LMI Institute Working Paper Series

In collaboration with the Center for Regional Economic Competitiveness

and the Projections Management Partnership

New Data Sources and Opportunities for Forecasting Labor Demand and Skills Mix

March 2023

Henry Renski, PhD
Professor of Regional Planning
University of Massachusetts Amherst

Working Paper 3

Center For Regional Economic Competitiveness
P.O. BOX 12546 | Arlington, VA 22209
T: 703.522.4980 | F: 703.832.8663

Introduction

This paper considers how emergent technologies will impact the demand for certain occupations and the skill, task and knowledge associated with these jobs. Unlike other studies that try to predict future labor demand for certain types of jobs, this paper focuses on measurement and methodological issues. In particular, I consider electronic (i.e. online) postings on job vacancies as a potential source of information that may help improve occupation-based forecasts. I focus on two separate types of forecasts. First, there are the official occupation-based employment projections produced by state Labor Market Information (LMI) offices. I also consider forecasting studies that try to measure how emergent technologies may affect the long-term labor demand for certain occupations and skills. I refer to these as Future of Work studies.

This paper opens with a brief review of the current approach for constructing sub-state occupational employment projections. It then discusses online job posting as a potential source of labor force information, before moving into potential applications and limitations. I conclude that while electronic job postings would be difficult to directly incorporate into the current (industry-derived) occupational projection system, they do have considerable value for documenting trends in the Skills, Tasks and Knowledge (STK) content of occupations. This, in turn, can greatly expand our understanding of occupational transformation that is at the heart of contemporary discussions of the future of work and the workforce impacts of automation and artificial intelligence. I close with some possible avenues for more detailed investigation. First, I use an occupational lens to outline a new method for measuring occupational transformations and changing skill requirements due to AI and automation. Second, I briefly consider the potential value of online job ads to develop region-specific matrices of occupational skills.

Occupational Employment Projections: A Primer

State Labor Market Information (LMI) analysts are responsible for developing projections for employment broken down into detailed occupations. Nevertheless, most states rely heavily on methodologies developed the national U.S. Bureau of Labor Statistics (BLS) for developing their own forecasts.

National projection methods

National employment forecasts follow a multi-step process. The first steps involve developing separate forecasts of the labor force and employment demand that ultimately lead to long-term industry employment projections. Occupational projections are derived from the industry employment projections. More specifically, the BLS uses its occupation by industry national staffing patterns matrix to distribute industry employment into occupations. It then adjusts the occupation shares within industries while considering anticipated trends in technology, business organizational practices, and offshoring.[1] These industry-occupation employment numbers are then aggregated across industries to get employment totals by occupation.

Along with employment change, the BLS also produces projections of occupational separations, commonly referred to as the separations model. Separations are estimates of workers who will leave an occupation because of either *exit* from the labor force (e.g. retirement), as well as separations due to *transfers* that occur when workers change occupations. Estimates of exits and transfers are based on regression models that use a variety of socio-demographic attributes from the Current Population Survey (CPS) to predict the likelihood of exit or transfer by occupation. These rates are then applied to the projected employment change to estimate the number of anticipated openings within an occupation.[2]

State and Sub-state Projections

Each state is responsible for developing its own projections, although most follow the general approach developed by U.S. Bureau of Labor Statistics. To help encourage sharing of information and best practices, the US Department of Labor's Employment and Training Administration (ETA) supports the State Employment Projections Managing Partnership (PMP) program. The PMP serves as a consortium and clearinghouse for state projections knowledge and data. It also provides a software suite containing the analytical models and much of the data that states use to produce employment projections. Like their federal counterparts, the state PMP process typically begins with the production of employment projections by industry, which are then converted into occupations using staffing patterns derived from the Occupational Employment and Wage Statistics (OEWS).[3] Most states seem to have adopted the recently revised BLS process for estimating separations and openings.[4] While they generally follow federal guidance, individual states are permitted to adjusted their industry and occupational employment projects based upon area expertise and local conditions.

---

[1] The BLS Occupational Employment and Wage Statistics (OEWS) survey is the main data source for staffing patterns.

[2] Annual openings are the sum of the projected net change plus transfers plus exits (Openings = Projection Change + Exits + Transfers).

[3] The BLS publishes a staffing patterns matrix for the nation as a whole, as well as for individual states (see https://www.bls.gov/oes/current/oes_research_estimates.htm). The BLS warns uses of limited reliability and large standard errors for many (smaller) states, and it is unclear how many states opt to use state-specific versus national shares for distributing industry employment to occupations.

[4] Presumably due to the limited sample sizes of the CPS, most states seem to favor using national-level regression coefficients of the likelihood of exit and separation by occupation, as opposed to developing their own.

Electronic job boards, job postings and real time LMI

Once upon a time people would turn to the "want ads" in the back of their local newspaper to discover job opportunities. Nowadays, this is all done online. Hiring businesses post job ads on their websites, Linked-In, or other online professional bulletin board service. There are also numerous specialized job boards targeting specific industries, occupations, or regions. These are often operated by national or regional wings of leading professional associations, such as the American Planning Association (APA) or the Northeast Public Power Association (NEPPA). There are also several private companies (Indeed.com, Glassdoor and Careerjet) offering job ad placement services that allow job seekers to review postings for free, but charge businesses to place ads or to place prominently featured premium ads. The National Labor Exchange (https://usnlx.com/) operates a free electronic job board service for employers that post to state workforce development agencies.

It didn't take long for workforce development and education officials to realize the value of the information contained in these postings and soon a cottage industry providing real-time labor market information was born. Two of the major providers of real-time labor market information (Burning Glass Technologies and EMSI) recently merged to form Lightcast – arguably now the leader in this space. Lightcast, and companies like it, usually offer a number of analytical and consulting services, but the heart of their real-time LMI business model is scraping hundreds of thousands of job postings from online sources, removing duplicates and correcting errors, classifying and coding responses to identify skills and occupations, and then re-packaging this information for workforce development agencies, corporate HR personnel, education and training providers, industry and professional association representatives, or anyone else with long-standing questions regarding trends in the labor force.

Uses of electronic job postings for occupational projections

Forecasting from Real-Time LMI

There are two uses of electronic job boards / real-time LMI that can help inform occupational forecasting: (1) tracking job openings/vacancies and (2) analyzing the skill content of occupations. Many LMI offices track and report aggregated data on job vacancies as a measure of occupational demand. For example, the state of California has a Burning Glass (now Lightcast) powered online job vacancy statistics dashboard that reports monthly vacancies for the nation, state, and metro areas in California.[5] They describe vacancies as indicating demand for particular types of occupations. In this sense, real-time LMI is used in similar ways to the current short-term (i.e. two-year) occupational projections offered by most state LMI offices. With a sufficient series of historic jobs vacancies data, one could conceivably develop short-term

[5] https://labormarketinfo.edd.ca.gov/data/help-wanted-online(hwol)/online-job-ads-data.html

forecasts by applying time series regression or similar tend-extrapolation that could distinguish seasonal aberrations from more enduring trends. [6] However, given the natural vagaries involved in coding and counting job descriptions, it might be better to have these forecasts remain as ranges as opposed to formal point estimates of employments or openings.

A real-time LMI approach has several advantages relative to conventional short-term occupational projections. The primary benefit is timeliness. Real-time LMI data is constantly being collected and processed. Therefore, forecasts using this information can be produced quickly and updated in a near automatic fashion as new data comes in. There is often a one-to two-year lag in the production of conventional short-term occupational forecasts. In many cases, by the time the projections are posted they are no longer technically forecasts but estimates of the present or recent past. Real-time LMI is also easier to tailor to specifical to regional labor markets, assuming that jobs posts are accurately tagged to employer locations.

There are also several major downsides to taking a pure real-time LMI approach. First, and most important, analysis of recent historical trends should not be used to develop long-term employment projections 10 to 20 years into the future. A real-time LMI approach based on time-series analysis is pure trend extrapolation and would only be reliable within a short forecast horizon. Data on job vacancies is also unable to distinguish separations due to exits versus separations due to occupational transfers, as does the current BLS separations model that most states use as the basis for short-term forecast.

There are also serious concerns regarding possible sample selection bias in job postings. Real-time LMI presents counts of online job postings. Not all job openings get posted online, and these will be underrepresented. Since most major job boards require payment by employers, it is more likely that low-wage positions may be underrepresented. There are also concerns over whether online postings represent real vacancies, or whether employers post job advertisements continuously as a fishing expedition – collecting resumes in the off-hand chance that something eventually opens (Cappelli, 2012).

Forecast accuracy is another potential concern with the adoption of new approaches and data sources. To be fair, predictive accuracy is a misnomer in the context of long-term projections. Projections will never be accurate predictions of employment numbers in the distant future. There are too many unknown factors. Projections are best understood as possible scenarios of the future under a set of assumptions. From this perspective, understanding how changes in key

[6] Two of the more common regression-based approaches for short-term forecasting include autoregressive integrated moving average (ARIMA) or exponential smoothing time-series analytical models.

assumptions may impact future employment numbers is far more illuminating than predictive accuracy, per se.

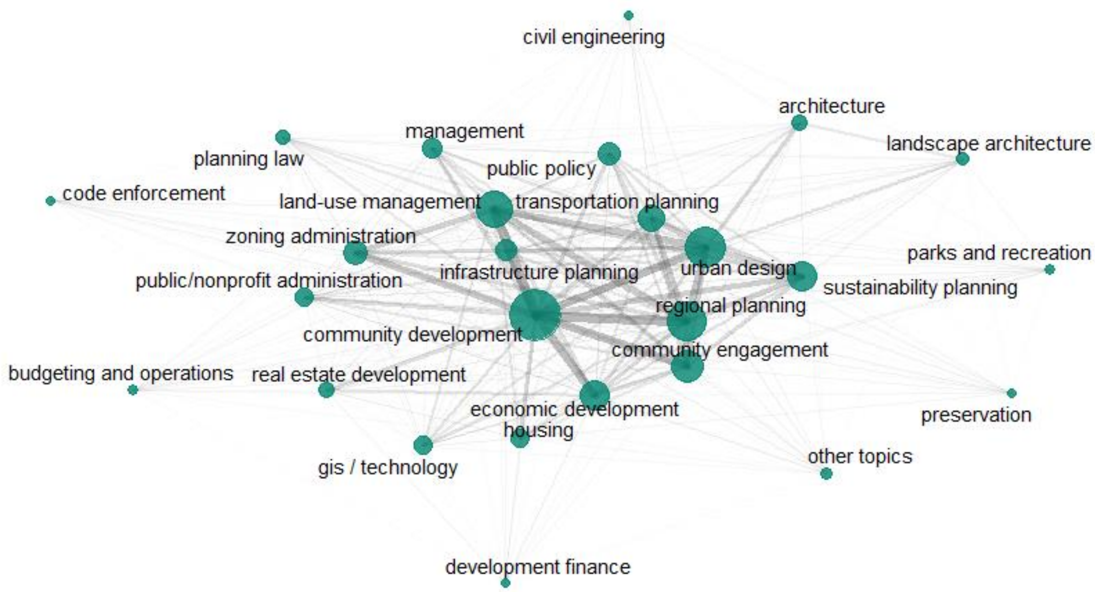
Accuracy can, and should, be evaluated in the development of short-term (two-year) forecasts. At a minimum to assess possible biases. To my knowledge, there has not yet been a widescale systematic analysis/comparison of how real-time LMI based employment trends compare to conventional LMI forecasts, versus actual employment counts by occupation.

Skills content analysis

A second, and perhaps ultimately more valuable, application involves synthesizing descriptive content from job ads to understand the demand for specific occupations as well as the skills, tasks and knowledge components that comprise them. Skills content analysis is a stream of applied research whose development has accelerated with the emergence of machine learning and other data science tools to analyze text-as-data. These methods can be used to extract relevant information from job announcements to create a structured dataset that can be used to measure the frequency of key terms (such as quantitative skills, or designed knowledge of software platforms) or used to identify associations among terms that are regularly co-mentioned in position descriptions. This allows investigators to develop a deeper understanding of the relationships between skills and occupations and among occupations with similar skill composition.

There are numerous examples of content analyses of job posting for a variety of occupations. For example, Renski, Barchers and Greene (forthcoming) recently analyzed online job postings for the City Planning profession based on data from the Planetizen online job board. In addition to tracking general trends across time and space, their study used correlational and network analytical methods to measure the interrelations among different forms of specialized planning knowledge (Figure 1). They found that community development, urban design, regional planning, and land-use management were the commonly mentioned forms of specialized knowledge and the central based on common co-citation with other areas. In short, while only a modest number of job ads specifically asked for community development planners, must planning jobs were looking for people with some background in community development.

Figure 1: A Network Map of Specialized Planning Knowledge



Source: Renski, Barchers and Green (forthcoming)

Another, more extensive, illustrative example is the jobs genome mapping project conducted by Burning Glass Technologies. Burning Glass (now Lightcast) used its massive collection of hundreds of millions of job ads to develop a comprehensive taxonomy of over 32,000 specific skills grouped into over 30 skill clusters (Figure 2). Skills were also classified by the relationship to each occupation: whether they were "defining" skills needed for performing day-to-day tasks, "distinguishing" skills pertaining to technical proficiencies, or the "necessary" skills that are prerequisites to each occupation's defining skills. This information has been used to provide deep skill profiles for each occupation, establish career pathways, and skill-based projections. However, since this data is proprietary, its use is restricted and its methods for de-duplication, coding, classification, and aggregation are somewhat opaque, with no independent evaluation of its accuracy or possible biases.

Figure 2: General skill categories identified by Lightcast

Source: https://lightcast.io/open-skills/categories

I've already mentioned that the risk of bias and selectivity is very much a concern when using job posting data to represent the present or future demand for occupations. Because of this, job postings are better considered a potential complement to, rather than pure substitute for, existing projection methods. A more technical concern is the possible miscoding of descriptive content in job postings. Classifying text as data typically involves first imposing some structure on the data. Job postings tend to follow a general cadence. There is typically a job title, the name of the employer, employer description, position description, and a list of qualifications. In this sense, job postings are easier to code and analyze than other forms of unstructured text (such as twitter posts). However, this structure differs from one job board to another, and the terminology used to describe job duties can also vary considerably from one employer to the next.

There are also questions as to whether the classifications that arise from job descriptions tell us anything meaningful or useful. The occupations, skills, and tasks defined by the BLS and used for staffing patterns or O*Net are very specifically defined and detailed. Job postings, by contrast, are not written with the LMI system in mind. Employers do not always use job titles that are easily matched to standard occupational titles and HR departments may be mistaken about what their workers actually do while on the job. The ads themselves are often purposely vague in their descriptions of job duties, desired skills, and experience expectations so as not to deter prospective candidates. The skill classifications that arise from job posts are often mush--making it difficult to identify shifting trends in skill requirements or the emergence of new task and skill combinations associated with the emergence of new occupations. One could make up for such vagaries with a larger corpus of ads. However, this could itself be a source of bias if detailed skill descriptions are based on a limited subset of only the most technical and specialized positions.

Skills, Tasks and Knowledge and the Future of Work

To our knowledge, content information from job postings has yet to be directly incorporated into the production of official sub-national forecasts. However, skills information from online job postings is already used widely throughout the workforce development and labor market information system, and thus *indirectly* makes its way into projections. For example, among Lightcast's clients are the U.S. BLS who use job postings data to identify missing elements and emergent skills in the O*Net system.

O*Net was originally developed as a more dynamic, web-based, replacement to the Dictionary of Occupational Titles (DOT) that was used to help job seekers and workforce development professionals identify jobs potentially suitable for different backgrounds and interests. O*Net has long since transcendence its predecessor its range of application. Of relevance to this study, O*Net classifies all occupations based upon their skills, tasks and knowledge (STK) requirements drawing a range of surveys, expert coding, job descriptions, and other sources. Because of its unique position as the primary source of public information on the STK content of occupations, O*Net has become the go-to source for understanding the similarities and inter-connections among occupations. This naturally leads to the development of career 'ladders' and 'lattices' that can help workers identify the skills and training needed to a lower-paying to a high-paying jobs (Nelson & Wolf-Powers, 2010) or large groups of complimentary occupations organized into 'clusters' that are used to target workforce and economic development resources to benefit the greatest mass of workers or businesses (Feser, 2003; Renski et al., 2007).

O*Net is at the heart of many recent studies on the Future of Work (FoW). Among the most well-known applications are the studies of Autor, Levy and Murnane (2003) who used DOT data and eventually O*Net to categorize occupational skills by whether routine/non-routine, manual/cognitive. Building upon this framework, Autor, Katz and Kearney (2010) and Goos and Manning (2007) show the emergent polarization of developed economies into "lovely" and "lousy" jobs, with a disappearing class of middle skilled jobs. Frey and Osborne (2017) use O*NET skills data to rank occupations by their susceptibility to automation. This paper produced the stunning conclusion that nearly 47% of all occupations were at risk of elimination due to automation, fueling much of the pre-pandemic fervor in the popular media about widespread job losses. More recently, Steven Hynes working in conjunction with the LMI institute adapt the Acemoglu and Autor (2011) framework to develop an updated automation exposure index.[7] This index is based on the rankings of sixteen O*Net characteristics ranging from the abstract analytical (e.g. creative thinking) to the routine manual (controlling machines and processes).

[7] https://www.lmiontheweb.org/automation-exposure-score/

The revised framework is supposed to better account for the potential of AI to substitute for knowledge work, instead of just traditional production.

Applying an organizational lens to understand the Future of Work

FoW studies are forecasts, even if many don't appear to be on the surface. Like all forecasts, FoW studies suffer from many limitations and possible biases. In the case of future technology, it is far easier to identify jobs that may be destroyed than it is to predict how existing occupations will be fundamentally transformed or even how new jobs might be created. The most noteworthy exceptions are deep-dive / crystal-ball style approaches develop different scenarios for the development of a certain technology (fully driverless trucks vs. platooning with one driver at the helm) and imagines how the ultimate deployment of such technologies may generate new opportunities throughout entire industries or for society as a whole (see Viscelli, 2016). The problem with industry-focused studies is their limited scope. While there may be some transferrable insights or research approaches, there are just too many industries, too many possible scenarios, and too much nuance for this approach to become the foundation for economy-wide projections.

While the present STK framework is generally adequate for classifying skills or tasks by their susceptibility to automation or offshoring, it does not account for how technology may lead to a recombination of tasks and skills across occupations. Perhaps the greatest revelation of the LMI automation exposure index is that far more occupations exist in the middle of the automatable distribution than at the poles. While anticipated automation will clearly impact wide swaths of the labor force, the ultimate effect will likely be more of occupational transformation than wholesale elimination. These new combinations may lead to the 'creation' of new occupational titles, or simply redefining STK content among existing occupations. In short, the name of the job stays the same while the nature of the job changes.

An organizational lens may hold some promise for quantitatively identifying likely pathways of occupational transformation – at least in the short and mid-term. It is important to remember that the allocation of work tasks are *decisions* made within organizations for the purpose of organizing labor for the provision of some good or service. Automation and other applications of AI will surely eliminate some tasks, but not others. It is up to management to decide how to effectively deploy automation as well as how to (re)organize the workforce for tasks that cannot or will not be automated.

In order to identify which occupations will be likely recombined, a good starting point may be identifying occupations that are common to particular organizations (e.g. companies or employers), comparing the STK content of those occupations, figuring out which are most susceptible to automation, and then re-allocating or combining the remaining tasks to form new occupations. The specific details of this approach still need to be developed. However, it seems that this can largely be done using existing LMI data sources and existing analytical methods. For example, O*Net can help identify the existing task and skill mix of occupations. Data synthesized from online job boards can identify trends and emergent skills. Skill and task-based automation indices can help identify the specific skills and tasks that are most likely to be automated or, conversely, the most resilient to automation. To identify the typical occupational mix of employers, it might be possible to develop an organization-based staffing patterns matrix through an analysis of state Unemployment Insurance (UI) records or perhaps microdata from the Occupational Employment Survey (OEWS). This data is confidential and therefore would need to be accessed from behind LMI firewalls or through other special access arrangements. In its absence, the existing occupation by industry staffing patterns series produced by the BLS could serve as a suitable proxy. The final step involves creating new occupational combinations, assuming that certain tasks will be automated and that the remaining tasks will be redistributed among the remaining work force. Conceptually, it stands to reason that the most likely new combinations will be among occupations where (non-automatable) skills, tasks and knowledge are most similar and grouped using conventional similarity/dissimilarity metrics and mathematical clustering methods.

Regionalizing the STK Occupation Content Model

A final potential application involves using online job descriptions to regionalize the O*NET STK content model. This might lead to more accurate regional forecasts with more specific contextualization. O*Net is essentially a national sample of occupational characteristics. They use job data from the NLx to show advertised vacancies down to the zip code level, but the STK content model itself assumes regional homogeneity within occupations. This may be a reasonable assumption for many occupations, especially those with more ubiquitous skills and well-defined tasks such as cashiers. However, this might be less likely to hold for workers with more specialized skills sets or highly variable requirements. STK data culled from electronic job descriptions might be able to help us identify such differences.

Conclusions

The current approach for developing long-term occupation employment forecasts has remained largely unchanged for the past several decades. This approach derives long-term occupation forecasts from industry employment projections using ratios taken from the industry staffing patterns matrix. In doing so, it treats occupations as benign entities whose rise and fall are purely the product of changes in the industries that they belong to. While simple to implement and literally guaranteed to produce consistent employment estimates with industry-based forecasts, this approach fails to recognize a possible revolution in the future or work brought about by continued automation, AI, and other forms of technological change. Industry is not the correct lens. We expect AI to transform jobs based upon what the workers do, regardless of what their employer actually makes.

Recent studies of the future of work (FOW) take a more task/skills based perspective. A growing number of studies from labor economics, sociology, industries studies and other disciplines model occupations as an inter-related set of work tasks with associated skill requirements. To understand how AI and other technological advances may affect the future workforce, they typically identify which component skills / tasks are most susceptible to automation. They then identify potentially at-risk occupations by summing up the risk level of their constituent skills. They build forecasts of occupational change from the ground up, whereas formal (industry-based) projection methods derive them from the top-down. However, these FOW studies typically do not develop numeric estimates of future employment levels by occupation. Nor do they legitimately account for other forms of occupational transformation. The authors (or the journalists popularizing their work) tend to equate a high level of AI exposure to a lost future job. And while some jobs will certainly no longer be necessary, a more likely scenario is that new technology will substitute for certain tasks, but not others. It may also create the need for new tasks, or entirely new occupations. It is the role of organizations and their managers to determine how to recombine tasks to carry out necessary functions, and therefore taking an organizational lens may help identify new occupational combinations.

Given the inherent uncertainty of developing forecasts in an environment of rapid technological change, it makes more sense to project occupations based upon future scenarios of the demand for skills and tasks under different technology regimes. Nevertheless, we are still a long way from integrating FOW perspectives into the official production of occupation employment projections. It would likely require the development of a totally new approach—one based on adding up occupational employment estimates from forecasts of demand for particular skills and tasks. New data sources, such as codified electronic job advertisements, hold some promise in this regard. Job ads are already scraped, coded, and routinely tracked to measure recent trends in occupational demand by state LMI offices. But while possibly well-suited for short-term projections of job openings and to measure the skill content of occupations, they do not cover incumbent employment and are less suited for long-term projections. We still lack a reliable historical time series data, as well as a verified crosswalk for aggregating skills/tasks into occupations. Furthermore, job ads are typically vague in their descriptions of some types of skills (e.g. communications skills) while highly detailed in others (e.g. must know Adobe Photoshop). They are also inherently unrepresentative without clear documentation or understanding of their biases. Resolving these fundamental concerns must take precedence prior to developing skills-based occupational employment projections.

Works Cited

Acemoglu, D., & Autor, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In D. Card & O. Ashenfelter (Eds.), *Handbook of Labor Economics* (Vol. 4, pp. 1043–1171). Elsevier. https://doi.org/10.1016/S0169-7218(11)02410-5

Autor, D. H. (2010). *The Polarization of Job Opportunities in the U.S. Labor Market: Implications for Employment and Earnings*. Center for American Progress.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*(4), 1279–1333.

Cappelli, P. (2012). *Why good people can't get jobs: The skills gap and what companies can do about it*. Wharton Digital Press.

Feser, E. J. (2003). What regions do rather than make: A proposed set of knowledge-based occupation clusters. *Urban Studies*, *40*(10), 1937–1958.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Goos, M., & Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in Britain. *Review of Economics and Statistics*, *89*(1), 118–133.

Nelson, M., & Wolf-Powers, L. (2010). Chains and Ladders: Exploring the Opportunities for Workforce Development and Poverty Reduction in the Hospital Sector. *Economic Development Quarterly*, *24*(1), 33–44. https://doi.org/10.1177/0891242409347721

Renski, H. C., Koo, J., & Feser, E. J. (2007). Differences in labor versus value chain industry clusters: An empirical investigation. *Growth and Change*, *38*(3), 364–395.

Viscelli, S. (2016). *The Big Rig: Trucking and the Decline of the American Dream*. Univ of California Press.