Appendix F: The EMSI “Complete Employment” Data Set

Introduction and Rationale
QCEW-derived data are a widely used resource in regional/local economic and labor market research. However, they exclude some payroll jobs and all non-payroll jobs such as proprietors and partners. Particularly in some sectors, such as agriculture, construction, professional/technical services, or real estate, non-payroll workers may have a very significant presence.

EMSI seeks to overcome this limitation of QCEW with its Complete Employment dataset. This data set attempts to record nearly every job in every sector at the local level.

Data Sources Used for EMSI Complete Employment
- EMSI Covered Employment
- Nonemployer Statistics (NES), Census Bureau
- Regional Economic Accounts and State/Local Personal Income reports, Bureau of Economic Analysis
- Current Employment Statistics, Bureau of Economic Analysis
- County Business Patterns (CBP), Census Bureau
- Ten-Year Industry Projections (BLS; various state agencies)

Employment Coverage of EMSI Complete Employment
EMSI’s Complete Employment dataset has a significantly expanded definition of “jobs” compared to EMSI Covered Employment. As a result, job numbers in most areas are significantly higher. Since many non-covered jobs are part-time, and a person can hold more than one job, this is to be expected.

Whereas EMSI Covered Employment matches the coverage of the Bureau of Labor Statistics’ Quarterly Census of Employment and Wages, EMSI Complete Employment uses Bureau of Economic Analysis data (www.bea.gov/bea/regional/) as its top-level benchmark, with a few adjustments. In addition to covered jobs taken care of by QCEW, BEA data attempt to count all types of paid employment.

Apart from farm employment, perhaps the most important class of workers excluded from QCEW but included in BEA data is the self-employed, which includes sole proprietorships and partnerships. Like covered jobs, self-employed jobs include both full- and part-time positions (e.g., one worker with a covered wage/salary job and self-employment on the side would be counted as two jobs). Unlike covered jobs, self-employed jobs from BEA data (1) may be reported geographically by place of residence or place of work (since they are based on the self-employed workers’ tax returns which may show a business or home address); and (2) represent the sum of all self-employed jobs existing at any time during the year, instead of an annual average. For these reasons, EMSI Complete Employment job figures may appear inflated to some researchers, especially those who are used to looking at covered employment only.
Industry Earnings in EMSI Complete Employment

In EMSI Complete, industry earnings are the total of three components:

- **Wages and salaries**
- **Supplements to wages and salaries:** Includes employer contributions to employee pension and insurance funds (private and/or government) as well as employer contributions to government social insurance.
- **Proprietors’ earnings:** For nonfarm proprietors, this is generally what is reported as “net-profit-less-loss” to the IRS on forms 1040 and 1065; however, see the “Smoothing the BEA’s Proprietor Earnings” section for important adjustments that EMSI makes to proprietors’ earnings.

Note that the “supplements to wages and salaries” component is defined very differently from the one used by the EMSI Covered Employment dataset, even though both are called “supplements.” For more information, see the Bureau of Economic Analysis’s glossary definition of “Earnings by Place of Work” (www.bea.gov/regional/definitions/).

Methodology for Creating EMSI Complete Employment

In order to capture its complete picture of historical industry employment and earnings, EMSI basically combines covered employment data from Quarterly Census of Employment and Wages (QCEW) produced by the Department of Labor with total employment data published by the Bureau of Economic Analysis (BEA), augmented with County Business Patterns (CBP) and Nonemployer Statistics (NES) published by the U.S. Census Bureau.

The process is not nearly as simple as this description makes it appear; there are literally hundreds of details that must be worked out in order for this process to yield valid data, and millions of data points must be processed. EMSI has spent thousands of staff hours gaining a complete understanding of the data sources involved, and even with several powerful computers, the process takes many hours to complete.

The primary steps in the process are as follows:

1. Perform the EMSI Covered Employment process (see above)
2. Create a dataset of “non-covered” jobs
   a. Unsuppress and extend Nonemployer Statistics (NES) to six-digit 2002 NAICS.
   b. Combine this modified NES with BEA State/Local Income Reports to produce farm and non-farm proprietor-by-industry estimates at the county level. For the most recent years, BEA/NES data may not be available, so EMSI uses projections to move the datasets forward. During this process we also “smooth” BEA proprietor earnings (see below).
   c. Estimate non-covered wage and salary (non-proprietor) employment by county using County Business Patterns.
3. Now the three basic data sets we need are built: (1) covered jobs, (2) non-covered proprietors, and (3) non-covered wage and salary jobs. Next EMSI combines the three to arrive at complete employment estimates by county. Again, there are hundreds of details to be worked out in the process, but they are proprietary to EMSI and beyond the scope of this document.
4. Geography adjustment: at this point, there may be aggregation error due to slightly different estimates at the county, state, and national levels. So EMSI performs a national adjustment to ensure that employment figures aggregate correctly for various geographies.

5. For the current and up to two most recent years, BEA data is projected from published data (which generally is one to two years old) using QCEW, which are updated more frequently.

6. After historical data are finalized, EMSI creates 10-year projections using 15-year past trends and projections produced by state/federal agencies. This process is beyond the scope of this document.

7. Finally, EMSI uses ZIP Code Business Patterns and USPS delivery statistics to create ZIP code industry employment estimates and projections. This process is beyond the scope of this document.

Smoothing the BEA’s Proprietor Earnings

There is one issue in BEA proprietors’ income that EMSI seeks to correct, and which can cause significant differences between earnings in BEA data and earnings in EMSI Complete Employment. In reporting proprietors’ income, the BEA has blurred the line between labor and capital (or “property”) income, while its primary intent is to report only labor income. To understand the problem further, let’s get some background first. While a salaried employee’s income is usually 100 percent labor income, proprietors are different because they are self-employed and may own capital or property used in their business. To take a simple example, a proprietor who owns a trucking business and owns the eighteen-wheeler he drives is receiving income from both his labor and his capital.

Proprietor income reported by the BEA is calculated from income tax forms and basically represents “net profit less loss.” Because of this, it can vary wildly from year to year for a given industry and county as proprietors experience boom and bust cycles or simply take a loss on capital investments. The industry total for a county can even be negative.

This poses two problems for EMSI’s data process:

- EMSI’s input/output model depends on data for a single “base year” to calculate impact scenarios. If data for that base year is atypical or varies greatly from year to year, model results will be unreliable. The model needs an “average” base year of data.

- Labor market researchers need to know if a given industry for proprietors is generally profitable and pays well. This can be obscured in year-to-year swings.

Admittedly, it is sometimes impossible to separate the two types of earnings for proprietors, but EMSI believes that the raw numbers reported by the BEA require some interpretation in order to be usable by input-output models and labor market researchers, especially when they exhibit significant year-to-year variations in the same county and industry. Accordingly, EMSI smooths proprietor earnings using Nonemployer Statistics (NES, another source for proprietor income) by analyzing the variation between NES and the BEA in industry-specific earnings across all states. In effect, EMSI gravitates toward the NES figure when this variation is high.