BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Employment and Unemployment Statistics

On Job Requirements, Skill, and Wages

Matthew Dey, U.S. Bureau of Labor Statistics
Mark A. Loewenstein, U.S. Bureau of Labor Statistics

Working Paper 513
March 2019

On Job Requirements, Skill, and Wages

Matthew Dey and Mark A. Loewenstein

U.S. Bureau of Labor Statistics

March 2019

Abstract

Occupations are bundles of inseparable skill requirements and tasks. We propose a novel approach for studying the relationship between wages and bundles of occupational skills and tasks. We predict occupational wages using a regression tree approach which also provides an empirically powerful aggregation scheme where detailed occupations with similar wages and job requirements are combined into 15 large occupation groups. Our empirical analysis is carried out on a dataset obtained by combining O*Net information on job attributes with the occupational wage and employment information from Occupational Employment Statistics. Not having a priori information on which O*Net variables belong in a wage equation, the first step in our analysis is to perform factor analysis on a number of O*NET categories that represent basic job skill requirements and job attributes. The second step is to use a regression tree to group the detailed SOC occupations into broader aggregates. These occupational aggregates are then used to non-parametrically analyze the well-known hollowing out phenomenon and the increase in log wage variance from 2005 to 2017.

Keywords: Skills, tasks, occupations, factor analysis, regression tree

JEL classification codes: J31, C43

Earlier versions of this paper were presented at the 2018 BLS-Census Conference and at the April IZA Labor Statistics Workshop on the Returns to Skill in the Labor Market. Any opinions and conclusions herein are those of the authors and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics. Stata and SAS versions of the analysis dataset are available for downloading.

Mark Loewenstein

Loewenstein.Mark@BLS.Gov Dey.Matthew@BLS.Gov

2 Massachusetts Ave., NE, Washington, D.C. 20212

(202)-691-7385

I. Introduction

There is a vast economic literature concerned with workers' accumulation of human capital as well as the relationship between wages and human capital, where human capital has been traditionally proxied by education. More recently, economists have begun to enrich the analysis of wages by introducing various job attributes. Most notably, attention has been placed on whether or not an individual's job is routine. Researchers have found that routine jobs pay less than jobs where workers have more independence in their decisions and actions. As computer processing power has become dramatically cheaper over time, jobs that are in the middle of the skill distribution have become more routine. Various authors have argued that this is an important factor behind the hollowing out of the wage distribution (see, for example, Autor, Levy, and Murnane (2003), Goos and Manning (2007), and Acemoglu and Autor (2011)).

Papers analyzing the relationship between job attributes and wages have utilized the US Department of Labor's Dictionary of Occupational Titles (DOT) and its successor the Occupational Information Network (O*Net). This has proved to be a fruitful line of research, but it is fair to say that these data have not been used very thoroughly. Relatively few researchers have availed themselves of the O*Net data. Those that have used the DOT or O*NET data have generally done so in a rather piecemeal fashion, choosing a handful of variables of interest and ignoring the remainder. One exception is Ingram and Neumann (2010), who merge demographic and wage information in the Current Population Survey (CPS) with job characteristic information in the DOT in order to investigate the effect that job skills have on wages and how job skills have changed over time. Rather than analyzing the separate effects of the myriad job characteristics in the DOT, Ingram and Neumann analyze returns to several latent characteristics that they obtain using factor analysis. Poletaev and Robinson (2008) and Robinson (2011) merge the wage and mobility information in the CPS and the Displaced Worker survey with job characteristic information in the DOT in order to investigate the likelihood and wage consequences of voluntary and involuntary mobility. Using factor analysis, these authors construct latent factors that they then use to measure the skill proximity of jobs.

With the few exceptions noted above, researchers have typically used the DOT and O*Net in a fairly ad hoc manner. In this paper, we attempt to use the O*Net variables in a systematic fashion. Furthermore, we merge the O*Net information with the U.S. Bureau of Labor Statistics' Occupational Employment Statistics (OES) dataset. Upon doing so, we are able to analyze the relationship between a host of job attributes and wages and how the prevalence of these job attributes are changing over time.

Our approach differs markedly from the one that is prevalent in the literature. We start by choosing O*NET categories that represent basic job skill requirements (e.g., deductive reasoning, oral expression, trunk strength) and job attributes (e.g., frequency of decision making). Together, these categories contain 148 O*NET variables. We then use factor analysis to reduce the number of O*NET variables, carrying out a separate factor analysis for each category of O*NET variables. Factor analysis in essence provides us with a convenient way to weight the variables within each of the O*NET categories and the factors we obtain are for the most part readily interpretable.

As several authors have noted, jobs are characterized by a bundling of attributes (for example, see Heckman, James and Jose Scheinkman (1987), Yamaguchi (2012), and Autor and Handel (2013). Thus, although the factors we obtain explain a great deal of the variance in occupational wages, coefficients on the individual factors in a wage equation must be interpreted with care. What makes conceptual sense is to estimate the returns to factor bundles. This suggests performing some type of cluster analysis on the factors that we have identified. A natural approach is to weight factors by their effect on wages, but this is complicated by the fact that the wage equation appears to be nonlinear with interaction effects. We therefore use a regression tree to group the detailed SOC occupations into broader aggregates. Each aggregate consists of detailed occupations paying similar wages and possessing similar factor quantities in so far as these factors are correlated with wages in the relevant part of the wage distribution. Taken together, the occupational aggregates obtained through the regression tree provide a clear picture of the types of jobs located at various parts of the wage distribution. They also enable us to examine the evolution of employment and wages over time using a nonparametric counterfactual analysis.

The paper is organized as follows. Section II briefly looks at the OES data by itself. The employment polarization researchers have found in other data in earlier periods is clearly evident in the OES data for the period 2005 -2017. Section III provides a brief look at the O*NET data and demonstrates that the O*NET variables are very powerful when it comes to explaining occupational wage variation. Section IV condenses the information on job attributes in O*NET using factor analysis. We are able to reduce an initial list of 148 O*NET variables to 21 factors. We examine the correlations among these factors as well as between the factors and the O*NET variables used by Acemoglu and Autor.

In Section V, we use a regression tree to group the 788 detailed SOC occupations into broader aggregates. Our analysis yields fifteen distinct occupational groups. Looking at the evolution of wages and employment from 2005 to 2017, one finds that the well-known hollowing out pattern that has been observed in previous years is vividly displayed in our fifteen occupational groups. The employment share of higher paying jobs with greater cognitive and decision making requirements increased, as did the share of low paying jobs requiring little in the way of cognitive or physical skills. The share of jobs requiring greater physical and psychomotor skills fell, as did the share of jobs requiring modest cognitive skills. We complete our analysis by looking at wage variance. We find that variance increased from 2005 to 2014 due to both shifts in occupational employment and changes in wages across the fifteen occupational groups. However, wage variance fell from 2014 to the end of our sample period due to an increase in the relative wage of the bottom occupational group.

Concluding remarks are presented in Section VI.

II. A Brief Look at the OES Data

The OES survey measures occupational employment and wages in the United States by geography and industry. The OES program surveys approximately 200,000 establishments per panel (every six months), and the entire sample is surveyed over a three year period. Each year of OES data therefore contains observations from about 1.2 million establishments. Each observation contains information on both the number of employees and the wages earned by workers in each occupation at an establishment.[1] In this draft, we use OES data from 2005 through 2017.[2]

The trend toward increasing inequality (as measured by an increase in the variance of log wages) and increasing polarization have received quite a bit of attention in the economic literature. This section briefly examines what OES data tell us about these phenomena.

Figure 1 depicts the evolution of major occupation employment shares during the 2005-2017 period. One sees that the share of sales and office employment fell steadily and substantially throughout the entire period. The shares of trades and blue collar employment fell sharply as result of the Great Recession and then levelled off with a slight recovery. In contrast, the shares of professional and service employment rose substantially during the Great Recession and then levelled off with a slight upturn (employment in these occupations did not increase during the Great Recession, but their employment share rose because employment in the other occupations fell.) Our occupation

---

[1] See Handwerker and Spletzer (2014) for a much more complete and thorough description of the OES. Handwerker and Spletzer use the OES to examine trends in wage variance.

[2] OES data are currently available through 2017. We exclude years prior to 2005 because of inconsistencies in the data.

breakdowns are a little different from those in Autor (2015). However, reassuringly, the patterns in the OES data cited above look comparable to those Autor finds using the American Community Survey.[3]

Figure 2 shows how employment of occupations in the various parts of the 2005 wage distribution has evolved from 2005 to 2017. One sees that the share of employment in the group of occupations paying average wages in the 20^{th} to 80^{th} percentile range in 2005 fell steadily. In contrast, the share of employment in the top and bottom paying groups in 2005 increased steadily throughout the period. The OES data thus provide clear evidence of labor market polarization: employment shares rose for occupations at the ends of the wage distribution and fell for those in the middle.

Since wages vary across occupations, changes in the composition of occupational employment lead directly to changes in the distribution of wages. In addition, changes in the relative payoffs to various occupations will also directly lead to changes in the wage distribution over time. Figure 3 shows how mean real wages would have evolved had occupational employment been fixed at 2005 levels. According to the OES, real wages moved unevenly from 2005 to 2007, fell from 2009 to 2014, and increased from 2014 to 2017. Wages across the bottom, middle, and top groups largely rose or fell together, but often by differing amounts. Interestingly, in 2014, real wages in the bottom and middle paying occupations were below their 2005 levels, while real wages in the top paying group were about two percent higher. However, from 2014 to 2017, percentage wage increases in the bottom paying occupational group outpaced those in the top and middle groups. By 2017, wages in the bottom paying occupational group were more than 6 percent above their level in 2005, wages in the top group were about 5 percent above their 2005 level, and wages in the middle paying occupation group were about two and a half percent higher than their 2005 level. Our results for the 2007-2012 subperiod are roughly similar to those in Autor (2015), who finds relatively flat wages for all groups when the

[3] Autor analyzes earlier years using data from the Census Integrated Public Use Microdata Series files. These years are not available in the OES.

endpoints for comparison are 2007 and 2012. However, as noted above, there is quite a bit of movement in wages during the years 2014- 2017.

III. A Look at the O*NET Data

The Occupational Information Network, which is known as O*Net, provides information on job content. The data are produced under the sponsorship of the Department of Labor's Employment and Training Administration. O*Net contains occupation-level measures of the knowledge and skills required by an occupation as well as on how work is carried out. As noted above, O*Net is the successor to the Dictionary of Occupational Titles (DOT). Initially, the information in the database was collected by occupational analysts. Over time, this information has been updated by surveys of both occupation experts and each occupation's worker population.

O*Net places job attributes into a number of categories. We use many, but not all of these categories. Specifically, we choose variables in categories that represent basic job skill requirements (e.g., deductive reasoning, oral expression, trunk strength) and job attributes (e.g., frequency of decision making).[4] However, we do not use variables in categories that describe occupation specific knowledge or interests (e.g., biology, chemistry, clerical) because these variables are not helpful in making cross-occupation comparisons. Using occupation specific job characteristics in a wage equation would be similar to simply using occupation dummies. One additional O*Net variable is of interest, the education level that is required for the job. This variable differs from the years of schooling variable found in demographic data sets, but one would expect the two variables to be positively correlated: we would expect to find individuals with more schooling sorted into jobs requiring more education. Required

[4] Another dataset with fairly substitutes on job attributes is the Qualification and Career Survey. Black and Spitz-Oener (2010) merge this dataset with wage information in the Administrative Social Security Records to analyze the evolution of women's and men's wages in West Germany between 1979 and 1999. Gathman and Schönberg (2010) use these two datasets to analyze occupational mobility.

education is coded in O*NET as a categorical variable – some high school, high school, some college, college degree, masters or Ph.D. For expositional convenience, we convert this into a continuous variable by computing the average years of education required for the occupation. For example, if the O*NET reports that 50% of the time an occupation requires a high school degree (assigned 12 years of education) and 50% of the time an occupation requires some college courses (assigned 13 years of education), the occupation is assigned 12.5 years of education.[5] However, our results below are insensitive to whether education is treated as categorical or continuous.

The 11 O*NET categories that we use and the variables in each category are listed in Table 1. The vast majority of the variables are self-explanatory, but explanations of all of them can be found on the O*NET website.[6] Note that the variables fall into several broad categories. The cognitive, physical, psychomotor, and sensory variables appear to measure the skills required by workers employed in an occupation. The information, interaction, mental, output, interpersonal, and structural variables would seem to describe the activities in which the workers in an occupation are engaged. Finally, the conditions variables appear for the most part to explain working conditions. We do not have a priori knowledge of which variables belong in a wage equation and therefore include all of the variables in our ensuing analysis.

To what extent do the O*NET variables explain wage variation across occupations? To answer this question, we merge the O*NET data into the 2016 Occupational Employment Statistics data.[7] The

[5] There are 12 education categories: Less than High School Diploma (8 years), High School Diploma (12), Post-Secondary Certificate or Some College Courses (13), Associate's Degree (14), Bachelor's Degree (16), Post-Baccalaureate Certificate (17), Master's Degree (18), Post-Master's Certificate (19), Professional Degree, Doctoral Degree, or Post-Doctoral Training (20).
[6] The Abilities variables can be found at https://www.onetonline.org/find/descriptor/browse/Abilities/, the Work Activities variables can be found at https://www.onetonline.org/find/descriptor/browse/Work_Activities/, and the Work Context variables can be found at https://www.onetonline.org/find/descriptor/browse/Work_Context/
[7] From 2005 to 2017, the OES has 788 time-consistent occupation codes. 44 of these codes are not found in the O*NET. For occupations that are not in the O*NET we find the closest (in terms of estimated occupational wage

first row in Table 2 shows the results of simply regressing the log of the average 2016 occupational wage against the education variable. In this equation and all that follow, we weight occupations by their total employment. Not surprisingly, an occupation's wage is strongly correlated with required education. The R squared in the regression indicates that education alone explains 66.5 percent of the variation in occupational wages.

Summary results of regressing occupational wages against the O*NET variables other than education are presented in the second row of Table 2; estimated coefficients and their standard errors can be found in Table 1. The O*NET variables as a group are exceptionally powerful in explaining occupational wage variation: the R squared in the regression is 0.933. However, the individual effects are difficult to interpret. Relatively few coefficients are significantly different from zero and a number of coefficients have signs contrary to what one would expect. This, of course, is not surprising since the O*NET variables are highly correlated with each other. As depicted in row 3 of Table 2, adding education to the equation adds little explanatory power, as the R squared only increases to 0.937.

There are a couple of potential reasons why the O*NET variables are highly correlated. First, many of the variables appear to measure similar things. Second, skills, job activities, and working conditions may not be randomly scattered across jobs, but instead may appear in patterns. Jobs invariably require a variety of skills and involve several tasks. A particular skill may have substantial value when combined with other skills and tasks, but have little value by itself. Indeed, as noted by Autor and Handel (2013), "tasks are a high-dimensional bundle of activities, the elements of which must be performed jointly to produce output. For example, flight attendants engage in both interpersonal and physical tasks, construction workers perform both analytical and physical tasks, and managers

premiums from a log wage regression on geographic area, detailed industry, and occupation) occupations to the missing occupation and assign the weighted average of the O*NET variables for these occupations to the missing occupation. This allows us to completely cover all employment in all time periods. Most of the missing occupations are residual or "All other" occupations.

perform both analytical and interpersonal tasks. In each case, these core job tasks cannot be unbundled; each worker occupying the job must perform them."

IV. Using Factor Analysis to Condense the Information in O*NET

We begin by addressing the first reason many of the O*NET variables are correlated, namely, that they measure similar things and partially addressing the bundling of skills and tasks. (We address bundling more fully in the next section). One potential way of dealing with the fact that the O*NET variables are correlated is simply to cherry pick variables on the basis of a priori intuition or experimentation. We take an approach that is less ad hoc: we significantly reduce the number of variables using factor analysis.

The aim of factor analysis is to generate latent factors that capture the variability among a larger number of observed, correlated variables. The factors that are generated have mean 0 and variance 1 and are uncorrelated with each other.

We perform factor analysis on each of the 11 O*NET categories listed in Table 1. We purposefully perform factor analysis within as opposed to across the various O*NET categories in order to facilitate interpretation. Interpreting the factors we obtain is straightforward because the O*NET categories have clear interpretations. In essence, factor analysis simply provides us with a convenient way to weight the variables within each of the O*NET categories. It also can be thought of as reducing the measurement error associated with any of the individual variables.

The factors and factor loadings are shown in Table 1 and Table 3 presents the correlation matrix of the 22 variables (years of education plus 21 factors). As Table 1 indicates, we are able to reduce the twenty-one cognitive variables to two factors that we label COGNITIVE1 and COGNITIVE2. We reduce the nine physical job attributes to one underlying factor and the ten psychomotor variables to one factor as well. Similar data reductions occur throughout all of the categories. All in all, we are able to boil down our initial list of 148 variables to 21 factors.

As mentioned previously, jobs typically involve a bundling of skills and activities. Note that this is reflected in our results above both in the factors themselves and in the correlations among the factors. A few observations concerning some of the factors and their correlations follow. The COGNITIVE1 factor captures a range of cognitive skills. The correlation between this variable and education is a very high 0.86. COGNITIVE1, MENTAL, which measures decision making job requirements), and education are all positively correlated with OUTPUT2, which measures interactions with computers. In contrast, PHYSICAL picks up a job's physical requirements. PHYSICAL's correlation with required education is a highly negative 0.53 and its correlation with COGNITIVE1 is a highly negative 0.69. On the other hand, PHYSICAL is strongly positively correlated with PSYCHOMOTOR, a factor that captures manual dexterity and related skills. Another variable of interest is STRUCTURAL1, which measures the extent to which jobs are non-routine and require independent decision making (one of this factor's major components is the variable structured versus unstructured work, which has previously received attention in the literature (for example, see Autor , Levey, and Murnane (2003) and Autor and Handel (2013)). This factor is highly correlated with both EDUCATION and COGNITIVE1. CONDITIONS1 captures hazardous and unpleasant working conditions. This variable is highly correlated with PHYSICAL, PSYCHOMOTOR, and jobs using machinery (OUTPUT1); it is negatively correlated with education and COGNITIVE1.

Following up on the initial work by Autor, Levy, and Murnane using the DOT to construct key job attributes, Acemoglu and Autor (AA) pick out several variables from O*NET to obtain similar constructs. To gain further insight into both the AA variables and the factors we have found, we present correlations between the AA variables themselves and between the AA variables and our factors in

Table 4.[8] From the table we see that the job attribute AA labeled ANALYTICAL is very highly correlated with the COGNITIVE1 factor (the correlation between the two is 0.85) and is essentially identical to the MENTAL factor (the correlation between the two is 0.97). It is also highly correlated with the INTERACTING1 factor (which measures management and supervisory activities). The AA INTERPERSONAL variable is highly correlated with these same variables (the correlations with COGNITIVE1, MENTAL, and INTERACTING1 are 0.67, 0.79, and 0.89, respectively). The AA ROUTINE COGNITIVE variable is highly correlated with the STRUCTURAL2 factor, which measures task repetitiveness (correlation = 0.88). The AA ROUTINE MANUAL and NON-ROUTINE MANUAL attributes are both highly correlated with the PSYCHOMOTOR factor (the correlations are 0.83 and 0.94, respectively). Finally, we see from Table 4 that the AA OFFSHORABILITY attribute is negatively correlated with the PHYSICAL and SENSORY1 (which captures a range of visual and auditory skills) factors. But the correlations of -0.68 and -0.67 are not as high as the others noted above.

It is worth noting that the AA ROUTINE MANUAL and AA NON-ROUTINE MANUAL variables are highly correlated with each other (the correlation is 0.84), which would seem to cast doubt on whether these variables are really capturing distinct types of manual jobs. There are two possible interpretations of this. One is that the two variables are essentially measuring the same job attribute. The second is that manual jobs often have both a routine and a non-routine component. The latter bundling argument would also seem to explain why the AA ANALYTICAL and INTERPERSONAL variables are highly correlated with each other (the correlation is 0.75): jobs that have an analytical component also may often have a significant interpersonal component. (We will return to this point in the next section where we explicitly construct occupational groupings consisting of jobs that have similar job attributes.)

[8] We construct these variables as described on page 1163 of Acemoglu and Autor (2011). To be consistent with our factors, we then normalize the AA measures so the weighted (by 2016 OES occupational employment) mean is 0 and the variance is 1.

Table 5 presents factor medians by major occupation and by 2005 wage groupings. The variation across occupations agrees quite well with one's intuition. By way of illustration, we highlight a few of these below.

As can be seen in Table 5, EDUCATION is highest in professional occupations (management, business, science, and the arts) and lowest in the blue collar (production, transportation and moving) and service occupations. Not surprisingly, the pattern for COGNITIVE1 and MENTAL is similar. In contrast, PHYSICAL is highest in the trades and blue collar occupations and lowest in the professional occupations. STRUCTURAL1 is highest in professional occupations and next highest in trades occupations (natural resources, construction, and maintenance), although there is less variation in this variable across occupation groups than in the preceding factors mentioned. STRUCTURAL2 is highest in sales and office and blue collar, with an overall level of variation across groups that is similar to STRUCTURAL1. Trades occupations are by far the most hazardous, as indicated by CONDITIONS1. The next most hazardous are the blue collar jobs, and the remaining occupations are all similar in their job hazards. OUTPUT1 follows a similar pattern.

It also of interest to categorize occupations on the basis of their 2005 wage and then see how the factors vary across the different wage groups. Toward this end, we have placed each occupation in one of three wage groups. The first group consists of occupations whose average wage is in the bottom 20 percent. The middle group consists of occupations whose average wage is between the 20th and 80th percentile. The top group consists of occupations whose average wage is in the upper 20 percent. Table 5 presents the relationship between 2005 wages and our variables of interest.

As expected, EDUCATION is highest for the top wage group, next highest for the middle group, and lowest for the bottom group. The same is true for COGNITIVE1 and MENTAL. The same pattern also holds for STRUCTURAL1 and INTERACTING1. Not surprisingly, PHYSICAL is clearly lowest in the top wage group. And PHYSICAL is lower in the bottom wage group than the middle. But interestingly, an examination of the data reveals that the PHYSICAL distribution functions for the bottom and middle wage groups cross at around a probability of 0.6. While a significant portion of occupations with a high physical requirement are in the bottom wage group, a substantial portion are in the middle group. It is also the case that occupations in the middle wage group tend to be more hazardous than those in either the top or the lower wage groups, as indicated by the CONDITIONS1 factor. (While the median value of CONDITIONS1 is slightly negative for the middle group, there is a significant right tail to the distribution which drives up the mean for this group, but keeps the median below zero).

Wage Regressions

We now estimate a wage regression in which the factors and education are the explanatory variables. In addition to estimating an equation without interactions, we also estimate an equation in which the factors are interacted. To facilitate interpretation of the latter, we transform the factors into variables that are non-negative (recall that the factors have mean 0).[9] The first column in Table 6 presents the non-interacted equation. Note that the R squared in the equation is 0.861. Recall that when education and all of the O*NET variables are included in the equation, the R squared is 0.937. Thus, little information is lost when the 148 individual O*NET variables are replaced by the 21 factors. One can test this formally since the factor scores are simply linear combinations of the O*NET variables. One cannot reject the hypothesis that the implied restrictions on the O*NET variables in the wage equation are invalid.

Factor analysis yields a method for aggregating the individual job attributes into broader categories that can be used to explain occupational wage variation. One question that comes to mind is

[9] Specifically, letting $minF$ denote the minimum value of the factor, the transformed factor is simply $F_i^* = F_i - minF$ for all i.

whether a simpler method might work just as well. An obvious alternative is to simply take an average across all of the variables in each O*NET category and then insert these averages as explanatory variables in the wage equation. When one does this, one obtains an R squared of 0.80. So while simple averages across the O*NET categories explain most of the variation in occupational wages, the factors do have greater explanatory power.

Examining the estimated coefficients and standard errors, we see that EDUCATION and COGNITIVE1 both have a positive effect on occupational wages. Other things the same, less structured jobs (STRUCTURAL1) pay more, as do jobs that involve the use of computers (OUTPUT2). It is also interesting to note that, other things the same, unpleasant, hazardous jobs (CONDITIONS1) pay more. This finding is noteworthy in light of the fact that researchers have found it notoriously difficult to find compensating wage differentials. Hazardous, unpleasant jobs tend to have other characteristics that are associated with lower wages. We suspect that we are able to tease out a positive effect for this variable because we are able to control for these other characteristics (still, see the discussion immediately below).

Note that a couple of factors in the wage equation have negative coefficients that appear puzzling at first sight. *Other things the same*, why should jobs that require more physical skills or certain sensory skills (SENSORY2) pay less? An observation is in order before answering this. As others have noted, and as the correlations in Table 3 indicate, skills and tasks tend to occur in combinations or, in other words, tend to be bundled. The coefficients in the wage equation therefore need to be interpreted with care. The coefficient for a given factor indicates the average wage return associated with that factor after taking into account the other factors the factor in question tends to be associated with. In reality, it is generally not possible for one factor to change by itself, so the factor’s coefficient tells us the result of a “what if” experiment that is impossible to perform.

The implicit assumption behind the equation in column 1 is that we can isolate the wage return associated with a given factor in isolation from the other factors. In other words, the equation does not allow for possible interaction effects. A crucial feature of any job is the amount of cognitive skills that it requires. We therefore choose to interact COGNITIVE1 with the other factors. Column 2 shows the wage equation that results when one keeps the interactions that are significant. Throwing out interactions that are not significant makes the estimated equation easier to interpret.[10] We see that the wage return associated with a job being unstructured is greater when the job also has a greater cognitive skill requirement. The same is true of jobs that require the use of a computer, interactions with external customers (INTERPERSONAL2), and direction of others (INTERPERSONAL3), and certain sensory skills. The opposite is the case for jobs requiring certain other sensory skills or greater physical strength. Note that the coefficient on PHYSICAL by itself is positive and significant and that on SENSORY2 is also positive (but not quite significant at the 5 percent level). So jobs requiring more physical skills offer a higher wage in cases where cognitive demands are low. But the return to physical skills falls as the cognitive skills required by a job increase.

V. Regression Tree

As noted above, jobs generally require that a variety of tasks be performed. Some tasks will tend to complement each other and therefore occur together. Conversely, others will generally not be seen together in the same job. Evidence of this is provided by the fact that the factors we have identified have some strong positive and negative correlations. A natural O*NET based classification system would be one that groups together occupations that have similar combinations of skills and job tasks. This suggests performing some type of cluster analysis on the factors that we have identified.

[10] We obtain this equation following an iterative procedure. We first interact all variables with COGNITIVE1. We then discard interactions that are not significant and re-estimate the equation. We then repeat the procedure until all interaction effects are significantly different from zero.

Clustering occupations into groups on the basis of the O*NET factors requires a metric that can be used to ascertain the similarity of factor combinations across groups. However, in addition to education, we have identified 21 distinct factors. When comparing any two occupations, some factor quantities may be quite similar and some may be quite different. What is needed is some way to weight the various factors. A natural approach is to weight factors by their effect on wages: a factor that has a large effect on wages would receive a higher weight than a factor that has a minor effect. This might suggest using the coefficients from a wage equation, but there is an important complication. The wage equation may well be nonlinear with important interaction effects.

In a preceding section, we interacted the COGNITIVE1 factor with other factors, and determined that several of these interactions had substantial effects. Not only may there be interactions involving factors besides COGNITIVE1, but there may be interactions involving more than two variables. Indeed, if bundling of skills and job tasks is important, this is likely to be the case. Furthermore, the interactions may vary at various points in the factor distributions. Estimating a regression equation that allows for all the possible interactions among the 21 factors plus education is an impossible task. Instead of estimating a regression equation, we therefore estimate a regression tree.

Although it can be computationally intensive to estimate, the idea behind the regression tree is conceptually simple. To set ideas, assume the wage in occupation i, w_i, is some function of explanatory variables $x_{1i}, x_{2i}, \ldots, x_{Ki}$ (in the present context, the x_{ki} variables are the 21 factors we have identified plus education):

(1) $$w_i = f(x_{1i}, x_{2i}, \ldots, x_{Ki})$$

The function f may be very complex and may also include an error term. Rather than attempting to estimate f, one can predict the wage by repeatedly bifurcating the set of occupations and taking means in the two groups.

For example, consider bifurcating on the first variable, x_{1i}. Specifically, choose a real number c_1 and define two sets of occupations $A(c_1) = \{i \mid x_{1i} \leq c_1\}$ and $B(c_1) = \{i \mid x_{1i} > c_1\}$. In addition, let

(2a) $\overline{w}_A(c_1) = \frac{\sum_{i \in A(c_1)} E_i \cdot w_i}{\sum_{i \in A(c_1)} E_i}$

(2b) $\overline{w}_B(c_1) = \frac{\sum_{i \in B(c_1)} E_i \cdot w_i}{\sum_{i \in B(c_1)} E_i}$

and form a predicted wage, $\widehat{w}_i(c_1)$, equal to $\overline{w}_A(c_1)$ or $\overline{w}_B(c_1)$ depending on whether the occupation is in set $A(c_1)$ or $B(c_1)$. The error in the predicted wage is given by

(3) $\varepsilon_i = w_i - \widehat{w}_i(c_1)$

and the unexplained variance in the wage is given by

(4) $Var(\varepsilon_i | c_1) = \frac{\sum_i E_i \cdot \varepsilon_i^2}{\sum_i E_i}$.

Let c_1^* denote the value of c_1 that minimizes the unexplained variance in (4).

Repeating this process, one can find values c_k^* and $Var(\varepsilon_i | c_k^*)$ for each factor k. Let k^* be the value of k that minimizes $Var(\varepsilon_i | c_k^*)$. The split on the variable k^* yields two broad groups with occupation i falling into the left (right) branch if $x_{k^*i} \leq (>) c_{k^*}^*$. By construction, the mean value of x_{k^*} in the left branch is lower than the mean value in the right branch. Whether the means of another variable x_j differ between the two branches will depend on whether x_j is correlated with x_{k^*}. If x_j is positively (negatively) correlated with x_{k^*}, then the mean value of x_j will be lower (higher) in the left branch than in the right branch.

The occupational groups $A(c_{k^*}^*)$ and $B(c_{k^*}^*)$ will likely both be large, but one can apply the above procedure again. Doing so allows one to break the set of occupations in $A(c_{k^*}^*)$ and $B(c_{k^*}^*)$ into two smaller subsets, yielding four groups in all. At the next level, one would have eight occupational groups. One can continue down any branch of the tree as far as is useful.[11]

Regression Tree Results

We now apply the regression tree procedure described above to the OES-O*NET data. COGNITIVE1 turns out to be the variable used in the first split of data. (Recall that this factor summarizes the reasoning as well as written and oral expression skills that are required for a job.) As shown in Table 7, the first bifurcation of the data yields two readily interpretable occupational groups. The first group (labelled A in the table) contains 484 occupations with a total employment in 2016 of about 92 million. The second group (labelled B in the table) consists of 304 occupations with a total employment in 2016 of about 49 million. The mean wage for the first group is markedly lower than that for the second group. In addition, the mean of the COGNITIVE1 factor is much lower in the first group. Besides COGNITIVE1, a number of variables differ between the two groups. Differences in means that are statistically significant at the 0.01 level are indicated by an asterisk. Our discussion below highlights some, but not all of these.

As seen in Table 7, the mean values of EDUCATION, MENTAL, INFORMATION (which measures the information inputs of jobs), INTERACTING1, and STRUCTURAL1 are all lower for occupation group A. In contrast, the means of PHYSICAL, PSYCHOMOTOR, and CONDITIONS1 are all higher in occupation A.

Table 8 shows the effect of splitting the two broad occupation groups, A and B, into four smaller groups. The 484 occupations in group A are split into a group (labelled A1) with 270 occupations and

[11] While we are primarily interested in using the regression tree as a classification tool, it is typically used as prediction device. The standard approach is to move far down the tree and then prune back to prevent over fitting. For this purpose, the data set is generally split into two pieces: a “training data set” is used to generate the tree and the remaining subset is used for cross-validation purposes. The tree is pruned back by eliminating branches that are least important in reducing the unexplained variance in the variable of interest. Typical applications involve very large data sets with a large number of variables. For a fuller discussion, see Varian (2014).

2016 employment of about 71 million and a group (labelled A2) with 214 occupations and 2016 employment of about 21 million. The mean wage in the first group is about 39 percent lower than in the second group. The mean value of COGNITIVE1 is similar for the two groups, but the higher paid group has jobs that are more challenging physically. Specifically, COGNITIVE2 (which summarizes awareness of and attention to the physical environment), PHYSICAL, PSYCHOMOTOR, OUTPUT1, and CONDITIONS1 are all higher in the higher paid group. Interestingly, MENTAL is higher as well.

The 304 occupations in occupation group B are split into a group (labelled B1) with 135 occupations and 2016 employment of about 28 million and a group (labelled B2) with 169 occupations and employment of about 21 million. The mean wage in the second group is 37 percent higher than that in the first group. EDUCATION and COGNITIVE1 are substantially higher in the higher paying group. Also higher are INFORMATION and MENTAL.

We next bifurcate each of the four occupation groups to get eight occupations. At this point, the occupations are still relatively large, but explain a great deal of wage and factor variation. Table 10 shows the R squared values from log wage and factor regressions on dummy variables for the occupation groups. The first split (two occupations) explains 60% of log wage variation, the first two splits (four occupations) explain 71% of log wage variation, and the first three splits (eight occupations) explain 78% of log wage variation. We see that certain factors are well-explained by the occupation groups, whereas the variation in other factors is not explained by the occupation groups. In particular, EDUCATION, COGNITIVE1, INFORMATION, MENTAL, and OUTPUT2 all have R squared values greater than 0.55, while INTERPERSONAL2 and STRUCTURAL2 have R squared values less than 0.10.

As table 9 indicates, there still is a fair amount of variation in wages and job characteristics above and beyond what is explained by the first eight occupation groups. We therefore choose to bifurcate further. The “best” stopping point is not entirely clear. One would like to get groups that are homogeneous, but still relatively large. We choose to stop bifurcating when a split would yield a group with fewer than 15 occupations. We end up with fifteen occupation groups. The extended tree is depicted in Figure 4. The fourth column of table 9 shows that 85% of variation in log wages is explained by our fifteen occupational groups. Table A.1 lists the five largest detailed occupations in the various regression tree groups.

Table 10 shows summary statistics for our fifteen occupation groups, which are ranked by their 2016 mean wage. The top row of the table shows which two-digit branch of the tree the occupation group is part of. For example, the lowest paid occupation (01) is part of the A1 branch of the tree while occupation 08 is part of the B1 branch. Table 10 lists the average real wage, the average level of education, and the average value of each of the twenty-one factors for each of the fifteen occupational groups. The vertical lines in the table divide the occupations into three main panels: the lowest paid occupation group (01) that stands on its own and makes up roughly 20 percent of the labor force, low to middle wage occupation groups (02 to 09) that make up about 60 percent of the labor force, and high wage occupation groups (10 to 15) that constitute about 20% of the labor force (this was not by construction; the groups fell out naturally). The horizontal lines divide factors that appear to be correlated into distinct groups. Also included in Table 10 are mean values for the six Acemoglu-Autor variables. Figures 5.a and 5.b present box and whiskers diagrams for a few variables that seem to us to be of particular interest and that we discuss below.

Eyeballing the table and figures, we see, not surprisingly, that occupational groups with higher wages generally require more education; a few exceptions will be noted below. Basic cognitive skills, as measured by COGNITIVE1, and decision making skills, as measured by MENTAL, follow a similar pattern, as do supervising activities, as measured by INTERACTING1. The same can be said of the STRUCTURAL1 variable, which measures the extent to which jobs are not routine. Jobs requiring physical skills, as measured by the PHYSICAL factor, and psychomotor skills, as measured by PSYCHOMOTOR, follow a different pattern. The lowest paying jobs require a modest amount of physical skills. But many of the jobs closer to the middle part of the wage distribution require substantial physical and psychomotor skills, and tend to be characterized by more hazardous and less pleasant working conditions, as measured by CONDITIONS1. It is striking that jobs in the 09 occupational group require lower education but pay a higher wage than jobs in the 02, 04, and 06 groups. Also note that higher paying jobs generally are low in their physical skill requirements.

There appear to be two types of jobs in the middle paying occupation groups. As noted above, some of the jobs in this group are physical in nature (specifically, jobs in occupation groups 03, 05, 07, and 09). Others (jobs in occupation groups 04, 06, and 08) tend to require average or moderate cognitive skills and involve the use of the phone, writing emails, and general computer use (INTERPERSONAL1). This latter subset of jobs require less in the way of cognitive skills than the high paying jobs, but more than the bottom paying group or the middle paying jobs that require substantial physical skills.[12]

Earlier we observed that the Acemoglu-Autor ANALYTICAL and INTERPERSONAL variables were highly correlated across OES detailed occupations as were their ROUTINE MANUAL and NON-ROUTINE MANUAL variables. As can be seen in Table 10, this high correlation persists in our aggregate occupation groups. Occupational groups in which AA ANALYTICAL is high (low) also are groups in which AA INTERPERSONAL is high (low) and AA ROUTINE MANUAL and AA NON-ROUTINE MANUAL are low (high). This is further evidence that jobs consist of a bundling of various attributes and that analyzing attributes in isolation may be problematic.

[12] Note that STRUCTURAL2, which is a measure of task repetitiveness, does not vary all that much among the fifteen occupational groups, but takes on its highest value in group 06 and its third highest value in group 04. This is consistent with the AA ROUTINE COGNITIVE variable, which takes its highest value in group 06 and its next highest in group 04. These jobs thus have an important element of routineness, but their main distinction seems to be that they require moderate cognitive skills.

Note from Table 9 that when one estimates a wage equation where the explanatory variables are dummies for the twenty-two major 2-digit SOC occupations, one obtains an R squared of 0.67. This is notably lower than the R squared that results when the fifteen groups from the regression tree are used as regressors. Besides showing the R squared that results when regressing the wage against the occupational groups, Table 9 also shows the R squared that results when regressing education and the various factors against the various occupational groupings. Interestingly, equations in which cognitive and decision making skills are the dependent variable have a higher R squared when our fifteen occupational dummies are the regressors than when the twenty-two major SOC groups are the regressors, indicating that cognitive and mental skills vary more systematically with our occupational groupings than with the SOC grouping. This presumably reflects the fact that the regression tree approach utilizes wages and that the cognitive and mental variables are closely associated with wages. In contrast, the SOC occupational groups do a somewhat better job of explaining physical and psychomotor skills.

It is not surprising that the regression tree groups yield a tighter fitting wage equation than the major SOC groups since wages are utilized in the regression tree splits. Of course, if one simply wants occupation groupings that yield a high R squared in the wage equation, there is a simpler approach: one need only rank occupations on the basis of their mean wages. We have done this and formed groups that are roughly the same size as the regression tree groups. These groups naturally do a great job of "predicting" wages, but are somewhat worse in explaining cognitive and mental job skills than the regression tree results. And they are especially poor in explaining physical skills. This is likely due to the fact that, as shown earlier, the wage equation is nonlinear in the physical skills, with interaction effects being important.

Our fifteen occupational groups were obtained without reference to the industries in which jobs are found. Table 11.a shows how average wages in our fifteen occupational groups vary across major industries. While there is some wage variation across industries, this is inordinately smaller than the wage variation across occupations. Regressing 2016 industry-occupation log wages on dummies for our fifteen occupational groups, one obtains an R squared of 0.84. Adding dummies for the industry sectors, the R squared only increases to 0.85 (when one only includes the industry variables, the R squared is 0.21). Adding interaction terms between the occupation groups and industry sectors still only increases the R squared to 0.88.

Table 11.b shows each occupation's share of industry employment for each of the sectors. Most notably, the bottom paying occupational group consisting of jobs with very low cognitive requirements and only modest physical skill requirements is used most intensively in the leisure and hospitality sector. Occupational groups where physical skills are more important figure prominently in construction, manufacturing, trade, and health services. Higher paying occupations with greater cognitive skill requirements tend to make up a higher share of employment in the information, financial, and professional services sectors.

Other things the same, large occupations will tend to have high representations in a given sector and small occupations will tend to have low representation. To correct for this, Table 11.c divides each occupation's share of industry employment by the occupation's share of total economy-wide employment. A value greater than 1 means that an occupation is over represented in a sector while a value less than 1 indicates that the occupation is under represented. The value of 4 for the bottom paying occupation in leisure and hospitality indicates how highly concentrated this group is in that sector. Similarly, high values for the middle paying occupations requiring physical skills tells us that these occupations are concentrated in natural resources, construction, and manufacturing. The higher paying occupations are concentrated in the information and financial services sector. The highest paying group, which is small in size and includes doctors, is very highly concentrated in health services.[13]

VI. Evolution over Time of Employment and Wages in the Occupation Groups

In Section II, we showed that the trend toward increasing employment and wage polarization was present in the 2005 – 2017 OES data. We now turn to the combined OES-O*NET data and examine how employment and wages in our fifteen occupation groups changed over time.[14]

We begin this section by briefly examining how the factors themselves have changed over time. The changes in the distributions of the factors from 2005-2017 are depicted in Table 12. To interpret the numbers in the table, consider COGNITIVE1. We first rank the occupations from lowest to highest in terms of their COGNITIVE1 score and continue adding occupations in the bottom group until 2005 employment in this group equals 20 percent of the employment in the economy as a whole.[15] We then add up the employment in these occupations in 2017, calculate their percentage of total employment, and subtract out 20%. From Table 12, one sees that employment in these occupations amounted to

---

[13] For comparison purposes, we perform the same calculations as in Table 11.c, but use two-digit SOC aggregate occupations in place of the occupations we obtained through the regression tree. The results are shown in Table 11.d.

[14] The O*Net data are updated slowly and partially over time. Like other users of O*Net and the DOT, we therefore must assume that the values of the O*Net variables in each of the detailed occupations are fixed over time (in our case at their 2016 values). Thus, the changes in factors that we document below stem solely from changes in employment across the detailed occupations. Black and Spitz-Oener determine that most of the task changes in their data occurred within occupation-industry cells and that these changes were largely associated with computerization. In assessing whether it is reasonable for us to take task values to be fixed within occupations, note two things. First, during the 1979-1999 period, there was widespread computer adoption in both the U.S. and Germany. However, by 2005, computer usage was very widespread and the trend toward increased computerization had slowed considerably. Second, our occupations are much more detailed that those in Back and Spitz-Oener. There are 788 detailed occupations in our analysis and 92 in Black and Spitz-Oener's. (By way of example, there are four different types of secretaries in our data). Employment shifts across closely related detailed occupations in our analysis would not show up at all in that of Black and Spitz-Oener.

[15] Since occupational employment is lumpy, we split the employment of the "marginal" occupation – the occupation where cumulative employment first exceeds 20% – proportionately into the bottom group and the middle group so that the bottom group contains exactly 20 percent of employment in 2005. We do the same for the median and the 80th percentile.

only 19.14 (20-0.86) percent of the labor market in 2017. The other numbers in columns 2 – 5 are calculated similarly. The final column in Table 12 shows the change in factor means from 2005 to 2017. It is important to remember that we are only measuring factor values in each detailed occupation at one point in time, so any changes are completely driven by changes in the occupational employment over this time period.

Looking at Table 12, we see that there has been a shift toward jobs requiring more education and more cognitive skills and tasks (viz. COGNITIVE1 and MENTAL) and away from jobs that require more physical skills (viz. PHYSICAL and PSYCHOMOTOR). There have been shifts toward jobs requiring general computer usage (OUTPUT2) and away from jobs that use heavier machinery (OUTPUT1) and that are more hazardous (CONDITIONS1). One also see that there have been shifts toward non-routine jobs that require independent decision making (viz. STRUCTURAL1) and management skills (INTERACTING1) and away from less routine, more repetitive jobs (viz. COGNITIVE3 and STRUCTURAL2). The Acemoglu-Autor variables tell a similar story. Employment has shifted into occupations that have higher levels of AA ANALYTICAL and AA INTERPERSONAL and away from occupations with higher levels of AA ROUTINE COGNITIVE, AA NONROUTINE MANUAL, and especially AA ROUTINE MANUAL.

The fifteen occupation groups generated by the regression tree analysis allow us to calculate simple counterfactual statistics that shed light on the changing nature of employment and wages over the recent past. The first panel in Table 13 depicts each occupation's total employment in 2005 and 2017, as well as the difference between the two levels. In addition, since our occupations are sorted by wage, we compute the cumulative difference in employment as we move from the lowest paying occupations to the highest paying occupations. The table shows a big increase, nearly 4 million jobs, in the employment of the lowest paying occupations. We then see relatively small changes for the next eight occupations, so that the nine lowest paying occupations grew by roughly 4.6 million jobs from 2005 to 2017. The next five higher paying occupations all grew substantially while the highest paid occupation grew slightly. In total, employment grew by approximately 11.5 million jobs from 2005 to 2017, slightly more than 34 percent of this increase is in the lowest paying occupation and nearly 60 percent is in the six highest paying occupations.

The second panel presents the occupational employment shares in 2005 and 2017 as well as the difference and cumulative difference between the two years. The pattern of employment share changes follows the same pattern as employment level changes. The share of employment in the lowest paying occupation grew by 1.3% from 18.3% to 19.6% of total employment. The employment shares of the next eight occupations all declined from 2005 to 2017. In total, the share of employment in these eight occupations declined by 4.4%. Since employment shares must sum to 100 percent, the shift away from the 2nd through 9th occupations must be completely accounted for by shifts into the lowest paying occupation and the six highest paying occupations. It follows that share of employment in the six highest paying occupations grew by 3.1%.

The third panel shows the counterfactual employment levels holding total employment fixed at the 2017 level, but using 2005 and 2017 employment shares. The pattern of the counterfactual employment levels is exactly the same as the pattern of the employment shares depicted in the previous panel. The third panel translates the shifts in employment shares into employment numbers. The reductions in the employment shares of the 2nd through 9th occupations implied a loss of slightly more than 6.2 million jobs from 2005 to 2017. The increase in its employment share implied a gain of slightly more than 1.8 million jobs in the lowest paying occupation while the increases in the employment shares of the six highest paying occupations implied a gain of slightly more than 4.4 million jobs.

The bottom row in the third panel can be graphed as the curve labelled “Total” in Figure 6. The hollowing out phenomenon is striking. Shifts in occupation employment shares between 2005 and 2017 imply an increase in employment in occupational group 01 offering a mean real wage below \$12.07, a reduction in employment in occupational groups paying a mean wage above \$12.07 and below \$26.67, and an increase in employment in occupational groups paying a mean wage greater than \$26.67. Recall that the bottom paying occupational group generally consists of jobs requiring little in the way of cognitive, physical, or other skills, the middle paying occupational groups experience employment declines are largely made up of jobs utilizing greater physical skills or moderate cognitive skills, and the top paying occupational groups largely consist of jobs with greater cognitive skill requirements (see Table 10).

An occupation's share of economy-wide employment may be increasing (decreasing) for either of two reasons. First, the occupation may be heavily concentrated in those industries that are growing the fastest (slowest). Second, the occupation may be making up a greater (lesser) share of employment in a host of industries. It is useful to distinguish between these reasons. To this end, let s_{ijt} denote the employment share of sector j and year t employment in occupation i. That is, letting E_{ijt} denote occupation i employment in sector j in year t, E_{jt} denote sector j employment in year t, and E_t denote total employment in year t. It follows that

$$(5) \quad s_{it} = \sum_j \frac{E_{ijt}}{E_t} = \sum_j \frac{E_{ijt}}{E_t} \times \frac{E_{jt}}{E_{jt}} = \sum_j \frac{E_{ijt}}{E_{jt}} \times \frac{E_{jt}}{E_t} = \sum_j s_{ijt} \times s_{jt}$$

where $s_{ijt} = \frac{E_{ijt}}{E_{jt}}$ and $s_{jt} = \frac{E_{jt}}{E_t}$.

The change in the employment share of occupation i from year 0 to year t is given by

$$(6) \quad s_{it} - s_{i0} = \sum_j s_{j0} \times \left(s_{ijt} - s_{ij0}\right) + \sum_j s_{ijt} \times \left(s_{jt} - s_{j0}\right)$$

The first term in the above equation is the change in the employment share of occupation i stemming from it being used more or less intensively in the various industries. The second term is the change in the occupation's share of overall employment stemming from changes in the industrial composition of employment. If the employment shares of the industries in which the occupation is used most intensively increase (decrease), then this term will be positive (negative).

The changes in employment implied by the shifting composition of employment across industries is portrayed in the curve labelled “Change in industry composition holding staffing pattern fixed at 2017 levels” in Figure 6. One sees that there has been a significant shift in the composition of employment toward industries employing the bottom paying occupation group 01. This reflects overall employment growth in the Leisure and Hospitality sector, a sector in which the bottom paying occupation is very heavily over-represented, as we have seen. The increased employment share of industries employing the bottom paying occupation has caused employment in this occupation group to increase by 1,818,000. The industry composition effect on employment in all of the other occupation groups has been either relatively flat or slightly negative.

Changes in employment caused by shifting staffing patterns are summarized by the curve labelled “Change in staffing patterns holding industry composition fixed at 2005 levels” in Figure 6. One sees that the effect of a shifting staffing pattern has been to reduce employment in all occupation groups with below average wages. In the case of the lowest paying group 01, the modest shift away from the bottom paying group has been strongly outweighed by the increase in employment caused by the greater employment share of industries that are heavy users of this occupation group. However, for all other employment groups in the bottom half of the wage distribution, the industry composition effect has been much weaker and either reinforces or does not outweigh the fall in employment share resulting from shifts in staffing patterns. Since changes in employment stemming from staffing pattern changes must sum to zero, employment in the top paying groups must have risen as a result of staffing pattern changes. Indeed, as can be seen in Figure 6, the staffing pattern effect is positive for all occupations paying above average wages.

Of course, the 2005 – 2017 period was punctuated by the Great Recession. It is natural to break the entire period into two sub-periods, the first from 2005 – 2010 and the second from 2010 – 2017. Figures 7.a and 7.b present the employment decomposition in the two sub-periods. Interestingly, the basic patterns looks quite similar. The employment shifts observed over 2005 – 2017 do not appear to simply be an artifact of the recession.

Let us now examine how wages changed over the 2005-2017 period. The average real wage increased by 6.2% over the period. This was a reflection partly of the shifts in employment across occupational groups discussed above and partly of changes in the real wage received by each occupational group. The fourth panel in Table 13 shows the real mean wages (in 2017 dollars) for the 15 occupations in 2005 and 2017 as well as the percent difference between the two years. Holding employment fixed at 2005 levels, average real wages would have increased by 3.4%. We see that the wages for the highest paying occupational groups 11 – 15 increased by more than the overall average increase of 3.4%. In contrast, the real wages for all but two of the lower and middle paying occupation groups 01 – 10 increased by less than the overall average. Interestingly, the percentage wage increase for the occupation group 01 exceeded that for all but one of the low and middle paying groups.

As noted above, the vertical lines in Table 13 roughly divide up employment into larger groups making up 20, 60, and 20 percent of the labor force. On average, wages in the bottom occupational group making up almost 20% of the labor force increased by 4.0%. Average wage growth in occupation groups 02 through 09 making up the middle 60 percent of the labor force was 1.5% and average wage growth in occupations 10 through 15 composing about 20 percent of the labor force was 5.8%.

As shown above, the OES-O*NET data provide evidence of both employment and wage polarization during the period 2005-2017. We conclude our analysis by looking at the variance of log wages over this period. The curve labelled “Total” in Figure 8 depicts the variance of log wages for each year from 2005 to 2017. This variance increased from 2005 to 2014 and then fell from 2014 to 2017.

It is straightforward to decompose the change in the variance into a component reflecting the effect of shifts in employment across occupational groups and a component reflecting variations in wages. The curve labelled “From change in employment shares” in Figure 8 depicts the change in the variance of log wages holding occupational wages fixed, but allowing employment in the occupations to change. The curve labelled “From change in within group variance” shows the effect of wage changes within occupational groups while the curve labelled “From change in between group variance” shows the effect of wage changes across occupation groups.

Looking at Figure 8, we see that shifting employment across occupations caused the variance in occupational wages to increase throughout the entire 2005 – 2017 period. Changes in average wages across the fifteen occupational groups also caused wage variance to increase through 2014. However, from 2014 on, changes in wages across occupations had a negative effect on wage variance. This effect, which primarily sprang from increases in relative wages at the bottom end of the wage distribution, outweighed the effect from employment shifts, causing the overall variance of log wages to fall in the last three years of the sample period. Finally, note that holding employment as well as average wages within occupational groups constant, changes in the distribution of detailed occupation wages within the occupational groups had practically no effect on the variance of log wages throughout the entire period.

V. Conclusion

The analysis in this paper demonstrates the payoff to combining the O*NET and OES data sets. OES is an excellent source of annual information on occupational employment and wages in the United States. O*NET is a rich source of information on occupational characteristics. We have used factor analysis to condense the O*NET information. To aid in the interpretation of the factors, we perform separate factor analyses for the various O*NET categories. Not surprisingly, there are high correlations among many of the factors. In addition to looking at the correlations among the factors we obtain, we also look at correlations involving the Acemoglu and Autor variables, which have become the standard in the literature. Interestingly, there is a very high correlation between the Acemoglu and Autor ROUTINE MANUAL and NON-ROUTINE MANUAL variables as well as between the Acemoglu and Autor ANALYTICAL and INTERPERSONAL variables. Either the variables in each case are measuring the same job attribute or they are measuring different attributes that are bundled together.

As a second methodological innovation, we use a regression tree to form broader occupational groups. While researchers have noted that jobs are best viewed as bundles of various activities requiring a number of complementary skills, they have not generally fully incorporated this feature into their analysis. Clustering occupations into groups on the basis of job characteristics requires a metric that can be used to ascertain the similarity of combinations of characteristics across groups. A natural approach is to weight characteristics by their effect on wages. Bundling also means that a wage equation is likely to be highly nonlinear with important interaction effects. Not only is a regression tree well suited for estimating this equation, but the estimation process naturally sorts occupations into larger groupings.

The O*NET variables taken altogether explain a high proportion of the observed variation in occupational wages. Not much wage information is lost when the individual O*NET variables are replaced by the factors. And very little wage information is lost in our final step where we use a regression tree to obtain fifteen occupational aggregates. The twenty-two major 2-digit SOC occupations are less successful in explaining occupational wage variation. They also are more heavily industry based than the occupational aggregates obtained through the regression tree.

Our OES-ONET data cover the period from 2005 to 2017. Employment and wage trends in the OES are similar to those found in other data sets. The hollowing out phenomenon pointed out by others for earlier years shows up clearly in the OES. It also is picked up quite clearly in the occupational groups that we have uncovered through our factor analysis and regression tree. Two distinct forces have been at work. Changing staffing patterns have led to employment shifts away from jobs with low or moderate cognitive skill requirements and toward jobs with greater cognitive skills requirements. Changes in the composition of industry employment have led to a shift toward employment in low wage jobs requiring little skill. As a consequence, the employment share of a large occupational group constituting about 20 percent of the labor force and requiring little in the way of cognitive, physical, or other skills and paying low wages increased over time. The employment share of the top paying occupational groups consisting of jobs with greater cognitive requirements also increased during the 2005-2017 period. In contrast, the share of middle paying occupational groups utilizing greater physical skills or moderate cognitive skills fell.

Wage variance increased from 2005 to 2014. Changes in employment shares across occupations caused the variance of wages to increase as did changes in wages across the various occupation groups. From 2014 to 2017, increases in the relative wage received by the bottom paying occupation group caused wage variance to fall, outweighing the increase stemming from the changes in employment shares across industries.

We believe the occupational aggregates we have obtained by applying factor analysis and a regression tree to the OES-O*NET data may have uses in other applications. Potential applications include analyses of occupational mobility, gender wage differences, and area wage differentials. We plan to explore this in future work.

References

Acemoglu, Daron and Autor, David, “Skills, Tasks and Technologies: Implications for Employment and Earnings.” Handbook of Labor Economics (2011), Ch. 12, pp. 1044 – 1169.

Autor, David ,” Why Are There Still So Many Jobs? The History and Future of Workplace Automation.” *Journal of Economic Perspectives*, Volume 29, Number 3 (Summer 2015), pp. 3–30.

Autor, David and Handel, Michael J., “Putting Tasks to the Test: Human Capital, Job Tasks, and Wages.” *Journal of Labor Economics*, Vol. 31, No. 2, The Princeton Data Improvement Initiative (Part 2, April 2013), pp. S59-S96.

Autor, David, Levy, Frank and Murnane, Richard J., “The Skill Content of Recent Technological Change: An Empirical Exploration.” *The Quarterly Journal of Economics*, Vol. 118, No. 4 (Nov., 2003), pp. 1279-1333.

Black, Sandra E. and Alexandra Spitz-Oener, “Explaining Women’s Success: Technological Change and the Skill Content of Women’s Work.” *The Review of Economics and Statistics*, Vol. 92, No. 1 (February 2010), pp. 187-194.

Gathman and Schönberg, “How general is specific human capital? A task-based approach.” *Journal of Labor Economics* 28, no. 1 (2017), pp.1–50.

Goos, Maarten, and Alan Manning, “The Rising Polarization of Work in Britain.” *The Review of Economics and Statistics*, Vol. 89, No. 1 (Feb., 2007), pp. 118-133.

Handwerker, Elizabeth Weber and Spletzer, James R., “Measuring the distribution of wages in the United States from 1996 through 2010 using the Occupational Employment Survey.” *Monthly Labor Review* (May 2014), pp. 1-19.

Heckman, James, and Jose Scheinkman. “The importance of bundling in a Gorman-Lancaster model of earnings.” *Review of Economic Studies* 54, no. 2, (1987) 243–55.

Ingram, Beth and Neumann, George R., “The Returns to Skill,” Working Paper, University of Iowa (2010).

Poletaev, Maxim and Robinson, Chris, “Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000.” *Journal of Labor Economics*, Vol. 26, No. 3 (July 2008), pp. 387-420.

Robinson, Chris, “Occupational Mobility, Occupation Distance and Specific Human Capital.” *Journal of Human Resources*, 2017.

Varian, Hal R.,”Big Data: New Tricks for Econometrics.” *Journal of Economic Perspectives*, Volume 28, Number 2 (Spring 2014), pp. 3–28.

Yamaguchi, Shintaro, “ Tasks and Heterogeneous Human Capital.” *Journal of Labor Economics*, Vol. 30, No. 1 (January 2012, pp. 1-53.)

Table 1. O*NET variable list

O*NET Category	Loading factors 1	2	3	4	Standardized scoring coefficients 1	2	3	4	2016 log wage regression results Estimate	SE	t-stat	Pr > |t|
Abilities: Cognitive Abilities (COGNITIVE)												
Oral Comprehension	92	19			0.1212	-0.0606			0.0198	0.0454	0.44	0.66
Written Comprehension	94	16			0.1172	-0.0770			0.0215	0.0419	0.51	0.61
Oral Expression	92	14			0.0771	-0.0649			-0.0409	0.0398	-1.03	0.30
Written Expression	95	10			0.1941	-0.2125			0.0824	0.0359	2.30	0.02
Fluency of Ideas	87	29			0.1463	-0.0592			-0.0878	0.0473	-1.85	0.06
Originality	84	29			0.0504	0.0488			0.0670	0.0457	1.47	0.14
Problem Sensitivity	78	45			0.0391	0.0579			0.1816	0.0403	4.51	0.00
Deductive Reasoning	90	34			0.0578	0.0391			0.0522	0.0492	1.06	0.29
Inductive Reasoning	86	35			0.0928	0.0109			0.0702	0.0429	1.63	0.10
Information Ordering	78	47			0.0042	0.1370			-0.0407	0.0419	-0.97	0.33
Category Flexibility	82	35			0.0500	-0.0228			-0.1504	0.0428	-3.52	0.00
Mathematical Reasoning	83	22			0.1114	-0.0482			0.0729	0.0354	2.06	0.04
Number Facility	78	23			0.0712	-0.0360			-0.0615	0.0352	-1.74	0.08
Memorization	72	30			0.0373	0.0101			0.0143	0.0325	0.44	0.66
Speed of Closure	64	62			-0.0032	0.1610			0.0143	0.0329	0.43	0.67
Flexibility of Closure	54	69			-0.0283	0.1702			-0.0813	0.0346	-2.35	0.02
Perceptual Speed	26	83			-0.0588	0.2643			0.0694	0.0374	1.85	0.06
Spatial Orientation	-51	55			-0.0606	0.1439			-0.1017	0.0299	-3.40	0.00
Visualization	14	78			-0.0913	0.2223			-0.0216	0.0281	-0.77	0.44
Selective Attention	34	75			-0.0563	0.1719			-0.0103	0.0431	-0.24	0.81
Time Sharing	19	58			-0.0484	0.1434			-0.0718	0.0354	-2.03	0.04

Table 1. O*NET variable list (continued)

	Loading factors				Standardized scoring coefficients				2016 log wage regression results			
O*NET Category	1	2	3	4	1	2	3	4	Estimate	SE	t-stat	Pr > |t|
Abilities: Psychomotor Abilities (PSYCHOMOTOR)												
Arm-Hand Steadiness	89				0.1573				-0.0062	0.0299	-0.21	0.84
Manual Dexterity	91				0.2161				0.0565	0.0297	1.90	0.06
Finger Dexterity	73				0.0164				-0.0415	0.0265	-1.56	0.12
Control Precision	93				0.0926				-0.0049	0.0252	-0.20	0.85
Multilimb Coordination	92				0.0938				0.0329	0.0238	1.38	0.17
Response Orientation	92				0.1603				-0.0645	0.0284	-2.27	0.02
Rate Control	90				0.1491				0.0260	0.0305	0.85	0.39
Reaction Time	91				0.1545				-0.0315	0.0299	-1.05	0.29
Wrist-Finger Speed	76				0.0178				0.0175	0.0188	0.93	0.35
Speed of Limb Movement	83				0.0356				-0.0115	0.0260	-0.44	0.66
Abilities: Physical Abilities (PHYSICAL)												
Static Strength	97				0.1671				-0.0534	0.0277	-1.93	0.05
Explosive Strength	58				0.0282				0.0313	0.0245	1.28	0.20
Dynamic Strength	97				0.2353				0.0720	0.0329	2.19	0.03
Trunk Strength	92				0.0387				-0.0053	0.0247	-0.22	0.83
Stamina	97				0.2268				-0.0422	0.0353	-1.20	0.23
Extent Flexibility	94				0.0691				-0.0762	0.0250	-3.05	0.00
Dynamic Flexibility	64				0.0205				0.1081	0.0382	2.83	0.00
Gross Body Coordination	97				0.1839				0.0942	0.0383	2.46	0.01
Gross Body Equilibrium	91				0.0762				-0.1007	0.0311	-3.24	0.00

Table 1. O*NET variable list (continued)

O*NET Category	Loading factors 1	2	3	4	Standardized scoring coefficients 1	2	3	4	2016 log wage regression results Estimate	SE	t-stat	Pr > |t|
Abilities: Sensory Abilities (SENSORY)												
Near Vision	8	64			0.0366	0.1116			0.1362	0.0332	4.10	0.00
Far Vision	69	34			0.1035	0.1702			0.0440	0.0271	1.62	0.11
Visual Color Discrimination	73	12			0.1851	0.1673			0.0116	0.0238	0.49	0.63
Night Vision	85	-34			0.1527	-0.0357			0.1207	0.0505	2.39	0.02
Peripheral Vision	84	-37			0.1346	-0.2624			-0.0425	0.0464	-0.92	0.36
Depth Perception	81	-27			0.1266	0.0129			-0.0012	0.0241	-0.05	0.96
Glare Sensitivity	84	-37			0.1047	-0.0764			0.0875	0.0324	2.70	0.01
Hearing Sensitivity	77	3			0.1094	0.0345			0.0351	0.0250	1.41	0.16
Auditory Attention	74	8			0.1056	0.0642			0.0735	0.0238	3.09	0.00
Sound Localization	88	-28			0.1985	0.0832			0.0312	0.0383	0.81	0.42
Speech Recognition	-20	84			0.0594	0.4061			-0.0770	0.0384	-2.01	0.05
Speech Clarity	-13	83			0.0613	0.3433			-0.0298	0.0385	-0.77	0.44
Work Activities: Information Input (INFORMATION)												
Getting Information	72				0.1759				0.0271	0.0224	1.21	0.23
Monitor Processes, Materials, or Surroundings	88				0.3227				0.0116	0.0219	0.53	0.60
Identifying Objects, Actions, and Events	86				0.2939				-0.0047	0.0199	-0.23	0.82
Inspecting Equipment, Structures, or Material	56				0.1320				-0.0138	0.0178	-0.77	0.44
Estimating the Quantifiable Characteristics of Products, Events, or Information	80				0.2133				-0.0427	0.0223	-1.91	0.06

Table 1. O*NET variable list (continued)

O*NET Category	Loading factors 1	2	3	4	Standardized scoring coefficients 1	2	3	4	2016 log wage regression results Estimate	SE	t-stat	Pr > |t|
Work Activities: Mental Processes (MENTAL)												
Judging the Qualities of Things, Services, or People	79				0.0577				0.0154	0.0200	0.77	0.44
Processing Information	89				0.1119				-0.0291	0.0233	-1.25	0.21
Evaluating Information to Determine Compliance with Standards	82				0.0536				0.0375	0.0166	2.25	0.02
Analyzing Data or Information	94				0.1620				0.0410	0.0237	1.73	0.08
Making Decisions and Solving Problems	94				0.1761				0.0175	0.0236	0.74	0.46
Thinking Creatively	87				0.0703				-0.0057	0.0205	-0.28	0.78
Updating and Using Relevant Knowledge	91				0.0995				-0.0188	0.0231	-0.81	0.42
Developing Objectives and Strategies	91				0.1533				0.0698	0.0212	3.30	0.00
Scheduling Work and Activities	87				0.1012				-0.0185	0.0192	-0.96	0.34
Organizing, Planning, and Prioritizing Work	90				0.1069				-0.0203	0.0251	-0.81	0.42
Work Activities: Work Output (OUTPUT)												
Performing General Physical Activities	70	-56			0.1588	-0.1151			-0.0270	0.0196	-1.37	0.17
Handling and Moving Objects	67	-63			0.0225	-0.4535			0.0318	0.0180	1.77	0.08
Controlling Machines and Processes	88	-27			0.1943	0.0770			-0.0015	0.0195	-0.08	0.94
Operating Vehicles, Mechanized Devices, or Equipment	77	-27			0.0725	-0.0094			0.0204	0.0211	0.96	0.33
Interacting With Computers	-9	85			0.1467	0.3980			0.0370	0.0158	2.34	0.02
Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment	72	26			0.0999	0.1258			0.0393	0.0141	2.80	0.01
Repairing and Maintaining Mechanical Equipment	93	-20			0.4311	-0.0113			-0.0270	0.0189	-1.43	0.15
Repairing and Maintaining Electronic Equipment	84	18			0.1592	0.2656			-0.0112	0.0168	-0.67	0.50
Documenting/Recording Information	5	64			0.0554	0.1740			-0.0493	0.0187	-2.63	0.01

Table 1. O*NET variable list (continued)

O*NET Category	Loading factors 1	2	3	4	Standardized scoring coefficients 1	2	3	4	2016 log wage regression results Estimate	SE	t-stat	Pr > |t|
Work Activities: Interacting with Others (INTERACTING)												
Interpreting the Meaning of Information for Others	68	39			0.0598	0.0093			-0.0374	0.0198	-1.89	0.06
Communicating with Supervisors, Peers, or Subordinates	74	39			0.0876	0.0147			0.0420	0.0242	1.73	0.08
Communicating with Persons Outside Organization	43	79			-0.1183	0.4248			0.0211	0.0190	1.11	0.27
Establishing and Maintaining Interpersonal Relationships	50	66			-0.0170	0.1699			-0.0288	0.0217	-1.32	0.19
Assisting and Caring for Others	25	28			-0.0139	0.0762			-0.0166	0.0166	-1.00	0.32
Selling or Influencing Others	23	67			-0.1051	0.2238			0.0023	0.0150	0.15	0.88
Resolving Conflicts and Negotiating with Others	64	61			-0.0322	0.2063			-0.0496	0.0177	-2.80	0.01
Performing for or Working Directly with the Public	-6	74			-0.1330	0.2953			-0.0163	0.0128	-1.27	0.20
Coordinating the Work and Activities of Others	89	17			0.1651	-0.1482			0.0611	0.0220	2.78	0.01
Developing and Building Teams	88	25			0.1432	-0.0551			-0.0080	0.0263	-0.31	0.76
Training and Teaching Others	81	22			0.0884	-0.0479			-0.0794	0.0206	-3.85	0.00
Guiding, Directing, and Motivating Subordinates	90	22			0.2712	-0.1821			0.0437	0.0211	2.08	0.04
Coaching and Developing Others	83	31			0.1095	0.0051			-0.0157	0.0227	-0.69	0.49
Provide Consultation and Advice to Others	82	37			0.1604	-0.0402			0.0727	0.0173	4.20	0.00
Performing Administrative Activities	56	52			0.0127	0.0695			-0.0249	0.0176	-1.41	0.16
Staffing Organizational Units	82	29			0.0831	0.0056			-0.0146	0.0167	-0.87	0.38
Monitoring and Controlling Resources	79	20			0.0891	-0.0721			-0.0211	0.0159	-1.32	0.19

Table 1. O*NET variable list (continued)

O*NET Category	Loading factors 1	2	3	4	Standardized scoring coefficients 1	2	3	4	2016 log wage regression results Estimate	SE	t-stat	Pr > |t|
Work Context: Interpersonal Relationships (INTERPERSONAL)												
Public Speaking	37	15	35		0.0317	0.0292	0.0226		0.0277	0.0207	1.34	0.18
Telephone	82	27	5		0.2380	0.1067	-0.1101		-0.0041	0.0209	-0.19	0.85
Electronic Mail	87	-11	22		0.3754	-0.2542	0.0900		-0.0034	0.0187	-0.18	0.85
Letters and Memos	80	17	18		0.1541	0.0297	-0.0358		0.0078	0.0208	0.38	0.71
Face-to-Face Discussions	48	6	46		0.0570	-0.0380	0.0984		-0.0715	0.0348	-2.06	0.04
Contact With Others	53	55	10		0.1216	0.1068	-0.0096		0.0840	0.0354	2.37	0.02
Work With Work Group or Team	34	22	62		0.0443	-0.0071	0.1531		-0.1005	0.0341	-2.95	0.00
Deal With External Customers	49	63	-6		0.0965	0.1521	-0.1113		-0.0186	0.0216	-0.86	0.39
Coordinate or Lead Others	36	17	78		0.0204	-0.0353	0.3482		0.0271	0.0293	0.92	0.36
Responsible for Others' Health and Safety	-51	25	63		-0.1905	0.0968	0.2770		0.0335	0.0231	1.45	0.15
Responsibility for Outcomes and Results	0	0	78		-0.0666	-0.1221	0.2713		0.0748	0.0245	3.05	0.00
Frequency of Conflict Situations	30	67	44		0.0154	0.2053	0.1483		-0.0296	0.0294	-1.01	0.31
Deal With Unpleasant or Angry People	5	89	5		-0.0935	0.4577	-0.1405		0.0538	0.0296	1.82	0.07
Deal With Physically Aggressive People	-11	68	21		-0.0823	0.1594	0.0099		-0.0944	0.0325	-2.90	0.00

Table 1. O*NET variable list (continued)

	Loading factors				Standardized scoring coefficients				2016 log wage regression results			
O*NET Category	1	2	3	4	1	2	3	4	Estimate	SE	t-stat	Pr > |t|
Work Context: Physical Work Conditions (CONDITIONS)												
Indoors, Environmentally Controlled	-65	-24	-14	13	-0.0337	-0.0009	-0.0203	0.0917	-0.0117	0.0158	-0.74	0.46
Indoors, Not Environmentally Controlled	84	13	1	-8	0.0959	-0.0060	-0.0868	-0.1095	-0.0072	0.0193	-0.37	0.71
Outdoors, Exposed to Weather	86	22	-29	-5	0.1758	0.1535	-0.4851	0.0199	0.0412	0.0229	1.80	0.07
Outdoors, Under Cover	84	16	-23	-2	0.0716	0.0136	-0.1301	-0.0370	-0.0697	0.0258	-2.70	0.01
In an Open Vehicle or Equipment	81	11	12	-8	0.0520	-0.0096	0.0209	-0.1091	-0.0137	0.0239	-0.57	0.57
In an Enclosed Vehicle or Equipment	75	-11	-36	-4	0.0577	0.0046	-0.1630	-0.0306	0.0358	0.0206	1.73	0.08
Physical Proximity	-15	63	-17	39	-0.0278	0.0282	-0.1006	0.1861	-0.0234	0.0217	-1.07	0.28
Sounds, Noise Levels Are Distracting or Uncomfortable	68	9	26	17	0.0035	0.0081	0.0287	0.0340	-0.0005	0.0192	-0.03	0.98
Very Hot or Cold Temperatures	81	36	7	-14	0.1058	0.0592	0.0271	-0.2545	-0.0193	0.0231	-0.83	0.41
Extremely Bright or Inadequate Lighting	85	17	12	17	0.0806	-0.0600	0.0232	0.0672	-0.0512	0.0278	-1.84	0.07
Exposed to Contaminants	68	39	30	33	0.0584	-0.0994	0.1395	0.1857	-0.0161	0.0200	-0.80	0.42
Cramped Work Space, Awkward Positions	69	31	27	39	0.0572	0.0002	-0.0311	0.2139	0.0396	0.0249	1.59	0.11
Exposed to Whole Body Vibration	81	9	19	2	0.0766	-0.0584	0.0699	-0.0098	-0.0159	0.0330	-0.48	0.63
Exposed to Radiation	1	1	8	73	-0.0150	-0.0097	-0.0273	0.1453	0.0471	0.0254	1.86	0.06
Exposed to Disease or Infections	-22	27	-16	76	-0.0332	0.0715	-0.1499	0.2536	0.0290	0.0178	1.63	0.10
Exposed to High Places	82	11	15	6	0.0900	-0.0777	0.0621	0.0316	0.0855	0.0297	2.87	0.00
Exposed to Hazardous Conditions	71	17	30	40	0.0419	-0.0800	0.0511	0.1487	-0.0299	0.0210	-1.42	0.16
Exposed to Hazardous Equipment	86	15	28	9	0.1496	-0.1249	0.1765	-0.0470	0.0421	0.0229	1.84	0.07
Exposed to Minor Burns, Cuts, Bites, or Stings	57	54	34	7	0.0103	0.0347	0.0911	-0.0310	-0.0104	0.0192	-0.54	0.59
Spend Time Sitting	-16	-94	-14	-4	0.0683	-0.3883	0.2728	0.1542	0.0148	0.0366	0.40	0.69
Spend Time Standing	14	94	19	2	-0.0791	0.3600	-0.0105	-0.1530	0.0117	0.0396	0.29	0.77
Spend Time Climbing Ladders, Scaffolds, or Poles	78	16	15	2	0.0783	0.0012	0.0153	-0.0992	-0.0338	0.0396	-0.85	0.39
Spend Time Walking and Running	19	89	13	5	-0.0388	0.1081	-0.0238	-0.0209	-0.0135	0.0260	-0.52	0.60
Spend Time Kneeling, Crouching, Stooping, or Crawling	48	64	20	20	0.0176	0.0641	0.0176	-0.0566	-0.0216	0.0299	-0.72	0.47
Spend Time Keeping or Regaining Balance	55	59	22	17	-0.0014	0.0760	-0.0120	-0.0061	-0.0044	0.0342	-0.13	0.90
Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls	38	43	66	13	-0.0169	-0.0386	0.3256	0.0012	0.0407	0.0238	1.71	0.09
Spend Time Bending or Twisting the Body	36	74	39	22	-0.0277	0.1440	0.1556	0.1046	0.0356	0.0309	1.15	0.25
Spend Time Making Repetitive Motions	-2	29	76	-6	-0.0431	-0.0025	0.2913	-0.1536	-0.0207	0.0248	-0.84	0.40

Table 1. O*NET variable list (continued)

O*NET Category	Loading factors 1	2	3	4	Standardized scoring coefficients 1	2	3	4	2016 log wage regression results Estimate	SE	t-stat	Pr > |t|
Work Context: Physical Work Conditions (continued)												
Wear Common Protective or Safety Equipment such as Safety Shoes, Glasses, Gloves, Hearing Protection, Hard Hats, or Life Jackets	56	40	36	40	0.0133	-0.0023	0.0381	0.1857	0.0070	0.0160	0.44	0.66
Wear Specialized Protective or Safety Equipment such as Breathing Apparatus, Safety Harness, Full Protection Suits, or Radiation Protection	59	18	19	58	0.0093	-0.0483	0.0099	0.2810	0.0189	0.0228	0.83	0.41
Work Context: Structural Job Characteristics (STRUCTURAL)												
Consequence of Error	49	31			0.0071	0.0916			-0.0107	0.0186	-0.57	0.57
Impact of Decisions on Co-workers or Company Results	88	14			0.4353	-0.0010			0.0294	0.0380	0.77	0.44
Frequency of Decision Making	78	19			0.1173	0.0597			0.0048	0.0332	0.15	0.88
Freedom to Make Decisions	85	-10			0.2532	-0.1753			0.0913	0.0310	2.94	0.00
Degree of Automation	-13	67			-0.0510	0.2215			0.0258	0.0243	1.06	0.29
Importance of Being Exact or Accurate	33	74			0.0646	0.3535			-0.0371	0.0294	-1.26	0.21
Importance of Repeating Same Tasks	0	74			-0.0773	0.3523			0.0376	0.0219	1.72	0.09
Structured versus Unstructured Work	77	3			0.2287	-0.0027			-0.0773	0.0345	-2.24	0.03
Level of Competition	45	5			0.0253	0.0110			0.0382	0.0190	2.01	0.04
Time Pressure	36	42			0.0280	0.1163			0.0384	0.0208	1.85	0.06
Pace Determined by Speed of Equipment	-34	26			-0.0378	0.1464			-0.0531	0.0238	-2.23	0.03

Note: There are 11 O*NET categories. The categories are listed in bold and our shorthand name for the category is in parentheses. Our criteria for the number of factors are the minimum eigenvalue is equal to 1 and the proportion of common variance accounted for by the retained factors is 0.95. The number of factors retained is the minimum number satisfying either criterion. Definitions of the loading factors and standardized scoring coefficients are found in the appendix. The dependent variable of the regression is the natural log of the mean occupational wage in 2016 and the regression includes 2016 occupational employment weights.

Table 2. Summary of 2016 log mean wage regressions

Regressors	R-squared
Years of education only	0.665
148 O*NET variables	0.933
148 O*NET variables and years of education	0.937
21 factors	0.849
21 factors and years of education	0.861
21 factors, years of education, and 7 cognitive 1 interactions	0.893
15 regression tree occupation group dummies	0.851
22 two-digit SOC group dummies	0.670

Note: In all regressions the dependent variable is the natual log of occupational mean wage in 2016 according to OES data. There are 788 occupations. The regressions include 2016 occupational employment weights.

Table 3. Correlation Matrix of O*NET based factors

Factor Name	EDUCATION	COGNITIVE1	COGNITIVE2	PHYSICAL	PSYCHOMOTOR	SENSORY1	SENSORY2	INFORMATION	INTERACTING1	INTERACTING2	MENTAL	OUTPUT1	OUTPUT2	INTERPERSONAL1	INTERPERSONAL2	INTERPERSONAL3	CONDITIONS1	CONDITIONS2	CONDITIONS3	CONDITIONS4	STRUCTURAL1
EDUCATION																					
COGNITIVE1	0.87																				
COGNITIVE2	0.17	0.03																			
PHYSICAL	-0.53	-0.69	0.35																		
PSYCHOMOTOR	-0.48	-0.64	0.55	0.83																	
SENSORY1	-0.08	-0.25	0.83	0.56	0.74																
SENSORY2	0.74	0.82	0.25	-0.49	-0.46	-0.04															
INFORMATION	0.61	0.51	0.55	-0.08	0.06	0.38	0.49														
INTERACTING1	0.62	0.60	0.37	-0.23	-0.18	0.15	0.52	0.71													
INTERACTING2	0.37	0.50	-0.18	-0.39	-0.47	-0.22	0.52	0.19	0.05												
MENTAL	0.83	0.80	0.33	-0.45	-0.34	0.13	0.67	0.83	0.80	0.39											
OUTPUT1	-0.19	-0.32	0.66	0.57	0.77	0.77	-0.21	0.33	0.17	-0.39	0.06										
OUTPUT2	0.67	0.76	0.17	-0.71	-0.49	-0.11	0.61	0.54	0.55	0.33	0.77	-0.05									
INTERPERSONAL1	0.66	0.81	-0.15	-0.74	-0.72	-0.34	0.67	0.24	0.28	0.65	0.59	-0.43	0.69								
INTERPERSONAL2	-0.15	-0.06	-0.06	0.18	0.02	-0.02	0.16	-0.05	-0.14	0.46	-0.15	-0.19	-0.22	0.02							
INTERPERSONAL3	0.36	0.27	0.37	0.13	0.10	0.24	0.32	0.45	0.65	-0.14	0.39	0.26	0.13	0.02	0.04						
CONDITIONS1	-0.21	-0.34	0.55	0.50	0.62	0.79	-0.28	0.14	0.05	-0.29	-0.01	0.73	-0.20	-0.32	-0.14	0.16					
CONDITIONS2	-0.55	-0.60	-0.09	0.70	0.42	0.04	-0.39	-0.38	-0.31	-0.26	-0.61	0.11	-0.76	-0.66	0.33	0.02	0.01				
CONDITIONS3	-0.37	-0.39	0.04	0.15	0.37	0.00	-0.40	-0.13	-0.19	-0.56	-0.31	0.30	-0.15	-0.45	-0.34	-0.10	0.01	0.02			
CONDITIONS4	0.28	0.15	0.36	0.22	0.28	0.25	0.28	0.51	0.20	0.04	0.25	0.25	0.07	-0.03	0.21	0.37	0.01	0.00	0.03		
STRUCTURAL1	0.61	0.58	0.29	-0.18	-0.17	0.21	0.55	0.55	0.46	0.43	0.63	0.06	0.41	0.52	0.16	0.44	0.12	-0.37	-0.41	0.34	
STRUCTURAL2	-0.07	0.05	0.09	-0.16	0.10	-0.01	-0.07	0.14	-0.07	-0.08	0.05	0.11	0.24	0.13	-0.03	-0.05	-0.01	-0.39	0.46	0.13	0.03

Note: The correlations are calculated using 2016 occupational employment weights. See Table 1 for the O*NET variables that contribute to each factor. Factor names come from their O*NET category and if the category includes 2 or more factors, the name includes a counter. For example, there is only one factor derived from the set of O*NET variables in the Physical Abilities category so that factor is simply named PHYSICAL. There are two factors identified in the Cognitive Abilities category so the factors are named COGNITIVE1 and COGNITIVE2, respectively.

Table 4. Correlations between Acemoglu-Autor variables and O*NET based factors

Acemoglu-Autor variable	ANALYTICAL	INTERPERSONAL	ROUTINE COGNITIVE	ROUTINE MANUAL	NON-ROUTINE MANUAL	OFFSHORABILITY
INTERPERSONAL	0.75					
ROUTINE COGNITIVE	-0.18	-0.30				
ROUTINE MANUAL	-0.40	-0.40	0.25			
NON-ROUTINE MANUAL	-0.38	-0.39	0.07	0.82		
OFFSHORABILITY	0.06	-0.13	0.07	-0.48	-0.60	
O*NET based factors						
EDUCATION	0.85	0.67	-0.26	-0.53	-0.51	0.15
COGNITIVE1	0.82	0.70	-0.12	-0.64	-0.68	0.27
COGNITIVE2	0.28	0.22	-0.03	0.36	0.52	-0.50
PHYSICAL	-0.50	-0.32	-0.08	0.63	0.79	-0.68
PSYCHOMOTOR	-0.39	-0.35	0.10	0.83	0.94	-0.67
SENSORY1	0.06	0.01	-0.12	0.51	0.77	-0.59
SENSORY2	0.68	0.66	-0.19	-0.54	-0.53	0.03
INFORMATION	0.77	0.63	-0.07	-0.02	0.03	-0.37
INTERACTING1	0.76	0.89	-0.27	-0.16	-0.20	-0.11
INTERACTING2	0.38	0.43	-0.07	-0.62	-0.46	-0.02
MENTAL	0.97	0.79	-0.18	-0.36	-0.33	-0.01
OUTPUT1	0.00	-0.05	-0.02	0.75	0.79	-0.68
OUTPUT2	0.80	0.54	0.09	-0.41	-0.48	0.28
INTERPERSONAL1	0.62	0.47	0.01	-0.72	-0.69	0.36
INTERPERSONAL2	-0.20	0.12	0.09	-0.19	-0.05	-0.34
INTERPERSONAL3	0.32	0.54	-0.23	0.05	0.03	-0.32
CONDITIONS1	-0.07	-0.11	-0.15	0.51	0.73	-0.40
CONDITIONS2	-0.61	-0.30	-0.17	0.31	0.36	-0.40
CONDITIONS3	-0.30	-0.42	0.46	0.68	0.37	0.01
CONDITIONS4	0.21	0.19	0.05	0.11	0.14	-0.53
STRUCTURAL1	0.57	0.56	-0.22	-0.34	-0.16	-0.20
STRUCTURAL2	0.02	-0.15	0.88	0.31	0.08	0.04

Note: The Acemoglu-Autor variables are constructed as described on page 1163 of Acemoglu-Autor (2011). To be consistent with our factors we normalize these measures so the weighted (by 2016 OES occupational employment) mean is 0 and variance is 1. The correlations are calculated using 2016 occupational employment weights.

Table 5. Factor medians by major occupation and 2005 wage groups

	Major occupation					2005 wage group		
Factor Name	Management, business, science, and arts (professional)	Service	Sales and office	Natural resources, construction, and maintenance (trades)	Production, transportation, and material moving (blue collar)	Bottom 20% of 2005 wage distribution	Middle 60% of 2005 wage distribution	Top 20% of 2005 wage distribution
EDUCATION	1.28	-0.71	-0.30	-0.56	-0.73	-1.11	-0.26	1.16
COGNITIVE1	1.05	-0.92	0.15	-0.81	-1.14	-0.93	-0.01	1.24
COGNITIVE2	0.20	-0.58	-0.75	1.44	0.60	-0.93	0.07	0.14
PHYSICAL	-0.85	0.72	-1.05	1.19	0.93	0.56	0.20	-1.21
PSYCHOMOTOR	-0.78	0.12	-0.59	1.30	1.28	0.11	0.24	-0.78
SENSORY1	-0.15	-0.43	-0.71	1.47	0.88	-0.66	-0.06	-0.10
SENSORY2	0.81	-0.38	0.06	-0.52	-0.94	-0.73	-0.05	0.76
INFORMATION	0.72	-0.65	-0.64	0.07	0.13	-0.95	-0.08	0.79
INTERACTING1	0.96	-0.35	-0.62	0.10	-0.30	-0.77	-0.25	1.17
INTERACTING2	0.35	-0.20	0.38	-1.06	-1.00	-0.21	0.16	0.48
MENTAL	1.06	-0.98	-0.29	-0.18	-0.44	-1.32	-0.11	1.23
OUTPUT1	-0.41	-0.34	-0.74	1.70	0.98	-0.37	-0.06	-0.31
OUTPUT2	0.69	-0.91	0.22	-0.48	-0.62	-0.99	0.12	0.90
INTERPERSONAL1	0.72	-0.89	0.83	-0.71	-1.17	-0.92	0.17	0.84
INTERPERSONAL2	-0.29	0.58	0.22	-0.32	-0.70	0.34	-0.04	-0.32
INTERPERSONAL3	0.42	-0.07	-0.61	0.29	0.09	-0.23	-0.05	0.49
CONDITIONS1	-0.50	-0.55	-0.57	2.22	0.87	-0.64	-0.17	-0.41
CONDITIONS2	-0.75	1.25	-0.97	0.47	0.24	1.21	0.11	-0.98
CONDITIONS3	-0.36	0.20	0.23	0.51	0.45	0.27	0.17	-0.33
CONDITIONS4	-0.10	-0.25	-0.50	0.17	-0.56	-0.57	-0.29	-0.19
STRUCTURAL1	0.55	-0.41	-0.02	0.36	-0.47	-1.02	0.16	0.91
STRUCTURAL2	-0.01	-0.32	0.63	-0.11	0.40	-0.32	0.23	0.19

Note: The medians are constructed using 2016 occupational employment. Major occupation categories are defined using the high-level aggregations suggested in Table 4 of the 2010 SOC User Guide (https://www.bls.gov/soc/soc_2010_class_and_coding_structure.pdf). The 2005 wage groups are defined so that exactly 20% of total 2005 employment is in the lowest and highest wage groups. This requires proportionately assigning the employment from the marginal occupations that cross that 20% and 80% thresholds.

Table 6. 2016 log wage regressions

Parameter	Without interactions		With Cognitive 1 interactions	
	Estimate	Standard error	Estimate	Standard error
Intercept	0.8302	0.1577	1.9349	0.1853
EDUCATION	0.0642	0.0083	0.0418	0.0082
COGNITIVE1	0.2010	0.0257	-0.3027	0.0633
COGNITIVE2	-0.0302	0.0183	-0.0212	0.0165
PHYSICAL	-0.0507	0.0222	0.0965	0.0398
PSYCHOMOTOR	0.0067	0.0253	0.0343	0.0240
SENSORY1	0.1391	0.0246	0.0498	0.0322
SENSORY2	-0.0705	0.0197	0.0702	0.0296
INFORMATION	-0.0053	0.0201	-0.0210	0.0183
INTERACTING1	0.0505	0.0188	0.0269	0.0174
INTERACTING2	0.0372	0.0157	0.0057	0.0146
MENTAL	0.0496	0.0316	0.1006	0.0296
OUTPUT1	-0.0779	0.0200	-0.0680	0.0184
OUTPUT2	0.0554	0.0217	-0.0800	0.0376
INTERPERSONAL1	-0.0266	0.0208	0.0531	0.0200
INTERPERSONAL2	-0.0423	0.0123	-0.0738	0.0218
INTERPERSONAL3	0.0575	0.0125	-0.1182	0.0223
CONDITIONS1	0.0867	0.0167	0.0829	0.0158
CONDITIONS2	-0.0072	0.0187	-0.0117	0.0176
CONDITIONS3	0.1032	0.0139	0.0902	0.0127
CONDITIONS4	-0.0018	0.0114	0.0218	0.0108
STRUCTURAL1	0.0839	0.0122	-0.0195	0.0215
STRUCTURAL2	-0.0103	0.0122	-0.0162	0.0111
Interactions with COGNITIVE1:				
PHYSICAL			-0.0732	0.0174
SENSORY1			0.0406	0.0123
SENSORY2			-0.0544	0.0111
OUTPUT2			0.0596	0.0144
INTERPERSONAL2			0.0224	0.0085
INTERPERSONAL3			0.0748	0.0099
STRUCTURAL1			0.0451	0.0097
R squared	0.861		0.893	

Note: The dependent variable in both regressions is the natural log of 2016 occupational mean wage according to the OES. Regressions include 2016 occupational employment weights. There are 788 occupations included in the regressions.

Table 7. Comparison of first two groups from regression tree

	Occupation Group				
	A		B		
Occupations	484		304		
2016 employment	91,926,974		49,349,126		
	Mean	SD	Mean	SD	
2016 log wage	2.74	0.30	3.57	0.37	
Factor Name	Mean	SD	Mean	SD	
EDUCATION	11.88	1.18	15.72	1.73	*
COGNITIVE1	-0.59	0.64	1.10	0.45	*
COGNITIVE2	-0.04	1.04	0.07	0.79	
PHYSICAL	0.41	0.87	-0.76	0.73	*
PSYCHOMOTOR	0.38	0.88	-0.70	0.79	*
SENSORY1	0.11	1.12	-0.20	0.58	*
SENSORY2	-0.45	0.75	0.85	0.63	*
INFORMATION	-0.34	0.84	0.63	0.81	*
INTERACTING1	-0.45	0.72	0.83	0.86	*
INTERACTING2	-0.23	0.95	0.43	0.77	*
MENTAL	-0.54	0.73	1.00	0.53	*
OUTPUT1	0.17	1.02	-0.32	0.79	*
OUTPUT2	-0.44	0.74	0.81	0.66	*
INTERPERSONAL1	-0.42	0.92	0.78	0.37	*
INTERPERSONAL2	0.10	0.89	-0.19	1.01	*
INTERPERSONAL3	-0.22	0.90	0.41	0.87	*
CONDITIONS1	0.18	1.13	-0.33	0.53	*
CONDITIONS2	0.36	0.92	-0.66	0.75	*
CONDITIONS3	0.22	0.96	-0.42	0.80	*
CONDITIONS4	-0.07	0.89	0.12	1.04	
STRUCTURAL1	-0.37	0.86	0.69	0.74	*
STRUCTURAL2	0.05	0.89	-0.09	0.91	

Note: The dependent variable in the regression tree is the natural log of the 2016 occupational mean wage according to the OES. We expand the dataset to include an observation for every ten workers in an occupation in order to account for the variance in the size of occupations. The * indicates that the difference in the factor means between group A and B are statistically significant at the 0.01 level. COGNITIVE1 defines the first split. Occupations where COGNITIVE1 is less than 0.466 are in group A.

Table 8. Comparison of first four groups of regression tree

	Occupation Group									
	A1		A2			B1		B2		
Occupations	270		214			135		169		
2016 employment	71,226,709		20,700,264			28,173,876		21,175,250		
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
2016 log wage	2.65	0.25	3.04	0.26		3.41	0.33	3.78	0.31	
Factor Name	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
EDUCATION	11.77	1.25	12.25	0.81	*	14.82	1.33	16.91	1.45	*
COGNITIVE1	-0.56	0.66	-0.70	0.56		0.78	0.19	1.53	0.30	*
COGNITIVE2	-0.48	0.68	1.47	0.52	*	0.00	0.70	0.17	0.90	
PHYSICAL	0.24	0.87	1.00	0.55	*	-0.67	0.71	-0.87	0.73	
PSYCHOMOTOR	0.10	0.75	1.35	0.53	*	-0.65	0.74	-0.78	0.85	
SENSORY1	-0.35	0.73	1.68	0.78	*	-0.14	0.59	-0.28	0.57	
SENSORY2	-0.49	0.78	-0.34	0.65		0.69	0.58	1.05	0.65	*
INFORMATION	-0.58	0.73	0.49	0.68	*	0.34	0.65	1.03	0.83	*
INTERACTING1	-0.60	0.63	0.09	0.75	*	0.62	0.79	1.11	0.85	*
INTERACTING2	-0.15	0.91	-0.52	1.02		0.38	0.77	0.51	0.76	
MENTAL	-0.71	0.66	0.06	0.64	*	0.75	0.44	1.34	0.45	*
OUTPUT1	-0.19	0.73	1.40	0.91	*	-0.20	0.80	-0.48	0.74	
OUTPUT2	-0.52	0.73	-0.15	0.70	*	0.68	0.66	0.98	0.61	*
INTERPERSONAL1	-0.34	0.97	-0.69	0.61	*	0.73	0.40	0.85	0.33	
INTERPERSONAL2	0.13	0.83	0.00	1.06		0.02	0.91	-0.47	1.06	
INTERPERSONAL3	-0.32	0.83	0.13	1.03		0.38	0.94	0.45	0.76	
CONDITIONS1	-0.20	0.83	1.47	1.06	*	-0.24	0.57	-0.44	0.45	
CONDITIONS2	0.46	0.96	0.00	0.63	*	-0.50	0.80	-0.89	0.61	
CONDITIONS3	0.19	0.90	0.34	1.13		-0.48	0.84	-0.33	0.72	
CONDITIONS4	-0.19	0.84	0.36	0.91	*	-0.08	0.66	0.39	1.35	
STRUCTURAL1	-0.56	0.80	0.27	0.75	*	0.55	0.71	0.88	0.75	
STRUCTURAL2	0.01	0.93	0.19	0.73		-0.18	0.96	0.03	0.83	

Note: See note for Table 7. The split for the A group occupations is defined by COGNITIVE2. Occupations where COGNITIVE2 is less than 0.785 are in group A1. The split for the B group occupations is defined by COGNITIVE1. Occupations where COGNITIVE1 is less than 1.213 are in group B1. The * in the A2 column indicate whether the factor means are statistically different (at the 0.01 level) between groups A2 and A1 whereas the * in the B2 column indicate whether the factor means are statistically different (at the 0.01 level) between groups B2 and B1.

Table 9. R squared from regressions using various occupation dummies

	Occupation level (number of occupations)					
	From the regression tree					
Dependent variable	1 level (2)	2 levels (4)	3 levels (8)	final (15)	SOC2 (22)	PCTL (15)
2016 log wage	0.60	0.71	0.78	0.85	0.67	0.98
EDUCATION	0.63	0.71	0.78	0.82	0.72	0.69
COGNITIVE1	0.66	0.71	0.82	0.86	0.75	0.65
COGNITIVE2	0.00	0.47	0.50	0.56	0.56	0.28
PHYSICAL	0.31	0.38	0.42	0.63	0.69	0.34
PSYCHOMOTOR	0.27	0.45	0.48	0.59	0.75	0.24
SENSORY1	0.02	0.51	0.51	0.65	0.68	0.22
SENSORY2	0.43	0.45	0.54	0.61	0.56	0.47
INFORMATION	0.24	0.43	0.56	0.62	0.58	0.51
INTERACTING1	0.39	0.47	0.54	0.59	0.55	0.55
INTERACTING2	0.11	0.13	0.19	0.22	0.57	0.14
MENTAL	0.55	0.65	0.83	0.86	0.64	0.76
OUTPUT1	0.06	0.37	0.38	0.49	0.67	0.19
OUTPUT2	0.41	0.44	0.57	0.68	0.66	0.55
INTERPERSONAL1	0.36	0.37	0.50	0.61	0.75	0.40
INTERPERSONAL2	0.02	0.05	0.07	0.20	0.52	0.15
INTERPERSONAL3	0.10	0.13	0.18	0.27	0.35	0.28
CONDITIONS1	0.06	0.38	0.38	0.60	0.73	0.20
CONDITIONS2	0.24	0.28	0.36	0.65	0.64	0.54
CONDITIONS3	0.10	0.11	0.13	0.17	0.54	0.18
CONDITIONS4	0.01	0.07	0.11	0.21	0.67	0.24
STRUCTURAL1	0.28	0.37	0.43	0.53	0.42	0.53
STRUCTURAL2	0.01	0.01	0.03	0.19	0.50	0.19

Note: The dependent variable is either the natual log of the 2016 occupational mean wage according to the OES or the factor value at the occupation. There are 788 occupations in each regression. The regressions include 2016 occupational employment weights. The first four columns represent different levels of aggregation within the regression tree with the column labeled final indicating our preferred specification of the tree. The column labeled SOC2 contains the 22 major occupation groups based on the SOC structure. The column labeled PCTL contains 15 occupations that are roughly the same size as the occupations generated by the regression tree, but only take into account the wage.

Table 10. Normalized factor values and Acemoglu-Autor variables by final regression tree occupations

	Final regression tree occupation code														
Variable	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
Branch of the tree	A1	A1	A1	A1	A2	A1	A2	B1	A2	B2	B1	B2	B1	B2	B2
Occupations	54	62	50	15	58	89	66	58	90	37	61	90	16	31	11
2016 mean wage	11.71	14.75	15.26	15.32	15.56	18.65	20.90	25.30	26.05	34.93	36.51	43.75	46.82	48.69	94.71
2016 employment share	19.2	11.0	5.6	3.1	3.7	11.4	4.8	10.6	6.1	3.1	6.6	6.2	2.7	5.0	0.7
COGNITIVE1	-0.91	-0.31	-1.37	-0.05	-1.13	0.04	-1.08	0.68	-0.15	1.45	0.93	1.59	0.81	1.49	1.89
SENSORY2	-0.78	-0.21	-1.48	0.15	-0.70	-0.08	-0.71	0.72	0.14	0.93	0.94	0.82	0.30	1.46	1.95
EDUCATION	-1.06	-0.42	-0.86	0.03	-0.62	-0.17	-0.60	0.55	-0.15	1.19	1.02	1.51	0.45	1.83	2.57
MENTAL	-1.28	-0.29	-1.13	-0.91	-0.23	0.08	-0.14	0.46	0.39	0.76	1.23	1.54	0.78	1.42	1.72
INTERACTING1	-0.75	-0.26	-1.13	-1.12	-0.13	-0.34	-0.33	0.32	0.55	0.20	1.03	1.40	0.89	1.35	1.39
INFORMATION	-1.03	-0.14	-0.85	-1.30	0.24	-0.04	0.34	0.07	0.82	0.06	0.75	1.01	0.52	1.67	1.92
STRUCTURAL1	-0.92	-0.40	-0.72	-0.40	-0.21	-0.16	0.22	0.30	0.62	0.99	0.54	0.24	1.70	1.50	2.26
INTERPERSONAL3	-0.38	-0.23	-0.38	-0.18	-0.33	-0.43	-0.25	0.08	0.72	-0.13	0.56	0.03	1.32	1.28	1.47
INTERACTING2	-0.44	0.13	-0.74	0.10	-0.87	0.26	-0.67	0.32	-0.26	0.60	0.42	0.25	0.66	0.80	0.96
INTERPERSONAL1	-0.88	-0.30	-1.13	0.60	-1.18	0.60	-0.86	0.70	-0.33	1.09	0.80	0.96	0.90	0.69	0.67
OUTPUT2	-0.99	-0.55	-1.41	0.04	-0.41	0.42	-0.42	0.56	0.21	0.97	1.07	1.44	0.59	0.68	0.80
COGNITIVE2	-0.79	-0.29	0.02	-0.55	1.22	-0.44	1.53	-0.19	1.71	-0.49	0.37	0.06	-0.15	0.58	1.25
SENSORY1	-0.69	-0.12	0.78	-0.61	1.17	-0.52	2.10	-0.33	1.73	-0.59	-0.03	-0.43	0.29	0.02	0.19
OUTPUT1	-0.49	0.15	0.55	-0.76	0.66	-0.23	1.86	-0.34	1.57	-0.92	-0.16	-0.54	0.22	-0.21	-0.23
CONDITIONS1	-0.58	-0.03	1.18	-0.40	0.47	-0.34	2.23	-0.43	1.49	-0.39	-0.14	-0.36	0.22	-0.57	-0.57
PSYCHOMOTOR	0.15	0.14	1.14	-0.48	1.03	-0.39	1.65	-0.66	1.33	-1.07	-0.74	-1.18	-0.43	-0.23	0.01
PHYSICAL	0.47	0.61	1.17	-0.78	0.75	-0.69	1.15	-0.66	1.05	-1.13	-0.81	-1.26	-0.40	-0.27	-0.69
INTERPERSONAL2	0.48	0.11	-0.54	-0.02	-0.17	-0.02	-0.20	0.18	0.25	-0.39	-0.26	-1.35	0.06	0.36	0.38
CONDITIONS2	1.17	0.89	0.43	-0.67	0.17	-0.82	-0.12	-0.41	-0.01	-1.04	-0.56	-1.21	-0.74	-0.47	-0.51
CONDITIONS3	0.43	-0.31	0.09	-0.09	0.51	0.44	0.59	-0.31	0.08	-0.46	-0.70	-0.33	-0.75	-0.29	-0.46
CONDITIONS4	-0.18	-0.36	-0.52	-0.21	0.09	0.08	0.04	-0.16	0.82	-0.12	0.06	-0.29	-0.12	1.41	1.97
STRUCTURAL2	-0.27	-0.64	0.11	0.42	0.10	0.94	0.11	-0.22	0.36	0.08	-0.30	-0.08	0.11	0.09	0.45

Key:
>= 1
>= 0.5 & < 1

Table 10 (continued). Normalized factor values and Acemoglu-Autor variables by final regression tree occupations

Variable	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
Acemoglu-Autor variable															
ANALYTICAL	-1.21	-0.27	-1.15	-0.78	-0.29	0.04	-0.26	0.45	0.30	0.82	1.26	1.68	0.41	1.39	1.59
INTERPERSONAL	-0.74	-0.20	-1.30	-0.60	-0.47	-0.20	-0.63	0.41	0.34	0.30	0.96	1.19	1.05	1.52	1.45
ROUTINE COGNITIVE	0.20	-0.42	0.00	0.50	0.29	0.82	-0.22	-0.30	0.23	-0.29	-0.49	-0.20	-0.55	-0.28	-0.21
ROUTINE MANUAL	0.23	0.18	0.92	-0.44	0.83	-0.04	1.48	-0.70	0.94	-0.98	-0.77	-0.95	-0.52	-0.51	-0.37
NON-ROUTINE MANUAL	0.04	0.19	1.32	-0.64	0.95	-0.36	1.94	-0.65	1.37	-1.02	-0.73	-1.01	-0.48	-0.55	-0.32
OFFSHORABILITY	0.19	-0.55	-0.13	0.69	-0.48	0.19	-0.97	0.36	-1.25	1.35	0.38	1.14	-0.30	-0.66	-0.63

Note: The final regression tree occupations are sorted in ascending order by 2016 mean wage. The first row of the table shows the branch of the tree that the occupation falls into. See Table 8 for more details. The shading and the ordering of the factors is meant to highlight patterns in the relationship between various factors and the regression tree occupations and hence wages.

Table 11.a. 2016 regression tree occupation by sector real mean wages

Regression tree occupation	Natural resources and mining	Construction	Manufacturing	Trade, transportation, and utilities	Information	Financial activities	Professional and business services	Education services	Health services	Leisure and hospitality	Other services	Government services	Total
01	12.06	14.04	13.18	11.64	11.22	13.60	12.11	13.71	12.04	11.14	11.60	14.75	11.71
02	14.75	18.85	15.94	13.32	19.87	16.13	13.60	14.84	14.27	15.59	15.36	16.84	14.75
03	12.88	20.28	15.85	16.39	15.87	13.43	13.33	15.06	13.52	11.73	13.00	16.68	15.26
04	17.42	16.64	16.87	15.44	16.52	16.15	15.80	14.53	14.08	13.34	14.58	16.87	15.32
05	15.97	18.19	16.72	15.66	18.36	16.06	13.51	14.87	15.25	13.24	15.09	17.98	15.56
06	27.50	26.28	18.99	17.34	22.13	18.35	18.27	18.65	17.72	18.70	16.20	20.00	18.65
07	22.47	23.22	21.00	20.97	28.76	17.95	19.13	20.20	19.07	16.67	19.14	21.32	20.90
08	29.12	25.29	28.86	25.82	29.03	26.22	25.07	25.09	22.29	23.52	23.37	25.72	25.31
09	28.15	25.05	25.25	30.98	27.52	23.65	22.73	21.76	25.27	25.09	22.21	26.76	26.05
10	42.47	32.99	37.40	35.73	33.74	41.65	37.34	34.15	26.92	25.12	26.01	32.44	34.93
11	43.80	44.08	42.63	38.69	44.29	39.06	41.37	29.12	30.91	29.29	30.80	36.95	36.51
12	51.18	35.82	45.94	43.26	49.96	48.37	44.95	38.06	44.64	32.56	34.28	39.83	43.75
13	58.60	51.57	57.19	46.70	44.83	39.56	58.59	46.02	39.75	39.38	40.40	42.19	46.82
14	71.68	59.68	59.88	59.44	76.55	65.83	66.87	45.82	38.44	49.43	50.59	45.71	48.69
15	94.38	91.21	99.97	98.33	108.79	105.60	105.64	65.12	97.25	85.15	84.71	78.25	94.71
Total	24.89	25.26	24.34	19.68	34.70	30.47	29.27	25.86	25.02	12.92	19.37	28.12	23.79

Note: Authors' calculations using OES data. Wages are in 2017 dollars.

Table 11.b. 2016 regression tree occupation share of sector employment

Regression tree occupation	Sector												Total
	Natural resources and mining	Construction	Manufacturing	Trade, transportation, and utilities	Information	Financial activities	Professional and business services	Education services	Health services	Leisure and hospitality	Other services	Government services	
01	2.1	0.6	5.3	20.9	5.8	11.4	9.0	6.0	18.9	77.8	15.3	3.9	19.3
02	4.0	21.6	12.2	20.2	1.8	0.9	4.3	12.0	11.6	7.5	10.7	3.9	11.0
03	28.3	5.2	9.9	11.0	1.3	1.7	8.2	0.6	0.4	2.7	6.8	3.0	5.6
04	2.0	3.2	1.7	2.0	2.4	3.8	3.9	4.1	4.8	1.0	5.5	4.0	3.1
05	1.6	1.3	14.8	2.3	0.5	0.7	5.0	2.3	0.9	1.5	15.2	2.8	3.7
06	7.7	10.6	13.4	9.7	15.0	17.8	13.6	6.6	19.2	2.0	10.3	11.0	11.4
07	27.5	13.6	7.2	8.9	2.3	4.5	2.3	1.0	0.6	1.0	9.5	5.6	4.8
08	4.1	3.5	5.7	11.0	11.2	14.8	12.0	28.5	7.7	2.7	4.9	11.4	10.6
09	8.0	27.6	11.2	3.6	10.7	0.8	3.3	1.2	3.7	0.6	3.1	22.4	6.1
10	1.1	1.5	2.8	1.7	8.5	9.3	3.8	4.9	3.2	0.4	2.3	2.9	3.1
11	3.4	4.2	5.1	2.1	17.3	7.1	10.3	16.9	4.1	1.0	6.7	12.9	6.6
12	4.4	3.6	5.5	2.4	14.5	12.2	15.0	7.6	3.5	0.5	3.0	8.2	6.2
13	2.9	3.1	2.2	2.7	5.3	10.9	3.2	0.6	1.1	1.1	5.1	2.2	2.7
14	2.8	0.3	2.7	1.2	3.1	3.8	5.8	7.4	16.8	0.2	1.4	4.9	5.0
15	0.1	0.2	0.2	0.1	0.3	0.3	0.3	0.3	3.6	0.0	0.2	1.0	0.7

Note: Authors' calculations using 2016 OES data. The entries show the percent of sector employment that is employed in the regression tree occupation. The total column shows the percent of overall national employment that is employed in the regression tree occupation. For example, 77.8% of Leisure and Hospitality employment is employed in occupation 01, but only 19.3% of total employment is employed in occupation 01.

Table 11.c. Relative importance of regression tree occupations to sector employment

Regression tree occupation	Sector											
	Natural resources and mining	Construction	Manufacturing	Trade, transportation, and utilities	Information	Financial activities	Professional and business services	Education services	Health services	Leisure and hospitality	Other services	Government services
01	0.1	0.0	0.3	1.1	0.3	0.6	0.5	0.3	1.0	4.0	0.8	0.2
02	0.4	2.0	1.1	1.8	0.2	0.1	0.4	1.1	1.1	0.7	1.0	0.4
03	5.0	0.9	1.8	2.0	0.2	0.3	1.5	0.1	0.1	0.5	1.2	0.5
04	0.6	1.0	0.6	0.7	0.8	1.2	1.2	1.3	1.5	0.3	1.8	1.3
05	0.4	0.4	4.0	0.6	0.1	0.2	1.3	0.6	0.2	0.4	4.1	0.7
06	0.7	0.9	1.2	0.8	1.3	1.6	1.2	0.6	1.7	0.2	0.9	1.0
07	5.7	2.8	1.5	1.8	0.5	0.9	0.5	0.2	0.1	0.2	2.0	1.2
08	0.4	0.3	0.5	1.0	1.1	1.4	1.1	2.7	0.7	0.3	0.5	1.1
09	1.3	4.5	1.8	0.6	1.8	0.1	0.5	0.2	0.6	0.1	0.5	3.7
10	0.3	0.5	0.9	0.6	2.7	3.0	1.2	1.6	1.0	0.1	0.7	0.9
11	0.5	0.6	0.8	0.3	2.6	1.1	1.6	2.5	0.6	0.1	1.0	2.0
12	0.7	0.6	0.9	0.4	2.3	2.0	2.4	1.2	0.6	0.1	0.5	1.3
13	1.1	1.1	0.8	1.0	1.9	4.0	1.2	0.2	0.4	0.4	1.9	0.8
14	0.6	0.1	0.5	0.2	0.6	0.8	1.2	1.5	3.4	0.0	0.3	1.0
15	0.2	0.2	0.3	0.1	0.4	0.4	0.4	0.5	5.1	0.1	0.3	1.4

Note: Authors' calculations using 2016 OES data. The entries in the table show the relative importance of an occupation within a sector. Precisely, the importance equals the share of sector employment employed in a regression tree occupation divided by the share of total employment employed in the same occupation.

Table 11.d. Relative importance of two-digit SOC (SOC2) to sector employment

	Sector											
SOC2	Natural resources and mining	Construction	Manufacturing	Trade, transportation, and utilities	Information	Financial activities	Professional and business services	Education services	Health services	Leisure and hospitality	Other services	Government services
11	0.8	1.2	1.1	0.7	1.4	1.8	1.5	1.0	0.7	0.5	1.0	1.2
13	0.7	0.7	0.8	0.4	1.5	3.8	2.1	0.4	0.3	0.1	1.1	2.1
15	0.3	0.1	0.8	0.3	6.7	1.7	3.3	0.6	0.2	0.0	0.3	0.9
17	2.0	0.7	3.6	0.3	0.8	0.0	3.0	0.1	0.0	0.0	0.1	1.7
19	2.6	0.0	1.2	0.1	0.1	0.0	2.3	1.7	0.6	0.0	0.4	3.8
21	0.0	0.0	0.0	0.0	0.0	0.1	0.2	1.7	3.8	0.0	1.9	3.5
23	0.7	0.0	0.1	0.0	0.4	1.0	4.6	0.1	0.0	0.0	0.2	3.6
25	0.0	0.0	0.0	0.0	0.1	0.0	0.1	9.6	0.4	0.0	0.3	0.4
27	0.0	0.1	0.5	0.5	12.4	0.3	1.6	1.5	0.1	1.0	1.6	0.5
29	0.1	0.0	0.0	0.3	0.0	0.1	0.3	0.4	5.7	0.0	0.0	0.8
31	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	6.5	0.0	0.5	0.5
33	0.1	0.0	0.0	0.2	0.1	0.2	1.7	0.5	0.2	0.5	0.4	8.1
35	0.0	0.0	0.0	0.2	0.2	0.0	0.1	0.4	0.3	7.6	0.3	0.1
37	0.2	0.2	0.2	0.2	0.1	0.6	3.0	1.2	0.7	1.6	0.8	0.7
39	0.1	0.0	0.0	0.1	0.7	0.1	0.2	0.5	3.3	1.7	6.4	0.9
41	0.1	0.2	0.3	3.5	1.1	1.6	0.5	0.0	0.0	0.4	0.5	0.1
43	0.5	0.6	0.6	1.3	1.1	2.4	1.3	0.6	0.9	0.3	0.9	1.1
45	87.9	0.0	0.8	0.7	0.0	0.0	0.3	0.1	0.0	0.1	0.1	0.8
47	5.4	15.7	0.4	0.1	0.1	0.1	0.4	0.1	0.0	0.0	0.1	1.2
49	1.5	2.3	1.3	1.5	2.3	1.3	0.5	0.3	0.2	0.3	4.2	1.1
51	0.9	0.3	8.0	0.5	0.2	0.0	0.8	0.0	0.1	0.1	1.1	0.3
53	1.8	0.5	1.1	2.8	0.2	0.3	0.9	0.3	0.1	0.3	1.4	0.6

Note: Authors' calculations using 2016 OES data. See note in Table 11.c describing the importance measures in the table.

Table 12. 2017 employment share changes by 2005 factor percentile groups

Variable	2005 factor percentile group: Bottom 20%	20-50	50-80	Top 20%	Change in means (2017 - 2005)
Real wage	1.39	-2.02	-2.30	2.93	1.412
O*NET based factor					
EDUCATION	-0.99	-1.56	-0.21	2.76	0.063
COGNITIVE1	-0.86	-1.47	-0.56	2.90	0.060
COGNITIVE2	0.23	0.46	0.08	-0.76	-0.007
PHYSICAL	0.30	0.21	0.69	-1.20	-0.019
PSYCHOMOTOR	0.51	1.48	0.18	-2.16	-0.042
SENSORY1	-0.08	2.00	-0.45	-1.46	-0.030
SENSORY2	-2.43	0.90	0.74	0.79	0.045
INFORMATION	-1.03	-0.40	-0.70	2.14	0.041
INTERACTING1	-1.51	-0.88	-0.09	2.49	0.048
INTERACTING2	-1.46	0.49	0.90	0.07	0.036
MENTAL	1.28	-3.33	-1.05	3.10	0.043
OUTPUT1	-0.32	1.84	0.53	-2.05	-0.042
OUTPUT2	0.24	-1.07	-0.54	1.38	0.023
INTERPERSONAL1	0.35	1.66	-0.43	-1.58	0.035
INTERPERSONAL2	0.32	0.11	-1.89	1.46	0.023
INTERPERSONAL3	-0.08	2.61	0.44	-2.97	0.031
CONDITIONS1	-0.26	-0.30	-1.23	1.79	-0.047
CONDITIONS2	-1.65	0.22	0.61	0.81	0.002
CONDITIONS3	-0.60	-1.44	1.68	0.36	-0.054
CONDITIONS4	-1.31	-0.68	1.47	0.51	0.040
STRUCTURAL1	-1.97	-0.30	0.96	1.30	0.057
STRUCTURAL2	0.62	1.13	-0.25	-1.50	-0.034
Acemoglu-Autor variable					
ANALYTICAL	0.91	-2.65	-0.97	2.71	0.047
INTERPERSONAL	-1.15	-1.63	0.88	1.90	0.061
ROUTINE COGNITIVE	0.33	0.91	0.05	-1.29	-0.035
ROUTINE MANUAL	0.49	1.74	0.76	-3.00	-0.071
NONROUTINE MANUAL	0.42	1.24	0.38	-2.05	-0.050
OFFSHORABILITY	0.05	-0.53	0.10	0.37	0.002

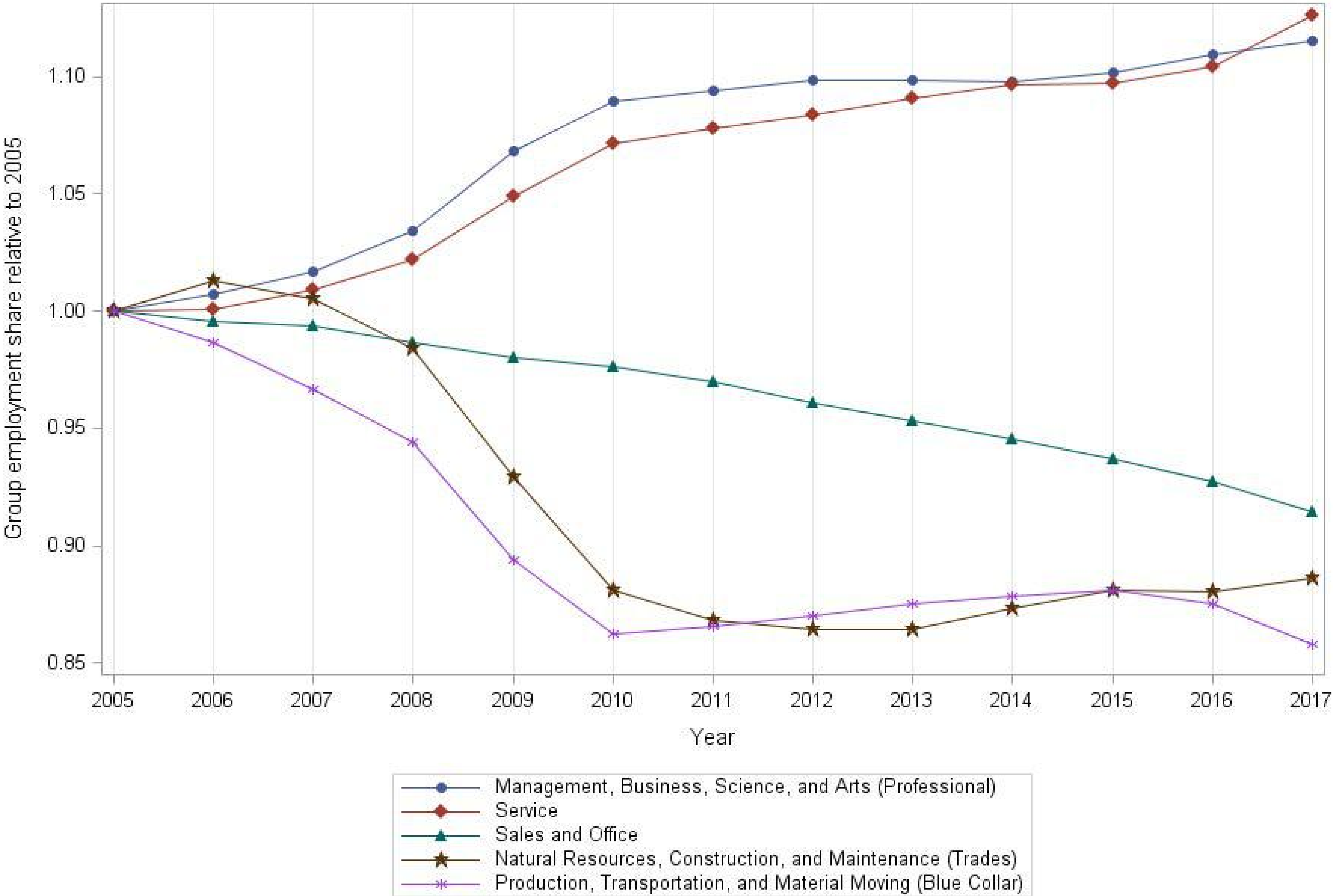
Note: See note for Table 5 describing the assignment of detailed occupations to a 2005 wage group. Factor means are weighted using detailed occupational employment in 2005 and 2017.

Table 13. Wages, employment, and employment shares by regression tree occupations

Variable	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	Total
	Regression tree occupation															
Employment levels (thousands)																
2005	24,050	14,988	8,171	4,621	5,577	15,191	6,859	15,138	8,593	3,545	7,996	6,800	3,272	5,611	912	131,323
2017	27,987	15,486	8,117	4,368	5,192	16,184	6,908	14,911	8,668	4,409	9,576	8,913	3,854	7,264	1,006	142,844
Difference	3,937	497	-54	-253	-385	993	50	-226	75	864	1,580	2,113	582	1,652	94	11,521
Cumulative difference	3,937	4,434	4,381	4,128	3,743	4,737	4,786	4,560	4,635	5,499	7,079	9,192	9,774	11,426	11,521	
Percent of total employment																
2005	18.3	11.4	6.2	3.5	4.2	11.6	5.2	11.5	6.5	2.7	6.1	5.2	2.5	4.3	0.7	
2017	19.6	10.8	5.7	3.1	3.6	11.3	4.8	10.4	6.1	3.1	6.7	6.2	2.7	5.1	0.7	
Difference	1.3	-0.6	-0.5	-0.5	-0.6	-0.2	-0.4	-1.1	-0.5	0.4	0.6	1.1	0.2	0.8	0.0	
Cumulative difference	1.3	0.7	0.2	-0.3	-0.9	-1.1	-1.5	-2.6	-3.1	-2.7	-2.1	-1.0	-0.8	0.0	0.0	
Counterfactual employment levels (thousands) holding total employment fixed at 2017 level																
2005	26,160	16,303	8,888	5,026	6,066	16,524	7,460	16,466	9,347	3,856	8,697	7,396	3,559	6,104	992	142,844
2017	27,987	15,486	8,117	4,368	5,192	16,184	6,908	14,911	8,668	4,409	9,576	8,913	3,854	7,264	1,006	142,844
Difference	1,827	-818	-770	-658	-874	-340	-552	-1,554	-679	553	879	1,517	295	1,160	14	
Cumulative difference	1,827	1,010	239	-419	-1,293	-1,632	-2,185	-3,739	-4,418	-3,865	-2,986	-1,470	-1,174	-14	0	
Real mean wages (2017 dollars)																
2005	11.61	14.96	15.50	14.62	16.01	18.97	21.19	25.20	26.04	35.38	35.32	42.02	45.19	46.10	85.38	22.87
2017 counterfactual holding occupational employment fixed at 2005 levels	12.22	15.07	15.84	15.76	16.18	19.09	21.34	25.90	26.46	35.73	36.71	44.23	45.56	50.08	96.14	
2017	12.07	15.10	15.68	15.74	16.01	18.98	21.39	25.70	26.67	35.12	37.22	44.53	47.46	49.65	96.33	24.28
Percent difference (2005 to 2017)	4.0	0.9	1.2	7.6	0.0	0.1	1.0	2.0	2.4	-0.7	5.4	6.0	5.0	7.7	12.8	6.2

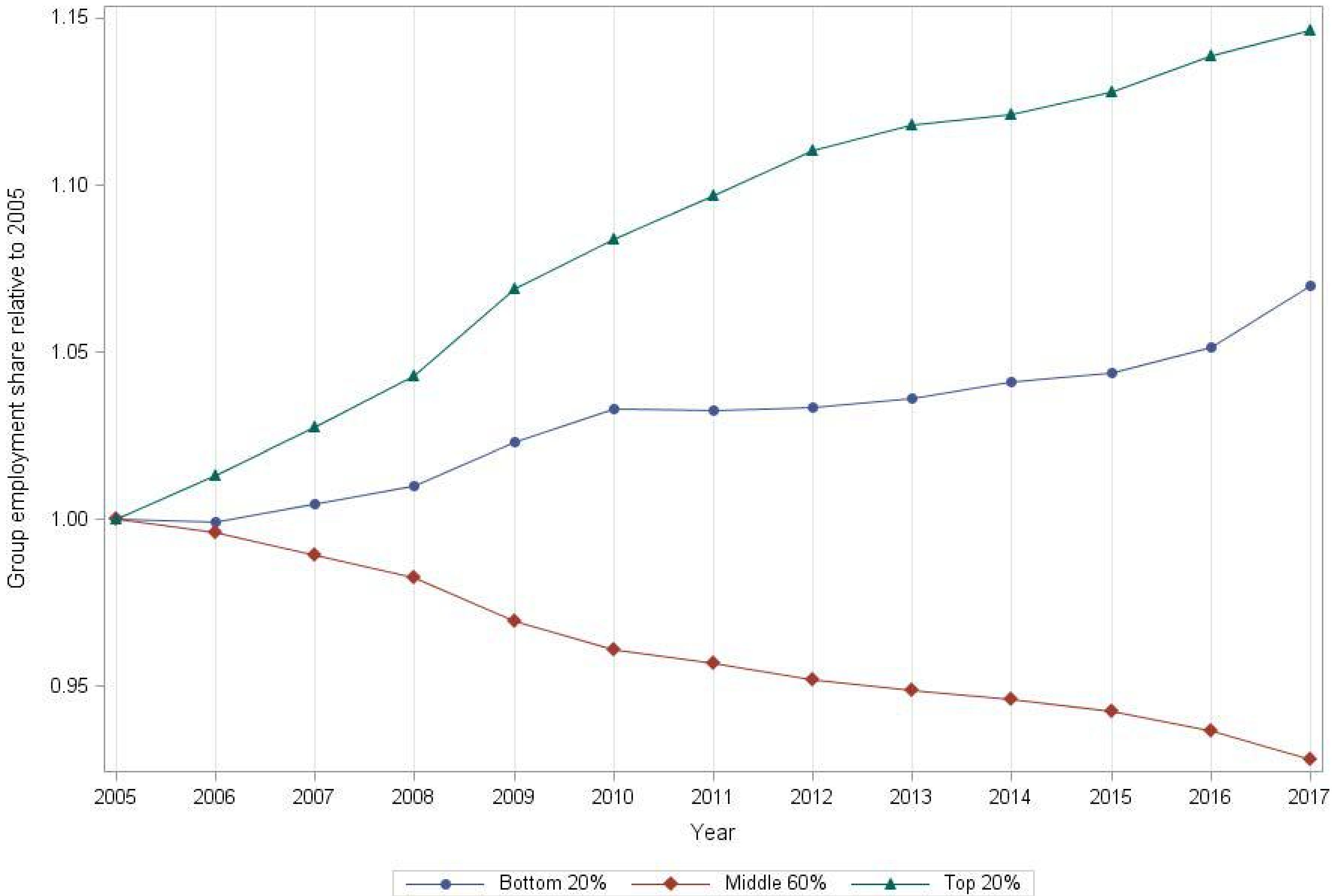
Note: Cumulative differences are summed from the lowest paying occupation (01) to the highest paying occupation (15).

Figure 1. Evolution of Major Occupation employment shares



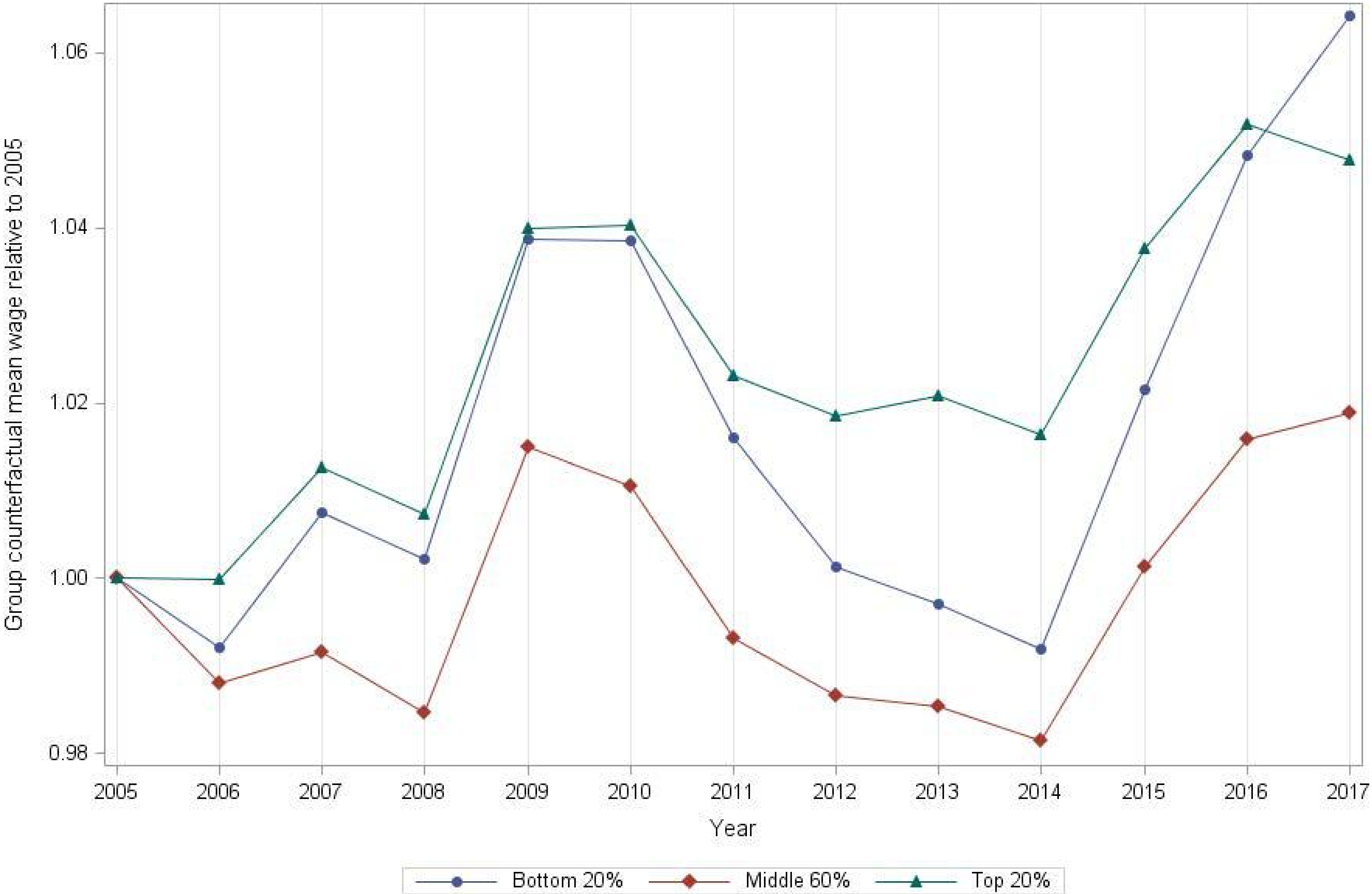
Note: Authors' calculations based on OES data.

Figure 2. Evolution of 2005 Wage Group employment shares



Note: Authors' calculations based on OES data.

Figure 3. Evolution of 2005 Wage Group counterfactual mean wages



Note: Authors' calculations based on OES data. The counterfactual mean wage holds detailed occupational employment levels fixed at 2005 levels, but allows occupational wages to change over time.

Figure 4. Regression tree details

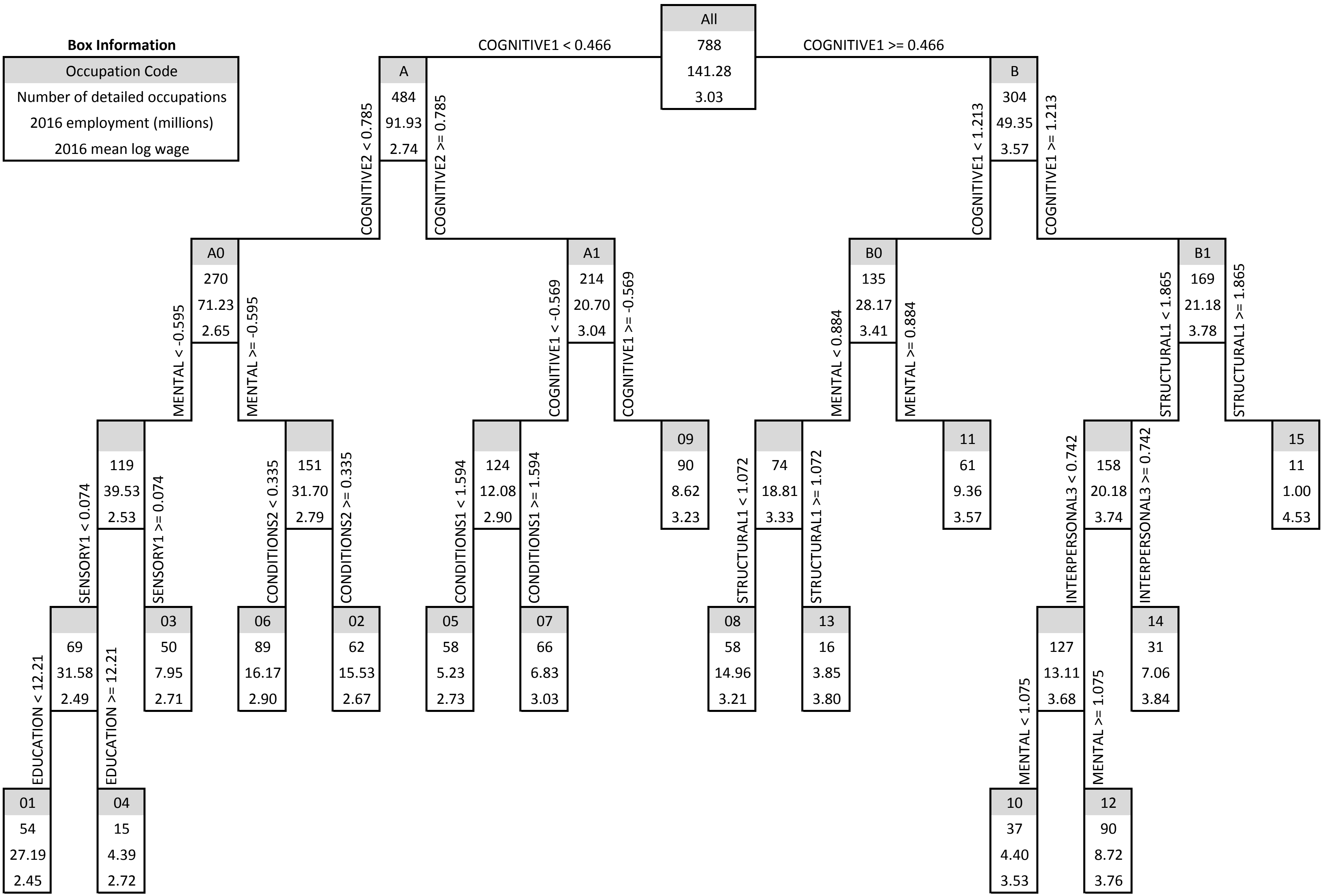
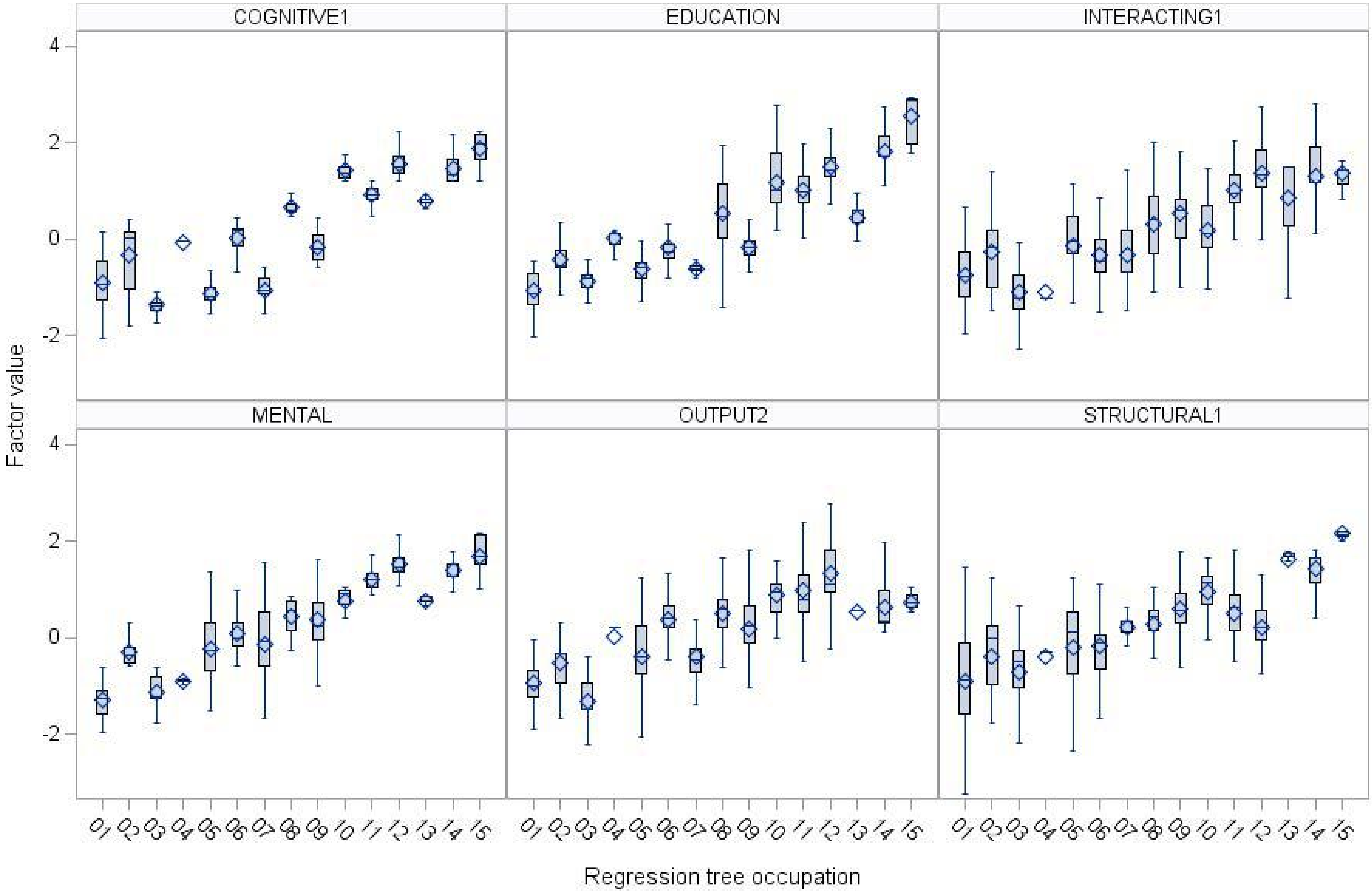
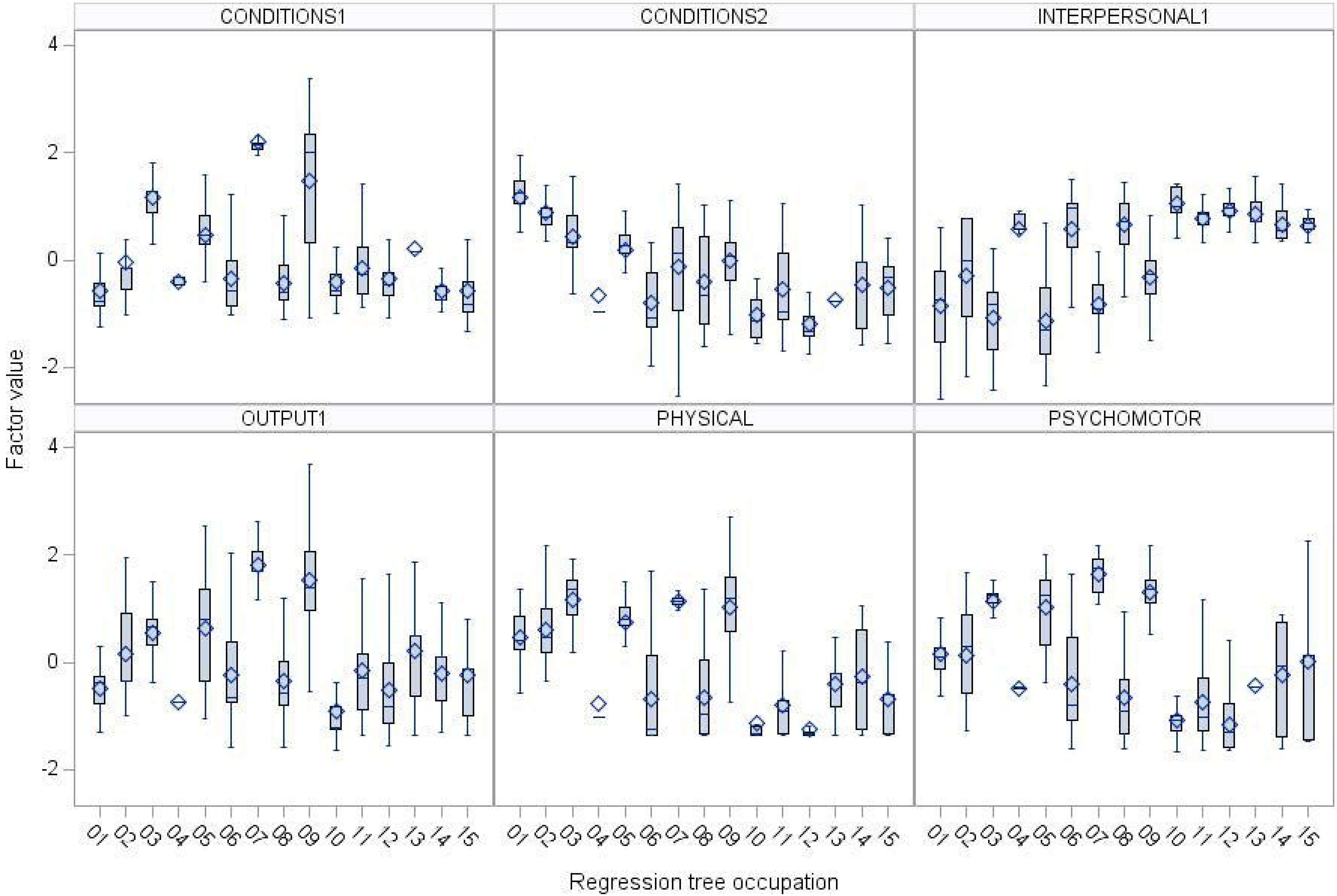


Figure 5.a. Distributions of key factors by regression tree occupations



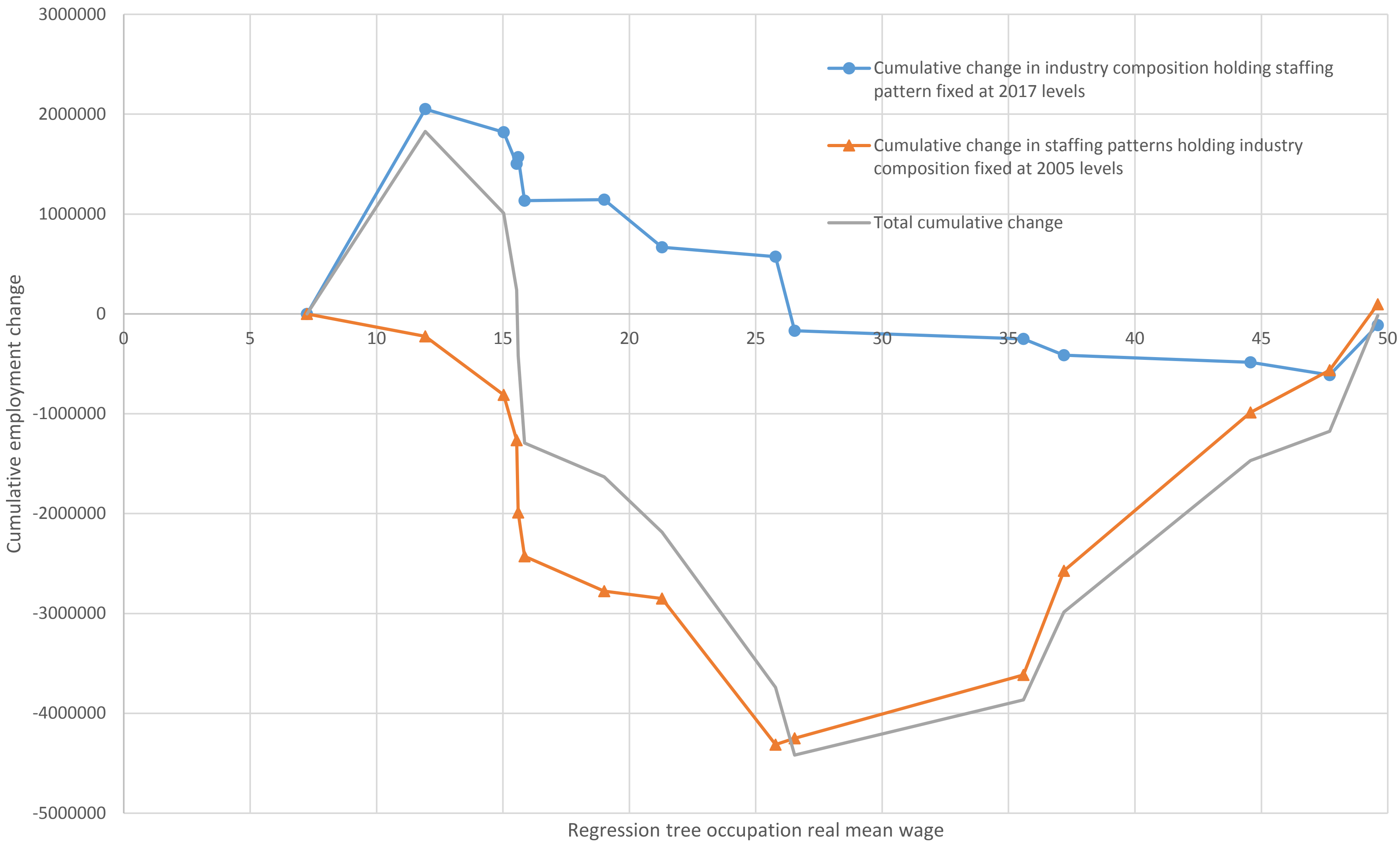
Note: Statistics are calculated using 2016 occupational employment weights. The EDUCATION variable is normalized years of education so that the weighted mean equals 0 and variance equals 1.

Figure 5.b. Distributions of key factors by regression tree occupations

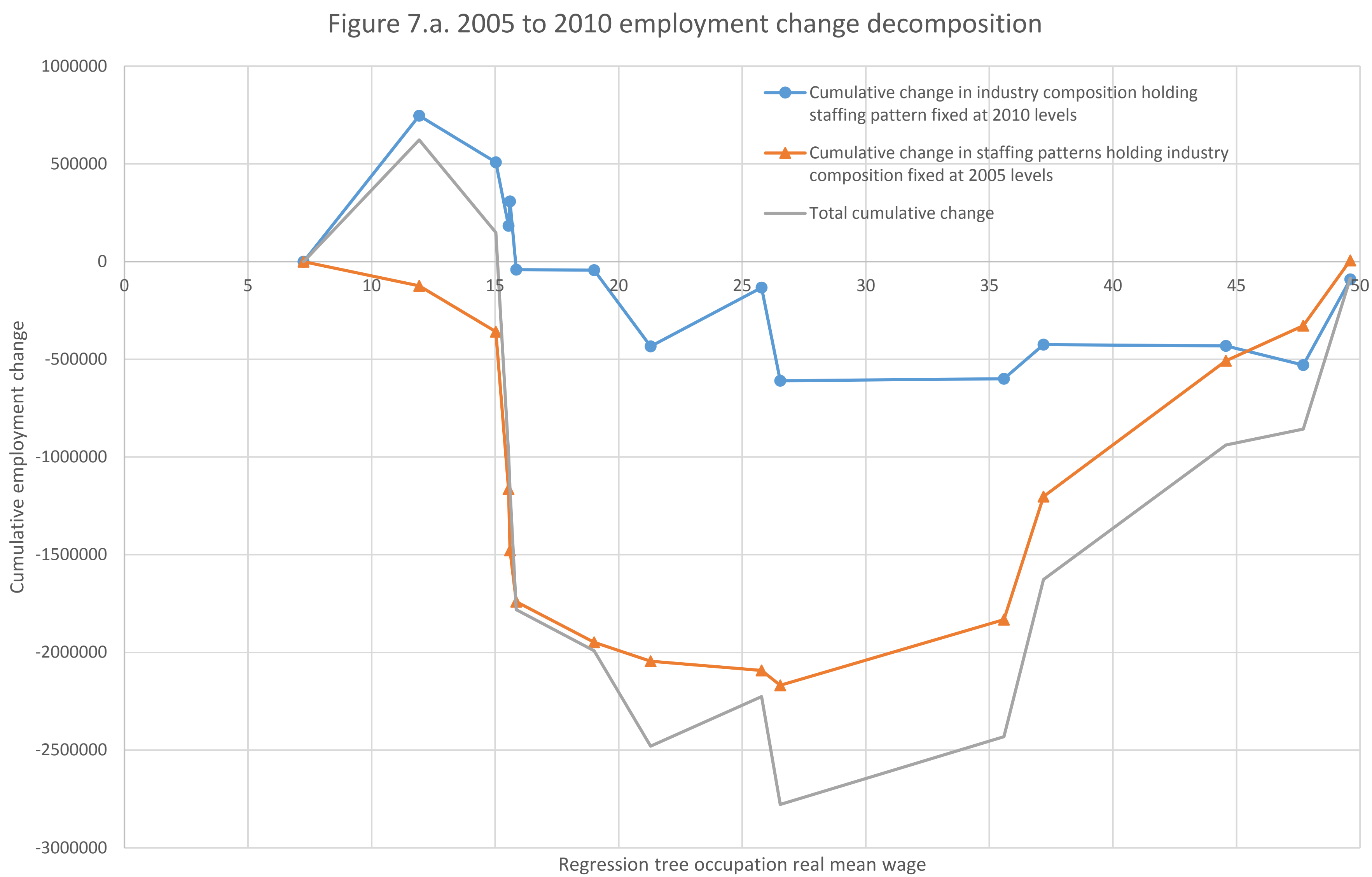


Note: Statistics are calculated using 2016 occupational employment weights.

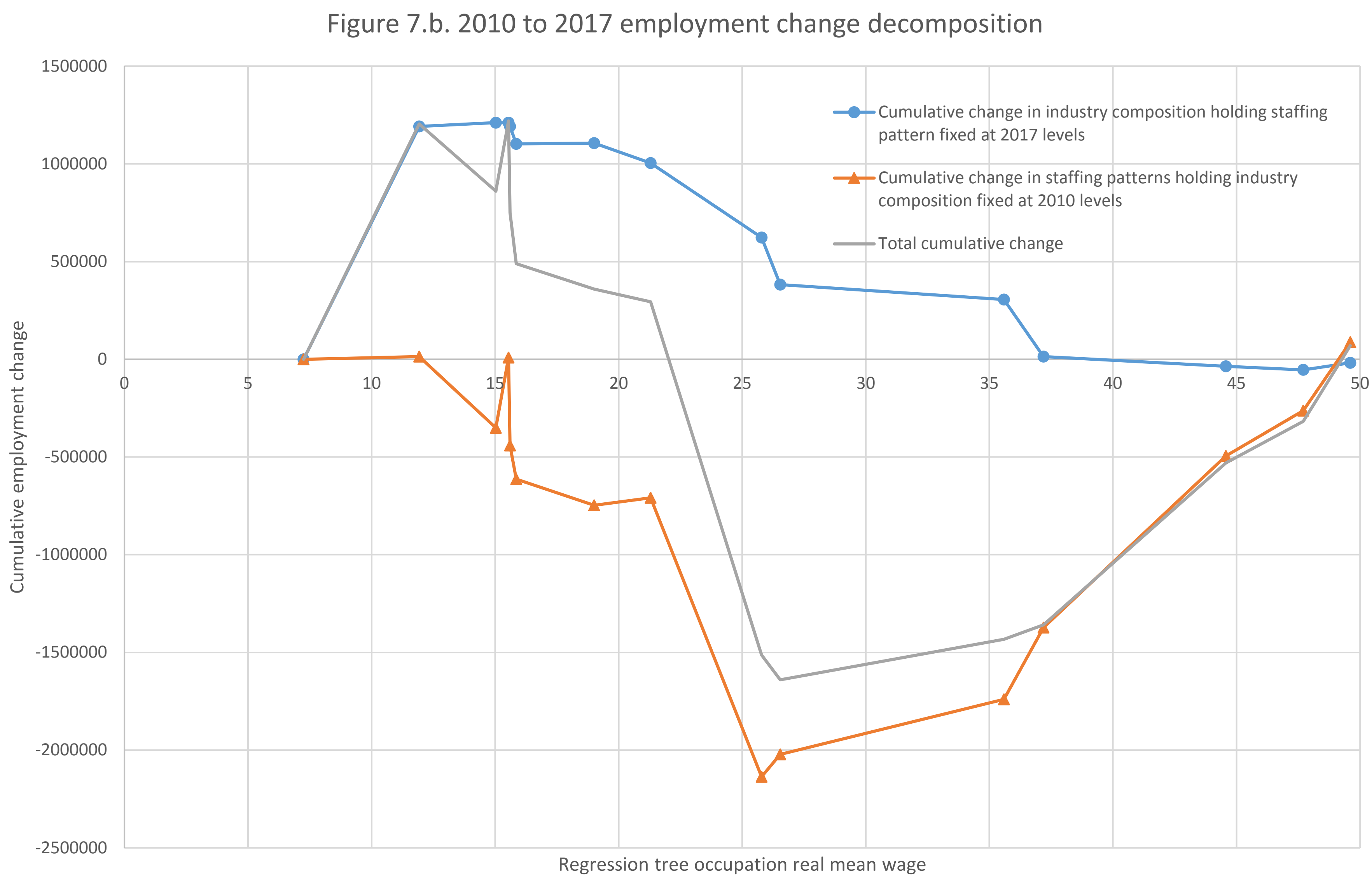
Figure 6. 2005 to 2017 employment change decomposition



Note: Each point in the graph represents a regression tree occupation. Changes for the regression tree occupations are summed from lower paying occupations to higher paying occupations (left to right). The graph begins at $7.25, the federal minimum wage in 2017. For scaling purposes, we do not include the highest paying occupation (15) in the graph.

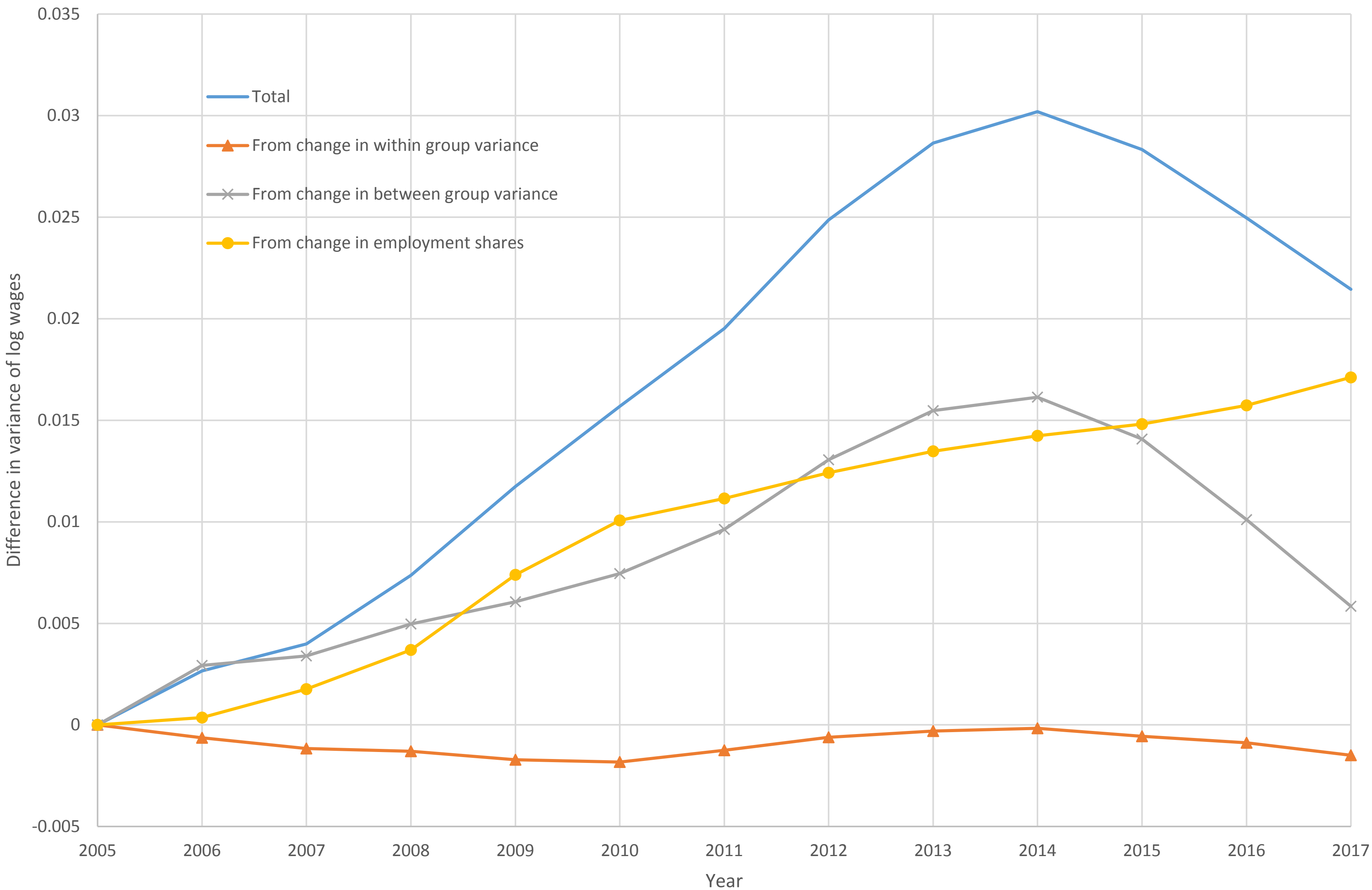


Note: Each point in the graph represents a regression tree occupation. Changes for the regression tree occupations are summed from lower paying occupations to higher paying occupations (left to right). The graph begins at $7.25, the federal minimum wage in 2017. For scaling purposes, we do not include the highest paying occupation (15) in the graph.



Note: Each point in the graph represents a regression tree occupation. Changes for the regression tree occupations are summed from lower paying occupations to higher paying occupations (left to right). The graph begins at $7.25, the federal minimum wage in 2017. For scaling purposes, we do not include the highest paying occupation (15) in the graph

Figure 8. Decomposition of difference in variance of log wages relative to 2005



Note: Wages are detailed occupational mean wages. We ignore any wage variation within detailed occupations. Groups in this decomposition are the 15 regression tree occupations described in the text.