

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

Linking the MEPS-IC to Administrative Records: A New Approach to Inferring Establishment Links

Thomas Hegland[1], Alice Zawacki[2], G. Edward Miller[1]
[1]Agency for Healthcare Research and Quality, [2]Census Bureau

thomas.hegland@ahrq.hhs.gov

FCSM, October 26, 2022

Disclaimer

Disclaimer: Any opinions and conclusions expressed herein are those of the authors and do not necessarily reflect those of the Agency for Healthcare Research and Quality, the Department of Health and Human Services, or the U.S. Census Bureau. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release.

Disclosure Review Board Approval Numbers CBDRB-FY22-047 and CBDRB-FY22-292; DMS project number 7514872.

The Medical Expenditure Panel Survey – Insurance Component (MEPS-IC)

- A survey of 40,000 business establishments
  - Approximately a 70% response rate
- Uses the Business Register as its survey frame
- Questions focus on health insurance and employment
  - Insurance offer rate, take-up rate, plan characteristics
- Would benefit from more data on workforce characteristics and parent firm characteristics
  - Employee demographics, marital status, household income
  - Benefit decisions may take place at the parent firm level
- Decennial Census, IRS Form 1040, and IRS W2 records include the desired information!

Central Data Linkage Challenge

- MEPS-IC data is at the establishment level
- IRS Form 1040 and W2 records do not contain establishment identifiers
  - W2s include EINs, but these do not uniquely identify establishments
  - Practical data issues can also yield situations where establishments have workers split across multiple EINs
- No solved training set of workers matched to establishments among multi-establishment firms

Solution Overview

- Establishment data: identity of parent firm, employment, payroll, and location
  - Information comes from Business Register and MEPS-IC
- Worker data: firm of employment, pay, location
  - From IRS W2s, 1040s, Decennial Census, RCF
- Use identity of parent firm to link establishment to all employees of its parent firm
  - This solves the match for single-establishment firms
- Search for the subset of the parent firm's workforce that best matches the establishment
  - Prioritize matching on total employment and payroll
  - Secondarily, prefer workers closer to the establishment

Pre-Match Preparation

- Match establishments to parent firms in the BR
  - Resolve verified single-establishment firms
- Identify employees of parent firms in W2 records using EINs
- Prepare employment and payroll targets reflecting number of W2s expected for each establishment and total payroll across them
  - Point-in-time (PIT) employment is captured by the MEPS-IC and the BR, but W2s reflect over-the-year (OTY) emp.
  - Derive OTY emp. targets using LASSO trained on single establishment firm data as a function of PIT employment, industry, and other characteristics
  - Payroll targets require only mild cleaning of BR information

Two Cases, Two Approaches

- Case 1: One Establishment from a Multi-Establishment Firm
  - Conceptually more complex solution method
  - Less computationally intensive solution method
- Case 2: Multiple Establishments from a Multi-Establishment Firm
  - Conceptually simpler solution method
  - More computationally intensive solution method
- Let's start with Case 2

Case 2: Multiple Establishments from a Multi-Establishment Firm

- Multiple establishments drawing from the same pool of parent-firm workers
  - Each establishment needs a targeted number of workers
  - Each establishment needs a targeted payroll total
  - Each worker can work only at one establishment
  - Not all workers need to be assigned to an establishment
- Begin by provisionally assigning each establishment its targeted number of workers
  - Initial assignments can be arbitrary
  - I loosely aim to minimize commute distances when forming initial assignments by giving each establishment the workers closer to it than any other establishment

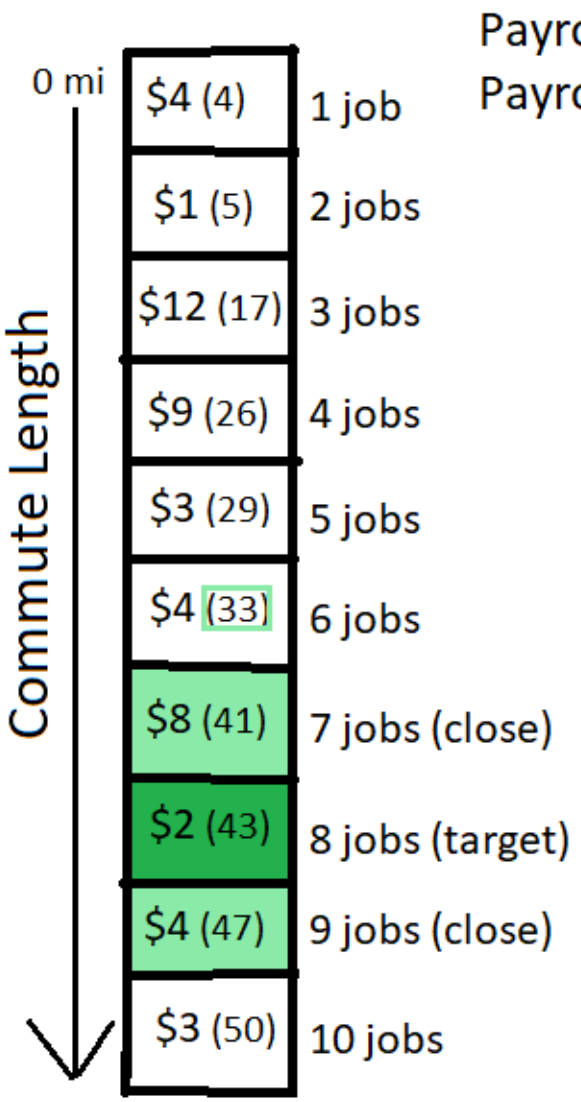
Case 2: Iterated Trading

- Iterate through establishments, allowing them to trade workers with other establishments and the pool of unassigned workers
  - Iterate until all estabs. have employment and payroll totals close to their targets
  - Failing that, stop after an immense number of trades have been completed
  - Caution: this can be very computationally intensive!
- Trades with other establishments
  - Trade must push both establishments closer to their payroll targets
  - Trade must not excessively increase estab. average commute distances
    - Threshold for permissible increase rises with the number of completed trading cycles
  - Exact trade made is chosen at random from the set of admissible trades
- Trades with the unassigned worker pool
  - Same conditions on trades for the establishment
  - All trades are admissible from the perspective of the unassigned worker pool
- Donations to and from the unassigned worker pool are also permitted
  - Donations may not push worker counts too far from estab. employment targets
  - Donations not allowed before completion of earliest rounds of trading

Case 1: One Establishment from a Multi-Establishment Firm

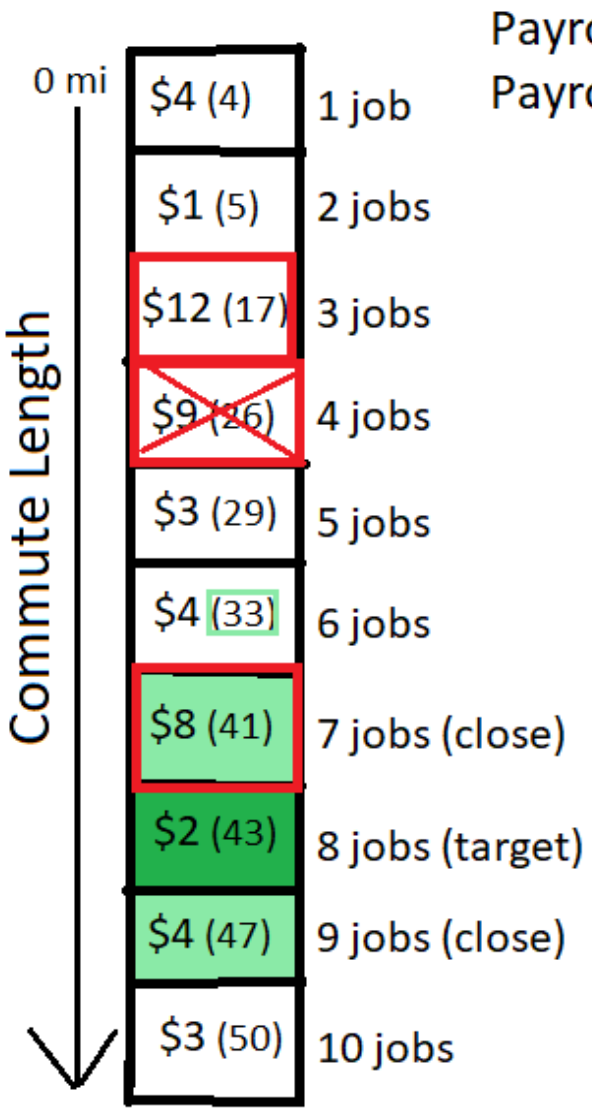
- Only one establishment and a pool of unassigned workers to consider
- We can aim for something more efficient than iterated trading
  - More precisely, we *must* do so for the solution to be feasible given many establishments fall into this case
- Note: the approach shown on the following slides is a simplification: additional efficiency gain tricks were used in practice
  - e.g. solution feasibility checks, batch processing of workers, etc.

Solution: One Establishment from a Multi-Establishment Firm



- Array workers in order of increasing distance from the establishment
- Calculate the running sum of payroll from worker 1 to each worker
- Check if the running sum of payroll is close to the payroll target for some worker *n* where *n* is close to the employment target
- If yes, choose one such *n* and assign workers 1 to *n* to the establishment

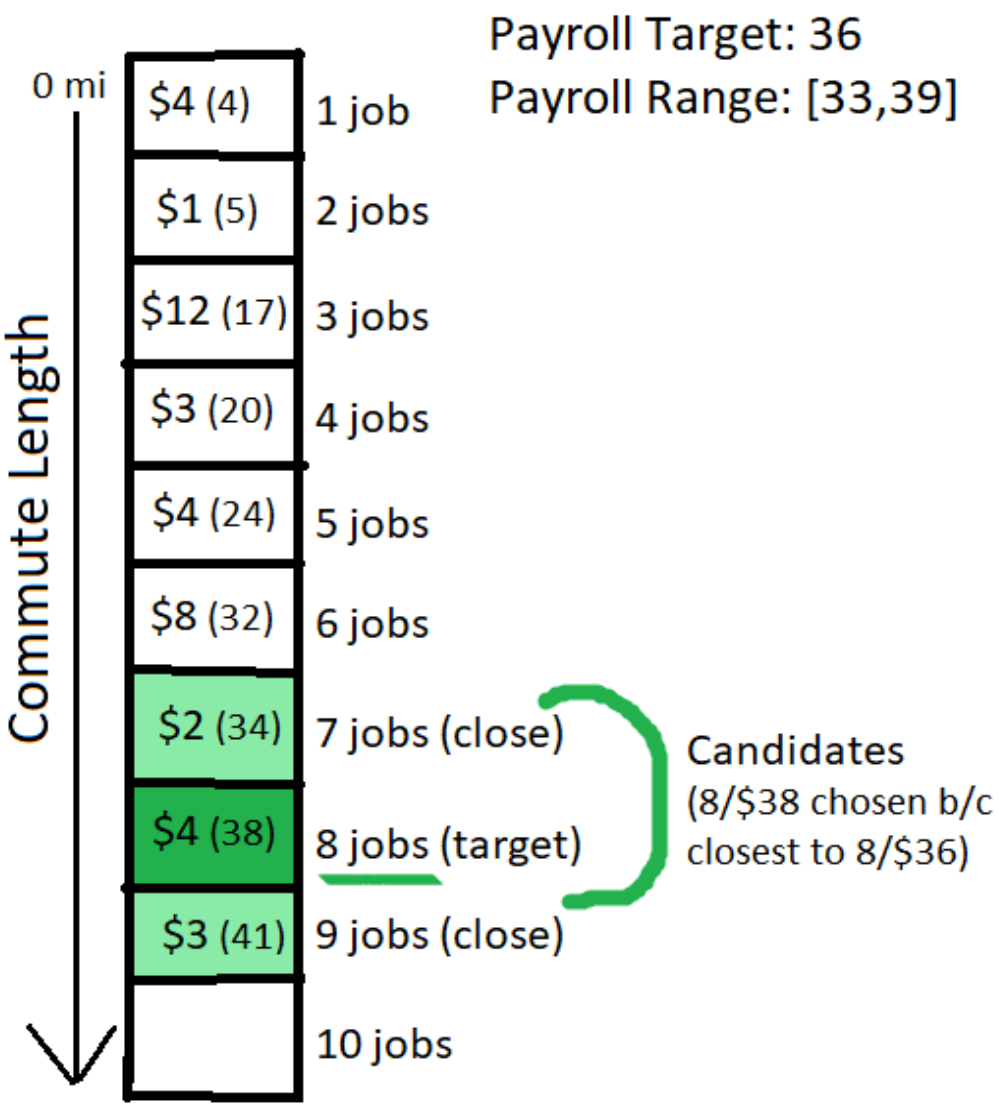
Solution: One Establishment from a Multi-Establishment Firm



Payroll Target: 36
Payroll Range: [33,39]

- If the running sum of payroll it too high, eject a high pay worker from the array
- If the running sum is too low, eject a low pay worker

Solution: One Establishment from a Multi-Establishment Firm



- Iterate until an admissible ordering is found, i.e. an ordering where there is some *n* close to the employment target with a running sum of payroll near the payroll target
- Use a penalty function to make the exact choice among admissible options

Solution: Final Steps

- Establishments have been matched to the set of workers they employ over-the-year
  - This is useful, but not the same employment concept as in the MEPS-IC and most other business surveys
- Construct weights that can be used to characterize the set of workers employed at the establishment at a given point-in-time
  - Conceptually, the weights should be the share of the year for which a given person was employed at the establishment
  - Develop weights using information about who is employed by the same firm in prior and future years, plus pay levels
- Check on completed matches and impose minimum match quality standards

Does it Work?

- 93% of establishments successfully match
  - Many match failures actually occur at the parent firm to W2s level
- Matched payroll and worker counts are generally close to their targets (within 5%)
  - This is true when considering the mean and median establishment with employment weights, and the median establishment in general
  - The mean establishment without employment weights sometimes deviates more due to sensitivity to very small establishments
- Implied turnover rate (OTY matched emp. / surveyed PIT emp.) also close to QWI-style turnover rate (derived from comparing tax records across years)
  - Turnover rates similarly close at the establishment-level and the parent firm-level

Does it Work?

- Minimal gap in match performance between single- and multi-establishment firms
- Tax-derived worker income distributions comparable to those in the American Community Survey after application of PIT weights
  - Mean 1040 Total Money Income within $2000 of mean ACS Total Family Income
  - Mean W2 pay within $2000 of mean ACS individual Wage and Salary Income
- Commute distance distribution similar to that in National Household Travel Survey benchmarks after trimming extreme commutes
  - Extreme commutes likely reflect measurement error

Discussion

- We develop a new approach for linking establishments to workers in tax data
- Our approach has a number of benefits:
  - Does not require hand linking a training set of workers within multi-establishment firms
  - Preserves information about relationship between workers: a given establishment might support only 1 highly paid manager along with other lower pay workers
  - Most common failure mode consists of assigning establishments a worker from the same firm, at the same level of pay, in a similar location to the worker that should have been assigned
  - Performance comparable to procedures used by other data programs
- Key limitation of the approach: computational intensity makes scaling to very large samples difficult
- Data currently spans 2005-2020: how will extending the data past 2020 work, given increased rates of work-from-home may affect importance of geographic location?

Conclusion: The MEPS-ICAR

- Final Data Product: The MEPS-IC with Administrative Records (MEPS-ICAR)
  - Spans 2005-2020 exclusive of 2007
  - Includes MEPS-IC establishment survey data linked to IRS W2 and 1040 Records + Decennial Census Records
- For more details on the linkage and on what data the MEPS-ICAR includes, see CES Working Paper Number CES-22-29
  - https://www.census.gov/library/working-papers/2022/adrm/CES-WP-22-29.html
  - Also published as MEPS Methodology Report 35
- Feel free to reach out with questions to: Thomas.Hegland@ahrq.hhs.gov
  - Co-author: Alice.M.Zawacki@census.gov
  - Co-author: Ed.Miller@ahrq.hhs.gov