# Amenities within industries and occupations[*]

*PRELIMINARY.*

**Saman Darougheh**     **Alexander Dietrich**
Danmarks Nationalbank   Danmarks Nationalbank

**Jonna Olsson**
Norwegian School of Economics

September, 2025

## Abstract

We discuss how to extract amenities using a standardized work-environment survey and demonstrate cross-sectoral difference in amenity exposure.

*Keywords:* Amenity choices, industries
*JEL Classification:* J16, J31, J13, J22, J24

# 1 Introduction

In recent years, research in the role of amenities –workplace characteristics other than pay– in determining jobs choices and career patterns has surged. While some researchers extract individual amenities using hand-crafted surveys, several countries offer large scale standardized work-environment surveys (Sweden, Denmark), surveying workers on many characteristics and settings within their specific workplace and employment situation. Extracting amenity information out of these surveys is not straight-forward, as many questions are similar in phrasing or highly correlated, asking for aggregation into meaningful amenity variables. However, a pure hands-off approach (such as principal component analysis) will aggregate (potentially very different) questions based on response correlations alone – which renders the interpretation of such components difficult.

In this note, we suggest an approach on how to extract amenity information out of these surveys. We then show that these amenities differ significantly across occupations and industries.

# 2 Methodology

**The data**   Our data stems from the 2021 wave of NOA-L (National Overvågning af Arbejdsmiljøet blandt Lønmodtagere), a national survey of the work environment performed by Arbjedstilsynet (The Danish Working Environment Authority). The survey has 25,520 employed respondents and consists of a wide range of questions related to working conditions and general physical/mental health.

These questions range from working hours to the level of noise during the work day and worry about being relocated. Most of the questions we use are asked on a Likert scale from 1-5 (for example, not worried at all of being relocated to extremely worried of being relocated), while a few are yes/no and a few are on a scale from 1-10.

## 2.1 The challenge

The goal is to estimate individual interpretable amenities out of the large set of questions provided by the work-environment survey, that can be used to estimate, for example, compensating differentials in wages. Below, we will outline several challenges of the data and lay out our approach.

First, several questions may ask differently for questions with a very similar notion. In section 2.2 we discuss how to reduce such questions into a single amenity variable. Second, respondents might not perceive an amenity question as relevant, which can lead to many missing amenity values, as the survey does not force a response or offers a "don't know/ does not apply" option. We discuss this in 2.3. Third, not all workplace

characteristics will translate into differences in compensating differentials. We discuss such exceptions in 2.4.

## 2.2 Reduction

One approach might be to manually select the question that fits best, according to the researchers subjective beliefs. Yet, in order to limit the extent to which an individual researcher's choices can affect research outcomes, "hands-off approaches" such as Principal Component Analysis (PCA) are often preferred. PCA is often used to extract underlying patterns out of a multitude of questions, for example to reduce the task requirements of different occupations, provided by O*NET, to a manageable size of job attributes. Typically, the components are then rotated such that the resulting components are orthogonal to each other. These rotations typically make it easier to interpret the resulting components.

A natural starting point would be to perform a PCA analysis on the set of $N$ questions available and set the number of extracted components $K$ either according to researcher's preference, or such that a specified share of total variance is explained.

One problem in such an approach lies in the fact that in general, all components will load (to some extent) on all questions, even after orthogonal rotations. This is an issue because even a small loading on an unrelated question might significantly change the value of the amenity. Several approaches exist to estimate sparse PCA —which attempt to set many loadings to zero. These approaches deal with this issue, but not the underlying problem: principal component analysis created components based on correlations, not based on content.

For example, consider a situation in which amenity-related questions belong naturally to $K$ amenities, and the questions belonging each amenity correlate most strongly with each other and less strongly with all other questions. In this situation, a PCA analysis with $K$ components could yield reasonable amenities. However, if questions belong to different amenities correlate more strongly with each other than within the ameniy block, some of the $K$ components might load on a mix of questions that do not belong (content-wise) to the same amenity.

This issue not only pertains to extracting amenities out of a set of (content-wise) unrelated questions, but also to extracting amenities out of questions that strongly belong to each other. Consider a researcher that attempts to estimate two amenities: "leadership quality" and "good relationship with colleagues" out of a larger set of questions. One might be tempted to estimate two principal components based on all questions pertaining to these two amenities. Yet, even after orthogonal rotation, one might arrive at two components that make intuitive sense, but do not resemble amenities: a first component that loads broadly on all related questions, and a second component that loads positively on "good manager"-related questions, and negatively on "good relationship with colleagues"-related questions (or vice-versa).

That is, one of the two components becomes a general component that captures whether any good characteristics exists in the workplace, and a second one that distinguishes which of the two – leadership or colleagues – is present. While the two components jointly span the space of two amenities, the interpretation of each of them in sense of classical amenities is difficult.

With all this in mind, we propose the following approach. For each amenity in mind, if a single question exists that maps very closely to that amenity, we choose that single question as our amenity measure. Whenever such a single question does not exist, but several questions can reasonable be used to compute the amenity (suggestively indicated by all of these questions highly correlating with each other), we extract a single principal component based on all relevant questions. Importantly, all questions within the same principal component analysis are asked in the same scale.[1]

## 2.3 Missing

Some questions have a high share of missing values, which may be due to respondents not willing to answer. Alternatively, some very specific work-environment related questions may appear non-nonsensical to respondents in certain occupations.

This puts a limit on the set of questions that can be used for amenity extraction, especially when one wants to study the effect of all amenities in a joint analysis. We proceed as follows: we exclude any question with a response rate of below 90% (of the employed population). Most of the remaining questions have a response rate between 95% and 99%.

## 2.4 Exceptions

We exclude two types of questions from our survey. First, we exclude questions that pertain to the worker's individual physical or mental health (such as "are you stressed", "are you anxious", or "are you tired"). While the answer to such questions might be partly due to work-environment factors, we are concerned that a large share of worker-level variation might be unrelated to their jobs.[2]

Second, we exclude questions about job events that occur very infrequently, such as sexual harassment, bullying, or injuries. We exclude these because they measure ex-post outcomes, not ex-ante risk. Even if aggregated to firm- or occupation- level, the low incidence rate will render such estimated probabilities very imprecise.

---

[1]Standard PCA based on ordinal data is troubling if the distances between different values are not perceived to be equal, but alternative approaches exist. We show in our application that estimating the components based on binary versions of all variables yields similar results.

[2]Indeed, even when collapsed to occupation-level analysis, it is difficult to exclude the possibility that workers with poor mental health self-select into certain occupations.

# 3 Implementation and results

We construct our amenity measure using a subset of 51 NOA-L questions that capture job-specific conditions. These questions range from working hours to the level of noise during the work day and worry about being relocated. Most of the questions we use are asked on a scale of 1-5 (for example, not worried at all of being relocated to extremely worried of being relocated), while a few are yes/no and a few are on a scale 1-10.

Consistent with the discussion in the previous section, some amenities directly correspond to a single question (for instance: "does your job include work during nights") while some amenities are generated from clusters of closely related questions, grouped with principal component analysis.[3]

We standardize all amenities. In the end, we have 23 amenities giving a broad perspective on the working conditions for the individual.

**Decomposition into occupations and industries**   To what extent do amenities vary by occupations or industries? Figure 1 displays a decomposition of the variance of each of our amenities into an industry component, an occupation component, and a selection component. Occupation and industry are defined at two-digit aggregation. The selection component captures the extent to which occupation-industry interactions capture differences in the variance beyond the additively separate occupation and industry components.

For most amenities, our three components explain only a minority of the variance — the range of the unexplained variance varies between 90% for relocation risk and 45% for physically demanding. Most of the explained variance stems from the sorting component: for very few amenities, a pure occupation or industry component captures a significant share of the amenities.

---

[3]For instance, we group all 14 questions in the survey modules "Management and collaboration", "Immediate manager", "Cooperation with your closest colleagues", and "Trust and fairness" into one amenity capturing the collegial environment.

# A  Tables

**Table 1:** Amenities with a 1-to-1 corresponding survey question

| Variable | Original Variable Name | Original Variable Description | Additional Info |
|---|---|---|---|
| Noise | S_STOEF_KRAFTIG_SF | | |
| Vibration | S_VIBRAT_UDSAT_SF | | |
| On-call duty | PV_RAADIGHED_V | Are you available outside working hours? | |
| Work on days off | PV_ARB_FRI_V | Do you think about work during leisure time? | |
| Hours flexibility | PV_TYPE_OVERARB_1 | | Set missing if missing PV_AR_OVER_V |
| Overtime necessary | PV_ARB_OVER_V | Do you work overtime? | Set to zero if AMN_PV_FLEX_DAYS == 1 |
| Emotional involvement | PV_FOEL_SITU_V | Are you feeling in control of the situation at work? | |
| Job loss risk | PT_BEKYM_ARBLOES_V | Are you worried about unemployment? | |
| Relocation risk | PT_BEKYM_FORFLYT_V | Are you worried about unwanted transfer? | |
| Physically demanding | EF_FYSISK_ANSTGRENG_SF | | |
| Night work | A_NATARBEJDE_01 | Fixed night work or changing times with night work | |
| Shift work | A_DOEGN_PRIM_ARB_SF | What time of day do you work | If value == 5 |
| Work hours: 40-60 | A_TIMER_PRIM_ARB | | 40 < hours < 60 |
| Work hours: 60+ | A_TIMER_PRIM_ARB | | hours > 60 |

*Notes.* The table lists the variable used to construct the

## Table 2: Amenities based on PCA

| PC | Amenity 1 | Amenity 2 | Amenity 3 | Amenity 4 | Amenity 5 |
|---|---|---|---|---|---|
| **Health(HA)** | | | | | |
| Pc1  No | | 0.710 HA KOMME SIG SF | 0.710 HA UDMATTET SF | | |
| **Workplace environment($I_{M}ILJOE$)** | | | | | |
| Pc1  No | | 0.710 Prevention a priority | 0.710 The employees involved | | |
| **Workplace health($I_{S}UNDHED$)** | | | | | |
| Pc1  No | | 0.710 Offers exercise facilities | 0.710 Offers small exercise activities | | |
| **Autonomy(PID)** | | | | | |
| Pc1  No | | 0.510 PI TRAEF BESLUT V | 0.510 PI BEFOEJ ANSVAR V | 0.510 PI INDFLYDELSE V | 0.470 PI RAEKKEF ARB V |
| **Development(PIL)** | | | | | |
| Pc1  No | | 0.710 PI UDVIK KOMP V | 0.710 PI LAERE NYT V | | |

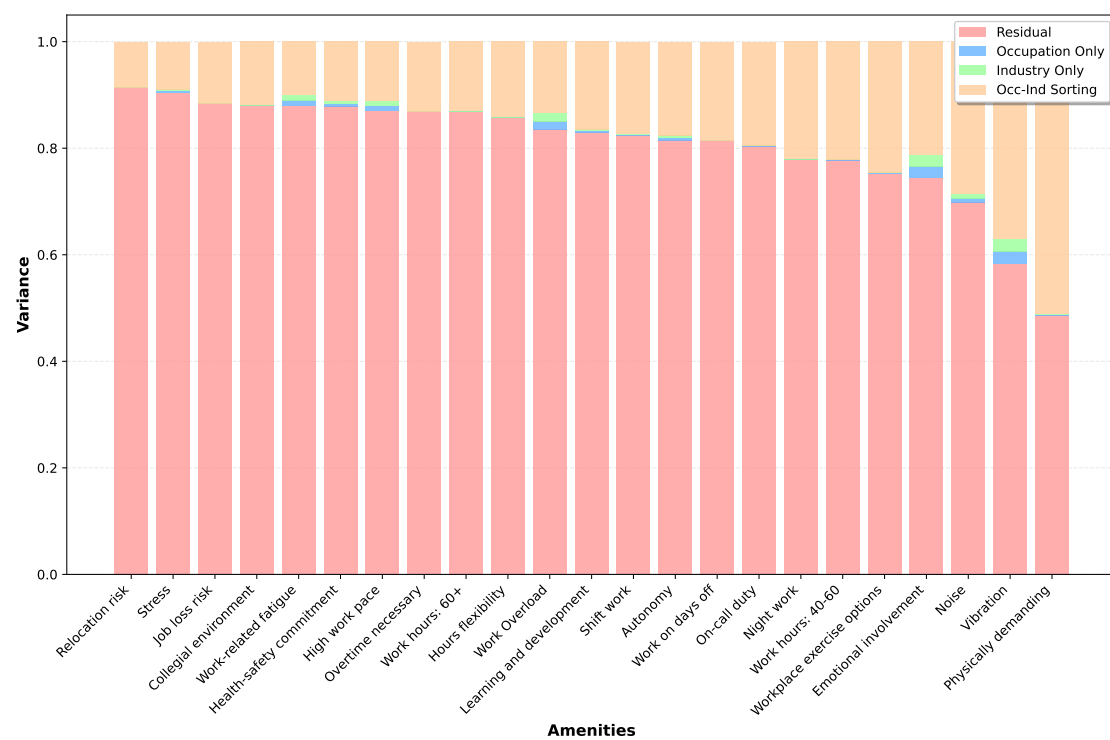| PC | Amenity 1 | Amenity 2 | Amenity 3 | Amenity 4 | Amenity 5 |
|---|---|---|---|---|---|
| **Supervisor/team relations($PR_{E}NV$)** | | | | | |
| Pc1  No | 0.300 PR LEDER TRIVSEL V | 0.300 PR MOTIVERE V | 0.290 PR KLARE MAAL V | 0.290 PR FORBEDR SAMARB V | 0.290 PR BEHAND RETFAER V |
| **BUSY($PV_{B}USY$)** | | | | | |
| Pc1  No | 0.440 PV BAGUD ARB V | 0.430 PV TIDSFRISTER V | 0.420 PV UVENT ARB V | 0.410 PV NAAR IKKE V | 0.380 PV TEMPO KVALI V |
| **Work tempo($PV_{T}EMPO$)** | | | | | |
| Pc1  No | −0.710 PV ARB TEMPO V | 0.710 PV PAUSER V | | | |

# B Figures

**Figure 1:** Fields decomposition of amenity variance into occupations and industries