Advanced Analytics for Audit Case Selection

Presented by:
Tim Gardinier, Manager, Data Processing Services

Audit Case Selection

Putting a scientific approach to selection into the Auditors hands
Goals of this Presentation

- Gain an understanding of how TACS got started, the theory of normal taxpayers within peer groups and outliers, and how models and features are designed.
- Gain an understanding of peer group selection and the process for model creation.
- Learn how "values" and "scores" are created.
- Understand how to interpret these values and scores using various graphical presentations of the data along with canned reports and how an understanding of these scores leads to the selection of an audit case.
- Learn how TACS was further developed for scoring and ranking pre-processed returns.
- Understand the term CISS.

The project business objective was to maximize the return of the DTF investment in tax return audit programs.

- Target tax returns with greatest potential payback.
- Provide information to support investigation, litigation, and settlement functions.
- Provide information to counter/prevent loss in the future.
The Audit Division wanted a better way to select cases and manage and/or prioritize their inventory

- Reduce screening time by focusing on attributes of non-compliant returns
- Identify characteristics of successful and unsuccessful audits
- Target tax returns that resemble those that historically produced successful audits
- Assist with assignment and management of cases by estimating...
  - case complexity
  - confidence in the selection
  - additional revenue

To accomplish these objectives NY realized it needed to take a more scientific approach to case selection

<table>
<thead>
<tr>
<th>Simple Edits &amp; Matching</th>
<th>Integrated Case Management</th>
<th>Compliance Data Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Warehouse &amp; Query Tools</td>
<td>Automated Case Creation (desk and field)</td>
<td>More scientific data driven approach to case selection</td>
</tr>
<tr>
<td>Data from internal systems</td>
<td>Link to data warehouse</td>
<td>Scoring, ranking, and classifying taxpayers based on behavior</td>
</tr>
<tr>
<td>Limited external data</td>
<td>Reliant on IT</td>
<td>Discovering new patterns of non-compliance</td>
</tr>
<tr>
<td>Manually working list of taxpayers</td>
<td></td>
<td>Interactive access to and analysis of tax data</td>
</tr>
<tr>
<td>Reliant on IT</td>
<td></td>
<td>Auditor/Collector-oriented tools independent of IT</td>
</tr>
</tbody>
</table>

Integrated internal sources
Enhanced external Data
Single View of Taxpayer
Some automation of desk audit cases
Reliant on IT
New York State Department of Taxation and Finance

Tax Audit and Compliance System (TACS)

A New Way to Select Audit Cases

TACS is a solution that enables taxing authorities to detect and investigate the fraudulent and abusive practices of taxpayers.

- Sorts through information on tens of thousands of tax returns in minutes
- Ranks taxpayers in relation to their peers
- Results are displayed in a variety of graphical formats that easily identify taxpayers who fall outside norms
- Supports further investigation
  - enables users to drill down into detailed information
- Integrated data mining, interactive data visualization and ROLAP drilldown reporting enable the system to display results as
  - reports
  - spreadsheets
  - bar graphs
  - two and three dimensional scatter plots
  - various other presentations
The foundation of TACS deviation detection is peer group profiling of similar entities using feature values and their profile scores.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature 1</td>
<td>13</td>
<td>850</td>
</tr>
<tr>
<td>Feature 2</td>
<td>1250</td>
<td></td>
</tr>
<tr>
<td>Feature 3</td>
<td>1.5</td>
<td></td>
</tr>
</tbody>
</table>

The profile model defines the content and structure of the profile. Scores for features roll up to groups, and groups to the composite.
The profile model Objective: Who, How, When, What to Profile and What to Report

Peer Class
- Tax Return Classification
- Industry??

Peer Group(s) Data
- Industry??
- Demographics
- Gross Receipts Range

Feature Generation
(Base, Computed, Derived)
- Profit Ratio
- % Expenses vs. Gross Receipts
- High Expense % & High Gross Receipts

Operations
- % of workstations
- # active peer groups
- Profile frequency

Profile Models
- COMPOSITE
- GB0001
- GB0002
- GB0003
- FB0001
- FBG004
- FB0006
- FB0002
- FB0007
- FB0003
- FB0008

Profile

Profile data

Reports
- Profile data
- ODS / Returns data

The metrics of the distribution of all entity values for a given feature are used to define the transformation of values into scores.

Sample Feature, "Average Dollars per ...........

Number of Entities

Feature Value

Feature Score

0 400 800 1200 1600 2000 2400 2800 3200

400 300 200 100 0

1820 370
Interactive visual data mining is provided by the integration of data visualization into the profile mode and into the report wizard.
Crystal Reports is the integrated report engine accessed through the extensible table / token / data driven Report Wizard.

Cluster IDs from profile data segmentation are dynamically added back to profile data in memory for iterative analysis and discovery.
Case Identification and Selection System (CISS)

Taking Audit Case Selection to Pre-processed Returns

Business Background
Categories of Questionable Returns

- Earned Income Tax Credit (EITC)
- Itemized Deductions
- Wage / Withholding
- Other
  - Dependent Child Care Credit
  - Real Property Credit
  - Suspect DB's (Prisoner, Identity Theft, Preparer, Bad Address)

There are 3 major programs for the questionable refund detection unit

The EITC audit issues mainly stem from the misrepresentation of income and/or dependents to maximize the credit.

- Potential Fraud – Maximizing the EITC Claim
  - Questionable Business Income
  - Gains/Losses - Schedule C
  - Questionable Wages And Withholdings
  - Ineligible Dependents
  - Preparer Fraud
  - Multiple Claims

- Domestic Issues
  - Custody Disputes
  - Duplicate Claims
  - Math Errors
  - Social Security Number Verification
Credit phases out as income increases

Claiming children raises the limit on income eligibility and increases the value of the credit

If married, income eligibility increases up to $1,000
As always, Itemized Deduction audit issues stem from the overstating of deductions to wipe out AGI and are hard to verify.

- Taxpayers claiming high deductions in areas of:
  - job expenses
  - personal mortgage interest
  - contributions

- Hard to verify deductions at the time of filing:
  - No requirement to provide proof of deductions claimed on Schedule A
  - Schedule A data not available at review time for paper filers
  - Limited third party data

- Most audits are performed post processing:
  - Deductions like job expenses and cash contributions need to be verified by correspondence with the taxpayer
  - Very labor intensive process

Falsifying employer withholding is a rapidly growing area of concern for the department and validation is difficult.

- Stolen social security numbers

- Fabricated wage and tax statements:
  - Easy to obtain W-2 statements with a valid employer EIN
  - Print their own W-2 on a computer to be submitted with their return

- Limited employer withholding information available at time of filing
Past enforcement efforts were not keeping up with the ever-changing compliance environment

- "Early warning" compliance detection was rudimentary
- Existing strategy was limited to "deferred detection" and "pay and chase"
- Batch selection systems were inflexible and difficult to change
- Case workflow was disjointed and mostly manual
- Auditors did not have the complete picture of why a case was selected
- Limited capability to analyze the effectiveness of the audit programs
- Limited capability to detect or predict new schemes

The project objective was to build a system to enhance existing audit case selection methods to detect questionable returns before processed

The Questionable Refund Detection unit wanted...

- A better way to identify questionable returns
- To question suspect returns before issuing refunds
- To improve the ability to collect on additional tax audit cases
- To issue refunds in a timely manner
- To make program management more flexible
- To leverage investments in data warehousing and business intelligence technologies
- To scientifically predict good audit candidates utilizing return filing patterns, case history, and other external indicators
- To improve their ability to detect new areas of fraud
NY realized it needed to find a way to detect a questionable return before issuing the refund avoiding a the "pay and chase" procedure.

Simple Edits & Matching
- Data Warehouse & Query Tools
  - Data from internal systems
  - Limited external data
  - Manually working list of taxpayers
  - Reliant on IT

Integrated Case Management
- Compliance Data Mining
  - Integrated internal sources
  - Enhanced external Data
  - Single View of Taxpayer
  - Some automation of desk audit cases
  - Reliant on IT

Predictive Compliance
- Automated Case Creation (desk and field)
  - Link to data warehouse
  - Reliant on IT
- Case selection at time of returns processing
  - Predict fraud based on historical behavior
  - Early detection of deviant non-compliant trend
  - Strong voluntary compliance effect
  - Independent of IT

Business rules and predictive models are applied to categorize and score returns received nightly
- Determine the "next best case"
  - Based on all returns received not on a batch or daily basis
  - Dynamic re-ranking of all potential cases nightly
- Identify screeners and auditor focus area
- Trigger automated downstream processes
  - Accept Refund Request
  - Automatic Stop
  - Audit Case Creation
  - Informational Letter
  - Manual Screening

Provides a web based application for return screening, audit case resolution, and answering taxpayer inquiries

CISS is an end-to-end system for audit case selection, resolution, and post processing year analysis

Powered by TACS – IBM's Tax Audit and Compliance System
The Rules Engine applies the business logic and predictive models to score returns and set workflow controls for downstream processes

- **Business rules**
  - Business logic that evaluates data from each current tax return, previous returns, audit history and other sources to make a determination about the action to be taken by the department regarding each return

- **Predictive modeling**
  - Predictive models utilize historical case results to predict such things as probability of fraud or taxpayer response
The Rules Engine scoring curve differs from the profile model scoring curve in that the user pre-defines the scoring curve points.

I-TACS Scoring – system derived based on data distribution of the peer group

P-TACS Scoring – user defined based on data analysis

The Rules Engine scoring process prioritizes cases to ensure the auditors and screeners are continually working the "Next Best Cases."

- Identify “next best case”
  - Selection of the best case based on all returns received not on a batch or daily basis
  - Dynamic re-ranking of all potential cases nightly
  - Browser based front end which provides a single point of entry for case identification, screening, analysis, selection, creation, and reporting
Externalized rules allow the user to manage the inventory and keep the volume of cases at a level that can be worked by the audit staff.

- Improved Program Management
  - Management of selection thresholds based on predicted resource availability
  - Current year case analysis capabilities to identify new areas of fraud with immediate implementation.
  - Externalized business rules provide greater flexibility in modifying existing edits and creating new ones
  - Post-year analysis of system edits to improve business rules, identify new areas for case review, and enhance the predictive model

Users now have the ability to implement business logic from existing compliance methods as well as “discover” new deviant patterns

- Advanced analytics takes detection to the next level
  - Augments return data with comprehensive indicators and scores to be used in rules engine
  - Utilizes predictive modeling to discover patterns in historical data for application to current returns
  - Utilizes profile modeling to identify new patterns and relationships in current returns
CISS incorporates a web-based information portal to support the return screening, audit resolution, and customer service processes.

- Improved user application for screening and auditing cases
  - **Enhanced System Screens**
    - Placing Needed Case Information In One Spot
    - Quick Access To Other System Screens
    - Reduced Clerical Support time
  - Automation Of Screening And Case Closing Functions
  - Improved Taxpayer Correspondence
    - Taxpayer Inquiry Letters Now Target The Specific Audit Issues
    - Utilized By Call Center For Refund Status Inquiries

An Example of a Case on an Auditor’s Worklist

To begin review process auditor would click on "Review" button

<table>
<thead>
<tr>
<th>Action</th>
<th>FID Name</th>
<th>EIN</th>
<th>Tax Year</th>
<th>Return Type</th>
<th>EIN</th>
<th>Status</th>
<th>Reason</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review</td>
<td>EXAMPLE</td>
<td>133-461-1060</td>
<td>2013</td>
<td>17-191</td>
<td>133-461-1060</td>
<td>Awaiting Resolution</td>
<td>100 JG WOODS AVE JG</td>
<td>100 JG WOODS AVE JG</td>
</tr>
</tbody>
</table>
Summary screen shows case scored 806 in the withholding category. Note key identifiers shows T/P not found on previous RP file and prime T/P filed another return this year.

Rule results screen shows which edits taxpayers return met and by clicking on the rule auditor gets a summary of data about the case.
Relevant data about the return is displayed: Note wages were found on our withholding system. However this is the second filing for the ID.

Code Description:
- Withholding Category

Full Rule Description or Calculation:
- Every return is evaluated for a variety of withholding issues based on income received for various jobs. Returns that score high for each of the following ratios have a high category score for potential fraud. Scored ratios include the wage ratio (Wage Reporting Wages / W-4 WC), Withholding Ratio (Wage Reporting Withholding / TIP Claimed Withholding (less suspicious as it approaches 100%).

Relevant Data:
- Wage Reporting System Wages: 42526
- Wage Reporting System Withholding Amount: 3213
- Total withheld: 3099
- NYS Tax Withheld: 1962
- Total tax withheld/Year: 0
- Total tax withheld/WC: 1137
- NYS WD: 42110
- Taxpayer Filing Request: 1149
- NYS DEC Claimed: 0
- DEC Amount: 0
- Withholding Category: 0.000
- Withholding Category Potential: 0.000
- Withholding Cat Next Best Case Value: 0.000

Results Screen shows that the case is in Process. Once the auditor works the case, the results will be posted here. Auditor will click on 'Deny Refund' button to continue.
Auditor has selected to deny the refund because the W/H was claimed by someone else and has selected to send an auto letter to T/P denying refund.

Auditor is assigning a Scheme number to the case and has entered an additional tax due amount of $1,947, a bill will be sent to the T/P.
The audit division has increased revenue by approximately $31 million from the previous year due to the implementation of CISS.

<table>
<thead>
<tr>
<th></th>
<th>Tax Year</th>
<th>2002</th>
<th>2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inquiry Letters Issued</td>
<td></td>
<td>55,776</td>
<td>78,000</td>
</tr>
<tr>
<td>Response Rate (%)</td>
<td></td>
<td>41%</td>
<td>35%</td>
</tr>
<tr>
<td>Refund Denials (%)</td>
<td></td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Cases Completed Per Day Per Auditor</td>
<td></td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Denied Refunds ($)</td>
<td></td>
<td>$49 Million</td>
<td>$80 Million</td>
</tr>
</tbody>
</table>

CISS provides the capability of adjusting thresholds such as the number of letters to send based on staffing limitations.

<table>
<thead>
<tr>
<th></th>
<th>Tax Year</th>
<th>2003</th>
<th>2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inquiry Letters Expired</td>
<td></td>
<td>27,458</td>
<td>49,966</td>
</tr>
<tr>
<td>Potential Denied Refunds ($)</td>
<td></td>
<td>$30.6 Million*</td>
<td>$70 Million **</td>
</tr>
</tbody>
</table>

*These cases were selected into the pool of potential letters to be sent, however were not sent due to the limitations of Audit to handle the additional potential responses

** On the assumption that the Category score were reduced to 500 and the probability of fraud were reduced to .74
New York State Department of Taxation and Finance

Next Steps.....

Build on our success from the previous years.

Future Compliance Initiatives

- Enhanced Refund Denial
  - Filing Status
  - Demographics Data
  - Preparer Analysis

- Expansion into other Audit Programs
  - Non-resident
  - Tax Shelters
Develop TACS for the Tax Compliance Division to aid in collection enforcement, risk analysis, and selection of next best collection case

- Understand the make up and complexity of cases
  - Determine what cases have the highest and lowest probability of being collected.
  - Determine the most effective and efficient collection path to assign to a case to maximize the chances of collection

- Streamline current risk assessment processes

- Enable managers to make smart and timely mission critical business decisions

Questions?
Interactive visual data mining is provided by the integration of data visualization into the profile mode and into the report wizard.