# General Purpose Business Establishment Synthesizer

*Model Design and Development*

***presented to***

*Ohio Travel Demand Model Users Group*

***presented by***

*Cambridge Systematics, Inc.*

*John Gliebe*

CAMBRIDGE SYSTEMATICS

Honoring the Past *Shaping the Future*

**November 18, 2022**

# Overview

- Motivation and Use Cases
- Establishment Evolution -- Births, Deaths, Migration
- Data: Longitudinal Establishment Records
- System Design and Components
- Performance
- Calibrating Behavior
- Next Steps Towards Usability

# Design Objectives

- General purpose business establishment synthesizer
- Analog to population synthesizers
- Respect TAZ control totals for employment by industry sector
- Reflect statistical trends found in longitudinal establishment records (QCEW)
- Preserve heterogeneity – carry forward establishments from prior years rather than synthesizing all new establishments for each year
- Implies an evolutionary model

# Use Cases

- Replace the local delivery/service commercial vehicle model (DCOM) in the statewide TDF with a model that includes explicit distribution/warehousing components and linkages to the long-distance freight movement model (ACOM).
- Modify the long-distance freight model (ACOM) in the statewide TDF which currently uses TAZ employment and productivity factors to disaggregate Freight Analysis Framework (FAF) commodity flows to TAZs to account for varying commodity production per employee by different firm sizes.
- Modify the 3C MPO model to convert the choice sets of the destination choice models from TAZ's to individual establishments (or households) using a richer set of mandatory and discretionary activity types corresponding to the detailed industry type of the synthetic establishments.

# Data: Longitudinal Establishment Records

- Mostly QCEW with reconciliation by ODOT staff
- Disaggregate establishment records used for two purposes:
  - As a basis for developing time series distributions of establishment births and deaths by industry category and the distribution of establishment by employer size class and industry category
  - As a starting point for synthesizing firms for a base-year model

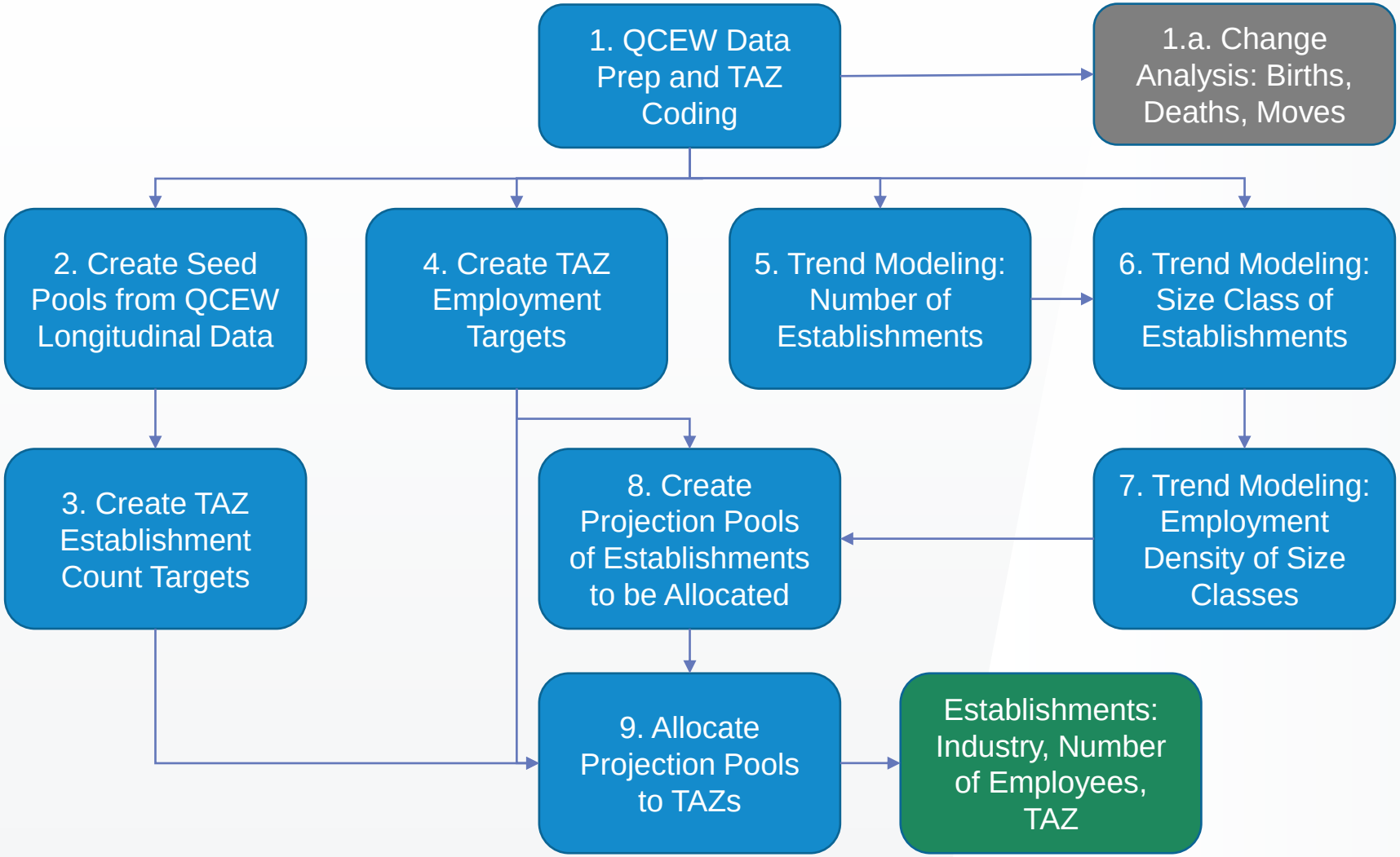
# Establishment Evolution

- Typical modeling approach is to simulate:
  - Births – new establishments (or move in from out of state)
  - Deaths – establishments go out of business (or move out of state)
  - Migration – geographic moves within the state

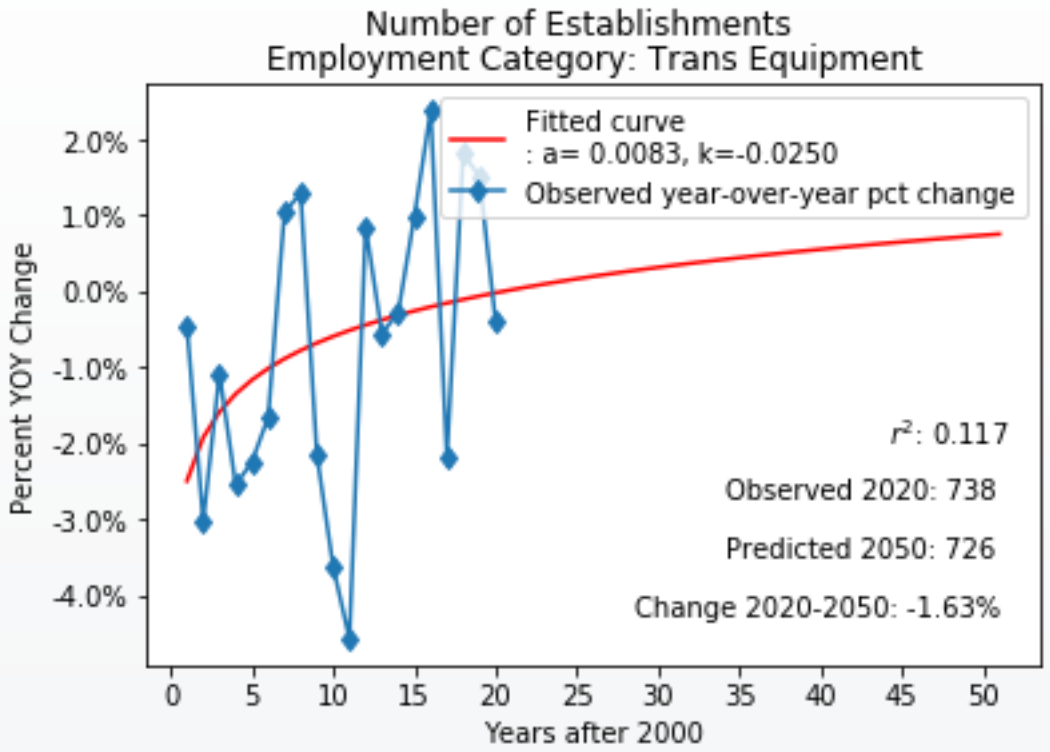
# Issues with Longitudinal Data

- For some establishments there are:
  - Gaps in records from year to year
  - NAICS recharacterization that change sector designations
  - Establishment IDs change but appears to be same business in same location (may be ownership changes)
  - Inconsistent reporting of employment
- Calculating separate birth and death rates proved unreliable
- Decided on simpler approach in which we modeled the yearly net change in establishment size distributions, rather than separate birth and death processes

# System Design and Components



# Number of Establishments Trend Modeling




- Fitted trends in year over year changes for each sector



CAMBRIDGE SYSTEMATICS

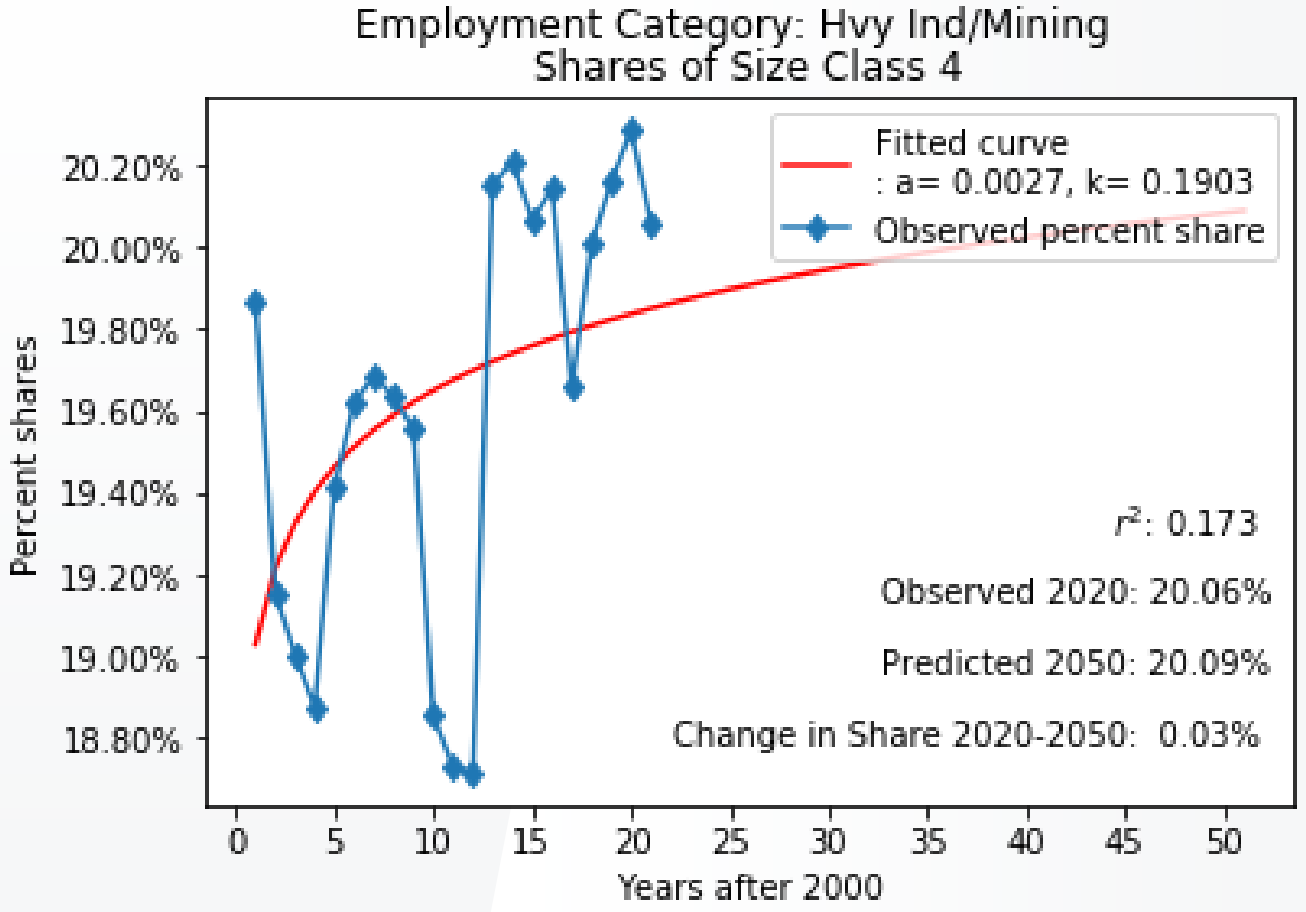
# Size Classes

**National Business Employment Dynamics Data by Firm Size Class**

- Size class 1 (1 to 4 employees)
- Size class 2 (5 to 9 employees)
- Size class 3 (10 to 19 employees)
- Size class 4 (20 to 49 employees)
- Size class 5 (50 to 99 employees)
- Size class 6 (100 to 249 employees)
- Size class 7 (250 to 499 employees)
- Size class 8 (500 to 999 employees)
- Size class 9 (1,000 or more employees)

# Employee Size Class Trends

- Fitted year-over-year changes in percent share of establishments within each size class
- Yes, they sum to 100% when projected!

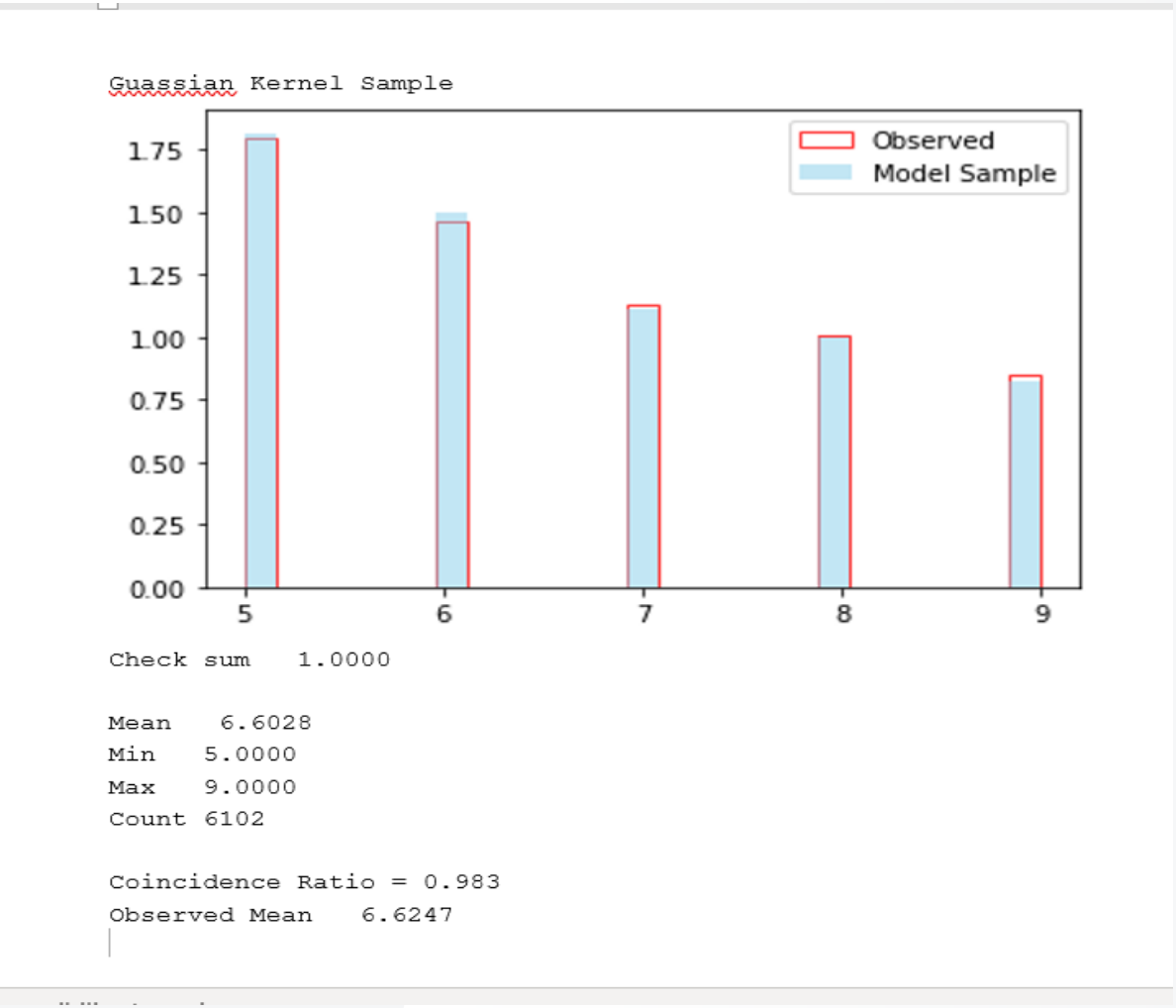





CAMBRIDGE SYSTEMATICS

# Employee Size Distribution Generator

- For each industry sector and size class, estimated three density functions and picked the one with the best fit
  - Truncated exponential
  - Gaussian kernel
  - Beta distribution






CAMBRIDGE SYSTEMATICS

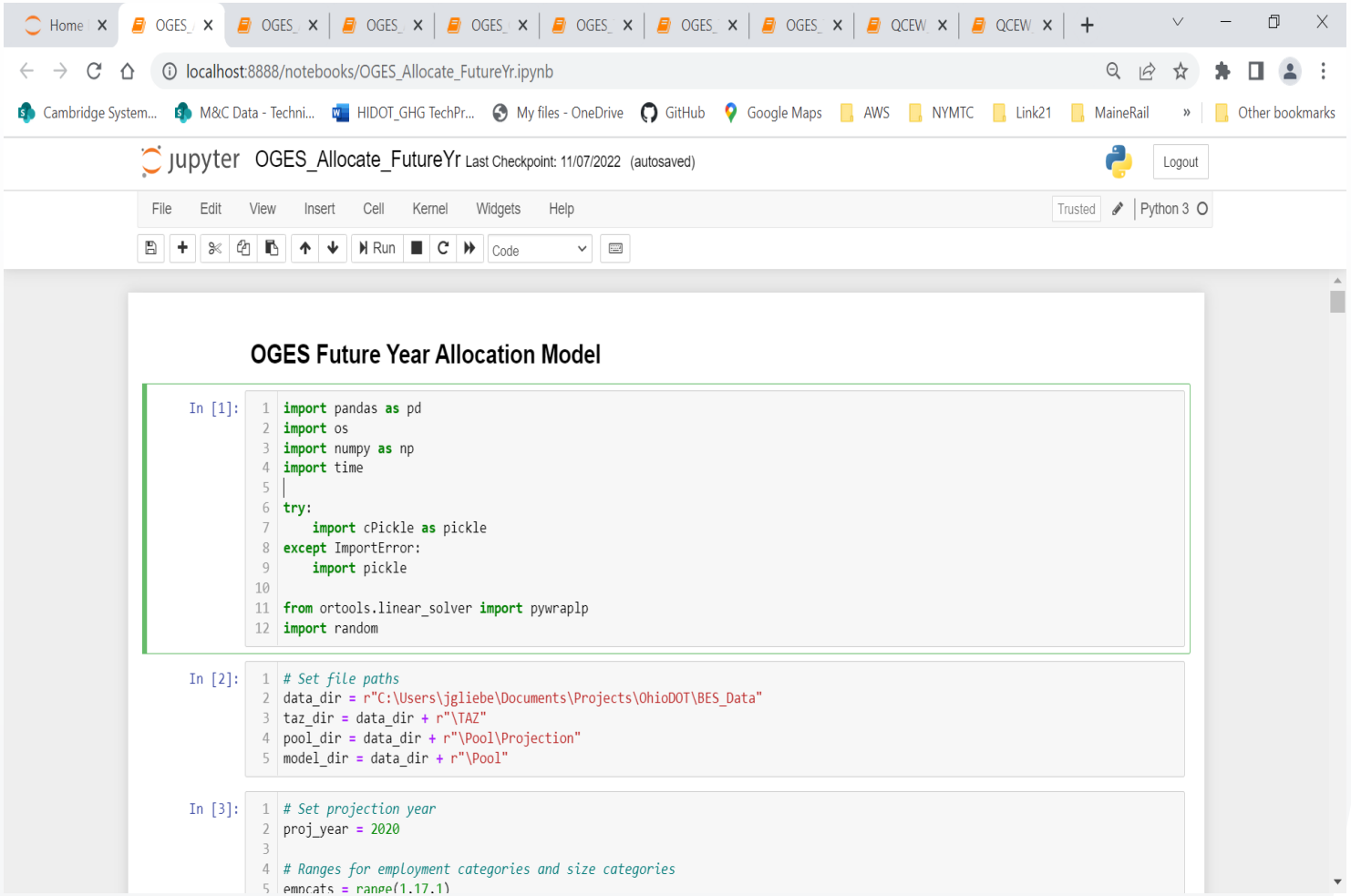
# Optimization Problem Formulation for Allocation Step

- Objective Function:
  - Maximize Sum of Total Allocated Employment +
  - Extra value for existing establishments allocated to base-year TAZ x Inertia Factor
- Constraints:
  - An establishment can be allocated to at most one TAZ
  - The sum of allocated establishments' employees cannot exceed TAZ targets
  - The size distribution of the allocated establishments deviates as little as possible from the projected size class distribution

# Future Year Employment for Existing Establishments

- Alt Future 1: Choose new employment from future-year size distribution
- Alt Future 2: Establishment grows in proportion to its base-year TAZ
- For each establishment, calculate employment both ways
- Optimization allows final employment value to vary between these two values

# TAZ Allocation Optimizer




- Each sector is a separate problem to solve
- "Multiple knapsack" formulation
- Implemented in Python
- Uses Google OR-Tools
- Planned migration from development version in Jupyter Notebooks to single executable



CAMBRIDGE SYSTEMATICS

# Problem Size

- For each sector, there may be several thousand establishments to allocate among several hundred TAZs
- For tractability, developed a batch processing approach

# Performance – Allocation and Run Times

```
Primary Metals, Batch 10 of 10
Pool size: 1314
Number of establishments in pool for this batch that have an input TAZ: 58
Solver time:   5.73 minutes
Solver number of constraints: 1289
Optimal solution found.
Value of Objective Function: 11446
Total Allocated Employment: 9229
Total TAZ Capacity Employment: 9275
Percent fulfilled: 99.50%


Finished!
Total time:   19.77 minutes

Results for Primary Metals:
Target employment:     103311
Allocated employment:  103214
Percent of total employment fulfilled: 99.91%
Number of TAZs not meeting target employment: 40
```

# Performance – Fit to Size Distribution Targets

Maximum Difference: 0.0165
RMSE: 0.0091
Coincidence Ratio: 0.931

Size Distributions:

| sclass | nestab | model_pct | target_pct | diff |
|---|---|---|---|---|
| 1 | 172 | 0.157078 | 0.150221 | 0.006857 |
| 2 | 126 | 0.115068 | 0.122755 | -0.007686 |
| 3 | 153 | 0.139726 | 0.145684 | -0.005958 |
| 4 | 219 | 0.200000 | 0.214231 | -0.014231 |
| 5 | 162 | 0.147945 | 0.139412 | 0.008533 |
| 6 | 167 | 0.152511 | 0.155973 | -0.003461 |
| 7 | 53 | 0.048402 | 0.044355 | 0.004047 |
| 8 | 39 | 0.035616 | 0.019094 | 0.016523 |
| 9 | 4 | 0.003653 | 0.008276 | -0.004623 |

# Allocation Results by TAZ

6]:

| | ecat | taz | targ_emp | model_emp | base_estb | model_estb | delta |
|---|---|---|---|---|---|---|---|
| 0 | 2 | 66 | 21 | 21 | 0 | 2 | 0 |
| 0 | 2 | 71 | 436 | 436 | 0 | 3 | 0 |
| 0 | 2 | 87 | 80 | 80 | 0 | 2 | 0 |
| 0 | 2 | 91 | 63 | 63 | 2 | 2 | 0 |
| 0 | 2 | 92 | 11 | 11 | 0 | 4 | 0 |
| 0 | 2 | 105 | 183 | 183 | 0 | 2 | 0 |
| 0 | 2 | 109 | 13 | 13 | 0 | 1 | 0 |
| 0 | 2 | 112 | 109 | 109 | 1 | 2 | 0 |
| 0 | 2 | 131 | 2 | 2 | 0 | 1 | 0 |
| 1 | 2 | 144 | 58 | 58 | 1 | 2 | 0 |
| 1 | 2 | 146 | 293 | 293 | 1 | 2 | 0 |

# Allocation Results by Establishment

18]:

| | ecat | uid | ytaz | xtaz | employees | sclass |
|---|---|---|---|---|---|---|
| 127 | 2 | 1782 | 0 | 66 | 6 | 2 |
| 93 | 2 | 787 | 66 | 66 | 15 | 3 |
| 27 | 2 | 3073 | 0 | 71 | 147 | 6 |
| 138 | 2 | 54 | 71 | 71 | 6 | 2 |
| 12 | 2 | 1177 | 71 | 71 | 283 | 7 |
| 90 | 2 | 2266 | 0 | 87 | 22 | 4 |
| 49 | 2 | 967 | 87 | 87 | 58 | 5 |
| 90 | 2 | 2318 | 0 | 91 | 21 | 4 |
| 64 | 2 | 886 | 91 | 91 | 42 | 4 |
| 157 | 2 | 1259 | 0 | 92 | 1 | 1 |
| 153 | 2 | 269 | 92 | 92 | 3 | 1 |

# Disposition Reports

Total allocated establishments: 1478
  Mean number of employees:  79.9

Total establishments with existing TAZ designation: 1215
  Mean number of employees:  84.9

Establishments assigned to same TAZ as the existing TAZ designation: 857
  Mean number of employees:  85.3

Establishments assigned to a different TAZ from the existing TAZ designation: 53
  Mean number of employees:  177.6

Establishments with an existing TAZ designation that were not assigned: 305
  Mean number of employees:  67.6

Establishments with an existing TAZ designation with more employees than the TAZ target: 237
  Mean number of employees:  88.7, Mean capacity:  37.4

Establishments with an existing TAZ designation that has zero target employees: 57
  Mean number of employees:  28.6

Establishments with existing TAZ designation and employees <= TAZ target but not assigned to that TAZ: 121
  Number of TAZs involved: 109
  Number of other establishments with existing TAZ designation allocated to those same TAZs: 139 found in 99 TAZs.
  Those 99 TAZs had a mean residual capacity of  146.4 that was fulfilled by establishments without this TAZ designation.
  Number of TAZs with capacity not allocated establishments with same existing TAZ: 10
  Number of establishments with existing TAZ designations left out: 10 with a mean employment of  25.4

# Interpretations as an Evolutionary Process

- New establishments added to round out the future-year size distribution and meet growth targets are considered “births.”
- Existing establishments that go unallocated are considered “deaths,” which commonly happens when a TAZ future-year target is significantly less than the base year or goes to zero.
- Existing establishments that are allocated to a different TAZ are considered geographic moves.
- The inertia parameter can be used to calibrate these rates against observed intra-state moving rates from the trend analysis.

# Next Steps Towards Usability

- Cleanup function to achieve 100% of TAZs meeting targets
- Calibration of Inertia Parameter vs. Intra-state move rates, which vary by industry and length of time horizon
- Fine tuning of future-employee size algorithm
- 3C MPO version with different and more sector definitions
- Run-time improvements – processing multiple sectors in parallel
- Implementation in Cube Catalog

# Q & A

CAMBRIDGE SYSTEMATICS