estimation.

5. Results

5.1. The Dependent Variable Should Be Fraction of Benefits Exhausted

We consider five alternative definitions for the dependent variable: *exhausted UI benefits* (dichotomous); *exhausted 90 percent of benefits* (dichotomous); *fraction of benefits exhausted* (continuous); *number of weeks claimed* (continuous); and, finally*, at least 26 weeks of benefits claimed* (dichotomous). In Table 5 we report predictive results from estimating five linear regression models, each containing the five WPRS model covariates and using one of the dependent variables just listed.

A continuous dependent variable – either the fraction of benefits exhausted or the number of weeks of claimed benefits – results in better predictive outcomes. Although not presented here, results from other model specifications indicate that the fraction of benefits exhausted is the best choice for a dependent variable. In all further analyses we estimate linear models using OLS with fraction of benefits exhausted as the dependent variable.

---

[14] For instance, about 40 percent of the observations on the fraction of benefits exhausted equal one in the Kentucky data.

5.2 Variables Measuring Local Economic Conditions

The WPRS model contains two variables measuring local economic conditions: the unemployment rate and aggregate employment growth rate in the claimant's industry of last employment. These variables provide sources of "false accuracy". It is quite possible that the local unemployment rate, while an important determinant of the duration of unemployment, will not provide useful information for sorting the claimants. While county level unemployment rates do vary somewhat at a point in time, most of the variation in unemployment rates is over time, not within relatively narrow geographic areas. Moreover, because virtually all of the claimants applying for UI at a given service delivery area (SDA) face the same unemployment rate, the regional variation in unemployment rates will not help separate among clients applying at the same SDA. The same points hold (although somewhat more weakly) for industry employment changes. Many claimants at a given office in a given week will often come from the same industry, due to the geographic sorting of industries and the occurrence of mass layoffs. Such claimants will all have the same value for the industry employment variable, which will therefore not aid in sorting among them, even though it may increase the fit of a profiling model.

Table 6 reports the sensitivity of the predictive performance of a model with the five WPRS covariates and fraction of benefits exhausted as the dependent variable to dropping either the local unemployment rate or the industry employment change variable. We also report results from models that omit the local unemployment rate and industry employment change variables, while adding instead local or regional dummies. We observe some increase in the predictive performance when unemployment rates are used among the regressors, but the performance increase is more noticeable when the local unemployment rates are replaced by regional dummy variables. The most likely explanation is that local unemployment rates are a proxy for the omitted regional variables. In a properly specified model, with no omitted variables, there is no need to include either local unemployment rates or industry employment changes. Consequently, the basic model to which we add additional regressors in the remainder of the analysis will be a basic WPRS model that includes only three regressors: education, tenure, and occupation. We omit unemployment rates and industry employment changes from all future specifications.

5.3 Additional Variables: There Is No Single Best Predictor

Assessing the relative importance of each variable included in the estimation allows us to evaluate whether each variable provides sufficient information to warrant inclusion in the model. It is a bit of an art form to come up with a specification that is both parsimonious and at the same time does not leave out any variables that may improve the predictive power of the model. The questions we address here are: What variables might be added to improve the predictive power of the model? Do these variables simply improve the statistical fit of the model, or do they alter the ordering of claimants?

In Table 7 we provide a description of the variables included in the 40 specifications we estimate using the Kentucky data. Model 0 is the three-covariate WPRS basic model specification. Models 1-7 augment the base specification with various measures of past UI benefit take-up. Other specifications up to Model 28 include, one at a time, economic status and transfer payment variables, previous wages, tenure squared, reason for job separation, enrollment in school at the time of filing the claim and employed at the time of filing the claim.[15] Model 29 adds local office variables. Models 31 to 38 combine some of the more successful variables identified previously. For models 31 to 38, each even numbered specification consists of the preceding odd number specification plus regional dummy variables.[16] The last specification, model 39, is the most elaborate one, including all of the available variables that might be expected to make a reasonable contribution to explaining UI benefit duration.

In putting together a best specification we have to keep in mind that, although the more covariates we add, the better the within-sample fit (as shown by the R-squared measure), increased performance comes at the cost of a more complex model, which is less easy to operate by the states. Table 8 reports in-sample measures of fit (R-squared) and out-of-sample measures of fit (fraction of benefits exhausted by predicted quintiles) for all of our model specifications. In every case we estimate linear models by OLS with fraction of benefits exhausted as the dependent variable. For simplicity, we only look at Kentucky data for fiscal year 1994.

It is clear that none of the specifications from Model 1 to Model 26 bring any improvements over the WPRS model. The first headway is made with the covariates “enrolled at time of claim” and “employed at time of claim.” With each of these two added covariates there is a significant jump in the R-squared as well. All specifications

---

[15] Models 18 and 19 include the weekly benefit amount received and the potential amount. We use potential (i.e., maximum allowable) weekly benefit amount because at the time of filing the claim the applicant will know the potential, but not the actual, benefits claimed. For conformity with other states, we tested specifications including either the potential or the actual weekly benefit amount claimed. All nominal variables are expressed in real terms using the CPI deflator with base year 1995.

[16] For example, model specifications 31 and 32 are equivalent, except that model 32 includes local office dummies while model 31 does not.

from Model 31 to Model 40 bring a large improvement over the base specification, both in terms of R-squared (about four times larger), and in terms of the out-of-sample predictive performance of the model. Not surprisingly, adding local office dummies, all else equal, improves the R-squared and the predictive performance.

Also not surprisingly, the more covariates we add to the model specification, the better the predictive performance. Nevertheless, the relationship between added regressors and performance increase is not linear, and at some point the cost of adding one more regressor will be less than the performance increase it entails. In other words, when picking the best specification we have to keep in mind that, although the more added covariates the better the fit, it all comes at the cost of a more complex model, less easy to operate by the states.

We consider the specification that keeps the best balance between performance and simplicity to be Model 36. In terms of predictive performance, Model 36 does almost as well as the most complex model, Model 39. Looking at the fraction of benefits exhausted within each predicted quintile, the differences between Model 36 and Model 39 show only at the third decimal point. The average difference in prediction differences across quintiles, our preferred measure of predictive fit, again shows Model 36 second only to the elaborate Model 39. What tilts the balance in favor of Model 36 is that, while Model 39 employs 28 covariates, Model 36 uses only about half as many – just 15 covariates.[17]

[17] When counting the covariates, dummies for the categories of a categorical variable are not counted as separate regressors. For instance, we count education as one regressor, although four dummies for educational categories enter in the estimation.

5.4 There Is No Improvement from Estimating Separate Regional Models

The economy of Kentucky is quite heterogeneous, with large metropolitan areas with expanding economies (such as Lexington), older, industrialized metropolitan areas (such as the Northern Kentucky portions of Cincinnati and Louisville), and a variety of rural areas. How important are regional differences in predicting the duration of recipiency? Should states use distinct models for various regions or does a single state model adequately predict recipiency duration? To our knowledge, no one else has documented in writing the relative improvement in the assignment that arises from the use of distinct models for different geographic regions, though some states other than Kentucky find such models useful based on their own analyses. Clearly the use of regionally distinct models greatly increases the complexity of implementation. Our analysis of the regional models has broader implications: if regional models are not important within Kentucky, this provides support for the view that our substantive findings for Kentucky likely generalize to other states.

The following question arises: How much does the added complexity improve the assignment of claimants to reemployment services? The answer, as evident from Table 9, is not very much.[18] There is no gain from estimating separate regional models once regional dummies are included in the model. Additional interaction terms do not improve the assignment of clients.

To provide evidence to this claim, we compare assignment results from two different experiments. In the first instance, we estimate separate models for the eight

[18] We estimate linear models by OLS using the 1994 Kentucky data with fraction of benefits exhausted as the dependent variable and the covariates from Model 36.

regions of Kentucky in our data. In the second instance, we estimate a single model jointly on all eight regions, with controls for each separate region. Although there is significant heterogeneity across regions in terms of economic conditions and other factors, the predictive difference between the regional models and the joint model is remarkably small. Across regions, the control variables have the same effect on the fraction of benefits exhausted.

5.5 Business Cycles and Data Quality

In Table 10 we report the results from estimating various models pooling all of the years in our data, and also separately year by year. We report results from both the basic WPRS model with three variables and our preferred specification, Model 36 from Tables 7 and 8.[19] We find substantial swings in predictive performance from year to year.

As a rule of thumb for identifying boom versus recession years we can compare the fraction of benefits exhausted in the entire sample during a particular year (as given by the random assignment measures) with the average for all years (reported in the column "All"). Recession years should have a larger than average fraction of benefits exhausted. Another rule of thumb indicates that during recession years about 60 percent of claimants (the top three quintiles in the Data panel) collect more than 90 percent of their benefits, while during boom years only about 40 percent of claimants (the top two quintiles in the Data panel) collect more than 90 percent of their benefits.

[19] We have already seen some of the results reported here. For instance, the first column from the base Maryland model estimated on pooled data (all years) is the same as column A from Table 6. The seventh column, 1994, from the best specification model is the same with the estimated model 36 from Table 8.

In terms of predictive success, profiling models for recession years like 1991 usually perform better than profiling models for boom years like 1994. This result also holds when an average boom period, 1993 to 1995, is compared with an average recession period, 1989 to 1992. It is reassuring that the claimants most likely to exhaust UI benefits, those most in need of reemployment services, are more easily identified in periods of recession, since the number of claimants treated with reemployment services cannot be proportional to the number of claimants because of capacity constraints. These patterns likely arise from the increasing diversity among UI claimants during downturns, as many workers with long employment spells become unemployed. It is easier to predict well when the variation in the data increases.

Nevertheless, one qualification is necessary. Models for recession years perform unambiguously better than models for boom years in the sparse 3-covariate basic WPRS specification. When the estimated model is our preferred specification – the 15-covariate Model 36 – the pattern of results changes somewhat. Although the average recession period of 1989 to 1992 continues to exhibit better predictive results than the average boom period of 1993 to 1995, the gap between the two periods becomes narrower. Moreover, examining the year-by-year results reveals that for a boom year like 1993, the predictive results in the best specification estimation (Model 36) can be as good as for most recession years and better than the results for 1992, a recession year. Because we know for a fact that, from 1993 on, the quality of the Kentucky data has improved, we believe that the better performance in 1993 comes from improved data quality. In the case of the sparsely specified basic WPRS model, improved data quality starting in 1993 did not help much as not much data was used in the first place.

These estimates do not answer the important question of how often to optimally re-estimate profiling models, but they do suggest that occasional re-estimation is useful, and that models with few covariates may optimally require more frequent re-estimation than models with many covariates. A complete answer to this question awaits future research.

5.6 Another Covariate: Delayed Filing

When presenting our initial draft at the Employment and Training Administration, a suggestion was made to examine the impact of delayed filing for UI benefits on the fraction of benefits exhausted. Unfortunately, the Kentucky data are not ideal for this task. For most claimants, we were able to construct this measure by examining the date last worked relative to the date the claimant filed for UI benefits. Some caseworkers, however, updated the date last worked variable when claimants had an employment spell within the benefit year and then had a subsequent unemployment spell. As a result, some of the estimated gaps between employment and filing for benefits are negative. With this caveat, however, we use the Kentucky data to examine the impact of filing late on the fraction of benefits exhausted.

Table 11 presents the definitions of the alternative versions of the delayed filing variable that we examine. We consider six categorical measures and three continuous measures in separate regressions. Table 12 presents the results of our analysis, in which we include these delayed filing variables in four of the models described in Table 7: the basic WPRS model (Model 0), the WPRS model with covariates for the previous years UI activity (Model 1), our preferred model (Model 36), and the model with the most extensive set of covariates (Model 39). In each case, the dependent variable consists of the fraction of benefits exhausted and we estimate a linear model using OLS.

For our preferred model with no delayed filing variable, claimants in the top quintile of predicted fraction exhausted actually collected an average of 74.1 percent of their benefits. Adding our measures of delayed filing generally improves the predictive performance of the model, with the fraction of benefits exhausted in the top quintile ranging from 74.0 percent to 75.1 percent. Using the difference between the top quintile and the bottom quintile, our preferred model does better with any of the delayed filing variables than it does without one. Moreover, the difference is often substantial.

Finally, note that much of the improvement is obtained when we use a simple dummy variable for whether or not the claimant waited 30 days or more to file a claim (Measure A in Table 11). Given its ease of implementation, this would appear to be a desirable variable to include in profiling models, although the evidence in support of including it in sparsely specified models, such as Models 0 and 1, is weaker than for the richer models.

5.7. Some Methodological Issues

We believe one of the major contributions of this work is to outline a method of evaluating the performance of competing profiling models that states can apply to their own data. Traditional measures of whether variables should be included in the model are generally not appropriate for the analysis of profiling models because in profiling only relative prediction matters. That is, in a profiling model, individuals are selected based on the order of the predicted values, not based on the actual predicted values themselves.

To see this point better, consider adding 0.3 to each of the predicted values for fraction of benefits exhausted from a model. In the case of OLS estimation, this would drastically reduce the R^2 because the regression line would no longer go through the center of the data. It would not, however, change the ordering of claimants when it came to allocating scare reemployment services slots. As a result, this reduction in R^2 would have no impact on the allocation of services. It is, of course, necessary that the inclusion of additional covariates increase the R^2 of an estimated equation, but, importantly, the additional covariates do not necessarily improve the ordering of claimants and may actually worsen the ordering over parts of the distribution.

Similarly, when evaluating the performance of a profiling model, it is crucial that the sample of claimants used for prediction consist of all claimants from randomly selected weeks, rather than a random sample of claimants from all weeks. In our empirical work, we randomly select a set of weeks and use all claimants from those weeks for prediction, leaving them out of the estimation entirely. This allows an evaluation of the profiling model in a realistic setting and minimizes any spurious improvement or degradation in the model's performance because of differing economic conditions across weeks.

6 Conclusions

In this paper, we conduct a systematic investigation of alternative profiling models using UI records from the Commonwealth of Kentucky. Our analysis provides a methodological template for similar analyses of data from other states. In addition, we provide six substantive guidelines for the specification of UI profiling models that follow from the empirical findings in our analysis. Although based on data from a single state, we argue that these findings have relevance for other states, in part because of our finding that separate regional models are not needed to capture the (very) heterogeneous labor markets within Kentucky. Following these six guidelines should result in substantial improvement in the performance of state the profiling models.

First, because of its simplicity and better performance, OLS estimation of linear models estimated is preferable to estimation of more sophisticated models such as the logit, probit or Tobit.

Second, a continuous variable, such as the ratio of benefits claimed to benefits entitled, captures more of the variation among UI claimants who do not exhaust their benefits. As a result, a continuous measure should be used instead of a dichotomous one as the dependent variable in the profiling model.

Third, there is no need to include either local unemployment rates (lagged) or employment changes in the industry of the claimant's last job (lagged), because these variables do not help improve the allocation of claimants.

Fourth, we suggest additional variables that improve the predictive performance of profiling models in the Kentucky data. We find no single best regressor that yields large improvements in predictive power. Instead, we find that improved performance beyond the basic WPRS model requires the addition of a set of additional variables. Although there is no scope for a miracle-working covariate, we can nevertheless identify relatively parsimonious specifications that substantially improve the fit of the model. Our preferred model specification from this analysis includes fifteen regressors, all of which are readily available on completed UI claim forms. These regressors include education, occupation, tenure and tenure squared, indicators for enrollment in school or employment at the time of claim, dummy variables for local offices, some measures for the economic status of the client and past UI claims, and a measure for wages in the year prior to unemployment.[20] A supplementary analysis indicates the value of adding a simple measure of delayed filing.

Fifth, there is no improvement in the quality of assignment to services from estimating separate regional models in Kentucky, once regional dummy variables are included. If this result generalizes to other states, it is good news, because it keeps the estimation and the assignment of claimants clear and simple. The evidence in favor of generalizability is the large degree of heterogeneity within the Kentucky economy – arguably larger than in most other states. If separate regional models do not add much in such a heterogeneous state, it suggests that they would add even less in more economically homogeneous states. The evidence against it is that some other states appear to find regional models helpful based on their own analyses.

Sixth, the business cycle does affect model performance. Predictive performance decreases in booms and increases in downturns, probably because claimants become more heterogeneous during downturns. Our evidence on the cyclical behavior of the model is consistent with the view that profiling models should be re-estimated occasionally, but a definitive answer to the question of how often to re-estimate the model must await further analysis.

[20] In our preferred specification, the economic status indicators are welfare and food stamp recipiency, JTPA eligibility status, and public transport required for work. The public transport measure may not add as much in states that include cities such as New York or Chicago, where many white-collar workers use public transit. We include two measures of past UI receipt: previous UI claims and previous UI benefits exhausted. We also include a variable for shift type.

References

Balducchi, David (ed.) 1996. Worker Profiling and Reemployment Services (WPRS) Systems: National Colloquium, Selected Papers and Materials. Washington, DC: Employment and Training Administration, U.S. Department of Labor.

Berger, Mark, Dan Black, Amitabh Chandra, and Steve Allen. 1997. "Kentucky's Statistical Model of Worker Profiling for Unemployment Insurance." *Kentucky Journal of Economics and Business* 16: 1-18.

Black, Dan, Jeffrey Smith, Mark Berger, and Brett Noel. 2002. "Is the Threat of Reemployment Services More Effective than the Services Themselves? Experimental Evidence from the UI System." NBER Working Paper No. #8825.

Eberts, Randall, and Christopher O'Leary. 1996. "Design of the Worker Profiling and Reemployment Services System and Evaluation in Michigan," Research Working Paper #96-41. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Eberts, Randall, Christopher O'Leary and Wandner, Stephen (eds). 2002. *Targeting Employment Services*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Kelso, Marisa L. 1998. "Worker Profiling and Reemployment Services Profiling Methods: Lesson Learned." UI Occasional Paper 99-5. Washington, DC: U.S. Department of Labor.

OECD. 1998. *Early Identification of Jobseekers at Risk of Long-term Unemployment*. Paris: Organization for Economic Co-Operation and Development.

O'Leary, Christopher, Paul Decker and Stephen Wandner. 1998. "Reemployment Bonuses and Profiling." W.E. Upjohn Institute Working Paper #98-51. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

O'Leary, Christopher, Paul Decker, and Stephen Wandner. 2002. "Targeting Reemployment Bonuses." In Randall Eberts, Christopher O'Leary and Stephen Wandner, eds., *Targeting Employment Services*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. 161-182.

Wandner, Stephen. 1997. "Early Reemployment for Dislocated Workers in the United States." *International Social Security Review*. 50(4): 95-112.

Wandner, Stephen and Jon Messenger (eds.). 1999. "Worker Profiling and Reemployment Services Policy Workgroup: Final Report and Recommendations." Working Paper. Washington, DC: U.S. Department of Labor.

Worden, Kelley. 1993. "Profiling Dislocated Workers for Early Referral to Reemployment Services." unpublished manuscript. Washington, DC: U.S. Department of Labor.

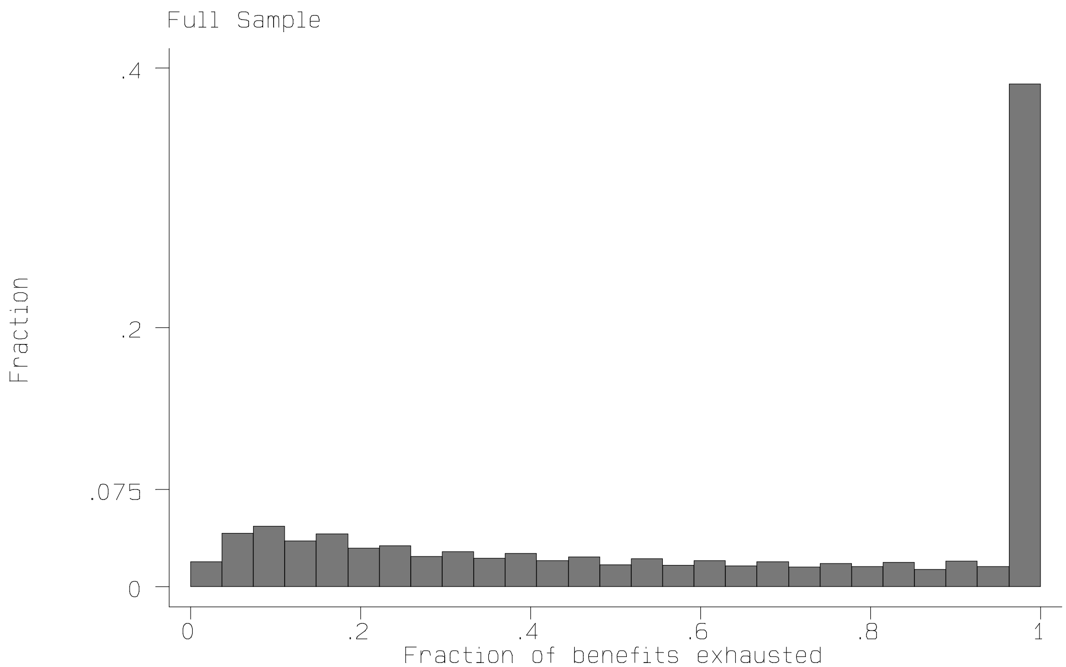

Figure 1: Fraction Benefits Exhausted in 1994

Note: Authors' calculations, Kentucky UI Claims data.

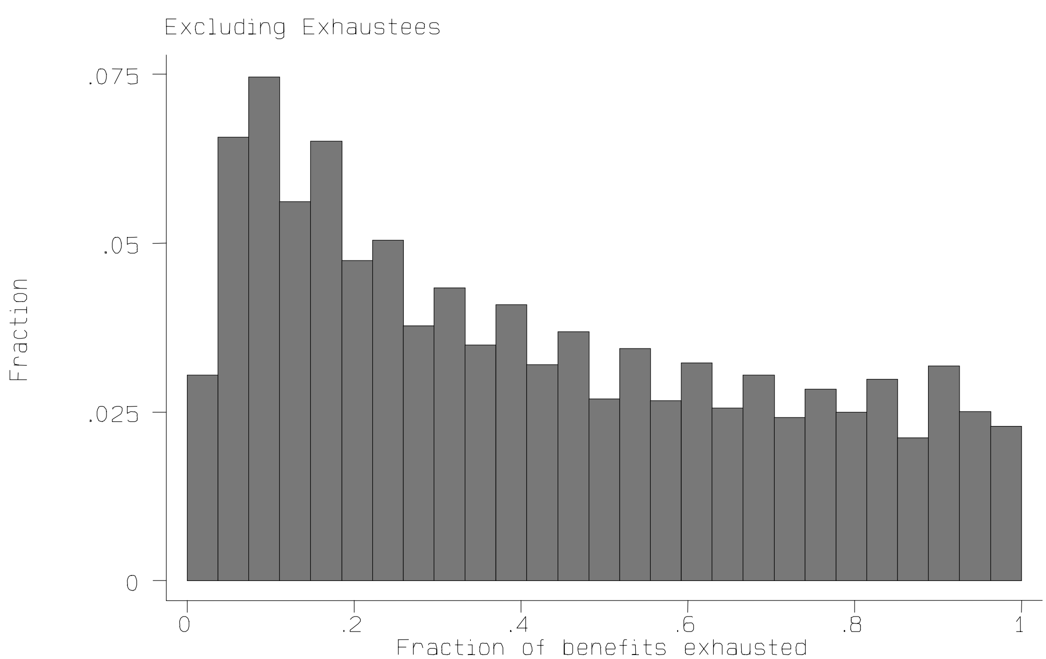

Figure 2: Fraction Benefits Exhausted in 1994

Note: Authors' calculations, Kentucky UI Claims data.

Table 1. WPRS model Covariates in the National, Maryland, and Kentucky Datasets.

	NATIONAL	WPRS	KENTUCKY
EDUCATION	*Categorical* Less than HS HS diploma Some college College	*Categorical* Less than HS HS diploma Some college Bachelors degree Masters degree/ Ph.D.	*Categorical* Less than HS HS diploma Some college Bachelors degree Masters degree/ Ph.D.
JOB TENURE	*Categorical* < 3 years 3 - 5 years 6 - 9 years > 10 years	*Continuous* Years of job tenure	*Continuous* Months of job tenure at main job (top coded at 99)
INDUSTRY Employment % change (monthly figures)	- SIC Division level - State level	- SIC Division level - SDA level	- SIC Division level (1 digit industry codes)
OCCUPATION	*Binary* 1 = growing 0 = declining or zero	*Categorical* (1-digit codes)	*Categorical* (1-digit codes)
UNEMPLOYMENT RATE (%) (monthly figures)	- State level	- SDA level	- SDA level

Source: Kelso (1998) for the national and Maryland data sets, and authors' manipulation of Kentucky data.

Notes:

SIC = Standard Industrial Classification
SDA = Service Delivery Area

Table 2. Profiling Variables

Variable	Used in our estimation on Kentucky Data?
"WPRS" model variables:	
Education variables	Yes
Job tenure	Yes
Industry (employment % change)	Yes
Occupation	Yes
Unemployment rate	Yes
Additional variables used by at least some states:[a]	
Weekly benefit amount	Yes
Wage replacement rate	Yes
Base year wage	Yes
Potential duration	*No*
Separation and claim filed dates	*No*
Ratio high quarter wage to base year wage	Yes
Number of base year employers	*No*
Month benefits began	Yes
Transfer payment recipiency	Yes
Pension recipiency	Yes
Claimant has phone	Yes
School enrollment	Yes
Separation from merger	Yes
Separation from plant closure	Yes
Worker has previous UI claim in recent past	Yes
Worker exhausted recent UI claim	Yes

Virtually all claimants in Kentucky data are eligible for 26 weeks of benefits and hence potential duration has no variation. In our sample, the separation date is too noisy to use, and we do not have precise information on the number of base year employers.

Additional variables included in our specifications:

Past UI claims	Average real dollar amount claimed in the past
	Number of weeks claimed in the past
	Average fraction of benefits exhausted in the past
	UI benefit exhaustion up to three year lags
Welfare indicators	Food stamps
	Welfare receipt
	JTPA eligibility
	Economically disadvantaged indicator
Other variables	Tenure squared
	Public transport required to get to work
	Type of shift worked
	Child support deductions

[a]Kelso (1998) and Berger et al. (1997).

Table 3. The Ability of Profiling Models to Predict UI Benefit Duration/Exhaustion

Fraction of Benefits Exhausted at the Top and Bottom of the Predicted Benefit Duration/Exhaustion Distribution for Various Profiling Models

	Top of Distribution	Bottom of Distribution	Difference
Pennsylvania model [a]	38.2	25.7	12.5
Washington model [a]	35.3	24.6	10.7
Kentucky model [b]	78.3	53.5	24.8
WPRS-type (Kentucky data, 1994) [b, c]	64.9	55.0	9.9
Preferred specification (Kentucky data, 1994) [b, d]	67.5	51.1	16.4

[a] The division for the Pennsylvania and Washington models is the top 25 percent predicted benefit exhaustion probabilities versus the bottom 75 percent. These are the maximum differences reported by O'Leary, Decker, and Wandner (1998).

[b] For Kentucky, the division is the top 60 percent of predicted benefit receipt durations and the remaining 40 percent.

[c] The WPRS-type specification reported here includes five covariates: education, occupation, tenure, local unemployment rate, and aggregate employment change in the claimant's industry.

[d] Our preferred specification, described as Model 36 in Table 7, includes education, occupation, tenure, plus a few more variables on past UI take-up, welfare recipiency and background.

Estimation and prediction for the Pennsylvania and Washington model are done using the same samples; the models for Kentucky and Maryland on Kentucky data are estimated on a 90 percent sub-sample of claimants with prediction based on the remaining 10 percent.

Table 4. Comparison of Alternative Functional Forms: OLS, Logit, Probit, and Tobit Predictive Performance from Estimation on Kentucky Data 1989-1995

Quintiles of Predicted Probability Distribution	DATA	RANDOM ASSIGNMENT
Q 5	0.997	0.637
Q 4	0.997	0.637
Q 3	0.742	0.636
Q 2	0.372	0.638
Q 1	0.102	0.638

DEPENDENT VARIABLE	EXHAUST			FRACTION	
Quintiles of Predicted Probability Distribution	OLS	LOGIT	PROBIT	OLS	TOBIT
Q 5	0.758	0.758	0.757	0.757	0.759
Q 4	0.699	0.698	0.698	0.696	0.697
Q 3	0.651	0.649	0.650	0.666	0.662
Q 2	0.624	0.625	0.625	0.617	0.617
Q 1	0.575	0.577	0.576	0.570	0.572

Differences Between Predicted Fractions of Benefits Exhausted

Top 80% - Bottom 20%	10.83	10.52	10.71	11.46	11.13
Top 60% - Bottom 40%	10.32	10.07	10.15	11.31	11.15
Top 40% - Bottom 60%	11.16	11.12	11.08	10.89	11.12
Top 20% - Bottom 80%	12.02	12.11	12.01	11.98	12.26
Average difference	11.09	10.95	10.99	11.41	11.42

Notes:
In the first panel we report the actual fraction of benefits exhausted by quintiles in the actual data and by randomly generated quintiles. In the second panel the fraction of benefits exhausted is reported by quintiles of the predicted benefit duration/probability to exhaust UI benefits. The third panel reports differences in predictive results across different percentiles of the distribution of predicted benefit duration/ exhaustion probabilities.

Explanatory variables

All models use the standard WPRS-type regressors: education, occupation, tenure, local unemployment rate (lagged) and employment changes in claimant's industry (lagged).

Dependent variables

EXHAUST is a discrete variable indicating whether unemployment benefits were exhausted (=1) or not (=0).
FRACTION is a continuous variable equal to the share of benefits claimed out of maximum allowed benefits.

Table 5. Comparison of Alternative Dependent Variables
Predictive Performance from OLS Estimation on Kentucky Data 1989-1995

Quintiles of Predicted Probability Distribution	FRACTION OF BENEFITS EXHAUSTED	EXHAUSTED BENEFITS	EXHAUSTED >= 90% BENEFITS	NUMBER OF WEEKS CLAIMED	AT LEAST 26 WEEKS CLAIMED
Q 5	0.757	0.758	0.760	0.756	0.754
Q 4	0.696	0.699	0.699	0.703	0.703
Q 3	0.666	0.651	0.651	0.659	0.648
Q 2	0.617	0.624	0.623	0.619	0.623
Q 1	0.570	0.575	0.574	0.570	0.578

Differences Between Predicted Fractions of Benefits Exhausted

Top 80% - Bottom 20%	11.46	10.83	10.93	11.44	10.45
Top 60% - Bottom 40%	11.31	10.32	10.52	11.15	10.14
Top 40% - Bottom 60%	10.89	11.16	11.38	11.38	11.23
Top 20% - Bottom 80%	11.98	12.02	12.38	11.89	11.61
Average difference	11.41	11.09	11.30	11.47	10.86

Notes:
In the first panel the fraction of benefits exhausted is reported by quintiles of the predicted benefit duration/probability to exhaust UI benefits. The second panel reports differences in predictive results across different percentiles of the distribution of predicted benefit duration/ exhaustion probabilities.

Explanatory variables

All models use the standard WPRS model regressors: education, occupation, tenure, local unemployment rate (lagged) and employment changes in claimant's industry of employment (lagged).

Table 6. Dropping Local Unemployment (UE) Rate and
Industry Employment (IE) Changes From the WPRS Model Specification
Predictive Performance from Estimation on Kentucky Data 1989-1995

Quintiles of Predicted Probability Distribution	A WPRS basic Model	B Adds both UE rate and IE changes	C Adds IE changes	D Adds UE rate	E Adds regional dummies	F Adds local office dummies
Q 5	0.743	0.757	0.747	0.757	0.758	0.764
Q 4	0.700	0.696	0.706	0.699	0.694	0.695
Q 3	0.662	0.666	0.660	0.662	0.667	0.661

Q 2	0.618	0.617	0.618	0.621	0.626	0.633
Q 1	0.585	0.570	0.577	0.568	0.564	0.557
Differences Between Predicted Fractions of Benefits Exhausted						
Top 80% - Bottom 20%	9.59	11.46	10.56	11.68	12.21	13.13
Top 60% - Bottom 40%	10.05	11.31	10.65	11.18	11.10	11.17
Top 40% - Bottom 60%	10.01	10.89	10.78	11.11	10.67	11.26
Top 20% - Bottom 80%	10.19	11.98	10.63	11.94	11.95	12.76
Average difference	9.96	11.41	10.65	11.48	11.48	12.08

Notes:
Linear model estimated by OLS. Dependent variable is fraction of benefits exhausted.
All models include the basic (reduced) WPRS regressors: education, occupation and tenure.
Each specification adds to the basic WPRS-type model variables as described below.

Model A:	Basic WPRS model specification: includes neither local unemployment rates nor industry employment changes.
Model B:	Includes both local unemployment rates and industry employment changes.
Model C:	Does not include local unemployment rates, but includes industry employment changes.
Model D:	Does not include industry employment changes, but includes local unemployment rates.
Model E:	Does not include unemployment and industry, but adds regional dummies.
Model F:	Does not include unemployment and industry, but adds local office dummies.

Table 7. Covariates Used in the 40 Model Specifications Analyzed for Kentucky

Model 0	**(WPRS)**	=	Education Occupation (1-digit codes) Tenure with last employer (and dummy for missing tenure)
Model 1	=	WPRS +	UI benefits exhausted last year UI benefits claimed last year
Model 2	=	WPRS +	UI benefits exhausted last year UI benefits claimed last year UI benefits exhausted two years ago UI benefits claimed two years ago
Model 3	=	WPRS +	UI benefits exhausted last year UI benefits claimed last year UI benefits exhausted two years ago UI benefits claimed two years ago UI benefits exhausted three years ago UI benefits claimed three years ago
Model 4	=	WPRS +	Indicator for previous UI benefit claims (in our sample)
Model 5	=	WPRS +	Average fraction of UI benefits exhausted in the past
Model 6	=	WPRS +	Average number of weeks claimed in the past
Model 7	=	WPRS +	Average UI benefit amount (real 1995 dollars) claimed in the past
Model 8	=	WPRS +	Welfare recipiency
Model 9	=	WPRS +	Food stamps recipiency
Model 10	=	WPRS +	Economically disadvantaged status
Model 11	=	WPRS +	JTPA eligibility
Model 12	=	WPRS +	Indicator for no phone
Model 13	=	WPRS +	Public transport for getting to work
Model 14	=	WPRS +	Quarterly wages within last year (real 1995 dollars)
Model 15	=	WPRS +	Tenure squared
Model 16	=	WPRS +	Type of working shift (shift 1,2,3, or rotating)
Model 17	=	WPRS +	Ratio of high quarter wage to average (base) year wage
Model 18	=	WPRS +	Weekly benefit amount (real 1995 dollars)
Model 19	=	WPRS +	Maximum benefit amount (real 1995 dollars)
Model 20	=	WPRS +	Wage replacement rule

Table 7. (cont.) Covariates Used in the 40 Model Specifications Analyzed for Kentucky

Model 21	**=**	WPRS +	Indicator for pension benefits recipiency
Model 22	=	WPRS +	Indicator for child support deductions
Model 23	=	WPRS +	Weekly amount of pension benefits received
Model 24	=	WPRS +	Month benefits began (continuous)
Model 25	=	WPRS +	Separation from plant closure
Model 26	**=**	WPRS +	Separation from merger
Model 27	=	WPRS +	Enrolled in school at the time of claim (+ dummy for missing enrollment)
Model 28	=	WPRS +	Employed at the time of claim (+ dummy for missing employment)
Model 29	=	WPRS +	Local office (32 dummies)
Model 30	=	WPRS +	Local office Claim filed out of state
Model 31	**=**	WPRS +	Indicator for previous UI benefit claim JTPA eligibility Quarterly wages within last year (real 1995 dollars) Tenure squared Employed at the time of claim (+ dummy for missing employment)
Model 32	=	WPRS +	Indicator for previous UI benefit claim JTPA eligibility Quarterly wages within last year (real 1995 dollars) Tenure squared Employed at the time of claim (+ dummy for missing employment) Local office
Model 33	=	WPRS +	UI benefits exhausted last year UI benefits exhausted two years ago Indicator for previous UI benefit claims Welfare recipiency Food stamps recipiency Public transport for getting to work Quarterly wages within last year (real 1995 dollars) Tenure squared Enrolled in school at the time of claim (+ dummy for missing enrollment) Employed at the time of claim (+ dummy for missing employment)
Model 34	=	WPRS +	UI benefits exhausted last year UI benefits exhausted two years ago Indicator for previous UI benefit claims Welfare recipiency Food stamps recipiency Public transport for getting to work Quarterly wages within last year (real 1995 dollars) Tenure squared Enrolled in school at the time of claim (+ dummy for missing enrollment) Employed at the time of claim (+ dummy for missing employment) Local office

Table 7. (cont.) Covariates Used in the 40 Model Specifications Analyzed for Kentucky

Model 35 = WPRS + UI benefits exhausted last year
Indicator for previous UI benefit claims
Welfare recipiency
Food stamps recipiency
Public transport for getting to work
JTPA eligibility
Quarterly wages within last year (real 1995 dollars)
Tenure squared
Enrolled in school at the time of claim (+ dummy for missing enrollment)
Employed at the time of claim (+ dummy for missing employment)
Type of working shift (shift 1,2,3, or rotating)

Model 36 = WPRS + UI benefits exhausted last year
Indicator for previous UI benefit claims
Welfare recipiency
Food stamps recipiency
Public transport for getting to work
JTPA eligibility
Quarterly wages within last year (real 1995 dollars)
Tenure squared
Enrolled in school at the time of claim (+ dummy for missing enrollment)
Employed at the time of claim (+ dummy for missing employment)
Type of working shift (shift 1,2,3, or rotating)
Local office

Model 37 = WPRS + UI benefits exhausted last year
Welfare recipiency
Food stamp recipiency
JTPA eligibility
High quarter wage to average (base) year wage
Quarterly wages within last year (real 1995 dollars)
Tenure squared
Enrolled in school at the time of claim (+ dummy for missing enrollment)
Employed at the time of claim (+ dummy for missing employment)

Model 38 = WPRS + UI benefits exhausted last year
Welfare recipiency
Food stamp recipiency
JTPA eligibility
High quarter wage to average (base) year wage
Quarterly wages within last year (real 1995 dollars)
Tenure squared
Enrolled in school at the time of claim (+ dummy for missing enrollment)
Employed at the time of claim (+ dummy for missing employment)
Local office

Model 39 = WPRS + Indicator for previous UI benefit claims
Average fraction claimed in the pas
UI benefits exhausted last year
UI benefits exhausted two years ago
Quarterly wages within last year
Weekly pension benefits
Weekly child support deductions
Tenure squared

	UI benefits exhausted three years ago	Enrolled in school at
the time of claim		
	Welfare recipiency	Employed at the
time of claim		
	Food stamps recipiency	Separation from
plant closure		
	JTPA eligibility	Separation from
merger		
	Indicator for economically disadvantaged	Month benefits
began		
	Indicator for no phone	Type of work shift
(1,2,3, rotating)		
	Indicator for public transport	Local office
	Maximum benefits	Claim filed out of
state		

Table 8. Predictive Performance from Different Model Specifications
Linear Model Estimated by OLS on 1994 Kentucky Data
The dependent variable is fraction of UI benefits exhausted.

Quintiles of Predicted Probabilities		**DATA**			**RANDOM ASSIGNMENT**			
Q 5		0.998			0.629			
Q 4		0.998			0.636			
Q 3		0.721			0.631			
Q 2		0.344			0.629			
Q 1		0.101			0.635			
MODEL	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Q 5	0.681	0.683	0.684	0.680	0.677	0.677	0.680	0.679
Q 4	0.643	0.639	0.643	0.647	0.651	0.645	0.641	0.641
Q 3	0.600	0.603	0.599	0.603	0.590	0.605	0.608	0.603
Q 2	0.597	0.596	0.599	0.595	0.610	0.600	0.600	0.602
Q 1	0.527	0.528	0.524	0.524	0.521	0.522	0.521	0.523
Differences Between Predicted Fractions of Benefits Exhausted								
Top 80% - Bottom 20%	10.33	10.24	10.68	10.68	11.12	11.00	11.10	10.82
Top 60% - Bottom 40%	7.98	7.96	8.01	8.35	7.42	8.15	8.25	7.87
Top 40% - Bottom 60%	8.78	8.53	8.94	8.94	9.04	8.52	8.42	8.40
Top 20% - Bottom 80%	8.95	9.14	9.27	8.75	8.35	8.39	8.76	8.64
Average Difference	**9.01**	**8.97**	**9.22**	**9.18**	**8.98**	**9.01**	**9.13**	**8.93**
R-squared	0.031	0.031	0.031	0.032	0.031	0.031	0.031	0.031
MODEL	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Q 5	0.679	0.686	0.681	0.682	0.681	0.682	0.677	0.681
Q 4	0.641	0.642	0.637	0.644	0.643	0.644	0.645	0.644
Q 3	0.605	0.600	0.604	0.602	0.600	0.597	0.616	0.592
Q 2	0.598	0.595	0.609	0.588	0.597	0.601	0.585	0.607
Q 1	0.525	0.525	0.518	0.535	0.527	0.524	0.525	0.523
Differences Between Predicted Fractions of Benefits Exhausted								
Top 80% - Bottom 20%	10.60	10.61	11.44	9.37	10.32	10.70	10.58	10.83
Top 60% - Bottom 40%	8.05	8.27	7.71	8.10	7.98	7.83	9.13	7.40
Top 40% - Bottom 60%	8.40	9.06	8.20	8.77	8.77	8.89	8.56	8.88
Top 20% - Bottom 80%	8.71	9.53	8.88	8.96	8.96	9.03	8.47	8.99
Average Difference	**8.94**	**9.37**	**9.06**	**8.80**	**9.01**	**9.11**	**9.18**	**9.03**
R-squared	0.031	0.032	0.031	0.033	0.031	0.031	0.033	0.031

Table 8. (cont.) Predictive Performance from Different Model Specifications
Linear Model Estimated by OLS on 1994 Kentucky Data
The dependent variable is fraction of UI benefits exhausted.

MODEL	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
Q 5	0.674	0.674	0.675	0.675	0.678	0.681	0.683	0.681
Q 4	0.652	0.648	0.652	0.652	0.650	0.644	0.640	0.644
Q 3	0.589	0.612	0.593	0.593	0.609	0.599	0.604	0.599
Q 2	0.610	0.591	0.616	0.616	0.597	0.598	0.593	0.598
Q 1	0.521	0.523	0.512	0.512	0.528	0.527	0.528	0.527
Differences Between Predicted Fractions of Benefits Exhausted								
Top 80% - Bottom 20%	10.98	10.77	12.19	12.19	10.52	10.33	10.24	10.33
Top 60% - Bottom 40%	7.29	8.74	7.61	7.61	8.31	7.91	8.20	7.91
Top 40% - Bottom 60%	9.01	8.57	9.04	9.04	8.61	8.78	8.64	8.78
Top 20% - Bottom 80%	8.14	8.10	8.20	8.20	8.24	8.92	9.14	8.95
Average difference	**8.86**	**9.05**	**9.26**	**9.26**	**8.92**	**8.99**	**9.06**	**8.99**
R-squared	0.031	0.032	0.039	0.039	0.032	0.031	0.032	0.031

MODEL	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
Q 5	0.684	0.681	0.680	0.705	0.703	0.707	0.707	0.716
Q 4	0.637	0.644	0.645	0.651	0.653	0.631	0.631	0.661
Q 3	0.602	0.599	0.598	0.642	0.633	0.620	0.620	0.630
Q 2	0.600	0.599	0.599	0.561	0.577	0.566	0.566	0.568
Q 1	0.526	0.526	0.527	0.490	0.482	0.523	0.525	0.473
Differences Between Predicted Fractions of Benefits Exhausted								
Top 80% - Bottom 20%	10.50	10.49	10.41	15.00	16.03	10.77	10.61	17.12
Top 60% - Bottom 40%	7.83	7.87	7.80	14.05	13.41	10.78	10.72	14.84
Top 40% - Bottom 60%	8.43	8.76	8.81	11.39	11.45	9.89	9.90	13.16
Top 20% - Bottom 80%	9.26	8.88	8.83	11.89	11.72	12.16	12.17	13.29
Average difference	**9.01**	**9.00**	**8.96**	**13.08**	**13.15**	**10.90**	**10.85**	**14.60**
R-squared	0.031	0.031	0.031	0.096	0.099	0.049	0.049	0.110

MODEL	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
Q 5	0.737	0.731	0.752	0.726	0.741	0.724	0.754	0.749
Q 4	0.676	0.664	0.677	0.660	0.684	0.650	0.646	0.680
Q 3	0.597	0.631	0.587	0.639	0.600	0.635	0.613	0.593
Q 2	0.570	0.540	0.558	0.550	0.560	0.554	0.564	0.568
Q 1	0.468	0.482	0.474	0.473	0.463	0.485	0.470	0.458
Differences Between Predicted Fractions of Benefits Exhausted								
Top 80% - Bottom 20%	17.71	15.92	16.98	17.07	18.31	15.59	17.41	18.94
Top 60% - Bottom 40%	15.09	16.43	15.62	16.37	16.38	15.05	15.37	16.07
Top 40% - Bottom 60%	16.15	14.63	17.52	13.94	17.20	12.93	15.09	17.48
Top 20% - Bottom 80%	15.94	15.19	17.79	14.57	16.48	14.31	18.03	17.41

Average difference	**16.22**	**15.54**	**16.98**	**15.49**	**17.09**	**14.47**	**16.47**	**17.47**
R-squared	0.128	0.113	0.130	0.116	0.134	0.107	0.124	0.139

Table 9. Regional Analysis Results
from Linear Model with Model 36 Covariates Estimated by OLS on 1994 Kentucky Data
Eight Models – One Estimated Separately for Each of the Eight Regions
One Overall Ranking – Based on Predictions from the Eight Separate Regressions

Quintiles of Predicted Probability Distribution	Overall Ranking	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7	Region 8
Q 5	0.738	0.742	0.685	0.669	0.719	0.787	0.773	0.784	0.743
Q 4	0.683	0.709	0.652	0.635	0.660	0.768	0.736	0.693	0.619
Q 3	0.597	0.609	0.605	0.589	0.550	0.678	0.684	0.656	0.578
Q 2	0.558	0.517	0.613	0.578	0.528	0.553	0.665	0.594	0.537
Q 1	0.472	0.429	0.538	0.474	0.441	0.499	0.494	0.546	0.413
Differences Between Predicted Fractions of Benefits Exhausted									
Top 80% - Bottom 20%	17.17	21.53	10.11	14.38	17.36	19.78	22.06	13.60	20.57
Top 60% - Bottom 40%	15.78	21.37	7.23	10.53	15.87	21.82	15.15	14.13	17.12
Top 40% - Bottom 60%	16.85	20.72	8.36	10.54	18.32	20.06	14.02	13.99	17.10
Top 20% - Bottom 80%	16.05	17.58	8.29	9.99	17.45	16.19	12.80	16.14	20.59
Average difference	**16.46**	**20.30**	**8.50**	**11.36**	**17.25**	**19.46**	**16.01**	**14.47**	**18.85**

One Model – Estimated Jointly for All Eight Regions, Including Regional Dummy Variables

Quintiles of Predicted Probability Distribution	Overall Ranking	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7	Region 8
Q 5	0.741	0.740	0.709	0.656	0.693	0.793	0.729	0.771	0.741
Q 4	0.673	0.709	0.618	0.638	0.676	0.733	0.758	0.737	0.645
Q 3	0.606	0.623	0.639	0.617	0.642	0.680	0.710	0.609	0.536
Q 2	0.557	0.517	0.570	0.553	0.462	0.558	0.624	0.635	0.545
Q 1	0.471	0.418	0.556	0.480	0.425	0.521	0.530	0.521	0.423
Differences Between Predicted Fractions of Benefits Exhausted									
Top 80% - Bottom 20%	17.36	22.91	7.77	13.63	19.27	17.03	17.54	16.63	19.44
Top 60% - Bottom 40%	15.95	22.31	9.22	12.08	22.68	19.63	15.51	12.75	15.68
Top 40% - Bottom 60%	16.27	20.55	7.50	9.69	17.49	17.68	12.20	16.56	19.18
Top 20% - Bottom 80%	16.44	17.34	11.27	8.38	14.18	16.96	7.33	14.53	20.39
Average difference	**16.51**	**20.78**	**8.94**	**10.94**	**18.41**	**17.82**	**13.14**	**15.12**	**18.67**

Table 10. Year-by-year Predictive Performance from Linear Models Estimated by OLS
on 1989-1995 Kentucky Data

	Assignment Mechanism	Quintiles of Predicted Probability Distribution	ALL	YEARS 1989	1990	1991	1992	1993	1994	1995	1989 To 1992	1993 To 1995
		Q 5	0.638	0.617	0.681	0.668	0.661	0.576	0.629	0.590	0.658	0.600
1).	Random	Q 4	0.637	0.618	0.673	0.666	0.655	0.572	0.636	0.583	0.657	0.597
	Assignment	Q 3	0.636	0.619	0.674	0.673	0.661	0.574	0.631	0.579	0.657	0.598
		Q 2	0.637	0.620	0.676	0.671	0.661	0.571	0.629	0.580	0.657	0.594

	Q 1	0.638	0.620	0.679	0.670	0.658	0.581	0.635	0.588	0.660	0.603
	Q 5	0.997	0.993	0.946	0.947	0.939	0.956	0.998	0.957	0.934	0.974
2). DATA	Q 4	0.997	0.993	0.946	0.947	0.939	0.956	0.998	0.957	0.934	0.974
	Q 3	0.742	0.701	0.946	0.947	0.939	0.577	0.721	0.648	0.934	0.636
	Q 2	0.372	0.346	0.431	0.426	0.392	0.307	0.344	0.315	0.406	0.329
	Q 1	0.102	0.092	0.116	0.117	0.101	0.099	0.101	0.095	0.113	0.100
	Q 5	0.781	0.745	0.882	0.820	0.801	0.803	0.741	0.714	0.803	0.744
3). MODEL 36	Q 4	0.712	0.686	0.741	0.755	0.729	0.772	0.684	0.655	0.729	0.678
BEST SPECIFICATION	Q 3	0.665	0.641	0.685	0.729	0.693	0.675	0.600	0.625	0.689	0.649
	Q 2	0.616	0.619	0.612	0.649	0.644	0.591	0.560	0.567	0.630	0.597
	Q 1	0.533	0.540	0.543	0.464	0.532	0.482	0.463	0.508	0.529	0.508
	Differences Between Predicted Fractions of Benefits Exhausted										
	Top 80% - Bottom 20%	16.02	13.30	18.69	27.39	18.45	22.80	18.31	13.23	18.34	15.86
	Top 60% - Bottom 40%	14.46	11.14	19.19	21.14	15.31	21.34	16.38	12.71	16.06	13.73
	Top 40% - Bottom 60%	14.16	11.58	19.84	17.35	14.25	20.48	17.20	11.78	14.99	12.57
	Top 20% - Bottom 80%	14.97	12.37	23.73	17.10	15.19	17.32	16.48	12.52	15.88	13.55
	Average difference	**14.90**	**12.10**	**20.36**	**20.74**	**15.80**	**20.49**	**17.09**	**12.56**	**16.32**	**13.93**

Table 10. (cont.) Year-by-year Predictive Performance from Linear Models Estimated by OLS on 1989-1995 Kentucky Data

Assignment Mechanism	Quintiles of Predicted Probability Distribution	ALL	YEARS 1989	1990	1991	1992	1993	1994	1995	1989 To 1992	1993 To 1995
	Q 5	0.743	0.681	0.801	0.784	0.761	0.706	0.681	0.663	0.760	0.682
WPRS	Q 4	0.700	0.657	0.765	0.723	0.702	0.700	0.643	0.656	0.720	0.639
4). MODEL	Q 3	0.662	0.648	0.727	0.724	0.678	0.682	0.600	0.602	0.693	0.626
BASE	Q 2	0.618	0.628	0.596	0.695	0.649	0.671	0.597	0.569	0.656	0.634
SPECIFICATION	Q 1	0.585	0.628	0.596	0.496	0.610	0.566	0.527	0.582	0.579	0.595
Differences Between Predicted Fractions of Benefits Exhausted											
Top 80% - Bottom 20%		9.59	2.55	12.64	23.54	8.76	12.33	10.33	4.10	12.77	5.03
Top 60% - Bottom 40%		10.05	3.40	16.85	14.78	8.46	7.74	7.98	6.48	10.64	3.44
Top 40% - Bottom 60%		10.01	3.43	14.34	11.51	8.59	6.35	8.78	7.49	9.70	4.20
Top 20% - Bottom 80%		10.19	4.06	13.05	12.44	10.18	5.12	8.95	6.08	9.80	5.89
Average difference		**9.96**	**3.36**	**14.22**	**15.57**	**9.00**	**7.89**	**9.01**	**6.04**	**10.73**	**4.64**

Notes:

For each of the estimated quintiles we report the fraction of benefits exhausted by claimants in that quintile. For the WPRS and best model specifications we also report differences between the predictions for the top 80 percent and bottom 20 percent , for the top 60 percent and bottom 40 percent, for the top 40 percent and bottom 60 percent, and for the top 20 percent and bottom 80 percent of the predicted values.

1. The random assignment mechanism forms quintiles at random.
2. The data distribution is computed using on the distribution of fraction of exhausted UI benefits in the raw data.
3. The best model specification is Model 36 described in Table 7.
4. The WPRS model base specification uses as covariates education, tenure, and 1-digit occupation codes.
   Year dummies are included in the models estimated using the pooled data (All, 1989 to 1992 and 1993 to 1995) .

Table 11: Definitions of Measures of Delayed Filing

Categorical measures of delay

A	0: # Days Elapsed <=30 1: # Days Elapsed > 30
B	1: 0 <= # Days elapsed <=14 2: 15 <= # Days elapsed <=30 3: 31 <= # Days elapsed <=60 4: 61 <= # Days elapsed <=180 5: 181<= # Days elapsed
C	1: 0 <= # Weeks elapsed <=1 2: 2 <= # Weeks elapsed <=6 3: 7 <= # Weeks elapsed
D	1: 0 <= # Days elapsed <=10 2: 11 <= # Days elapsed <=15 3: 16 <= # Days elapsed <=30 4: 1 <= # Months elapsed <=2 5: 3 <= # Months elapsed
E	1: 0 <= # Weeks elapsed <= 1 2: 2 = # Weeks elapsed 3: 3 <= # Weeks elapsed <= 4 4: 5 <= # Weeks elapsed <= 6 5: 7 <= # Weeks elapsed <= 8 6: 9 <= # Weeks elapsed
F	1: # Months elapsed = 0 2: # Months elapsed = 1 or 2 3: # Months elapsed >= 3

Continuous measures of delay

G	# Days elapsed
H	# Weeks elapsed
	# Months elapsed

K	

Table 12. Predictive Performance from Adding Delayed Filing Measures.
Four Selected Model Specifications. OLS Estimation of Linear Model on 1994 Kentucky Data
(Delay = Time Gap Between Losing a Job and Filing for UI benefits)

Model 0 (WPRS)

		Categorical measures of delay						**Continuous measures of delay**		
	Model 0	A	B	C	D	E	F	G	H	K
Q5	0.681	0.695	0.695	0.691	0.700	0.693	0.700	0.691	0.691	0.694
Q4	0.643	0.643	0.648	0.649	0.641	0.653	0.643	0.653	0.651	0.653
Q3	0.600	0.609	0.607	0.610	0.610	0.600	0.606	0.602	0.606	0.593
Q2	0.597	0.561	0.556	0.558	0.556	0.561	0.560	0.563	0.561	0.567
Q1	0.527	0.540	0.542	0.541	0.542	0.542	0.541	0.540	0.540	0.541
Differences Between Predicted Fractions of Benefits Exhausted										
Top 80% - Bottom 20%	10.33	8.78	8.45	8.61	8.44	8.45	8.61	8.75	8.75	8.55
Top 60% - Bottom 40%	7.98	9.88	10.07	10.05	10.09	9.71	9.92	9.76	9.88	9.24
Top 40% - Bottom 60%	8.78	9.93	10.33	10.01	10.10	10.55	10.22	10.37	10.22	10.60
Top 20% - Bottom 80%	8.95	10.69	10.67	10.11	11.27	10.42	11.23	10.10	10.12	10.48
Average difference	**9.01**	**9.82**	**9.88**	**9.69**	**9.97**	**9.79**	**10.00**	**9.75**	**9.75**	**9.72**
R-squared	0.031	0.044	0.044	0.044	0.045	0.044	0.045	0.044	0.044	0.044

Model 1

		Categorical measures of delay						**Continuous measures of delay**		
	Model 1	A	B	C	D	E	F	G	H	K
Q5	0.683	0.693	0.693	0.690	0.700	0.692	0.700	0.691	0.693	0.694
Q4	0.639	0.642	0.648	0.648	0.639	0.653	0.642	0.651	0.649	0.649
Q3	0.603	0.612	0.608	0.611	0.611	0.600	0.606	0.605	0.605	0.596
Q2	0.596	0.562	0.557	0.558	0.556	0.561	0.560	0.560	0.561	0.568
Q1	0.528	0.540	0.542	0.541	0.542	0.542	0.541	0.541	0.541	0.541
Differences Between Predicted Fractions of Benefits Exhausted										
Top 80% - Bottom 20%	10.24	8.69	8.45	8.59	8.43	8.45	8.61	8.63	8.64	8.53
Top 60% - Bottom 40%	7.96	9.80	10.05	10.05	10.10	9.68	9.90	9.86	9.83	9.17
Top 40% - Bottom 60%	8.53	9.64	10.16	9.92	10.00	10.49	10.24	10.29	10.25	10.33
Top 20% - Bottom 80%	9.14	10.47	10.45	10.07	11.29	10.29	11.29	10.20	10.48	10.54
Av. Dif.	**8.97**	**9.65**	**9.78**	**9.66**	**9.96**	**9.73**	**10.01**	**9.75**	**9.80**	**9.64**
R-squared	0.031	0.044	0.044	0.044	0.045	0.044	0.045	0.044	0.044	0.044

Table 12 (cont.) Predictive Performance from Adding Delayed Filing Measures.
Four Selected Model Specifications. OLS Estimation of Linear Model on 1994 Kentucky Data
(Delay = Time Gap Between Losing a Job and Filing for UI benefits)

Model 36 (Best specification)

		Categorical measures of delay						**Continuous measures of delay**		
	Model 36	A	B	C	D	E	F	G	H	K
Q5	0.741	0.746	0.742	0.744	0.744	0.740	0.743	0.751	0.749	0.746
Q4	0.684	0.674	0.681	0.675	0.678	0.679	0.675	0.674	0.675	0.678
Q3	0.600	0.592	0.591	0.597	0.595	0.597	0.597	0.592	0.594	0.590
Q2	0.560	0.580	0.578	0.573	0.578	0.576	0.578	0.575	0.574	0.574
Q1	0.463	0.456	0.456	0.459	0.453	0.457	0.455	0.456	0.456	0.460
Differences Between Predicted Fractions of Benefits Exhausted										
Top 80% - Bottom 20%	18.31	19.24	19.22	18.88	19.61	19.13	19.33	19.23	19.16	18.69
Top 60% - Bottom 40%	16.38	15.29	15.46	15.60	15.68	15.59	15.51	15.69	15.75	15.38
Top 40% - Bottom 60%	17.20	16.73	17.02	16.64	16.89	16.64	16.54	17.12	17.06	17.04
Top 20% - Bottom 80%	16.48	17.08	16.55	16.81	16.75	16.28	16.66	17.65	17.44	17.02
Av. Dif.	**17.09**	**17.09**	**17.06**	**16.98**	**17.23**	**16.91**	**17.01**	**17.42**	**17.35**	**17.03**
R-squared	0.134	0.146	0.146	0.145	0.146	0.146	0.146	0.145	0.145	0.145

Model 39 (Most covariates)

		Categorical measures of delay						**Continuous measures of delay**		
	Model 39	A	B	C	D	E	F	G	H	K
Q5	0.749	0.762	0.758	0.762	0.764	0.759	0.768	0.751	0.749	0.746
Q4	0.680	0.663	0.665	0.657	0.656	0.667	0.653	0.674	0.675	0.678
Q3	0.593	0.597	0.599	0.605	0.604	0.600	0.603	0.592	0.594	0.590
Q2	0.568	0.576	0.576	0.574	0.573	0.572	0.575	0.575	0.574	0.574
Q1	0.458	0.450	0.450	0.450	0.451	0.450	0.449	0.456	0.456	0.460
Differences Between Predicted Fractions of Benefits Exhausted										
Top 80% - Bottom 20%	18.94	20.00	19.94	19.89	19.82	19.99	20.11	19.23	19.16	18.69
Top 60% - Bottom 40%	16.07	16.13	16.11	16.24	16.26	16.44	16.27	15.69	15.75	15.38
Top 40% - Bottom 60%	17.48	17.21	17.03	16.59	16.75	17.25	16.85	17.12	17.06	17.04
Top 20% - Bottom 80%	17.41	19.10	18.59	18.99	19.28	18.67	19.80	17.65	17.44	17.02
Av. Dif.	**17.47**	**18.11**	**17.91**	**17.93**	**18.03**	**18.09**	**18.26**	**17.42**	**17.35**	**17.03**
R-squared	0.139	0.151	0.152	0.151	0.152	0.152	0.152	0.151	0.151	0.151