Statistisk sentralbyrå
Statistics Norway

Estimation of training costs for the Norwegian Labour Cost Survey (LCS)

Susie Jentoft

TALL SOM FORTELLER

NOTATER / DOCUMENTS

2019/5

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

Published 16 January 2019

ISBN 978-82-537-9871-4 (electronic)

Symbols in tables	**Symbol**
Category not applicable	.
Data not available	..
Data not yet available	…
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0.0
Provisional or preliminary figure	*
Break in the homogeneity of a vertical series	—
Break in the homogeneity of a horizontal series	|
Decimal punctuation mark	.

Preface

Labour costs are no longer estimated using a traditional survey in Norway but instead are calculated based on register and alternative data sources. Estimates for 2016 are required to be reported to Eurostat. This document outlines the estimation for one category of indirect labour cost: those associated with training costs. This work has been a collaboration project between the Division for Methods and the labour costs working group which is now within the Division for structural business statistics.

Statistisk sentralbyrå, 21 December 2018

Christian Thindberg

Abstract

The traditional Norwegian Labour Cost Survey (LCS) has been replaced by alternative data sources in 2016. Most of the key costs are easily available and of a high quality through Norway's extensive registers. Indirect labour costs associated with training are not however covered by registers and therefore need to be estimated. We investigated alternative data sources for estimating these costs, and develop an estimation based on data from the Continued Vocational Training Survey. We matched the populations and definitions to ensure comparability. Three estimation methods were tested for estimating costs in 2016: a rate model, regression and nearest neighbour imputation. The rate model provided the best estimates and produced totals that were comparable with previous LCS results. Rate models were used to predict training costs per employee and register data on the number of employees was used to calculate total costs. The estimated total indirect labour costs for training in 2016 were approximately NOK 10 billion (excluding apprentices).

Contents

1. Background

Statistics Norway decided to discontinue the traditional Labour Cost Survey (LCS) in 2016 and instead will deliver figures to Eurostat based on administrative and alternative data sources. Most of the data on labour costs collected for LCS are now accessible directly from Norway's newly established employee/employer register called *A-ordning*, established in 2015. However, costs associated with the training of employees are not collected by any administrative data sources in Norway and this is a post required to be reported to Eurostat at a macro-level.

Training costs represent a small fraction of the total labour costs. In 2012, it was estimated that an average of 9200 NOK per full-time equivalent of a total 683 000 NOK (approximately 1 per cent) was spent on training costs. The population for LCS consists of all businesses with 10 or more employees, in the industry groups B-N, P-S (explanation given in Appendix A). This was around 31 000 businesses in 2016.

In this study we use an alternative data source, the Continued Vocational Training Survey (CVTS), combined with administrative data to test three different estimation methods for training costs for 2016. These are compared to predicted estimates of training costs based on previous LCS data for comparison. We do not wish to use previous LCS data in the final estimation as it is becoming increasingly outdated, with no plans for future surveys.

2. Description of data sources

CVTS is a European legislated survey run every 5-years. The last two surveys were in 2010 and 2015, and these are used as the main data source for the new estimation methods. This survey includes similar questions to LCS on both training participation and costs associated with training. Training costs reported from this survey generally include Personal Absence Costs (PAC) however this should not be included in the LCS post on training as it is included indirectly in other areas. Therefore, PAC are excluded from CVTS data when used for estimation in LCS. A previous study by Eurostat compared variable and population definitions between LCS and CVTS[1] (Eurostat, 2014). They concluded that LCS has a broader definition of training and is asked in a way that includes both formal and more informal training. Therefore, when replacing LCS with CVTS data there may be an underreporting of training costs under the LCS definition. This is believed to be relatively small and is not adjusted for in this study.

The costs associated with apprentices should be reported as part of the training costs in LCS however is distinguishable as a separate post. CVTS does not collect data on costs associated with apprentices. From 2015, Statistics Norway has access to administrative data from *A-ordning* on costs associated with apprentices which can be used directly. This cost has therefore been taken from this data source, integrated with the business population, and summed. As it represents a separate post for delivery to Eurostat it is excluded in the general estimation procedure and previous LCS estimates have excluded apprentices' costs to make it more comparable.

In addition, we use data from the business register (VOF) in the estimation which includes the number of companies, the number of employees and turnover within NACE (classification of economic activities) groups. A summary of the data sources and key differences is given in Table 2.1.

[1] Eurostat (2014). Future of the CVTS data collection. WORKING GROUP LABOUR MARKET STATISTICS, Luxembourg. *Eurostat/F3/LAMAS/31/14*

Table 2.1. Summary of data sources used in estimation with comparison of coverage and definitions

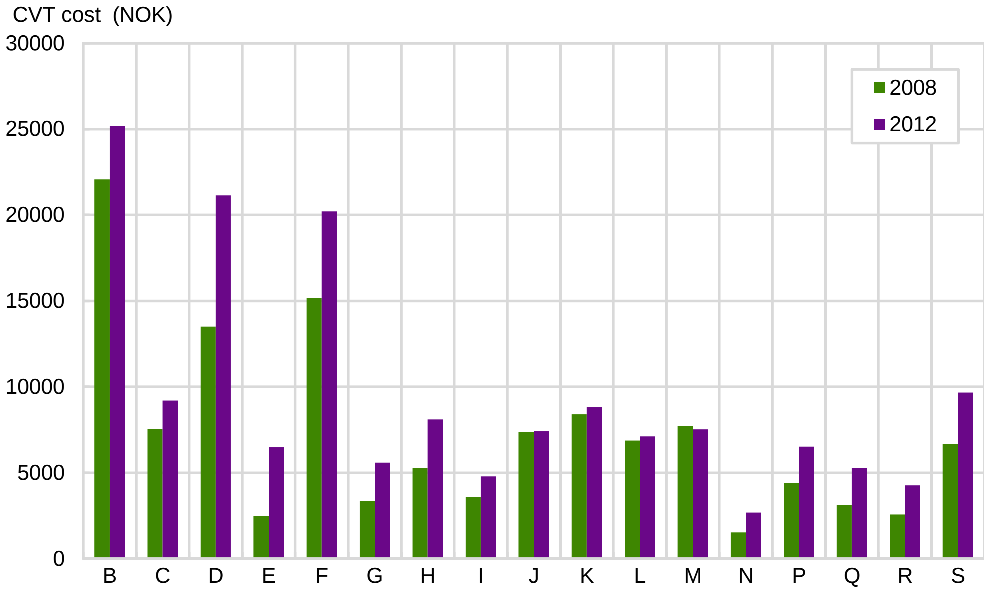
Data source	LCS	CVTS	A-ordning	VOF
Coverage	Sample	Sample	Population	Population
Size coverage	>=10 employees	>=10 employees	All	All
Nace coverage	B-N, P-S	B-N, R-S	All	All
Frequency	Every 4 years (2008, 2012)	Every 5 years (2010, 2015)	Continuous – monthly	Continuous – monthly
Reference period	Year	Year	Month – summed to year	Month or year
Apprentice cost	Included	Excluded	Included	No costs
Training definition	Formal and informal training	Formal training	N/A	N/A

Source: Statistics Norway.

3. Completion of population

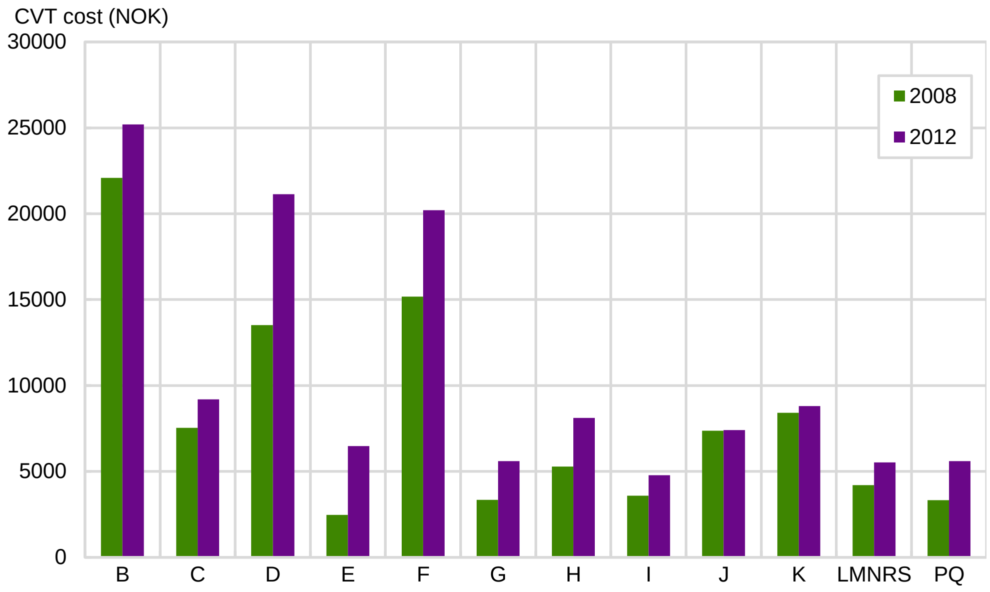
Data used for modelling from CVTS does not include NACE groups P (Education) and Q (Human health and social work activities). We therefore need to adjust the estimation method in some cases to allow estimates that include these groups. Figure 3.1 gives a comparison of training cost per employee for each of the industry groups based on previous LCS data where P and Q are included in data collection. Figure 3.2 gives average training costs in broader groups whereby L (Real estate activities), M (Professional, scientific and technical activities), N (Administrative and support service activities), R (Arts, entertainment and recreation) and S (Other service activities) are combined and P and Q are combined. These 2 combined groups have similar training cost per employee and therefore the group LMNRS is used in the following methods to estimate P and Q.

Figure 3.1. Average training cost per employee by main NACE group using LCS data



Source: Statistics Norway.

Figure 3.2. Average training cost per employee with broader NACE groups (LMNRS and PQ)



Source: Statistics Norway.

4. Estimation methods for 2016

In addition to certain population and variable definition differences between LCS and CVTS, there is a time difference. That latest CVTS was for the reference year 2015, however, we want to deliver figures for 2016 and in addition, have the possibility for training costs to be estimated on a yearly basis. Three estimation methods were tested and are summarized in the following sections.

4.1. Method 1: Predicted rate estimation with number of employees

We used a simple linear extrapolation of a rate estimation for cost per employee to calculate training costs for 2016 (excluding apprentice). The extrapolation used the cost per employee, r_h, within standard industry strata, h, using CVTS data (2010 and 2015):

$$r_h = \frac{\sum_{i\in s_h} c_{ih} w_{ih}}{\sum_{i\in s_h} n_{ih} w_{ih}}$$

where $c_{ih}$is the training costs for business i in strata h from survey data, $w_{ih}$is the associated adjusted weight for business i, and $n_{ih}$is the number of employees from the business population register. Weights were adjusted so that

$$\sum_{i\in s_h} n_{ih} w_{ih} = \sum_{i\in U_h} n_{ih}.$$

We chose to extrapolate the rate per employee for 2016 rather than the total cost as this should be more robust while still considering recent changes and trends in employee numbers. We used a simple linear extrapolation based on the previous two surveys' data to predict the rate per employee for year 2016 within each stratum ($r_{h,2016}$) as

$$r_{h,2016} = r_{h,2010} + \frac{6}{5}\left(r_{h,2015} - r_{h,2010}\right).$$

The rate was then converted to a total training cost, $c_{h,2016}$, within the strata and to the population c_{2016}:

$$c_{h,2016} = r_{h,2016} \sum_{i\in U_h} n_{ih,2016}$$

and

$$c_{2016} = \sum_{h=1}^{H} c_{h,2016}.$$

4.2. Method 2: Regression

Regression modelling was used to predict 2016 training costs using CVTS data from 2015 and register data from VOF on the number of employees and turnover. This was to test whether we could predict 2016 without linear prediction but using real changes in turnover and employee numbers to reflect training cost changes. A weighted linear model was fitted on 2015 CVTS (sample) data within broader industry groups, g, described in appendix A, as we did not have enough (non-zero) observations within the standard strata used in the previous method. The following model was used within groups, including an interaction term

$$y_i = \log(c_i) = \beta_0 + \beta_1 \log(n_i)\log(x_i) + \beta_2 \log(n_i) + \beta_3 \log(x_i) + \varepsilon_i$$

where n_i is the number of employees, and x_i is the total turnover for company i for the year 2015. We used a weighted linear model that minimises

$$\sum w_i \varepsilon_i^2$$

where w_i is the adjusted survey weight for company i. Coefficients from the model were then used to predict y_i for all companies in the population for 2016. Both n_i and x_i came from 2016 VOF data for prediction. Predicted cost were summed for all companies within strata, h, for comparison with other methods by

$$c_{h,2016} = \sum_{i \in U_h} \exp(\hat{y}_{hi}).$$

Prior to modelling we tested whether to include both dependent variables with an interaction term using AIC backward selection and found that the inclusion of an interactive term provided the best fit. During modelling, we observed some extreme outliers that may impact significantly on the estimation and excluded them from the main modelling. We classified outliers when: $\text{studentized residual}_i > 10$. For those in 2015 with turnover = 0, an alternate model was used including only number of employees as the dependent variable.

4.3. Method 3: Nearest neighbour imputation

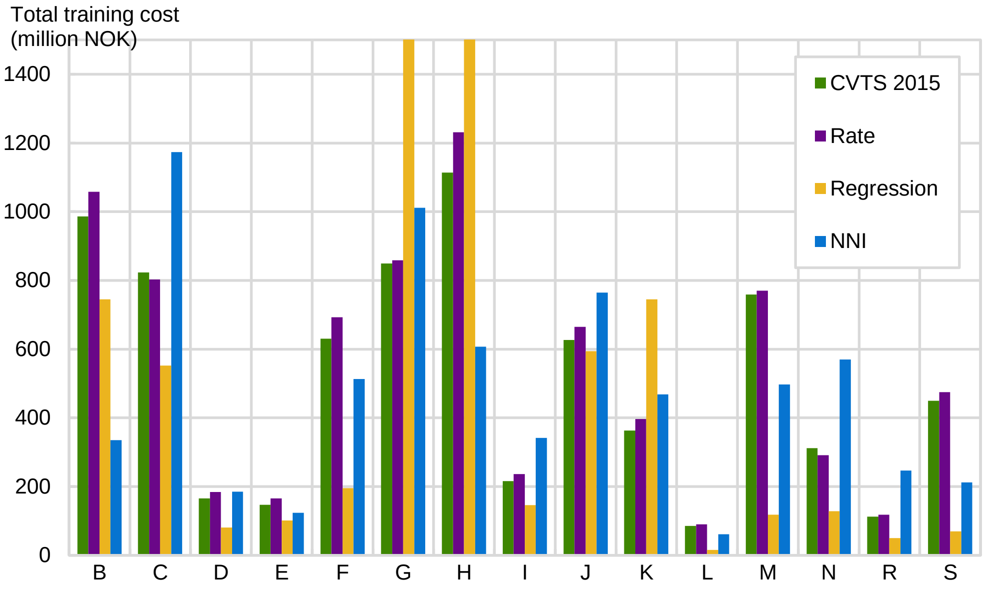
The distribution of training costs was not easy to fit as there are many companies that report 0 training costs. This creates a skewed distribution which effects normal assumptions associated with regression methods. We tested an alternate non-parametric method based on the distribution of the actual data using a cold-deck, nearest neighbour imputation for the entire population. We wanted to test this method to see if it would produce reasonable estimates in addition to maintaining a similar structured distribution to our survey data. We used predictive mean matching to find the nearest neighbour based on the same regression models described in method 2. Predicted values for each company i, based on 2016 register data were matched to the nearest predicted value based on the CVTS sample data from 2015. The observed cost value for the sample unit was then used directly as the donor for the population unit in 2016. We hoped by doing this matching with prediction we would take account of both inflation factors while using observed values to better maintain the distribution of costs.

5. Results

5.1. Method comparison

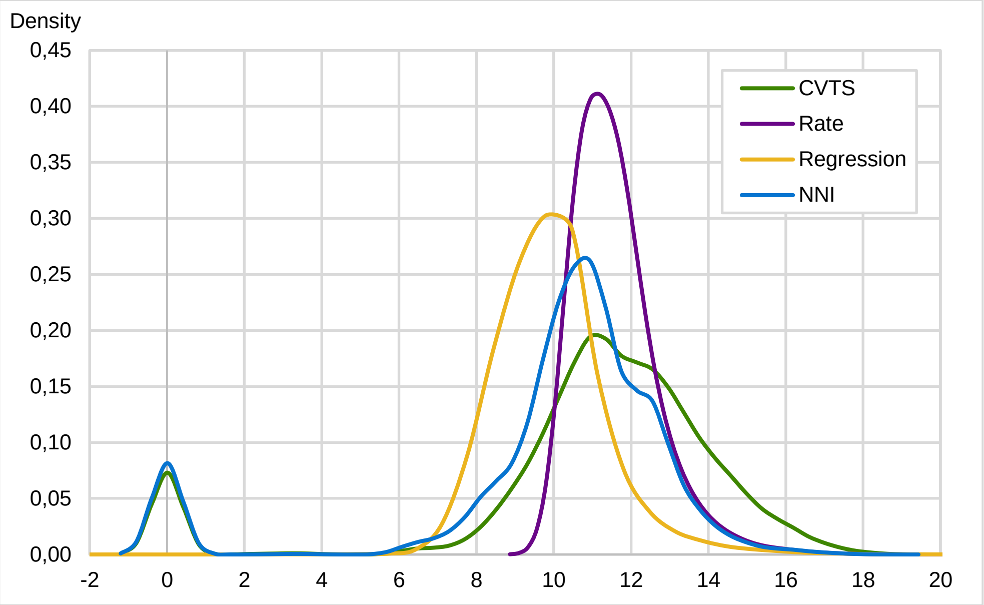
Results of the three methods for estimation of training costs within each of the industry strata are shown in figure 5.1 along with estimates from CVTS 2015 using official survey estimation weights (adjusted to our population). In regression estimation there were 2 NACE-groups (G & H) where extremely large values were estimated and are not shown on the figure. This appears to be when there was a breakdown in the correlation between training costs and number of employee and turnover for some larger companies. Figure 5.2 shows the distribution of predicted training costs at a company level. Method 3, NNI, appears to maintain the distribution in a way most resembling that of the CVTS data, however overall results appear very inconsistent with the previous CVTS. Given we are most interested in the *point estimate* for the total rather than the distribution or percentiles we perceive the rate estimate to be the best estimate as it is most consistent with previous years while reflecting some progression.

Figure 5.1 Comparison of total training costs within NACE groups using three estimation methods and compared with weighted results from CVTS



Source: Statistics Norway.

Figure 5.2. Distribution of training costs (log) in CVTS 2015 compared with three estimation methods

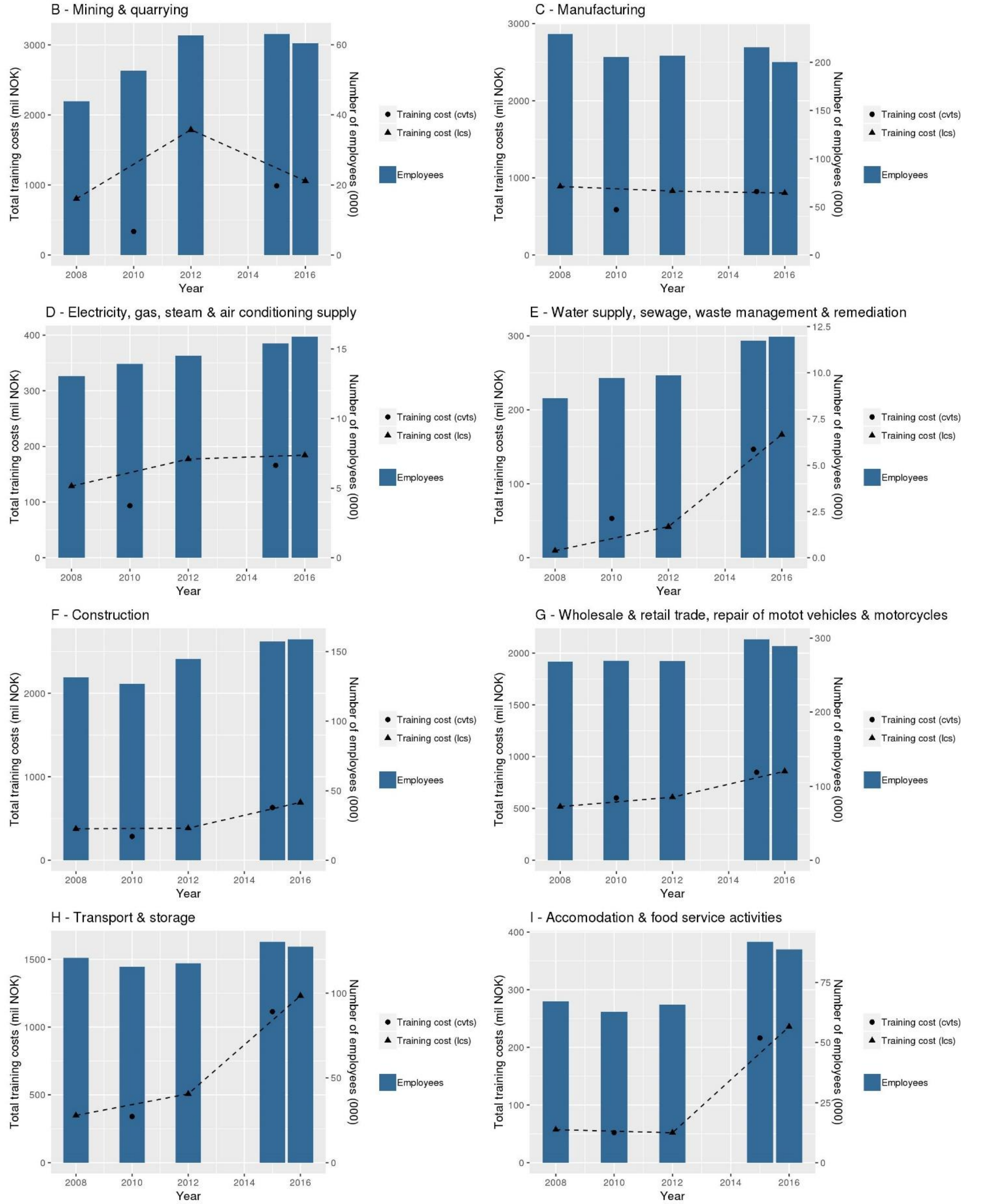


Source: Statistics Norway.

5.2. Comparison with LCS

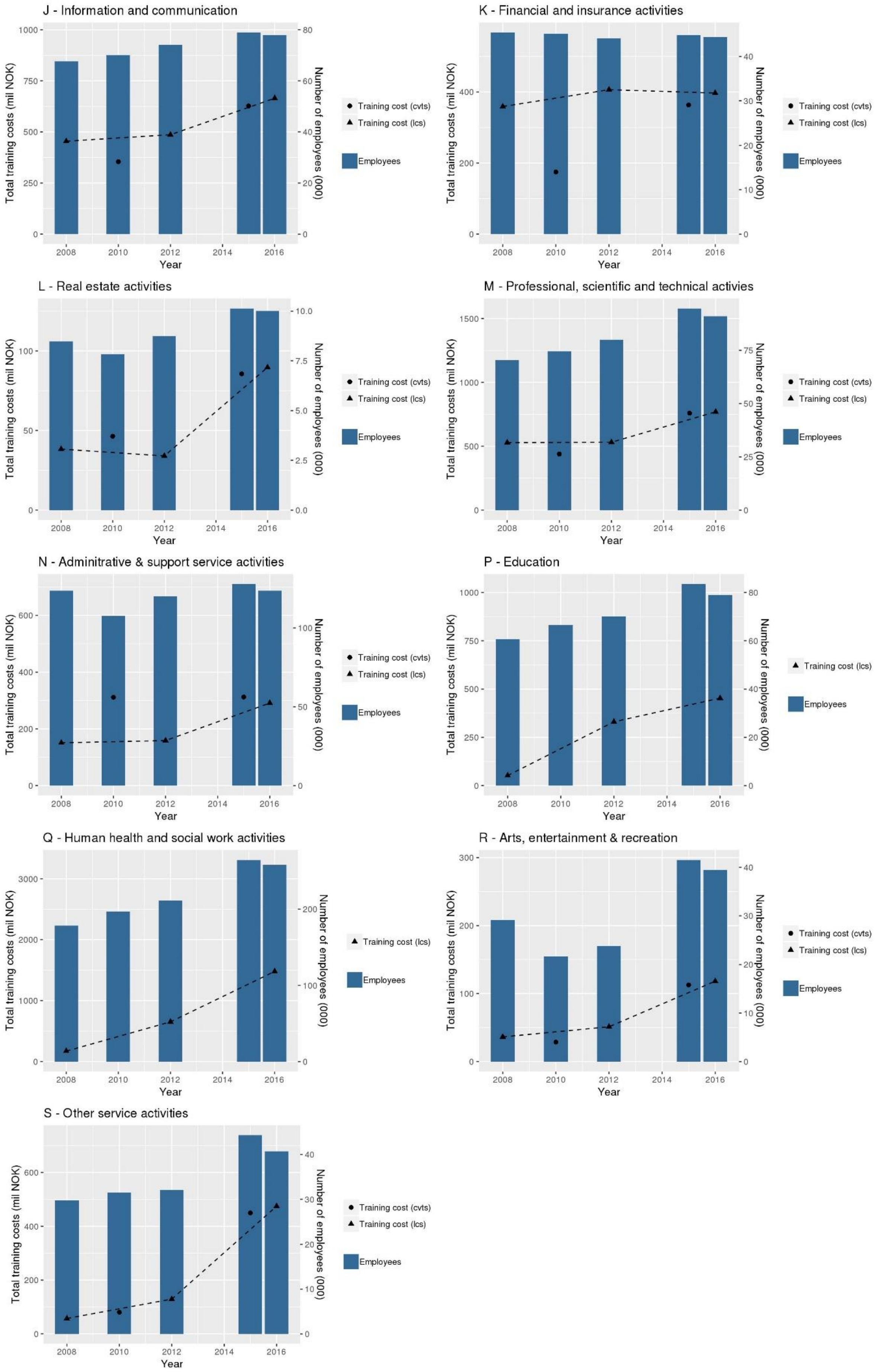
Estimates for 2016 are shown in Figure 5.3 and compared with previous LCS estimates (2008 & 2012) by industrial classification groups. Many of the estimates appear to be reasonable in comparison with the previous data source. For example, industrial group G - Wholesale trade, LCS and CVTS appears to have been well aligned in previous surveys and the estimation for 2016 is in line with what we would expect from both sources. However, in group B – Mining and quarrying – there is likely a difference in what the two data sources are measuring and a break in the series is introduced due to changing the data source.

Figure 5.3. Comparison of total training costs from LCS and CVTS including LCS estimate for 2016 within standard industrial classification groups. Total number of employees (register) is included for additional comparisons



Source: Statistics Norway.

Figure 5.3 continued



J - Information and communication

K - Financial and insurance activities

L - Real estate activities

M - Professional, scientific and technical activies

N - Adminitrative & support service activities

P - Education

Q - Human health and social work activities

R - Arts, entertainment & recreation

S - Other service activities

Source: Statistics Norway.

5.3. Total costs

Total training costs for Norway are given in Table 5.1. Overall with the new estimation and data source we see an increase in training costs similar to that expected based on previous years' data. The new data source for costs associated with apprentices also shows an increase but a relatively stable percentage of the total training costs.

Table 5.1 Total training costs based on LCS in 2008 and 2012 and estimation using CVTS in 2016. Million NOK

	2008	2012	2016
Total training cost (without apprentices	4975	7165	9968
Total training cost (including apprentices)	8208	11941	14667

Source: Statistics Norway.

6. Discussion

Of the three estimation methods tested we believe that a rate model using the number of employees from register data gives the most reasonable statistics and has been chosen as the preferred method. The new data source (CVTS) shows some differences, particularly in the industry group: Mining and quarrying. CVTS appears to have an overall lower level of training costs in these groups compared to that previously recorded by LCS. This may be in part due to differences in definition.

We conclude, given that LCS has not been undertaken in 2016, that CVTS provides the best data source for training costs in Norway and allows Statistics Norway to produced adequate statistics for this post at a macro-level for Eurostat. A rate model using number of employees and a linear extrapolation for cost per employee produces reasonable estimates and allows statistics at a country level, for main industrial groups and at a NACE 2 level. This method may also be used for yearly estimations and updated with new CVTS data after the next survey round.

Appendix A: Standard industrial classifications

Strata (h)	Group (g)	Name	Inclusion in LCS (Norway)
A	n/a	Agriculture, forestry and fishing	No
B	Industry	Mining and quarrying	Yes
C	Industry	Manufacturing	Yes
D	Industry	Electricity, gas, steam and air conditioning supply	Yes
E	Industry	Water supply, sewerage, waste management and remediation activities	Yes
F	Construction	Construction	Yes
G	Trade	Wholesale and retail trade; repair of motor vehicles and motorcycles	Yes
H	Trade	Transportation and storage	Yes
I	Trade	Accommodation and food activities	Yes
J	IKT	Information and communication	Yes
K	IKT	Financial and insurance activities	Yes
L	Prof	Real estate activities	Yes
M	Prof	Professional, scientific and technical activities	Yes
N	Prof	Administrative and support service activities	Yes
O	n/a	Public administration and defence; compulsory social security	No
P	Prof	Education	Yes
Q	Prof	Human health and social work activities	Yes
R	Prof	Arts, entertainment and recreation	Yes
S	Prof	Other service activities	Yes
T	n/a	Activities of household as employers; undifferentiated goods- and services-producing activities of households for own account	No
U	n/a	Activities of extraterritorial organisations and bodies	No

List of figures

List of tables