# Creating a Data-Driven Audit Division:

# New York City's Experience

Joy Chen

Data Integrity Group, Office of Tax Policy

NYC Department of Finance

Presented at FTA Conference on Revenue Estimating and Tax Research
Des Moines, Iowa
September 2009

## Creating a Data-Driven Audit Division: An Overview

- Background: "TAPE" division reorganization
- Required resources: personnel and technical infrastructure
- Goals
  - Establish good data foundations
  - Facilitate preliminary modeling
- Successes to date
  - Better data infrastructure
  - Case studies of preliminary modeling
- Future directions
  - Sophisticated modeling and data mining
  - Streamlined audit process

## Background: "TAPE" Division Reorganization

- Re-organization to merge Audit, Enforcement, and Tax Policy divisions
- Elimination of field vs. desk auditor distinction
- Instrumental in creating impetus for more data-driven approach to audit operations
  - Data Integrity Group, under Tax Policy, resides with Audit

## Required Resources: Personnel

- Current state of job market
- Senior analysts – economists or public policy experts with strong economics background
- Analyst – data documentation and basic statistical reports
- Strong programming skills; highly trained in data analysis

## Required Resources: Database and Analytical Tools

- IBM AIX db2 warehouse
- SAS
  - Base SAS – bulk of data work
  - SAS/STAT – more sophisticated research
  - Other advanced tools for data-cleaning and predictive modeling
    - SAS Enterprise Miner, DataFlux suite, etc.

## Goal I: Good Data Foundation

- Processes standardization and streamlining
- Documentation
  - Data calendar
  - Data dictionaries
- Cleansing
  - Understanding nature and source of data
  - Flexible programming

## Goal II: Preliminary Modeling

- If-then selection
  - Converting audit issues into data parameters
- Data matches
  - Between various sources
- Computing potential assessments
  - Useful to estimate effectiveness of model; requires clean data

## Early Success: Better Foundation

- Organized data warehouse
- Standardized data transmission protocols
- Expansion in data availability
- Improvement in data quality
- Affect both direct access and modeling capabilities

## Early Success: Case Studies

- Commercial check-cashing project
  - Use data tools to roll-up transactional data to analytic level
  - Generate potential audit candidates
- Restaurant project
  - Federal-state-city data match
  - Longitudinal study to target truly audit-worthy candidates

## Future Directions: Modeling

- Progression into more sophisticated, statistics- or theory-driven models
- Automated/limited scope audits
- Targeted audits to maximize productivity

## Future Directions: Streamlined Audit Process

- Better population to screen
- Appropriate assignment based on clearly defined complexity levels
- Greater transparency and clearer communication to taxpayers
  - Audit issue identification at start of audit