

LMI Training Institute
June 16-20, 2003

Prepared by the Staff of the LEHD Program

June 2, 2003

Chapter 2

QWI Weights

2.1 Guide to the QWI Weights

2.1.1 Motivation

The Quarterly Workforce Indicators (QWI) contain detailed labor market information, some of which is similar to released reports by the Bureau of Labor Statistics (BLS). For example, beginning of period employment (B) calculated from Unemployment Insurance (UI) records might be expected, at least conceptually, to be comparable to BLS/ES-202 first month employment. However, various factors such as different rules of inclusion, reporting discrepancies at the firm level, and timing related computation issues result in significant differences in both level and trend. The weights are designed to address these issues, insuring statistics on the QWI agree closely with similar statistics in BLS published reports.

2.1.2 Definitions

Indexes

$s = 1, \dots, S$ geographical area, usually states of the U.S.

$t = 1, \dots, T_s$ quarter of the year, beginning in 1990:1 and running to the last quarter data is available for geographic entity s

$j = 1, \dots, J_{s,t}$ firm index, where $J_{s,t}$ represents the number of firms in state s in period t

$u = 1, \dots, U_{s,j,t}$ unit index, where $U_{s,j,t}$ represents the number of units at firm j in state s in period t

Variables

There are two main sources of data used to calculate the weights. The target totals come from BLS published month 1 employment. The other inputs are taken directly from the Employer Characteristics File (ECF), the main source of firm-level data inside LEHD. The ECF is built primarily from ES-202 data, but also contains firm-level employment and payroll information calculated from the UI worker level data.

$L_{s,t}$ = BLS published month 1 employment (private) for state s in period t

$O_{s,j,u,t}$ = ECF ES-202 month 1 employment for unit u in firm j in state s in period t

$B_{s,j,t}$ = ECF UI beginning of period employment for firm j in state s in period t

$E_{s,j,t}$ = ECF UI end of period employment for firm j in state s in period t

$M_{s,j,t}$ = ECF UI employment for firm j in state s in period t

$$V_{s,j,t} = \begin{cases} B_{s,j,t}, B_{s,j,t} > 0 \\ E_{s,j,t}, B_{s,j,t} = 0, E_{s,j,t} > 0 \\ M_{s,j,t}, B_{s,j,t} = 0, E_{s,j,t} = 0, M_{s,j,t} > 0 \\ 0, \text{ otherwise} \end{cases}$$

Functions

In order to classify firms and units into various categories several functions are defined.

$$P(j) = \begin{cases} 1, M_{s,j,t} > 0, O_{s,j,t} > 0 \\ 2, M_{s,j,t} > 0, O_{s,j,t} = 0, \text{ firm } j \text{ on ECF in at least one } t \\ 0, \text{ otherwise} \end{cases}$$

$$C(u) = \begin{cases} 1, \text{ ownership code}=5 \text{ and first digit of SIC} \neq 9 \\ 0, \text{ otherwise} \end{cases}$$

$$C_{s,j,t} = \begin{cases} 1, \left[\sum_u C(u) \right] > 0 \\ 0, \text{ otherwise} \end{cases}$$

2.1.3 Simple Weight

In order to get a better understanding of the QWI weight let's create a simple one first. The target we would like to hit is first month BLS employment using $B_{s,j,t}$ calculated from the UI data.

Although the two measurements of employment should be similar, there are several reasons why the sources differ. First, coverage is not the same, resulting in no UI reported information for many active firms on the ES-202 (however, the converse is rarely true, only a small amount of reported employment is contained within firms on the UI that are not on the ES-202). This leads to totals of $B_{s,j,t}$ at the state level being strictly less than totals produced from the ES-202. However, even if coverage was exactly the same and there were no reporting differences we would still expect employment totals to differ.

Remember, $B_{s,j,t}$ is calculated from individual level data and has a reference date as of the beginning of the first month of the quarter, while BLS first month employment is reported as of the 13th of the first month. Thus, some workers employed on the 1st of the month will no longer be employed on the 13th and vice-versa. In addition a more subtle difference arises due to the way information is reported on the UI versus the ES-202. For example, in order to estimate $B_{s,j,t}$ a firm must be present on the UI for both period t and $t - 1$. This implies that for new firms active on the first and the 13th of the month $B_{s,j,t} = 0$ while $O_{s,j,t} > 0$.

Equation 2.1 below represents our objective, scale B such that the state total is the same as BLS first month employment.

$$L_{s,t} = w0 * \sum_j B_{s,j,t} \quad (2.1)$$

$$w0_{s,t} = \frac{L_{s,t}}{\sum_j B_{s,j,t}} \quad (2.2)$$

The resulting weight, $w0$, is the ratio of BLS first month employment in state s to the sum of B for every firm in state s in time period t . Depending on the information we have about the data generating process for each series the above weight might be sufficient. However, if the relationship varies significantly by observable characteristics within a state and time period then we may want to calculate a separate weight along those dimensions.

2.1.4 Employer Characteristics File (ECF) Weight

The ECF weight expands on the ideas presented in the previous section and is made up of two components. The first component is similar to the simple weight above, while the second component uses detailed firm level information available on the ECF.

$$w1_{s,t} = \frac{L_{s,t}}{\sum_{j \in \{P(j)=1\}} \sum_{u \in \{C(u)=1\}} O_{s,j,u,t} + \sum_{j \in \{P(j)=2\}} B_{s,j,t} * C_{s,j,t}} \quad (2.3)$$

$$w2_{s,j \in \{P(j)=0\},t} = 0 \quad (2.4)$$

$$w2_{s,j \in \{P(j)=1\},t} = \left[\frac{O_{s,j,t}}{V_{s,j,t}} \right] \quad (2.5)$$

$$w2_{s,j \in \{P(j)=2\},t} = \left[\frac{V_{s,j,t}}{V_{s,j,t}} \right] \quad (2.6)$$

$$qwi_unit_weight_{s,j,t} = w1_{s,t} * w2_{s,j,t} \quad (2.7)$$

Keep in mind that the weight, $qwi_unit_weight_{s,j,t}$, is positive whenever $P(j) > 0$, even though only firms with $C_{s,j,t} = 1$ are included in the denominator of $w1$. Therefore, we are implicitly assuming $[w1_{s,t} | C_{s,j,t} = 1] = [w1_{s,t} | C_{s,j,t} = 0]$, or that the adjustment factor $w1$ calculated using only private firms (ownership code=5) is similar to the value of $w1$ for public entities (ownership code $\neq 5$).

A trim of the weight is implemented to minimize the impact of large differences in reported employment at the firm level. The upper and lower bounds for $w2_{s,j,t}$ are presently set at .7 and 1.3, respectively. To ensure that the weighted sum, $B_{s,t}$, equals the target number, $L_{s,t}$, the weight formulas are modified using the function defined below.

$$R_{s,j,t} = \begin{cases} .7, 0 < w2_{s,j,t} < .7 \\ 1, .7 \leq w2_{s,j,t} \leq 1.3 \\ 1.3, w2_{s,j,t} > 1.3 \\ 0, \text{ otherwise} \end{cases}$$

The appropriately modified equations 2.3 and 2.7 are shown below.

$$w1_{s,t}^* = \frac{L_{s,t}}{\sum_{j \in \{P(j)=1\}} \sum_{u \in \{C(u)=1\}} O_{s,j,u,t} * R_{s,j,t} + \sum_{j \in \{P(j)=2\}} B_{s,j,t} * C_{s,j,t} * R_{s,j,t}} \quad (2.8)$$

$$qwi_unit_weight_{s,j,t}^* = w1_{s,t}^* * w2_{s,j,t} * R_{s,j,t} \quad (2.9)$$

2.1.5 Final QWI Weight

The ECF weight, calculated in the previous section, differs slightly from the weight used to calculate the QWI indicators due to edits made to B after the ECF is created. In order to improve the accuracy of the QWI estimates, flows of workers from one firm to another are suppressed based on predecessor/successor information contained in the Entity Demography File (EDF). These relationships typically arise when a firm, for whatever reason, obtains a new SEIN but continues to operate with predominately the same workforce. If not corrected, the switch to the new SEIN will result in a large number of "false" accessions and separations. The edits applied during QWI processing suppress the flows, while at the same increasing B , a primary input to $qwi_unit_weight^*$.

Several other edits are also implemented, but these have the opposite effect of decreasing B relative to the values on the ECF. Records are removed if any of the following conditions are met; 1. missing data (sex, age, industry, etc. This should never actually occur since all missing data is imputed earlier in the processing sequence), 2. a person with more than 1000 job spells over the sample period, 3. age is less than 14 or greater than 99.

To account for these edits a correction factor, qwi_wcf , is calculated during part A of the Unit Flows File (UFF) sequence. The correction factor is then multiplied by $qwi_unit_weight^*$, thus creating qwi_final_weight . The formula below gives the details of the calculation of qwi_wcf . Notice that B , ECF UI beginning of period employment is replaced with \hat{B} , QWI edited first period employment.

$$qwi_wcf_{s,j,t} = \frac{L_{s,t}}{\sum_j \hat{B}_{s,j,t} * qwi_unit_weight^*_{s,j,t}} \quad (2.10)$$

The final QWI weight, given below, is simply the product of the correction factor and the weight calculated directly from the ECF. In practice, the correction factor is typically very small ($\sim .98$), resulting in a final weight very close to $qwi_unit_weight^*$.

$$qwi_final_weight_{s,j,t} = qwi_wcf_{s,j,t} * qwi_unit_weight^*_{s,j,t}$$

2.1.6 Directions for the Future

Several improvements to the weights were suggested in comments received from Phil Hardiman and Richard Kihlthau of California. They point out two situations where statistics generated as part of the QWI should not be compromised to agree with the information on the ES-202. In general this occurs when actual firm reports are missing and the employment and payroll data are then imputed. In this case, it may not be desirable to adjust reported information on the QWI to match imputed information on the ES-202.

Currently, QWI processing does not utilize any information on data quality contained in various imputation/error flags available on the ES-202 (research on an appropriate method of standardization is currently underway). The use of these variables will allow the weight generation program to ignore ES-202 employment and payroll information when the flags suggest the data is in error. The two cases they mention are outlined below with comments.

Case 1

1. ES-202 payroll imputation flag indicates the value is reported
2. ES-202 payroll closely equals QWI total wages
3. ES-202 first month employment is imputed

This case occurs when a firm reports payroll on their quarterly tax return to avoid an administrative penalty, but does not report ES-202 monthly employment and an automated system generates imputed estimates based on historical employment. This case can be integrated into the weights generation program by ignoring the imputed ES-202 employment value and treating it as if it were missing.

Case 2

1. ES-202 payroll imputation flag indicates

2. ES-202 payroll closely equals QWI total wages
3. ES-202 first, second, and third month employment equals M , or the total number of wage records on the UI.

This case occurs if the firm reports ES-202 first, second, and third month employment as the total number of wage records instead of the actual employment on the 13th of each month. However, this is a legitimate report if the firm had no employee turnover during the period and all employees were actively working at the firm on the 13th of each month. The relationship between B , E , and M can be used to distinguish at least some of the valid reports from those that are invalid (for example, if $B = M = E$ then it is likely the ES-202 report is correct and no adjustment to the weights is necessary). Further research will be necessary before proceeding with this modification to the weights.