OPPORTUNITY INDUSTRIES

TECHNICAL APPENDIX

CHAD SHEARER
ISHA SHAH

December 2018

Metropolitan Policy Program at BROOKINGS

Table of Contents

Research approach

The research that undergirds the Opportunity Industries report began with two key objectives. First, the research sought to identify the industries that disproportionately contain or "concentrate" good jobs in local labor markets. Second, the research sought to estimate how the distribution of good jobs across industries, and the growth of these jobs over time, enable people's movement between jobs in different occupations or industries and to assess how such movements support upward earnings mobility.

The shortage of ready-made information or analytic methods that fit the needs of the research presented several challenges. In particular, the special interest in the spatial variation of job quality and mobility require granular data on local labor markets and labor market dynamics, such as occupational mobility and careers, that are hard to find or do not currently exist.

To overcome these challenges, researchers adopted an inductive analytic approach. First, researchers compiled as much granular, historical data on local labor markets as possible. With these data in hand, researchers used statistical models to capture variation in certain measures of interest, namely occupational mobility and compensation, relative to control variables. These measures of co-variation are then used to estimate occupational mobility and job quality in each metropolitan area.

Developing the estimates used in the research therefore entails four main tasks, described in greater in what follows:

1. Compiling a database that describes in detail (a) the current and historical distribution of jobs across occupations within each industry in each metropolitan area and (b) the characteristics of the workers who hold these jobs.
2. Projecting job growth and associated job openings by occupation in each metropolitan areas using (a) public and private projections of industry and occupational job growth and (b) estimates of labor turnover within occupations.
3. Using these historical data and projections to, first, model the likelihood of occupational switching and, second, use these likelihoods to simulate potential career pathways available to incumbent workers in each metropolitan area over a 12-year period.
4. Using historical data to estimate workers' compensation based on their personal characteristics.

This approach was first applied by the authors of a 2015 Brookings study of labor market opportunity in metro New Orleans. The authors of the present Opportunity Industries report have updated and improved on the data and methods from this prior report to produce this expanded analysis.

The resulting estimates of occupational and mobility and job quality paint a nuanced and varied picture of labor market opportunity across the nation's largest metropolitan areas, leading to intuitive findings and implications that reflect the stylized facts of these regional economies. The assumptions made by these procedures and methods are supported by the literature, by statistical analyses of the fitness of the findings at multiple stages of the research process, and by the representative nature of the dynamics the findings ultimately illustrate.

These methods therefore lead to estimates that suggest how local labor market opportunity may vary across time and space. Hopefully, this approach it sets a helpful precedent on which future research on labor market dynamism and mobility may improve.

Data sources and procedures

Because no one data source contains information on metropolitan areas' industrial and occupational job structure, job openings, worker characteristics, compensation, and the future of each of these dimensions, these data had to be created by assimilating data and projections from a variety of sources, many of which are listed below. For each of these data products, the authors made efforts to draw on well-documented sources and established literature on data cleaning and estimation procedures.

1. Occupational staffing patterns and job openings

The proprietary labor market data firm Economic Modeling Specialists, Inc. (Emsi) provided most of the data on current and historical industry and occupational staffing patterns that provided the basis for much of this research. Emsi's data contain the number of jobs by industry and by occupation at the county level, estimated from datasets published by the federal government in concert with state labor market offices.

Emsi's staffing patterns are used for two purposes: estimating job openings and, later in the research process, to assign jobs by occupation to industries. Data on job openings are crucial to the predicting the likelihood of occupational transitions and to extend the applicability of data on occupational mobility to sub-state areas. Unfortunately, data on occupational job openings are not readily available.

There are two sources of job opening. First, a job opening is created when a business decides to hire for a position that did not previously exist. This is the job growth source. Second, a job opening is created when an incumbent worker moves leaves an occupation. This is the turnover source. Turnover occurs when a worker leaves a job in an occupation for a job in another occupation, or when a worker permanently leaves the labor force, for the purposes of retirement, for example.

Historical data for the job growth source of job openings are generated by interpolating monthly county-level occupational employment data from Emsi's annual estimates using a restricted cubic spline and then subtracting the previous month's number of jobs from the current month's.

Data on future job openings from the job growth source are projected out to 2027 using Moody's Analytics's county-level industry job growth projections and Bureau of Labor Statistics (BLS) Occupational Employment Projections. The Moody's Analytics projections forecast industry job growth based on a proprietary macroeconomic forecast model. The BLS employment projections then provide a forecast of the change in job composition *within* industries *by occupation.* Pairing these two projections together therefore provides some sense of future job growth by occupation.

Estimates for job openings from the turnover sources were generated from the authors' analysis of monthly Current Population Survey (CPS) data. The CPS data were matched and cleaned using methods described in the following sections, below. Using methods similar to those developed by BLS economists, the authors predict job exits by occupation based on the demographics of current job holders.[1] The CPS data are pooled over different time periods to provide average rates of turnover and exiting by month for the past and future, again, using methods similar to those developed by BLS economists.

A comparison of these summed estimates of monthly occupational job openings from across these various sources at the national level to data from the BLS's Job Openings and Labor Turnover Survey (JOLTS) reveal that the present methods yield job openings estimates that closely match estimates provided by JOLTS, in terms of absolute number of openings and rate of openings. As there is no source for sub-state occupational job openings, the authors are unable to validate estimates for those areas. The author's projections of job openings by occupation from the turnover source are much higher than comparable BLS's projections because the authors use a smaller unit of time to measure turnover.

2. Worker characteristics

These data on county-level occupational staffing patterns by industry and estimates of historical and projected job opening by occupation provide a nuanced picture of past and future labor market structure and labor demand in local areas. These estimates are crucial to fitting national patterns estimated in later steps to local circumstances.

However, fitting national patterns to local circumstances requires a parallel look at the workers who comprise local labor markets and their characteristics. The characteristics of the labor force in each metropolitan area included in this analysis come from the public-use microdata sample (PUMS) for the American Community Survey (ACS) 1-year estimates over the years 2012 to 2016, collected directly from the U.S. Census Bureau. PUMS data for these five years were pooled and the person weights from these samples were summarized into individual personas along the following personal dimensions: age in years, sex, race (black or non-black and Hispanic or non-Hispanic), educational attainment, and occupation.

These personas and their metropolitan-level person weights are used in later steps to better fit modeled predictions to the local labor force.

These same ACS data are also used the source for data on employer-sponsored health care insurance coverage, as described in greater detail, below.

3. Occupational mobility

A major challenge for any study of labor flows, such as job or occupational mobility, is finding appropriate data. People can make work transitions quickly and frequently. Detecting all of these transitions requires frequent and detailed observation.[2] Though some governments maintain administrative data that track labor movements, in the U.S. researchers typically must rely on household surveys. However, because longitudinal surveys are cumbersome for both subjects and interviewers, especially at the short intervals required, longitudinal surveys tend to be small. This makes it difficult to extract robust insights on certain labor flows, especially if one is interested in variations across time and space.

Fortunately, The Current Population Survey (CPS) manages to overcome most of the challenges of studying labor flows and boasts several advantages over alternative sources. The CPS is a household survey run jointly by the U.S. Bureau of Labor Statistics and the U.S. Census Bureau. It is one of the nation's largest and longest-running surveys. The monthly version of the survey covers roughly 60,000 occupied households—meaning dwelling units—and is representative of U.S. and state populations.[3] It has served as the nation's source of the monthly unemployment rate since 1948.

The CPS has a unique longitudinal design and detailed questionnaire that makes it particularly well suited for labor flows analysis. A household rotates into the survey for four consecutive months, then out of the survey for eight months, and then returns to the survey for another four consecutive months before leaving the sample permanently.[4] In the initial month of each four-month rotation, respondents are asked about their employment status, including their employer, industry, and occupation. In subsequent months, respondents are asked a series of questions about what, if anything, about their employment status has changed. This makes it possible to confidently identify changes in respondents' labor market status, such as a different employer or occupation from the prior month.

The size, frequency, detail, and quality controls of the CPS yield advantages over alternative sources of data on occupational transitions. The CPS surveys enough households that it is representative of the population and workforce at the national and state levels. The survey's monthly frequency over four consecutive months reduces survey attrition and makes it possible to detect most labor market transitions. The survey also queries demographic and educational characteristics of respondents, making it possible to distinguish labor market behaviors among different population groups. Finally, the design of the questionnaire and back-end quality control permit high confidence in such identifications.

Alternative sources provide few of these advantages. Longer-range longitudinal surveys like the National Longitudinal Survey of Youth or Panel Study of Income Dynamics are too small and infrequent to robustly identify all of their subjects' labor market transitions. Resume data are an exciting potential source of career data in the future but so far researchers have not been able to overcome all of resumes' inherent selection bias, in terms of population and content, and unstructured nature. And no other source provides as comprehensive and representative information about subjects' personal characteristics and broader labor force trends.

However, the CPS is not perfect and does require some care to make it maximally useful. Despite its longitudinal design and detailed information about the work status of respondents, the CPS was not originally intended as a longitudinal source. It therefore requires procedures that help us take advantage of this design. Further, like all surveys, the CPS must deal with survey attrition and non-response—problems that have grown worse for almost all household surveys and polls in recent years. In cases of non-response to a question or inconsistent answers, CPS interviewers and data coders strive for completeness rather than accuracy, which can introduce error for studies like this one.

To start with, monthly records of CPS respondents must be matched across months. To do this, the authors follow the procedures developed by Madrian and Lefgren of using administrative record numbers, which should enable a guaranteed "mechanical" match.[5] For some years, the authors apply supplemental procedures proposed by researchers at the Minnesota Population Center to fill in fields needed for mechanical matching.[6] Occasionally, mechanical matching procedures result in a bad match due to data entry errors during survey administration. Therefore, following these earlier studies, mechanical matches are confirmed using sex, race, age, and education.[7]

Next, records that do not survive these matching procedures for month-to-month matching are disqualified and dropped from the analysis. Because these non-matching records are more common for some population groups, the authors use an inverse weighting technique similar to the one proposed by Bollinger and Hirsch to avoid introducing sample bias from dropping the disqualified records.[8] In this procedure, the longitudinal survey weight of each matched pair of monthly records is inflated in proportion to the likelihood of that respondent not matching in that month.

With these unbiased matches in hand, confirmed and reweighted cross-monthly records are then analyzed to identify occupational mobility. This step also requires some careful cleaning and reweighting procedures.

The CPS began using what it calls a "dependent coding" scheme in 1994 to record industry and occupation for survey respondents. This dependent coding scheme yields more reliable data on these dimensions than the earlier independent coding scheme. Under dependent coding, a survey respondent is asked about their employer, their employer's industry, and their own occupation only in the initial month of each survey rotation. In subsequent months, changes to these fields are "dependent" upon the respondent's affirmation of some change in their work status or history, such as a new employer or a change in the respondent's job duties or activities. If one of these dimensions has changed, then a new occupation is recorded based on the same series of question at the initial month. Otherwise, the respondent's industry and occupation are recorded as the same as the prior month. In the event of a gap in survey coverage for a household or particular respondent during a given month of a rotation, or if the answer to a survey question is ambiguous, the CPS attempts to impute a response to complete the survey record.

For this present analysis, the authors adopted a highly restrictive approach to vetting these data on occupational mobility from the matched cross-monthly records. First, the authors identified any record in which industry or occupational data had been independently coded in the second of the paired monthly records. These are matched records where fields should have been dependently coded in the second month, thus the independently coded industry and occupation fields are suspect. Second, the authors dropped any matched records where industry or occupation fields had been imputed using procedures that estimate these fields from data not originally provided by the respondent. Third, the authors dropped all matched records in which the CPS implemented new industry and occupation coding schemes during the second of the two matched monthly records so as not to confuse coding changes with changes in the respondent's occupation. As with the monthly record matching procedure, the longitudinal weight of the records retained following these cleaning procedures are inflated based on the record's probability of having been dropped due to independent coding or imputation. This should correct potential sample bias introduced by the cleaning procedures.

As a result of these procedures, the authors are able to retain roughly 7.9 million matched month-to-month records covering the period from January, 2001 through June, 2018. Of these records, slightly more than 300,000 reflect an occupational change. This implies that transitions from one occupation to another occupational, at least at the most detailed level of occupational disaggregation available from the CPS, are rather rare—numbering less than one out of every 25 workers on a monthly basis, and far less in some months.

The final step in the process of culling data on occupational mobility from the CPS for the analyses that follow is translating Census Bureau industry and occupational coding schemes into the coding schemes followed by the BLS and other federal statistical bureaus. To do this, the authors created a series of concordance tables linking Census industry and occupation coding schemes from different time periods into contemporary North American Industrial Classification System (NAICS) industry codes and Standard Occupational Classification (SOC) System occupation codes. For the purposes of this research, industry codes were translated into only their two-digit NAICS equivalent—the highest level of industry aggregation available in the NAICS scheme. However, occupational detail is of special interest for this project since one goal is to reveal occupational mobility with as much precision as possible—a way of getting to the most accurate estimates of mobility. Occupational codes have become more detailed over time, which means that some older Census occupation codes have been split into multiple SOC codes. Rather than sacrifice occupational detail, matched month-to-month CPS records for split codes are multiplied by the number of splits for the original code. This one-to-many matching on occupational codes does lead to a small upward bias in long-run estimates of the average rate of occupational mobility in affected occupations but is well worth the greater overall occupational precision.

As a result of this occupational recoding, the authors are able to look at mobility and compensation in 532 distinct occupations.

In addition to these cleaning procedures, the authors make an effort to identify or estimate the geographic location of CPS respondents since geographic variation is an interest of this research. All CPS records from the period included in this analysis include state identifiers and most also include metropolitan area and sometimes even county identifiers. Those records without a metropolitan or county identifier are given a unique, time-specific geographic weight that reflects the average location of that survey respondent in the state it is identified with. These geographic weights are used in subsequent analyses to approximate the job openings within a giving commuting distance of that respondent.

4. Compensation

The final data input required for this research are microdata on individuals' wages, hours worked, and employment benefits. Most of these data come from the same monthly CPS data described above. In the final month of each of a household's two survey rotations, respondents are interviewed about their hours worked, hourly or weekly wages, and eligibility for overtime and other types of non-wage compensation. These records, referred to as the Outgoing Rotation Groups (ORG) file, are the basis for the analyses of hours and compensation described below. As noted earlier, the comprehensive nature of the CPS, including the information it contains on individual characteristics, makes it especially useful source for this information. The information itself is also more granular, in terms of providing compensation for different units of labor, than alternative sources that provide only annual compensation. However, using the CPS for these data yields a smaller sample and diminished ability to analyze geographic variation in compensation within occupations.

Data on the provision of benefits come from the ACS. The primary interest here is whether respondents are covered by employer-sponsored health care insurance, which the authors consider a proxy for other types of employment benefits, such as paid leave. The ACS is a superior source for this information because, like the CPS, it contains data on individual characteristics as well as industry and occupation. The ACS is also a much larger survey than the CPS, meaning it does lend itself to analyses of geographic variation in insurance coverage within occupations. However, one limitation of the ACS is that it is not clear whether a respondent is covered by their own employer's coverage or that of a parent or spouse.

Analytic methods

With the data described above in hand, the authors began a four-step process of modeling career pathways and occupational job quality for workers within each of the nation's 100 largest metropolitan areas.

1. Modeling local career pathways from data on occupational mobility

The authors used regression analysis to estimate the probability that an incumbent worker would switch from his or her present occupation to another. This was done for each unique pair of occupational transitions observed in the CPS data on occupation mobility. Although there are over 282,000 potential transitions, the data on occupational mobility reflect fewer than 12,000 pairs of occupations where more than one respondent made a transition.

In each regression for these occupational transitions, the universe was any respondent who reported working in the origin occupation in the prior month. The dependent variable indicated whether the respondent worked in the destination occupation in question during the current month. The authors used logistic regressions that condition the probability of switching between the two occupations on the destination occupation's share of job openings in the vicinity of the respondent in the current month and the respondent's personal characteristics, including age in years, sex, race, and level of education.

In the authors' view, this is the best approach to estimating the probability of occupational switching given the task at hand: using national data to develop localized estimates of the probability of occupational switching. For this purpose, the authors were not concerned with explanatory power or statistical significance. Instead, the concern was predictive power. Most of the successful regressions do have reasonable explanatory power. Based on the authors' review of literature, most also have better predictive power than regressions conditioned on the similarity of occupations based on their shared tasks or human capital. However, models for some pairs of occupations had too few observations to produce valid results. These models were omitted from further steps.

2. Constructing occupational transition matrices

The results of these regressions were then applied against data on the metropolitan area's workforce characteristics and projected future job openings in order to derive monthly persona-specific occupational transition matrices. In these matrices, each cell contains the probability that a person meeting the characteristics of that persona would switch from a given origin occupation (the row) to a given destination occupation (the column) at a given month in the future in the metropolitan area.

This process began by defining a universe of personas. Each persona is defined by the personal characteristics included in the occupational transition regressions. There is one persona for every combination of age, sex, race, and education included in the regressions, resulting in many thousands of personas for each metropolitan area. For example, one persona represents a 35-year-old black non-Hispanic male with a bachelor's degree. Another represents a 50-year-old Hispanic female with some college experience. The authors calculate a person weight for each row each persona's matrix using data from the ACS to represent the number of workers that belong to each persona that report working in the origin occupation in the metropolitan area.

Using the successful occupational transition regressions from above, the authors determined the probability that a persona will transition between a given pair of occupations given the personal characteristics it represents and a destination occupation's projected share of job openings in a given future month in the metro area. For pairs of occupational transitions where regressions were unsuccessful, the authors simply assigned the observed rate of transitions from the CPS data rather than a probability estimated from a regression. For pairs of occupations where no transitions can be observed from the CPS data, a transition probability of zero was assigned. Finally, the probability of not transitioning from a given occupation to another (contained in the cells on the matrix's diagonal) was set equal to one minus the sum of the other cells in the row.

This process was done for every persona and for every month from 2017 to 2027, resulting in several hundred thousand matrices, each conveying transition probabilities for every pair of 532 occupations that describe every job in the metropolitan area. Each matrix conveys the likelihoods that a person fitting the characteristics of that persona will switch between any two of the occupations in a given month in the metropolitan area, and each persona's matrix has a weight that references the number of actual workers who fit the persona's definition in the metropolitan area.

3. Estimating occupational transitions in career pathways

Finally, to estimate the cumulative conditional probability of that a worker represented in a persona who begins in any one occupation at the end of 2016 will end up in any other by the end of 2027, these monthly persona-specific matrices are multiplied against each other in what is known as a Markov Chain. This begins by multiplying the matrix for January 2017 against the matrix for February 2017. The matrix product of this multiplication indicates the probability that a worker starting in a given occupation in December 2016 will transition to any other occupation come February, 2017. This matrix product is then multiplied against the matrix for the month of March 2017, whose product is multiplied against the matrix for April 2017, and so on until December 2027.

The final matrix product that results from this process, and the person weight assigned to the original matrices, as described above, forms the basis for this report's analysis of local career pathways.

This approach cannot perfectly describe career pathways. It rests on an assumption that a worker's probability of transitioning into another occupation depends only on the worker's occupation in the prior month (along with the variables factored into the regressions, of course). Actually, transitions likely also depend on the worker's tenure in their occupation and may also depend on the occupations the worker has previously worked in. Basically, this approach cannot capture the importance of a person's resume in predicting his or her future occupational transitions. However, such information is at least partially factored into the analysis given that the real workers and transitions represented in the underlying CPS data were in fact able to obtain those jobs and make those transitions. This approach seems reasonable in the absence of data that truly represent actual careers.

This approach is also supported by the literature. Academic researchers and private-sector business consultants commonly use this sort of matrix multiplication to simulate exactly these sorts of state-change dynamics, and many academics have applied this approach to similar questions of labor flows. Two central issues in these studies are identification and embeddability.[9] Identification refers to whether one has accurately identified the true probability of a transition. This is why in the present case so much care has been taken to clean and parse the data on occupational mobility that undergirds this analysis, to appropriately condition estimate transition probabilities, to rely only on transition regressions that yield robust results, and to create unique matrices for each persona.

Embeddability refers to whether the transition matrix containing transition probabilities can accurately describe a continuous process of change. In this case, the transition matrices used in the analysis are so large and the probability of transition between occupations in a given month so low that the question of whether the matrices are truly embeddable is not a concern. Nonetheless, the authors again take great care to create period-specific transition matrices conditioned on projected job opening rates in each metropolitan area so as to ensure these matrices accurately reflect a continuous process of change.

Modeling earnings and benefits

In the final phase of this analysis, hours, wages, and benefits were estimated for all starting and ending occupations for each persona. Estimates for each these measures were derived from regression analyses of the CPS ORG data or ACS data described above. Each of these regressions are premised on the Mincerian earnings function, a model developed by economist Jacob Mincer that labor economists commonly use to explain wages. In Mincer's model, education and work experience explain wages.

The regressions used in the present set of analyses build from Mincer's relatively simple model. Age is substituted as a proxy for work experience. Additionally, regressions for each measure include variables indicating sex, race, and sector of employment since labor market outcomes like wages are known to vary along these lines, even for workers in the same occupation. For the wage and benefit regressions, full- or part-time status is also used as an explanatory variable. For the benefit regressions, state fixed effects are included. (Benefit regressions only include data from years after the implementation of the Affordable Care Act). All regressions were carried out for each occupation.

Estimating each measure involved a different model specification. To estimate hours, the authors used an ordered logistic regression. To estimate probability of receiving employ-sponsored health care insurance—a proxy for other employment benefits—the authors used a logistic regression. To estimate hourly wage, the authors used a generalized linear model where the wage is modeled as its logarithm. Additionally, the authors estimated workers' exempt or non-exempt status to determine their overtime eligibility. This logistic regression excluded demographic factors besides age and included state fixed effects.

4. Estimating job quality by industry

The results of all these estimates—occupational transitions, hours worked, wages earned, overtime eligibility, and benefits—are, at last, evaluated together to determine whether the average worker in a given persona, given current occupation, and given future occupation has a good, promising, or other job. The result is a nuanced accounting of the proportion of workers within a given occupation who hold jobs that meet the criteria for good or promising jobs.

A worker is assigned to a "good" tbased on the probability that the worker's job meets these conditions:

1. The worker's combination of hourly wage and hours worked per week, plus overtime pay, over 52 weeks must be equal to or greater than the median annual earnings of full-time, year-round sub-baccalaureate workers in the reference metropolitan area.
2. The worker must be covered by employer-sponsored health insurance.

Workers who do not have a good job are assigned to a promising job based on the probability that they will reach a good job by the start of 2028 according to the estimates of cumulative conditional probability of transition described above and the workers probability of meeting the same criteria for good jobs in 2028.

The proportion of workers who do not hold good or promising jobs are assigned instead to a category of "other" jobs.

In the final step, all these estimates are combined and assimilated for each occupation by industry. This is done by "mapping" the results of all these various analyses, which have been done by occupation and at the industry-sector level, back to detailed six-digit NAICS codes using Emsi's occupational staffing patterns, described above.

Future research

Recent advances in data completeness, computing, and analysis make it possible to document labor market trends like never before. Many of these trends appear concerning, and policymakers and practitioners are therefore grappling with difficult new questions about change labor market demand and the future of labor market opportunity.

This research effort attempts to capitalize on advances in data and analysis to shed new insight on the possible future of labor market dynamism in local areas and how such dynamics might be tilted so as to improve outcomes for individuals and communities. Importantly, this research builds on prior researchers' efforts to describe and model various labor market processes. As a result of these efforts and these other advances, the present effort is therefore able to put forward new methods, approaches, and insights that begin to advance our understanding of the labor market dynamics that enable and inhibit upward economic mobility.

Yet there remains ample room for improvement, refinement, more thorough testing, and a deepening of this research. Though there is considerable reason to believe the results of this analysis are suggestive of real trends and processes, this conjecture must be put to the test through further research and monitoring. New data sources, especially resume data, may eventually help bridge the gap between the inductive techniques used here and deductive methods of analysis to better test and base assumptions or even extend the power of the approaches used in this research. Finally, assuming this research is at least somewhat true to life, further research should be done to understand what its estimates of labor market mobility can reveal about the determinants of upward mobility and the types of human capital accumulation that will help workers advance in today's economy.

References

Bollinger, Christopher and Barry Hirsch. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching." Journal of Labor Economics, 24(3): 483-519, 2006.

Madrian, Brigitte and Lars John Lefgren. "A Note on Longitudinally Matching Current Population Survey (CPS) Respondents." National Bureau of Economic Research Working Paper No. T0247. 1999.

Moscarini, Guiseppe and Kaj Thomsson. "Occupational and Job Mobility in the U.S." Scandinavian Journal of Economics, 109(4): 807–836. 2007.

Rivera Drew, Julia, Sarah Flood, and John Robert Warren. "Making Full Use of the Longitudinal Design of the Current Population Survey: Methods for Linking Records across 16 Months." Journal of Economic and Social Measurement, 39(3): 121–144. 2014.

Shearer, Chad, Amy Liu, Natalie Holmes, and John Ng. "Opportunity Clusters: Identifying pathways to good jobs in metro New Orleans." Brookings. 2015.

Singer, Burton and Seymour Spilerman. "The Representation of Social Processes by Markov Models." American Journal of Sociology, 82(1): 1-54. 1976.

U.S. Census Bureau, 1-Year American Community Survey (ACS) Public Use Microdata Sample (PUMS), 2012-2016.

U.S. Census Bureau and U.S. Bureau of Labor Statistics, Basic Monthly Current Population Survey (CPS) microdata, 1994-2018.

Notes

[1] Michael G. Wolf and C. Brett Lockard, "Occupational separations: a new method for projecting workforce needs," U.S. Bureau of Labor Statistics, 2018. Available at https://www.bls.gov/opub/mlr/2018/article/occupational-separations-a-new-method-for-projecting-workforce-needs.htm.
[2] Guiseppe Moscarini and Kaj Thomsson, "Occupational and Job Mobility in the US," Scandinavian Journal of Economics, 109(4), 2007.
[3] U.S. department of Labor and U.S. Department of Commerce. "Current Population Survey Design and Methodology." Technical Paper 66. 2006.
[4] Ibid.
[5] Brigette Madrian and John Lars Lefgren, "A Note on Longitudinally Matching Current Population Survey (CPS) Respondents," NBER, 1999.
[6] Julia Rivera Drew, Sarah Flood, and John Robert Warren, "Making Full Use of the Longitudinal Design of the Current Population Survey: Methods for Linking Records Across 16 Months," Journal of Economic and Social Measurement, 2014.
[7] Challenging mechanical matches on these characteristics is crucial for deriving the most accurate estimates of occupation mobility as possible. Any record that is not a perfect match has a high potential of biasing estimates of occupational mobility upwards since the occupation of incorrectly matched records is not likely to match. Because rates of mobility are low at high levels of occupational disaggregation, the bias from such errors can be significant. The authors find that education is effectively treated as an immutable characteristic for the purposes of the CPS's two four-month survey rotations: respondents appear to be asked about educational attainment only at the beginning of each rotation, so it should not change from month to month. The authors therefore disqualify naively matched records where education does not match.
[8] Christopher Bollinger and Barry Hirsch, "Match Bias from Earnings Imputation in the Current Population Survey: the Case of Imperfect Matching," Journal of Labor Economics, 24(3), 2006.
[9] Burton Singer and Seymour Spilerman, "The Representation of Social Processes by Markov Models," American Journal of Sociology, 82(1), 1976.

Metropolitan Policy Program
at BROOKINGS

1775 Massachusetts Avenue, NW
Washington, D.C. 20036-2188
telephone 202.797.6139
fax 202.797.2965
www.brookings.edu/metro