Wage Imputation in the OES Survey: A Model-Assisted Approach

Jane G. Osburn

Office of Occupational Statistics and Employment Projections,
U.S. Bureau of Labor Statistics

1. Introduction

Wage imputation in the Bureau of Labor Statistics Occupational Employment Statistics (OES) Survey requires a process that matches non-respondent and respondent establishments on characteristics that best predict the wage levels of the non-respondent establishments in a given MSA /Industry / Establishment Size cell. The mean wage distribution of the establishments in the donor cell is then imputed to the non-respondent. Currently, the OES Survey procedures first define the donor cell for a given non-respondent Establishment /Occupation cell by the same semi-annual panel /MSA / four/five digit NAICS Industry / Establishment-size (2) as the non-respondent, and cells are then collapsed across industry and size groupings in the case of insufficient response. If insufficient response still exists after collapsing across industry and size groupings, the base level strata default from MSA to State, the donor cell is reformed by State / four/five digit NAICS Industry / Establishment Size (2), and the process repeats as for the MSA base-level strata.[1] The base-level strata eventually default to USA in a similar manner.

In those instances in which State or USA are the base-level strata of the donor cell used in an imputation, the current method in effect uses the average wage levels of the State or USA as a proxy for the wage level of the non-respondent establishment. For those states that contain areas with widely varying wage levels, some of the imputation outcomes could be improved.

A number of experimental wage imputation procedures are examined, the first of which replaces the base-level strata currently used in the OES Program with groupings of MSAs based on statistical estimates of area wage levels. A linear mixed model is used to estimate the wage level of each area in the U.S., and the estimates are used to guide the formation of groupings of MSAs with similar wage levels. Three successively more aggregated groupings are formed, including a thirteen-level grouping, 'MS' group one, a five-level grouping, 'MS' group two, and a three-level grouping, 'MS' group three. 'MS' groups one and two take the place of the State, USA base-level strata of the current method, and wage imputation otherwise proceeds exactly as in the OES method. 'MS' group three adds an additional level.[2]

Five other estimators based on a similar notion are also examined. Three of the estimators use a two-stage estimation approach in which the estimates of area wage levels are used as a predictor of the establishment wage level along with other predictors including the total employment of the establishment and an OES-data proxy for the establishment average wage. Both of these predictors are available from the Quarterly Census of Employment and Wage Program, which also provides the basis for the OES sample frame.[3] This approach is then revised for the case of the best-performing estimator, in favor of a more efficient approach that uses the proxy QCEW wage as a predictor in the linear mixed model.

Simulation is used to examine the relative performance of the estimators. The results suggest that the best estimator predicts the wage level of establishments using a linear mixed model that contains the random effect of areas along with the effects of industry, establishment size, occupation, and an OES-data proxy for the establishment average wage available from the QCEW Program.

---

[1] The OES strata include Ownership in NAICs 611, Educational Services, and NAICS 622, Hospitals.
[2] Note that MSA groups one, two, and three each contain all MSAs.
[3] BLS Statisticians Polly Phipps and Daniell Toth proposed the use of establishment wage information available from the QCEW Program as a predictor of establishment wage levels.

The following section discusses wage imputation in the OES Survey under the current and experimental procedures. Section three discusses the data produced by the Occupational Employment Statistics Survey. Section four outlines the estimation of area wage levels using the OES data in a linear mixed model, discusses six experimental estimators of the establishment wage level, and describes how the model estimates are used to form donor cells under each model. Section five describes a simulation study that examines the relative performance of the estimators. Section six examines the results. Section seven offers some conclusions and directions for future research.

2. Overview of the Current and Proposed Methods of Wage Imputation

Current OES procedures first identify respondent establishments for a given semi-annual panel and occupation from the same MSA/ four/five digit NAICS Industry /Size(2) cell as the non-respondent establishment. If the cell contains a sufficient number of respondents for the occupation, the mean wage distribution of the cell is imputed to the non-respondent Establishment /Occupation cell. In the absence of sufficient response, the cells are combined into successively more aggregated industry and size groupings, and the mean wage distribution of the aggregated grouping /Occupation cell is used. If cells defined by MSA /All Industries /Occupation still contain insufficient response, the base-level strata defaults from MSA to State, the cell is reformed by State /four/five digit NAICS Industry / Size (2) /Occupation, and the procedure repeats as for the MSA base-level strata. If cells defined by State /All Industries /Occupation still contain insufficient response, the base-level strata defaults to USA, cells are reformed by USA /four/five digit NAICS Industry /Size (2) /Occupation, and the procedure repeats as for the State base-level strata.

Each of the experimental estimators replace the base-level strata currently used in the OES Program with three successively more aggregated establishment groupings that are based on a statistical model. 'MS' group one is a thirteen-level grouping, 'MS' group two is a five-level grouping, and 'MS' group three is a three-level grouping. Wage imputation follows the same rules used in the current OES procedures, with the exception that 'MS' groups one and two take the place of the base-level strata State, USA in the current OES method, and 'MS' group three adds an additional level to the base-level strata.

3. Data from the Occupational Employment Statistics Survey

The Bureau of Labor Statistics Occupational Employment Statistics Survey collects data over a three year cycle on the wages and occupational category of each employee in a total of approximately 1.2 million establishments, spanning the non-farm private and public sectors in the U.S. For each occupation employed in the establishment, survey respondents record the number of employees earning wage rates that fall within each of twelve wage intervals. The average wage of an Establishment /Occupation cell is the sum over wage intervals of the product of the employment in the interval and the 'midpoint', or estimated mean wage of the interval, divided by the sum of employment across the wage intervals. The wage interval midpoints that are applied to establishments surveyed during different quarters of the three year period are updated using the Employment Cost Index (ECI), such that all establishments' wages reflect the final year /quarter wage levels in the completed dataset.[4]

4**. Six Models of the Establishment Wage Differential**

4. 1 A1 Estimator

The A1 estimator uses an estimate of the area wage differential as a proxy measure of the wage differential of each establishment in the area. A nested error linear mixed model containing the effects of industry, occupation, and establishment size, and the random effects of area is used to estimate the Empirical Best Linear Unbiased (EBLUP) estimator of the area wage effect;

[4] The midpoints of the wage intervals are estimated using data from the Bureau of Labor Statistics National Compensation Survey.

$$\ln w_{ijkl} = Industry_j + Occupation_k + Establishment\ size_l + Area_i + e_{ijkl}$$
$$area_i \sim N(0,\sigma_\upsilon^2)$$
$$e_{ijkl} \sim N(0,\sigma_e^2)$$

where i indexes areas, j indexes industries, k indexes occupations, and l indexes establishments
Industry is a three digit NAICS industry classification
Occupation is a six digit Standard Occupational Classification
Establishment size is a six category establishment size grouping

$\ln w_{ijkl}$ is the log of the mean wage by Establishment /Occupation cell

The area effects and residual errors are assumed to be iid random variables, independent of each other.

The model is estimated using Pseudo-Maximum Likelihood, in which each Establishment /Occupation observation is weighted by a measure of sample-weighted occupational employment that is scaled to account for the dependence of the wage distribution on areas.[i] The parameters are estimated using restricted maximum likelihood.[5] The EBLUP estimator corresponds to the set of parameter values that minimizes the mean squared error of prediction (MSPE) of $\ln w$[6]. The EBLUP of the area wage effect is the weighted average residual for the area, discounted by the estimated share $\hat{\gamma}_i$ of between-area variance in the overall variance;

Let

wtemp = scaled, sample-weighted occupational employment
eo = Establishment /Occupation cell

$$\text{EBLUParea}_i = \hat{\gamma}_i \left[\frac{\sum_{eo \in Area\ i} wtemp_{ijkl} * \left(\ln \overline{w}_{ijkl} - \overline{x}'_{ijkl}\hat{\beta} \right)}{\sum_{eo \in Area\ i} wtemp_{ijkl}} \right] \quad \text{where} \quad \hat{\gamma}_i = \frac{\hat{\sigma}_\upsilon^2}{\hat{\sigma}_e^2 + \hat{\sigma}_\upsilon^2}\ .^7$$

The standard errors are estimated using the delta method as described in Goldstein (Sec.3.24).

The A1 estimator uses the estimates of the area wage effect (EBLUParea) to assign all establishments in each MSA to 'MS' groups one, two, and three, where 'MS' group one has 13 levels, each of which contains MSAs with very similar average wage levels, 'MS' group two has five levels, each of which contains MSAs within a broader range of wage levels, and 'MS' group three has three levels, each of which contains MSAs with wage levels within a broad range.

4.2. Four Experimental Estimators of the Establishment Wage Level

A variety of additional estimators are based on a two-stage approach that uses a measure of the establishment wage differential as the dependent variable and explores the use of predictors including the estimated area wage effect

[5] Unlike maximum likelihood, the REML method of estimating the variance components accounts for the loss of degrees of freedom that results from the estimation of β. See discussion in Rao (2003), pp.100-102.
[6] Royal (1976) derived the best linear unbiased predictor for any linear function.
[7] The estimate $\hat{\sigma}_{ei}^2$ in the denominator of $\hat{\gamma}_i$ is the approximate variance of the mean residual for area i. Thus defined, it is a cluster-level variance, on the same scale as the estimated model variance $\hat{\sigma}_\upsilon^2$.

from the model of Section 4.1 (EBLUParea), establishment total employment, and an OES-data proxy for the establishment average wage that is available from the QCEW program.

The dependent variable used in each model is a measure of the establishment wage differential (subsequently Estabdiff_v2) formed from residuals obtained from the model of Section 4.1 by differencing the fixed portion of the wage estimate from the log wage. The R1A estimator attempts to predict estabdiff_v2 using both the EBLUParea and the proxy QCEW wage in conjunction with a measure of establishment employment. The R1B estimator incorporates the EBLUParea as a predictor along with a measure of establishment employment. The R1C model uses the proxy QCEW establishment wage and establishment employment as predictors;

$$R1A;\ \ Estabdiff_{v2} = Intercept + \beta_1 * qcewdiff_{OES\_PROXY} + \beta_2 * EBLUP_{area} + \beta_3 * emp$$
$$R1B;.\ Estabdiff_{v2} = Intercept + \beta_1 * EBLUP_{area} + \beta_2 * emp$$
$$R1C:\ Estabdiff_{v2} = Intercept + \beta_1 * qcewdiff_{OES\_PROXY} + \beta_2 * emp$$

Each model was estimated by NAICS industry /establishment size using a general linear model.[8] For each model, the model-predicted establishment wage differential was used to assign each establishment to groupings similar to those described above for the A1 estimator, resulting in three complete sets of establishment groupings that form the base level strata under each model.

A fifth estimator, R1F, uses the OES-data proxy QCEW wage differential directly to assign establishments to the groupings.

4.3 R1A_r1 Model

The two-stage approach outlined above is quite inefficient. A revised version of the R1A model, R1A_r1, augments the model of Section 4.1 to include the proxy QCEW wage differential as a predictor;

$$\ln\_wg\_dif_{ijkl} = industry_j + occ_k + area_i + estab\,size_l + qcew\_dif_l + e_{ijkl}$$

where

i indexes areas
j indexes industries
k indexes occupations
l indexes establishments
ln_wg_dif = the deviation of the log establishment /occupation wage from the grand mean of the log establishment /occupation wage
qcew_dif = the deviation of the log establishment QCEW wage from the grand mean of the log QCEW establishment wage.

The explanatory power of the proxy QCEW wage predictor is most clearly demonstrated by examining its effect in a two-fold nested error model (Stukel and Rao, 1999) that accounts for the dependence of wages on both areas and establishments;

$$\ln\_wg\_dif_{ijkl} = industry_j + occ_k + area_i + estab_l(area_i) + qcew\_dif_l + e_{ijkl}$$

In this model, the effects of area and of establishment conditional on area are each assumed to be iid random variables, independent of each other;

---

[8] The models were estimated at the four, three, and two digit NAICS levels and analyses for detailed groupings were replaced with analyses using more aggregated groupings as needed. For the purposes of the simulation discussed in Section 6, the models are estimated using data that excludes data from the State for which the simulations are conducted. In actual practice, the models should be estimated using all of the data.

$Area_i \sim N(0, \sigma_v^2)$
$Estab_{l|i} \sim N(0, \sigma_u^2)$
$e_{ijkl} \sim N(0, \sigma_e^2)$

The EBLUP of the conditional establishment wage differential in the two-fold model is the average percentage by which the wages of occupations in the establishment differ from that predicted by the mixed model, discounted by the share of between-establishment variance, (γ_{li}), in the total (between- plus within-establishment) wage variance;

$$\text{EBLUP}_{\text{estab l|i}} = \bar{u}_{li} = \gamma_{li}\left[\left(\ln \bar{w}_{li} - \bar{x}'_{li}\tilde{\beta}\right) - \gamma_i\left(\ln \bar{w}_{i\gamma} - \bar{x}'_{i\gamma}\tilde{\beta}\right)\right]$$

or

$$\text{EBLUP}_{\text{estab l|i}} = \bar{u}_{li} = \gamma_{li}\left(\ln \bar{w}_{li} - \bar{x}'_{li}\tilde{\beta}\right) - \gamma_{li}\gamma_i\left(\ln \bar{w}_{i\gamma} - \bar{x}'_{i\gamma}\tilde{\beta}\right)$$

where

$$\bar{w}_{i\gamma} = \sum_{l\in i}\gamma_{li}\bar{w}_{li} \Big/ \sum_{l\in i}(\gamma_{li})$$, similarly for $\bar{x}_{i\gamma}$

$$\hat{\gamma}_{li} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_e^2 + \hat{\sigma}_u^2}$$ [9]

$$\hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2 \Big/ \sum_{l\in i}\gamma_{li}}$$

The EBLUP of the area wage effect is similar to that discussed earlier, with the exception that the shrinkage parameter $\hat{\gamma}_i$ takes into account the lower level random effect.

The two-fold model was estimated twice, including and excluding the OES-data proxy QCEW wage differential.[10] The inclusion of the proxy QCEW wage reduced the size of the area and establishment variance components by over ninety percent and eighty percent, respectively.

The EBLUP of the establishment wage differential obtained from this model is of limited direct use in the current study, because it represents the portion of the (unconditional) establishment wage differential for which we have no model. That is, it is the portion of the unconditional establishment wage effect that is not explained by the area wage effect and the proxy-QCEW variable together. The establishment wage effect estimated in this model would become useful, however, if we were to find an additional predictor that helps explain it.

The revised estimator, model R1A_r1, uses the estimated parameters of the model to predict the unconditional establishment wage differential;

R1A_r1 = EBLUP of area wage effect + ($\beta_1 \times$ QCEW_wage_differential)

The model-predicted establishment wage differentials (R1A_r1) are used in a way similar to the procedures discussed earlier. All establishments, non-respondents and respondents, are assigned to groupings on the basis of the

[9] Discussion here assumes equal residual error variances across areas and abstracts away from survey weights .
[10] A nested error model containing a single random effect of area was estimated first using a large dataset. The 2-fold model was then estimated using a sample of residuals from this estimation, holding the area variance component constant at the level of the first estimation.

model-predicted establishment wage differential, and these are used to form the successively more aggregated, three-tier base-level strata in the same way as discussed earlier.

Wage imputation then proceeds as in the previously discussed method. Non-respondent Establishment /Occupation cells from a given panel are wage-imputed using the mean wage distribution of the panel respondents in the same model-defined group /Industry /Size /Occupation cell or appropriately collapsed cell.

5. The Simulation

Simulation is used to test the various model alternatives against the OES method. The main elements of the simulation include the establishment groupings ('MS' groups one, two, and three) under each model, the simulated sample, the simulated non-respondents, and the simulated respondents.

The simulation uses data from the May 2009 OES survey round for all size four establishments located in Louisiana, and all size seven establishments located in New York. The simulated sample for a given State is comprised of all of the respondent establishments from the State /Size cell. Stratified random samples from the simulated sample are used to identify a set of simulated non-respondent establishments that makeup a similar proportion of establishments and a similar Industry /Size composition of establishments to the set of actual non-respondent establishments in the OES sample. The simulated respondents include all units in the simulated OES sample not identified as simulated non-respondents.

The simulation for each model is guided by the 'Imputation dataset', which contains the simulated non-respondent establishment data stripped of all wage distributions but retaining identifiers including a non-respondent indicator, industry and establishment size identifiers, establishment total employment, the occupational employment of each occupation employed in the establishment, and the 'MS' group one, two, and three assignments under each of the models A1, R1A, R1A_r1, R1B, R1C, and R1F.

For each model, the identifiers are used to match recipients and donor cells such that the units included in the donor cell are drawn from the same semi-annual panel, Occupation, Industry, and 'MS' group one, two, or three as the non-respondent Establishment /Occupation cell. The mean wage distribution of the donor cell is then used to apportion the recipient cell occupational employment among the OES wage intervals.[11]

For each simulation and for each of the six experimental models and the OES method, the mean imputed wage by Establishment /Occupation is calculated using the wage interval means in conjunction with the apportioned occupational employment.

Performance measures including 1) average relative error $\overline{(ARE)}$, 2) average relative bias $\overline{(ARB)}$, and 3) variance are calculated by averaging over the simulations for a cell;

Let

$\overline{w}_{sim}$ = mean imputed wage by Sim /MSA /Establishment /Occupation

$\overline{w}_{actual}$ = mean "true" wage by MSA /Establishment /Occupation

nocc = number MSA /Establishment /Occupation cells in summary cell

t = un-weighted occupational employment by MSA /Establishment /Occupation

$$\overline{A}\overline{R}\overline{E} = \left(\frac{1}{\sum_{1}^{nocc}(t)} \right) \sum_{1}^{nocc} \left(\frac{1}{100} \right) \sum_{sim=1}^{100} \left(t * \left| \frac{\overline{w}_{sim}}{\overline{w}_{actual}} - 1 \right|_{OCC} \right)$$

[11] In the actual OES procedures, non-respondent establishments are first employment-imputed and then wage imputed. The simulation by-passes the employment-imputation step in favor of using the actual occupational employment totals of the simulated non-respondents that are available by virtue of the simulation design.

$$\overline{A}\overline{R}\overline{B} = \left(\frac{1}{\sum_{1}^{nocc}(t)} \right) \sum_{1}^{nocc} \left(\frac{1}{100} \right) \sum_{sim=1}^{100} \left[t * \left(\frac{\overline{w}_{sim}}{\overline{w}_{actual}} - 1 \right) \right]$$

Let

$w_{\mathrm{mod},sim}$ = unweighted mean wage of a FIPS /MSA /Occupation cell for a given simulation, calculated using dataset S, the union of the imputed non-respondent and respondent Estab /OCC cells in the simulated sample

$\overline{w}_{\mathrm{mod}}$ = unweighted mean of $w_{\mathrm{mod},sim}$, calculated over all simulations

$$= \frac{1}{s} \sum_{sim=1}^{S} \left(w_{\mathrm{mod},sim} \right)$$

The simulation variance is

$$SV = \frac{1}{(S-1)} \sum_{sim=1}^{S} \left(w_{\mathrm{mod},sim} - \overline{w}_{\mathrm{mod}} \right)^2$$

6. Results

Table 1 contains the results by State /OCC2 (two digit Standard Occupational Classification code) and by State /MSA for the relative error and bias measures.

In New York overall, the weighted average relative error for the A1, R1B, and R1C estimators is about .25, and the similar measures for the R1A, R1A_r1, R1F and OES estimators are .23, .22, .24 and .28, respectively. The R1A_r1 estimator has lower error than the OES estimator in 20 of the 22 occupational groups, while the A1, R1A, r1B, r1C, and r1F estimators have lower relative error in 13, 18, 15, 15, and 17 occupational groups, respectively.

The R1A_r1 estimator is biased downward about one percent in New York, while the R1B, R1C, R1F, and OES estimators over-predict wages to some extent. The R1C and OES estimators are biased upward about three percent, the R1B estimator is biased upward about four percent, and the R1F estimator is biased upward about two percent. The A1 and R1A estimators are biased downward about five percent and six percent, respectively. The R1B and R1F estimators are less biased than the OES estimator in 11 of the 22 occupational groups, while the A1, R1A, R1A_r1, and R1C estimators are less biased than the OES estimator in 17, 13, 14, and 12 occupational groups, respectively.

In Louisiana overall, the weighted average relative error is .22 for the A1 estimator, .18 for the R1A estimator, .17 for the R1A_r1 estimator, .25 for the R1B and OES estimators, and .20 for the R1C and R1F estimators. The R1A estimator has lower error than the OES estimator in the case of 17 of the 22 occupational groups, the R1A_r1 estimator has lower error than the OES estimator in the case of 21 of the 22 occupational groups, and the A1, R1B, R1C, and R1F estimators have lower relative error than the OES estimator in 14, 12, 18, and 20 occupational groups, respectively.

The R1A and A1 estimators are biased upward about three percent in Louisiana overall, the R1A_r1 estimator is biased upward about four percent, the R1B, R1C, and R1F estimators are all biased upward about six percent, and the OES estimator is biased upward about ten percent. The R1A and R1A_r1 estimators have lower bias than the OES estimator in the case of 19 of the 22 occupational groups, and the A1, R1B, R1C, and R1F estimators have lower bias than the OES estimator in 20, 17, 14, and 18 of the 22 occupational groups, respectively.

One source of upward bias in the OES estimator is the use of the ‘State’ base level imputation strata. In States that contain one or more large, high-wage areas in addition to numerous smaller, lower wage areas, the mean wages of the ‘State’ donor cell are dominated by the large, high-wage area(s) in the cell, causing upward bias in the imputed wage for most areas. One would expect this effect to be stronger for New York, due to a somewhat larger median wage gap between NYC and other areas (about 25% versus 19% between New Orleans/Baton Rouge and other areas in LA) and its larger share of State employment.[12]

The results do not confirm this expectation. Looking at two digit SOC occupational groups, the results show that, on average, the OES estimator over-predicts the wages of 16 of the 22 occupational groups in the New York data, while it over-predicts the wages of 19 of the 22 groups in the OES data for Louisiana. All of the experimental estimators exhibit a more balanced pattern of over- and under-prediction for both states. In New York, the A1 and R1F estimators each overpredict the wages of 7 groups, while the R1A, R1A_r1, R1B, and R1C estimators over-predict the wages of 6, 12, 11, and 8 groups, respectively. In Louisiana, the R1A estimator over-predicts the wages of 14 groups, and the R1B, R1C, A1, R1A_r1, and R1F estimators over-predict the wages of 18, 17, 15, 19, and 18 groups, respectively.

In both New York and Louisiana, all estimators are least biased for most of the largest occupations and those earning close to the minimum wage, including Office and Administrative Support (43), Food Service and Preparation (35), and Personal Care and Service (39) occupations. In Louisiana, all estimators are generally more biased for the professional occupations that contain much lower employment in Louisiana, including Managers (11), Business and Financial Operations (13), Computer and Mathematical Science (15), Architecture and Engineering (17), and Life, Physical, and Social Science (19), Community and Social Service (21), Legal (23), and Education, Training, and Library (25) occupations. The OES estimator is severely upward biased (.50+) for both Business and Financial Operations (13) and Sales (41) occupations. In New York, the estimators generally had low amounts of bias for all occupations with the exception that all estimators over-predict the Healthcare Practitioner and Technical (29) Occupations, most models under-predict Arts, Design, Entertainment, Sports and Media (27) occupations, and most models seriously over-predict the wages of Building and Grounds workers (37). In addition, all of the experimental estimators under-predict the wages of Healthcare Support occupations (31) in both the Louisiana and New York data.

The results also seem to reflect the sensitivity of wages of many local-level public sector occupations to a combination of fiscal and political pressures. All estimators over-predict the wages of Education, Training, and Library occupations and Social and Community Service occupations in Louisiana, whose public school systems rank in the fourth quartile on spending per student and whose fiscal commitment to anti-poverty measures is similar.[13]

Turning to the results by area, in New York the OES estimator is clearly upward biased in most of the lower wage areas including all of the Balance of State areas, Glen Falls, Utica, Syracuse, Buffalo, and Binghamton, while it is downward biased in NYC. In Louisiana, the pattern of biases of the OES estimator is similar, and the A1, R1A, and R1A_r1 estimators are all significantly less biased than the OES estimator in all areas.

Table 2 contains, for each estimator, the value (S) of the simulation variance at which MSA /Occupation cells that together account for a given percentage of total sample employment have values of simulation variance less than or equal to S. For example, in the New York simulation of the R1A model, MSA /Occupation cells that together account for ninety eight percent of New York, size-seven-establishment sample employment have simulation variance less than or equal to 5.65. The R1A_r1, R1A, R1C, and R1F estimators all have lower variance than the OES estimator at all points in the distribution except for the last .5%, and the R1B and A1 estimators have lower variance than the OES estimator at all but three points.

7. Conclusions and Directions for Future Study

Overall, the results suggest that the R1A_r1 estimator outperforms the other estimators, showing that an estimator that augments information about area wage levels with auxiliary information about the average wage of the establishment can outperform an estimator that utilizes either one of these alone. The analyses are currently being re-run using the actual QCEW wage in place of the OES data proxy.

[12] BLS statisticians Polly Phipps and Daniel Toth have identified a second source of wage bias resulting from non-response patterns of large establishments.
[13] Auxiliary analyses show that the average magnitude of the wage bias for the detailed occupations in each group, for each model, mirrors the results for the more aggregated groups.

References

Asparouhov, Tihomir 2005. "Sampling Weights in Latent Variable Modeling," *Structural Equation Modeling*, Vol. 12., No.3., pp.411-434.

Eltinge, John L. and Jeffrey M. Gonzalez 2009. "Sensitivity of Inference Under Imputation," Federal Committee on Statistical Methodology Research Conference, Washington, DC.

Firebaugh, Glenn 1978. "A Rule for Inferring Individual-Level Relationships from Aggregate Data," *American Sociological Review,* Vol.43., No.4., pp.557-572.

Goldstein, Harvey 2011. *Multilevel Statistical Models*, 4th Edition, John Wiley and Sons, Ltd., Chichester, West Sussex, UK.

Greene, William H. 2000. *Econometric Analysis*, Fourth Edition. Prentice Hall Press, New Jersey.

Grilli, Leonardo and Monica Pratesi. 2002. "Weighted Estimation in Multilevel Models to Allow for Informativeness of the Sampling Design," *Working Paper 2002 /01,* Dipartimento di Statistica, Università degli Studi di Firenze.

Hox, Joop. 2002. *Multilevel Analysis: Techniques and Applications,* Lawrence Erlbaum Associates, Mahwah, NJ.

Pfeffermann, D., C.J. Skinner, D.J. Holmes, H. Goldstein, J. Rasbach "Weighting for Unequal Selection Probability in Multilevel Models," *Journal of the Royal Statistical Society*, B, 1998 Vol.60., Part 1. pp.23-40.

Prasad, N.G.N and J.N.K.Rao. 1990. "The Estimation of Mean Square Error of Small Area Estimators," *Journal of the American Statistical Association*, Vol.85., pp.163-171.

Rao, J.N.K. 2003. *Small Area Estimation,* John Wiley and Sons, New Jersey

Royal, R.M. 1976. "The Linear Least-Squares Prediction Approach to Two-Stage Sampling," *Journal of the American Statistical Association*, Vol.71.,pp.657-664.

Shoemaker, Owen and William Johnson. 1999. "Estimation of Variance Components for the U.S. Consumer Price Index," *Proceedings of the Section on Business and Economic Statistics, American Statistical Association.*

Snijders, Tom and Roel Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling,* Sage Publications, Thousand Oaks, CA.

Stukel, D.M. and Rao, J.N.K. 1999. "Small Area Estimation Under Two-fold Nested Error Regression Models," *Journal of Statistical Planning and Inference*, Vol.78., pp. 131-147.

Endnotes

[i] See Pfeffermann, et.al (1998). Weighting is not needed in a linear mixed model that includes as covariates each of the strata components of a self-weighting survey, in which the probability of selection is the same for all units within a stratum. The parameter estimates of the linear mixed model are biased if units at any level (i.e. individual establishment/occupation observations at level 1, MSAs at level 2) are selected with unequal probabilities in ways that are not accounted for by the linear mixed model. While the OES Survey is self-weighting within MSA/Industry/Size strata, the model used here includes these strata components as covariates only at somewhat aggregated levels. For example, a three digit Naics industry is included rather than the four or five digit Naics industry used in sampling.

A correction offered by Pfeffermann, et.al (1998) involves scaling the level 1 sample weights such that the naïve estimate of MSA employment, $\hat{N}_i$, equals actual total reported employment for the area. With this scaling, the variance of the cluster-level latent variable is determined by the true sample size of level 1 units (see Asparouhov (2005). Goldstein (2011) also provides a discussion of the approach. Grilli and Pratesi (2002) conducted extensive simulations demonstrating that the approach outlined above works well, minimizing both bias and the increase in variance that results from the use of survey weights.

Table 1

OES Wage Imputation Simulation Study: Results of Six Experimental Models and the OES Estimator

1. New York, size seven establishments

OCC Title	Number Estab/ OCC Cells	Average Relative Error							Weighted Average Relative Error							Average Relative Bias							Weighted Average Relative Bias						
		(Simple average over sims of MEAN(ABS (WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL)							Simple average over sims of Σ(ABS (WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / Σ(EMP)							Simple average over sims of Σ((WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / Σ(EMP)							Simple average over sims of Σ((WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / Σ(EMP)						
		A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES
ALL	117735	0.27	0.24	0.23	0.26	0.25	0.25	0.30	0.25	0.23	0.22	0.25	0.25	0.24	0.28	-0.02	-0.04	0.00	0.05	0.03	0.03	0.07	-0.05	-0.06	-0.01	0.04	0.03	0.02	0.03
Managers	5030	0.25	0.19	0.19	0.19	0.19	0.22	0.23	0.24	0.19	0.19	0.18	0.19	0.21	0.22	-0.07	-0.02	0.02	0.00	-0.03	-0.05	0.10	-0.08	-0.03	0.01	0.00	-0.04	-0.07	0.08
Bus.&Fin. Operations	5952	0.31	0.28	0.23	0.29	0.29	0.22	0.42	0.31	0.27	0.22	0.28	0.29	0.22	0.42	0.11	0.03	0.03	0.15	0.04	0.06	0.28	0.13	0.05	0.05	0.17	0.08	0.07	0.30
Computer & Math	6073	0.27	0.26	0.21	0.25	0.27	0.22	0.25	0.25	0.24	0.20	0.24	0.25	0.21	0.23	0.00	-0.03	0.05	0.11	0.00	0.01	0.08	-0.02	-0.05	0.04	0.11	-0.01	0.00	0.06
Architecture and Eng.	6452	0.35	0.27	0.26	0.28	0.27	0.23	0.28	0.30	0.24	0.28	0.26	0.28	0.23	0.29	0.19	0.04	0.08	0.13	0.05	0.03	0.04	0.07	-0.03	-0.04	0.02	-0.06	-0.06	-0.07
Life, Physical, and Social Science	6893	0.27	0.21	0.20	0.24	0.27	0.37	0.25	0.25	0.19	0.19	0.22	0.27	0.35	0.22	-0.01	0.01	0.03	0.09	0.08	0.20	0.07	-0.03	0.02	0.02	0.10	0.10	0.19	0.04
Community and Social Service	3096	0.18	0.14	0.15	0.15	0.16	0.15	0.20	0.17	0.14	0.15	0.15	0.16	0.14	0.19	-0.06	-0.03	-0.01	-0.02	-0.03	-0.05	0.06	-0.06	-0.02	-0.01	-0.01	-0.04	-0.06	0.06
Legal	1227	0.23	0.17	0.21	0.44	0.16	0.15	0.18	0.20	0.15	0.16	0.45	0.10	0.10	0.18	0.03	0.10	0.14	0.17	0.09	0.06	0.07	0.03	0.10	0.13	0.19	0.07	0.05	0.08
Education, Training, and Library	27476	0.28	0.25	0.26	0.29	0.27	0.26	0.33	0.27	0.25	0.24	0.28	0.26	0.24	0.30	-0.03	-0.07	-0.02	0.07	0.06	0.05	0.07	-0.08	-0.09	-0.04	0.05	0.05	0.03	0.02
Art, Design, Entertain., Sprts, Media	5170	0.38	0.36	0.32	0.44	0.33	0.36	0.39	0.42	0.38	0.33	0.48	0.33	0.38	0.41	-0.13	-0.19	-0.02	-0.06	-0.15	-0.16	-0.15	-0.11	-0.16	-0.03	0.03	-0.17	-0.13	-0.15
Healthcare Prac. and Tech.	10551	0.41	0.42	0.33	0.28	0.44	0.47	0.41	0.42	0.42	0.34	0.28	0.45	0.48	0.42	0.28	0.28	0.20	0.13	0.28	0.33	0.29	0.29	0.29	0.21	0.14	0.30	0.35	0.30
Healthcare Support	3208	0.13	0.18	0.16	0.14	0.19	0.17	0.17	0.13	0.19	0.16	0.13	0.19	0.18	0.16	-0.04	-0.07	-0.03	-0.02	-0.07	-0.07	0.02	-0.06	-0.08	-0.03	-0.03	-0.08	-0.08	0.01
Protective Service	3342	0.16	0.16	0.18	0.18	0.18	0.17	0.29	0.15	0.14	0.17	0.17	0.16	0.15	0.29	0.06	0.00	0.07	0.03	0.02	-0.04	0.15	0.05	-0.02	0.06	0.02	0.01	-0.04	0.15
Food Preparation and Serving	1422	0.26	0.22	0.18	0.14	0.22	0.20	0.24	0.26	0.21	0.18	0.13	0.21	0.19	0.23	-0.15	-0.10	0.04	-0.04	-0.13	-0.09	0.06	-0.13	-0.08	0.04	-0.03	-0.12	-0.08	0.08
Building and Grounds, Cleaning and M	567	0.55	0.36	0.32	0.50	0.35	0.38	0.21	0.55	0.36	0.32	0.50	0.35	0.39	0.21	0.37	0.19	0.22	0.30	0.16	0.02	0.01	0.37	0.19	0.23	0.30	0.16	0.02	0.01
Personal Care and Service	2376	0.10	0.13	0.15	0.18	0.09	0.12	0.12	0.10	0.13	0.14	0.17	0.08	0.11	0.12	-0.03	-0.09	0.09	-0.12	-0.01	-0.06	0.06	-0.03	-0.09	0.09	-0.12	-0.01	-0.06	0.05
Sales and Related	1500	0.45	0.33	0.23	0.33	0.28	0.25	0.65	0.24	0.28	0.23	0.30	0.21	0.20	0.36	0.22	-0.16	-0.11	-0.07	-0.18	-0.19	0.30	0.05	-0.25	-0.18	-0.21	-0.17	-0.15	-0.11
Office and Administrative Spt.	5874	0.20	0.16	0.15	0.15	0.17	0.21	0.20	0.19	0.16	0.14	0.15	0.16	0.20	0.20	0.00	0.02	-0.02	0.03	0.00	0.05	0.05	0.00	0.03	-0.02	0.03	0.00	0.06	0.05
Farming, Fishing, and Forestry	218	0.24	0.21	0.20	0.29	0.24	0.27	0.34	0.22	0.20	0.19	0.29	0.22	0.27	0.33	-0.10	-0.05	0.00	-0.07	-0.21	-0.18	-0.22	-0.08	-0.04	-0.01	-0.06	-0.19	-0.17	-0.21
Construction and Extraction	4160	0.25	0.19	0.15	0.15	0.22	0.20	0.33	0.25	0.19	0.15	0.15	0.22	0.20	0.33	-0.04	-0.03	0.02	-0.02	-0.13	-0.12	0.18	-0.04	-0.02	0.02	-0.02	-0.13	-0.11	0.18
Installation, Maint., and Repair	4253	0.19	0.14	0.11	0.19	0.17	0.13	0.15	0.19	0.14	0.10	0.18	0.17	0.14	0.14	-0.04	-0.04	-0.03	-0.08	-0.08	-0.05	-0.05	-0.04	-0.03	-0.02	-0.08	-0.08	-0.06	-0.05
Production	9815	0.18	0.18	0.16	0.15	0.20	0.19	0.24	0.17	0.18	0.15	0.15	0.20	0.18	0.24	-0.08	-0.02	0.04	-0.02	-0.03	-0.06	0.01	-0.07	-0.02	0.03	-0.02	-0.03	-0.07	0.01
Transportation and Mat. Moving	3080	0.22	0.22	0.26	0.28	0.28	0.29	0.27	0.22	0.22	0.25	0.28	0.27	0.29	0.26	-0.05	-0.08	-0.10	-0.15	-0.16	-0.19	-0.09	-0.03	-0.09	-0.10	-0.13	-0.13	-0.16	-0.06

1. New York, size seven establishments

Area Name	Number Estab/ OCC Cells	Average Relative Error							Weighted Average Relative Error							Average Relative Bias							Weighted Average Relative Bias						
		(Simple average over sims of MEAN(ABS (WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL)							Simple average over sims of Σ(ABS (WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / Σ(EMP)							Simple average over sims of Σ((WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / Σ(EMP)							Simple average over sims of Σ((WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / Σ(EMP)						
		A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES
	6568	0.16	0.14	0.15	0.14	0.17	0.14	0.23	0.16	0.138	0.14	0.14	0.16	0.14	0.235	0.02	-0.03	0.02	0.06	-0.02	-0.02	0.11	0.02	-0.02	0.03	0.06	-0.02	-0.02	0.11
Balance of	1547	0.20	0.25	0.20	0.20	0.25	0.23	0.35	0.2	0.251	0.2	0.2	0.25	0.23	0.355	0.12	0.08	0.05	0.13	0.04	0.04	0.35	0.12	0.08	0.05	0.13	0.04	0.04	0.35
State	5839	0.28	0.12	0.15	0.17	0.16	0.18	0.23	0.27	0.117	0.14	0.17	0.15	0.18	0.228	0.19	0.02	0.06	0.03	0.06	0.09	0.12	0.19	0.02	0.05	0.03	0.06	0.08	0.12
Areas	9604	0.23	0.22	0.20	0.19	0.22	0.20	0.39	0.22	0.215	0.18	0.2	0.21	0.18	0.389	0.04	0.13	0.14	0.07	0.14	0.07	0.28	0.06	0.12	0.12	0.06	0.13	0.04	0.26
	5920	0.29	0.25	0.20	0.28	0.23	0.25	0.37	0.22	0.224	0.19	0.28	0.19	0.23	0.27	0.05	-0.12	-0.07	-0.06	-0.13	-0.20	0.06	0.00	-0.15	-0.09	-0.10	-0.13	-0.19	-0.07
Albany-Schenectady-Troy, NY	7056	0.23	0.17	0.20	0.19	0.18	0.18	0.25	0.2	0.159	0.18	0.18	0.16	0.15	0.25	-0.03	0.01	0.06	-0.02	-0.02	0.03	0.06	-0.04	-0.01	0.04	-0.02	-0.04	0.02	0.06
Binghamton, NY	434	0.10	0.14	0.16	0.12	0.22	0.17	0.29	0.1	0.141	0.16	0.12	0.22	0.17	0.293	-0.02	0.02	0.08	0.06	0.11	0.08	0.27	-0.02	0.02	0.08	0.06	0.11	0.08	0.27
Buffalo-Niagara Falls, NY	8500	0.31	0.18	0.19	0.31	0.23	0.24	0.39	0.29	0.182	0.18	0.3	0.23	0.25	0.368	0.03	0.04	0.06	0.03	0.08	0.13	0.21	-0.02	0.04	0.04	0.03	0.08	0.13	0.16
Elmira, NY	637	0.26	0.25	0.16	0.15	0.29	0.24	0.26	0.26	0.254	0.16	0.15	0.29	0.24	0.258	-0.10	-0.21	-0.15	-0.10	-0.27	-0.20	-0.24	-0.10	-0.21	-0.15	-0.10	-0.27	-0.20	-0.24
Glens Falls, NY	5315	0.08	0.16	0.14	0.12	0.18	0.18	0.27	0.08	0.157	0.14	0.12	0.18	0.18	0.268	0.02	-0.09	0.09	0.00	-0.06	-0.13	0.17	0.02	-0.09	0.09	0.00	-0.06	-0.13	0.17
Ithaca, NY	12588	0.23	0.24	0.25	0.30	0.34	0.30	0.27	0.23	0.242	0.25	0.3	0.34	0.3	0.27	-0.22	-0.22	-0.20	0.18	0.10	0.04	0.09	-0.22	-0.22	-0.20	0.18	0.10	0.04	0.08
Kingston, NY	1028	0.20	0.20	0.15	0.22	0.17	0.16	0.11	0.2	0.205	0.15	0.22	0.17	0.16	0.105	-0.18	-0.11	-0.05	-0.21	-0.06	-0.04	-0.04	-0.18	-0.11	-0.05	-0.21	-0.06	-0.04	-0.04
Nassau-Suffolk, NY Metro Div	9756	0.31	0.23	0.24	0.22	0.20	0.19	0.30	0.26	0.2	0.23	0.21	0.18	0.17	0.246	0.21	0.09	0.11	0.07	0.02	0.01	0.06	0.16	0.08	0.12	0.05	0.02	0.01	-0.01
New York-Wayne-White Plains, NY-N	13120	0.36	0.30	0.30	0.33	0.31	0.31	0.32	0.35	0.287	0.29	0.32	0.3	0.3	0.323	0.05	0.05	0.06	-0.03	0.00	0.03	-0.06	-0.02	0.01	0.00	-0.07	-0.05	-0.03	-0.12
Poughkeepsie-Newburgh-Middletowr	2271	0.17	0.11	0.17	0.14	0.11	0.13	0.15	0.18	0.113	0.16	0.15	0.11	0.12	0.151	0.07	-0.02	0.08	0.02	-0.04	-0.05	0.07	0.08	-0.01	0.09	0.03	-0.04	-0.05	0.07
Rochester, NY	8081	0.24	0.27	0.24	0.22	0.17	0.22	0.32	0.26	0.275	0.19	0.2	0.17	0.17	0.244	-0.11	-0.11	0.11	0.06	0.02	0.06	0.07	-0.17	-0.12	0.08	0.04	0.06	0.06	-0.03
Syracuse, NY	11489	0.24	0.27	0.22	0.18	0.28	0.31	0.26	0.23	0.255	0.21	0.17	0.27	0.31	0.256	0.09	0.07	0.05	0.01	0.11	0.13	0.14	0.09	0.07	0.05	0.01	0.12	0.12	0.15
Utica-Rome, NY	7982	0.10	0.14	0.14	0.15	0.13	0.11	0.21	0.1	0.135	0.14	0.15	0.13	0.12	0.215	0.00	-0.07	0.05	-0.07	-0.02	-0.04	0.15	-0.01	-0.07	0.05	-0.07	-0.03	-0.03	0.15

Table 1

OES Wage Imputation Simulation Study: Results of Six Experimental Models and the OES Estimator

Louisiana, size four establishments

OCC Title	Number Estab/ OCC Cells	Average Relative Error (Simple average over sims of MEAN(ABS (WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL)							Weighted Average Relative Error Simple average over sims of ∑(ABS (WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / ∑(EMP)							Average Relative Bias Simple average over sims of ∑((WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / ∑(EMP)							Weighted Average Relative Bias Simple average over sims of ∑((WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / ∑(EMP)						
		A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES
ALL	67467	0.23	0.18	0.17	0.25	0.20	0.20	0.25	0.22	0.18	0.17	0.25	0.20	0.20	0.25	0.02	0.02	0.04	0.06	0.06	0.05	0.09	0.03	0.03	0.04	0.06	0.06	0.06	0.10
Managers	4089	0.39	0.30	0.32	0.38	0.30	0.33	0.41	0.39	0.30	0.32	0.38	0.31	0.33	0.41	0.05	0.02	0.09	0.11	0.11	0.13	0.15	0.05	0.02	0.10	0.12	0.11	0.13	0.13
Bus.&Fin. Operations	3270	0.54	0.33	0.39	0.63	0.39	0.49	0.72	0.54	0.32	0.39	0.64	0.38	0.49	0.73	0.30	0.17	0.25	0.43	0.21	0.32	0.50	0.31	0.16	0.26	0.44	0.21	0.34	0.52
Computer & Math	1029	0.37	0.43	0.40	0.32	0.40	0.46	0.46	0.34	0.44	0.38	0.33	0.41	0.43	0.47	0.10	0.24	0.20	0.06	0.21	0.34	0.25	0.09	0.30	0.18	0.08	0.27	0.32	0.26
Architecture and Eng.	5257	0.27	0.26	0.27	0.28	0.32	0.27	0.34	0.27	0.26	0.26	0.30	0.32	0.27	0.34	0.02	0.08	0.08	0.06	0.13	0.09	0.15	0.05	0.10	0.10	0.10	0.16	0.11	0.17
Life, Physical, and Social Science	1097	0.38	0.29	0.28	0.37	0.40	0.31	0.24	0.36	0.28	0.25	0.37	0.38	0.28	0.25	0.11	0.01	0.13	0.09	0.17	0.12	0.06	0.14	0.02	0.08	0.09	0.16	0.07	0.05
Community and Social Service	1938	0.35	0.30	0.25	0.37	0.38	0.33	0.36	0.40	0.32	0.27	0.41	0.43	0.32	0.38	0.16	0.12	0.02	0.22	0.22	0.15	0.25	0.23	0.17	0.06	0.28	0.29	0.18	0.28
Legal	1018	0.26	0.38	0.26	0.41	0.44	0.37	0.18	0.26	0.38	0.26	0.41	0.44	0.37	0.18	-0.05	0.20	0.16	0.25	0.33	0.25	0.04	-0.04	0.20	0.16	0.25	0.33	0.25	0.04
Education, Training, and Library	1152	0.28	0.10	0.15	0.39	0.17	0.17	0.33	0.28	0.10	0.15	0.39	0.17	0.17	0.33	0.26	0.06	0.10	0.35	0.13	0.13	0.26	0.26	0.06	0.10	0.35	0.13	0.13	0.26
Art, Design, Entertain., Sprts, Media	2380	0.34	0.21	0.22	0.33	0.29	0.36	0.37	0.34	0.20	0.21	0.33	0.28	0.36	0.37	0.07	-0.05	0.06	0.12	0.12	0.13	0.21	0.07	-0.06	0.06	0.13	0.11	0.12	0.21
Healthcare Prac. and Tech.	3528	0.25	0.25	0.24	0.24	0.26	0.24	0.32	0.25	0.25	0.25	0.24	0.26	0.24	0.32	0.11	-0.01	0.06	0.08	0.08	0.09	0.16	0.12	0.00	0.06	0.08	0.08	0.09	0.16
Healthcare Support	1476	0.21	0.17	0.17	0.22	0.18	0.19	0.24	0.21	0.17	0.18	0.21	0.18	0.19	0.24	-0.09	-0.05	-0.06	-0.12	-0.04	-0.06	-0.13	-0.09	-0.07	-0.07	-0.12	-0.04	-0.07	-0.13
Protective Service	1430	0.19	0.18	0.17	0.20	0.24	0.26	0.34	0.21	0.20	0.17	0.23	0.25	0.28	0.35	0.07	0.02	0.01	0.07	0.02	0.07	0.19	0.10	0.04	0.02	0.10	0.04	0.09	0.20
Food Preparation and Serving	4620	0.07	0.06	0.07	0.09	0.08	0.08	0.09	0.07	0.06	0.07	0.08	0.08	0.07	0.10	0.01	0.01	0.03	0.02	0.04	0.04	0.04	0.01	0.02	0.03	0.02	0.04	0.04	0.04
Building and Grounds, Cleaning and Main.	1003	0.18	0.10	0.09	0.18	0.09	0.12	0.17	0.18	0.10	0.10	0.18	0.09	0.12	0.17	0.09	-0.02	0.03	0.04	0.00	0.04	0.06	0.09	-0.02	0.03	0.04	0.00	0.04	0.07
Personal Care and Service	2435	0.18	0.14	0.11	0.20	0.14	0.14	0.17	0.18	0.15	0.11	0.21	0.14	0.14	0.17	0.00	0.00	0.01	0.03	0.02	0.02	-0.01	0.00	0.00	0.01	0.04	0.03	0.02	-0.01
Sales and Related	2463	0.44	0.23	0.22	0.50	0.26	0.24	0.67	0.46	0.22	0.22	0.51	0.24	0.24	0.75	0.14	-0.08	0.01	0.17	-0.04	-0.01	0.45	0.17	-0.08	0.02	0.17	-0.06	-0.01	0.51
Office and Administrative Spt.	6089	0.19	0.16	0.17	0.21	0.18	0.21	0.22	0.19	0.16	0.17	0.22	0.18	0.21	0.23	-0.05	0.00	0.02	-0.05	0.04	0.04	0.05	-0.05	0.00	0.03	-0.04	0.04	0.04	0.06
Farming, Fishing, and Forestry	322	0.35	0.71	0.15	0.23	0.28	0.18	0.20	0.35	0.71	0.15	0.23	0.28	0.18	0.20	-0.06	0.57	-0.02	-0.13	-0.15	-0.08	-0.10	-0.06	0.57	-0.02	-0.13	-0.15	-0.08	-0.10
Construction and Extraction	6636	0.24	0.22	0.20	0.24	0.22	0.21	0.22	0.24	0.22	0.20	0.24	0.21	0.21	0.22	-0.04	0.03	0.07	0.01	0.06	0.03	0.04	-0.04	0.04	0.08	0.01	0.06	0.03	0.05
Installation, Maint., and Repair	5411	0.29	0.20	0.17	0.35	0.22	0.22	0.32	0.29	0.20	0.17	0.35	0.22	0.21	0.31	0.02	0.02	0.01	0.07	0.06	0.04	0.12	0.02	0.02	0.01	0.07	0.07	0.04	0.12
Production	7187	0.27	0.22	0.19	0.29	0.22	0.23	0.25	0.26	0.21	0.18	0.29	0.22	0.22	0.25	-0.03	-0.04	-0.03	-0.01	-0.02	-0.02	0.08	-0.03	-0.04	-0.03	-0.02	-0.03	-0.02	0.07
Transportation and Mat. Moving	3637	0.28	0.19	0.18	0.31	0.22	0.21	0.33	0.23	0.20	0.18	0.28	0.23	0.22	0.30	0.08	0.10	0.08	0.13	0.11	0.09	0.13	0.03	0.10	0.08	0.09	0.11	0.09	0.10

Louisiana, size four establishments

Area Name	Number Estab/ OCC Cells	Average Relative Error (Simple average over sims of MEAN(ABS (WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL)							Weighted Average Relative Error Simple average over sims of ∑(ABS (WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / ∑(EMP)							Average Relative Bias Simple average over sims of ∑((WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / ∑(EMP)							Weighted Average Relative Bias Simple average over sims of ∑((WAGE_EST - WAGE_ACTUAL) / WAGE_ACTUAL) * EMP / ∑(EMP)						
		A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES	A1	R1A	R1A_r1	R1B	R1C	R1F	OES
|	2638	0.14	0.11	0.12	0.14	0.13	0.14	0.18	0.14	0.11	0.12	0.14	0.13	0.14	0.18	0.04	0.05	0.08	0.05	0.09	0.11	0.11	0.04	0.05	0.07	0.05	0.10	0.11	0.11
Balance of State Areas	3480	0.21	0.16	0.18	0.23	0.18	0.19	0.30	0.22	0.16	0.18	0.23	0.18	0.19	0.30	0.07	0.02	0.07	0.11	0.07	0.07	0.22	0.07	0.02	0.06	0.11	0.07	0.07	0.23
|	5268	0.22	0.24	0.16	0.30	0.25	0.24	0.28	0.22	0.24	0.15	0.31	0.25	0.24	0.29	0.04	0.07	0.04	0.18	0.12	0.09	0.18	0.05	0.08	0.05	0.19	0.12	0.10	0.19
|	6103	0.30	0.25	0.25	0.34	0.29	0.26	0.28	0.31	0.25	0.24	0.35	0.30	0.25	0.28	0.07	0.06	0.08	0.10	0.13	0.11	0.09	0.08	0.07	0.08	0.11	0.15	0.10	0.09
Alexandria, LA	3295	0.14	0.11	0.13	0.24	0.14	0.15	0.15	0.13	0.11	0.12	0.24	0.14	0.15	0.15	0.02	-0.01	0.03	0.12	0.04	0.04	0.08	0.02	0.00	0.03	0.13	0.04	0.05	0.08
Baton Rouge, LA	8771	0.22	0.19	0.17	0.25	0.20	0.20	0.26	0.22	0.19	0.17	0.26	0.20	0.20	0.26	0.01	0.01	0.03	0.07	0.04	0.03	0.10	0.01	0.00	0.03	0.07	0.03	0.02	0.10
Houma-Bayou Cane-Thibodaux, LA	6091	0.20	0.18	0.15	0.22	0.21	0.18	0.21	0.20	0.18	0.15	0.22	0.22	0.19	0.21	0.00	0.06	0.04	0.03	0.07	0.08	0.10	0.00	0.07	0.05	0.03	0.08	0.08	0.10
Lafayette, LA	6487	0.24	0.17	0.18	0.24	0.21	0.21	0.24	0.22	0.18	0.18	0.22	0.21	0.21	0.23	0.09	0.05	0.07	0.09	0.11	0.09	0.13	0.07	0.05	0.08	0.07	0.11	0.10	0.12
Lake Charles, LA	6163	0.18	0.12	0.12	0.19	0.15	0.17	0.21	0.18	0.12	0.12	0.19	0.14	0.17	0.22	-0.02	-0.05	0.00	-0.02	0.02	0.02	0.07	-0.01	-0.05	0.00	-0.01	0.02	0.02	0.07
Monroe, LA	3328	0.14	0.13	0.12	0.15	0.12	0.13	0.14	0.14	0.13	0.12	0.15	0.12	0.13	0.13	0.00	-0.04	0.01	0.00	0.02	0.01	0.03	0.00	-0.04	0.01	0.00	0.02	0.01	0.04
New Orleans-Metairie-Kenner, LA	9084	0.30	0.20	0.20	0.31	0.22	0.22	0.34	0.30	0.19	0.20	0.32	0.21	0.22	0.34	-0.01	0.01	0.02	0.01	0.02	0.01	0.05	0.00	0.01	0.02	0.01	0.02	0.01	0.05
Shreveport-Bossier City, LA	6759	0.21	0.19	0.18	0.23	0.19	0.20	0.22	0.21	0.19	0.19	0.23	0.20	0.20	0.22	0.05	0.05	0.06	0.06	0.07	0.08	0.07	0.05	0.06	0.07	0.07	0.08	0.09	0.07

Table 2

OES Wage Imputation Simulation Study: Results of Six Models and the OES Estimator

Simulation Variance

1. New York, size seven establishments

Percentage Employment for which simulation variance is <= table entry	Model						
	A1	R1A	R1A_r1	R1B	R1C	R1F	OES
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
30	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
70	0.00	0.00	0.00	0.00	0.00	0.00	0.00
75	0.00	0.00	0.00	0.00	0.00	0.00	0.00
85	0.01	0.01	0.01	0.01	0.01	0.01	0.02
90	0.12	0.16	0.10	0.15	0.14	0.15	0.20
95	2.00	0.60	0.81	1.41	0.75	0.78	1.59
98	5.65	3.30	3.94	5.02	3.53	4.55	8.98
99	11.02	8.58	9.63	9.80	7.35	10.32	16.89
99.5	20.21	14.99	15.22	16.54	14.63	15.15	30.82
100	1110.34	1790.14	1513.22	1513.22	1824.61	1841.19	1267.13

2. Louisiana, size four establishments

Percentage Employment for which simulation variance is <= table entry	Model						
	A1	R1A	R1A_r1	R1B	R1C	R1F	OES
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
30	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
70	0.00	0.00	0.00	0.00	0.00	0.00	0.00
75	0.00	0.00	0.00	0.00	0.00	0.00	0.00
85	0.11	0.08	0.08	0.15	0.08	0.11	0.17
90	0.48	0.36	0.32	0.58	0.37	0.41	0.64
95	2.38	1.53	1.31	2.39	1.72	1.74	2.60
98	9.03	5.99	4.95	10.03	7.98	7.57	9.70
99	20.94	13.00	11.89	23.80	17.50	14.31	22.66
99.5	34.78	24.42	25.39	40.60	35.87	25.32	47.30
100	2207.65	1620.61	1701.49	2083.04	1622.91	1184.54	673.38