A grade of membership approach to grouping occupations

Steven Whitlow

March 23, 2021

Abstract

I group occupations based on their task content by estimating a hierarchical Bayesian model. In contrast to traditional classification methods, the Grade of Membership (GoM) model I implement allows for occupations to have partial membership in multiple groups. I apply the model to occupational task content from O*NET. After a k-fold cross-validation exercise, I estimate a version of the model with seven distinct extreme types using variational inference. Most occupations are characterized by a high degree of membership in one primary type, but also a substantive level of mixing over types. I show that the estimated memberships are intuitive and can explain much of the variation in occupational mean wages. Using data from the Displaced Worker Supplement of the CPS, I find that wage changes after a displacement event depend both on the direction and distance of post-displacement occupational switches, as argued by Robinson (2018). Workers who switch occupations suffer higher wage losses; in addition, those who make a downward (upward) occupational switch suffer higher (lower) wage losses, with the magnitude depending on the distance of the switch. In particular, workers who switch in the direction of low-skill service occupations suffer steep wage losses. Switching narrowly defined occupations is associated with wage losses after controlling for the direction and distance of the switch, suggesting occupational specificity of human capital, as argued by Kambourov and Manovskii (2009).

1 Introduction

A large literature in labour economics examines the returns to job tenure and the degree to which accumulated skills are transferable across different jobs, recognizing that human capital is unlikely to be purely general.[1] Since microdata typically include information on the industries and occupations worked by individuals, it is natural to consider the extent to which human capital accumulation is industry- or occupation-specific: for instance, Neal (1995) finds large wage losses for displaced workers who switch industries, while Kambourov and Manovskii (2009) find large returns to occupational tenure, even for narrowly defined occupations. An alternate but related approach to considering human capital as occupation-specific is to treat human capital as being specific to the tasks that are performed on the job. The Dictionary of Occupational Titles (DOT) and its successor O*NET are two datasets from the U.S. Department of Labor which provide information on the degree to which different occupations utilize different tasks. In this sense, a workers' occupation can be thought of as a noisy proxy for the tasks they perform on the job, allowing for the estimation of task-specific human capital accumulation processes (Yamaguchi 2012; Sanders 2016; Guvenen et al. 2020; Lise and Postel-Vinay, forthcoming). In addition, the job task literature has also considered how changes in labour demand may result in alternate assignments of workers to tasks. For instance, technological advances have resulted in a decline in the relative demand for routine tasks which can be easily automatized (Autor, Levy, and Murnane 2003; Acemoglu and Autor 2011).

Although the U.S. Census classification has hundreds of occupational codes at the three-digit level, so too are there hundreds of task measures in O*NET. This implies that some sort of dimensionality reduction procedure must be performed on the job task measures before bringing them to the data.The approaches towards reducing the dimensionality of the task space of the DOT and O*NET data taken by the existing literature can be divided into two groups: those which result in continuous measures of task intensity and those which result in discrete classifications.

In this paper, I propose an alternate method of dimensionality reduction in which I estimate a Grade of Membership (GoM) model on the data from O*NET. The GoM model is a hierarchical Bayesian model which treats observations as mixtures over a given number of latent extreme types and is part of a larger family of mixed membership models (Airoldi et al. 2014). Latent Dirichlet Allocation (LDA), proposed as a text classification method by Blei, Ng, and Jordan

[1] See Sanders and Taber (2012) for a review of this literature.

(2003), is likely the most familiar model of this class to economists. LDA has been used to classify central bank communication (Hansen and McMahon 2016; Hansen, McMahon, and Prat 2018), news articles (Larsen and Thorsrud 2019), and patterns of CEO time use behaviour (Bandiera et al. 2020). The GoM model can be considered a close cousin of LDA; while LDA is applied to counts, such as the number of appearances of a given word in a document, the GoM model focuses on discrete responses to survey data. As opposed to LDA, GoM models have largely escaped the attention of economists.[2]

This can be thought of a hybrid of the existing two approaches: although, like discrete classifications, GoM estimation divides occupations into groups, it produces "soft" classifications in which occupations can be considered as partial members of multiple groups. GoM also has several advantages relative to other approaches which result in continuous measures. First, existing methods using data from the DOT or O*NET are limited in the fact that their task intensity measures are ordinal, since they are based on categorical responses to survey data.[3] One common technique to address this is to replace the task intensity measure for each occupation with its percentile rank; even with this transformation, it is still unclear the extent to which one can assess the distance between occupations in terms of task content. However, classifications produced by a GoM model are simply point estimates for parameters of a multinomial distribution for each occupation, which are cardinal measures and clearly sum to one by construction. Secondly, it tends to give intuitive results without the need to aggregate the descriptors into groups using a priori knowledge; much of the existing literature has been criticized on the grounds of the subjectivity inherent in the process of dimensionality reduction.[4]

I solve the model using variational inference; specifically, I use a variational Expectation Maximization algorithm (Wang and Erosheva 2015). I examine the estimated classifications and latent extreme occupational types in detail, arguing that both sets of estimates are highly intuitive despite the fact that the GoM model does not require the use of a priori knowledge. I also demonstrate that the estimated membership vectors can explain most of the variation in occupational mean wages in the 2000 U.S. Census. I then take the GoM estimates to data

[2]One exception is Munro and Ng (2020), who propose a model that can be seen as a simplification of the GoM model where individuals are assigned to only one latent type. GoM models have also been applied in marketing (Varki, Cooil, and Rust 2000; Dotson, Büschken, and Allenby 2020).

[3]For an alternative using different data, see Stinebrickner, Stinebrickner, and Sullivan (2019), who use data in which respondents state the percent of time on the job they perform on various tasks. The advantage of my approach using O*NET data is that it can be used in conjunction with common survey datasets such as the CPS or PSID.

[4]For an example of such a criticism, consider Autor (2013): "Researcher discretion again becomes paramount in this data construction process, and some transparency is inevitably lost. While I have found that task measures distilled from DOT and O*NET can serve as powerful proxies for occupational tasks, I am at best only moderately comfortable with these tools because their complexity and opacity places little discipline on how they are applied and interpreted."

from the Displaced Worker Supplement of the Current Population Survey and examine how occupation switches made by workers following a job displacement event map onto changes in wages. Poletaev and Robinson (2008) argue that the wage changes of displaced workers following occupational switches are driven by changes in the underlying skill or task portfolio as opposed to the loss of occupation-specific human capital.[5] I use as a starting point the argument of Robinson (2018) that the change in wage following an occupational switch should derive in part from the *distance* and *direction* in the task space associated with the switch. The GoM model I estimate invites simple measures of distance and direction. I define the distance of an occupational switch as the Hellinger distance between the estimated membership vectors of a workers' occupations before and after a displacement event. I determine the direction of the switch based on whether a worker moves down the occupational wage ladder predicted by the estimated membership vectors. I show that, as expected, the magnitude of wage changes following displacement depends on the distance of the switch, whereas the sign depends on the distance. After I control for both the distance and direction of occupational switches in the reduced task space, occupational switches at the narrow 3-digit level following displacement are associated with wage declines of about six percent, which suggests that occupational specificity of human capital may be an important aspect of accounting for wage losses, as argued by Kambourov and Manovskii (2009). Finally, I argue that the most severe wage losses are borne by workers who switch to low-skill service occupations following displacement.

The remainder of this paper is as follows. In Section 2, I provide an overview of existing approaches in the literature to reduce the dimensionality of the DOT and O*NET task data. In Section 3, I discuss the O*NET data and introduce the Grade of Membership model. I also introduce the variational Expectation Maximization algorithm developed by Wang and Erosheva (2015) to solve the model. In Section 4, I first perform a cross-validation exercise to select the number of extreme types in the GoM model before estimating the model. I then examine the estimates in detail. In Section 5, I take the estimates of the GoM model to data on displaced workers and show that I can use these estimates to construct coherent measures of distance and direction which can be used to explain changes in wages following displacement. Section 6 concludes.

2 Literature review

In this section, I describe the existing literature on deriving occupations' task intensity from sources such as the DOT or O*NET. I first discuss approaches which result in continuous

[5] Gathmann and Schönberg (2010) also find evidence for the task specificity of human capital.

measures of task intensity.[6] One strategy is to use a priori knowledge to divide variables in the task data into groups and assume that these subsets measure the degree to which a given occupation utilizes different tasks. For instance, in Autor, Levy, and Murnane (2003) routine manual activity is inferred from the DOT's measure of the level of finger dexterity used in the job. Other papers which construct measures from means of groups of O*NET intensity measures include Peri and Sparber (2009) and Deming (2017). Yamaguchi (2012) selects groupings of task measures but constructs the measure of task intensity by using Principal Components Analysis (PCA) within groups instead of taking a mean. Other papers, in the interest of allowing for variables in the data to reflect the intensity of multiple tasks, perform PCA on a larger set of measures without first dividing them into groups (Ingram and Neumann 2006; Poletaev and Robinson 2008; Robinson 2018). However, PCA can result in principal components which can sometimes be hard to interpret; Lise and Postel-Vinay (forthcoming) rotate the principal components so as to satisfy exclusion restrictions that make the resulting factors easily interperetable.[7]

An alternative to the methods described above is to instead divide occupations into mutually exclusive groups. Since most occupation classification schemes are hierarchical, one strategy is to aggregate by groupings of occupations; for instance, Cortes (2016) uses two-digit classifications from the U.S. Census. Grigsby (2019) constructs fifteen separate occupation groups by first dividing occupations into terciles based on their educational requirements and then dividing each of those three groups into five using a k-means clustering algorithm using O*NET data.

3 Data and methodology

3.1 O*NET data

I use data from the Occupational Information Network, better known as O*NET, to measure skill requirements for each occupation. O*NET collects a wide range of information on occupational characteristics through surveying both individuals working in given occupations as well as professional analysts. I focus on the data from three of the O*NET files, namely abilities, skills, and work activities.[8]

[6]Note that here I focus on approaches to reduce the dimensionality of the task data into a small number of underlying skills. If one is only interested in the differences between two occupations, one could use measures like angular separation (Gathmann and Schönberg 2010).

[7]For example, one restriction is that the O*NET score for "mathematics knowledge" reflects only underlying cognitive skills.

[8]As of O*NET 24.3, the version I use, the data on work activities is derived from surveys of workers, while the data on abilities and skills is derived from surveys of occupational analysts. In earlier incarnations of O*NET worker activities were also inferred from surveys of analysts.

For each descriptor, survey respondents are asked two questions. First, they are asked to rate the degree the given task or skill is important on a Likert-like scale, where a response of "1" corresponds to "Not Important" and "5" corresponds to "Extremely Important". Second, respondents are asked to rate the level of the given task or skill required for the given job by providing an integer between 1 and 7, where 7 represents the highest level of the task or skill, except if they had already ranked the descriptor as "not important", in which case a value of zero is imputed for the respondent. For the level question, O*NET guides responses by labelling three of the possible responses with an example of an action that would require that level of a given skill. Figure 1 provides an example of the O*NET questions for the Oral Comprehension descriptor from the Skills file. I choose to estimate the model using the responses for the level category associated with each descriptor.[9]

While the GoM model can easily be extended to allow for multiple responses for a given variable to be associated with each observation, O*NET unfortunately does not report the raw survey responses. Instead, they report the mean response for the level of each descriptor across all respondents; hence, I must rescale this continuous measure back into a discrete one. Since most applications of O*NET data use the mean responses directly, I am not aware of any best practices in the literature for rescaling these data back into discrete values. I could do so by grouping descriptors into discrete categories based on arbitrary thresholds; however, this is potentially problematic since the distributions of mean responses vary across descriptors. While this could be reflective of real differences in the underlying levels of given skills or tasks across occupations, this may also be a result of respondents using the examples for each question as an anchor for their own responses.[10] Though this should not affect the estimates of group membership by occupation, it would make the estimated subpopulation parameters difficult to interpret.

I choose to normalize these ratings by transforming the means of responses into a categorical variable which takes values $X_{i,j} \in \{1, 2, 3\}$ based on the position of occupation i in the descriptor-specific distribution of mean level ratings for descriptor j. For instance, $X_{i,j}$ would take a value of 1 if occupation i is in the first tercile of mean level ratings for descriptor j. The responses for each occupation should thus be interpreted as reflecting the levels of ability in a given skill or task required *relative* to all other occupations in the O*NET data.

As of O*NET 24.3, there are $N = 968$ occupations in the data. There are 52 descriptors of

[9] I have tried using the values for importance instead, which gives qualitatively similar results. While I could use both values, I have not done so since this would greatly expand the dimensionality of the parameter space, which is already very large.

[10] The ambiguity of the O*NET level scales has been criticized in the past (see, e.g. Autor 2013). In addition, the precise levels for which the questionnaire gives examples varies across descriptors; to the extent respondents use these examples as an anchor, this is problematic for comparing levels across descriptors.

abilities, 35 descriptors of skills, and 41 descriptors of work activities, giving a total of $J = 128$ descriptors. Beyond the transformation described above, I do not perform any other cleaning on the data. In principle, I could first group descriptors into categories, similar to the previous literature described in Section 2, to reduce the size of J before taking the model to the data. I choose not to do this so as to demonstrate that the GoM model is capable of delivering intuitive classifications without resorting to a priori knowledge.

3.2 Grade of membership model

The Grade of Membership model dates to Woodbury, Clive, and Garson Jr (1978), originally being proposed as a statistical aide to medical diagnosis. In its initial formulation, the latent individual degrees of membership were treated as fixed constants; the model itself can be solved by maximum likelihood.[11] Erosheva (2002) instead proposes a hierarchical Bayesian framework in which the individual parameters are realizations from a Dirichlet distribution, showing that such a formulation can be thought of as part of a larger class of independently proposed mixed membership models, including LDA and a similar approach developed by Pritchard, Stephens, and Donnelly (2000) for use with genetic data. This Bayesian formulation has since been used in categorizing scientific publications Erosheva, Fienberg, and Lafferty (2004), political ideologies of survey respondents (Gross and Manrique-Vallier 2014), and types of disabilities from survey data Erosheva, Fienberg, and Joutard (2007).[12] Note that in this paper I refer to the Bayesian approach of Erosheva (2002) as the "Grade of Membership model", as this is the model I take to the O*NET data.

As a reminder, occupations are indexed by $i = 1 \ldots, N$ and descriptors by $j = 1, \ldots, J$. Denote the vector of responses for occupation i as $\boldsymbol{X}_i = \{x_{i,1}, \ldots, x_{i,J}\}$. Assume there are K extreme profiles of which each occupation can take partial membership, which I refer to as *extreme types*, or more colloquially, *groups* of occupations. Define $\boldsymbol{\lambda}_i = (\lambda_{i,1}, \ldots, \lambda_{i,K})$ as the latent distribution over extreme profiles for occupation i, where $\sum_{k=1}^{K} \lambda_{i,k} = 1 \; \forall \; i$. Further, let $\theta_{j,k,l}$ be the probability that extreme profile k is associated with response l for descriptor j, where $\boldsymbol{\theta}_{j,k} = \{\theta_{j,k,1}, \theta_{j,k,2}, \theta_{j,k,3}\}$ is the vector of those probabilities. Evidently, $\sum_{l=1}^{3} \theta_{j,k,l} = 1 \; \forall \; j, k$.

Since the GoM model is generative, it is intuitive to describe the model by specifying the hypothetical process through which the data are generated. With the above definitions, I am now in a position to describe the generative process, which is as follows (Erosheva, Fienberg, and Joutard 2007):

[11] Manton, Woodbury, and Tolley (1994) develop a fixed-point approach to solve this formulation of the GoM model.

[12] Gormley and Murphy (2009) extend the model for use with ranked data, applying this to exit poll data from a presidential election in Ireland, a country which uses ranked choice voting for presidential elections.

1. For each occupation i in the data, $i = 1, \ldots, N$, draw $\boldsymbol{\lambda}_i \sim \text{Dirichlet}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_K\}$.

2. For each occupation i and each descriptor j:

   (a) From the distribution over groups for occupation i, draw a group for the descriptor. Formally, let $Z_{i,j} \sim \text{Multinomial}(\boldsymbol{\lambda}_i)$. $Z_{i,j}$ represents the group for which occupation i can be associated with with respect to descriptor j.

   (b) From the distribution over responses for that group, draw a response. Formally, let $X_{i,j} \sim \text{Multinomial}(\theta_{j,Z_{i,j}})$ be the response for descriptor j associated with occupation i.

It is instructive to compare the generative process for the GoM model to that of LDA. In LDA, the equivalent of groups are "topics", and instead of descriptors, there are "words"; the topic-word distribution can be expressed as a J dimensional object for each topic, where J is the number of words. In the GoM model these subpopulation parameters are $J \times L$, where J is the number of questions in the survey and L is the number of possible responses for each question. This is because each question for each group in GoM has its own multinomial distribution; in LDA, the bag-of-words assumption implies that after conditioning on topic, each word is drawn from the same topic-word distribution. This difference makes GoM better suited for applications to discrete survey data.[13]

3.3 Estimation

Likelihood. Before discussing the estimation procedure I use, I first write down the likelihood function for the model. Given the latent membership vector $\boldsymbol{\lambda}_i$, the conditional probability of observing response $x_{i,j}$ can be written as (Erosheva, Fienberg, and Lafferty 2004; Gross and Manrique-Vallier 2014):

$$\Pr(X_{i,j} | \boldsymbol{\lambda}_i, \boldsymbol{\theta}) = \sum_{k=1}^{K} \lambda_{i,k} \theta_{j,k,X_{i,j}}.$$

Note that implicit in the generative process for GoM described above is the assumption that the response variables $x_{i,j}$ are independent conditional on the latent membership vector $\boldsymbol{\lambda}_i$.

[13] By reshaping the $X_{i,j}$ into "words", one could estimate an LDA model using this data. For instance, if the skill "Active Learning" takes value 2 for some occupation, that could be written as one word: `activelearning2`. By creating a vector of "words", one could then apply LDA. This approach is used by Draca and Schwarz (2020). However, the local independence assumption is violated in the case of survey data: for instance, with the O*NET survey data, exactly one "word" which contains the string `activelearning` will appear in the vector associated with each occupation. This makes the results difficult to interpret.

This is referred to as the local independence assumption (Holland and Rosenbaum 1986). This assumption implies that one can write the conditional probability of occupation j as

$$\Pr(\boldsymbol{X_i}|\boldsymbol{\lambda}_i,\boldsymbol{\theta}) = \prod_{j=1}^{J}\sum_{k=1}^{K}\lambda_{i,k}\theta_{j,k,X_{i,j}},$$

and using the fact that occupations are independent:

$$\Pr(\boldsymbol{X}|\boldsymbol{\lambda},\boldsymbol{\theta}) = \prod_{i=1}^{N}\prod_{j=1}^{J}\sum_{k=1}^{K}\lambda_{i,k}\theta_{j,k,X_{i,j}}. \tag{1}$$

In the original formulation of GoM, (1) was estimated directly. However, recall that in the Bayesian framework of Erosheva (2002), the latent membership vectors are distributed as Dirichlet with parameters $\boldsymbol{\alpha}$; denoting this distribution as as $D_{\boldsymbol{\alpha}}$, one can write the marginal likelihood by integrating over the latent membership vectors:

$$\Pr(\boldsymbol{X}|\boldsymbol{\alpha},\boldsymbol{\theta}) = \prod_{i=1}^{N}\int_{\Delta^{K-1}}\prod_{j=1}^{J}\sum_{k=1}^{K}\lambda_{k}\theta_{j,k,X_{i,j}}D_{\boldsymbol{\alpha}}(d\lambda). \tag{2}$$

Unfortunately, (2) does not have a closed form solution and evaluating it is intractable in general, because it involves integration over the $K-1$ dimensional simplex. This implies that some sort of approximation is required to calculate the posterior and thus fit the model.

Variational inference. The two primary options for approximate inference for GoM models are Markov Chain Monte Carlo (MCMC) and variational inference. I choose to use a variational inference method to estimate the model.[14] An MCMC solution method was proposed in the context of GoM models by Erosheva (2002); since MCMC is familiar to economists, I do not describe it here. Variational inference is an approximate estimation technique from machine learning commonly used in models with latent variables and which is often far less computationally expensive than MCMC. While I provide only a brief summary, Blei, Kucukelbir, and McAuliffe (2017) provide a introduction to VI methods for a statistical audience. In short, while MCMC revolves around sampling from the posterior, variational inference involves replacing the true posterior with a simple analytical approximation. Applications of variational inference within economic contexts include Ruiz, Athey, and Blei (2019) and Draca and Schwarz (2020).

Variational inference approximates the true posterior over the latent variables with a simpler variational distribution. A variational distribution which assumes that the latent variables

[14] In particular, I use the `R` package `mixedMem` (Wang and Erosheva 2015).

are independent is said to be part of the mean field variational family (Blei, Kucukelbir, and McAuliffe 2017). A mean field approximation approximation of the posterior for the GoM model is given by (Wang and Erosheva 2015; Wang, Matsueda, and Erosheva 2017)

$$q(\boldsymbol{\lambda}, \boldsymbol{Z}|\boldsymbol{\phi}, \delta) = \prod_{i=1}^{N} \text{Dir}(\boldsymbol{\lambda}_i|\boldsymbol{\phi}_i) \prod_{j}^{J} \text{Mult}(Z_{i,j}|\delta_{i,j}), \tag{3}$$

where $\boldsymbol{Z}$ is the matrix containing the group-descriptor assignments for each occupation and δ and $\boldsymbol{\phi}$ are referred to as variational parameters. Note that these variational parameters vary across occupations. One can now consider reformulating (2) as an expectation over the variational approximation of the posterior (Wang, Matsueda, and Erosheva 2017):

$$\begin{aligned}
\log(\Pr(\boldsymbol{X}|\boldsymbol{\alpha}, \boldsymbol{\theta})) &= \log\left[\int_{\lambda}\int_{\boldsymbol{Z}} \Pr(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\lambda}|\boldsymbol{\alpha}, \boldsymbol{\theta}) d\boldsymbol{\lambda} d\boldsymbol{Z}\right] \\
&= \log\left[\int_{\lambda}\int_{\boldsymbol{Z}} \frac{q(\boldsymbol{\lambda}, \boldsymbol{Z}|\boldsymbol{\phi}, \delta)}{q(\boldsymbol{\lambda}, \boldsymbol{Z}|\boldsymbol{\phi}, \delta)} \Pr(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\lambda}|\boldsymbol{\alpha}, \boldsymbol{\theta}) d\boldsymbol{\lambda} d\boldsymbol{Z}\right] \\
&= \log \mathbb{E}_q\left[\frac{\Pr(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\lambda}|\boldsymbol{\alpha}, \boldsymbol{\theta})}{q(\boldsymbol{\lambda}, \boldsymbol{Z}|\boldsymbol{\phi}, \delta)}\right].
\end{aligned} \tag{4}$$

Applying Jensen's inequality to (4), one can obtain an expression for a lower bound on the evidence, referred to as the ELBO:

$$\log(\Pr(\boldsymbol{X}|\boldsymbol{\alpha}, \boldsymbol{\theta})) \geq \underbrace{\mathbb{E}_q \log(\Pr(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\lambda}|\boldsymbol{\alpha}, \boldsymbol{\theta})) - \mathbb{E}_q \log(q(\boldsymbol{\lambda}, \boldsymbol{Z}|\boldsymbol{\phi}, \delta))}_{\text{ELBO}(q)}. \tag{5}$$

In fact, an exact relationship between the left and right hand side of (5) can be found (Jordan et al. 1998; Blei, Kucukelbir, and McAuliffe 2017):

$$\log(\Pr(\boldsymbol{X}|\boldsymbol{\alpha}, \boldsymbol{\theta})) = \text{ELBO}(q) + \text{KL}\left(q(\boldsymbol{\lambda}, \boldsymbol{Z}|\boldsymbol{\phi}, \delta)||\Pr(\boldsymbol{\lambda}, \boldsymbol{Z}|\boldsymbol{X})\right), \tag{6}$$

where the second term on the right hand side is the Kullback-Leibler divergence between the true posterior and its variational approximation. Intuitively, as the KL divergence between the true posterior and its approximation $q(\cdot)$ approaches zero, maximizing the ELBO becomes equivalent to maximizing the true posterior.

Wang and Erosheva (2015) use a variational Expectation-Maximization algorithm to fit the model. In the E-step, $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ are held constant and the ELBO is maximized with respect to the occupation-level variational parameters $\boldsymbol{\phi}$ and δ. In the M-step, the Newton-Raphson method is used to maximize the ELBO with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$, holding fixed $\boldsymbol{\phi}$ and δ. Wang, Matsueda, and Erosheva (2017) describe in detail the variational EM algorithm for the GoM model.

Variational inference has several advantages relative to MCMC. As noted above, MCMC tends to be more computationally expensive, especially with large datasets. Secondly, note that mixed membership models such as the GoM model are invariant to label switching (permutations of the group labels). With MCMC, special care must be taken to prevent label switching during estimation, but since variational inference does not rely on sampling from the posterior, the results only depend on the starting values for $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$, as well as the variational parameters $\boldsymbol{\phi}$ and δ (Wang and Erosheva 2015).

That said, the variational EM algorithm only guarantees convergence to a local mode. In general, the ELBO is non–convex, so different initializations may result in different optima.[15] Blei, Kucukelbir, and McAuliffe (2017) note that, in the context of mixed membership models, many of these seemingly different modes are in fact the result of innocuous permutations of the group labels, and thus represent equivalent models. However, the best practice is to try as many initializations as are computationally feasible. In my k-fold cross-validation exercise in the next section, I consider 24 different initializations for each fold. That said — despite the simplicity of the mean-field approximation (3) — in cases where both MCMC and variational methods have been used to estimate GoM models, the results have been shown to be qualitatively similar, suggesting that the variational approximation is reasonable (Erosheva, Fienberg, and Joutard 2007; Wang and Erosheva 2015).

4 Grade of Membership results

Model selection. I now proceed to estimate the model. The choice of the number of groups, K, to use in estimation is a classic model selection problem. To select the number of groups, I divide the data into five folds and perform a cross-validation exercise. Note that, as I described in Section 2, model selection in applications of O*NET data has often been informal, making use of a priori knowledge. Since the goal of the estimation is to cluster occupations such that the resulting partial membership vectors have economic meaning — rather than, for instance, minimize prediction error — formal model selection is not as important as in many machine learning applications.[16]

As noted in Section 3, the variational EM algorithm requires me to specify starting values for both $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$.[17] I follow the approach suggested in Wang and Erosheva (2015). Firstly, I

[15] Figure 2 of Blei, Kucukelbir, and McAuliffe (2017) provides a graphical example of this.

[16] For instance, I could arbitrarily choose $K = 2$, in which case occupations are assigned partial membership in what could be labelled "manual" and "cognitive" groups. This is clearly a distinction that has economic meaning, but as I show shortly, the held-out ELBO is much lower when $K = 2$ compared to higher levels of K.

[17] In all initializations, $\boldsymbol{\phi}$ and δ are initialized uniformly across groups, as recommended by Wang and Erosheva

choose starting values for $\boldsymbol{\alpha}$ by initializing occupation-group distribution as symmetric Dirichlet: $\boldsymbol{\alpha} = \{\tilde{\alpha}, \ldots, \tilde{\alpha}\}$ for some constant $\tilde{\alpha}$. Secondly, I initialize the group-descriptor distribution $\boldsymbol{\theta}$ by taking draws from a symmetric Dirichlet distribution with parameter $\tilde{\theta}$ to form each slice $\boldsymbol{\theta}_{j,k}$.[18]

I try two different starting values for the occupation-group parameters: $\tilde{\alpha} \in \{0.6, 1.6\}$. I try four different Dirichlet parameters for initializing the group-descriptor distribution: $\tilde{\theta} \in \{\frac{2}{3}, 1, \frac{4}{3}, \frac{5}{3}\}$. Finally, I try three different seeds for the random number generator used to take draws from the Dirichlet($\tilde{\theta}$) distribution. I estimate the model using each possible combination of these initialization parameters, so there are a total of $2 \times 4 \times 3 = 24$ different initializations.

For the cross-validation exercise, since the GoM is an unsupervised model, the value of interest is the held-out ELBO.[19] The procedure I follow for the cross-validation exercise is as follows. First, I randomly divide the data into five folds of approximately equal size: $f \in \{1, \ldots, 5\}$. Then, for a given number of types $K \in \{2, \ldots, 7\}$ and testing fold f_{test}, I estimate the model using each of the 24 different initializations via variational EM using all folds except f_{test} as a training set. In total, with 24 initialization points, 5 folds, and 6 alternatives for K, the cross-validation exercise involves estimating $24 * 5 * 6 = 720$ GoM models.

For each initialization, after obtaining fitted values $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\theta}}$, I then calculate the held-out ELBO for fold f_{test} as follows. First, I obtain classifications (formally, the variational parameters) for the occupations in f_{test} by applying a single E-step, holding fixed the estimated $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\theta}}$ obtained from the training set. Second, I calculate the ELBO for the occupations in f_{test} using the fitted values from the training set $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\theta}}$ and the classifications from the E-step.

I define the held-out ELBO for a given K and fold f as the maximum held-out ELBO across all initializations. The held-out ELBO for a given K is thus the mean of these maximum held-out ELBO values.[20] Figure 2 demonstrates the results of this process. Among the values I test, the held-out ELBO is strictly increasing in K, so that $K = 7$ has the highest mean held-out ELBO. That said, the difference in the mean held-out ELBO associated with moving from $K = 6$ to $K = 7$ is relatively small relative to the gains at lower levels of K. In principle, I could estimate the model using higher values of K; I choose not to do so in part because the cross-validation process rapidly becomes more expensive as K expands and in part because I have found anecdotally that the groups become harder to interpret with a high value for K.

(2015).

[18] Since $\boldsymbol{\theta}$ is $J \times K \times 3$, $\boldsymbol{\theta}_{j,k}$ is a vector with three elements. Therefore, I take a total of $J \times K$ draws from this Dirichlet($\tilde{\theta}$) distribution in total to initialize the group-descriptor distribution.

[19] See Wang, Matsueda, and Erosheva (2017) for a different held-out ELBO testing exercise which involves one training and one testing set and selects the initialization simultaneously with K.

[20] This implies that different initializations may be used across folds within a single K. If I instead calculated the mean of the held-out ELBO values for each initialization and K defined the held-out ELBO for a given K as the maximum of these, I get very similar results.

I estimate the model with $K = 7$ using the full sample at each of the 24 initializations and select the initialization with the highest ELBO. The variational EM algorithm gives as output point estimates of the occupation-type distributional parameters $\hat{\boldsymbol{\alpha}}$, point estimates of the descriptor-type distributional parameters $\hat{\boldsymbol{\theta}}$, and point estimates of the variational parameters $\hat{\boldsymbol{\phi}}$ and $\hat{\delta}$.[21] Note that although the classifications of each occupation are not direct outputs of the EM algorithm, they can be backed out by using their implied posterior mean as a point estimate (Wang and Erosheva 2015):

$$\hat{\lambda}_{i,k} = \frac{\hat{\phi}_{i,k}}{\sum_{k=1}^{K} \hat{\phi}_{i,k}}.$$

Labelling of extreme types. With the estimates of $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\theta}}$ in hand, I can now proceed to examine the estimated classifications in detail, beginning with labelling the extreme types $k \in \{1, \ldots, 7\}$. Note that the creation of group labels in mixed membership models is a subjective process; the group labels I give each extreme type have no meaning in and of themselves beyond aiding with intuition. Further, since mixed membership models are invariant to permutations of the order of groups, I first reorder the extreme types in descending order of the coefficients on group membership derived from an OLS regression of occupational mean wages on the estimated membership scores by occupation.[22]

I use two sources of information in labelling the extreme types. Since O*NET provides titles for each occupation, I am able to use the titles of occupations very close to each extreme type to aid in labelling. I list the ten occupations with the highest degree of membership in each group in Table 1. For additional information on the character of extreme types, I use information from the $\hat{\boldsymbol{\theta}}$ matrix. In particular, for each extreme type k I examine the ten highest elements of the $\hat{\boldsymbol{\theta}}_{k,j}$ vector for $j \in \{1, 2, 3\}$. These are the descriptors of skills or tasks most likely to be in the first, second, and third tercile respectively of the distribution across occupations for that skill or task. I reproduce the list for the $\hat{\boldsymbol{\theta}}_{k,3}$ vectors in Table 2. The results for the first and second terciles are reproduced in Table A1 and Table A2 respectively.

I now discuss my group labels. The first two extreme types can be considered as "abstract" occupations and roughly correspond to the common "non-routine cognitive" grouping in the task literature. The first type, which I label "Scientific", is characterized by a high level of skills

[21]Note that a limitation of variational inference is that it does not provide standard errors. It is in principle possible to obtain nonparametric estimates of the standard errors. Chen, Wang, and Erosheva (2018) use a bootstrap procedure to obtain standard errors. In this case the computational expense of such an approach would be prohibitive. Although other alternatives have been proposed, including one based on the delta method (Giordano et al. 2017), I follow existing applications of mixed membership models in economics and abstract away from the estimation of standard errors.

[22]I discuss this regression in detail at the end of this section.

and tasks often described as "cognitive", especially of skills related to problem solving, but often requires a significant degree of manual skills as well. Examples of occupations close to this extreme type, with degrees of membership above 0.99, include materials or biomedical engineers, physical medicine physicians, and soil and plant scientists. The second type, which I label "Other Abstract", also requires high cognitive ability, but typically demands less ability in manual skills and tasks. Occupations with a high degree of membership in this extreme type include social scientists such as economists and political scientists, management occupations, and lawyers.

The next three occupations I describe as "mixed" because that they typically feature moderate required levels of cognitive skills. The first of these three, which I label "Skilled Manual" occupations, are in addition highly likely to require skill in operating, maintaining, and repairing equipment. Examples of occupations with a high degree of membership in this type include hydroelectric plant technicians and forest firefighters. The second of these types I label "Skilled Service": these occupations unsurprisingly demand skill in interpersonal tasks or skills such as "assisting and caring for others" and "social perceptiveness". Examples of Skilled Service occupations include practical nurses and midwives. I label the last of these set of extreme types "Routine Cognitive", since the occupations with high degrees of membership in this type also tend to be classed as having high relative levels of routine task content in the measure of Autor and Dorn (2013). These occupations are distinguished by requiring moderate levels of cognitive skills; examples close to the pure type include travel agents, tax preparers, and procurement clerks.

The final grouping of pure types I refer to as "low-skill". First, "Low-Skill Manual" occupations tend to require manual ability, but are much less likely to demand competency in cognitive tasks relative to the "Skilled Manual" group. They are also less likely to demand specialized skills such as repairing equipment. Occupations close to this extreme type include many often called "routine manual", and include production workers, rock splitters, and insulation workers. The final extreme type I label "Low-Skill Service". Occupations close to this type include food servers, fast food cooks, and orderlies.

Analysis of results. To this point, I have examined only occupations which are closest to one of the extreme types. However, it would be undesirable if all occupations were estimated as mixing over only one extreme type; if that were the case, there would be little reason to estimate the GoM model relative to a simpler classification method such as k-means clustering. Figure 3 plots the empirical distribution of estimated maximal degrees of membership — that is, $\max_k \hat{\lambda}_{i,k}$ — across all occupations in the O*NET data. It is clear that most occupations are characterized by a substantial degree of mixing over types. However, there is a spike around $\max_k \hat{\lambda}_{i,k} \approx 1$, indicating that some occupations are well characterized by one of the extreme types.

In Table 3, I give examples of the estimated grades of membership for a selection of other occupations. To provide a baseline for my results, I also reproduce the skill measures for these sample occupations which are used in Autor and Dorn (2013) and Lise and Postel-Vinay (forthcoming). The classifications are highly intuitive: for instance, the GoM classification gives both accountants and telemarketers high degrees of membership in the Routine Cognitive group, but the secondary group for accountants is Other Abstract while the secondary group for telemarketers is Low-skill Service.

To examine the classifications in more detail, I use data from the 2000 U.S. Census to obtain estimates of wages and employment by occupation. I merge the GoM classifications and the Census data using the crosswalk from Sanders (2016), which links O*NET-SOC codes to the 2000 Census occupation codes. I retain only full-time full-year workers who report a positive income.[23] 485 of the 968 occupations in the O*NET dataset have a counterpart Census code in the Sanders (2016) crosswalk.

I first plot visually the classification results. Since the estimated degrees of membership vector $\hat{\boldsymbol{\lambda}}_i$ has seven dimensions, I first combine the extreme types into the three categories reflecting the broad groups I discussed above: "Abstract", "Mixed", and "Low-Skill", so that I can then plot where occupations lie on the implied simplex directly. Since multiple occupations may have very similar estimated mixtures over types, I round the estimated membership scores to two decimal places and merge occupations which have identical rounded membership scores for legibility purposes.

The results are shown in Figure 4. The size of each point is determined by the sum of total employment at that area of the simplex expressed as a percent of the total workforce in 2000. The colour of the point is determined by the mean log hourly wage of all workers in occupations located at that area of the simplex, relative to the mean log hourly wage of all workers. The figure shows that although a large minority of workers are employed in occupations on the vertices of the simplex, most occupations are estimated to have substantial degree of mixing over the three broad groups. Further, as one would expect, mean wages relative to the overall mean wage are low in the vertex associated with "Low-Skill" occupations and high in the vertex associated with "Abstract" occupations.

Only a handful of occupations are estimated to have substantial degrees of membership in both Abstract and Low-Skill types, as indicated by the fact that few occupations are located

[23] Specifically, I drop all workers who report working less than 35 hours per week or less than 35 weeks in the previous year.

on the lower edge of the simplex or the lower half of the interior of the simplex. This is intuitive, since it would be peculiar if many occupations were estimated as being mixtures over the two categories. However, it implies that the assumption that the parameterization of the occupation-group distribution as Dirichlet is at odds with the data. The assumption that the membership vectors are distributed as Dirichlet, while common in mixed membership models, prevents realistic patterns of correlation across groups. The correlated topic model of Blei and Lafferty (2007) addresses this limitation of LDA by replacing the Dirichlet distribution with a logistic normal one. While this allows for correlations between topics that are not possible with the Dirichlet parameterization, the loss of the conjugacy of the Dirichlet distribution makes the correlated topic model much less tractable. Since Gibbs sampling is no longer possible, variational inference is the only feasible estimation method of the model. The restrictions on correlations across topics have not prevented the use of LDA in many text classification contexts. In general, the unrealistic correlations imposed by the Dirichlet parameterization pose problems only to the extent to which the researcher wishes to use the estimated parameters to generate new data. I am not aware of a similar analogue to the GoM model, though the GoM model and LDA are similar enough that a similar extension in principle should be possible; I leave this to future research.[24]

I now examine the link between the estimated and classifications and wages in more detail by estimating a series of descriptive regressions using the Census data. I consider the following regression, in which mean log wages in occupation i are given by:

$$w_i = \beta_0 + \sum_{k=1}^{K-1} \beta_k \lambda_{i,k} + \epsilon_i, \tag{7}$$

where $\lambda_{i,k}$ is occupation i's degree of membership in group k and ϵ_i is an error term. Note that the degrees of membership for each occupation are probabilities, so that they are bounded between zero and one and sum to one. Since a regression which included all of the estimated degrees of membership would be perfectly collinear, I am forced to omit one; I choose to omit the Low-Skill Service group. The coefficients on the degrees of membership, β_k can be interpreted as the estimated log difference in wages between an occupation with full membership in the extreme type k and the omitted extreme type.

Although it is indeed possible to consider descriptive regressions of mean log wages on existing skill measures, (7) is likely to be more informative. It is based on measures which have a clear interpretation, rather than being ordinal. Secondly, existing methods – be they

[24] An alternative would be to reduce the number of extreme types until the Dirichlet parameterization appears compatible with the estimated classifications, such as by choosing $K = 2$.

averaging groups of descriptors selected a priori, PCA, or k-means clustering — impose linear relationships on the task data. In contrast, a synthetic occupation generated from one of the extreme types in the GoM model is highly likely to look like real occupations in the O*NET data.

The results are listed in Table 4. The estimated degrees of membership can explain much of the variation in mean wages across occupations. In column 3 of Table 4, I instead run an analogue of (7) on individual-level data using education and age controls. Controlling for education compresses the estimated coefficients, indicating sorting of high-skill workers into extreme types with higher wages, but the ranking is mostly the same.

5 Application: Wage losses following job displacement

Overview and data. I now examine the extent to which viewing occupation changes through the lens of the estimated GoM model can assist in understanding wage losses following a job displacement event. Note that accurately assessing the costs of job loss involves carefully specifying the counterfactual; for instance, if one wishes to identify the cost relative to non-displacement, a control group is needed (Jacobson, LaLonde, and Sullivan 1993; Davis and von Wachter 2011). In contrast, the goal of this exercise is descriptive, namely to determine if the GoM model can assist in explaining why different displaced workers experience different changes in wages following reemployment. For this analysis, I examine log hourly wages instead of weekly wages.[25]

Several papers have examined occupational changes following job displacement in the context of a ladder in which occupations are ranked vertically by their mean wage (Huckfeldt 2016; Raposo, Portugal, and Carneiro 2019; Forsythe 2020).[26] As one expects, moves up (down) this ladder are associated with wage gains (losses). Poletaev and Robinson (2008) and Robinson (2018) examine occupational mobility for displaced workers through the lens of changes in task mixes, where tasks are estimated using PCA. Of particular interest for this paper is Robinson (2018), who argues that wage changes depend both on the *distance* and *direction* of occupational moves in the task space. The distance between two occupations is defined by Robinson as Euclidean distance in the estimated task space, while direction is simply the signs associated

[25]Clearly, changes in weekly wages depend on changes in hourly wages and changes in hours worked. Much of the literature has focused recovery of hours following job displacement events; Farber (2017) notes that a large portion of earnings losses for displaced workers is driven by inability to find full-time employment. I choose to use hourly wages as my dependent variable so as to more clearly focus on the impact of changes in the bundles of tasks which workers supply to firms following displacement.

[26]See, e.g. Groes, Kircher, and Manovskii (2015) for a study which uses a similar framework to examine occupational mobility more generally outside of the specific setting of job displacement.

with the difference between the task bundles of workers' current and former occupation. While the definition of distance is clear, the definition of direction is somewhat ambiguous except in cases where the sign of the change in a workers' task bundle are the same across all factors.

In this section, I use an alternate definition of direction: the difference between the fitted values associated with the current and former occupation obtained from the regression (7) using the 2000 U.S. Census data.[27] A similar exercise using traditional task measures is in theory possible, but it would be less informative since such an OLS regression would have no clear interpretation beyond showing that the factors can explain much of the variation in wages. On the other hand, as I argued in Section 4, a OLS regression using the cardinal measures of the degree of mixing over types by occupation has a clear interpretation. On the other hand, an exercise analogous to Robinson (2018) would not be possible using "hard" classification methods, since there is no partial membership of groups and therefore no measure of distance.[28] In my case, I use the Hellinger distance between the estimated membership vectors of workers' current and former occupations as a measure of distance. While I could use the Euclidean distance, the Hellinger distance is convenient to interpret due to the fact that it is bounded between zero and one:[29]

$$d(new, old) = \frac{1}{\sqrt{2}}\sqrt{\sum_{k=1}^{K}\left(\sqrt{\lambda_{new,k}} - \sqrt{\lambda_{old,k}}\right)^2}. \tag{8}$$

To examine the occupation switching patterns of displaced workers, I use the Displaced Worker Supplement to the Current Population Survey. Although the DWS has been administered in the January of each even year since 1984, I use the data from the years 2004-2010.[30] In my sample, I only include individuals who respond as having experienced a job displacement event in the three years previous to the survey, and report a positive current wage both at the time of the survey and at the time of displacement. A job displacement event in the DWS is defined as when a worker's employment ends because the company or plant at which they were employed closed, because their position or shift was abolished, or because of "insufficient work". Individuals are asked to report how long ago (in years) their displacement event occured,

[27]Note that I use the results of the Census data for this purpose rather than the CPS since the sample sizes of the CPS are too small to obtain accurate estimates of mean wages by occupation. Specifically, I use the regression in column 1 of Table 4 to obtain fitted values for each occupation.

[28]One could take a different approach, e.g. by estimating switching costs which vary by source and destination occupation, but that is beyond the scope of what I do here.

[29]Specifically, a worker who does not switch occupations, or switches to an occupation with identical mixing over types, has a Hellinger distance of zero. A worker who switches to a new occupation with no overlap with their old occupation, i.e. an occupation that mixes with positive probability over none of the extreme types which their old occupation mixed with positive probability, has a Hellinger distance of one.

[30]The reason I focus on these four years of the DWS is because they use the 2000 Census occupation codes. This allows me to use the results from the previous section directly in the subsequent analysis.

as well as both their current wage and their wage at the time of displacement. So as to avoid respondents who give unrealistic wages from affecting my results, I drop all those who report a current or former hourly wage which is less than 75 percent of the prevailing federal minimum wage. Finally, I deflate the hourly wages using the CPI-U.[31]

After applying my sample restrictions, I am left with 2378 observations; of these, 2223, or over 93 percent, have a match for both their current and former occupation in the Sanders (2016) crosswalk. I present summary statistics in Table 5, dividing workers by whether or not they make an occupational switch and the implied direction of the switch with respect to the occupational ladder constructed using the membership vector estimates. Note that occupations can have the same ranks in the ladder based on fitted values from the regression in Table 4, since multiple occupations can have the same estimated membership vectors. This will occur when a worker switches between two occupations located on the same vertex of the $K-1$ dimensional simplex. As a result, not all workers will be within the set of upward and downward switchers. I therefore separate workers into two groups based on whether or not they make a downward switch.[32] The log difference of the current and former wage is much lower for those who switch occupations and have a negative direction relative to those who do not switch or do not have a negative direction. Workers who do not make an occupational switch tend to have higher pre-displacement wages than those who do. While the literature has gone into considerable detail examining the selection into occupation switching, including following a displacement event (Forsythe 2020), my purposes in this application are descriptive; in particular, I focus on examining how movements in the Δ^{K-1} membership space following a job displacement event map onto changes in hourly wages.

Results. I now proceed to my results. In Table 6, I first focus on regressions with only a direction indicator which are similar to Table 4 of Huckfeldt (2016).[33] This allows me to compare my measure of direction, formed from the occupational ladder implied by the fitted values obtained by regressing mean wages by occupation on the estimated membership vectors, to the common measure of direction in the existing literature based on a ladder of mean wages by occupation. The results of the first two columns are very similar to those in Huckfeldt (2016).

[31] In particular, I follow the process of Forsythe (2020): since the DWS is administered in January, current wages are deflated using the CPI-U value for January of the year of the survey. Since the DWS asks respondents how many years ago they lost their job, former wages are deflated using the mean CPI-U value for the implied year of displacement.

[32] Note that a worker who makes a switch that is neither upward or downward will have an associated Hellinger distance of zero.

[33] Huckfeldt (2016) also includes a recession indicator. Since I use only a subset of the available DWS years and thus only have one recession year in my data, I do not include this indicator. Further, Huckfeldt (2016) uses the Autor and Dorn (2013) occupation codes; I can simply use the occupation codes in the DWS data directly, since they are consistent across the period I examine.

Notably, without controlling for the direction of switches, occupational switches — defined as changes in occupation at the 3-digit level — are associated with wage losses (Column 1). After controlling for direction using the traditional measure, making a downward switch is associated with steep wage losses, whereas the coefficient on the occupational switching indicator, which now represents making an upward or sideways move, becomes positive and insignificant (Column 2). The magnitude of this wage loss – about 14.5 log points — is very similar to that found in Huckfeldt (2016), even though Huckfeldt uses weekly wages whereas I use hourly wages.

In Column 3, I replace the mean wage ladder measure of direction to the direction measure formed from the wage ladder implied by the membership vectors; the magnitude of the estimated coefficient for this indicator is very similar to the traditional measure of direction. Finally, in column 4, I include both measures of distance. They both remain negative and significant, implying that the two measures are only moderately collinear.

Next, in Table 7 I estimate regressions that account for both the distance and direction of occupational switching. In Column 1, I estimate a regression which accounts for the distance of occupational switches but not the direction by including the Hellinger distance defined in (8) as well as the indicator for occupation switching. While the coefficient on the occupation switch indicator is negative, as expected, the coefficient on the Hellinger distance is very close to zero.

In Column 2 of Table 7, I account for distance as well by including the indicator for whether a downward move in the wage ladder is implied by the estimated membership vectors. I also include an interaction between the direction and distance terms. The coefficient on the Hellinger distance is now positive and significant – implying that a displaced worker who makes a large positive switch in occupation actually experiences wage gains. On the other hand, the coefficient on the Hellinger–negative switch indicator is negative and significant, as expected. This suggests that wage changes following an occupational switch depend on both the distance and direction of the switch. Note that the negative switch indicator by itself is positive but insignificant; intuitively, occupational switches that only result in marginal changes in a worker's group membership vector should not impact wages except to the extent human capital is occupation-specific.

As it turns out, there does appear to be evidence that the loss of occupational-specific human capital plays a role in wage losses resulting from occupational switches following job displacement. Recall that in Table 6, I only controlled for the direction of occupational switches. Adding the coefficients on the Hellinger distance and its interaction with the distance indicator, I now control for *both* distance and direction, allowing for the occupational switching indicator to be interpreted as a "fixed cost" in wage terms associated with switching occupations which is separate from both the distance or direction of the switch. After doing this, the coefficient on occupational switches once again becomes negative and significant. The coefficient on the indicator implies that switching occupations following job displacement, even if both the destination and origin occupation have identical membership vectors, is associated with an additional loss in hourly wages of 6 log points.[34]

To analyze occupational changes following displacement in more detail, I examine how patterns of movement within the membership simplex are associated with patterns of wage changes. Since the membership vector has seven dimensions examining changes in it directly, beyond specifying a distance metric, tend to be difficult. I therefore consider changes in the primary type of the occupation in which workers are employed. Let p be the primary type of occupation i if $\max_k \hat{\lambda}_{i,k} = \hat{\lambda}_{i,p}$, i.e. if the maximal degree of membership for occupation i is $\hat{\lambda}_{i,p}$. I can then proxy changes in occupations that result in major changes in the task bundle supplied by displaced workers as occupational changes that result in changes of the primary type. The key limitation of this approach is that, for workers that mix over two types, small changes in the membership vector can result in changes in the primary type.[35]

Table 8 lists the mean log differences between current and pre-displacement wages for each combination of current and former primary type. Of the $7 \times 7 = 49$ possible combinations, 48 are observed in the DWS data. In Figure 5, I plot the data using an alluvial diagram. In particular, the size of the flows between primary types reflects the (weighted) share of the total sample that is observed making such a transition. In addition, the flows are shaded so as to reflect the conditional mean log difference in wages reported in Table 8. It is clear that there is a large degree of movement across primary types; however, the extremes of wage changes appear to be driven by movement to and from occupations which are primarily classified as low-skill service.

To investigate this in more detail, in Column 3 of Table 7 I run another regression on the log difference of wages, but this time using an alternate direction. I define this alternate direction as being negative if and only if an individual works in an occupation which has a primary type other than low-skill service prior to displacement and following displacement works in an

[34] Forsythe (2020) uses an alternate measure of occupational distance, namely the difference in log mean wages between the current and former occupation. Although I do not report the results here, I get similar results with respect to the occupation switch indicator if I use an analogous measure which only uses the variation in membership vectors, namely the difference between the fitted values from the regression in (7).

[35] For instance, consider a worker who is initially employed as a transit or intercity bus driver, an occupation which has a membership of 0.514 in the low-skill manual type and 0.484 in the low-skill service type. If, following displacement, this worker becomes a taxi driver or chauffeur — with membership of 0.530 in the low-skill service type and 0.468 in the low-skill manual type — they would be recorded as changing primary types, even though the Hellinger distance between the two membership vectors is quite small.

occupation which has low-skill service as its primary type. Hence, many of the occupational switches classified as "negative" in the previous definition of direction will now be considered as neutral, changing the interpretation of the regressors. Looking at the results, the coefficient on the Hellinger distance without the special direction measure becomes less positive, though it is still significant, reflecting the fact that it now includes some negative switches in addition to the set of positive switches in the DWS data.

However, the other coefficients are remarkably similar. In particular, consider a worker who makes a maximal move down the occupational ladder implied by the estimated membership vectors. By taking the sum of the coefficients in Column 2 of Table 7, this worker is expected to incur wage losses of 19.3 log points. Now, consider a worker who makes a maximal move such that their primary occupational type switches to low-skill service. Summing the coefficients in Column 3 of Table 7, one finds that such a worker is expected to incur wage losses of 21.5 log points. In other words, although wage losses associated with occupational changes following a job displacement event can be thought of as the result of movements down the occupational wage ladder, they can also be thought of — at least during the period of the 2000s I study — as predominantly movements towards the low-skill service occupations that tend to occupy the lowest rungs of the occupational wage ladder. While a large literature considers the increased employment in low-skill service jobs as part of trends of job polarization driven by technological change (Autor and Dorn 2013), it has largely been focused on a long-term view of changes in the labour market. More recently, attention has been given to extent to which polarization may be driven by business cycle fluctuations (Jaimovich and Siu 2020; Heathcote, Perri, and Violante 2020). In addition, Cortes et al. (2020) argue that the bulk of the decline in the employment share in routine occupations are a result of declining inflows from unemployment and nonparticipation. This can be related to the model of Huckfeldt (2016), in which firms become more selective in hiring in recessions, increasing the probability that workers who are displaced are forced to move down the occupational ladder.

6 Conclusion

In this paper, I propose a new method to identify occupations' task bundles from O*NET survey data. I estimate a hierarchical Bayesian model, the Grade of Membership model, which allows occupations to have partial membership in multiple latent types. Unlike many existing approaches, I am not required to make a priori assumptions about which tasks, or groupings of tasks, are most important to understand differences between occupations. Although mixed membership models tend to be quite complex to estimate, I am able to estimate the model using variational inference, an approximate estimation technique from the machine learning literature.

Although latent Dirichlet allocation, a closely related algorithm, has seen increasing attention among economists in recent years, the Grade of Membership model has yet to see widespread use. The application in this paper is among the first in economics and related fields. The classifications estimated by the Grade of Membership model are intuitive, as are the archetypes of occupations represented by the estimated extreme types. This suggests that the Grade of Membership model is worth consideration by researchers who seek to identify latent structure inherent not only in the task data of O*NET, but in discrete survey data more broadly.

I demonstrate the utility of the estimates from the model by using estimated membership vectors to define intuitive and coherent measures of the distance and direction of occupational switches. I then test the utility of these measures using data from the Displaced Workers Supplement of the CPS. I show that, as expected, the magnitude of wage changes associated with occupational switches following job displacement depends on the distance of the move, whereas the sign depends on the implied direction of the wage change. In addition, hourly wage losses associated with occupational switches of about six percent remain *after* controlling for both the distance and direction of an occupational change. This strongly suggests an important role of occupation-specific human capital, as argued by Kambourov and Manovskii (2009). Finally, a large proportion of workers who suffer large wage losses following job displacement appear to be workers switching into low-skill service occupations. Since it has been shown that occupational downgrading following job displacement occurs predominantly during recessions (Huckfeldt 2016), this result implies that the link between business cycle dynamics and job polarization is a fertile ground for future research.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings". In *Handbook of Labor Economics*, 4:1043–1171. Elsevier.

Airoldi, Edoardo M., David Blei, Elena A. Erosheva, and Stephen E. Fienberg. 2014. *Handbook of Mixed Membership Models and Their Applications*. CRC press.

Autor, David H. 2013. "The "Task Approach" to Labor Markets: An Overview". *Journal for Labour Market Research* 46, no. 3 (): 185–199.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market". *American Economic Review* 103 (5): 1553–97.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration". *The Quarterly journal of economics* 118 (4): 1279–1333.

Bandiera, Oriana, Andrea Prat, Stephen Hansen, and Raffaella Sadun. 2020. "CEO Behavior and Firm Performance". *Journal of Political Economy* 128 (4): 1325–1369.

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe. 2017. "Variational Inference: A Review for Statisticians". *Journal of the American Statistical Association* 112, no. 518 (): 859–877. arXiv: `1601.00670`.

Blei, David M., and John D. Lafferty. 2007. "A Correlated Topic Model of Science". *The Annals of Applied Statistics* 1, no. 1 (): 17–35. arXiv: `0708.3601`.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation". *Journal of machine Learning research* 3 (Jan): 993–1022.

Chen, Yen-Chi, Y. Samuel Wang, and Elena A. Erosheva. 2018. "On the Use of Bootstrap with Variational Inference: Theory, Interpretation, and a Two-Sample Test Example" (). arXiv: `1711.11057 [stat]`.

Cortes, Guido Matias. 2016. "Where Have the Middle-Wage Workers Gone? A Study of Polarization Using Panel Data". *Journal of Labor Economics* 34, no. 1 (): 63–105.

Cortes, Guido Matias, Nir Jaimovich, Christopher J. Nekarda, and Henry E. Siu. 2020. "The Dynamics of Disappearing Routine Jobs: A Flows Approach". *Labour Economics* 65 (): 101823.

Davis, Steven J., and Till von Wachter. 2011. "Recessions and the Costs of Job Loss". *Brookings Papers on Economic Activity*: 1.

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market". *The Quarterly Journal of Economics* 132, no. 4 (): 1593–1640.

Dotson, Marc R., Joachim Büschken, and Greg M. Allenby. 2020. "Explaining Preference Heterogeneity with Mixed Membership Modeling". *Marketing Science* 39, no. 2 (): 407–426.

Draca, Mirko, and Carlo Schwarz. 2020. "How Polarised Are Citizens? Measuring Ideology from the Ground-Up": 88.

Erosheva, Elena A. 2002. "Grade of Membership and Latent Structure Models with Application to Disability Survey Data". PhD Thesis, PhD thesis, Carnegie Mellon University, Department of Statistics.

Erosheva, Elena A., Stephen E. Fienberg, and Cyrille Joutard. 2007. "Describing Disability through Individual-Level Mixture Models for Multivariate Binary Data". *The Annals of Applied Statistics* 1, no. 2 (): 502–537.

Erosheva, Elena, Stephen Fienberg, and John Lafferty. 2004. "Mixed-Membership Models of Scientific Publications". *Proceedings of the National Academy of Sciences* 101 (suppl 1): 5220–5227. pmid: 15020766.

Farber, Henry S. 2017. "Employment, Hours, and Earnings Consequences of Job Loss: US Evidence from the Displaced Workers Survey". *Journal of Labor Economics* 35, no. S1 (): S235–S272.

Forsythe, Eliza. 2020. "Occupational Job Ladders and Displaced Workers": 51.

Gathmann, Christina, and Uta Schönberg. 2010. "How General Is Human Capital? A Task-Based Approach". *Journal of Labor Economics* 28 (1): 1–49. JSTOR: 10.1086/649786.

Giordano, Ryan, Runjing Liu, Nelle Varoquaux, Michael I. Jordan, and Tamara Broderick. 2017. "Measuring Cluster Stability for Bayesian Nonparametrics Using the Linear Bootstrap" (). arXiv: 1712.01435 [stat].

Gormley, Isobel Claire, and Thomas Brendan Murphy. 2009. "A Grade of Membership Model for Rank Data". *Bayesian Analysis* 4 (2): 265–295.

Grigsby, John. 2019. "Skill Heterogeneity and Aggregate Labor Market Dynamics": 92.

Groes, Fane, Philipp Kircher, and Iourii Manovskii. 2015. "The U-Shapes of Occupational Mobility". *The Review of Economic Studies* 82, no. 2 (): 659–692.

Gross, Justin H., and Daniel Manrique-Vallier. 2014. "A Mixed-Membership Approach to the Assessment of Political Ideology from Survey Responses". *Handbook of mixed membership models and their applications*: 119–140.

Guvenen, Fatih, Burhan Kuruscu, Satoshi Tanaka, and David Wiczer. 2020. "Multidimensional Skill Mismatch". *American Economic Journal: Macroeconomics* 12, no. 1 (): 210–244.

Hansen, Stephen, and Michael McMahon. 2016. "Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication". *Journal of International Economics* 99:S114–S133.

Hansen, Stephen, Michael McMahon, and Andrea Prat. 2018. "Transparency and Deliberation within the FOMC: A Computational Linguistics Approach". *The Quarterly Journal of Economics* 133 (2): 801–870.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2020. *The Rise of US Earnings Inequality: Does the Cycle Drive the Trend?* Federal Reserve Bank of Minneapolis.

Holland, Paul W., and Paul R. Rosenbaum. 1986. "Conditional Association and Unidimensionality in Monotone Latent Variable Models". *The Annals of Statistics*: 1523–1543.

Huckfeldt, Christopher. 2016. "Understanding the Scarring Effect of Recessions". *Working paper*.

Ingram, Beth F., and George R. Neumann. 2006. "The Returns to Skill". *Labour economics* 13 (1): 35–59.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers". *The American economic review*: 685–709.

Jaimovich, Nir, and Henry E. Siu. 2020. "Job Polarization and Jobless Recoveries". *Review of Economics and Statistics* ().

Jordan, Michael I., Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. 1998. "An Introduction to Variational Methods for Graphical Models". In *Learning in Graphical Models*, ed. by Michael I. Jordan, 105–161. Dordrecht: Springer Netherlands.

Kambourov, Gueorgui, and Iourii Manovskii. 2009. "Occupational Specificity of Human Capital*". *International Economic Review* 50 (1): 63–115.

Larsen, Vegard H., and Leif A. Thorsrud. 2019. "The Value of News for Economic Developments". *Journal of Econometrics*, Annals Issue in Honor of John Geweke "Complexity and Big Data in Economics and Finance: Recent Developments from a Bayesian Perspective", 210, no. 1 (): 203–218.

Lise, Jeremy, and Fabien Postel-Vinay. Forthcoming. "Multidimensional Skills, Sorting, and Human Capital Accumulation". *American Economic Review*.

Manton, Kenneth G., Max A. Woodbury, and H. Dennis Tolley. 1994. *Statistical Applications Using Fuzzy Sets*. Vol. 225. Wiley-Interscience.

Munro, Evan, and Serena Ng. 2020. "Latent Dirichlet Analysis of Categorical Survey Expectations" (). arXiv: 1910.04883.

Neal, Derek. 1995. "Industry-Specific Human Capital: Evidence from Displaced Workers". *Journal of labor Economics* 13 (4): 653–677.

Peri, Giovanni, and Chad Sparber. 2009. "Task Specialization, Immigration, and Wages". *American Economic Journal: Applied Economics* 1 (3): 135–69.

Poletaev, Maxim, and Chris Robinson. 2008. "Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000". *Journal of Labor Economics* 26, no. 3 (): 387–420.

Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data". *Genetics* 155, no. 2 (): 945–959. pmid: 10835412.

Raposo, Pedro, Pedro Portugal, and Anabela Carneiro. 2019. "The Sources of the Wage Losses of Displaced Workers: The Role of the Reallocation of Workers into Firms, Matches, and Job Titles". *Journal of Human Resources* (): 0317–8667R3.

Robinson, Chris. 2018. "Occupational Mobility, Occupation Distance, and Specific Human Capital". *Journal of Human Resources* 53, no. 2 (): 513–551.

Ruiz, Francisco J. R., Susan Athey, and David M. Blei. 2019. "SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements" (). arXiv: 1711.03560 [cs, econ, stat].

Sanders, Carl. 2016. "Skill Uncertainty, Skill Accumulation, and Occupational Choice".

Sanders, Carl, and Christopher Taber. 2012. "Life-Cycle Wage Growth and Heterogeneous Human Capital". *Annual Review of Economics* 4, no. 1 (): 399–425.

Stinebrickner, Ralph, Todd Stinebrickner, and Paul Sullivan. 2019. "Job Tasks, Time Allocation, and Wages". *Journal of Labor Economics* 37, no. 2 (): 399–433.

Varki, Sajeev, Bruce Cooil, and Roland T. Rust. 2000. "Modeling Fuzzy Data in Qualitative Marketing Research". *Journal of Marketing Research* 37 (4): 480–489.

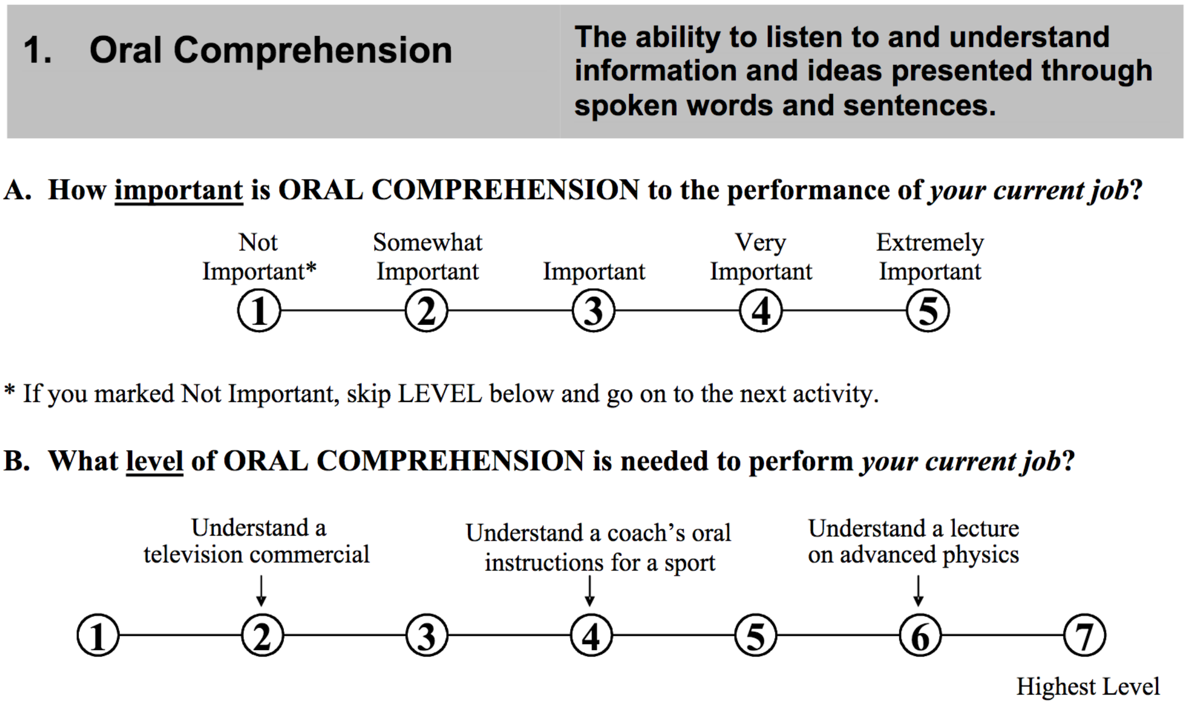
Wang, Y Samuel, and Elena A Erosheva. 2015. "Fitting Mixed Membership Models Using mixedMem": 21.

Wang, Y. Samuel, Ross L. Matsueda, and Elena A. Erosheva. 2017. "A Variational EM Method for Mixed Membership Models with Multivariate Rank Data: An Analysis of Public Policy Preferences". *Annals of Applied Statistics* 11, no. 3 (): 1452–1480. arXiv: 1512.08731.

Woodbury, Max A., Jonathan Clive, and Arthur Garson Jr. 1978. "Mathematical Typology: A Grade of Membership Technique for Obtaining Disease Definition". *Computers and biomedical research* 11 (3): 277–298.

Yamaguchi, Shintaro. 2012. "Tasks and Heterogeneous Human Capital". *Journal of Labor Economics* 30, no. 1 (): 1–53.

Figure 1: Sample O*NET survey question



1. Oral Comprehension	The ability to listen to and understand information and ideas presented through spoken words and sentences.

A. How <u>important</u> is ORAL COMPREHENSION to the performance of *your current job*?

Not Important* — 1
Somewhat Important — 2
Important — 3
Very Important — 4
Extremely Important — 5

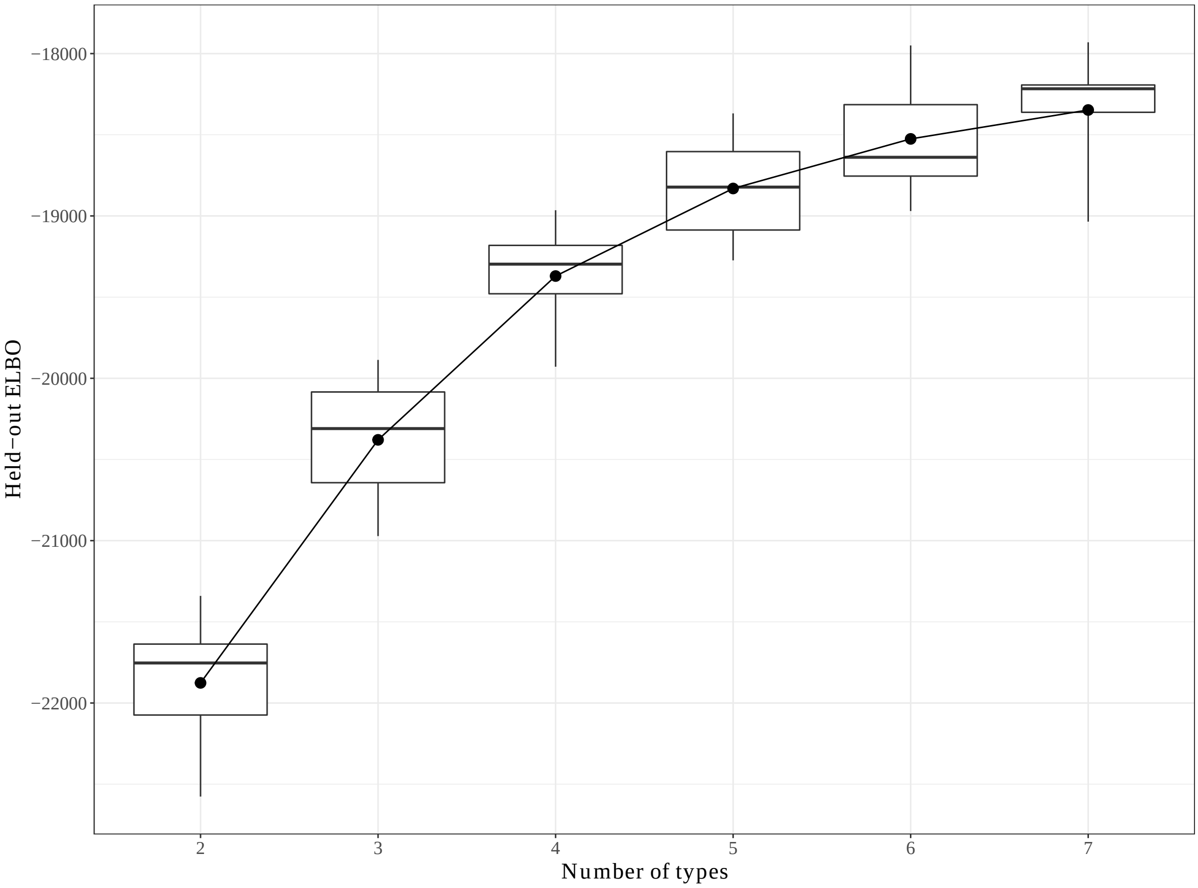
* If you marked Not Important, skip LEVEL below and go on to the next activity.

B. What <u>level</u> of ORAL COMPREHENSION is needed to perform *your current job*?

1
2 — Understand a television commercial
3
4 — Understand a coach's oral instructions for a sport
5
6 — Understand a lecture on advanced physics
7 — Highest Level

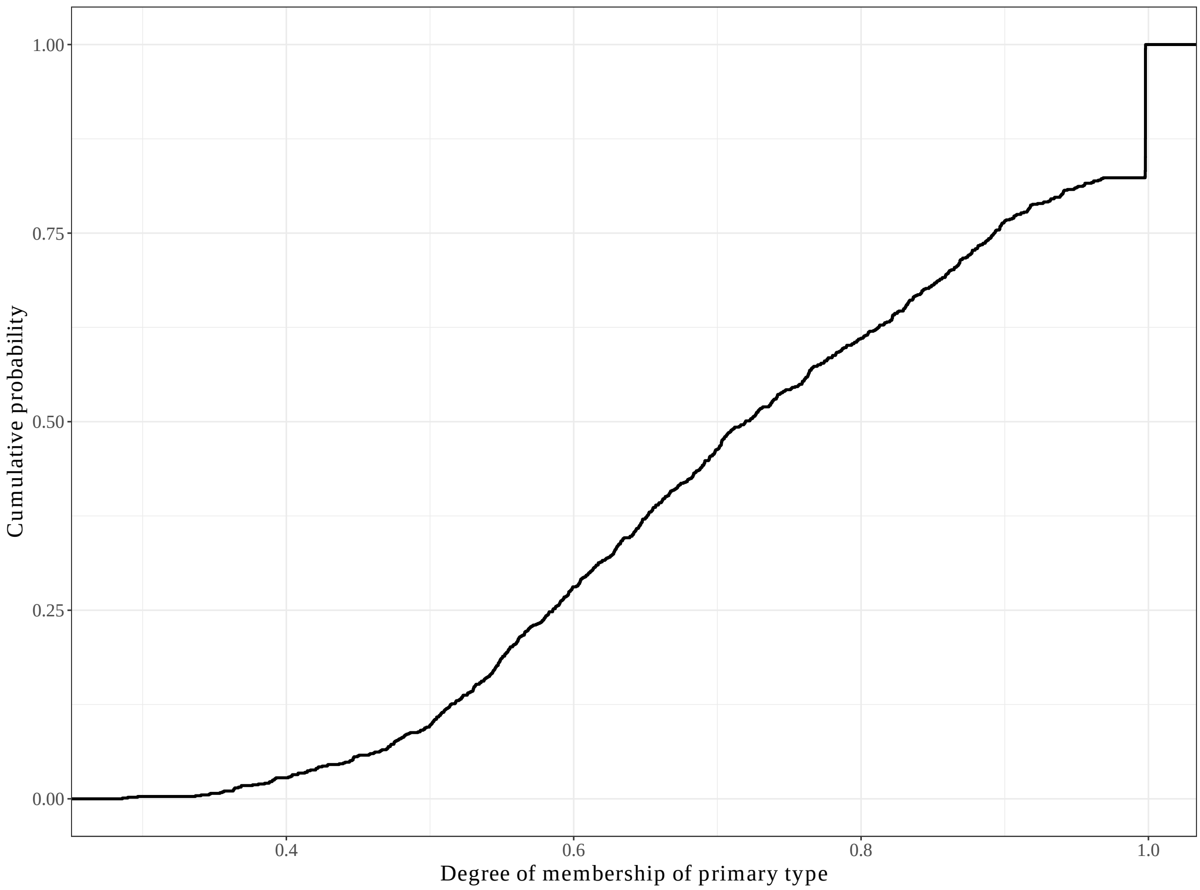
Source: Retrieved from the O*NET website (`https://www.onetcenter.org/questionnaires.html`).

Figure 2: Results of cross-validation exercise



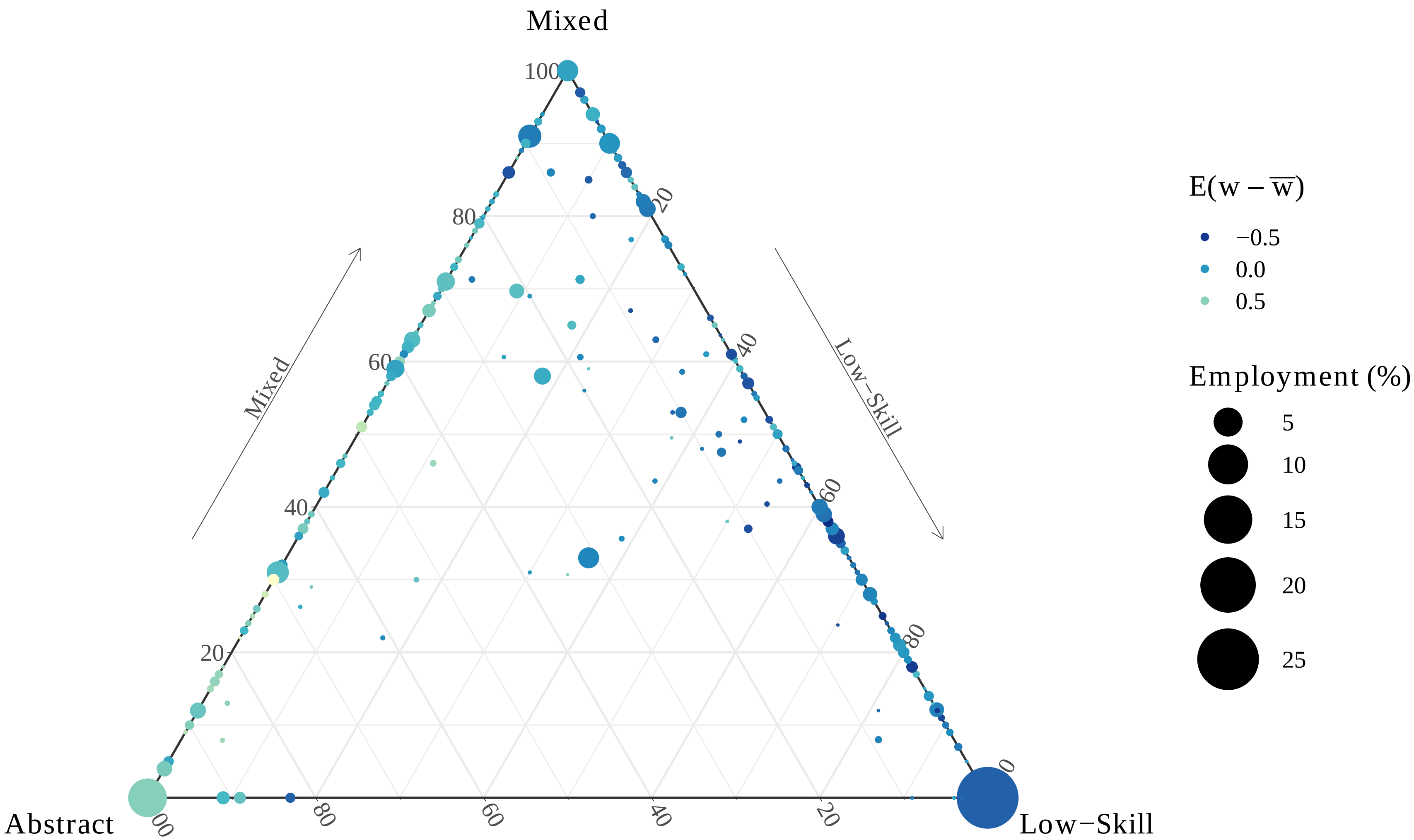
Notes: Held-out ELBO for each of the five folds as described in text. For each value of K, the points represent the mean held-out ELBO across all five folds.

Figure 3: Empirical distribution of maximal degrees of membership



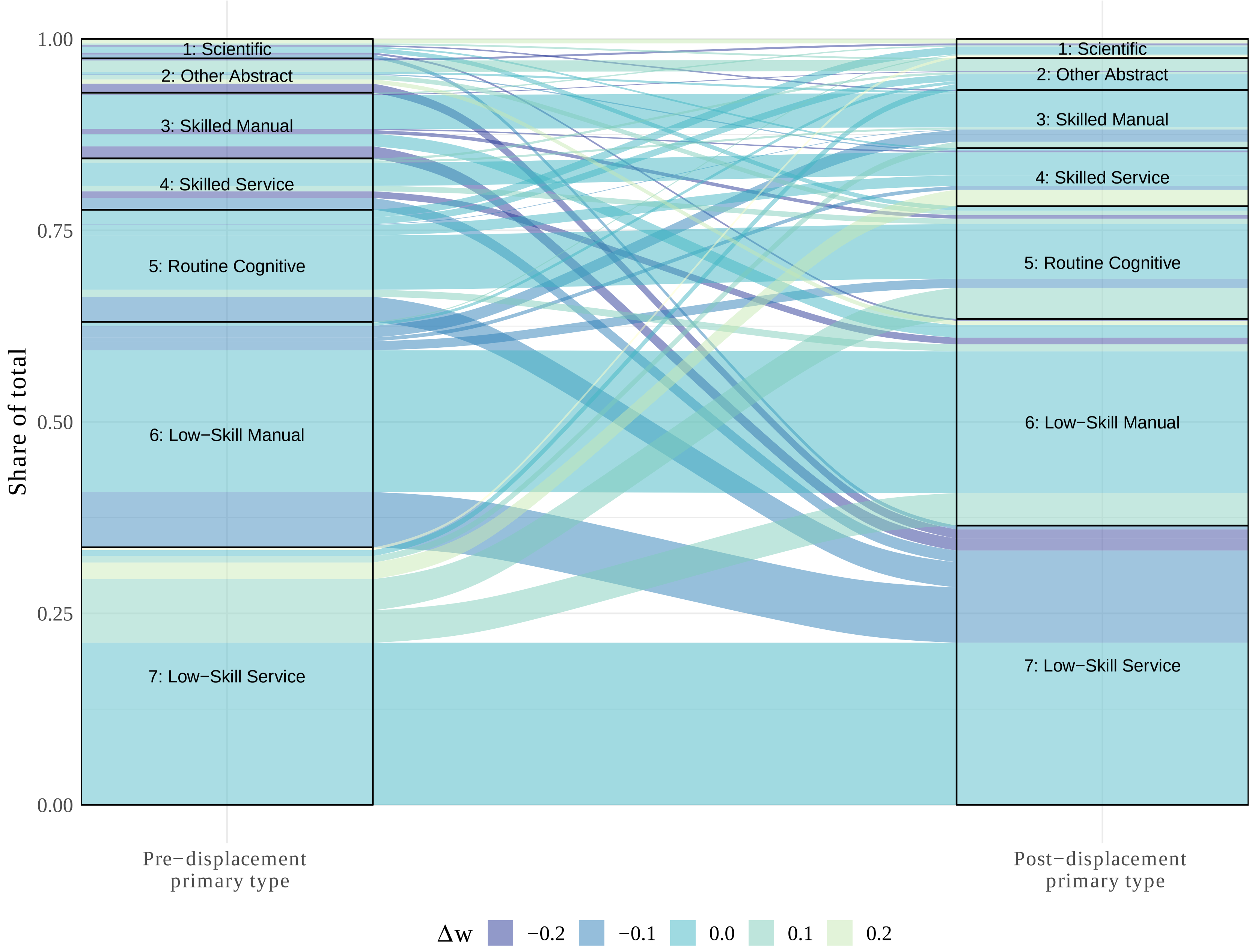
Notes: This figure plots the empirical distribution function of the estimated maximal degrees of membership, $\max_k \hat{\lambda}_{i,k}$, across occupations.

Figure 4: Ternary plot of wages and employment shares across three broad groups



Notes: Occupations located on lines parallel to the right edge of the simplex have the same degree of membership in abstract (Scientific and other Abstract) occupation groups. Occupations located on lines parallel to the bottom edge of the simplex have the same degree of membership in mixed (Skilled Manual, Skilled Service, and Routine Cognitive) occupation groups. Occupations located on lines parallel to the left edge of the simplex have the same degree of membership in low-skill (Low-skill Manual and Low-skill Service) occupation groups. The size of each point is determined by the share of the total workforce in the 2000 US Census who work in occupations at that location of the simplex (rounded to two decimal places). The colour of each point is determined by the log difference between the mean log wage of workers at that location of the simplex and the mean wage across all workers in 2000. All calculations use the Census final weight.

Figure 5: Switching between primary types following job displacement



Notes: The size of the flows between primary types reflects the weighted share of the total sample that is observed making such a transition. The flows are shaded so as to reflect the mean log difference in wages among workers who make a given transition. All results are calculated using the DWS weights.

Table 1: Occupations with highest estimated degree of membership by group

Group	Occupations with highest degree of membership
1: Scientific	Materials Engineers (0.998); Biomedical Engineers (0.998); Biochemists and Biophysicists (0.998); Physical Medicine and Rehabilitation Physicians (0.998); Soil and Plant Scientists (0.998); Materials Scientists (0.998); Biochemical Engineers (0.998); Food Scientists and Technologists (0.998); Biofuels/Biodiesel Technology and Product Development Managers (0.998); Hydrologists (0.998)
2: Other Abstract	Treasurers and Controllers (0.998); Industrial-Organizational Psychologists (0.998); Education Administrators, Postsecondary (0.998); Financial Examiners (0.998); Management Analysts (0.998); Auditors (0.998); Forestry and Conservation Science Teachers, Postsecondary (0.998); Natural Sciences Managers (0.998); Labor Relations Specialists (0.998); Epidemiologists (0.998)
3: Skilled Manual	Non-Destructive Testing Specialists (0.998); Forest and Conservation Technicians (0.998); Hydroelectric Plant Technicians (0.998); Water and Wastewater Treatment Plant and System Operators (0.998); Medical Appliance Technicians (0.998); Forest Firefighters (0.998); Petroleum Pump System Operators, Refinery Operators, and Gaugers (0.965); Mates- Ship, Boat, and Barge (0.96); First-Line Supervisors of Construction Trades and Extraction Workers (0.941); Electrical and Electronics Repairers, Commercial and Industrial Equipment (0.938)
4: Skilled Service	Midwives (0.998); Licensed Practical and Licensed Vocational Nurses (0.998); Medical Assistants (0.967); Psychiatric Technicians (0.941); Funeral Service Managers (0.911); Recreational Therapists (0.898); First-Line Supervisors of Correctional Officers (0.866); Police, Fire, and Ambulance Dispatchers (0.862); Occupational Therapy Assistants (0.858); Chefs and Head Cooks (0.845)
5: Routine Cognitive	Insurance Sales Agents (0.998); Insurance Underwriters (0.998); Eligibility Interviewers, Government Programs (0.998); Tax Preparers (0.998); Procurement Clerks (0.998); Travel Agents (0.998); Credit Authorizers (0.998); Brokerage Clerks (0.956); Credit Checkers (0.948); Cargo and Freight Agents (0.917)
6: Low-skill Manual	Helpers–Production Workers (0.998); Rock Splitters, Quarry (0.998); Sawing Machine Setters, Operators, and Tenders, Wood (0.998); Insulation Workers, Floor, Ceiling, and Wall (0.998); Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic (0.998); Molding, Coremaking, and Casting Machine Setters, Operators, and Tenders, Metal and Plastic (0.998); Textile Winding, Twisting, and Drawing Out Machine Setters, Operators, and Tenders (0.998); Loading Machine Operators, Underground Mining (0.998); Reinforcing Iron and Rebar Workers (0.998); Roustabouts, Oil and Gas (0.998)
7: Low-skill Service	Food Servers, Nonrestaurant (0.998); Cooks, Fast Food (0.998); Locker Room, Coatroom, and Dressing Room Attendants (0.998); Shampooers (0.998); Ushers, Lobby Attendants, and Ticket Takers (0.998); Models (0.998); Hosts and Hostesses, Restaurant, Lounge, and Coffee Shop (0.998); Orderlies (0.998); Counter Attendants, Cafeteria, Food Concession, and Coffee Shop (0.998); Combined Food Preparation and Serving Workers, Including Fast Food (0.998)

Notes: This table lists the occupation titles of the O*NET-SOC Codes associated with the highest estimated degrees of membership in each of the 7 estimated groups. The estimated degree of membership in the group for each occupation is listed in parentheses.

Table 2: Descriptors most likely to be in third tercile by occupation extreme type

Group	Descriptors of skill or task most likely to be in third tercile
1: Scientific	Category Flexibility (1.000); Complex Problem Solving (1.000); Deductive Reasoning (1.000); Inductive Reasoning (1.000); Information Ordering (1.000); Mathematical Reasoning (1.000); Systems Analysis (1.000); Systems Evaluation (1.000); Number Facility (1.000); Mathematics (1.000)
2: Other Abstract	Writing (1.000); Written Expression (0.969); Judgment and Decision Making (0.963); Active Learning (0.955); Speaking (0.938); Speech Clarity (0.928); Developing Objectives and Strategies (0.921); Systems Analysis (0.916); Complex Problem Solving (0.914); Organizing, Planning, and Prioritizing Work (0.910)
3: Skilled Manual	Operation and Control (1.000); Repairing (0.999); Equipment Selection (0.995); Repairing and Maintaining Electronic Equipment (0.992); Inspecting Equipment, Structures, or Material (0.987); Equipment Maintenance (0.984); Visual Color Discrimination (0.984); Repairing and Maintaining Mechanical Equipment (0.977); Operation Monitoring (0.963); Glare Sensitivity (0.953)
4: Skilled Service	Assisting and Caring for Others (1.000); Social Perceptiveness (1.000); Service Orientation (1.000); Performing for or Working Directly with the Public (0.919); Monitor Processes, Materials, or Surroundings (0.735); Problem Sensitivity (0.730); Speed of Closure (0.715); Coordination (0.710); Identifying Objects, Actions, and Events (0.697); Negotiation (0.695)
5: Routine Cognitive	Near Vision (0.504); Speech Recognition (0.401); Interacting With Computers (0.395); Performing Administrative Activities (0.395); Number Facility (0.341); Service Orientation (0.339); Negotiation (0.318); Programming (0.314); Mathematical Reasoning (0.296); Active Listening (0.290)
6: Low-Skill Manual	Control Precision (1.000); Multilimb Coordination (1.000); Reaction Time (1.000); Rate Control (1.000); Static Strength (0.991); Trunk Strength (0.982); Extent Flexibility (0.976); Dynamic Strength (0.969); Speed of Limb Movement (0.966); Response Orientation (0.963)
7: Low-Skill Service	Performing for or Working Directly with the Public (0.506); Dynamic Flexibility (0.478); Assisting and Caring for Others (0.433); Trunk Strength (0.414); Explosive Strength (0.407); Stamina (0.349); Dynamic Strength (0.312); Extent Flexibility (0.273); Gross Body Coordination (0.267); Selling or Influencing Others (0.245)

Notes: This table lists, for each extreme profile, the descriptors of skills or tasks for which it is most likely a high level of ability is required (within the third tercile of the distribution across occupations for that skill or task). The estimated probability $\hat{\theta}_{j,k,3}$ is in parentheses.

Table 3: Comparison of classifications for sample occupations to alternate measures

Occupation	GoM classification	DD skill measure	LPV skill measure
Economists	Other Abstract: 0.998	Abstract: 5.562	Cognitive: 0.635
		Routine: 1.932	Interpersonal: 0.679
		Manual: 0.183	Manual: 0.285
Accountants	Routine Cognitive: 0.630	Abstract: 7.620	Cognitive: 0.551
	Other Abstract: 0.369	Routine: 5.993	Interpersonal: 0.535
		Manual: 0.016	Manual: 0.324
Telemarketers	Routine Cognitive: 0.518	Abstract: 2.151	Cognitive: 0.147
	Low-skill Service: 0.480	Routine: 1.586	Interpersonal: 0.330
		Manual: 0.195	Manual: 0.000
Dentists, General	Scientific: 0.738	Abstract: 4.619	Cognitive: 0.621
	Skilled Manual: 0.260	Routine: 4.086	Interpersonal: 0.704
		Manual: 2.329	Manual: 0.614
Ship Engineers	Skilled Manual: 0.707	Abstract: 3.291	Cognitive: 0.519
	Low-skill Manual: 0.242	Routine: 4.225	Interpersonal: 0.423
	Skilled Service: 0.050	Manual: 4.724	Manual: 0.917
Roofers	Low-skill Manual: 0.857	Abstract: 1.184	Cognitive: 0.235
	Skilled Manual: 0.142	Routine: 7.394	Interpersonal: 0.290
		Manual: 4.955	Manual: 0.653

Notes: The first column lists for each occupation the estimated grades of membership for that occupation, where grades of membership below 0.01 have been suppressed. The second column lists the skill measures for each occupation from Autor and Dorn (2013), and the third column lists the skill measures for each occupation from Lise and Postel-Vinay (forthcoming).

Table 4: Descriptive regressions of log wages on estimated memberships (2000 U.S. Census)

Dependent variable:	Occupational mean wage		Individual wage
	(1)	(2)	(3)
Scientific	1.044^{**}	0.498^{**}	0.728^{**}
	(0.056)	(0.071)	(0.082)
Other Abstract	0.900^{**}	0.420^{**}	0.567^{**}
	(0.045)	(0.060)	(0.060)
Skilled Manual	0.545^{**}	0.385^{**}	0.435^{**}
	(0.048)	(0.044)	(0.060)
Skilled Service	0.490^{**}	0.166^{**}	0.286^{**}
	(0.070)	(0.077)	(0.073)
Routine Cognitive	0.490^{**}	0.163^{**}	0.253^{**}
	(0.043)	(0.049)	(0.065)
Low-Skill Manual	0.354^{**}	0.487^{**}	0.304^{**}
	(0.039)	(0.034)	(0.055)
Education controls	—	Yes	Yes
Age controls	—	—	Yes
N	485	485	935,131
R^2	0.597	0.712	0.280

Notes: Columns 1 and 2: robust standard errors in parentheses. Column 3: standard errors clustered at the occupation level in parentheses. Education controls are categorical and consist of four groups (high school droputs, high school graduates, some college, and bachelor's degree or greater). Age controls consist of a quadratic in age. A single asterisk * represents significance at the 5 percent level and two asterisks ** represent significance at the 1 percent level.

Table 5: Summary statistics for displaced workers by occupation switching and direction

Occupation switch:	No	Yes	
Direction:	—	Upward / Neutral	Downward
	(1)	(2)	(3)
Former log wage	1.904	1.714	1.822
	(0.446)	(0.359)	(0.374)
Current log wage	1.902	1.708	1.702
	(0.440)	(0.356)	(0.362)
Log difference in wages	-0.002	-0.006	-0.119
	(0.233)	(0.293)	(0.322)
Hellinger distance	0.000	0.573	0.617
	(0.000)	(0.339)	(0.287)
Age	41.881	39.350	39.979
	(12.273)	(13.089)	(12.743)
Education:			
Dropout	0.122	0.096	0.075
	(0.328)	(0.294)	(0.263)
High school graduate	0.447	0.376	0.416
	(0.498)	(0.485)	(0.493)
Some college	0.317	0.408	0.387
	(0.466)	(0.492)	(0.487)
College+	0.114	0.120	0.123
	(0.318)	(0.326)	(0.328)
N	663	767	793

Notes: Standard deviations in parentheses. Definitions of direction of switch are as defined in the text. Results are calculated using the DWS weights. Hourly wages are deflated using the CPI-U as described in the text. Workers who report their current or former hourly wage to be less than 75% of the prevailing federal minimum wage are dropped from the sample.
Source: 2004-2010 Displaced Worker Supplement of the CPS.

Table 6: Wage losses following displacement: distance only

	(1)	(2)	(3)	(4)
Constant	-0.014	-0.015	-0.014	-0.014
	(0.014)	(0.014)	(0.014)	(0.014)
Switch occupation	-0.053^{**}	0.022	0.016	0.043^{*}
	(0.013)	(0.016)	(0.016)	(0.017)
Switch $\downarrow$ (Mean wage ladder)	—	-0.145^{**}	—	-0.102^{**}
		(0.018)		(0.019)
Switch $\downarrow$ (Implied by MV)	—	—	-0.137^{**}	-0.087^{**}
			(0.018)	(0.020)
Year fixed effects	Yes	Yes	Yes	Yes
N	2223	2223	2223	2223

Notes: Robust standard errors in parentheses. Downward switch indicators constructed as described in the text. A single asterisk * represents significance at the 5 percent level and two asterisks ** represent significance at the 1 percent level. Results are calculated using the DWS weights. Hourly wages are deflated using the CPI-U as described in the text. Workers who report their current or former hourly wage to be less than 75% of the prevailing federal minimum wage are dropped from the sample.

Table 7: Wage losses following displacement: distance and direction

	(1)	(2)	(3)
Constant	-0.014	-0.013	-0.013
	(0.014)	(0.014)	(0.014)
Switch occupation	-0.051^{**}	-0.060^{*}	-0.066^{**}
	(0.020)	(0.024)	(0.020)
Hellinger distance	-0.003	0.136^{**}	0.078^{*}
	(0.030)	(0.041)	(0.034)
Switch $\downarrow$ (Implied by MV)	—	0.029	—
		(0.035)	
Hellinger $\times$ Switch $\downarrow$	—	-0.285^{**}	—
		(0.059)	
Switch to LS	—	—	0.003
			(0.078)
Hellinger $\times$ Switch to LS	—	—	-0.217^{*}
			(0.104)
Year fixed effects	Yes	Yes	Yes
N	2223	2223	2223

Notes: Robust standard errors in parentheses. Downward switch indicators constructed as described in the text. A single asterisk * represents significance at the 5 percent level and two asterisks ** represent significance at the 1 percent level. Results are calculated using the DWS weights. Hourly wages are deflated using the CPI-U as described in the text. Workers who report their current or former hourly wage to be less than 75% of the prevailing federal minimum wage are dropped from the sample.

Table 8: Mean of log differences in real hourly wage by former and current primary types

Former \ Current	1: Scientific	2: Other Abstract	3: Skilled Manual	4: Skilled Service	5: Routine Cognitive	6: Low-Skill Manual	7: Low-Skill Service
1: Scientific	0.120	0.058	−0.284	−0.054	−0.022	−0.405	−0.116
	(0.334)	(0.120)	(0.183)	(0.168)	(0.266)	(0.301)	(0.232)
2: Other Abstract	−0.255	0.100	−0.019	−0.167	0.042	0.189	−0.207
	(0.264)	(0.251)	(0.171)	(0.065)	(0.290)	(0.250)	(0.382)
3: Skilled Manual	0.042	−0.650	−0.000	−0.322	−0.280	−0.083	−0.266
	(0.171)	(0.406)	(0.221)	(0.203)	(0.258)	(0.295)	(0.297)
4: Skilled Service	—	0.089	0.069	−0.033	0.015	−0.225	−0.153
		(0.256)	(0.203)	(0.242)	(0.354)	(0.310)	(0.260)
5: Routine Cognitive	−0.006	−0.001	−0.101	−0.062	−0.056	0.055	−0.193
	(0.263)	(0.285)	(0.186)	(0.280)	(0.264)	(0.330)	(0.399)
6: Low-Skill Manual	0.099	−0.037	−0.140	−0.175	−0.190	−0.043	−0.160
	(0.203)	(0.213)	(0.312)	(0.199)	(0.406)	(0.266)	(0.356)
7: Low-Skill Service	0.300	−0.015	0.058	0.146	0.036	0.003	−0.008
	(0.243)	(0.381)	(0.278)	(0.294)	(0.280)	(0.322)	(0.243)

Notes: Standard deviations by group in parentheses. Results are calculated using the DWS weights. Hourly wages are deflated using the CPI-U as described in the text. Workers who report their current or former hourly wage to be less than 75% of the prevailing federal minimum wage are dropped from the sample. Note that the reason no mean wage change associated with a transition from a Skilled Services occupation to a Scientific occupation is reported is because no workers in the data were observed making this transition.
Source: 2004-2010 Displaced Worker Supplement of the CPS.

Table A1: Descriptors most likely to be in first tercile by occupation extreme type

Group	Descriptors of skill or task most likely to be in first tercile
1: Scientific	Dynamic Flexibility (0.861); Peripheral Vision (0.728); Explosive Strength (0.684); Night Vision (0.675); Glare Sensitivity (0.633); Installation (0.618); Sound Localization (0.535); Assisting and Caring for Others (0.489); Performing for or Working Directly with the Public (0.488); Speed of Limb Movement (0.479)
2: Other Abstract	Arm-Hand Steadiness (1.000); Control Precision (1.000); Extent Flexibility (1.000); Manual Dexterity (1.000); Multilimb Coordination (1.000); Rate Control (1.000); Reaction Time (1.000); Finger Dexterity (1.000); Static Strength (1.000); Peripheral Vision (1.000)
3: Skilled Manual	Selling or Influencing Others (0.4); Communicating with Persons Outside Organization (0.343); Performing for or Working Directly with the Public (0.328); Speech Clarity (0.298); Establishing and Maintaining Interpersonal Relationships (0.236); Resolving Conflicts and Negotiating with Others (0.223); Service Orientation (0.216); Social Perceptiveness (0.209); Speech Recognition (0.198); Identifying Objects, Actions, and Events (0.19)
4: Skilled Service	Operations Analysis (0.428); Sound Localization (0.418); Repairing (0.345); Equipment Maintenance (0.330); Category Flexibility (0.294); Spatial Orientation (0.294); Explosive Strength (0.281); Technology Design (0.281); Programming (0.274); Mathematics (0.248)
5: Routine Cognitive	Extent Flexibility (1.000); Gross Body Equilibrium (1.000); Multilimb Coordination (1.000); Performing General Physical Activities (1.000); Reaction Time (1.000); Response Orientation (1.000); Static Strength (1.000); Gross Body Coordination (1.000); Rate Control (1.000); Stamina (1.000)
6: Low-skill Manual	Active Learning (1.000); Active Listening (1.000); Critical Thinking (1.000); Negotiation (1.000); Oral Comprehension (1.000); Oral Expression (1.000); Persuasion (1.000); Reading Comprehension (1.000); Service Orientation (1.000); Social Perceptiveness (1.000)
7: Low-skill Service	Analyzing Data or Information (1.000); Complex Problem Solving (1.000); Deductive Reasoning (1.000); Inductive Reasoning (1.000); Systems Analysis (1.000); Systems Evaluation (1.000); Updating and Using Relevant Knowledge (1.000); Developing Objectives and Strategies (1.000); Flexibility of Closure (1.000); Judging the Qualities of Things, Services, or People (1.000)

Notes: This table lists, for each extreme profile, the descriptors of skills or tasks for which it is most likely a low level of ability is required (within the first tercile of the distribution across occupations for that skill or task). The estimated probability $\hat{\theta}_{j,k,1}$ is in parentheses.

Table A2: Descriptors most likely to be in second tercile by occupation extreme type

Group	Descriptors of skill or task most likely to be in second tercile
1: Scientific	Rate Control (1.000); Response Orientation (0.958); Reaction Time (0.945); Operation and Control (0.898); Control Precision (0.873); Multilimb Coordination (0.849); Depth Perception (0.817); Manual Dexterity (0.798); Handling and Moving Objects (0.789); Controlling Machines and Processes (0.747)
2: Other Abstract	Monitor Processes, Materials, or Surroundings (0.493); Technology Design (0.484); Near Vision (0.476); Programming (0.470); Performing for or Working Directly with the Public (0.469); Speed of Closure (0.467); Problem Sensitivity (0.451); Flexibility of Closure (0.445); Far Vision (0.433); Interacting With Computers (0.431)
3: Skilled Manual	Active Learning (1.000); Deductive Reasoning (1.000); Judgment and Decision Making (1.000); Fluency of Ideas (1.000); Originality (1.000); Reading Comprehension (1.000); Written Expression (0.996); Critical Thinking (0.987); Oral Comprehension (0.978); Written Comprehension (0.973)
4: Skilled Service	Reaction Time (1.000); Rate Control (1.000); Depth Perception (1.000); Operation and Control (0.998); Control Precision (0.969); Fluency of Ideas (0.968); Troubleshooting (0.958); Installation (0.958); Number Facility (0.948); Repairing and Maintaining Mechanical Equipment (0.932)
5: Routine Cognitive	Critical Thinking (0.982); Inductive Reasoning (0.915); Writing (0.901); Active Learning (0.896); Written Expression (0.892); Oral Comprehension (0.891); Judgment and Decision Making (0.881); Oral Expression (0.874); Reading Comprehension (0.85); Deductive Reasoning (0.834)
6: Low-skill Manual	Visual Color Discrimination (0.633); Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment (0.542); Science (0.541); Visualization (0.534); Monitor Processes, Materials, or Surroundings (0.529); Inspecting Equipment, Structures, or Material (0.515); Far Vision (0.515); Quality Control Analysis (0.509); Flexibility of Closure (0.499); Assisting and Caring for Others (0.498)
7: Low-skill Service	Night Vision (1.000); Installation (1.000); Peripheral Vision (1.000); Sound Localization (0.986); Glare Sensitivity (0.984); Multilimb Coordination (0.979); Equipment Maintenance (0.976); Repairing (0.959); Rate Control (0.892); Reaction Time (0.889)

Notes: This table lists, for each extreme profile, the descriptors of skills or tasks for which it is most likely a moderate level of ability is required (within the second tercile of the distribution across occupations for that skill or task). The estimated probability $\hat{\theta}_{j,k,2}$ is in parentheses.