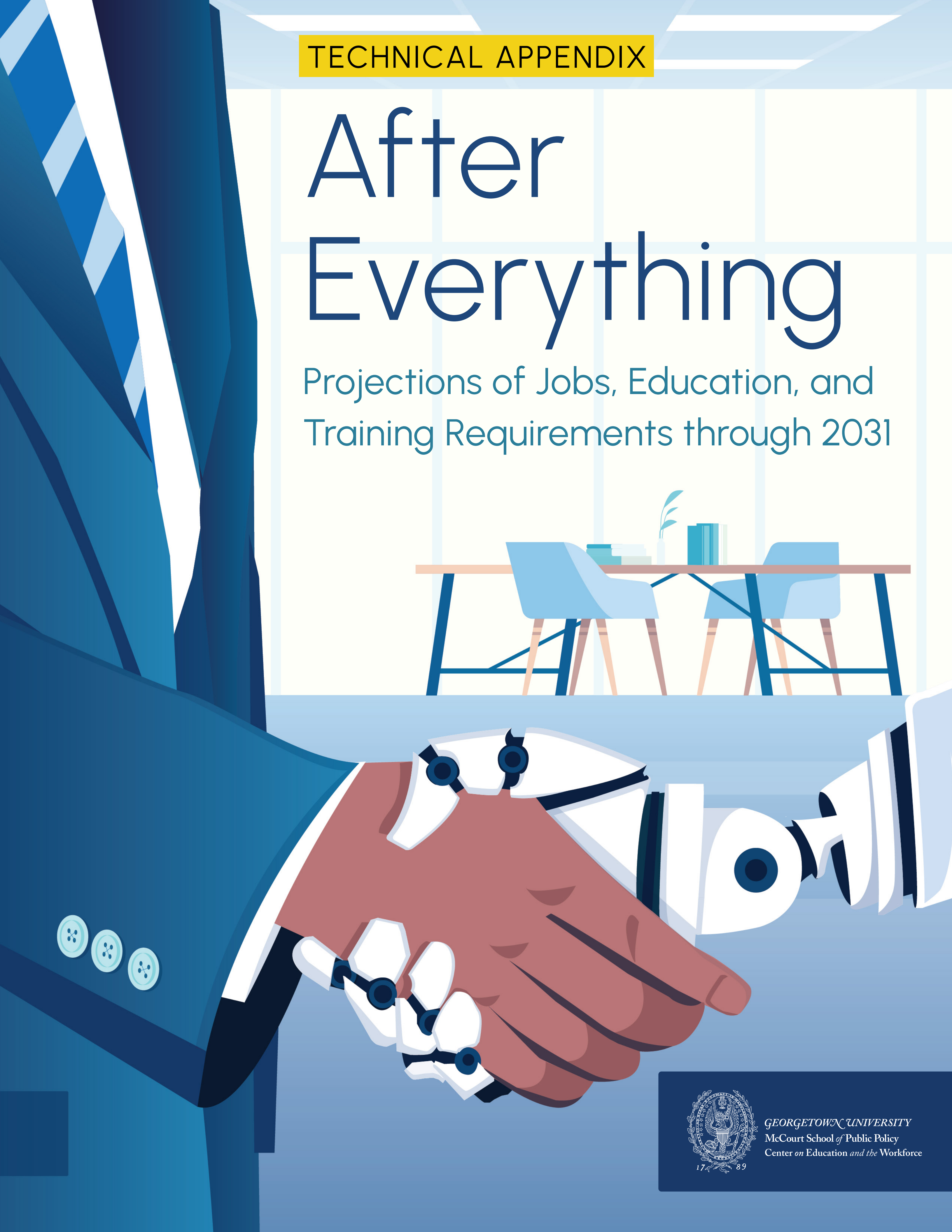


TECHNICAL APPENDIX

After Everything

Projections of Jobs, Education, and Training Requirements through 2031

GEORGETOWN UNIVERSITY
McCourt School *of* Public Policy
Center *on* Education *and the* Workforce

Table of Contents

Building Capacity for Projecting Educational Demand

This appendix documents the methodology used by the Georgetown University Center on Education and the Workforce to project educational demand within the US economy.

Our methodology produces forecasts using data from two private analytics companies. We use occupational forecasts provided by Lightcast that are calibrated to total employment forecasts from IHS Markit (Figure 1). We then feed the data into a model that we created more than a decade ago and have repeatedly refined. In the model, we use gross domestic product (GDP) and employment projections from IHS Markit as feedstock for an input-output (I/O) model developed by Lightcast.[1]

We tested the robustness of this modeling procedure using several methods:

- Evaluation of model fit: We compare the root mean squared errors (RMSE) and the coefficient of variation between models to monitor the scope of outliers.
- In-sample forecasting: We estimate the results of our model using only a portion of the sample, and then using the result to predict outcomes for the remainder of the sample in order to test the extent to which the model accurately predicts known events. In addition, we judge the extent of the variation between observed and predicted outcomes for several different lag lengths in the forecast period.
- Comparison with alternative approaches: We forecast educational demand using a Markov transition probabilities process and compare it to our time-series approach.

Our procedure for modeling of educational demand has several advantages over more static approaches that only provide information for a base year and a projection year. That is because the model includes

- changes in the occupational distribution;
- disaggregation of education clusters;
- changes in the education distribution across occupations;
- lessons from earlier macroeconomic shocks and estimates of those shocks on national job creation; and
- annualized forecasts beyond a base and a projection year.

Measuring Educational Demand: A Four-Step Method

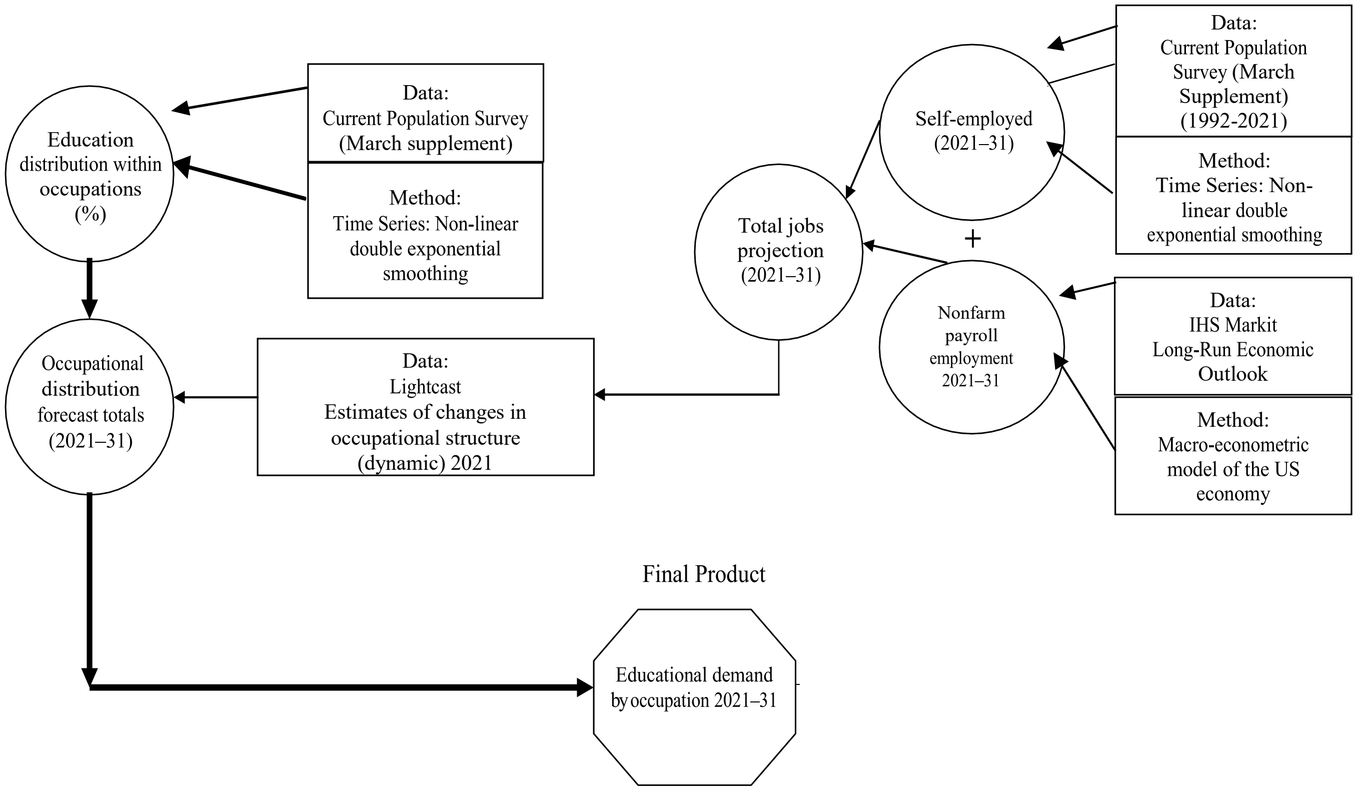
We have a four-step approach to forecasting educational demand:

1. Forecast educational distributions within occupations
2. Estimate long-term employment projections
3. Estimate change in the occupational structure
4. Project educational demand through 2031

Figure 1 demonstrates a four-step model to forecast educational demand.

[1] The Lightcast input/output model uses government sources to produce detailed industry and occupational employment data adjusted according to the most current and detailed labor-market information.

Figure 1: Process for Projecting Demand



Step One: Forecast Educational Distributions within Occupations

The first step in our projections process is to forecast changes in educational attainment within occupational groups using a time-series method. To understand the dynamic change in education requirements for occupations, we use time-series data from the Current Population Survey (CPS) March Supplement to define the proportion of persons with a particular education level within occupations. The proportion of educational requirements observed for each occupation level is then forecast 10 years into the future. We estimate the proportion of workers within occupations by eight educational attainment levels:

1. Less than high school
2. High school diploma
3. Some college, but no degree
4. Associate's degree
5. Bachelor's degree
6. Master's degree
7. Professional degree
8. Doctoral degree

Using the CPS data, we then establish trends since 1992 (when level of educational attainment was first added to the survey) for each of these educational attainment levels within 22 occupational groups drawn from the Bureau of Labor Statistics' Standard Occupational Classification (SOC) system.[2] These include the following occupational groupings:

1. Management
2. Business and financial operations
3. Computer and mathematical sciences
4. Architecture and engineering
5. Life, physical, and social sciences
6. Community and social services
7. Legal
8. Education, training, and library
9. Arts, design, entertainment, sports, and media
10. Healthcare practitioners and technical
11. Healthcare support
12. Protective services
13. Food preparation and serving
14. Building and grounds cleaning and maintenance
15. Personal care and services
16. Sales and related
17. Office and administrative support
18. Farming, fishing, and forestry
19. Construction and extraction
20. Installation, maintenance, and repair
21. Production
22. Transportation and material moving

The March CPS Supplement is a nationally representative, cross-sectional data set that provides information on the socioeconomic characteristics of the American population. It presents detailed information on demographic characteristics and labor-market behavior for about 50,000 households. We have chosen to use the CPS instead of the much larger American Community Survey (ACS) due to the longer history of the CPS data set.[3] Our time-series methodological framework benefits from our use of the longest-term data set possible with information pertaining to the educational and occupational characteristics of the population.[4] The relatively lengthy period of observations in the CPS data set also

[2] Changes in the Standard Occupational Classification (SOC) codes during the time period were bridged using a crosswalk developed at Westat, Inc.

[3] The Current Population Survey (CPS) has been conducted by the US Census Bureau and US Bureau of Labor Statistics for more than 50 years (although only a shorter segment is available to us due to definitional changes). The US Census Bureau's American Community Survey (ACS) is less than 30 years old. It was first conducted in 1996 in a subsample of US counties. The US Bureau of Labor Statistics has only recently transitioned to the ACS as the standard from which to derive educational clusters.

[4] The Current Population Survey (CPS) fulfills this requirement, although the authors recognize that sample-size bias might

makes it easier to demonstrate skill-biased technological change within occupations—that is, the changes that occur as the proportion of highly-skilled workers within an occupation increases with time.

The March CPS Supplement details the highest education level attained and the occupation of survey respondents. We use data on the weighted percentage of workers with a particular education level employed in a specific occupation to estimate the realized demand for education within that occupation.

We used a probabilistic crosswalk developed at Westat, Inc., to bridge the extensive changes to the Standard Occupational Classification (SOC) codes that occurred in 2002. This crosswalk relies on probabilities derived from the CPS practice of double-coding occupations over three years to provide a means of empirical comparison between the two systems.[5]

We assume that each of the time-series variables in the model represents one observation of an underlying data-generating process. We also assume that this process consists of both stochastic and deterministic series.[6] Our data set includes 39 initial observations (one for each year from 1983 to 2021).[7,8]

We calculate the percentage change in the educational distribution within occupations over time. Our objective is to find an economic model that is parsimonious, plausible, and informative, and that best represents past information to generate conceivable forecasts of educational demand within occupations. To conduct these forecasts, we employ nonlinear exponential smoothing, with the restriction that the estimated proportions for each education level sum to one for each of the years in the forecast horizon.[9]

necessitate pooling of data from some years for comparisons by state or other smaller jurisdictions.

[5] These data can be found in the US Census Bureau's Technical Paper 65, "The Relationship between the 1990 Census and Census 2000 Industry and Occupational Classification Systems," 2003. The program for this crosswalk is available upon request.

[6] The stochastic random process must be modeled.

[7] Due to a change in the definition of educational attainment in the Current Population Survey (CPS) data set starting in 1992, our sample size is reduced to allow for the greater degree of specificity in the definitions of educational attainment.

[8] Over the years, there have been several changes in Current Population Survey (CPS) methodology. In 1992, the education variable started including more details on the level of degree, so we have estimated the time series model starting from 1992, as defined by the data. We do, however, include information on educational demand from 1983 to obtain a longer history of the changing education structure, while recognizing that some of the details on level of degree that could be observed in 1992 were not available in 1983.

[9] Exponential smoothing is a time-series method in which past observations of a series are used to forecast the future. It is a variant of a moving-average process that places relatively greater emphasis on the most recent past and includes information on the time trend in the data.

Using Exponential Smoothing to Measure Educational Demand

We first use a non-linear exponential smoothing technique to estimate the educational distributions across time. We chose a non-linear smoothing technique because of its simplicity in design and application, its desirable small-sample properties, and the exponentially declining emphasis it places on more distant observations.

When applying exponential smoothing, each smoothed term is a weighted average of the current observation and the smoothed value of the previous observation (Equation 1). From a forecasting perspective, this method is reliable, easy to apply, and functional. This framework has the advantage of allowing us to produce dynamic out-of-sample forecasts with minimal in-sample sum of squared prediction errors.

$$S_t = a\, y_t + (1 - a)S_{t-l} \qquad \text{(Equation 1)}$$

l= lag length

y_t = the raw data sequence beginning in time t=0

$0 \leq a \leq 1$ where a is the dampening parameter/smoothing constant.

Exponential smoothing attaches greater significance to the most recent observations, with an exponential decline in the contribution of older observations.[10] This aligns with our belief that recent events should be the more important predictors of future values while taking longer-term trends into account. The rate at which the weights of older observations decline is determined by the size of the smoothing constant selected. The closer this value is to 1, the less significant are older observations. With noisy data, smaller values should be used for the smoothing constant.

Where the data exhibit trend or cyclical characteristics, a double exponential smoothing method is more appropriate.[11] In these cases, the exponential smoothing model contains two weighing factors, α and γ, and is defined by two equations (Equation 2 and Equation 3) that account for changing trends. Specifically,

$$S_t = \alpha\, y_t + (1-\alpha)(S_{t-l} + b_{t-1}), \qquad 0 \leq \alpha \leq 1 \qquad \text{(Equation 2)}$$

$$b_t = \gamma(S_t - S_{t-l}) + (1-\gamma)\, b_{t-1} \qquad 0 \leq \gamma \leq 1 \qquad \text{(Equation 3)}$$

Again, S_t represents the smoothed values, b_t is an estimate of the trend, and y_t is the raw data sequence of observations. The dampening parameters, α and γ, are determined by the extent to which we want to emphasize the contributions of the most recent and oldest observations. The closer the dampening parameters are to 1, the greater the relative role of more recent observations in determining the future forecast values.

[10] The typical econometric procedure involves decomposing the series into a trend with seasonal and irregular components. ARMA, Box-Jenkins framework, and (G)ARCH modeling are also generally accepted univariate methods.

[11] A Holt-Winters smoothing technique (triple exponential smoothing) may be used where the data display seasonality, but this method is more appropriate for monthly or quarterly data.

Step Two: Estimate Long-Term Employment Projections

Occupational growth is closely connected to the growth of the entire economy through industry productivity and overall economic expansion. In step two, we purchase data on the macroeconomic outlook from IHS Markit. Using a model of the US economy that includes 1,500 separate equations, IHS Markit forecasts unemployment rates, inflation rates, gross domestic product (GDP), and overall nonfarm payroll employment.

We generate projections of the demand for education in the US economy by incorporating assumptions about the patterns in job growth and job creation within the larger macro economy. Using this data and IHS Markit's model, we produce forecasts of job creation through 2031.[12]

Total jobs in this model consist of nonfarm payroll employment plus solo proprietors and self-employed workers.[13] We obtain the employment totals for solo proprietors and self-employed workers from government sources.[14]

Step Three: Estimate Change in the Occupational Structure

We partner with Lightcast to provide forecasts of employment by occupation, adjusted for current and projected industry job gains and losses, and calibrated to national forecasts of job decline and job growth through 2031. Data are purchased from Lightcast, a private company that procures occupational data by SOC code from various government sources. Utilizing a proprietary occupational taxonomy, we obtain six-digit SOC detail on the size of occupations over a 10-year projected time horizon at both the national and state levels.

We obtain estimates of changes in occupational distribution over time from Lightcast.[15,16] In our modeling process, all of Lightcast's forecasts in step 3 conform to IHS Markit's totals obtained in step 2.

---

[12] IHS Markit uses a macro-econometric model of the US economy that incorporates information on the business cycles into final estimates of national and state levels of occupational demand.

[13] Total jobs in general consist of nonfarm payroll employment, solo proprietors, the self-employed, unpaid family workers, agricultural employees, and paid private household workers. Unpaid family workers, agricultural employees, and paid private household workers are excluded from our calculations of total employment.

[14] Projections of self-employment (unincorporated) are appended to nonfarm payroll employment to obtain forecasts of total job creation in the economy. Proprietor ownership and the educational demand associated with this subset of workers are estimated separately using data from the US Bureau of Economic Analysis (BEA) on non-farm and farm proprietors. These values include both incorporated and unincorporated self-employed workers and are based on IRS tax data that reflect the address from which the proprietor's individual tax return is filed (usually the proprietor's residence).

[15] Structural change in the US economy, including recent substantial reductions in manufacturing and retail employment, can have a substantial impact on the occupational mix. The Quarterly Census of Employment and Wages (QCEW) and the monthly Current Employment Statistics (CES) surveys, both from the US Bureau of Labor Statistics, provide information from an industry perspective.

[16] Lightcast combines data, updated on a quarterly basis, from more than 80 government and private-sector sources. These data capture occupational growth trends and information about skill-biased technological change. Total employment is subdivided into non-farm payroll employment and self-employed workers. The former is derived from the Quarterly Census of Employment and Wages (QCEW) and reflects the occupational distribution of the Occupational Employment and Wage Statistics (OEWS) surveys. The latter are derived from the Current Population Survey (CPS) and reflect the occupational distribution of the CPS and the American Community Survey (ACS).

Step Four: Project Educational Demand through 2031

In this final step, we build on the three previous steps to estimate the number of jobs within each occupation at each education level. We later sum each education level across occupations to get an estimate of national educational demand.

Our approach to estimating educational demand has many positive attributes:

- **It allows for possible change in the educational distribution within occupations**. Historically, the assumption of a fixed distribution of education within occupations has consistently underestimated the demand for higher education. Such an assumption ignores information about trends that should be used to improve projections of educational demand. We use the actual education characteristics of American workers in our calculations and make no assumptions regarding entry-level requirements. In fact, entry-level requirements for today's jobs are almost universally higher than entry-level requirements in the past.
- **It allows for possible changes in the occupational distribution.** We assume that structural changes in the macroeconomy affect the occupational distribution of jobs in the US economy. For example, long-term reductions in the manufacturing industry are reflected in reductions in occupations that are unique to or dominant within that industry. By accounting for changes in the occupational distribution across the economy, we allow for structural changes in occupational staffing ratios.[17]
- **It accounts for the effects of macroeconomic shocks and business cycles.** While adhering to general, long-term full employment, we allow for short-term fluctuations and departures from the steady state as reflected in boom-and-bust cycles, including recessions.
- **It allows for annual forecasts.** Our approach allows us to estimate the progression in educational demand for every year of the 10-year forecast instead of only at the beginning and end of the forecast horizon.

Robustness Testing

We have defined three procedures to test for the robustness or stability of our model, where robustness is defined as the degree to which a model performs in a predictable fashion.

Procedure One: In-Sample Model Performance

We estimate the model equation for a portion of the available data, then compare these data to the observed values for that time period to test the predictive success of our model's estimates. We perform this calculation first in a recursive fashion, assuming that our model's accuracy is an increasing function of the number of data points and inversely related to the length of the forecast horizon. We use the root mean squared error (RMSE) to calculate the extent of differences between the actual and fitted values. The RMSE is a good measure of the extent to which the model predicts the response, where lower values of RMSE indicate better fit in a nested context.[18] We find that the model fit improves with

[17] The changing mix of workers within each industry over time will also influence the educational requirements of the occupations within that industry.

[18] The RMSE is the square root of the variance of the residuals. It is therefore a comparison of observed data points to

smaller estimates for the RMSE as the sample size increases. This is in keeping with our assumptions about model predictability. As expected, the RMSE increases in and around recession years as these periods are generally more difficult to predict. But our model estimates tend to smooth variables along the time horizon and, as a result, even out variability.

Procedure Two: Standard Coefficient Testing

Calculating the statistical significance of the smoothing parameters and transition probabilities is a standard part of our robustness tests. Additionally, we test for coefficient stability in a model framework. To facilitate comparison among equations, we calculate the coefficient of variation within each model. The model RMSE and mean of the predicted variable are both expressed in the same unit, so taking their ratio makes the resultant statistic unit-neutral. Similar to the above, we find coefficient parameters are statistically significant, with little variability.

Procedure Three: Stability of Estimates Derived from Alternative Model Assumptions and Processes

We analyze step one of the estimation procedure from two alternative perspectives, but come to similar conclusions. Both the simplified time series exponential smoothing approach and the Markov transition probability matrices model produce estimates of the proportion of education level within occupations that are highly plausible and similar in size.

---

the model's predicted values. The mean absolute percentage error (MAPE) is also a popular method used to evaluate forecasts from simple models such as these.

State Projections

Once national projections were determined, we forecast educational demand for all 50 states and the District of Columbia that connect to the national totals. We utilize the input/output model developed by Lightcast and various government sources to produce detailed industry and occupational employment data adjusted according to the most current and detailed labor-market information. Our state totals include nonfarm payroll jobs only, while our national totals include both nonfarm payroll jobs and self-employed workers.[19]

Educational demand is obtained using distributions from both the ACS data (for state distributions) and CPS data (for national distributions) within occupations.

1. We forecast the educational distribution within each occupation for the nation from 2021 through 2031 with CPS data using an exponential smoothing process.
2. We separately define the educational distribution within each occupation for each state and the District of Columbia in 2021.
3. National growth rates of education demand within occupations through 2031 are applied to state occupation forecasts to obtain the forecasted educational distribution within each occupation for each state in 2031.
4. Each state has a unique forecast of education demand since its forecast is based on current education demand distributions that are then advanced by a standard national rate of growth within occupations.

Annual Job Openings

Job openings discussed in this report are an annualized flow of job openings over a 10-year time horizon. Some job openings result from a newly created opportunity. Others result when an individual permanently leaves a job—whether for personal reasons, retirement, injury, or even death. To project openings, the US Bureau of Labor Statistics (BLS) estimates the fraction of separations that occur when workers permanently leave an occupation or the workforce, as opposed to separations that occur when workers move to other job locations or get a promotion. The BLS estimate of openings does not count people who change jobs but stay in the same occupational cluster. The BLS website contains a technical description of the regression analysis and assumptions the agency used to estimate permanent labor force exits and job movement.[20] We agree with the new BLS approach and have adopted it as part of this report. Our projections for total jobs openings and net new jobs do not always align with those of BLS, however.

The current separations methodology represents a significant change in how BLS estimates replacement jobs. Before the 2016 to 2026 projections, it used a cohort-component method to determine job separations. BLS discontinued this method after identifying several statistical and conceptual issues. In actuality, the forecasts produced using the previous method (illustrated by the last four rows of Table 1)

[19] Nonfarm payroll jobs account for approximately 80 percent of the workers who contribute to Gross Domestic Product (GDP) in the US. Total jobs also include proprietors, private household employees, unpaid volunteers, farm employees, and the unincorporated self-employed, workers who are not employed by a corporation or other entity. The number of self-employed workers was estimated for the national report but the data were not considered accurate enough at the state level to provide credible long-term projections, so they are not included in the state job projections.

[20] US Bureau of Labor Statistics, "Occupational Separations Methodology," 2022.

resulted in the under-representation of the number of job openings, because of a disproportionate representation of churn in existing jobs rather than new job openings.

In practical terms, the change in methodology results in higher annual average numbers for job openings based on retirements, separations, and new growth compared to the past, but lower estimates compared to the Job Openings Labor Turnover Survey (JOLTS). This means we are not making an apples-to-apples comparison when we look at estimates from previous years, as shown by the following table. Note, in particular, the significant changes in job replacements and total job openings.

Table 1. Comparisons of job forecasts from the US Bureau of Labor Statistics

	Net new jobs, 10-year totals (in millions)	Job replacements, 10-year totals (in millions)	Total job openings, 10-year totals (in millions)	Occupational openings, annual average (in millions)
2021-2031	8.3			19.5
2020-2030	11.9			18.5
2019-2029	6			17.5
2018-2028	8.4			19.7
2016–2026	11.5			18.7
2014–2024	9.8	36.7	46.5	
2012–2022	15.6	35	50.6	
2010–2020	20.5	34.3	54.8	
2008–2018	15.3	35.6	50.9	

Source: US Bureau of Labor Statistics, "Employment Projections" news releases, various years.

Compared with previous years, the new methodology results in very different estimates of job openings. The new methodology results in significantly improved projections of job replacements and therefore is a better measure of employment demand, in our judgment. However, we caution that this significant correction in the methodology must be considered when making comparisons to past projections.

After Everything: Projections of Jobs, Education, and Training Requirements through 2031 can be accessed online at: *cew.georgetown.edu/Projections2031*.

GEORGETOWN UNIVERSITY
McCourt School *of* Public Policy
Center *on* Education *and the* Workforce

instagram.com/GeorgetownCEW

twitter.com/GeorgetownCEW

linkedin.com/company/GeorgetownCEW