

## **Specifications for Ratio Method Imputation for QCEW (Quarterly census of Employment and Wages) Program**

**1.0 Purpose:** The purpose of this document is to provide the general programming specification for implementing the ratio imputation method for the QCEW data on private and government establishments.

### **1.1 Cell (stratum) structures:**

The cell structures to be used in the imputation processing for private sector is: ST/2-digit NAICS/6 size classes/ownership =5.

The cell structure for government sector is given in the paper titled “Simulation of Imputations by ratio method for QCEW Government Data” dated August 14, 2009.

The size classes are defined by using the third month employment of **previous quarter (PQ)** of the **imputation quarter (CQ)**. For example, if the imputation is for **CQ-2008Q1** then the **PQ** is 2007Q4 (First four digits are the year, Q is for quarter and number following Q stands for the quarter number, 1 being January, February and March Quarter and 4 means October, November and December Quarter).

We define six size classes based on 3<sup>rd</sup> month of the previous quarter as follows:

Size Class 0: employment 0 (zero) (excluded from this imputation. This class has its own imputation method)

Size Class 1: employment 1 – 4

Size Class 2: employment 5 – 9

Size Class 3: employment 10 – 19

Size Class 4: employment 20 – 49

Size Class 5: employment 50 – 99

Size Class 6: employment 100+

### **1.2 Data Required:**

Imputation procedures require at least two consecutive quarters (including the CQ) of data for size classes 1 thru 6 and five consecutive quarters (including CQ) for zero size class imputation. **The zero size class imputation is specified in a separate document titled “Proposed Methodology for Imputation of Missing Data for Zero Size Class Units in QCEW” dated August 14, 2009.**

### **1.3 Identifying Influential Observations**

Before computing cell ratios, it is essential to identify influential observations and exclude them in the formation of cell trends. To be used in cell ratio computation the unit

(establishment) must meet the following criterion. The unit must have reported values for the variables (see Appendix-A) m3\_empl, tot\_wages from PQ; and m1\_empl, m2\_empl, m3\_emp and tot\_wages from CQ; these reported values are necessary to compute individual unit ratios. In other words, the respective indicators namely m3\_ind, totw\_ind from PQ and m1\_ind, m2\_ind, m3\_ind and totw\_ind from CQ must equal 'R'. These establishments should be selected in a table and sorted by cell structure.

The following formulae are used to define the individual unit ratios.

$$(1) \quad re_t = \frac{empl_t}{empl_{t-1}} \quad t = 4, 5, 6$$

$$(2) \quad rw = \frac{cw}{pw}$$

where,

- empl<sub>t</sub>*            reported employment for month t
- re<sub>t</sub>*                ratio of month employment for month t to month t-1
- cw*                 current quarter reported wages
- pw*                 previous quarter reported wages
- rw*                 ratio of wages

**Note:** Imputation is done using two consecutive quarters. The labeling of t = 1 to the first month of the PQ, and t = 6 to the last month of the CQ. Computation of employment ratios for months t = 4, 5, 6 corresponding only to the CQ months.

Once a unit ratios for employment and wages are calculated, they are compared to table1 or dataset to determine if it is an influential observation. This table or dataset is from the Appendix B – Influential Observations for Imputing QCEW Data.

Table 1 Cut off values for ratios in each size class for employment and wages for upper bounds

Size	Employment Ratio-1	Employment Ratio-2	Employment Ratio-3	Wage Ratio
1: 1 – 4	5	5	5	9
2: 5 – 9	5	5	5	6
3: 10 – 19	5	5	5	6
4: 20 – 49	5	5	5	5
5: 50 – 99	5	5	5	4
6: 100 +	5	5	5	3

Following example illustrates the computation of ratios.

**Example 1: Computation of ratios for individual establishment**

Assume that the record of employment and wages for some establishment for the last two quarters looks as follows:

Previous quarter				Current quarter			
M1_empl (m1)	m2_empl (m2)	m3_empl (m3)	Tot-wage (w_p)	m1_empl (m4)	m2_empl (m5)	m3_empl (m6)	Tot_wage (w_c)
6	6	7	50,000	7	10	9	620,000

We compute the employment ratios for the establishment:

$$re_4 = \frac{m4}{m3} = \frac{7}{7} = 1 \quad re_5 = \frac{m5}{m4} = \frac{10}{7} = 1.43 \quad re_{65} = \frac{m6}{m5} = \frac{9}{10} = 0.9$$

and the wage ratio:

$$rw = \frac{w_c}{w_p} = \frac{620,000}{50,000} = 12.4$$

As per cut-off values in table 1, the rw in this example is greater than 6 (size class 2) therefore, this unit is an influential observation that will not be included in computation of cell ratios.

### Example2:

In this example one of the reported values in the current quarter is zero.

Previous quarter				Current quarter			
m1_empl (m1)	m2_empl (m2)	m3_empl (m3)	Tot-wage (w_p)	m1_empl (m4)	m2_empl (m5)	m3_empl (m6)	Tot_wage (w_c)
6	6	7	50,000	0	10	9	60,000

We compute the employment ratios for the establishment:

$$re_4 = \frac{m4}{m3} = \frac{0}{7} = 0 \quad re_5 = \frac{m5}{m4} = \frac{10}{0} = \text{undefined} \quad re_{65} = \frac{m6}{m5} = \frac{9}{10} = 0.9$$

and the wage ratio:

$$rw = \frac{w_c}{w_p} = \frac{60,000}{50,000} = 1.2$$

Even though one of the ratio re4 equals 0 and other re5 is undetermined, **the unit is included in the computation of cell ratios in the current simulation.**

## 1.4 Computation of cell ratios:

In this section the ratios for the entire cells of the establishments are computed and called cell ratios. All ratios are computed on the basis of employment (employment cell ratio) and wages (wage cell ratio). Each cell will be denoted by  $h(t)$ ; where  $h$  is the cell and  $t$  is the month within the 2 consecutive quarters.

### 1.4.1 Employment cell ratios:

For the purpose of employment imputation for each cell  $h$ , three employment cell ratios are calculated for every month of the current quarter:  $RE_{h(4)}$ ,  $RE_{h(5)}$ ,  $RE_{h(6)}$ . A cell ratio is a quotient of the sum of employment for all the reported units excluding the influential observations of the imputation month and the sum of employment of all the reported units excluding the influential observations of the prior month. The formula for employment cell ratio is:

$$(3) \quad RE_{h(t)} = \frac{\sum_{h(t) \cap (\text{Rep-Outliers})} \text{employment}}{\sum_{h(t-1) \cap (\text{Rep-Outliers})} \text{employment}} \quad t=4, 5, 6$$

Outliers are influential observations and Rep are the reported values for all the six variables. For example:

$h(t) \cap (\text{Rep-Outliers})$  – the set of all establishments in cell  $h$  during month  $t$  which are reported units, and are not influential observations

### 1.4.2. Wage Cell Ratios:

For wage imputation, the wage ratio is computed for each cell by dividing the sum of wages of all reported units of the current quarter excluding influential observations by the sum of wages of all reported units of the previous quarter excluding influential observations. The formula for wage cell ratio is:

$$(4) \quad RW_{h(\text{current})} = \frac{\sum_{h(CQ) \cap (\text{Rep-Outliers})} \text{TotalWages}}{\sum_{h(PQ) \cap (\text{Rep-Outliers})} \text{TotalWages}}$$

Where, **Rep –outliers** are units which are reported units excluding influential observations.

## 1.5 Imputing Employment and Wages:

Imputation starts with employment imputation followed by the wage imputation. The imputation process is illustrated through the following example:

### Example 3:

Let establishment xx have the following record of employment for the first quarter:

m1_em pl	m2_em pl	m3_em pl	tot_wag e1	m4_em pl	m5_em pl	m6_em pl	tot_wag e2
3	3	4	28000	?	?	?	?

Questions marks mean that we need to impute values there.

#### **1.5.1 Steps used in imputation of employment and wages:**

Impute employment for m4\_empl, m5\_empl, m6\_empl using the employment cell ratios for the month(s) of interest.

$$\begin{aligned}m4\_empl &= RE_{h(4)} * m3\_empl; \\m5\_empl &= RE_{h(5)} * m4\_empl; \\m6\_empl &= RE_{h(6)} * m5\_empl; \\tot\_wage2 &= RW_h * tot\_wage1;\end{aligned}$$

Suppose that the  $RE_{h(4)} = 1.0$ ,  $RE_{h(5)} = 1.0$  and  $RE_{h(6)} = 1.25$  then outcome of employment will be as follows by using the above equations.

$$\begin{aligned}m4\_empl &= 1.0 * 4 = 4 \\m5\_empl &= 1.0 * 4 = 4 \\m6\_empl &= 1.25 * 4 = 5\end{aligned}$$

Now assume that  $RW_h = 1.05$  then outcome of wages imputation is given by  $tot\_wage2 = 1.05 * 28,000 = 29,400$

m1_em pl	m2_em pl	m3_em pl	tot_wag e1	m4_em pl	m5_em pl	m6_em pl	tot_wag e2
3	3	4	28,000	4	4	5	29,400

#### **1.5.2 Rounding of imputed values:**

Generally, the imputed employment results in decimal value by imputation method. Use a random number generator routine to round the employment to an integer. During process of rounding the imputed value should be separated into two parts, the integer and the remaining fraction. Assign a random number to every imputed unit for comparison. If the random number is greater than the remaining fraction then use integer value; if the random number is less than the remaining fraction then add 1 to the integer value.

### Example 4:

Suppose that the resulted imputed employment is 3.18, then using a random number generator routine there is a 18% probability the employment is rounded upward to four and 82% probability it is rounded to three.

Imputed value: 3.18   Integer value: 3   Remaining Fraction: 0.18

Suppose the random number is 0.003 then

$0.003 \leq 0.18$  then the imputed units value will be 4

Now suppose that the random number is 0.6775 then

$0.6775 > 0.18$  then the imputed units value will be 3

**Note:** Using the rule of 5 is likely to bias the employment values downward as vast majority of the imputed units have employment of less than five and the rate of change is generally less than one percentage point (i.e.,  $4 \times 1.01 = 4.04$ ).

## **1.6   Exceptions to the imputation**

As shown earlier there is a possibility of computing zero ratio value or a ratio that is undetermined at a unit level. In the present research these units have been included for computation of the cell ratios. **During the simulation no situation arose in which a cell ratio for a particular month was zero or undetermined, but in production this is a possibility. SMS recommends in these cases (ratio = 0 or undefined) use the last reported value of the establishment.**

**Establishments with no prior available data are essentially birth units. For them, use the last reported or imputed value going forward. That is, if second month employment is given and the third month employment is missing, then use second month employment to impute for third month employment. However, do not impute employment for the first month of the current quarter; this assumes, the establishment came into existence in the second month of the current quarter.**

## Appendix-A

### QCEW imputation simulation in-scope conditions:

The data items which were used in this study are defined below.

The variable names used in the following conditions are as follows.

Variable	Description
State_fips	State Fips code data field on LDB
Yr_qtr	year quarter data field on LDB
cyr	Current year of processing (First 4 digits of yr_qtr data field on LDB)
cqtr	current quarter of processing (Last digit of yr_qtr data field on LDB)
sic	standard industry code (LDB data field)
naics	North American Industry Code Standards(data field on LDB)
m3_empl(pm)	Third month employment of previous quarter(data field in LDB)
tot_wage(pwage)	Total quarterly wages (data field on LDB)
m1_empl	First month employment of a quarter of processing year quarter( data field on LDB)
m2_empl	Second month employment of a quarter of processing year quarter( data field on LDB)
m3_empl	Third month employment of the processing year quarter (data field on LDB).
m1_ind	First month employment indicator of the processing quarter (data field on LDB).
m2_ind	Second month employment indicator of the processing quarter (data field on LDB).
m3_ind	Third month employment indicator of the processing quarter (data field on LDB).
Tot_wage	Total quarterly wages (data field on LDB)
Totw_ind	Total wage indicator (data field on LDB)
own_code	Owner ship code (data field on LDB)
size_yrqrtr	size of establishment based on the last month's employment of previous quarter (defined field).
Naics2_yrqrtr	Two digit naics (first two digits of naics data field)

### In-Scope conditions used at the time of data extraction from the LDB for the processing quarters are as follows:

We excluded all the records having industry classification of private households and public administration. That is prior to 3rd quarter of year 2003, all the records having a sic code of either 8811 or sic in the range 9099 to 9800 are excluded; after the 2<sup>nd</sup> quarter of 2003, all records having the NAICS code of either 814110 or NAICS code in the range of 919999 and 930000 are excluded. Lastly, only private units are included in simulation. That is the records with owner ship code of 5 are included. The other ownership codes, i.e., Federal, State and Local Government data simulation is done separately.