Employment Data for Regional Labor Market Analysis

Application to the Agriculture, Food and Natural Resources Career Cluster

July 30, 2022

Report prepared for Office of Career, Technical, and Adult Education
Virginia Department of Education

July 30, 2022

ACKNOWLEDGEMENTS

The author, Terance J. Rephann, would like to give special thanks to Trailblazers Project Director, Kathryn Piper Crespin, for contributing to this report. Amy Muldoon assisted with document preparation. Any errors or omissions are the responsibility of the author.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Educational administrators and researchers face an almost bewildering array of state and local employment data sets to choose from when analyzing the local economy and educational program alignment with job opportunities. This paper examines some of the challenges in using regional data for local labor market analysis and illuminates some of the limitations in data availability and quality, using Agriculture, Food, and Natural Resources career cluster employment as an example.

Acquiring an accurate picture of the labor market demand for the Agriculture, Food, and Natural resources career cluster is challenging. The farming profession has and will continue to undergo profound changes due to newer, more advanced technologies, which means the types of skills required will continue to evolve, especially in the direction of more technical, sophisticated skills. In addition, the farming industry is inherently more difficult to capture within available data sources due to such complications as seasonal fluctuations in employment and higher levels of self-employment. Finally, career clusters, including Agriculture, Food, and Natural Resources, may not always closely align with the industry employment with which they are most commonly associated. For example, most employment opportunities for the Agriculture, Food, and Natural Resources career cluster are outside the Agriculture and Forestry industry sector, with the Professional, Scientific, and Technical Services sector accounting for 32% of employment. All of these factors can make it more difficult for CTE professionals and counselors to gauge changes in career opportunities for the career cluster.

The limitations of all available data sources, whether public, private, or open-source, contributes to the difficulty in gaining an accurate regional outlook of labor market demand in the Agriculture, Forestry, and Natural Resources career cluster.

Six publicly available data sets provide industry and occupational employment information for the Agriculture, Food, and Natural Resources career cluster. Each excels in some aspect, such as timeliness or completeness of data, but not in all; so one source cannot provide a full analysis of the career cluster. In addition, data suppression methods used to prevent disclosure of sensitive data make it difficult to find data for more disaggregated industries and smaller geographical units. Finally, most public regional data sources do an incomplete job of accounting for career cluster employment. Thus, exclusive reliance on such data sources in program development and career planning may undercount opportunities, particularly for industries where small businesses and agricultural workers form a relatively large share of industry employment.

Several private vendors of regional employment data are available that attempt to overcome many of the limitations of public employment data sources. These products provide user-friendly online software platforms that provide easy access to industry and occupational employment data and tools for data visualization and reporting. However, private vendors employ data estimation techniques to fill in missing data (data that was suppressed in the publicly available data sets), and the accuracy of these estimations is questionable, particularly for detailed industry employment analysis and for smaller geographic units.

In light of these limitations, CTE professionals can take several steps to gain a more accurate regional picture of employment in the Agriculture, Food, and Natural Resources career cluster:

- Because no single data source is perfect, consulting multiple sources can help provide a fuller, more accurate picture of labor market demand. For example, sector-specific data series, such as farm employment data from the USDA, which is based on a Census of Agriculture conducted every five years, may help to fill the gaps in publicly available industry and occupational employment data.
- Carefully evaluating the methodology of the data source is critical to ensure it is clearly explained and provides enough detail to indicate how the data is measured and what is being measured.
- Focusing on those skills that are currently in demand across clusters, occupations, and industries to inform strategic planning may be prudent rather than relying solely on the SOC classification system, which may lag behind. Local employers, business and industry representatives, and student High-Quality Work-Based Learning experiences continue to be an invaluable resource for staying informed about what occupational skills are currently in greatest demand.

1. INTRODUCTION

This report provides a roadmap for CTE staff to better understand how regional employment can be used in regional labor analysis given data availability and quality limitations, particularly in relation to the Agriculture, Food, and Natural Resources career cluster. It examines some of the challenges in using the data for local labor market analysis, including the constant upskilling of jobs and job categories due to rapid technological change, the differential alignment of CTE career clusters with individual industries, and limitations in employment data availability and quality. It also compares and contrasts several competing public and private employment data sources and describes their strengths and weaknesses. An example of how to use the data to analyze Agriculture, Food, and Natural Resources career cluster employment is included.

The report is divided into five sections. The first section describes Virginia's Agriculture, Food, and Natural Resources career cluster, its composition, relative size, growth, and change. The second section shows how the cluster aligns with various industries and how the Agriculture and Forestry industry aligns with various career clusters. The third section examines the advantages and limitations of six public sources of regional industry and occupational employment data. The fourth section provides a general overview of several competing private employment data products and methods used to overcome some of the limitations of public data by expanding the universe of coverage or estimating undisclosed data. It also describes the potential pitfalls of relying on such data, particularly for smaller areas and more granular industries. The fifth section provides a comparison of public and private data sources for the Commonwealth of Virginia and a representative rural county, Wythe County, in Southwest Virginia. The report ends with a summary and conclusions.

2. AGRICULTURE, FOOD, AND NATURAL RESOURCES CAREER CLUSTER: SIZE, COMPOSITION, AND CHANGE

The Agriculture, Food, and Natural Resources career cluster is one of 17 Virginia CTE career clusters. Jobs in this cluster account for only about 2% of Virginia CTE career cluster wage and salary jobs covered by Unemployment Insurance (UI) according to 2018 Bureau of Labor Statistics (BLS) Occupational Employment and Wage Statistics (OEWS) data. This career cluster ranks 13th among the 17 Virginia CTE career clusters in terms of size (number of jobs), with Business Management and Administration (15%), Marketing (13%), Hospitality and Tourism (11%), Architecture and Construction (10%), and Health Science (8%) representing over half of the total. According to Advance CTE (n.d.), an organization representing state CTE officials, this cluster concentrates on "the production, processing, marketing, distribution, financing, and development of agricultural commodities and resources including food, fiber, wood products, natural resources, horticulture, and other plant and animal products or resources." The Agriculture, Food, and Natural Resources career cluster encompasses seven pathways: (1) Agribusiness Systems, (2) Animal Systems, (3) Environmental Service Systems, (4) Food Products and Processing Systems, (5) Natural Resources Systems, (6) Plant Systems, and (7) Power, Structural, and Technical Systems.

Table 1 shows the CTE career pathways and nonduplicated occupations included in the Agriculture, Food, and Natural Resources career cluster in Virginia. A nonduplicated list results from each occupation being assigned to only one CTE career cluster and pathway. It is recommended for use in analyzing labor market needs, including current employment levels and employment projections, because it avoids double-counting. However, a duplicated list provides a fuller picture of occupational opportunities for use in career planning.

Table 1. Virginia Agriculture, Food, and Natural Resources Career Cluster Pathways and Nonduplicated Occupations

Pathway	SOC Code	SOC Title
Food Products and Processing Systems	191012	Food Scientists and Technologists
Food Products and Processing Systems	194011	Agricultural and Food Science Technicians
Food Products and Processing Systems	513021	Butchers and Meat Cutters
Food Products and Processing Systems	513022	Meat, Poultry, and Fish Cutters and Trimmers
Food Products and Processing Systems	513023	Slaughterers and Meat Packers
Food Products and Processing Systems	513091	Food and Tobacco Roasting, Baking, and Drying Machine Operators and Tenders
Food Products and Processing Systems	513092	Food Batchmakers
Food Products and Processing Systems	513099	Food Processing Workers, All Other

Pathway	SOC Code	SOC Title
Plant Systems	191013	Soil and Plant Scientists
Plant Systems	373012	Pesticide Handlers, Sprayers, and Applicators, Vegetation
Plant Systems	373013	Tree Trimmers and Pruners
Animal Systems	291131	Veterinarians
Animal Systems	292056	Veterinary Technologists and Technicians
Animal Systems	319096	Veterinary Assistants and Laboratory Animal Caretakers
Animal Systems	392011	Animal Trainers
Animal Systems	392021	Nonfarm Animal Caretakers
Power, Structural, and Technical Systems	493041	Farm Equipment Mechanics and Service Technicians
Natural Resources Systems	119121	Natural Sciences Managers
Natural Resources Systems	191032	Foresters
Natural Resources Systems	194093	Forest and Conservation Technicians
Natural Resources Systems	333031	Fish and Game Wardens
Natural Resources Systems	454021	Fallers
Natural Resources Systems	454022	Logging Equipment Operators
Natural Resources Systems	454023	Log Graders and Scalers
Environmental Service Systems	172081	Environmental Engineers
Environmental Service Systems	173025	Environmental Engineering Technicians
Environmental Service Systems	192041	Environmental Scientists and Specialists, Including Health
Environmental Service Systems	194091	Environmental Science and Protection Technicians, Including Health
Environmental Service Systems	372021	Pest Control Workers
Environmental Service Systems	474041	Hazardous Materials Removal Workers
Environmental Service Systems	518031	Water and Wastewater Treatment Plant and System Operators
Environmental Service Systems	537081	Refuse and Recyclable Material Collectors
Agribusiness Systems	119013	Farmers, Ranchers, and Other Agricultural Managers
Agribusiness Systems	451011	First-Line Supervisors of Farming, Fishing, and Forestry Workers

The Agriculture, Food, and Natural Resources career cluster is projected to experience only modest growth based on CTE Trailblazers calculations using the 2018 to 2028 Virginia Employment Commission (VEC) Occupational Projections. The cluster is projected to grow just 4% compared to average occupational growth of 7% among all occupations over the period. However, this career cluster includes both fast-growing and slow-growing occupations. For example, Veterinarians, Veterinarian Techs, and Veterinarian Assistants are projected to grow 31-34% over the period, while Farmers, Ranchers, and Other Agricultural Workers are projected to lose employment (-2%).

This slow growth is due to the increasing role of technological change. New technology is transforming production practices and reshaping the demand for labor skills. Two examples in farming and food processing illustrate how the production landscape is changing. First, farmers are increasingly combining data and automation in a process described as "precision farming." For crop production, this can involve the use of yield sensors, variable rate seeders, maps, data, and algorithms to identify the optimal levels and assignments of inputs (e.g., fertilizer, water, herbicide, and pesticide). It can also involve the use of new equipment, such as autonomous systems and robotics (e.g., driverless tractors, robotic fruit harvesters, weeding, spraying and pruning controllers, and drones for overhead farm monitoring and mapping). Similar changes are occurring in food manufacturing industries. For example, many poultry and livestock processing facilities are incorporating new technologies, such as three-dimensional (3D) scanners, automated cutting machines, automated deboning machines, and new conveyor systems to increase the speed, quality, and safety of their operations.

Both of these developments will have profound effects on future labor demand in the Agriculture, Food, and Natural Resources cluster. They will speed the shift from less skilled to more skilled workers and create challenges for current workers and educators. CTE career counselors may find it increasingly challenging to convince students that the new Agriculture, Food, and Natural Resources vocations are not the grueling, dirty, and dangerous jobs of yesteryear but rather high-tech professions requiring knowledge of science, mathematics, and technology (National Academies of Sciences, Engineering, and Medicine 2021). In the most recent update of the Standard Occupational Classification (SOC) system in 2018, for example, Agricultural and Food Science Technicians (SOC 2010 19-4011) were split into two new occupations—Agricultural Technicians (SOC 2018 19-4012) and Food Science Technicians (SOC 2018 19-4013)—reflecting a shift to a more specialized scientific and technical skill set for each. Similarly, some Agricultural Equipment Operators (SOC 45-2091) may need to learn new skills in precision agriculture to maintain employment in farming. Workers in meat processing, such as Meat, Poultry, and Fish Cutters and Trimmers (SOC 51-3022), may need to learn new skills in sensor and robotics repair to remain in the industry. Although these types of skills may more often be associated with occupations such as Electro-Mechanical and Mechatronics Technologists and Technicians (SOC 17-3024) or with career clusters such as Manufacturing and STEM, they are becoming increasingly important in the Agriculture, Food, and Natural Resources cluster as well.

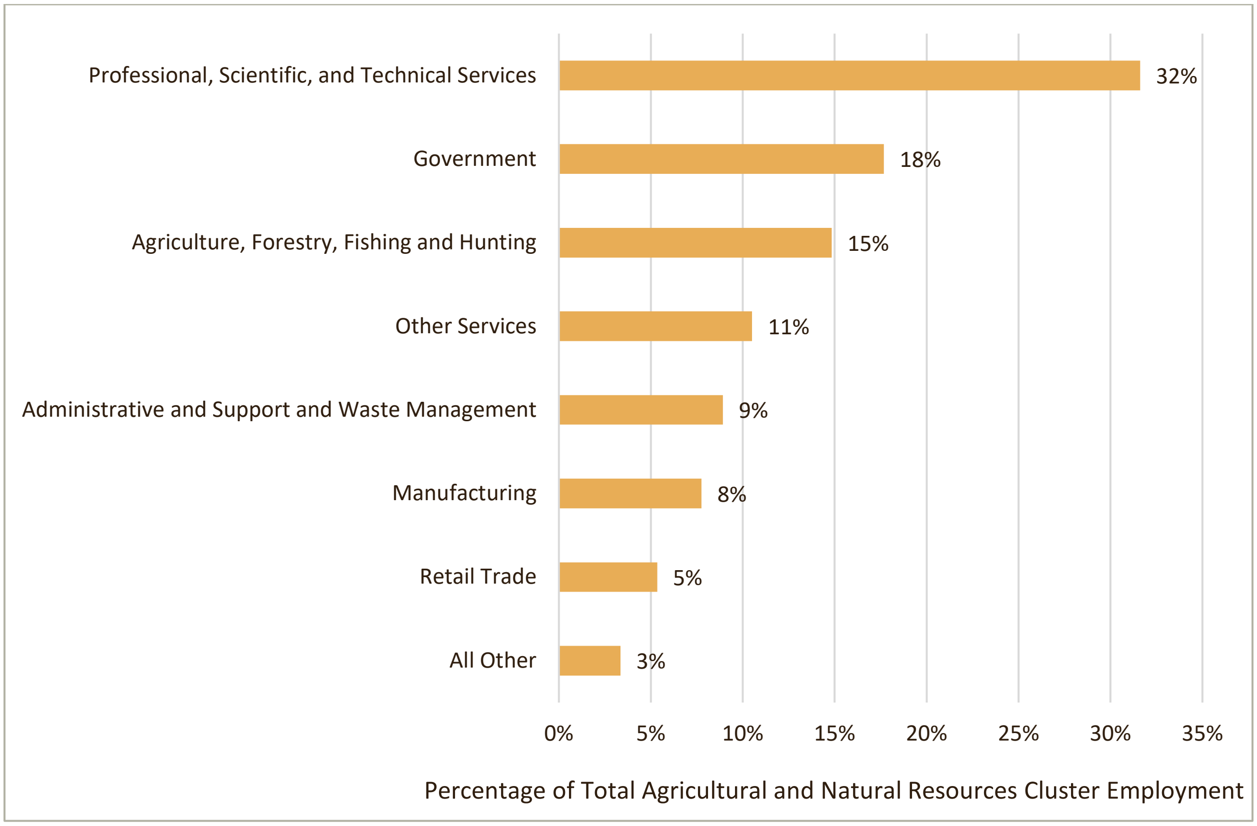
3. AGRICULTURE AND FORESTRY INDUSTRY VERSUS AGRICULTURE, FOOD, AND NATURAL RESOURCES CLUSTER

Although Agriculture, Food, and Natural Resources career cluster occupations may commonly be associated with the Agriculture and Forestry industries, the occupations are employed in a much wider range of industries, and the industries employ workers drawn from a range of career clusters. One definition of Agriculture and Forestry industries, which draws on definitions used by the U.S. Department of Agriculture Economic Research Service and U.S. Forestry Service, includes agriculture and forestry growing and harvesting; agricultural and forestry services; food, beverage, textiles, and apparel processing; and wood and paper products manufacturing. In addition, selected agricultural and forest production distribution activities are included (Rephann 2008). The industry can be partitioned into three subcategories consisting of production (the first two categories), processing (the second two categories), and distribution.

The Agriculture, Food, and Natural Resources career cluster includes some production, processing, and distribution occupations. Agribusiness Systems is closely connected to production agriculture, while Natural Resources Systems is linked to forestry. The Food Product and Processing Systems pathway has a close nexus with agriculture processing and wholesale/retail trade, but it also includes occupations that are part of the agriculture and forestry supply chain (and, thus, lie outside the Agriculture and Forestry industry boundaries). The Animal Systems and Power, Structural, and Technical Systems pathways include occupations that provide goods and services that allow agribusinesses to operate, such as Veterinarians and Farm Equipment Mechanics and Service Technicians, while occupations within the Environmental Service Systems pathway tend to serve an even wider variety of industries. On the other hand, the industries used in Rephann (2008) include some forestry product processing and farm and forestry product distribution industries, such as sawmills, pulp and paper mills, and warehousing and storage. The workers in these industries are concentrated in CTE career clusters such as Manufacturing and Transportation, Distribution, and Logistics rather than in Agriculture, Food, and Natural Resources.

Figure 1 shows a mapping of Virginia Agriculture, Food, and Natural Resources CTE career cluster occupations onto industries. It indicates that most Virginia workers in this career cluster are employed outside the Agriculture and Forestry production industries. About half are employed within the Professional, Scientific, and Technical Services and Government industries. Just 15% are in the Agriculture, Forestry, Fishing and Hunting industry, 11% in Other Services, 9% in Administrative and Support and Waste Management, and 8% in Manufacturing. The remaining 8% are employed in other industries.

Figure 1. Virginia Agriculture, Food, and Natural Resources CTE Career Cluster Employment by Industry

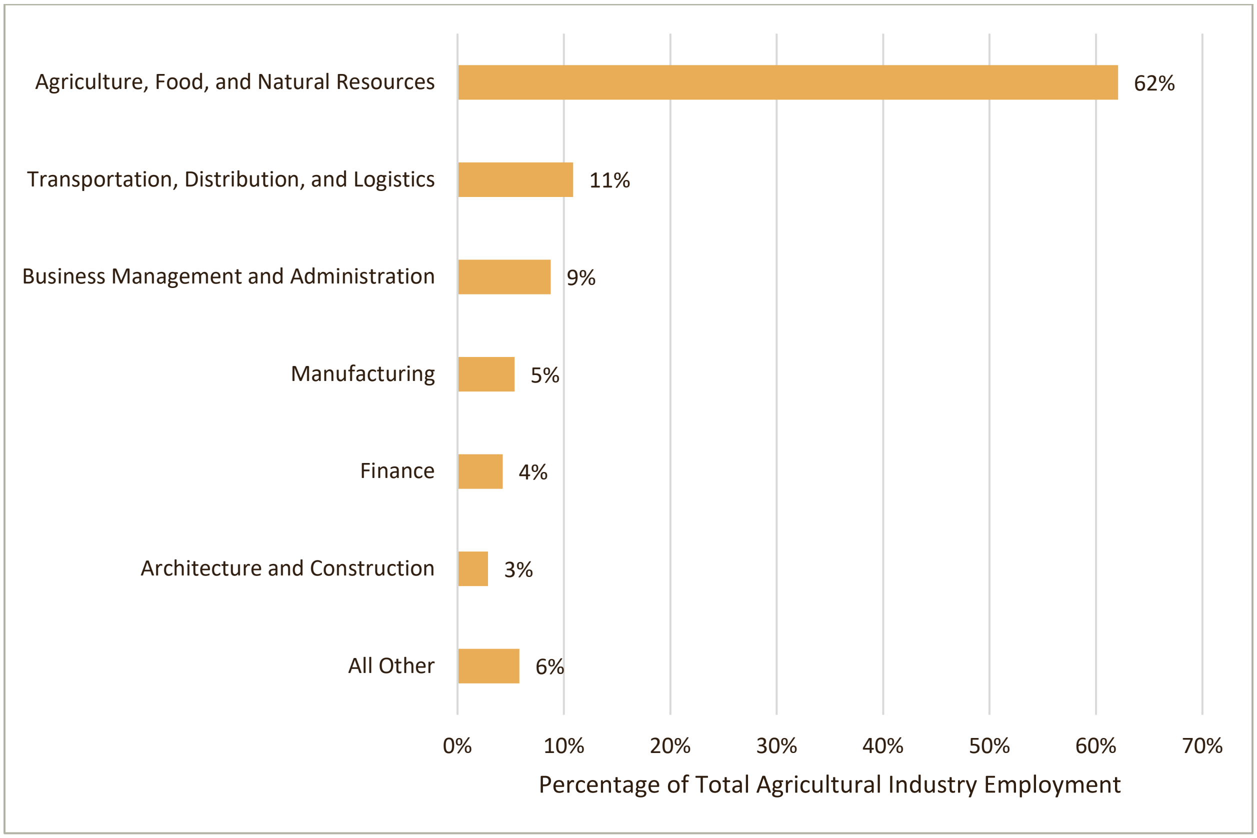


Data Source: Lightcast™ (formerly Emsi Burning Glass) industry data and a national industry-occupation matrix from the Bureau of Labor Statistics.[i]

Note: Estimated Virginia Agriculture, Food, and Natural Resources CTE career cluster employment by two-digit NAICS industry for 2020 using Lightcast data and Bureau of Labor Statistics National Industry-Occupation Matrix

On the other hand, Agriculture and Forestry industry production activity aligns closely with the Agriculture, Food, and Natural Resources career cluster (see **Figure 2**). Looking at industry rather than career cluster employment, over 60% of Agriculture and Forestry industry production workers are drawn from the Agriculture, Food, and Natural Resources cluster. An additional 11% are drawn from the Transportation, Distribution, and Logistics cluster, 9% from the Business Management and Administration cluster, and 5% from the Manufacturing cluster. The remaining 13% are drawn from other career clusters.

Figure 2. Virginia Agricultural Industry Employment by CTE Career Cluster



Note: Estimated Virginia Agriculture and Forestry industry production employment (NAICS=111-113, 115) by CTE career cluster for 2020 using Lightcast data and Bureau of Labor Statistics National Industry-Occupation Matrix.

4. PUBLIC REGIONAL INDUSTRY AND OCCUPATIONAL EMPLOYMENT DATA

Three U.S. government agencies provide local industry employment data for public use.[ii] The data series differ not only in terms of the agency of collection but also the sampling and estimation methods, timeliness of data, completeness of data (e.g., inclusion of farmers), industry detail, geographical detail, and data suppression and noise injection methods used to prevent disclosure of sensitive data. They include the Bureau of Labor Statistics (BLS), the U.S. Census Bureau, and the Bureau of Economic Analysis (BEA). The BLS is a statistical data collection and analysis agency within the U.S. Department of Labor. It is widely recognized as the primary source of employment data for the United States, states, and localities, and it partners with state employment agencies to collect, analyze, and report data. The second agency tasked with reporting geographical industry employment data is the U.S. Census Bureau, located within the U.S. Department of Commerce. It is the nation's largest statistical data collection and analysis agency for demographic and economic data. It also publishes a source of national, state, and local employment data called County Business Patterns (CBP) using different data and estimation methods than the other agencies. The BEA is an agency also located in the U.S. Department of Commerce that is most commonly recognized for reporting information on U.S. gross domestic product as part of its National Income and Product Accounts (NIPA). Lesser known are its regional economic accounts, which provide state and local estimates for several measures parallel to national accounts, including employment data. The employment series is the most comprehensive measure of the geographical industry employment.

All of the employment data sets have certain features in common. Each reports the number of jobs, not the number of workers. Also, the total job figures include both part-time and full-time jobs. Moreover, geographical job figures are reported by place of work rather than place of residence. The geographical data products are subject to suppression rules in order to conceal confidential business employment and payroll information and to avoid presenting data that fails to meet agency data accuracy standards. Lastly, each source reports employment on an industry level. Industries are identified using the NAICS (North American Industrial Classification System) hierarchical numbering system. NAICS uses two to six digits to characterize industries with greater degree of specificity for higher digit representations. The first two digits represent the economic sector; the third digit is the subsector; the fourth digit corresponds to industry group and the fifth and sixth digits indicate the industry. An illustration from the Agriculture and Forestry industries shows the utility of this hierarchical classification scheme. In analyzing food processing employment activity, two-digit codes 31-33 represent the Manufacturing sector, the three-digit code 311 represents Food Manufacturing, the four-digit code 3115 indicates Dairy Product Manufacturing, the five-digit code 31151 is Dairy Product (except Frozen) Manufacturing, and the six-digit code 311511 constitutes Fluid Milk Manufacturing.

4.1 Public Regional Industry Employment Data

Four geographical industry-level data sets are available from the three federal agencies (BLS, U.S. Census Bureau, and BEA) (see **Appendix A**). They include:

- **Quarterly Census of Employment and Wages (QCEW)**. This industry employment and wage data series is available on a quarterly basis for the nation, states, metropolitan areas, and localities (counties and their equivalents). It is based on quarterly payroll tax information provided by business and government employers that have wage and salary employees covered by state and federal unemployment insurance (UI) programs. According to the BLS, the reporting captures approximately 95% of all wage and salary workers. Although it represents close to a census of workers covered by UI insurance, it does not count employees that are exempt from the UI programs. Among the workers falling into this category are some agricultural (e.g., farm contract workers) and domestic workers, self-employed individuals in unincorporated businesses, other business proprietors, members of the U.S. Armed Forces, railroad workers (who are covered by railroad unemployment insurance), and workers for some nonprofit organizations. The QCEW provides employment data for 6-digit level NAICS industries, but they are often subject to suppression at the local level to protect business confidentially.
- **State and Metro Area Employment and Earnings (SAE)**. The BLS provides monthly nonfarm payroll employment estimates. It is the most frequently available series, published on the first Friday after the close of the reporting month. The data are based on a monthly survey of approximately 700,000 non-agricultural establishments reporting UI-covered employees in QCEW, and the estimates are benchmarked with the QCEW. The SAE data product excludes important categories of workers, including many agricultural workers, self-employed individuals, and military members. But it makes adjustments to cover certain categories of workers in non-UI-covered jobs, including workers at railroads, nonprofit organizations, and a few other categories of jobs. It provides primarily major industry sector employment information. Because the monthly survey this data set is based upon samples of a smaller number of establishments than the other data products, it is more geographically restricted than they are, providing data only for states and metropolitan areas. It is also subject to substantial industry-level revisions due to its reliance on some non-UI-covered data sources that have a reporting lag.
- **County Business Patterns (CBP)**. CBP data are based on employer data from Census Bureau business surveys (e.g., Economic Census) and federal administrative records. The coverage includes wage and salary workers enrolled in Social Security. It excludes proprietors, most agricultural workers, members of the U.S. Armed Forces, railroad workers, domestic workers, and some workers in nonprofit organizations. Unlike other employment data sources that provide annualized figures, CBP employment represents March employment tallies. CBP provides employment data disaggregated for detailed (6-digit NAICS) industries. In addition to the suppression of employment data for these sectors, which is done by some competing industry employment data series as well, the CBP employs a statistical "noise infusion" process that applies random statistical noise to reported data in order to maintain business confidentiality. CBP also uses disclosure flags indicating the amount of error associated with

employment levels reported (i.e., 0 to less than 2%, 2 to less than 5%, and greater than or equal to 5%).

- **Local Area Personal Income and Employment (LAPI)**. The BEA provides the most comprehensive picture of total employment as part of its LAPI product. The QCEW wage and salary employment information is an important component of the employment estimation. Nationally, QCEW-covered wage and salary employment represented about 73% of all LAPI employment in 2020, with coverage by industry varying. The biggest source of discrepancy was the business proprietor category (including the self-employed). However, the BEA uses other data sources to estimate workers who are not covered by unemployment insurance, including proprietors and farm workers, and also uses other data cleaning and imputation procedures. For example, information on the number of proprietors is drawn from IRS data, while farm employment gaps are estimated using information from the U.S. Department of Agriculture (USDA) Agricultural Resource Management Study (ARMS) and Census of Agriculture. Because of the substantial additional estimation and processing needed, LAPI data are available with a long time lag, typically about 11 months. Moreover, it provides local employment data at only a two-digit NAICS level. The comprehensive coverage of farm sector and other agricultural workers makes it more appropriate than other employment data series for agricultural industry analysis. Unfortunately, it does not break out industries with agriculturally-dependent subsectors, such as agricultural services and food and beverage manufacturing, in sufficient detail to identify employment levels.

4.2 Public Regional Occupational Employment Data

BLS and its state employment agency partners also produce two occupational data sets:

- **Occupational Employment and Wage Statistics (OEWS)**. The BLS produces estimates of occupational employment for states, metropolitan areas, and the balance of state territory located in non-metropolitan areas. Estimates are provided for over 800 occupations, which are arranged within a six-digit Standard Occupational Classification (SOC) system with a hierarchical structure similar to NAICS. The estimates are drawn from a three-year rolling survey of 1.2 million nonfarm establishments based on QCEW records. Thus, the survey has the same wage and salary UI coverage as QCEW and excludes those categories of workers not found there, such as proprietors, many farmworkers, domestic workers, and the military.
- **Occupational Projections**. The BLS also provides occupational employment estimates for a baseline year and employment projections 10 years beyond the baseline. Baseline coverage includes workers counted in the SAE (i.e., UI-covered employment and some uncovered jobs) plus selected other categories of jobs not included in the SAE (e.g., the unincorporated self-employed) from information derived from Census Bureau statistics. Thus, the employment coverage is not as broad as LAPI. The BLS converts industry employment estimates to occupations using a national industry-occupation matrix showing industry staffing patterns derived from OEWS data for wage and salary jobs and other residual categories (self-employed, some agricultural and domestic workers). Staffing patterns for each industry are adjusted by BLS to reflect expected industry changes in occupational staffing in the future for wage and salary

workers and for the self-employed for the purposes of projecting employment. BLS provides only national occupational employment projections. BLS state labor market information (LMI) affiliates, such as the Virginia Employment Commission (VEC), issue state and regional (e.g., Local Workforce Development Area) projections based on a BLS methodological template, but state affiliates are allowed some latitude in drawing on other state data sources to improve the accuracy of the projections. Like other data products, both the BLS and VEC suppress selected occupational employment data because of confidentiality restrictions and data quality issues. BLS and state LMI affiliate agencies caution that occupational estimates and projections are not "exact counts" or "forecasts" but rather approximate indicators of "levels and trends" and should be used in combination with other information about local economies.[iii]

All six of the publicly available data sets described above have some limitations in providing industry and occupational employment information for the Agriculture, Food, and Natural Resources career cluster at the regional level. For example, several of the sets exclude categories of workers common in agricultural industries—such as business proprietors, self-employed individuals, and many types of agricultural workers—by focusing primarily on those workers who are covered by unemployment insurance programs (QCEW and OEWS), enrolled in Social Security (CBP), or working only in non-agricultural business establishments (SAE). Even when these special categories of agricultural workers are incorporated into the data (e.g., LAPI), there can be tradeoffs for the enhanced occupational coverage, including less granular industry employment data or a longer wait to get the information. Additionally, when employment data are made publicly available for substate geographic areas (e.g., VEC occupational projections for Local Workforce Development Areas), they are often subject to higher rates of data suppression simply because there aren't enough employers or workers within the geographic area to provide an accurate assessment of employment levels or to avoid violating data confidentiality.

All of these challenges can be an understandable source of frustration among CTE professionals attempting to utilize these publicly available data sources to assess labor market demand in agriculture, especially when it seems like there is a greater degree of "missing" or suppressed data in the Agriculture, Food, and Natural Resources cluster compared to other CTE career clusters, particularly at the regional level. It is important to keep in mind, however, that the accurate collection of data related to farm employment is complicated by several factors that are uniquely prevalent in agricultural industries, including a greater degree of seasonal fluctuations in employment, greater numbers of undocumented workers, and higher levels of self-employment, part-time employment, unpaid family employment, and hobby employment (Hasenstab 2019). To better understand how some of these factors affect local labor market demand in the Agriculture, Food, and Natural Resources cluster and to supplement the information that might be missing from the data series described above (series that attempt to provide data across *all* industries and occupations), CTE professionals might consider consulting sector-specific data series.

Several U.S. agencies provide local employment data from surveys that highlight particular sectors of the economy. For instance, the USDA provides data on farm employment at the local level based on a Census of Agriculture conducted every five years. In addition to providing a complete count of U.S. farms, the Census of Agriculture report provides statistics on land use and ownership, operator characteristics, production practices, and income and expenditures, and it facilitates the identification of trends over time by offering numerous tables and graphs that compare the most recent census to the one conducted five years prior. The U.S. Energy Information Administration provides information on underground and surface coal mining employment. The U.S. Census conducts economic surveys for various industries on a five-year schedule that provide employment estimates by detailed industry. The Census Bureau also produces an annual series called Nonemployer Statistics (NES) that provides local counts of businesses that have no paid employees (a measure of self-employed workers). In some instances, information from these alternative federal data sources is used in estimating employment by other agencies, such as the BEA LAPI.

5. PRIVATE AND OPEN-SOURCE INDUSTRY AND OCCUPATIONAL EMPLOYMENT DATA

In recent years, more local industry and occupational employment data have become available that overcome many of the limitations of the public data described above. They include data products offered by private firms as well as open-source databases created by academic researchers. Lightcast (formerly Emsi Burning Glass) and Chmura Economics and Analytics are vendors of the most comprehensive labor market data products. IMPLAN and REMI offer area industry employment data, either as stand-alone products (imputed QCEW by IMPLAN and eREMI by REMI) or as part of a wider suite of regional economic impact modeling software. In addition, there are several open-source databases available that provide imputations of federal employment data products, such as County Business Patterns (CBP) that includes Eckert et al. (2021) and Upjohn Institute's WholeData. [iv]

The private sector products attempt to remedy limitations of existing public data by performing one or more of the following services: (a) estimating suppressed values in public geographical industry employment data, (b) performing industry crosswalks to create continuity in detailed industry employment due to changes in NAICS definitions over time, (c) accelerating the timeline of detailed employment data availability by updating it on a quarterly or even monthly basis, (d) providing greater geographical detail (e.g., down to county and even zip code region), and (e) covering a more comprehensive employment universe than workers covered by unemployment insurance (e.g., business owners, agricultural workers, and military personnel). These data products are available on online platforms that facilitate analysis, presentation, economic impact modeling (i.e., "what if" type analyses), and data download. Private vendors use company imputation techniques that rely on both public and private/proprietary data sources and algorithms. Open-source providers offer their data sets for free and provide more complete documentation of the underlying data estimation methodologies. Rapid growth in machine learning technologies have assisted the development of improved imputation methods.

There are certain commonalities in the published imputation methods (see **Appendix B**). They exploit information about undisclosed employment data cells from other sources. These sources include industry hierarchy data (e.g., employment data available for a higher, more aggregated level of industry in the region), geographical hierarchy data (e.g., industry totals available for a higher level of geography, such as for the state or the nation), information on the size range of an undisclosed employment number, and the number of establishments by size category (Isserman and Westerveldt 2006; Eckart et al. 2021; Jaromczyk et al. 2021). Temporal data (employment data from time periods before and after the time period of the undisclosed cell) are sometimes also used because industry employment tends to be stable over time (Zhang and Guldmann 2013). Lastly, unsuppressed sector data from other public data sources (e.g., using CBP to help uncover suppressed values for QCEW) can be utilized. Values of the undisclosed cell size that are consistent with the various boundaries (or constraints) can be found using algorithmic search techniques. The imputation methods also differ in the type of methodology used to estimate the missing cells, varying from ad-hoc methods to search optimization algorithms based on linear programming or even more complex and computationally intensive machine learning tools.

Methodological details offered by private data vendors are not as complete as academic research descriptions, but some provide outlines of their basic approach.

Public data sources are subject to various types of errors (U.S. Department of Labor, Bureau of Labor Statistics. n.d.). These include sampling error when the data (such as SAE and OEWS) are derived from survey data. In addition, non-survey errors may occur when coding and recording data, when employing methods to adjust for missing or underreported data, and in other ways. Private data sources introduce additional layers of potential error in the process of estimating undisclosed data.

In an analysis of industry employment imputation products, Carpenter, Van Sandt, and Loveridge (2021) conclude that private data sources are "of unknown accuracy." One obstacle to evaluating some of the private products is the limited documentation available on the imputation methods used because they are considered proprietary, which "limits validity testing and research reproducibility." Another obstacle is that the data products sometimes develop estimates from multiple data sources, which makes it difficult to identify a benchmark public employment data source for evaluation purposes.

There are two ways to assess the accuracy of data imputations. One way is to compare the estimates to non-disclosed public employment data through a data sharing agreement with a government agency. For example, one could gain access to undisclosed employment data from a state LMI agency, but the conclusions drawn from the analysis may not be generalizable to the entire universe of state employment data (Cao 2018). Alternatively, researchers could obtain access to undisclosed data from the federal agencies. In either event, disclosure restrictions would hamper the researchers' ability to share some detailed findings from the research (Carpenter, Van Sandt, and Loveridge 2021). For these reasons, several researchers have used simulation to replicate the properties of public employment data sets like CBP and QCEW, including their suppression characteristics. These simulated data sets offer certain analytical advantages, such as an ability to vary features of the simulated data and ascertain the contexts in which various imputation methods improve the fit. However, Isserman and Westerveldt (2006) caution that federal suppression rules are difficult to mimic and are likely to be modified further by federal agencies if data imputations are successful in approximating undisclosed employment data. As if to underline this point, shortly after Eckart et al. (2021) released their first CBP imputed database in 2018, the Census Bureau changed its disclosure methodology.[v]

The handful of evaluations of employment imputation techniques suggest that they tend to perform poorly in comparison to actual unsuppressed data and simulated data, particularly for NAICS industry employment data more disaggregated than the three-digit NAICS level and for smaller geographical units. Three studies compare imputed data to undisclosed public administrative data (Carpenter, Van Sandt, and Loveridge 2021; Zheng n.d.; Holan et al. 2008), while two others examine data imputations for simulated data with data characteristics and suppression patterns similar to public data (Cao 2018; Zhang and Guldmann 2013).

Carpenter, Van Sandt, and Loveridge (2021) compare imputed industry employment data sets to unsuppressed public data available through a Federal Statistical Data Center research agreement for use in academic research. They find that statistical analysis using privately imputed data creates substantial bias in statistical estimates (up to 82% attenuation in coefficient estimates at low-digit NAICS levels) relative to unsuppressed data and recommend against using such data for analysis below a three-digit NAICS industry level in regression analysis. Zheng (n.d.) compares the results of an imputation procedure used for Indiana QCEW employment data to a confidential state employment data set from the Indiana LMI agency and finds that the imputed values for above the two-digit NAICS level deviated from the actual values by an average of more than 50%. Holan et al. (2008) shows slightly better results, with only 29% of QCEW wage data varying 10% or more from the true values. Cao (2018) utilizes several different multiple imputation methods (a statistical method commonly used to estimate missing survey data) with synthetic QCEW data but finds that they tend to perform poorly for industries with low rates of employment disclosure.

Although not yet a subject of academic research and evaluation, occupational employment estimates likely have even lower reliability. Regional occupational employment estimates are the product of (a) area industry employment estimates and (b) area industry occupational staffing patterns. The hazards of estimating missing data for the former were described previously. However, regional occupational staffing patterns need to be imputed as well since they are not available from public sources.[vi] For example, Lightcast utilizes OEWS and QCEW information in imputing local industry staffing patterns.[vii] Occupational employment estimates present a more formidable assessment challenge since public administrative data for comparison purposes are unavailable. The only source of data is BLS occupational data (OEWS), which is only available for geographies such as such state and metropolitan area for a three-year average period. Moreover, generating occupational employment simulation data for assessing imputation methodologies is likely to be even more challenging than performing the same exercise for industry employment data.

6. PUBLIC AND PRIVATE EMPLOYMENT DATA COMPARISONS

The prevalence of geographical employment suppression for different levels of geographical and industry disaggregation for a representative employment data set (in this case, QCEW) is shown in **Table 3** for the Commonwealth of Virginia. It shows the percentage of industry employment cells that are suppressed for three increasingly more disaggregated geographies (the state, Local Workforce Development Areas (LWDAs), and counties/independent cities) and industry levels (two- to six-digit NAICS levels). The table shows that employment suppression does not occur at a two-digit level for larger geographical units, such as the state or LWDAs. However, 14% (approximately one in seven industries) of six-digit NAICS employment numbers are suppressed at the state level. The problem becomes more pronounced for LWDAs and counties/independent cities. For LWDAs, 41% of six-digit employment numbers are suppressed. For counties, suppression occurs for 17% of two-digit NAICS industries, which grows to 62% of six-digit NAICS industries. This problem is more common for rural (nonmetropolitan) counties than urban (metropolitan) counties. For metropolitan counties, 59% of six-digit NAICS industries have undisclosed employment data; this rises to 71% of industries for nonmetropolitan counties.

Table 2: Suppression Incidence (Percentage of Total Number of Industries) for Virginia QCEW Private Employment, 2020 Annual Data

			NAICS LEVEL		
	2-digit	3-digit	4-digit	5-digit	6-digit
State	0	2	5	7	14
Local Workforce Development Areas	0	12	26	35	41
Counties	17	35	48	59	62
Metropolitan counties	12	31	45	56	59
Nonmetropolitan counties	24	42	57	69	71

The practical differences between the various employment data series can be illustrated by showing industry values and suppression features for various industry levels of aggregation for a single year for different geographies. In the following examples, the more aggregated geographical unit is the Commonwealth of Virginia, and the more granular unit is a representative nonmetropolitan county located in southwestern Virginia (Wythe County). Wythe County is approximately one hour driving distance along I-81 from Roanoke City to the north and Bristol City to the south. It has a relatively large agricultural sector, including a sizable beef cattle herd and thirty dairy farms. Forage crops and corn are also important. The county also hosts several small- and mid-sized enterprises in the food and beverage manufacturing industries.

Table 4 compares state-level estimates for two- to six-digit NAICS industries from four public data sets (LAPI, QCEW, SAE, and CBP) and two private data sets (Lightcast and IMPLAN). LAPI provides the most comprehensive data coverage, including UI-covered and UI-uncovered wage and salary workers, proprietors, and military members. QCEW and CBP provide more limited coverage, restricted to UI-covered jobs in QCEW and social-insurance-covered jobs in CBP. Thus, the total jobs counts are significantly lower than LAPI, representing only approximately three-quarters of its total. SAE provides survey-based estimates of UI-covered and some UI-uncovered workers and is intermediate between LAPI and QCEW or CBP, but closer to the latter. As mentioned previously, CBP employment reporting is slightly different from the other employment data series. Employment cells are injected with various levels of statistical noise, ranging from less than 2% (most industries) to from 2% to 5% or more (Forestry, Fishing, and Related Activities; Utilities; and Commercial Bakeries).

Employment estimates from two private vendor sources are also provided for comparison purposes. Descriptions of these private data sets (Lightcast and IMPLAN) are included in **Appendix C**. Both vendors provide a level of comprehensive coverage of state employment similar to LAPI but use different data imputation processes. The Lightcast total figures (released in July 2022) show total and two-digit industry employment numbers that match closely with available LAPI industry totals. IMPLAN shows a slightly higher total employment estimate and generally more variance from LAPI published industry employment figures.

Table 3. Virginia Employment in 2020 by Two-Digit and Selected Three- to Six-Digit NAICS Industries

NAICS	Industry Description	LAPI	QCEW	SAE	CBP	Lightcast[viii]	IMPLAN[ix]
	Total Employment (Number of Jobs)	**5,100,799**	**3,744,148**	**3,857,100**	**3,857,100**	**5,124,672**	**5,211,312**
111-112	Farm	45,441	8,523	--	--	45,772	49,366
113-115	Forestry, Fishing, and Related Activities	13,271	3,989	--	4,298	13,292	14,440
21	Mining, Quarrying, and Oil and Gas Extraction	8,598	5,108	7,000	5,182	8,576	8,473
22	Utilities	11,537	10,771	10,900	14,471	11,487	11,341
23	Construction	294,411	200,956	202,600	197,430	291,929	311,679
31-33	Manufacturing	247,150	232,457	234,100	241,221	246,203	253,191
311	Food Manufacturing	--	30,347	--	31,002	32,389	45,675
311511	Fluid Milk Manufacturing	--	1,744	--	871	1,780	1,820
311812	Commercial Bakeries	--	1,818	--	2,855	2,106	--
312	Beverage and Tobacco Product Manufacturing	--	10,165	--	11,069	10,743	10,845
312111	Soft Drink Manufacturing	--	2,176	--	2,481	2,202	--

NAICS	Industry Description	LAPI	QCEW	SAE	CBP	Lightcast[viii]	IMPLAN[ix]
42	Wholesale Trade	117,028	105,514	106,200	104,750	115,952	116,441
44-45	Retail Trade	464,589	384,363	387,400	425,169	463,019	421,149
48-49	Transportation and Warehousing	219,690	126,966	131,900	120,952	221,828	225,534
51	Information	79,722	64,907	65,400	97,012	81,365	83,240
52	Finance and Insurance	231,060	142,058	155,700	166,939	230,948	237,557
53	Real Estate and Rental and Leasing	226,003	53,461	54,000	56,906	225,641	249,039
54	Professional, Scientific, and Technical Services	584,123	438,950	442,300	509,368	579,477	628,490
55	Management of Companies and Enterprises	85,650	80,000	80,800	82,124	85,166	84,337
56	Administrative and Support and Waste Management and Remediation Services	311,697	233,161	234,700	280,701	308,402	313,980
61	Educational Services	108,336	58,957	87,800	82,412	120,296	83,963
62	Health Care and Social Assistance	504,124	433,371	439,700	447,937	503,346	499,275
71	Arts, Entertainment, and Recreation	91,427	42,418	44,000	66,686	91,248	93,665
72	Accommodation and Food Services	305,749	278,341	280,800	372,701	304,735	342,281
81	Other Services	291,597	122,675	180,400	171,523	300,522	341,160
...	Government and Government Enterprises	859,596	694,263	711,300	35,612	853,142	842,709
...	Federal Civilian	206,208	186,552	186,600	--	209,341	202,211
...	Military	121,549	--	--	--	121,549	149,320
...	State and Local	531,839	507,711	524,700	--	522,252	491,178
99	Unclassified	--	22,940	--	476	22,326	--

Table 5 compares county-level (Wythe County) industry employment estimates from three public data sets (LAPI, QCEW, and CBP). (Regional data for SAE is available only for metropolitan areas, and private vendor data was not available for comparison.) They illustrate that data suppression is a much more significant problem for rural Wythe County than for the Commonwealth of Virginia and is more prevalent for higher-digit NAICS industries than for two-digit sectors. Some two-digit sectors are suppressed for each of the series; however, they are not always the same sectors because agencies use different suppression criteria.

In the case of QCEW, industry employment information is suppressed if there are fewer than three industry establishments or one establishment accounts for more than 80% of employment in the industry. In addition, QCEW suppresses industry employment data in order to ensure that the values of undisclosed industries cannot be ascertained using information from disclosed sectors. This results in the suppression of three two-digit industries: Utilities, Management of Companies and Enterprises, and Educational Services. Information for more detailed industries is more likely to be suppressed. For example, employment for Food Manufacturing (311), Fluid Milk Manufacturing (311511), and Commercial Bakeries (311812) is not disclosed.

CBP information is suppressed if an industry has three or fewer establishments. In addition, CBP reports employment totals with various amounts of noise. Most employment estimates are infused with less than 2% error, but Manufacturing; Transportation and Warehousing; Information; and Educational Services all have 2% to 5% error, and Mining, Quarrying and Oil and Gas Extraction; Utilities; and Wholesale Trade have more than 5% error. CBP employment data are suppressed for several industries.

LAPI does not report its exact suppression rules. But data are not disclosed for six two-digit sectors, including Utilities; Transportation and Warehousing; Professional, Scientific, and Technical Services; Management of Companies and Enterprises; Educational Services; and Health Care and Social Assistance.

Table 5: Wythe County Employment in 2020 by Two-Digit and Selected Three- to Six-Digit NAICS Industries

NAICS	Industry Description	LAPI	QCEW	CBP
	Total Employment (Number of Jobs)	**14,640**	**10,718**	**9,533**
111-112	Farm	885	32	--
113-115	Forestry, Fishing, and Related Activities	61	16	(D)
21	Mining, Quarrying, and Oil and Gas Extraction	133	113	131
22	Utilities	(D)	(D)	23
23	Construction	666	404	364
31-33	Manufacturing	2,395	2,325	2,119
311	Food Manufacturing	--	(D)	(D)
311511	Fluid Milk Manufacturing	--	(D)	(D)
311812	Commercial Bakeries	--	(D)	(D)
312	Beverage and Tobacco Product Manufacturing	--	736	(D)
312111	Soft Drink Manufacturing	--	728	(D)
42	Wholesale Trade	209	166	505
44-45	Retail Trade	2,077	1,784	1,879

NAICS	Industry Description	LAPI	QCEW	CBP
48-49	Transportation and Warehousing	(D)	131	408
51	Information	68	50	71
52	Finance and Insurance	402	215	250
53	Real Estate and Rental and Leasing	572	77	59
54	Professional, Scientific, and Technical Services	(D)	159	162
55	Management of Companies and Enterprises	(D)	(D)	(D)
56	Administrative and Support and Waste Management and Remediation Services	658	424	304
61	Educational Services	(D)	(D)	78
62	Health Care and Social Assistance	(D)	1,109	1,210
71	Arts, Entertainment, and Recreation	150	73	76
72	Accommodation and Food Services	1,166	1,087	1,524
81	Other Services	764	245	364
...	Government and Government Enterprises	2,357	2,190	--
...	Federal Civilian	86	85	--
...	Military	89	--	--
...	State and Local	2,182	2,105	--
99	Unclassified	--	54	--

(D)=Not Disclosed

Virginia occupational projection industry-level data produced by the Virginia Employment Commission are more difficult to compare to other public industry employment data sets because employment figures report aggregate public and private employment together. As mentioned earlier, VEC occupational projections are based on a BLS methodological template, and BLS reports that they make certain adjustments in their occupational data to expand coverage beyond the mainly UI wage and salary worker coverage of OEWS to include the self-employed and other categories of workers. However, a comparison of VEC 2018 occupational projection industry employment estimates with the 2018 LAPI employment estimates shows that coverage is still substantially lower in the VEC occupational data products because they do not account for most business owners, military personnel, and some other job categories. Total estimated employment for baseline Virginia 2018 employment in the VEC data is 4,179,857 compared to LAPI's 5,298,892 for the same year (79% of the LAPI total).

Coverage is also significantly lower for industries where many Agriculture, Food, and Natural Resources cluster workers are found. For example, Virginia occupational projection baseline 2018 employment for Forestry, Fishing, and Related Activities (NAICS 113-115) was 4,619 and for Professional, Scientific, and Technical Services (NAICS 54), it was 427,684. These employment numbers compare to parallel industry private employment figures found in LAPI of 14,332 for Forestry, Fishing, and Related Activities and 573,989 for Professional, Scientific, and Technical Services. The former Virginia occupational projection employment figures represent less than 33% and 75% of LAPI employment figures in those two industries, respectively. Thus, VEC baseline and projected employment figures likely fail to account for a substantial number of Agriculture, Food, and Natural Resources career cluster employment opportunities.

7. SUMMARY AND CONCLUSIONS

This report examines the challenges of using regional employment data in analyzing jobs associated with CTE career clusters, drawing on the Agriculture, Food, and Natural Resources career cluster as an example. The report finds that caution should be made in making inferences from available public and private data sources.

Occupations within the Agriculture, Food, and Natural Resources career cluster are rapidly transforming due to technological change, and the skills required by workers in related fields are undergoing a considerable amount of redefinition and recalibration as a result. Also, career clusters may not always closely align with the industry employment with which they are most commonly associated. For example, most employment opportunities for the Agriculture, Food, and Natural Resources career cluster are outside the Agriculture and Forestry industry sector, with the Professional, Scientific, and Technical Services sector accounting for 32% of employment and Government making up 18%. The Agriculture, Forestry, Fishing, and Hunting industry is third at 15% of employment. On the other hand, most jobs in Agriculture and Forestry production draw from occupations in the Agriculture, Food, and Natural Resources career cluster. Still, about 38% of workers in Agriculture and Forestry production industries come from occupations in other career clusters, reflecting the diverse skills and occupational specialties needed to produce agricultural and forestry commodities.

There are certain features of employment data that would be most ideal for the purposes of regional labor market analysis. The ideal data source would provide a (a) comprehensive count of the total number of jobs in the regional economy, (b) be available on a timely basis, (c) provide rich geographical detail, and (d) offer granular industry detail. Unfortunately, no single existing public data series has all of these characteristics. Individual public employment data sources excel along one or more of these dimensions but not all. For example, BEA LAPI is the most comprehensive data source and closest estimate of the total number of jobs by aggregate industry available in a local economy, while BLS SAE is the only labor market data available on a monthly basis. Both BLS QCEW and U.S. Census CBP provide granular geographical industry employment detail down to a six-digit NAICS level. Unfortunately, the former provides industry estimates only for metropolitan areas (and the state nonmetro residual area), while industry employment for the latter is subject to data disclosure rules that result in a significant amount of data suppression, particularly for more disaggregated industries and smaller geographical units.

Most public regional data sources do an incomplete job of accounting for Agriculture, Food, and Natural Resources career cluster employment, as is the case for other career clusters not examined in this report (e.g., Construction). Some data sources capture only a portion of jobs available in the industry sectors where many Agriculture, Food, and Natural Resources occupations are found, such as Agriculture, Forestry, Fishing, and Hunting and Professional, Scientific, and Technical Services. BLS data sources provide full coverage of jobs enrolled in state and federal unemployment insurance programs. However, coverage of uncovered agriculture and forestry workers, some government employees, the self-employed, and other business proprietors is incomplete. The VEC occupational projections add some categories of uncovered UI workers to their baseline employment estimates but miss others, such as business proprietors and the military, resulting in an undercount of approximately 21% of jobs in

Virginia. Thus, exclusive reliance on this data source in program development and career planning may undercount opportunities, particularly for industries where small businesses and agricultural workers form a relatively large share of industry employment. The VEC occupational data set is also subject to similar data disclosure restrictions as the industry employment data series. Regional occupational employment estimates and projections are not provided for some occupations because of confidentiality rules and because data may not meet quality standards.

Several private vendors of regional employment data are available that attempt to overcome many of the limitations of public employment data sources. These products provide user-friendly online software platforms that provide easy access to industry and occupational employment data, tools for data visualization and reporting, economic impact modeling, and access to additional job data, such as local job openings culled from other sources. The employment and occupational data improve upon public data by offering comprehensive job coverage, timely reporting, and granular industry detail for local economies. The private vendors generate these data products by utilizing proprietary imputation (data estimation) techniques to fill in the missing data, with some data products requiring a multi-step series of imputations to produce the final data set. Several studies have examined how successful these data imputation methods are in accurately estimating the missing data. This research has examined the performance of imputed employment data compared to confidential public data and to simulated data with employment and suppression characteristics similar to the publicly available data. **This research collectively suggests that caution should be used in making inferences from imputed data, particularly for more detailed industry employment categories and for smaller geographical units**. Although no research was examined that looked at chain imputations (i.e., data sets that rely on imputed industry data in combination with imputed industry occupational staffing patterns to produce occupational employment patterns and projections), it would be reasonable to expect the error and uncertainty is compounded. Moreover, if users transform geographical data in other ways (e.g., by aggregating imputed local data to the regional level), this could potentially introduce more error into the aggregated data.

In light of these findings, CTE professionals can take several steps when assessing regional labor market demand in the Agriculture, Food, and Natural Resources career cluster to ensure the well-roundedness of their analysis and insight:

- First, no single data source is perfect, so be aware of the limitations of the data used and the reasons behind those limitations. Gaps in publicly available industry and occupational employment data, while frustrating, also represent a commitment on the part of the issuing agencies to uphold standards of data accuracy and privacy. When not enough data are available to uphold these standards—for example, at the regional or local level where counts of people and organizations tend to be smaller—agencies may opt to withhold rather than present data in which they have low confidence or that sacrifice the privacy of individuals or businesses in the area. Labor market data developed by private vendors may fill in some of the information gaps in publicly available data and are certainly worth consulting if available; however, the data estimation and imputation techniques employed by these private firms may potentially introduce additional error into the estimates. The fact that federal statistical agencies have not, once again, overhauled their data suppression methodologies to thwart the imputation attempts of private vendors provides some indirect evidence that the private data vendors have not been all that successful in replicating undisclosed federal employment data.

- Second, pay attention to the methodological transparency of the data source. Does the source provide enough detail to allow one to ascertain what is measured and how it is measured by providing at least the basic contours of the imputation procedure and information on how the data imputation methods were validated? Additionally, to what degree does the data provider account for differences in the reliability of the various data it provides? For example, one private data vendor mentioned in this report employs an imputation indicator within its data product to help users distinguish the publicly reported data from the privately imputed data.
- Third, given differences in data set characteristics and reliability, consult multiple sources of data and information to develop a more robust and well-rounded assessment of labor market demand.
    - Publicly available data can be supplemented with data from private vendors if available.
    - Occupational data by CTE career cluster (e.g., Trailblazers occupational projections) can be supplemented with employment data by economic sector and industry (e.g., LAPI).
    - Data series designed to cover employment across all industries (e.g., QCEW, SAE, CBP, and LAPI) can be supplemented with industry-specific employment data (e.g., Census of Agriculture).
    - Data that seem sparse at the local level or at more detailed levels of occupations or industries can be supplemented with data for larger geographic areas and broader occupational or industrial groups. For example:
        - Review statewide labor market trends in addition to regional trends.
        - Review the occupational outlook for broad occupations (e.g., Agricultural and Food Scientists-SOC 191010) or minor occupational groups (e.g., Life Scientists-SOC 191000) if data are not available in your geographic area for a specific detailed occupation (e.g., Food Scientists and Technologists-SOC 191012).
        - Review economic sector data at the two-or three-digit NAICS code level (e.g., Food Manufacturing-NAICS 311) if data are not available for a specific industry at the six-digit NAICS code level (e.g., Fluid Milk Manufacturing-NAICS 311511).
    - Data pertaining to the Agriculture, Food, and Natural Resources cluster can be supplemented with data pertaining to other related career clusters.
    - Data pertaining to Virginia or any of its regions can be supplemented with data pertaining to other states or regions.
    - Information from labor market data sets can be supplemented with information from other types of sources, such as academic reports, professional association websites, and expert advice from local business and industry representatives.

- Fourth, new equipment and technology, data applications, and automated processes are quickly transforming production practices and reshaping the demand for labor skills within the Agriculture, Food, and Natural Resources career cluster. Although new occupations are added to occupational classification systems like the SOC over time to reflect evolving occupational skill sets, out of practical necessity, these systemic updates are typically undertaken only once per decade. This pace often does not truly reflect the speed with which new technology transforms the demand for certain occupational skills. **Therefore, in addition to reviewing occupational projections and industry forecasts, identify what job skills are currently in demand <u>across</u> clusters, occupations, and industries to account for the rapid technological change associated with emerging agricultural industries.** Local employers, business and industry representatives, and student High-Quality Work-Based Learning experiences continue to be invaluable resources for staying informed about what occupational skills are currently in greatest demand.

APPENDICES

APPENDIX A: Public Geographical Industry Employment Data Set Features

Data Set Name	Agency	Geographic Units	Reporting Interval	Reporting Lag	Industry Detail	Source	Coverage	URL
Quarterly Census of Employment and Wages (QCEW)	BLS	County, Metro Areas, Workforce Investment Areas, Other*	Quarterly	Six months	NAICS 6-digit	UI administrative records of nearly 11 million establishments	UI-covered population (Coverage for state UI and Unemployment Compensation for Federal Employees (UCFE)). Excludes proprietors and selected other sectors (e.g., military, railroad workers)	https://www.bls.gov/cew/ https://virginiaworks.com/Quarterly-Census-of-Employment-and-Wages-QCEW/index
State and Metro Area Employment (SAE)	BLS	Metro Areas	Monthly	Less than a month	NAICS Sector	Survey of approximately 700,000 establishments	Non-farm wage and salary employment, both jobs covered by UI and some excluded	https://www.bls.gov/sae/
County Business Patterns (CBP)	Census Bureau	Counties, Metro Areas	1st quarter (March)	16 months	NAICS 6-digit	Census Bureau surveys and administrative records	Employment covered by Social Security	https://www.census.gov/programs-surveys/cbp/data/datasets.html
Local Area Personal Income (LAPI)	BEA	County, Metro Areas	Annual	11 months	NAICS 2-digit	Assembled from variety of government data sources, including QCEW	All public and private sector employment, including proprietors and military	https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas

Data Set Name	Agency	Geographic Units	Reporting Interval	Reporting Lag	Industry Detail	Source	Coverage	URL
Occupational Employment and Wage Statistics (OEWS)	BLS	Metro areas and nonmetro residual, Local Workforce Development Areas, Other*	Annual (3-year average)	9 months	SOC 6-digit	3-year rolling survey of QCEW employers	UI-covered employment.	https://www.bls.gov/oes/tables.htm https://virginiaworks.com/Occupational-Employment-Statistics-OES
Occupational Employment Projections**	BLS	Local Workforce Development Areas**	Annual	About 3 years	SOC 6-digit	OEWS and other public data (e.g., Census Bureau CPS)	UI-covered employment with some supplemental adjustments for uncovered employees, including self-employed	https://virginiaworks.com/Occupational-Projections

** Some substate data provided by BLS state affiliate (e.g., Virginia Employment Commission)*

*** Data provided by BLS state affiliate (e.g., Virginia Employment Commission)*

APPENDIX B: Employment Imputation Methods by Research Study

		INFORMATION USED						
Study	**Computational Method**	**Data Series**	**Industry Temporal Data**	**Geographical Hierarchy**	**Industry Hierarchy**	**Industry Employment Range**	**Industry Establishment Size Distribution**	**Other Unsuppressed Industry Employment Data Series**
Cao (2018)	Multiple Imputation	QCEW	X	X	X			
Eckert et al. (2021)	Linear Programming	CBP		X	X	X		
Holan et al. (2008)	Multiple Imputation	QCEW	X	X	X			
IMPLAN (n.d.)	Ad Hoc, Iterative Proportional Fitting	QCEW	X	X	X			X
Isserman and Westervelt (2006)	Simulated Annealing	CBP		X	X	X	X	
Jaromczyk et al. (2021)	Linear Programming	CBP	X	X	X		X	
Zhang and Guldmann (2013)	Linear Programming	CBP	X	X	X	X	X	
Zheng (n.d.)	Machine Learning Algorithm (unspecified), Iterative Proportional Fitting, and Other	QCEW		X	X	X		X

APPENDIX C: Lightcast and IMPLAN Industry and Occupational Employment Data

Lightcast. Lightcast (formerly Emsi Burning Glass) provides local industry and occupational employment data as part of its Analyst software platform. The firm offers several data series, among which are a wage and salary series comparable to the QCEW with suppressed data imputation. A second series adjusts QCEW to include other UI-uncovered wage and salary jobs, and a third and fourth series add self-employed individuals and other proprietors, respectively. The addition of these other categories makes the coverage similar to BEA's LAPI. For the industry employment data, the estimation relies on a proprietary algorithm, which is not provided in methodological description materials.

Lightcast also provides occupation job counts and projections data for localities. The projections differ from BLS projections in several ways, and the firm cautions that the projections are not directly comparable to the ones produced by the BLS because of methodological and other differences. First, the projections are based on Lightcast baseline employment data for expanded employment categories (described above) with industry imputations, whereas the BLS projections are based on covered wage and salary employment that have been supplemented with estimates of some uncovered workers and self-employed individuals. Second, the firm regionalizes national staffing patterns using data imputations that rely on OEWS and QCEW data. Lightcast also uses its own internal projection methods to project forward. Lastly, baseline and projected occupational staffing patterns are applied to industry baseline employment and projections. These methods produce six-digit SOC occupation detail that is usually missing in regional estimates and estimates that are not available at the local level. However, they involve several layers of estimation and imputation that may result in greater uncertainty and error.

IMPLAN. IMPLAN provides regional employment and occupational data as part of its online input-output economic impact software product. In addition, it offers a QCEW database of fully populated, six-digit industry employment data for localities. The data provides disclosure codes indicating whether the value provided was obtained from the BLS or estimated. The company also provides a local industry employment series using an IMPLAN sectoral scheme that provides estimates of wage and salary employment that is adjusted for proprietors, military, and other non-UI-covered sectors. These data are offered in an IMPLAN industry scheme for use in input-output modeling, which in many instances has NAICS six-digit equivalents. The firm's methodological "Frequently Asked Questions" on their website provides some details on the imputation method. Industry imputation for QCEW is performed using a combination of disclosed NAICS employment data from the previous year, CBP industry employment data (when disclosed, for QCEW-suppressed industry employment), and state or U.S. industry data. Estimation of the number of proprietors and non-covered jobs is provided with BEA LAPI data. IMPLAN now also provides regional occupational data as well.

REFERENCES

Advance CTE. n.d. Agriculture, food & natural resources. (Accessed August 17, 2022).

Bureau of Economic Analysis. n.d. Local area personal income and employment methodology. https://www.bea.gov/system/files/methodologies/LAPI-Methodology.pdf (Accessed July 13, 2022).

Cao, Jian. 2018. Multiple imputation methods for large multi-scale data sets with missing or suppressed values. Florida State University. Dissertation submitted to the Department of Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Carpenter, Craig Wesley, Anders Van Sandt, and Scott Loveridge. 2021. Measurement error in US regional economic data. *Journal of Regional Science* 62: 57-80.

Eckert, Fabian, Teresa C. Fort, Peter K. Schott, and Natalie J. Yang. 2021. Imputing values in the US Census Bureau's County Business Patterson. Working Paper 26632. Cambridge, MA: National Bureau of Economic Research.

Hasenstab, Maria. 2019. Nonfarm Payrolls: Why Farmers Aren't Included in Jobs Data. Federal Reserve Bank of St. Louis. https://www.stlouisfed.org/open-vault/2019/july/nonfarm-payrolls-why-farmers-not-included (Accessed August 8, 2022).

Holan, Scott H., Daniell Toth, Marco A.R. Ferreira, and Alan F. Karr. 2008. Bayesian multiple imputation with implications to data confidentiality. Technical Report Number 171. Research Triangle Park, NC: National Institute of Statistical Sciences.

IMPLAN. n.d. https://support.implan.com/hc/en-us/sections/115000320233-From-the-Data-Team. https://support.implan.com/hc/en-us/articles/360059868114-Employment-Wages-by-NAICS-Imputed-CEW-Data- (Accessed July 13, 2022).

Isserman, Andrew M. and James Westervelt. 2006. 1.5 million missing numbers: Overcoming employment suppression in County Business Patterns data. *International Regional Science Review* 29, 3: 311-335.

Jaromczyk, Jerzy, Samantha Cohen, Houtian Ge, Miguel Gomez, Sara Rehkamp, and Jing Yi. 2021. Imputing suppressions in public data: The case of County Business Patterns. Paper presented at the 2021Annual Meeting of the Agricultural and Applied Economics Association, August 1-3, Austin, Texas.

Lightcast (formerly Emsi Burning Glass). n.d. Data methodology. https://kb.emsidata.com/methodology/ (Accessed July 13, 2022).

National Academies of Sciences, Engineering, and Medicine 2021. Creating the future workforce in food, agriculture, and natural resources: Proceedings of a workshop. Washington, DC: The National Academies Press.

Rephann, Terance. 2008. The economic impact of agriculture and forestry on the Commonwealth of Virginia. Charlottesville, VA: Weldon Cooper Center for Public Service, University of Virginia.

U.S. Census Bureau. n.d. County Business Patterns (CBP) methodology. https://www.census.gov/programs-surveys/cbp/technical-documentation/methodology.html (Accessed July 13, 2022).

U.S. Department of Labor, Bureau of Labor Statistics. n.d. *Handbook of methods.* https://www.bls.gov/opub/hom/ (Accessed July 13, 2022).

Zhang, Sumei and Jean-Michel Guldmann. 2013. A regression-constrained optimization approach to estimating suppressed information using time-series data: Application to County Business Patterns 1999-2006. *International Regional Science Review* 38, 2: 119-150.

Zheng, Ping. n.d. Filling in the holes: Generating point estimates for QCEW suppressed data. Indiana Business Research Center at Indiana University's Kelley School of Businesshttps://www.incontext.indiana.edu/2020/jan-feb/article2.asp

ENDNOTES

[i] Virginia Lightcast employment data are used because it better counts agriculture-related and proprietor employment than BLS/VEC employment data as described elsewhere in this report. A national industry-occupational matrix was used because it has less employment data suppression than the VEC Virginia matrix.

[ii] Several U.S. agencies also provide local employment data from other surveys that highlight particular sectors of the economy. For instance, the U.S. Department of Agriculture (USDA) provides figures on farm employment based on a Census of Agriculture conducted every five years. The U.S. Energy Information Administration provides information on underground and surface coal mining employment. The U.S. Census conducts economic surveys for various industries on a five-year schedule that provide employment estimates by detailed industry. The Census Bureau also produces an annual series called Nonemployer Statistics (NES) that provides local counts of businesses that have no paid employees (a measure of self-employed workers). In some instances, information from these alternative federal data sources is used in estimating employment by other agencies, such as the BEA LAPI.

[iii] See, for example, the Pennsylvania LMI (Department of Labor and Industry Center for Workforce Information & Analysis) description of their occupational employment projection statistics. dhttps://www.workstats.dli.pa.gov/Products/ShortTermForecasts/Pages/default.aspx

[iv] Eckart's CBP database is available at: http://fpeckert.me/cbp/. Upjohn Institute's WholeData is found here: https://www.upjohn.org/data-tools/wholedata. WholeData uses an algorithm described in Isserman and Westerveldt (2006).

[v] Eckart et al's (2021) original technique is no longer valid because of changes in CBP suppression methodology. Jaromczyk et al. (2021) published a revised imputation technique that utilizes information about another data constraint in view of the new U.S. Census Bureau employment data suppression rules.

[vi] BLS makes available state occupational staffing patterns for "research purposes," which are subject to substantial suppression.

[vii] A methodological description can be found here: https://kb. Emsidata.com/methodology/occupation-employment-process/

[viii] LightcastTM Analyst 2022. Retrieved July 6, 2022.

[ix] IMPLAN® model, 2020 Data, data provided by IMPLAN Group LLC, IMPLAN System (data and software), 16905 Northcross Dr., Suite 120, Huntersville, NC 28078 www.IMPLAN.com.