

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.