Modeling Employer Decisions With Synthetic Firms in HISIM2: New Methods and Data

American Society of Health Economists

8th Annual Conference

Alexandra Minicozzi

Health, Retirement, and Long-Term Analysis Division

Why Does CBO Create Synthetic Firms in Its New Health Insurance Simulation Model (HISIM2)?

To model health insurance coverage, CBO also has to model whether or not employers offer employment-based health insurance.

Most people who have access to employment-based insurance enroll in it.

Health Insurance Coverage of People Under Age 65, 2015
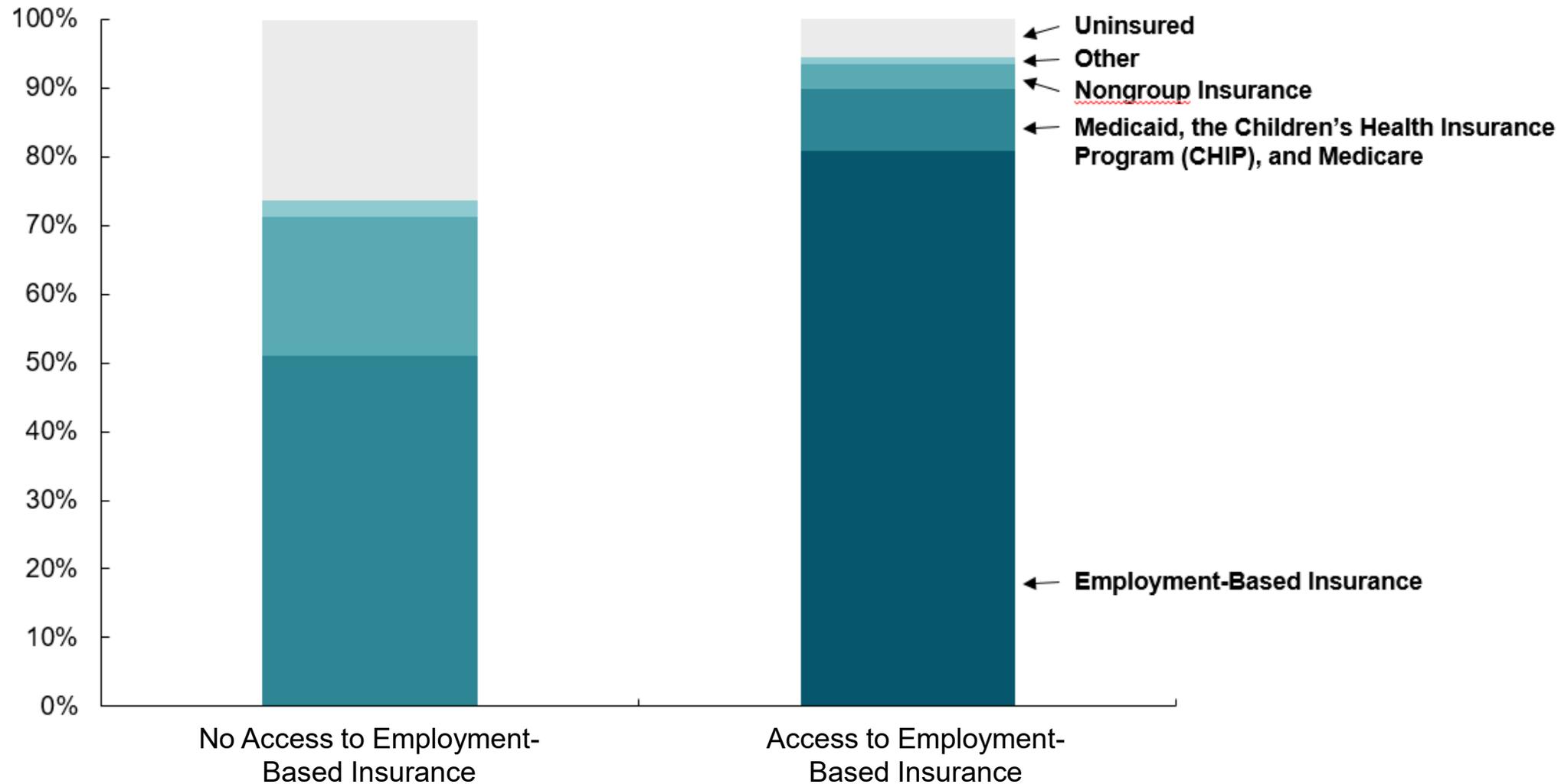
In HISIM2, CBO imputes the type of plan, if any, that a firm offers in the base year of the model.

For subsequent years, firms may change their decisions within the model, although switching a decision from the previous year has a cost.

When deciding whether to offer employment-based insurance, employers consider the characteristics of their workforce.

Firms' Objective Function

Firm k's workers' average willingness to pay for alternative ℓ

A randomly assigned shifter that changes the value of alternative ℓ for firm k and depends on the alternative ℓ^* chosen in the previous year

$$U_{k\ell} = WTP_{k\ell} - NetCost_{k\ell} + \gamma_{k\ell}(\ell^*)$$

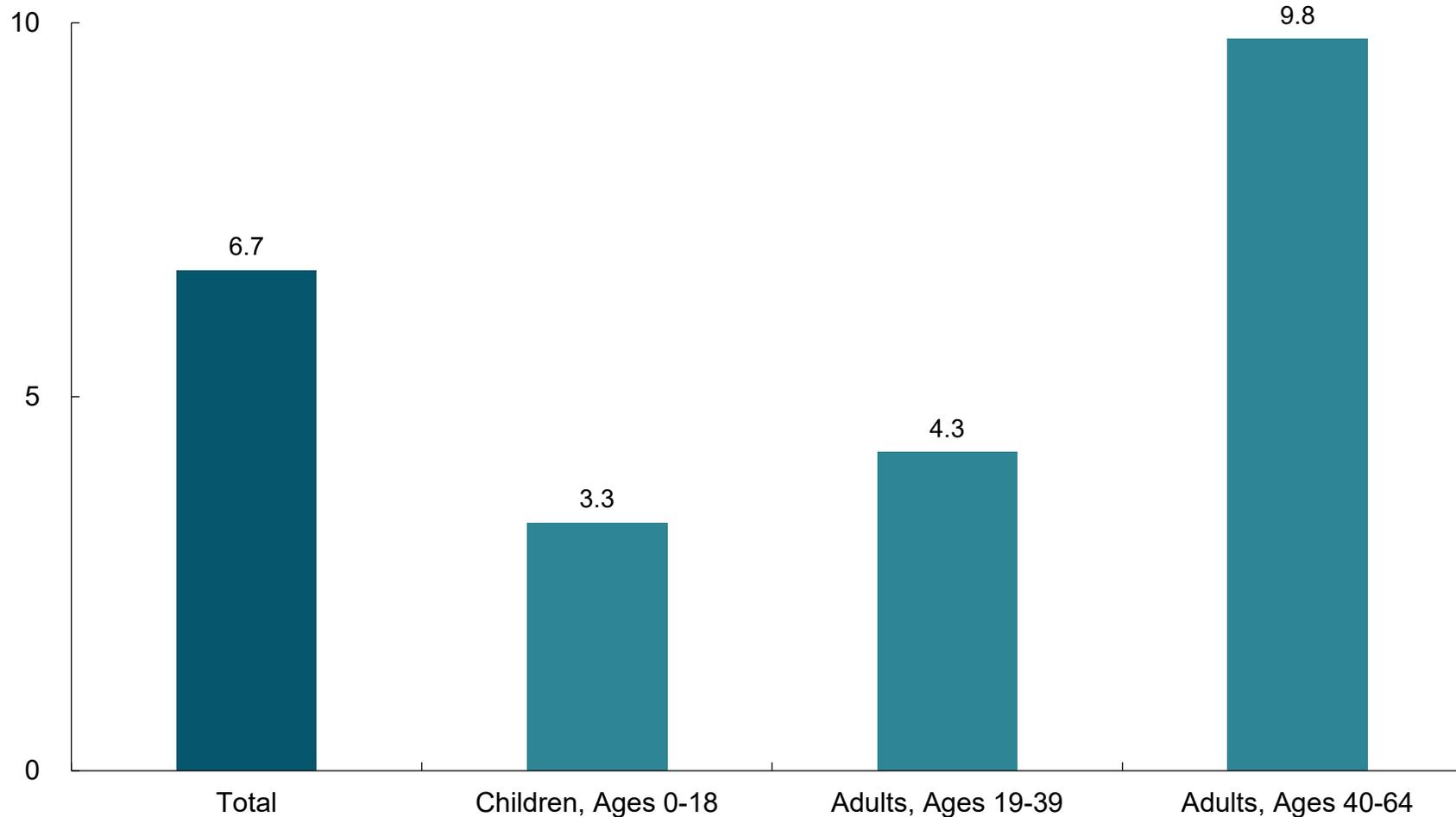Objective function value of alternative ℓ for firm k

Firm k's expected cost of offering alternative ℓ

Older workers consume more medical care, on average, than younger workers.

In HISIM2, they are willing to pay more for employment-based insurance than younger workers but are also more expensive to insure.

Average Health Care Spending for People Under Age 65 With Private Health Insurance, 2015

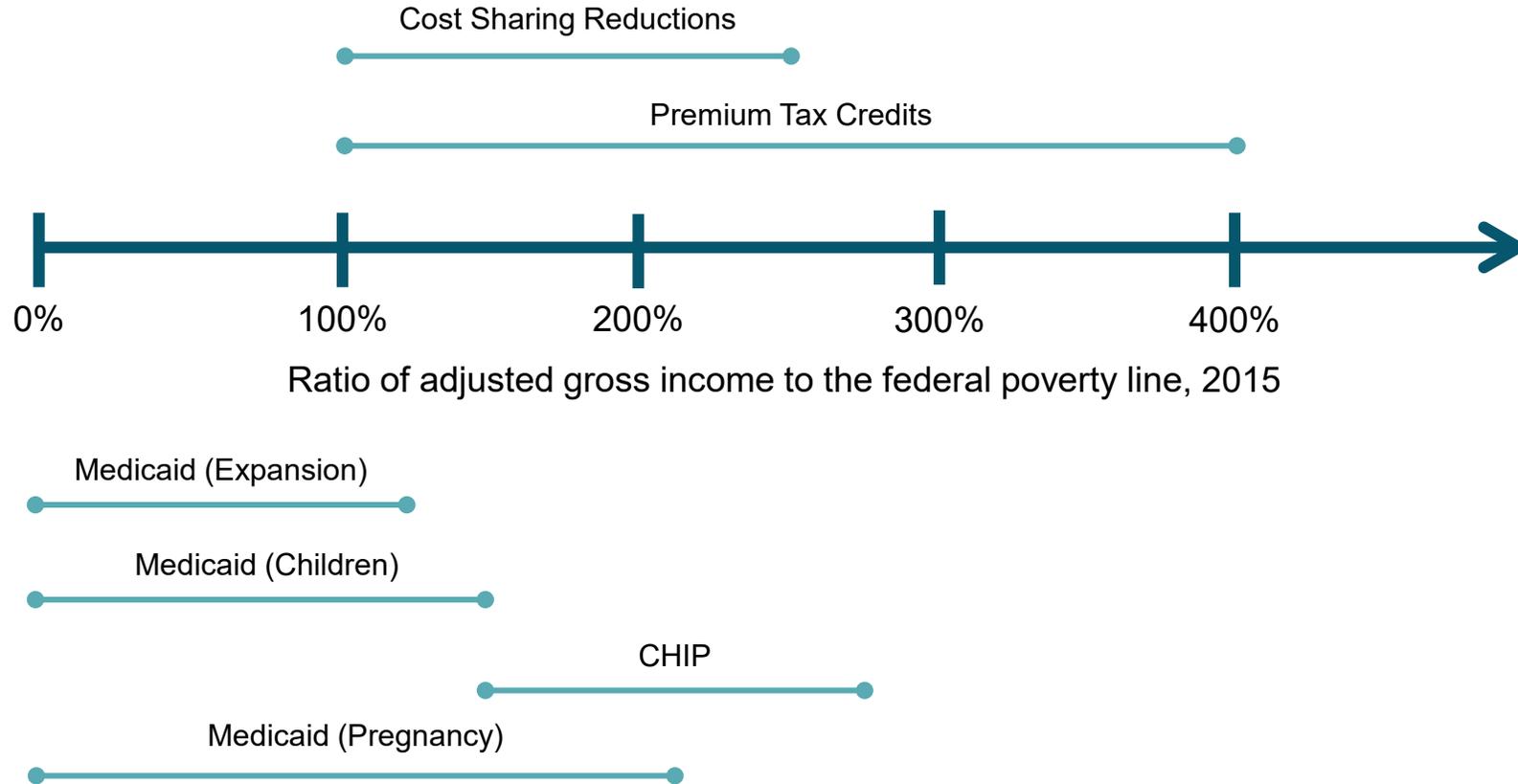Jessica Banthin and others, *Sources and Preparation of Data Used in HISIM2—CBO's Health Insurance Simulation Model*, Working Paper 2019-04 (Congressional Budget Office, April 2019), www.cbo.gov/publication/55087.

In HISIM2, lower-earning workers who qualify for Medicaid, CHIP, or marketplace subsidies are not willing to pay as much for employment-based insurance as higher-earning workers.

The value of the tax exclusion for employment-based insurance is also smaller for lower-earning workers than for higher-earning workers.
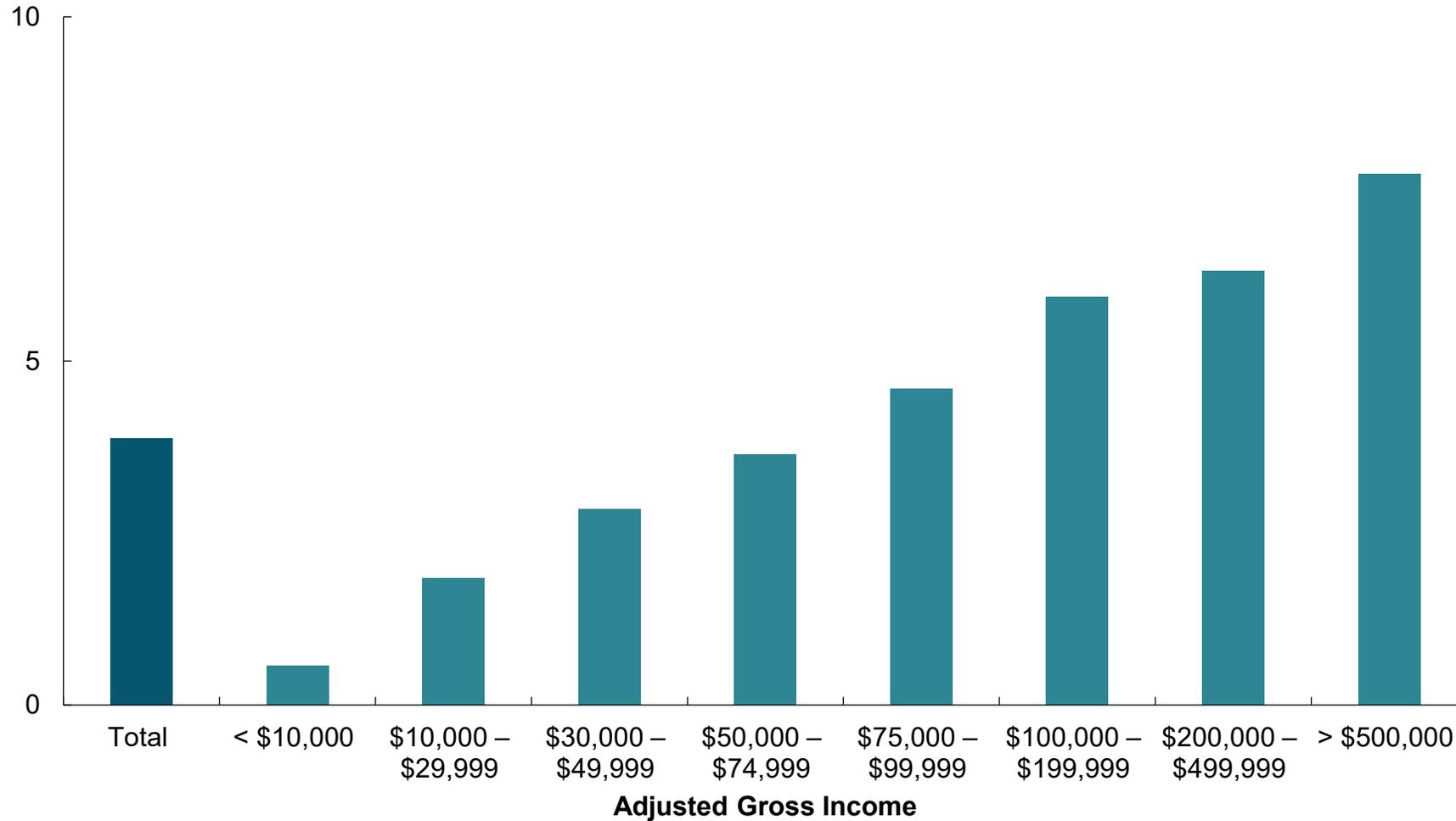
Eligibility for Subsidies Across Incomes

Cost Sharing Reductions

Premium Tax Credits

0%   100%   200%   300%   400%

Ratio of adjusted gross income to the federal poverty line, 2015

Lower-income workers may qualify for Medicaid, CHIP, or marketplace subsidies.

Medicaid (Expansion)

Medicaid (Children)

CHIP

Medicaid (Pregnancy)

Eligibility for marketplace subsidies is based on different definitions of household and income from eligibility for Medicaid and CHIP coverage. Income limits for Medicaid and CHIP eligibility vary by state and are sometimes higher or lower than shown.

Average Value of the Tax Exclusion per Return, 2019

Thousands of 2019 Dollars

The tax exclusion for employment-based insurance is more valuable for higher-income workers.

Adjusted Gross Income

CBO

CBO's HISIM2 model is based on the Current Population Survey (CPS), which identifies workers but does not provide information on their coworkers except for a total count.

Creating synthetic firms allows CBO to use the CPS while leveraging tax data that do provide information on coworkers.

Synthetic firms are created by randomly selecting coworkers for each worker in the CPS.

The probability that a coworker is selected is estimated from the tax data.

Example of How Synthetic Firms Are Created in HISIM2

HISIM2 Workers

Age	Income
<30	$25,000
<30	$40,000
30–54	$50,000
30–54	$68,000
>54	$60,000
>54	$80,000
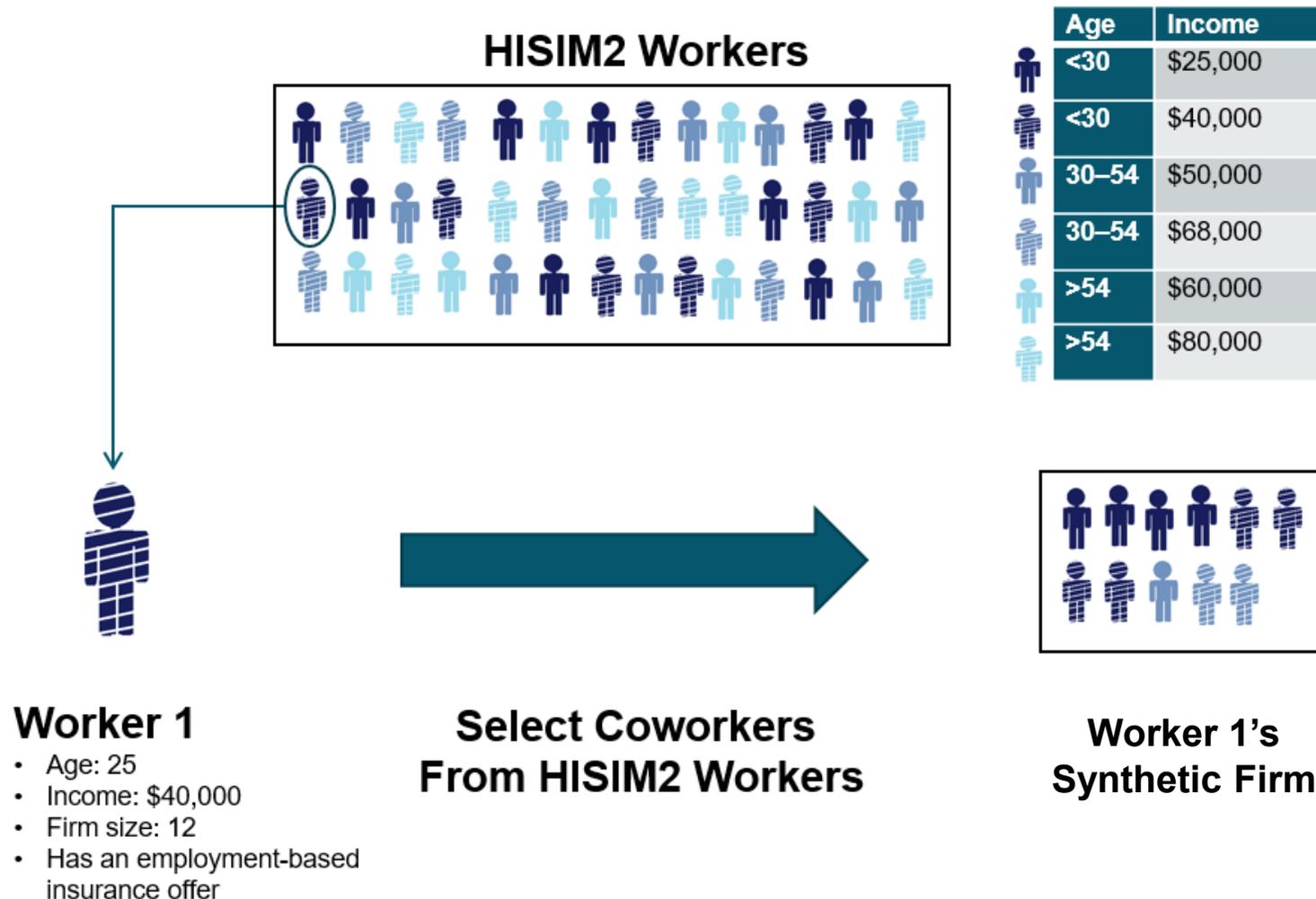
Each HISIM2 worker is given a synthetic firm composed of other HISIM2 workers.

Worker 1's Synthetic Firm

Select Coworkers From HISIM2 Workers

Worker 1
- Age: 25
- Income: $40,000
- Firm size: 12
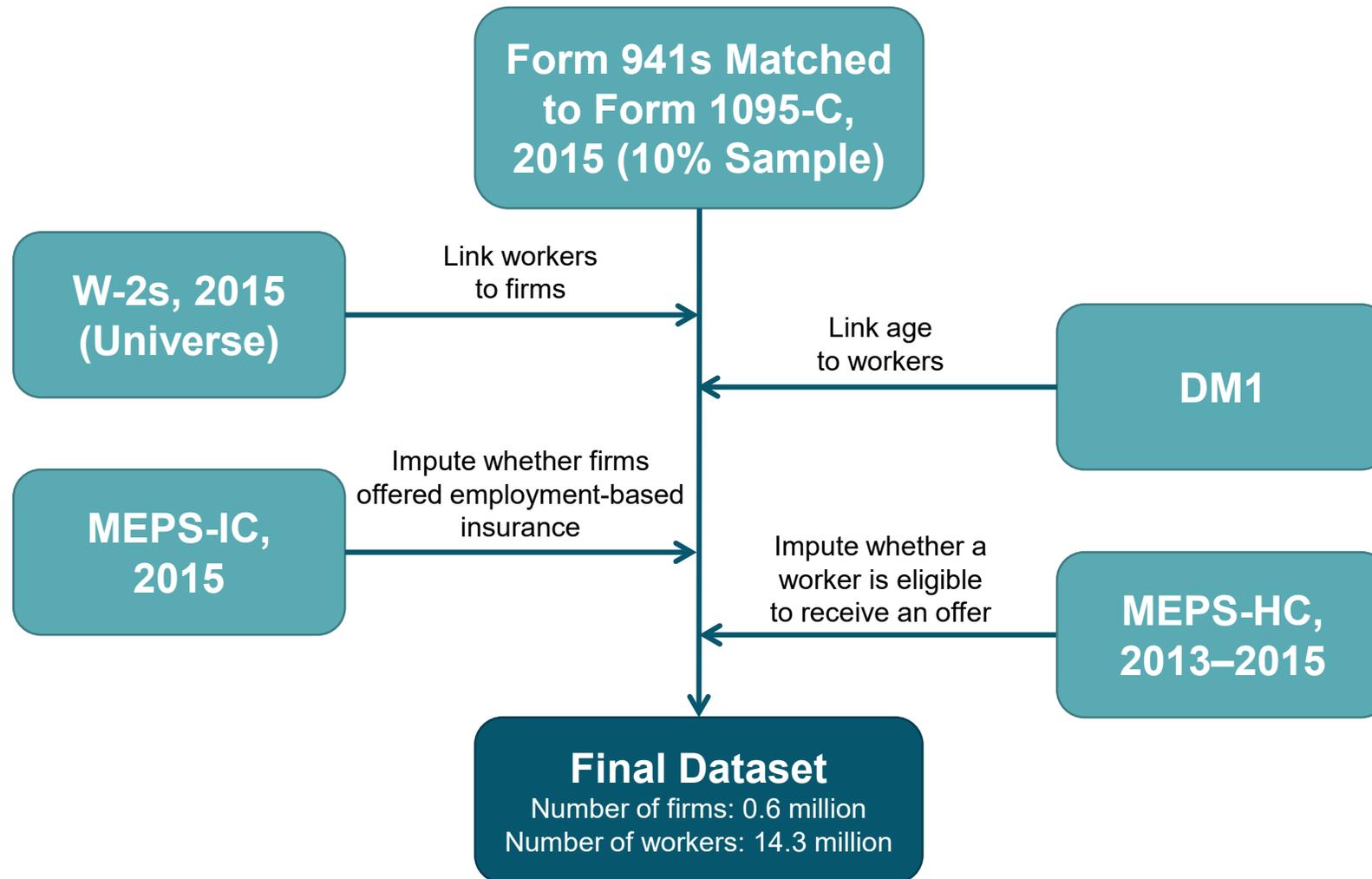- Has an employment-based insurance offer

Newly Available Tax Data

Newly available tax and Social Security Administration data provide detailed information on the age and earnings of an employer's workers.

Form 941s, 1095-Cs, and W-2s, which provide information on employers and workers, are linked using employer identification numbers.

Those data are supplemented with data from other sources, such as the Social Security Administration's DM1 data, which provide information on workers' ages.

Creating the Dataset for Assigning Coworkers to Synthetic Firms

Form 941s Matched to Form 1095-C, 2015 (10% Sample)

W-2s, 2015 (Universe) — Link workers to firms →

Link age to workers ← **DM1**

MEPS-IC, 2015 — Impute whether firms offered employment-based insurance →

Impute whether a worker is eligible to receive an offer ← **MEPS-HC, 2013–2015**

Final Dataset
Number of firms: 0.6 million
Number of workers: 14.3 million

Coworkers are assigned to synthetic firms on the basis of estimates from linked Form 941, Form 1095-C, W-2, and Social Security Administration DM1 data.

Form 1095-C is required only for applicable large employers (50 or more full-time-equivalent workers). Code DD on the W-2 is required only for employers with 250 or more workers.

The sample consists of 0.6 million employers out of a total of 6.0 million employers.

Employers do not always report whether or not they offered employment-based insurance.

CBO imputes whether employers that do not report offering employment-based insurance actually offered it.

Share of Firms Offering Employment-Based Insurance, 2015

Firm Size	Number of Firms, Form 941s (2015)	Share of Firms Offering Employment-Based Insurance	
		Pre-imputation (Percent)	Post-imputation (Percent)
Fewer Than 25 Workers	5,308,800	7	35
25 to 99 Workers	463,600	39	77
100 or More Workers	158,200	96	96

Firms That Offer Employment-Based Insurance

Firm Size	Number of Workers in Sample	Worker Age	Workers Earning Less Than $25,000 (Percent)	Workers Earning $25,000 to $49,000 (Percent)	Workers Earning $50,000 or More (Percent)	Total (Percent)
Fewer Than 25 Workers	958,100	Younger than 30	1.3	0.5	0.1	2.9
		30 to 54	1.7	1.5	1.1	4.3
		55 or older	0.7	0.6	0.5	1.8
		Total	**3.7**	**2.6**	**1.7**	**8.0**
25 to 99 Workers	1,515,400	Younger than 30	2.3	0.9	0.3	3.5
		30 to 54	2.3	2.4	2.1	6.8
		55 or older	0.8	0.8	0.8	2.4
		Total	**5.4**	**4.1**	**3.2**	**12.7**
100 or More Workers	9,411,900	Younger than 30	11.5	5.2	2.2	18.9
		30 to 54	11.1	14.5	18.0	43.6
		55 or older	4.7	5.1	6.8	16.6
		Total	**27.3**	**24.8**	**27.0**	**79.1**
Total	**11,885,500**					

Firms that offer employment-based insurance tend to be larger and have older, higher-earning workers.

Firms That Do Not Offer Employment-Based Insurance

Firm Size	Number of Workers in Sample	Worker Age	Workers Earning Less Than $25,000 (Percent)	Workers Earning $25,000 to $49,000 (Percent)	Workers Earning $50,000 or More (Percent)	Total (Percent)
Fewer Than 25 Workers	1,202,000	Younger than 30	8.4	2.8	0.6	11.8
		30 to 54	11.8	9.2	6.1	27.1
		55 or older	5.1	3.5	2.9	11.5
		Total	**25.3**	**15.5**	**9.6**	**50.4**
25 to 99 Workers	362,100	Younger than 30	3.2	1.0	0.3	4.5
		30 to 54	3.1	2.7	2.1	7.9
		55 or older	1.1	0.9	0.8	2.8
		Total	**7.4**	**4.6**	**3.2**	**15.2**
100 or More Workers	820,100	Younger than 30	7.3	2.0	0.6	9.9
		30 to 54	8.7	5.4	4.0	18.1
		55 or older	3.3	1.8	1.4	6.5
		Total	**19.3**	**9.2**	**6.0**	**34.5**
Total	**2,384,200**					

Firms that do not offer employment-based insurance tend to be smaller and have younger, lower-earning workers.
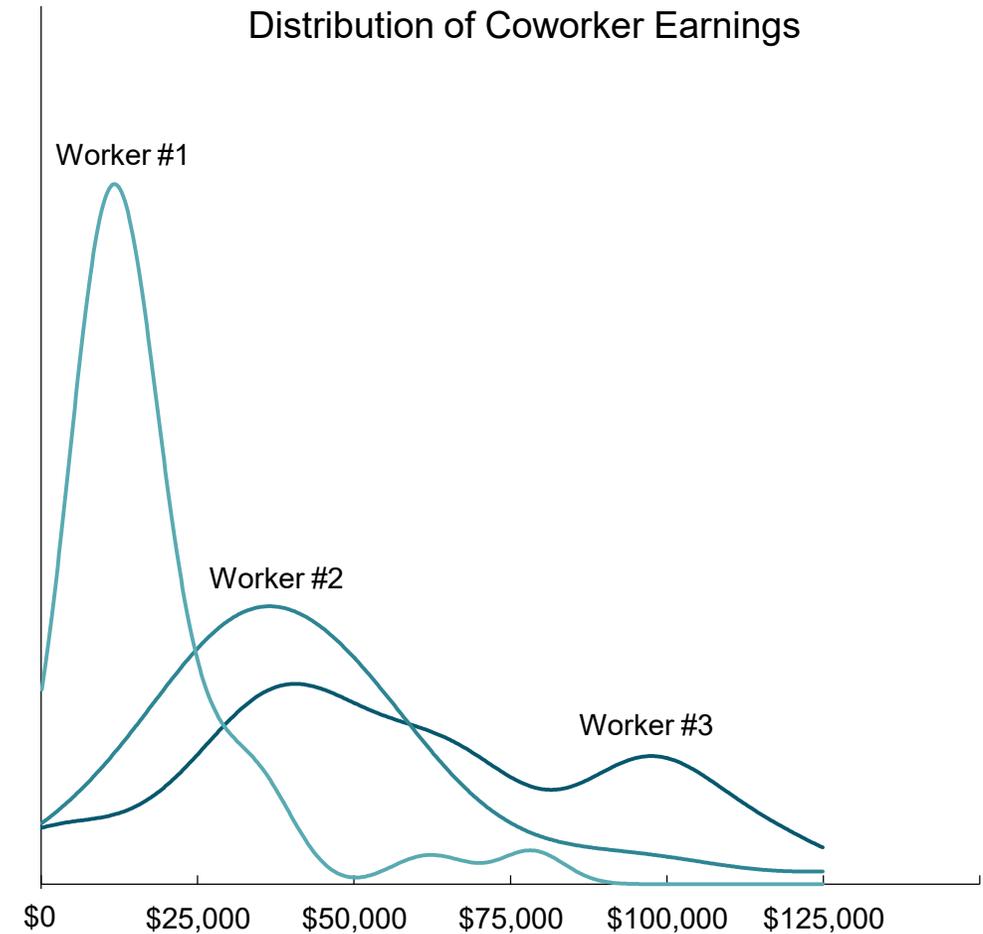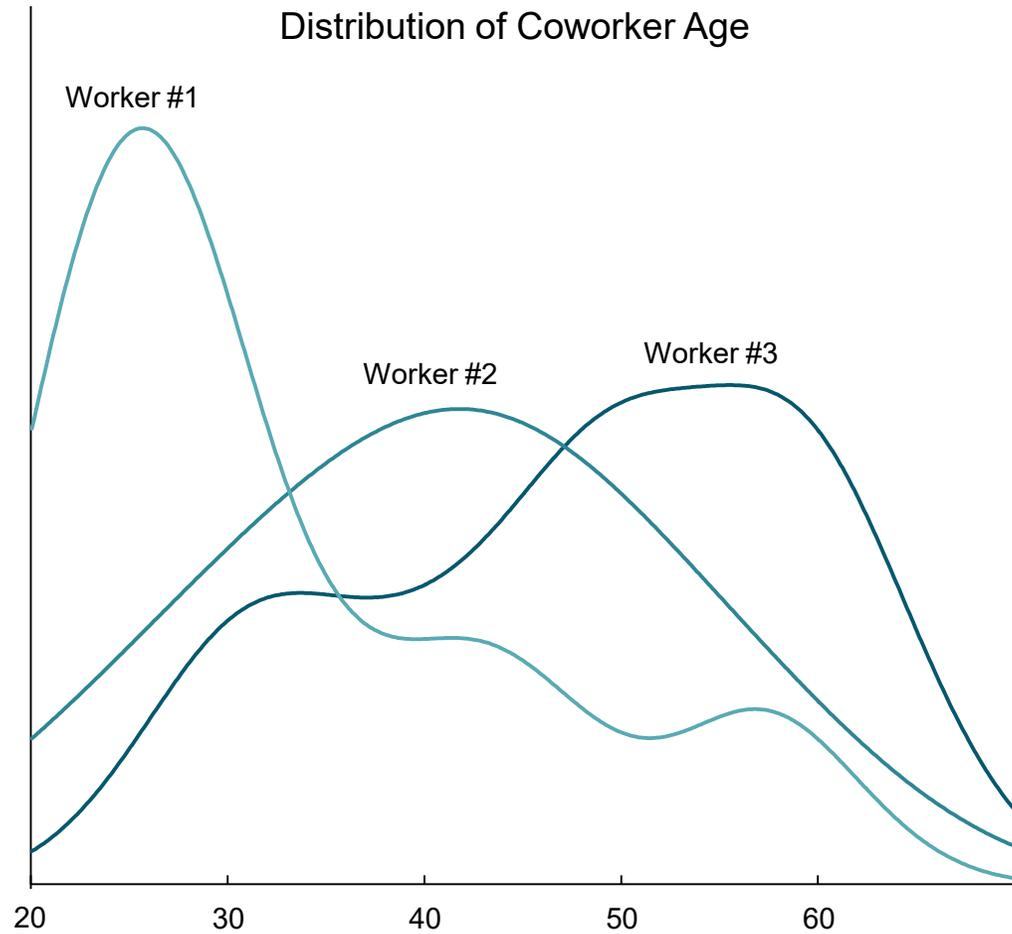
Identifying Different Types of Employers

Two workers may be similar but have very different coworkers.

CBO identifies different types of employers. For each type of employer, CBO estimates selection probabilities and the probability that workers are employed by that type of employer.

That process ensures that the synthetic firms for two similar workers can differ in a realistic way.

Similar Workers Can Be Employed at Different Types of Firms

Distribution of Coworker Age

Distribution of Coworker Earnings

The different types of employers are identified using the k-means clustering algorithm.

The k-means clustering algorithm is a popular machine learning algorithm that classifies observations into k clusters of observations that are similar to each other.

Clustering Algorithms Are Used to Classify Observations Into Different Types

k-Means Clustering Algorithm

1.   Choose the number of clusters k.

2.   Select an initial centroid for each cluster.

3.   Repeat (a)-(b) until the centroids do not change.

   a.   Assign firms to the cluster with the closest centroid.

   b.   Recompute the centroid of each cluster.

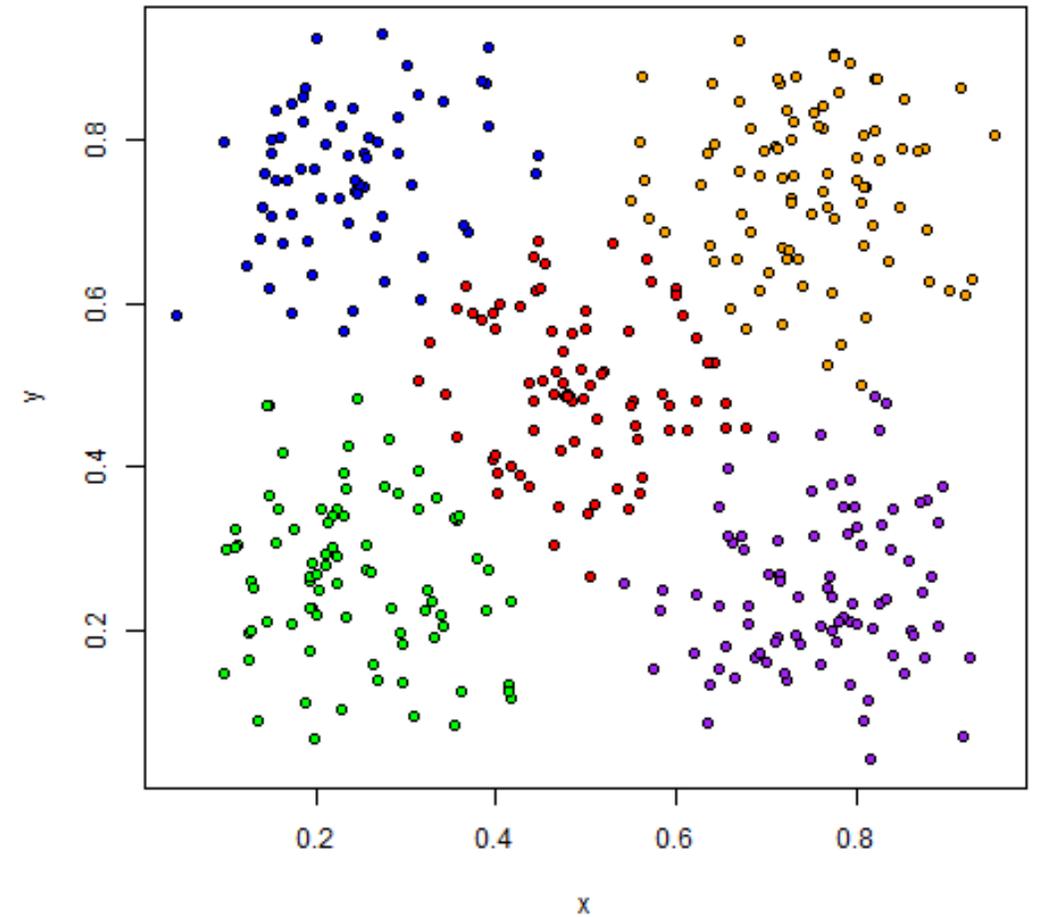The user specifies how many clusters the algorithm should identify.

CBO identified as many types of employers as possible without introducing the risk that personally identifiable information would be disclosed.

The User Specifies How Many Clusters the Algorithm Should Identify

Two Clusters

Five Clusters

The data on this slide were randomly generated.
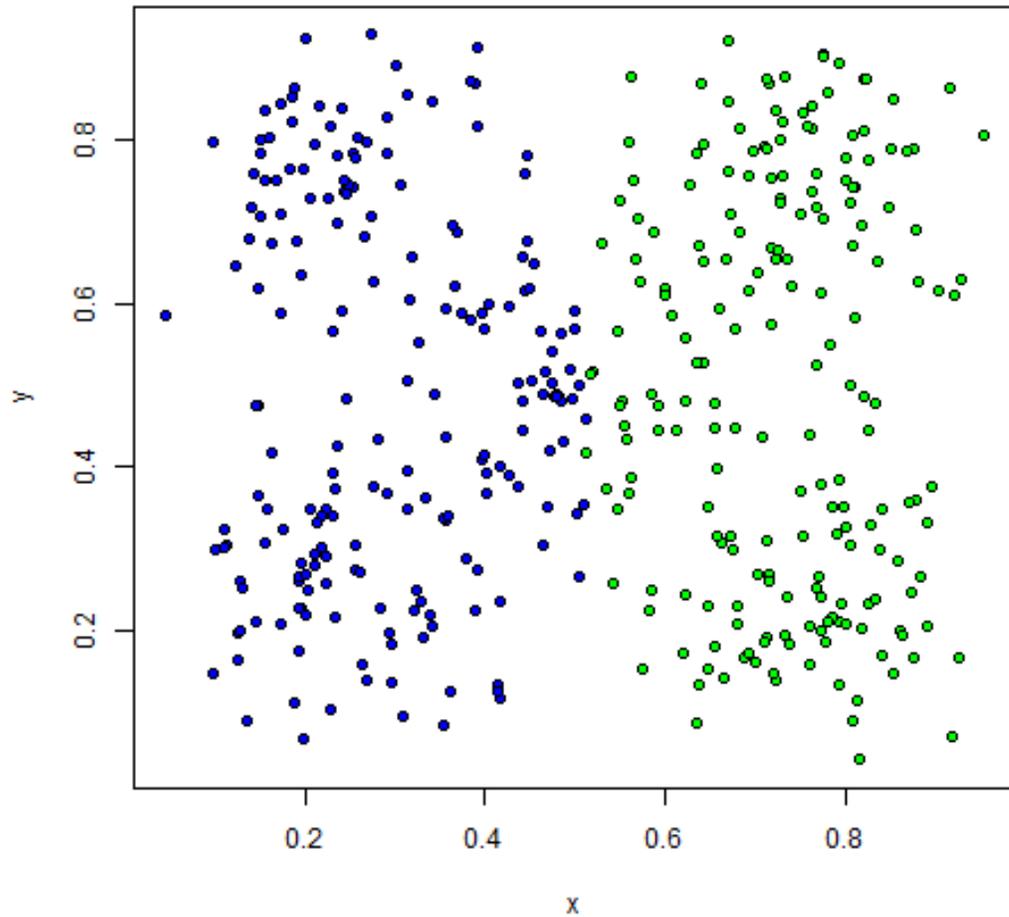
The user also specifies a way to measure the distance between observations.

CBO measures the distance between two employers as the Euclidean distance between their cumulative joint distributions of worker age and earnings.

Employers Are Clustered on the Basis of Their Workers' Age and Earnings

- Share of firm j's workers of age a or less and income y or less:

$$\hat{F}_j(a, y)$$

$$a \in \{<30, 30-54, 55+\}, \qquad y \in \{<\$10k, \ldots, \$90k+\}$$

- Centroid of cluster c:

$$\mu_c(a, y) = \frac{1}{n_c} \sum_{j \in S_c} \hat{F}_j(a, y)$$

Number of firms in cluster c

Set of firms in cluster c

- Distance between firm j and cluster c:

$$d(j, c) = \sum_{a \in A} \sum_{y \in Y} \left(\hat{F}_j(a, y) - \mu_c(a, y) \right)^2$$

Conclusions

- Using these data and methods, CBO constructs synthetic firms that mimic the real-world variation between firms.

- This approach improves CBO's ability to model employers' decisions in HISIM2.

Citations

Congressional Budget Office, *Federal Subsidies for Health Insurance Coverage for People Under Age 65: 2019 to 2029* (May 2019), www.cbo.gov/publication/55085.

Congressional Budget Office, *HISIM2—The Health Insurance Simulation Model Used in Preparing CBO's Spring 2019 Baseline Budget Projections* (April 2019), www.cbo.gov/publication/55097.

Jessica Banthin and others, *Sources and Preparation of Data Used in HISIM2—CBO's Health Insurance Simulation Model*, Working Paper 2019-04 (Congressional Budget Office, April 2019), www.cbo.gov/publication/55087.

Gareth James and others, *An Introduction to Statistical Learning* (Springer, 2013), pp. 385–401, https://tinyurl.com/ycxjja7u.

CBO thanks the staff of the Joint Committee on Taxation for their contributions to this analysis and presentation.