# OES Time Series
## Alternative designs and estimation methods

## Research Team

Matt Dey, Steve Miller, and Dave Piccone

November 21, 2014

BLS
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

www.bls.gov

# Goal of this research

- Redesign the sample and estimation methods so the design supports both cross-section (or annual) estimates and time series estimates without additional sample units

BLS

# A Brief History

- Began with my research
  - Tested on Alabama
  - One alternative sample design and one alternative estimation method
  - Promising Results
  - Presented to TAC in November 2012
- Major milestones of time series team
  - Created a simulated OES population
  - Created/Tested two alternative sample designs
    - Second alternative due to suggestion of TAC
    - Other suggestions made by TAC were incorporated into design
  - Created/Tested an alternative method for employment and wage estimates

BLS

# Current Sample Design

- Frame is created from QCEW records
  - Over 7 million in-scope establishments
- Sample size = 1.2 million establishments
  - 400,000 establishments per year (bi-annually)
  - Large sample size to cover detailed areas and industries
- Allocation, 2 goals:
  - Minimum allocation (maximize publishable estimates)
  - Power Neyman allocation (efficiency)
- Selection, stratified PPS
  - Strata = homogeneous cells (state x area x industry)
  - PPS = Probability Proportionate to Size (Size = Employment)
  - **All establishments selected in previous 5 panels are ineligible for selection in current panel**

BLS

# Two Alternative Sample Designs

- Major deviations from current sample design:
  - An establishment can be sampled in consecutive years
  - Annual sampling
  - Any given annual sample represents entire universe
  - Less detailed strata definitions
- Similarities with current sample design:
  - Uses similar allocation procedures
  - Uses probability proportionate to size sampling
  - Same annual sample size (approx. 400,000 establishments)

BLS

# Repeated Cross Section (RCS) Design

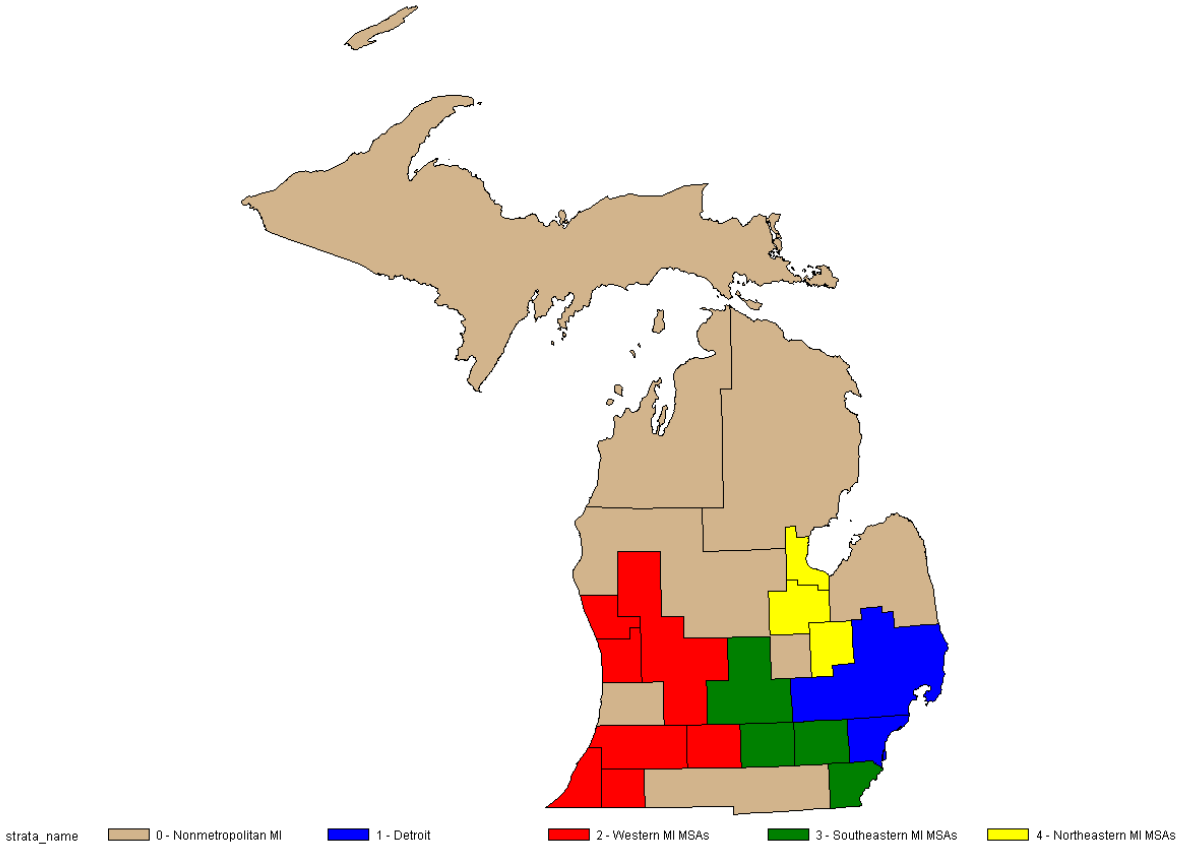
- Stratification: State x Aggregate area x Industry
  - Aggregate areas: collapse similar areas together within state, based on area size and geographic orientation

    *Example:* MI has 20 MSA/BOS areas → 5 aggregate areas

    CA has 33 MSA/BOS areas → 10 aggregate areas
- Allocation: Minimum + Power Neyman
  - Minimum Allocation: one per stratum
- Selection: Probability Proportionate to Size (PPS)
  - Size = establishment's employment
- Year-to-Year Overlap: Random

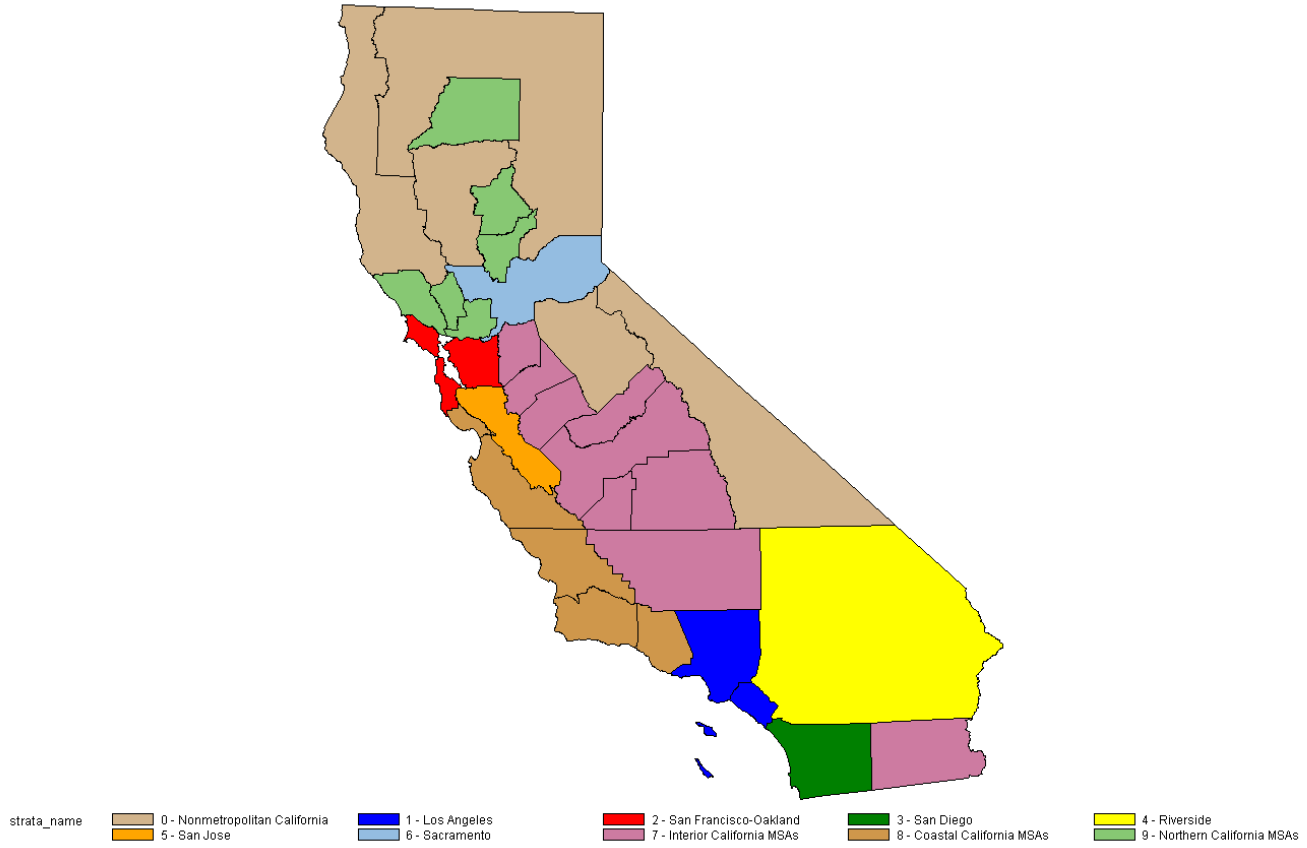
BLS

# Michigan Aggregate Areas




BLS

# California Aggregate Areas




BLS

# Two-Thirds Rotating Panel (PAN) Design

- Stratification: State x Aggregate area x Industry
- Allocation: Minimum + Power Neyman
  - Minimum Allocation: three per stratum
- Selection: Probability Proportionate to Size (PPS)
  - Size = establishment's employment
- Year-to-Year Overlap: two-thirds of sample is forced to overlap

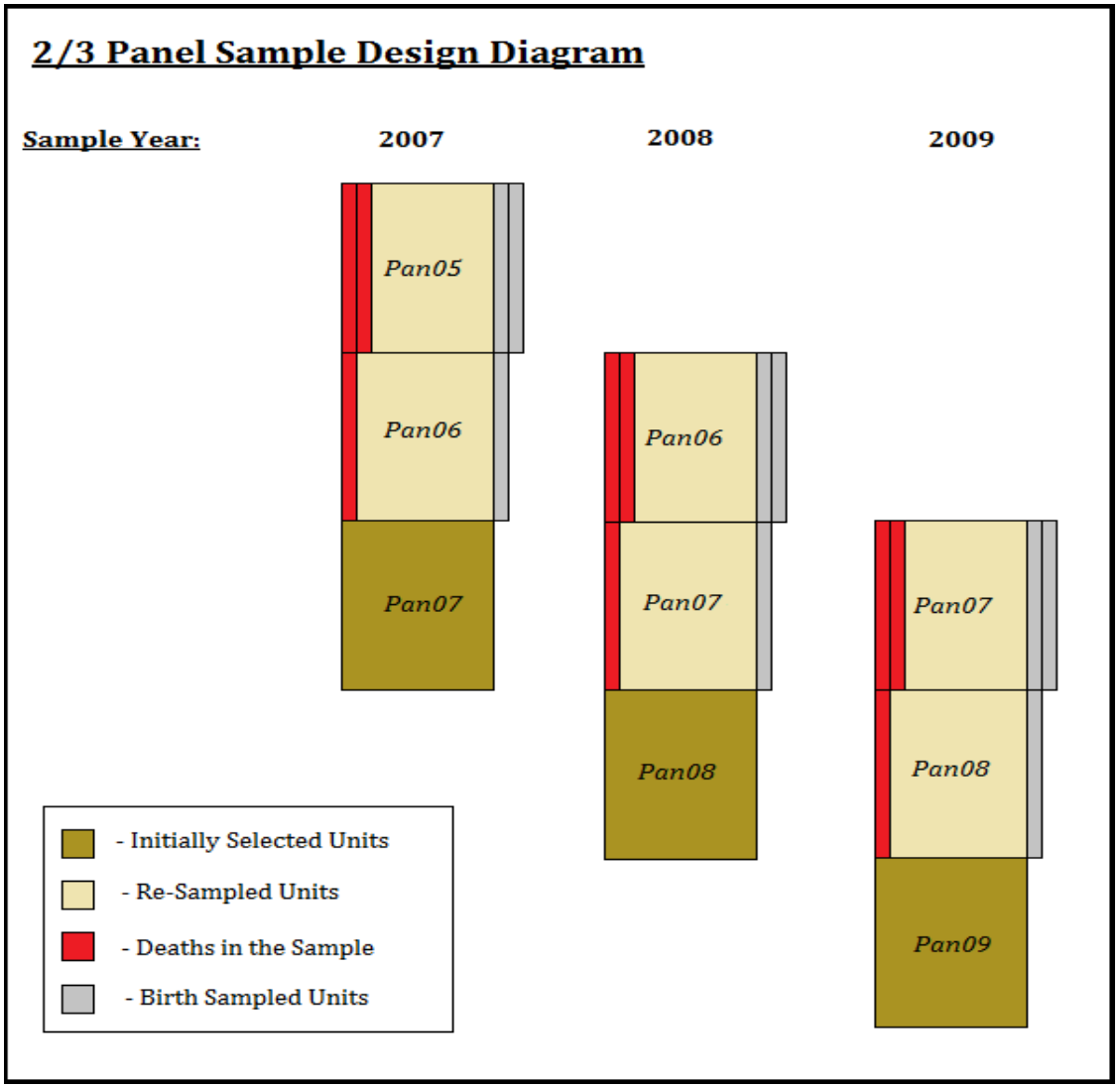
BLS

# Two-Thirds Rotating Panel (PAN) Design

- Some Details:
  - Frame is randomly split into three identical sub-frames
  - Each year only one sub-frame has new allocation and sample
  - Each year two sub-frames use the same sample from last year
  - 3 year cycle so that for any given year:
    - One third of the sample is newly allocated/selected
    - One third of the sample was allocated/selected last year
    - One third of the sample was allocated/selected two years ago
  - Overlapping samples are updated by removing deaths and sampling births

BLS

# Two-Thirds Rotating Panel (PAN) Design



# Current Estimation Methods

- 6 panels of OES data are combined
  - Employment updated by benchmarking to QCEW
  - Wages updated by Employer Cost Index (ECI)
- Impute for non-response
- Weight adjustments for atypical data
- Occupation Employment and Mean Wages
  - Different geography, industry, and ownership detail levels
- Direct estimation (design-based)
  - Sample is weighted up to make estimates
  - Employment = weighted total
  - Wages = weighted mean

BLS

# Alternative Estimation Method

- Prediction Theory – use OES respondents to build a model to predict occupational employment/wages for non-responding/non-sampled establishments
  - End Result: Every establishment on frame has occupational employment/wage data
- Modeled Estimates (instead of design-based)
  - Sample weights used in wage adjustment model
- Occupational Employment and Mean Wages
  - Different geography, industry, and ownership detail levels
  - Also can produce change estimates

BLS

# Model-Based Estimation Method (MB3): Intuition

- Goal: predict the staffing pattern and wages of non-sampled units in the target population using sampled establishments (over the previous three years) in the same four or five digit industry and state
- In the estimator, sample units are weighted according to their proximity to the non-sampled unit in four dimensions
  - Penalty for different six-digit industry
  - Penalty for different detailed area within the state
  - Penalty for employment differences
  - Penalty for sampled in a previous year
- Similar to current employment imputation procedure but without hierarchical structure

BLS

# MB3 Example

- Estimation cells
  - Observationally equivalent establishments
    - Defined by detailed industry - detailed area - employment (continuous)
  - Some contain lots of establishments, some a single unit
  - Predictions are identical for all establishments within cell
  - For computational efficiency purposes only – identical results if processed establishment-by-establishment
- Consider the following estimation cell:
  - New Single-Family Housing Construction (236115)
  - Located in Birmingham, AL
  - With 30 workers
- See table 1, 2.a, and 2.b

BLS

**Table 1 -- Characteristics and Weights of the Ten Best Matches**

| Match | Area | Strata Area | Sample Year | Reported Employment | Employment Weight | Area and Year Weight | Total Weight | Relative Weight |
|---|---|---|---|---|---|---|---|---|
| 1 | Birmingham | Central Alabama MSAs | 2007 | 49 | 0.75949 | 1.00000 | 0.75949 | 0.25845 |
| 2 | Tuscaloosa | Central Alabama MSAs | 2007 | 22 | 0.84615 | 0.66667 | 0.56410 | 0.19196 |
| 3 | Gadsden | Central Alabama MSAs | 2006 | 58 | 0.68182 | 0.44444 | 0.30303 | 0.10312 |
| 4 | Auburn-Opelika | Southern Alabama MSAs | 2007 | 23 | 0.86792 | 0.33333 | 0.28930 | 0.09845 |
| 5 | Mobile | Southern Alabama MSAs | 2006 | 27 | 0.94737 | 0.22222 | 0.21052 | 0.07164 |
| 6 | Huntsville | Northern Alabama MSAs | 2006 | 26 | 0.92857 | 0.22222 | 0.20635 | 0.07022 |
| 7 | Montgomery | Southern Alabama MSAs | 2006 | 38 | 0.88235 | 0.22222 | 0.19608 | 0.06672 |
| 8 | Balance of State Area | Balance of State Areas | 2006 | 39 | 0.86957 | 0.22222 | 0.19324 | 0.06576 |
| 9 | Balance of State Area | Balance of State Areas | 2005 | 29 | 0.98305 | 0.11111 | 0.10923 | 0.03717 |
| 10 | Huntsville | Northern Alabama MSAs | 2005 | 28 | 0.96552 | 0.11111 | 0.10728 | 0.03651 |

Note: Never-responding unit in the 2007 frame is a New Single-Family Housing Construction Company located in Birmingham, AL with 30 workers.

Note: All matches are from the same detailed industry as the never-responding unit.

BLS

**Table 2.a -- Reported Employment for Carpenters by Wage Interval and All Other Workers**

| | | | Carpenters | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Match | Reported Employment | Relative Weight | B | C | D | E | F | G | H | All Other Workers |
| 1 | 49 | 0.25845 | | 3 | 5 | 3 | 4 | 1 | | 33 |
| 2 | 22 | 0.19196 | | 1 | 5 | 1 | 1 | | | 14 |
| 3 | 58 | 0.10312 | 1 | 8 | 3 | 5 | | | | 41 |
| 4 | 23 | 0.09845 | | | | | | | 4 | 19 |
| 5 | 27 | 0.07164 | 5 | 7 | 1 | 2 | | | | 12 |
| 6 | 26 | 0.07022 | | 1 | 1 | | | | | 24 |
| 7 | 38 | 0.06672 | | | 2 | | 1 | | | 35 |
| 8 | 39 | 0.06576 | | | 9 | 8 | 1 | 1 | | 20 |
| 9 | 29 | 0.03717 | | | 2 | 1 | 1 | | | 25 |
| 10 | 28 | 0.03651 | | | | 3 | 3 | | | 22 |

**Table 2.b -- Employment Shares for Carpenters by Wage Interval and All Other Workers**

| | | | Carpenters | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Match | Reported Employment | Relative Weight | B | C | D | E | F | G | H | All Other Workers |
| 1 | 49 | 0.25845 | | 0.06 | 0.10 | 0.06 | 0.08 | 0.02 | | 0.67 |
| 2 | 22 | 0.19196 | | 0.05 | 0.23 | 0.05 | 0.05 | | | 0.64 |
| 3 | 58 | 0.10312 | 0.02 | 0.14 | 0.05 | 0.09 | | | | 0.71 |
| 4 | 23 | 0.09845 | | | | | | | 0.17 | 0.83 |
| 5 | 27 | 0.07164 | 0.19 | 0.26 | 0.04 | 0.07 | | | | 0.44 |
| 6 | 26 | 0.07022 | | 0.04 | 0.04 | | | | | 0.92 |
| 7 | 38 | 0.06672 | | | 0.05 | | 0.03 | | | 0.92 |
| 8 | 39 | 0.06576 | | | 0.23 | 0.21 | 0.03 | 0.03 | | 0.51 |
| 9 | 29 | 0.03717 | | | 0.07 | 0.03 | 0.03 | | | 0.86 |
| 10 | 28 | 0.03651 | | | | 0.11 | 0.11 | | | 0.79 |

BLS

# Wage Estimation under MB3

- Estimating interval mean wages
  - Exploit the fact that the current sample from either RCS or PAN design is representative of the population
  - Create aggregate occupation and area cells comprised of similarly paid occupations and detailed areas
    - For each occupation, determine what interval the median wage falls into using current sample weighted information (e.g., nurses' wages fall in G, postsecondary economics teachers in H, fast food cooks in A)
    - Similarly, for each detailed area, determine what interval the median wage falls into using current sample weighted information (e.g., Boston wages fall in E, Chicago in D, Birmingham in C)
    - Usually around 40 to 50 aggregate occupation x area cells

BLS

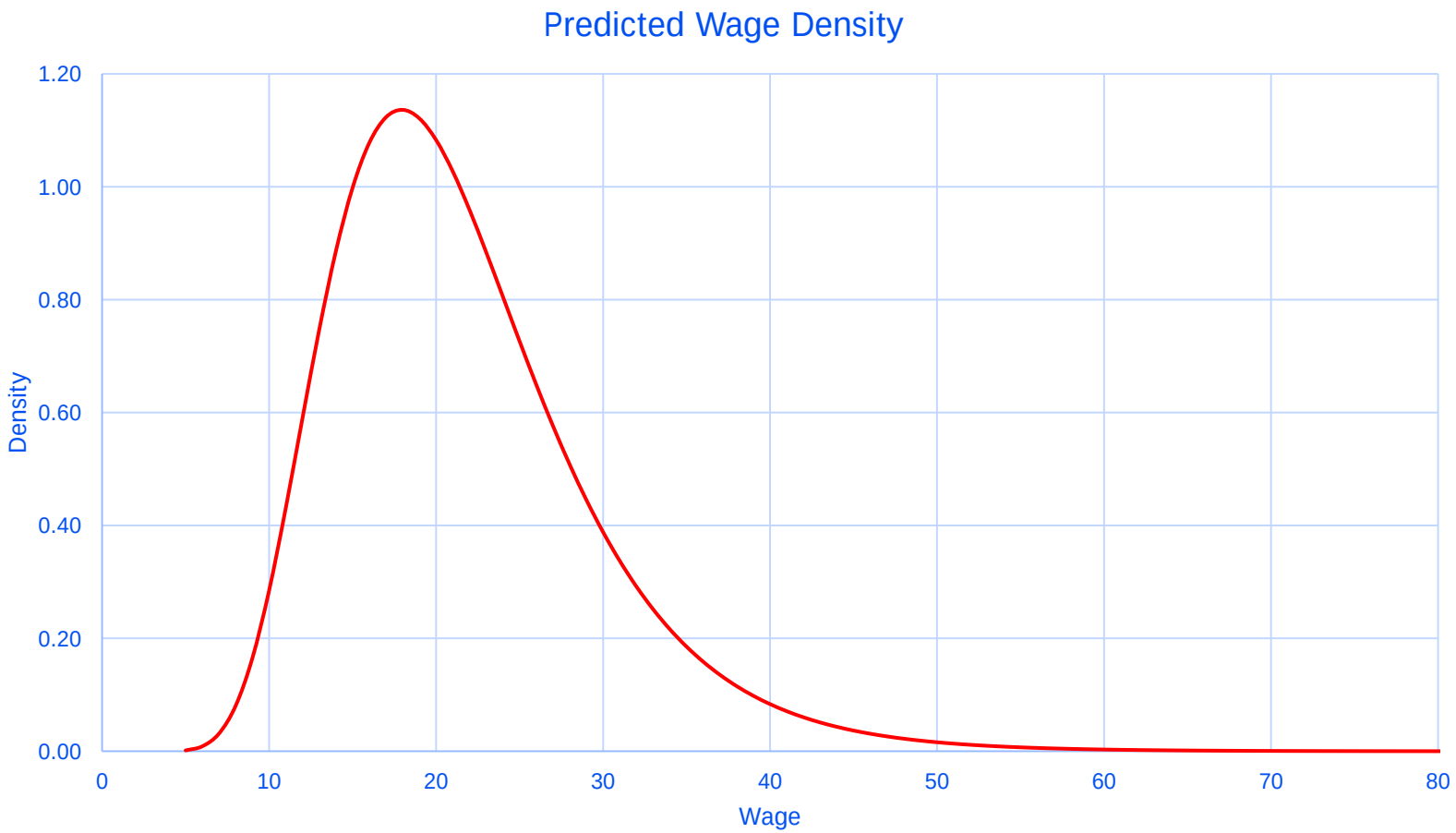
# Wage Estimation under MB3 (cont'd)

- Estimating interval mean wages
  - Assume the wages in each aggregate occupation x area cell follow a unique lognormal distribution
  - Estimate parameters of each lognormal distribution using maximum likelihood estimation
  - Graphs
  - Then directly compute interval means
  - See Table 3.a

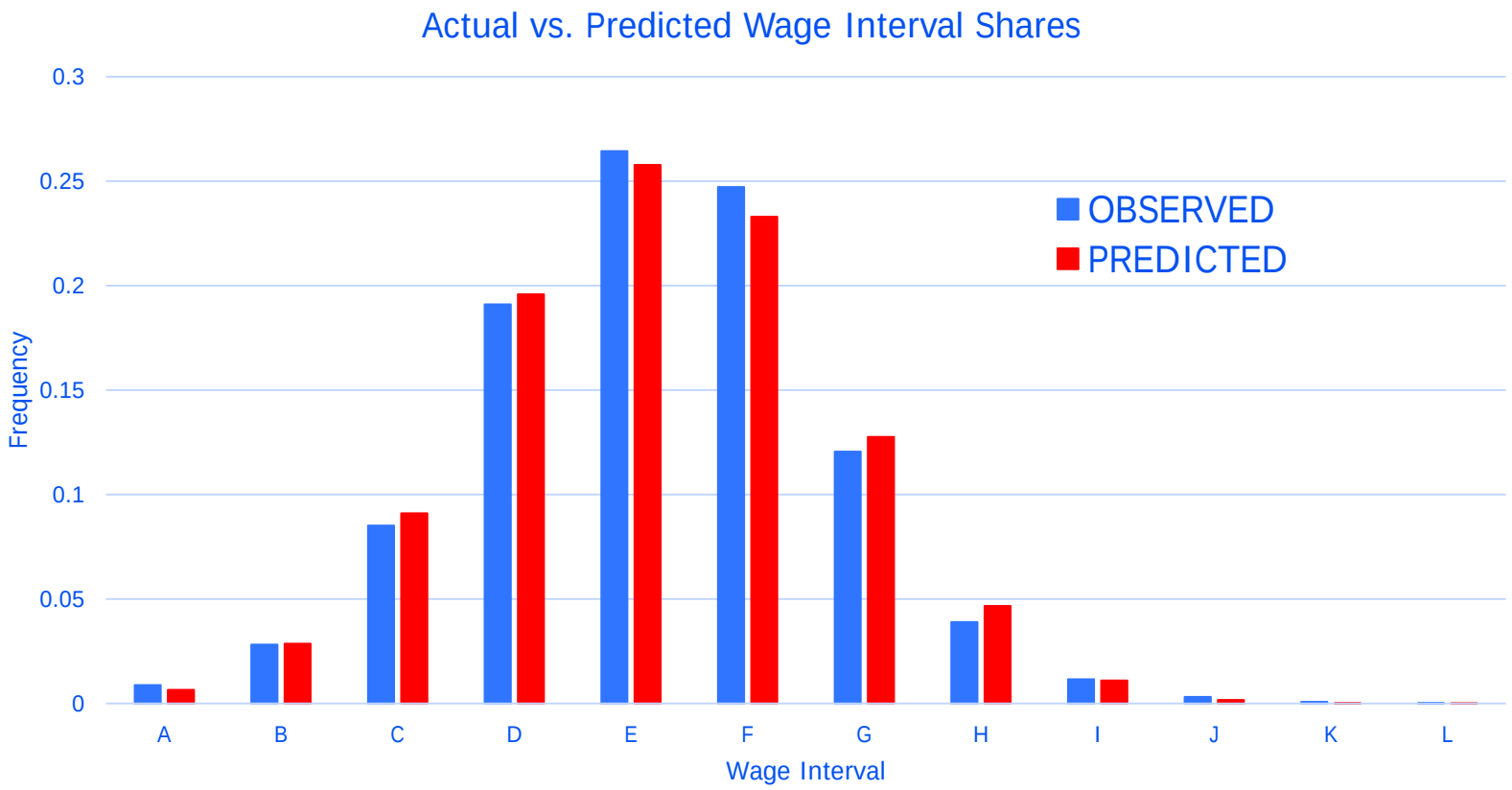
BLS

# Middle Wage Occupation Group

# Middle Wage Occupation Group



# Middle Wage Occupation Group



# Three Predicted Wage Densities

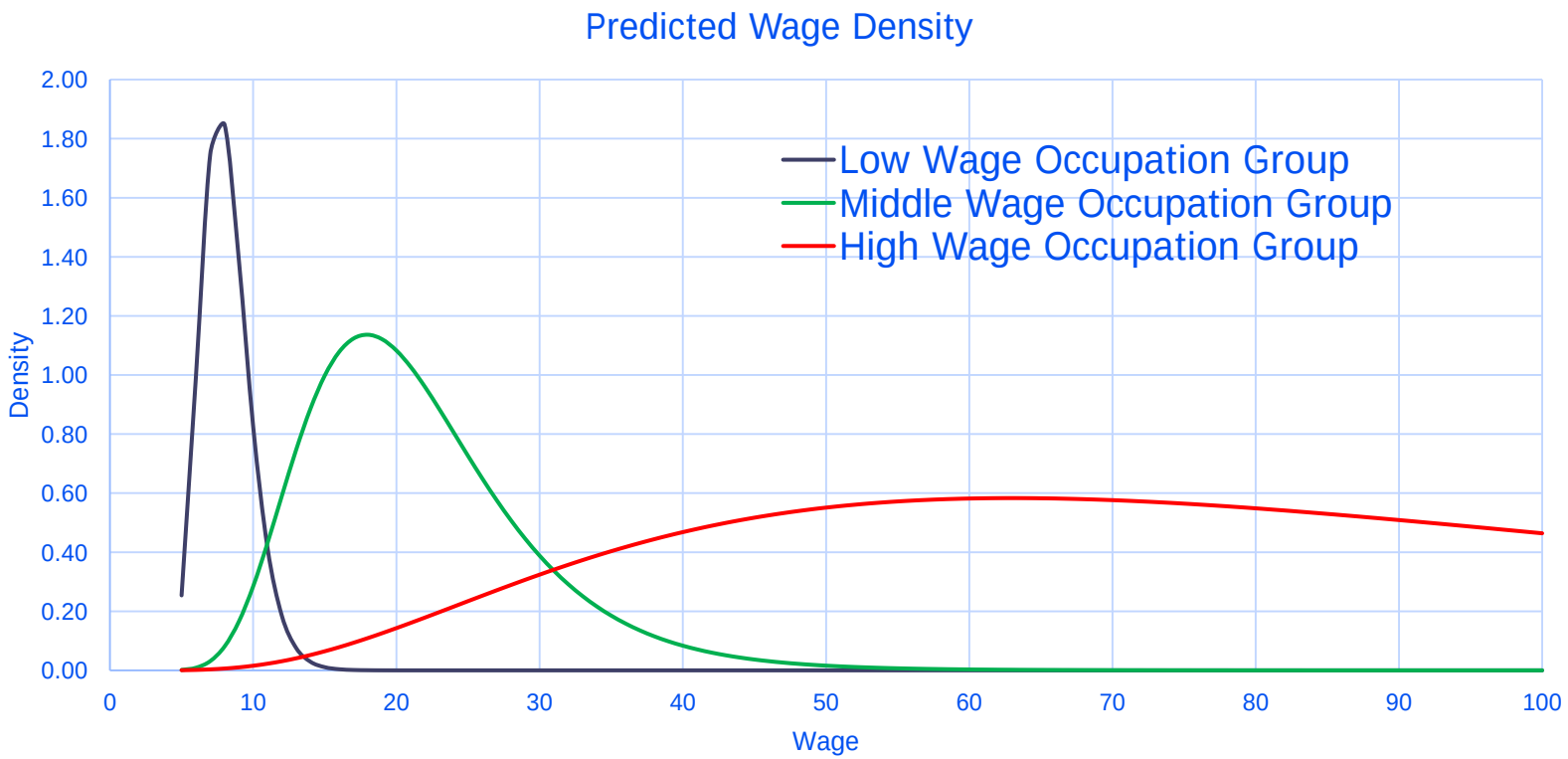


**Table 3.a -- Estimated Interval Mean Wages and Adjustment Factors for Carpenters**

| | | | Carpenters | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Match | Area | Sample Year | B | C | D | E | F | G | H | Adjustment factor |
| 1 | Birmingham | 2007 | | 10.88 | 13.67 | 17.14 | 21.51 | 27.05 | | 0.99989 |
| 2 | Tuscaloosa | 2007 | | 10.88 | 13.67 | 17.14 | 21.51 | | | 1.05298 |
| 3 | Gadsden | 2006 | 8.64 | 10.83 | 13.60 | 17.06 | | | | 1.17902 |
| 4 | Auburn-Opelika | 2007 | | | | | | | 33.74 | 1.07270 |
| 5 | Mobile | 2006 | 8.64 | 10.83 | 13.60 | 17.06 | | | | 1.13414 |
| 6 | Huntsville | 2006 | | 10.89 | 13.69 | | | | | 1.06110 |
| 7 | Montgomery | 2006 | | | 13.69 | | 21.59 | | | 1.09948 |
| 8 | Balance of State Area | 2006 | | | 13.60 | 17.06 | 21.42 | 26.96 | | 1.16484 |
| 9 | Balance of State Area | 2005 | | | 12.22 | 15.25 | 19.09 | | | 1.18476 |
| 10 | Huntsville | 2005 | | | | 15.25 | 19.09 | | | 1.08931 |

Note: Intervals correspond to sample year. For example, in 2007 interval E includes wages from $15.25 to $19.24 per hour, but includes wages from $13.50 to $16.99 per hour in 2005.

BLS

# Wage Estimation under MB3 (cont'd)

- Under MB3 approach, matches may differ from non-sampled unit in four dimensions:
  - Area
  - Industry
  - Year
  - Employment
- Need to adjust wages of matches into current, local dollars
  - Standard log wage regression on current data
  - Main effects for occupation, area, industry by strata area, and employer size
  - Coefficients vary across time

BLS

# Wage Estimation under MB3 (cont'd)

- Adjustment factor
  - Ratio of predicted wage for an occupation in non-sampled establishment to predicted wage for an occupation in the match
  - For every occupation reported by a match there is a unique adjustment factor
  - See table 3.a
- Interval shifts
  - Adjusted wages in current dollars
  - May need to shift wage interval employment if adjusted wage falls into a different interval
  - See table 3.b and 4.a

BLS

**Table 3.a -- Estimated Interval Mean Wages and Adjustment Factors for Carpenters**

| | | | Carpenters | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Match | Area | Sample Year | B | C | D | E | F | G | H | Adjustment factor |
| 1 | Birmingham | 2007 | | 10.88 | 13.67 | 17.14 | 21.51 | 27.05 | | 0.99989 |
| 2 | Tuscaloosa | 2007 | | 10.88 | 13.67 | 17.14 | 21.51 | | | 1.05298 |
| 3 | Gadsden | 2006 | 8.64 | 10.83 | 13.60 | 17.06 | | | | 1.17902 |
| 4 | Auburn-Opelika | 2007 | | | | | | | 33.74 | 1.07270 |
| 5 | Mobile | 2006 | 8.64 | 10.83 | 13.60 | 17.06 | | | | 1.13414 |
| 6 | Huntsville | 2006 | | 10.89 | 13.69 | | | | | 1.06110 |
| 7 | Montgomery | 2006 | | | 13.69 | | 21.59 | | | 1.09948 |
| 8 | Balance of State Area | 2006 | | | 13.60 | 17.06 | 21.42 | 26.96 | | 1.16484 |
| 9 | Balance of State Area | 2005 | | | 12.22 | 15.25 | 19.09 | | | 1.18476 |
| 10 | Huntsville | 2005 | | | | 15.25 | 19.09 | | | 1.08931 |

Note: Intervals correspond to sample year. For example, in 2007 interval E includes wages from $15.25 to $19.24 per hour, but includes wages from $13.50 to $16.99 per hour in 2005.

BLS

**Table 3.b -- Adjusted Interval Mean Wages for Carpenters**

| | | | Carpenters | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Match | Area | Sample Year | B ($7.50 to $9.49) | C ($9.50 to $11.99) | D ($12.00 to $15.24) | E ($15.25 to $19.24) | F ($19.25 to $24.49) | G ($24.50 to $30.99) | H ($31.00 to $39.24) | Interval Shift |
| 1 | Birmingham | 2007 | | 10.88 | 13.67 | 17.14 | 21.51 | 27.05 | | |
| 2 | Tuscaloosa | 2007 | | 11.46 | 14.39 | 18.05 | 22.65 | | | |
| 3 | Gadsden | 2006 | | 10.19 | 12.77 | 16.03 | 20.11 | | | Yes |
| 4 | Auburn-Opelika | 2007 | | | | | | | 36.19 | |
| 5 | Mobile | 2006 | | 9.80 | 12.28 | 15.42 | 19.35 | | | Yes |
| 6 | Huntsville | 2006 | | 11.56 | 14.53 | | | | | |
| 7 | Montgomery | 2006 | | | 15.05 | | 23.74 | | | |
| 8 | Balance of State Area | 2006 | | | | 15.84 | 19.87 | 24.95 | 31.40 | Yes |
| 9 | Balance of State Area | 2005 | | | 14.48 | 18.07 | 22.62 | | | |
| 10 | Huntsville | 2005 | | | | 16.61 | 20.79 | | | |

Note: Intervals now correspond to the current year. The interval shift column indicates whether wages (and subsequently employment) moved from one interval to another.

BLS

**Table 4.a -- Adjusted Employment Shares for Carpenters by Wage Interval and All Other Workers**

| | | | Carpenters | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Match | Reported Employment | Relative Weight | B ($7.50 to $9.49) | C ($9.50 to $11.99) | D ($12.00 to $15.24) | E ($15.25 to $19.24) | F ($19.25 to $24.49) | G ($24.50 to $30.99) | H ($31.00 to $39.24) | All Other Workers |
| 1 | 49 | 0.25845 | | 0.06 | 0.10 | 0.06 | 0.08 | 0.02 | | 0.67 |
| 2 | 22 | 0.19196 | | 0.05 | 0.23 | 0.05 | 0.05 | | | 0.64 |
| 3 | 58 | 0.10312 | | 0.02 | 0.14 | 0.05 | 0.09 | | | 0.71 |
| 4 | 23 | 0.09845 | | | | | | | 0.17 | 0.83 |
| 5 | 27 | 0.07164 | | 0.19 | 0.26 | 0.04 | 0.07 | | | 0.44 |
| 6 | 26 | 0.07022 | | 0.04 | 0.04 | | | | | 0.92 |
| 7 | 38 | 0.06672 | | | 0.05 | | 0.03 | | | 0.92 |
| 8 | 39 | 0.06576 | | | | 0.23 | 0.21 | 0.03 | 0.03 | 0.51 |
| 9 | 29 | 0.03717 | | | 0.07 | 0.03 | 0.03 | | | 0.86 |
| 10 | 28 | 0.03651 | | | | 0.11 | 0.11 | | | 0.79 |

BLS

# Wage Estimation under MB3 (cont'd)

- Putting it all together
  - For each match we have occupation - (current) wage interval employment shares (table 4.a) and mean wages (table 3.b)
  - Predicted employment for an occupation-interval is weighted sum over matches
    - Sum of Employment Share x weight x employment of non-sampled unit
  - Predicted total wage for an occupation-interval is weighted sum over matches
    - Sum of Employment Share x mean wage x weight x employment of non-sampled unit
  - See table 4.b

BLS

**Table 4.b -- Weighted Employment Levels for Carpenters by Wage Interval and All Other Workers**

| | | | Carpenters | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Match | Reported Employment | Relative Weight | B ($7.50 to $9.49) | C ($9.50 to $11.99) | D ($12.00 to $15.24) | E ($15.25 to $19.24) | F ($19.25 to $24.49) | G ($24.50 to $30.99) | H ($31.00 to $39.24) | All Other Workers | Total |
| 1 | 49 | 0.25845 | | 0.47 | 0.79 | 0.47 | 0.63 | 0.16 | | 5.22 | 7.75 |
| 2 | 22 | 0.19196 | | 0.26 | 1.31 | 0.26 | 0.26 | | | 3.66 | 5.76 |
| 3 | 58 | 0.10312 | | 0.05 | 0.43 | 0.16 | 0.27 | | | 2.19 | 3.09 |
| 4 | 23 | 0.09845 | | | | | | | 0.51 | 2.44 | 2.95 |
| 5 | 27 | 0.07164 | | 0.40 | 0.56 | 0.08 | 0.16 | | | 0.96 | 2.15 |
| 6 | 26 | 0.07022 | | 0.08 | 0.08 | | | | | 1.94 | 2.11 |
| 7 | 38 | 0.06672 | | | 0.11 | | 0.05 | | | 1.84 | 2.00 |
| 8 | 39 | 0.06576 | | | | 0.46 | 0.40 | 0.05 | 0.05 | 1.01 | 1.97 |
| 9 | 29 | 0.03717 | | | 0.08 | 0.04 | 0.04 | | | 0.96 | 1.12 |
| 10 | 28 | 0.03651 | | | | 0.12 | 0.12 | | | 0.86 | 1.10 |
| **Predicted Employment** | | | | 1.27 | 3.35 | 1.59 | 1.93 | 0.21 | 0.56 | 21.09 | 30.00 |
| **Predicted Mean Wages** | | | | 10.67 | 13.69 | 16.70 | 20.99 | 26.54 | 35.76 | | |

BLS

# Testing the Alternative Methods

- Created a fictional population where occupational employment and wages were known for all establishments
  - Covered 18 states
  - Include small / medium / large (within limit) states across all census regions
  - Responders use their data directly
  - Non-responders use mixing algorithm to borrow from across both time and space
  - Covers 2004 to 2009
- Created population = truth for testing

BLS

# Testing the Alternative Methods

- Used a Simulation Study
- Selected 100 samples using current, RCS, and PAN sample designs
- Create estimates using current, and proposed model methods
- Did this 6 times – 2004 to 2009
- Calculate statistics showing how well each method does at estimating yearly and change estimates

BLS

# Summary of Results

- Let's look at tables of results (separate document)
- The alternative methods systematically outperform the current OES methods
- Current methods performs the best when occupational estimates are stable
- The model based estimates using the RCS design have the best overall results
- The model based estimates using the PAN design have comparable (and sometime better) results than the model based estimates using the RCS design for estimates of change

BLS

# Next Steps

- Research phase is coming to an end
  - Methods for measuring reliability still need to be developed
  - Possible testing of alternative allocations
- Decision phase is beginning
  - Should we pursue these methods?
  - How would we implement these methods?
    - What tools are needed for implementation?
    - What is the most efficient way to implement these methods?
    - How will we determine the aggregate areas to sample by?
    - Etc.

BLS

# Thanks!