Model-based Estimates for Farm Labor Quantities

Lu Chen[*,#], Nathan B. Cruze[+], Linda J. Young[#]

[*]National Institute of Statistical Sciences (NISS)

[#]United States Department of Agriculture
National Agricultural Statistics Service (NASS)

[+]NASA Langley Research Center

FCSM Session C-3: Advances in Small Area Estimation
October 25, 2022

Disclaimer and Acknowledgment

The findings and conclusions in this presentation are those of the authors and should not be construed to represent any official USDA or US Government determination or policy.

- This research was supported by the U.S. Department of Agriculture, National Agricultural Statistics Service.
- The authors thank Theresa Varner and Jennifer Rhorer NASS, USDA for their contributions of the success of the 2020 publication.
- They also acknowledge Andreea Erciulescu for her contributions in the early stage of the research.

Outline

Motivation

Data and Modeling Procedure

Models

Case Study

Concluding Remarks

Motivation

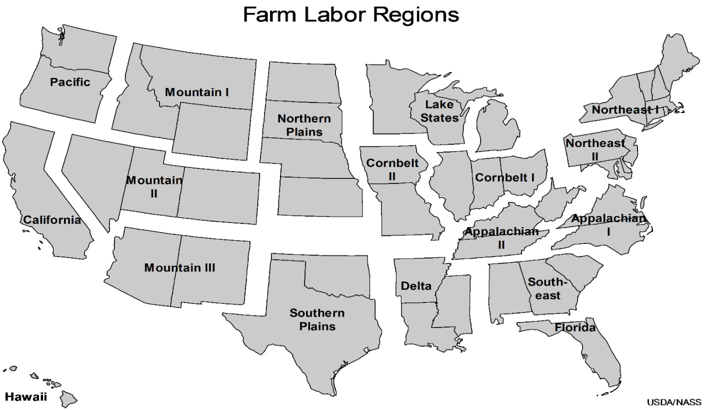
"Improving Farm Labor Estimates using small area models."

- The Farm Labor Survey conducted by NASS, USDA
- Important official statistics to various data users
- Required tabulations at different levels
- Sparse sample in some states for some cells
- Multiple data sources available

Question: How to construct modeling process to produce reliable and coherent estimates with measures of uncertainty for all required tabulations in the publication?

Motivation: Quantities of Interest

- Regional and US level estimates:



- **NASS Worker Types**; the Standard Occupational Classification (SOC)

Traditional USDA NASS Official Statistics

Agricultural (Farm) Labor Survey

- Time: biannual official statistics for four quarters
  - May (April and January) and November (October and July)

- Quantities: **number of workers, hours/week, wage rate**
  - Expert assessment
  - Point estimates only (no measures of uncertainty*)

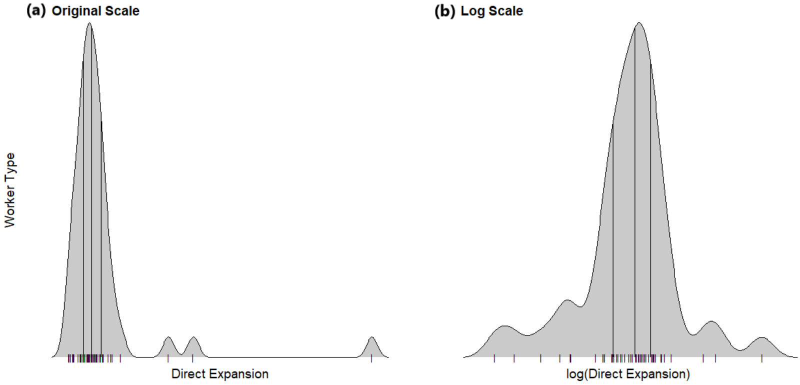
- Domains: **region, US, worker-type**
  - Different worker types: field, livestock, supervisor and others
  - Aggregation based on finer geographical or worker-type domain

*quality measures were historically published for some selective *survey* estimates.

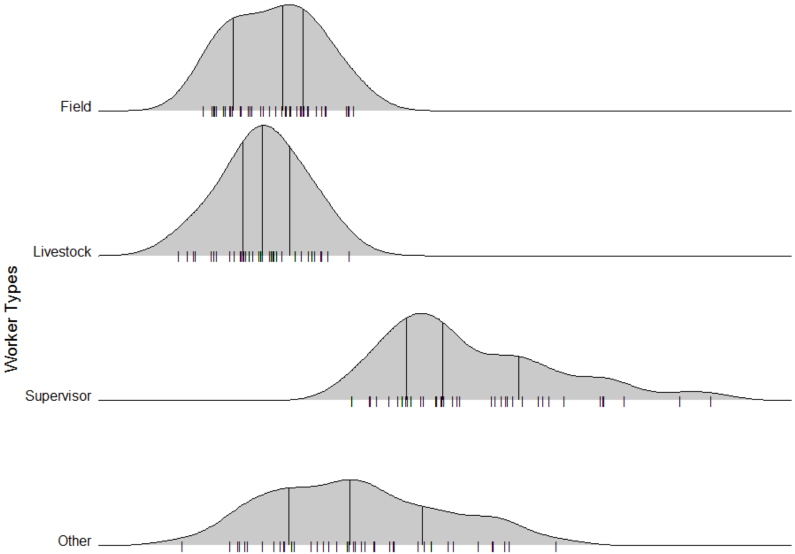
Current Modeling Application

- Model estimates as the key indicators for official statistics
  - Hierarchical Bayesian sub-area small area models produce all levels estimates by different NASS worker types
  - Associated measures of uncertainty published on quality measures
  - Harmony among nested levels and consistent ratio definitions
    - Geographic: State $\rightarrow$ Regional $\rightarrow$ US
    - Worker types: field + livestock, all hired
  - Transparent and reproducible method
  - Increase precision and reliability
- Published articles: Chen et al. (2022); Young and Chen (2022)

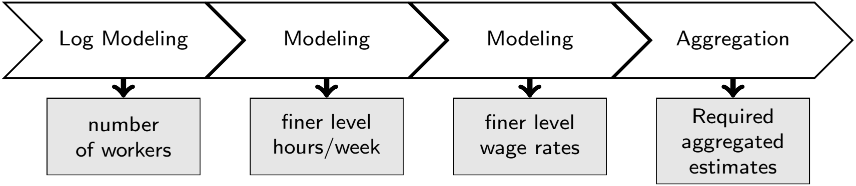
Direct Estimates — Number of Workers



Direct Estimates — Wage Rates by Types



Modeling Procedure



Input

- Finer level survey summaries: state $\times$ worker-type
- Previous year, same quarter, the official values and sample sizes: state $\times$ worker-type

Output

- Finer level and domain-level estimates
  - Point estimates, measures of uncertainty, distributions

Notation

- $i = 1, \ldots, m$ index for areas (i.e. regions)
- $j = 1, \ldots, n_i$ index for sub-areas (i.e. states) within area i
- $k = 1, ...K$ index for different NASS worker types
- $\hat{y}_{ijk}, \ \hat{\sigma}^2_{ijk,y}$ Farm Labor direct estimates by worker types
- x_{ijk} known auxiliary information: the previous year, same quarter, official estimates; number of positive responses; and worker types

Model for Number of Workers

The sub-area model:

$$\hat{\theta}_{ijk} = \log(\hat{y}_{ijk})|\theta_{ijk} \overset{ind}{\sim} N(\theta_{ijk},\ \hat{\sigma^*}^2_{ijk}),\ k = 1, \ldots, K,$$

$$\theta_{ijk}|\boldsymbol{\beta}, \nu_i, \sigma^2_\mu \overset{ind}{\sim} N(\mathrm{x}'_{ijk}\boldsymbol{\beta} + \nu_i, \sigma^2_\mu), j = 1, \ldots, n_i,$$

$$\nu_i|\sigma^2_\nu \overset{iid}{\sim} N(0,\ \sigma^2_\nu),\ i = 1, \ldots, m,$$

$$\boldsymbol{\beta} \sim MN(\hat{\boldsymbol{\beta}},\ 1000 \times \hat{\Sigma}_{\hat{\beta}}),$$

$$\sigma^2_\mu \sim \text{Uniform}(R^+),\ \sigma^2_\nu \sim \text{Uniform}(R^+),$$

where $\hat{\sigma^*}^2_{ijk} = (\hat{y}_{ijk})^{-2}\hat{\sigma}^2_{ijk,y}$ serves as estimate for the sampling variances.

- Goal:
  - State $\times$ type worker: $y^{wk}_{ijk} = \exp(\theta_{ijk})$

Model for Hours and Wage Rates

The sub-area model for hours/week and wage rates (Erciulescu et al. 2020):

$$\hat{\theta}_{ijk}|\theta_{ijk} \overset{ind}{\sim} N(\theta_{ijk},\ \hat{\sigma}^2_{ijk}),\ k = 1, \ldots, K,$$

$$\theta_{ijk}|\boldsymbol{\beta}, \nu_i, \sigma^2_\mu \overset{ind}{\sim} N(\mathrm{x}'_{ijk}\boldsymbol{\beta} + \nu_i, \sigma^2_\mu), j = 1, \ldots, n_i,$$

$$\nu_i|\sigma^2_\nu \overset{iid}{\sim} N(0,\ \sigma^2_\nu),\ i = 1, \ldots, m,$$

$$\boldsymbol{\beta} \sim MN(\hat{\boldsymbol{\beta}},\ 1000 \times \hat{\Sigma}_{\hat{\beta}}),$$

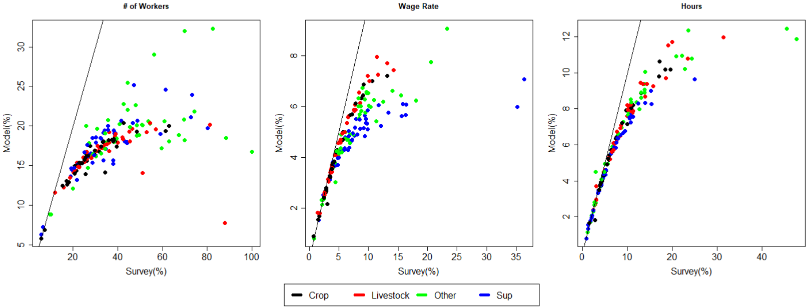
$$\sigma^2_\mu \sim \text{Uniform}(R^+),\ \sigma^2_\nu \sim \text{Uniform}(R^+),$$

- Goal:
  - State $\times$ type wage rate: $y^{wg}_{ijk} = \theta_{ijk}$
  - State $\times$ type hours/week: $y^{hr}_{ijk} = \theta_{ijk}$

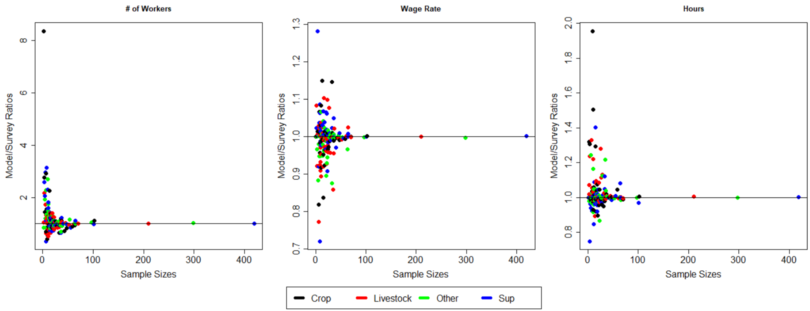
Case Study: 2022 January

- Example:
  - 44 states within 18 regions by worker types
  - Number of workers; hours/week; wage rates
  - Goal: regional and US level estimates by worker types or combined worker types
- Computation:
  - Rjags: 10,000 MCMC samples and 2,000 burn-in, 3 chains, each thinned every 8 samples, resulting in a number of 3,000 samples for inference
  - Convergence diagnostics are conducted: Rhat ≤ 1.1 and effective sample sizes are around 3,000

CV Comparisons by Worker Types at State Level



Model Effectiveness by Worker Types at State Level

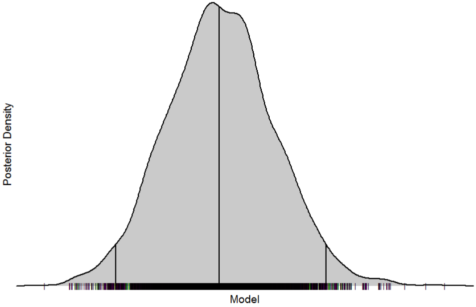


Number of Worker: Posterior Distribution

- **US $\times$ all hired worker** estimates, computed at the h^{th} draw:

$$y^{wk,(h)} = \sum_{k=1}^{K} \sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{ijk}^{wk,(h)},$$

  where $h = 1, \ldots H$ are the draws.
- Posterior distribution based on $y^{wk,(h)}$:

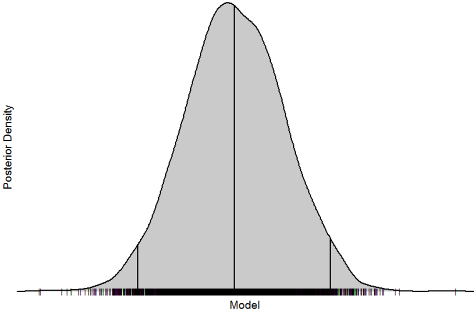


Wage Rate: Posterior Distribution

- **US** $\times$ **all hired worker** wage rate estimates, computed at the h^{th}:

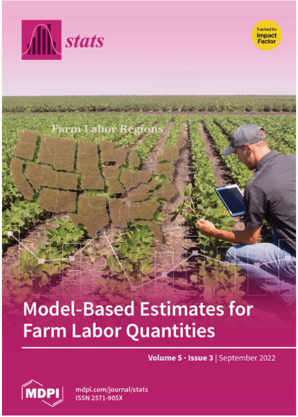
$$y^{wg,(h)} = \frac{\sum_{k=1}^{K} \sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{ijk}^{wk,(h)} y_{ijk}^{hr,(h)} y_{ijk}^{wg,(h)}}{\sum_{k=1}^{K} \sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{ijk}^{wk,(h)} y_{ijk}^{hr,(h)}},$$

  where $h = 1, \ldots H$ are the draws.
- Posterior distribution based on $y^{wg,(h)}$:



Concluding Remarks

- NASS incorporated model-based estimates into the official Farm Labor publication since 2020
- Increased the accuracy and improved the precision of estimates
- Harmony among nested levels and worker types
- Fast computation time within production window



Reference

Chen, L., Cruze, N. B., and Young, L. J. (2022).
Model-based estimates for farm labor quantities.
Stats, 5(3):738–754.

Erciulescu, A. L., Cruze, N. B., and Nandram, B. (2018).
Benchmarking a triplet of official estimates.
Environmental and Ecological Statistics, 25:523–547.

Erciulescu, A. L., Cruze, N. B., and Nandram, B. (2019).
Model-based county level crop estimates incorporating auxiliary sources of information.
Journal of the Royal Statistical Society: Series A (Statistics in Society), 182(1):283–303.

Erciulescu, A. L., Cruze, N. B., and Nandram, B. (2020).
Statistical challenges in combining survey and auxiliary data to produce official statistics.
Journal of Official Statistics, 36(1):63–88.

Rao, J. N. K. and Molina, I. (2015).
Small Area Estimation.
2015 John Wiley and Sons, Inc.

USDA NASS.
Farm labor publication.
https://usda.library.cornell.edu/concern/publications/x920fw89s.

Young, L. J. and Chen, L. (2022).
Using small area estimation to produce official statistics.
Stats, 5(3):881–897.

Thank You!

lu.chen@usda.gov