Agricultural Survey Quarterly Meeting

Joshua Moll,
Research Economist
Employment Security Department
Labor Market and Economic Analysis
Program Evaluation, Research & Analysis

Presented 02/01/2019

Employment Security Department
WASHINGTON STATE

Agenda

- Survey background
- Overview of H-2A program
- Employer estimation
- Moving forward

Survey background

- What:
  - Wage rates and employment practices for agricultural worksites in Washington state
- Why:
  - U.S. Department of Labor (USDOL) requires survey every year for occupations and activities that involve temporary foreign workers
  - USDOL uses survey results to establish wage rates and employment standards for agricultural employment contracts

Survey background

- Who:
  - Agricultural business under certain industry codes (NAICS)
  - Agricultural workers involved in apple and cherry harvesting
- How:
  - Survey development and administration
  - Data collected is aggregated and analyzed by ESD
  - Results provided on Employment & Training Administration (ETA) 232 forms

Overview of H-2A program

- Regulated by USDOL
- Used when there is a perceived shortage of domestic workers
- Employment is seasonal or temporary
- Employment of H-2A worker must not negatively impact wages and employment practices for similarly employed domestic workers

Employer estimation

- ESD requirements
- Caveats of estimation
- Estimation method
  - Method overview
  - Method assumptions
  - Analytical steps
- Results of application
  - Industry estimation
  - Crop estimation
  - Crop variety estimation

Employer estimation: ESD requirements

- USDOL/ETA form 232 requires:
  - Total number of employers contacted during the survey
  - Total number of respondents
  - Total number of U.S. workers reported by employers
  - Estimated number of U.S. workers
  - Estimated number of employers
    - Estimated number of crop variety growers

Employer estimation: caveats of estimation

- ESD administrative databases are limited:
  - Unemployment Insurance covers employers by NAICS and worksite location
    - Recorded by NAICS industry, not by crop or crop-variety
    - Single worksites can produce multiple crops and crop varieties
  - Reporting lag
  - Administrative databases do not tell us who qualifies for the survey

Employer estimation: method overview

- Classical capture-recapture estimators:
  - Classical experiment is to study the demographic characteristics of an animal population and determine the population size
  - Animals are captured, marked with a tag and released back into the population
    - The operation gets repeated several times
  - Each animal is associated with a capture history
    - Capture histories indicate a “catch” or a “miss” by a binary vector (1 or 0)
- General form of a population size estimator:
  - $\widehat{N} = n + \mu_0$
    - n, is the number of units caught at least once
    - μ_0, is the estimated number of units missed
    - $\widehat{N}$, is the estimated population size

Employer estimation: method overview continued…

- History of capture-recapture and applications:
  - Originally developed in the field of wildlife management (*Petersen, 1896*)
    - Petersen estimator
  - Gained popularity with a treatment by Chapman (*Chapman, 1951*) in the field of ecology
  - Log-linear treatment of capture-recapture estimators was later applied by Fienberg and Cormack (*Fienberg, 1972; Cormack, 1989*) to deal with heterogeneity of individual behaviors, which can bias estimators of abundance
  - Has been further applied to fields such as: epidemiology, the evaluation of census undercount and software testing (*International Working Group for Disease Monitoring and Forecasting, 1995 a,b; Darroch, Fienberg, Glonek & Junker 1993; Wohlin, Runeson & Brantestam 1995; Ebrahimi 1997; Briand, El Emam, Freimut & Leiterberger, 2000*)

Employer estimation: method overview continued…

- Log-linear models for capture recapture:
  1) Determine the probability of a unit to experience a capture history
     - Example: Determine the likelihood of a crop-variety firm responding to the surveys
  2) From understanding the probability of capture, the expected number of units having a capture history can be determined
  3) The expected number of units having a capture history then is re-expressed as a log-linear model
     - Expression as a log-linear model aids in reducing inherent bias from the data and allows the fitting of a regression model to estimate abundance
  4) Fit a log-linear model
     - Poisson regression, deals with count data
     - Helps us identify bias, correct any bias found and produce a stable estimate
     - Enables the estimation of firms missed during the search occasions
  5) Abundance estimation
     - Produces final abundance estimate
     - Uses the number found at least once and the estimated number missed

Employer estimation: method overview continued…

- Base types of general linear models:
  - M_0: all capture occasions are independent with a common probability of being caught
  - M_t: each capture occasion has it's own capture probability (temporal effect or change)
    - Best suited for three or more search occasions
  - M_b: a unit's behavior changes after the first capture (behavioral effect or change)
    - Best suited for three or more search occasions

Employer estimation: general model requirements and assumptions

- General model requirements:
  - Have at least two capture occasions
    - *Example:* Two agricultural survey iterations
  - Capture occasions occur over a short period of time
  - Search procedures are conceptually equivalent
    - *Example:* Survey forms and the type of search being conduct are the same
- Assumptions:
  - Population in question is closed:
    - The population is finite
    - Immigration into the population area is negligible
    - Mortality rates are negligible
    - *Example:* the size of the closed population does not drastically vary over a short period of time

Employer estimation: overview continued…

- Log-linear model fitted with a Poisson Regression for capture-recapture experiments (M_0):
  1) Probability of a unit to experience a capture history,ω,:
     - $\Pr(\omega) = (1-p)^{t-\sum \omega_j} p^{\sum \omega_j}$
       - t = *capture occasions*
       - p = *single capture probability to all units*
       - $\sum \omega_j$ = *the number of times the unit is caught*
  2) Therefore, the expected number of units in the population having a capture history ω is:
     - $\mu_\omega = N(1-p)^{t-\sum \omega_j} p^{\sum \omega_j}$
  3) Expected frequency re-expressed as a log-linear model:
     - $\mu_\omega = exp\left(\log(N\left((1-p)^t\right) + \sum \omega_j \log\left(\frac{p}{1-p}\right)\right)$
  4) Fit a log-linear model:
     - $E(Y) = exp(X\beta)$
       - Y is equeal to the $(2^t - 1) \times 1$ vector of the observed frequencies n_ω
       - X *is a* $(2^t - 1) \times 2$ design matrix
       - $\beta = (\gamma, \beta)^t$
  5) Abundance estimate:
     - $\widehat{N} = n + exp(\gamma)$
     - $exp(\gamma) = exp\left(\log(N\left((1-p)^t\right)\right) = N(1-p)^t = N \times Pr(\omega_0) = \mu_0$
       - ω_0 = the unobservable capture history of zero capture
       - μ_0 = the expected number of units never captured

Employer estimation: analytical steps

Descriptive statistics

- Transform data to a usable format (matrix of capture histories)
  - Assign binary indicator for each capture occasion
- Produce descriptive statistics for capture-recapture data

Model fitting

- Fit various loglinear models for a closed population
  - M_0, M_t, M_b

Model selection

- Produce fit statistics for the number of captures on each capture occasion and model performance.
  - AIC, BIC, standard error, etc.
- Using model fit statistics select the model to be used for estimation

Abundance estimate

- Apply the selected model to compute the abundance estimation and 95% confidence interval of a closed population

Employment Security Department
WASHINGTON STATE

Results: application to estimate industry firm abundance

- Method was applied to survey data collected from 2015 and 2017:
  - 2015 and 2017 data was made compatible in order to apply this technique
    - 2017 survey data was far more granular in terms of what crop-varieties were allowed to be report
  - Comparison against adjusted 2017 average annual firm counts by six digit NAICS code from QCEW
    - QCEW firm counts were adjusted to meet the scope of the survey
    - Ratios of eligibility were extracted from 2015 (74%) and the most recent 2018 (80%) survey disposition records and then averaged
    - Therefore, on average 77% (0.77) are considered eligible under the scope of the survey
      - Example: $100_{total\ firms} \times 0.77_{eligible} = 77_{adjusted\ firms}$

Results: Industry estimates

Industry (NAICS)	Adjusted 2017 QCEW firm count	Abundance estimate	AE	APE	Low 95	Hi 95
Other vegetable and melon farming	225	181	44	20%	128	284
Apple orchards	588	549	39	7%	483	633
Grape vineyards	156	149	7	4%	118	201
Berry (except strawberry) farming	176	180	4	2%	137	253
Fruit and tree nut combination farming	18	13	5	28%	8	>37.5
Other noncitrus fruit farming	713	695	18	3%	625	782
All other miscellaneous crop farming	209	217	8	4%	129	442

Results: Crop estimates

Crop	Abundance estimate	Low 95	Hi 95	Occasion 1 (2015)	Occasion 2 (2017)	Both occasions
Apples	943	830	1086	316	292	98
Berries	249	191	344	61	87	22
Cherries	759	665	880	235	276	86
Grapes	266	200	379	70	76	20
Pears	513	418	649	131	159	41

Results: Crop variety estimates

Crop	Variety	Abundance estimate	Low 95	Hi 95	Occasion 1 (2015)	Occasion 2 (2017)	Both occasions
Apple	Braeburn	105	42	>315	11	18	2
Apple	Cripps pink	113	45	>338	5	25	2
Apple	Fuji	360	247	577	61	81	14
Apple	Gala	646	506	859	133	159	33
Apple	Golden delicious	439	324	634	82	110	21
Apple	Granny smith	455	278	865	54	74	9
Apple	Honeycrisp	476	327	757	56	113	15
Apple	Red delicious	423	310	618	63	121	20

Results: Crop variety estimates continued…

Crop	Variety	Abundance estimate	Low 95	Hi 95	Occasion 1 (2015)	Occasion 2 (2017)	Both occasions
Berry	Blueberry	182	117	328	35	46	9
Berry	Raspberry	69	51	104	22	33	11
Berry	Strawberry	37	20	105	9	12	3
Cherry	Dark red	444	332	641	40	200	18
Cherry	Red	725	551	1001	167	118	28
Cherry	Yellow	441	308	685	57	111	16
Pear	Bartlett	400	308	547	83	121	26
Pear	Bosc	469	200	>1406	18	57	3
Pear	D’anjou	355	248	557	60	86	15

Moving forward

- January 31st, 2019:
  - Survey administration and data collection closed
  - Worker survey response rate: 42.91%
  - Employer survey response rate (1/20/2019): 42.14%
- February 28th, 2019:
  - University of Washington delivers final survey data set to LMEA
- March, 2019:
  - Agricultural survey quarterly meeting to discuss worker estimation method (announcement of date and time will follow shortly)
  - Final employer and worker survey analysis and estimation
- April, 2019:
  - Conference call with all stakeholders presenting final results
    - Feedback period of approximately one week
  - Submission of final results to USDOL
    - Publication of final results is contingent upon USDOL
  - Begin administrative planning for 2019 survey iteration

References

- Briand, L.C., El Emam, K., Freimut B.G. & Leiterberger O. (2000). “A comprehensive evaluation of capture-recapture models for estimating software defect content. *IEEE Transactions on Software Engineering,* **26,** 518-540
- Chapman (1951). “Some properties of the hypergeometric distribution with applications to zoological census.” *University of California Public. Stat.* **1,** 131-160.
- Cormack, RM (1989). “Loglinear models for Capture-Recapture.” *Biometrics,* **45,** 395-413.
- Darroch, J.N., Fienberg, S.E., Glonek, G. & Junker, B. (1993). “A three sample multiple capture-recapture approach to the census population estimation with heterogeneous catchability.” *Journal of the American Statistical Association,* **88,** 1137-1148.
- Ebrahimi, N. (1997). “On the statistical analysis of number of errors remaining in a software design document after inspection.” *IEEE Transactions on Software Engineering,* **26,** 529-532.
- Fienberg, S.E. (1972). “The multiple recapture census for closed populations and incomplete 2^k contingency tables.” Biometrika, 59, 591-603.

References continued…

- IWDGMF (International Working Group for Disease Monitoring and Forecasting) (1995a). “Capture-recapture and multiple-record systems estimation, I: History and theoretical development. American Journal of Epidemiology, 142, 1047-1058.
- IWDGMF (International Working Group for Disease Monitoring and Forecasting) (1995b). “Capture-recapture and multiple record systems estimation, II: Applications in human diseases. American Journal of Epidemiology, 142, 1059-1068.
- Petersen, C.G.J. (1896). “The yearly immigration of young plaice into the Limfjord from the German Sea.” Rep. Danish Biol. Sta. 6, 1-48.
- Wohlen, C., Runeson, P. & Brantestam, B. (1995). “An experimental evaluation of capture-recapture in software inspection.” Software Testing, Verification and Reliability, 5, 213-232.
- Rivest, L.P. & Baillargeon, S. (2007). “Rcapture: Loglinear Models for Capture-Recapture in R”. *Journal of Statistical Software,* **19**(5).
- Rivest L.P. % Baillargeon, S. (2014). “Rcapture: Loglinear Models for Capture-Recapture Experiments. *R package version 1.4-2.* https://cran.r-project.org/web/packages/Rcapture/index.html
- Rivest, L.P. & Levesque, T. (2001). “Improved Log-linear Model Estimators of Abundance in Capture-Recapture Experiments.” *Canadian Journal of Statistics*, **29**, 555-572.

Contact information

Steven Ross, Director
Employment Security Department
Labor Market Information
Labor Market and Economic Analysis
(360) 507-9615
sross@esd.wa.gov

Gustavo Avilés, Manager
Employment Security Department
Program Evaluation, Research & Analysis
Labor Market and Economic Analysis
(360) 507-9552
gaviles@esd.wa.gov

Joshua Moll, Research Economist
Employment Security Department
Program Evaluation, Research & Analysis
Labor Market and Economic Analysis
(360) 507-9554
jmoll@esd.wa.gov

Employment Security Department
WASHINGTON STATE