Data Intelligence:

Mining Data to Identify

Under-reporters and Non-filers

Office of Tax Policy
New York City Department of Finance

Presented at FTA Conference on Revenue Estimating and Tax Research
San Antonio, TX
September 2014

Data Intelligence: Mining Data to Identify Under-reporters and Non-filers

Background

- Data Intelligence Group (DIG) was created in late 2008 and fully staffed in Spring 2009.
- Modeling is intended to increase the revenues collected from firms other than the largest, which are on regular audit cycles, and to identify good audit candidates among those firms that have not been audited previously.

Data

- Tax data (Federal, State, City)
- Historical assessment data (Federal, State, City)
- Other data (licenses, third-party reporting, 1099Ks, FINCEN, etc.)

Data Intelligence: Mining Data to Identify Under-reporters and Non-filers

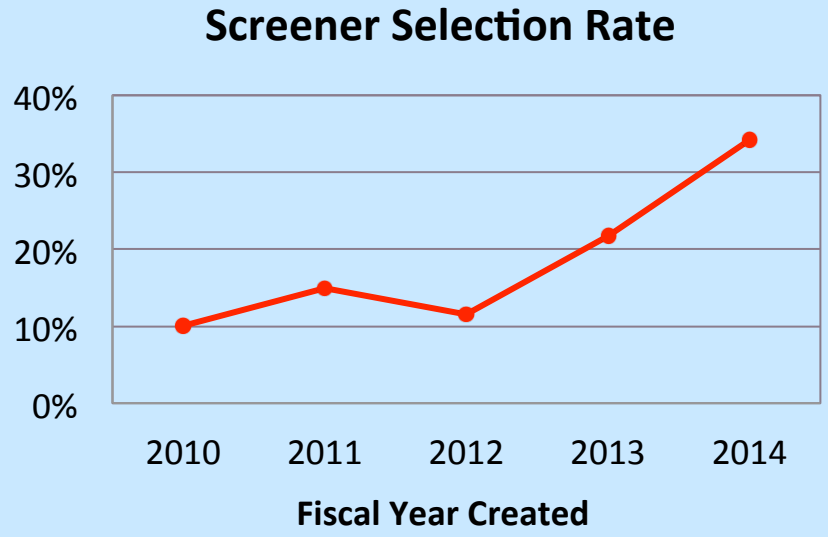
Modeling

- **Developed over 250 models, generating total closed assessments of $379m.**
- **Under-reporter models**
  - Comparison of line values (within and across returns) for under-reporters
  - Some of our best under-reporter models
    - GCT investment capital: $12m
    - GCT BAP property factor/real estate rented: $16m
    - UBT addback of partner payments: $6m
  - Predictive modeling (decision tree, logistic regression)
- **Non-filer models**
  - Over 2,600 cases created using our sophisticated non-filer protocol, with about $59m in closed assessments
  - FINCEN data – about $13m in closed assessments

Data Intelligence: Mining Data to Identify Under-reporters and Non-filers

Measuring a Model's Effectiveness

- Median or mean assessment
- Screener selection rate = # returns screeners selected/# returns identified by model
- Audit selection rate = # returns Audit selected/# returns screeners selected
- Audit assessment rate = # returns assessed/# returns Audit selected
- Overall assessment rate = # returns assessed/# returns identified by model



Finance

Data Intelligence: Mining Data to Identify Under-reporters and Non-filers

Lessons Learned / Next Steps

- **Hire staff with a passion for data.**
- **Data infrastructure is critical.**
- **Measuring results is key, but can create incentives for distortion.**
- **Model organization/management can be daunting.**
  - Consolidation of models
  - Organize by tax year or in-date?
- **Next step: drillable dashboard**