Welcome

Please help yourself to breakfast.
Agenda:

8:00-8:45am: Registration & Breakfast
8:45-8:55am: Welcome
8:55-9:25am: Daniel Zabronsky, Deputy Associate Commissioner, SSA
9:25-9:55am: Q&A and Discussion
9:55-10:00am: Closing
Predictive Modeling to Support Program Integrity at the Social Security Administration

February 12, 2019
Organizational Overview of SSA’s Office of Analytics, Review and Oversight
Key OAI Areas of Expertise

- Predictive modeling
- Data analytics
- Natural Language Processing
- Application development
- Business process modeling and improvement
- Automated case processing
- Special studies and evaluations
- Agency internal surveys
- Enterprise wide program management
Office of Analytics and Improvements

5-Step Approach to Modeling Methodology

➢ Objective identification
  Most important step

➢ Data development
  Step that takes the most time

➢ Variable selection

Steps that most people focus on

➢ Model estimation

➢ Performance assessment
Office of Analytics and Improvements

Define Model Objectives
- Know Business Issues
- Know Tools
- Know Data Infrastructure

Clean Data
- Learn Data and its properties
- Apply integrity rules, standardize, missing values
- Time-Properties

Create Model Development Datasets
- Test 20%
- Validate 10%
- Train 70%

Variable Selection
- Traditional Approaches
- Data Mining Approaches

Model Estimation
- Regression
- Decision Tree
- Neural Networks

Performance Assessment
- Out-of-Sample Performance assessment using Test data

Based on out-of-sample performance identify additional variables
Based on out-of-sample performance re-estimate model
Out-of-Sample Performance assessment using Validate data
Key Similarities between OAI Predictive Models that Support SSA Program Integrity

➢ Estimated using logistic regression approach (standard approach used to model binary rare events)

➢ Estimated using SAS software

➢ Use SSA legacy systems data at the individual record level

➢ Periodically re-estimated

➢ Estimated with data that contain various limitations

➢ Operate in a sensitive political environment

➢ Large magnitude of impact involved in any major model changes
Key Differences between OAI Predictive Models

- Model objectives (dependent variable) differ
- Models are estimated on different populations
- Some models are stand-alone applications; some are integrated within SSA legacy Systems
- Models use different data files
- Models were estimated at different times
- Models are re-estimated on different schedules
Keys to Successful Predictive Model Design, Development, and Implementation at SSA

➢ Knowledge of analytical tools & techniques

➢ Knowledge of agency data infrastructure

➢ Knowledge of agency business processes, policies, and procedures

➢ Close collaboration with operational, policy, and systems business partners
OAI Predictive Modeling Case Study

Predictive Modeling Case Study: SSI Redetermination Model
SSI Program Overview

➢ Supplemental Security Income (SSI) is a Federal income supplement program administered by SSA
  • Funded by general tax revenue (not OASDI taxes)
  • Provides cash to meet basic needs for food, clothing and shelter
  • ~$50 billion in benefits paid each year to ~8 million beneficiaries

➢ Two key eligibility requirements
  • Beneficiary must be elderly, disabled, or blind
  • Beneficiary must not exceed income and asset criteria

➢ Each eligibility requirement has an associated program integrity process
  • Medical—Continuing Disability Reviews (CDRs)
  • Financial—SSI Redeterminations (RZs)
SSI Redetermination Scoring Model

- SSA is required to conduct SSI redeterminations which are reviews of SSI beneficiaries continued eligibility for SSI payments based on income, resources, and living arrangements.

- The SSI Redetermination Scoring Model allows SSA to efficiently prioritize SSI redeterminations to ensure that the most productive and cost effective reviews are the highest priority reviews.
SSI Redetermination Scoring Model

SSI Redetermination Scoring Model Key Facts

➢ The SSI Redetermination scoring model was first implemented in the late 1970’s and is continually monitored and periodically re-estimated. The models cover the entire SSI population.

➢ The model is estimated through logistic regression and generalized linear modelling using historical data on SSI redetermination completions with the dollar amount of predicted overpayment as the dependent variable.

➢ The current model contains about sixty independent variables. Key independent variables include SSI beneficiary income, resource, and living arrangement variables.
SSI Redetermination Scoring Model

➢ The SSIRM model is a two-stage model. The first step, which estimates the probability of SSI overpayment standard logistic regression model:

\[(P) \text{ OP} = f(\text{indvar1, indvar2, indvar3, indvar4, etc...})\]

➢ The second step, which estimates the predicted dollar amount of SSI is standard Ordinary Least Squares (OLS) model:

\[\text{OPAMT} = f(\text{indvar1, indvar2, indvar3, indvar4, etc...})\]

➢ For each case on the SSI rolls, step one and step two results are combined to get predicted SSI overpayments for each case:

\[\text{OP} = (P) \text{ OP} \times \text{OPAMT}\]
SSI Redetermination Scoring Model

➢ The data used to estimated the SSIRM model is the SSI Chang Rate Study data

➢ SSI Change Rate Study data consists of about 25,000 SSI cases of a randomly selected sample of the SSI universe. These Change Rate cases undergo a redetermination and overpayment and underpayment dollar amounts are computed for each case based on the results of the completed redetermination.

➢ The SSIRM is re-estimated each year with new SSI Change Rate Study Data
SSI Redetermination Scoring Model

The SSIRM model contains three types of variables typically found in predictive models:

➢ Binary or Categorical Variables
   ▪ Can only take on a value of zero or one
   ▪ Example: Eligibility for other income maintenance programs

➢ Discrete Variables
   ▪ Can take on a fixed range of values
   ▪ Example: Number of prior redeterminations

➢ Continuous Variables
   ▪ Can take on a value of any real number (within reason)
   ▪ Example: Other program payment amounts
SSI Redetermination Scoring Model

➢ The model is run in the SSI payment system and cases are prioritized by model score for case selection for an SSI redetermination

➢ Software within the SSI payment system generates case selection diaries that alert the SSA field office to initiate the SSI redetermination

➢ Other types of SSI redeterminations not selected by the model occur each year; these include unscheduled reviews, age 18 redeterminations and Limited Issue redeterminations

➢ Software within the SSI payment system includes statistical sampling to select cases for the Change Rate Study data set used for model re-estimation and ongoing model validation and assessment
OAI Model Monitoring, Validation & Assessment

➢ A key component of the OAI predictive models is the implementation of a continuous monitoring approach to ensure ongoing Model estimation, validation and performance assessment

▪ Annual integrity samples are developed and implemented to ensure sufficient data to both provide model validation and support periodic model re-estimation

▪ Statistical performance assessment are undertaken to provide continuous monitoring of model accuracy and reliability

▪ OAI engages in analytical research to assist with ongoing enhancements and improvements to the models

▪ OAI is heavily involved in the Agency’s IT Modernization efforts to continue to strengthen the data used for analytics modeling purposes
Questions?