Are jobs more polarized in ICT firms?

Petri Böckerman,* Seppo Laaksonen** and Jari Vainiomäki***

Abstract

We perform decompositions and regression analyses that test the routinization hypothesis and implied job polarization at the firm level, instead of the aggregate, industry or local level as in prior studies. We examine the technology explanations for routinization and job polarization using firm-level ICT factors derived from a factor analysis of indicators for the adoption of ICT at the firm level as explanatory variables in regressions for wage bill shares of different education and occupation groups. Our results for the abstract and routine occupation group are consistent with the routinization hypothesis at the firm level, which relates to ICT adoption. The service sector share is independent of ICT at the firm level, but our decompositions show that increasing service sector is related to the demand effect at the aggregate economy level, i.e. shifts in production between firms and the entry of new service intensive firms. In summary, we find evidence for disappearing middle (routine work) due to technological change both at the firm and at the aggregate level.

Acknowledgements: We are grateful for financial support from the Palkansaaja Foundation and Academy of Finland Strategic Research Council project Work Inequality and Public Policy (number 293120). We thank Pekka Laine from Statistics Finland for help in constructing the data for this paper.

Manuscript
October 2015

Affiliations of the Authors:

* Turku School of Economics and Labour Institute for Economic Research, Finland, and IZA
** University of Helsinki, Finland
*** University of Tampere, Finland

** Corresponding author's address
Jari Vainiomäki
School of Management
33014 University of Tampere
Finland
E-mail: jari.vainiomaki@uta.fi

1. Introduction

The prominent view in the literature is that the driving force for the increase of wage differentials, education premiums and skill-upgrading observed in industrialized countries since the 1970s/1980s has been skill-biased technological change (SBTC), see e.g. Bound and Johnson (1992); Acemoglu (2002) for a review. Card and DiNardo (2002), and Acemoglu and Autor (2010) have pointed out several anomalies that are difficult to explain with the standard SBTC model. The latter authors presented a task-based model of technological change to explain these anomalies. Earlier Autor, Levy and Murnane (ALM) (2003) focussed on computer-based production technologies, and also argued for an approach making a distinction between the tasks that workers do in their jobs, and the skills they have in order to perform these tasks. Adoption of computers at the workplace changes the demand for tasks performed at the jobs. Computers substitute for routine tasks, but complement non-routine analytical and interactive tasks. ALM showed that the decline in the price of computers leads to increasing demand for routine task input in total, including computer capital, but a reduction in routine task human labour input. This process is dubbed as the routinization hypothesis (or routine-biased technological change). The concurrent increase in analytical non-routine tasks increases the relative demand for educated workers, because they have a comparative advantage in these tasks and computer usage. On the other hand, the demand for intermediate skilled (educated) declines, because they often work in routine task intensive jobs.

Autor, Katz and Kearney (2006) and Goos and Manning (2007) argued that the routinization process may lead to job polarization, where employment growth concentrates to the low and high skill (wage) occupations, whereas the jobs at the middle of the skill distribution are diminished.[1] The low wage (education) non-routine manual and service jobs are probably not directly affected by computerization, but non-homothetic preferences in product demand may lead to increasing demand for low paid services and non-routine manual jobs.

[1] Jenkins (1995) provides an early predecessor of current literature on job polarization.

Acemoglu and Autor (2010) present empirical evidence for their task-based model using aggregate-level regressions for wage changes of different skill groups (defined by sex, education and experience). They measure the relative advantage of the skill groups in performing abstract, routine or service tasks by the shares of each skill group in abstract, routine, and service occupations prior to the computer era. The increase over time (that proxies for technological change) in the coefficients for these initial abstract and service shares compared to routine shares in explaining wage changes is consistent with their task-based model for technological change, routinization and polarization.[2]

ALM (2003) presented industry-level evidence for their model. They used task measures derived from the Dictionary of Occupation Titles (DOT) and showed that analytical and interactive non-routine task inputs rise more and routine task inputs decline more in industries that invest more heavily in computer capital. Michaels, Natraj and Van Reenen (2010) used country-industry panel data, and showed that wage bill shares (and relative wages) for both high and low education levels are positively related to industry ICT capital, whereas those for the middle educated are negatively related to ICT. This pattern is consistent with job polarization. Also Goos et al. (2010) used country-industry data to show that polarization is a pervasive phenomenon in 16 European countries, and further that the routine task intensity of occupations is a more important driver for employment polarization than offshorability.

Autor and Dorn (2013) apply a spatial equilibrium model to show that the falling cost of computing together with preference for variety in consumption can explain both the rise in service occupations, and polarization in employment and wages. Using commuting-zone data they find that local labour markets that initially specialized in routine tasks adopt more information technology, experience growth in service occupations and experience wage polarization between 1980 and 2005. An alternative explanation to the decline in routine (manufacturing) work is import penetration from, or offshoring to, low cost countries. Autor et al. (2013) exploit the commuting-zone

[2] It should be noted that Acemoglu and Autor (2010) describe their empirical exercise as "…highly preliminary – indeed, it is intended as an example of an empirical approach rather than a test of the theory…".

variation in rising Chinese import competition between 1990 and 2007 to examine recent labor market trends in terms of employment and wages in the US local labour markets. They document that increasing import penetration explains one-quarter of the aggregate decline in US manufacturing employment, and reduces wages in local labour markets. Autor et al. (2013b) show that highly trade-exposed and highly technology-exposed commuting zones are dissimilar in their occupational employment structure. Automation and trade therefore affect local labour markets differentially in terms of employment structure and wages.

In this paper we study the routinization hypothesis and the implied job polarization in the Finnish private sector using the Harmonized Wage Structure Statistics (HWSS) data of Statistics Finland at the firm level. As the discussion above indicated most (if not all) empirical studies of this issue have been conducted using different types of aggregate data (total economy, industries, locations). Our point of departure in this paper is to examine changes in the structure of labour demand at the firm level in order to study the technology explanation for routinization and job polarization at the micro level, where the actual labour demand decisions are made. In this way we account for the effects from the compositional changes in product demand on the structure of employment or wages. The compositional changes in production confound aggregate studies if they are driven by other factors than technology (for example by offshoring or consumer preferences). Furthermore, we are able to examine routinization and polarization at the firm level using firm-specific adoption of specific Information and Communication Technologies (ICT) as explanatory variables for employment structure in firm-level regressions. Using firm-level ICT indicators allows us to establish a more direct link between observed occupational or educational changes and computer-based technological change than aggregate studies based only on time effects or industry-level aggregates. Using micro data we are also able to perform decompositions of changes in educational and occupational employment (wage bill) shares into within firms, between firms, and entry-exit components to gain knowledge about the likely sources of these changes.

2. Data

Harmonized Wage Structure Statistics (HWSS) data of Statistics Finland combines the annual wage structure statistics data into a harmonized panel data, where all important wage measures and classifications, such as industry and occupation, are consistent across the years and sectors. The core of the annual wage structure statistics are the firm and individual level wage surveys of employer federations for their member firms. Statistics Finland augments these with samples of non-member firms and sectors not covered by the employer data. The harmonized data is available for the private sector annually from 1995 onwards. In this paper we exclude the recession years following the financial crisis after 2008.

Harmonization over time is needed because of the differences and changes in collective wage contracts and classifications used both over time and across sectors. The annual harmonization across different sectors takes into account the differences in wage concepts and compensation components used in different collective agreements. For example, hourly and monthly paid are made comparable.

In the panel data education, occupation and industry variables are harmonized over time to the latest versions of standard classifications of Statistics Finland. Formal education is available from a comprehensive register of completed degrees. The industry of firms is available at the 5-digit level but used in the analyses at the 2-digit level. Occupation codes in the primary data are converted into international ISCO 2001 codes at the 5-digit level but used in the analyzes at the 3-digit level. Unfortunately, it is not possible to completely harmonize some occupations for white-collar manufacturing workers over the break point 2001-2002 due to the classification change in the primary employer survey data. Hence, we either perform all our estimations using separate data before and after this break point using the periods 1995-2001 and 2002-2008, or focus only on the latter period.

This longitudinal data for the years 1995-2008 contains some 600 000–750 000 employees per year and about 28 000 firms exist in the data for at least one year during the period 1995-2008. Using sampling weights, these data are representative of the total private sector, except for the smallest firms, exempted from the wage surveys of employer associations and Statistics Finland.

Our wage concept, the “hourly wage for regular working time” includes basic pay and various supplements for working conditions and performance-pay paid on a regular basis. It does not include overtime pay or one-off items, such as holiday and annual performance bonuses. In addition to wages we observe regular working hours per month for each worker in these wage statistics. Because the employer firm of each person is known, we are able to calculate the total number of employed persons and their total monthly wage bill for each firm in these statistics. Finally, we observe the education level and the occupation of each person, so we are able to disaggregate these measures by education and occupation in order to examine the employment structure at the firm level.

The wage bill is divided into three education groups (low, intermediate and high) and into three occupation groups (abstract, routine and service occupations). We have also constructed similar shares for hours worked and employed persons, but the results for these are similar to the wage bill, so we report only those. The low education group consists of those with basic compulsory education only. The high education group consists of those with a university level bachelor’s degree or more. The intermediate group consists of all degrees in between these, i.e. from vocational to non-university higher degrees that usually involve two to four years of education. Our occupational grouping is an application of the classification presented in Acemoglu and Autor (2010) to the Finnish ISCO occupations. The abstract group includes managers, professionals and technicians; the routine group includes occupations for sales, clerical, production and operator’s work; and services include occupations involving work in protection, food preparation, building and grounds, cleaning and personal care and services.

Using this data we first detect the pattern of employment polarization in the Finnish private sector.[3] As Figure 1 shows the changes in the employment shares by initial occupational wage deciles has been U-shaped in both periods 1995-2001 and 2002-

[3] Earlier Finnish evidence on polarization at the aggregate level is provided in Asplund et al. (2011), Mitrunen (2013) and Böckerman et al. (2013).

2008, similar to the polarization pattern documented for the UK in Goos and Manning (2007). On the other hand, we find no indication of wage polarization in Finland in Böckerman et al. (2013), where wage growth increases almost linearly with the initial wage level.

Figure 1 here

We augment the wage and employment data with the firm-level variables for the adoption of Information and Communication Technology. Our data on the use of information technology and electronic commerce in firms originates from the Statistics Finland survey “Use of Information Technology in Enterprises” (ICT survey). The survey is a stratified random sample of firms in the sampling frame of Business Register of Statistics Finland. It covers all large firms (100 employees or more) and a random sample of smaller firms with more than 5 employees. Because also the HWSS data is a sample for smaller firms, matching of these two data reduces the number of firms in the linked data substantially. Therefore, our results are representative only for the larger firms. The variables used from this survey describe the usage of ICT in firms, for example, the Internet, intranet, broadband, home pages, services offered via home pages, electronic commerce, and electronic data interchange (EDI). The full list and explanations for variables is provided in Appendix 1.

The variables that describe various aspects of ICT are highly correlated as they measure the underlying characteristics of firms that affect the adoption of new technologies. We use factor analysis to compress this information into latent factors, which we use as explanatory variables in our regressions. This alleviates multicollinearity and variance inflation in the estimated models, because the factors are orthogonal. We use principal factors method and based on the eigenvalues three factors are adequate to describe the common variance of the ICT indicators. The cumulative variance explained is 71%. Factor loadings are presented in Appendix 2. We call Factor 1 EDI as it loads on variables related to the usage of electronic data interchange by the firm for various purposes (sending and receiving invoices or orders, or sending transport documents). Factor 2 loads on a large number of variables related to broadband or mobile access to Internet, the firm having a web site, and that the firm orders or sells through the computer networks. This factor also loads on the firm having an enterprise resource planning (ERP), but we call this the Internet factor for short. The third factor loads on two variables indicating whether the firm shares supply chain management (SCM) data with suppliers or customers.

Furthermore, we match task input measures at the occupation level from Goos, Manning and Salomons (GMS) (2010) into the wage data. They provide measures for the abstract, routine and service intensity for each 2-digit occupation. The measures are derived from the 2006 version of the Occupational Information Network (ONET) database, which provides the occupational attributes and characteristics of workers in 812 US SOC (Standard Occupational Classification) occupations. GMS (2010) manually convert these to the International Standard Classification of Occupations (ISCO), so we can match the GMS (2010) measures to our data at the 2-digit ISCO level. GMS (2010) use 96 ONET variables related to worker characteristics, worker requirements and work activities to create their measures for the Abstract, Routine and Service task intensities of different occupations. The task information is gathered from job incumbents, occupational analysts and occupational experts, who evaluate how important these task variables are in each occupation on a scale from 1 (not important at all) to 5 (extremely important). The 96 ONET variables are divided into one of three groups of Abstract, Routine and Service tasks. Abstract task variables measure things like critical thinking and complex problem solving. Routine task variables measure things like manual dexterity, finger dexterity and operation monitoring. Service task variables measure assisting and caring for others, service orientation, and establishing and maintaining interpersonal relationships. The actual task measures are averages of these variables for each SOC occupation. They are converted to an ISCO occupation, using US employment in SOC cells as weights. Each task measure is normalized to have zero mean and unit standard deviation and they are available at the 2-digit ISCO level from GMS (2010).

Figure 2 documents the development over time of the task importance variables as weighted means across 3-digit occupations. The decline of Routine importance shows that the employment in routine intensive occupations has decreased steadily over the years. Increasing Abstract and Service importance correspondingly imply increasing employment in abstract and service intensive occupations. Defining Routine intensity as the ratio of Routine importance to the sum of Abstract and Service Importance as in

Goos et al. (2011), these trends imply a decreasing trend in Routine intensity. The pattern is consistent with the Routinization hypothesis of ALM (2003).

Figure 2 here

3. Decompositions for employment and wage bill shares

The main variables of interest in our examinations are the shares in the total wage bill by educational and occupational groups at the firm level, and their changes over time. To obtain preliminary information about the possible sources of the changes in the employment structure, we present firm-level decompositions for the changes in wage bill shares.[4] This decomposition augments the Berman, Bound and Griliches (1994) industry-level decomposition to an unbalanced panel of firms with entry and exit.[5] The aggregate change in the wage bill share of a worker group defined by education or occupation (indexed by g) can be decomposed as follows:

$$\Delta P^A = \sum_i \Delta S_i \bar{P}_i + \sum_i \Delta P_i \bar{S}_i + w_t^N \left(P_t^N - P_t^S\right) + w_{t-s}^D \left(P_{t-s}^S - P_{t-s}^D\right)$$

where $P = \dfrac{E^g}{E}$, $P_i = \dfrac{E_i^g}{E_i}$, $S_i = \dfrac{E_i}{E}$, $w_t^N = \dfrac{E_t^N}{E_t^A}$ and $w_{t-s}^D = \dfrac{E_{t-s}^D}{E_{t-s}^A}$.

P is the aggregate share of the skill group in total wage bill (denoted by E), P_i is the corresponding share in firm *i* (*i* = 1,...,N), S_i is the share of firm *i* in aggregate wage bill, Δ indicates change over the period (t-s, t), and bar an average over the period's initial (t-s) and final year (t) values. Superscripts indicate the sums or shares for all firms (A), surviving firms (S), entering firms (N) and exiting firms (D). It can be shown that the entry and exit effects can also be written as

[4] We have also performed the decompositions for employment shares and working hour's shares, but the results are essentially similar to those for wage bill shares that we report.
[5] See Vainiomäki (1999) for a detailed derivation of this decomposition augmented to include entry and exit effects.

$$ENTRY = w_t^N \left(P_t^N - P_t^S\right) = \left(P_t^A - P_t^S\right)$$

$$EXIT = w_{t-s}^D \left(P_{t-s}^S - P_{t-s}^D\right) = \left(P_{t-s}^S - P_{t-s}^A\right).$$

These effects therefore depend on the deviation of the entering and exiting plant's average skill group shares from that of continuing plants. The entry effect is positive and greater the *higher the group's share in new plants* is compared to continuing plants ($P_t^N \geq P_t^S$). Similarly, the exit effect is positive and greater the *lower the group's share is in exiting plants* compared to continuing plants ($P_{t-s}^S \geq P_{t-s}^D$). But it is noteworthy that the entry effect is also given by the simple difference between the group's aggregate wage bill share for all firms and continuing firms in the *final* year of the period. Similarly, the exit share is given by the simple difference in the shares for continuing firms and exiting firms in the *initial* year of the period.

The other two terms are standard from industry-level decompositions. The first sum is the *between firms effect*, which captures the shifts of employment (wage bill) between firms with different average wage bill shares. It is positive if the wage bill shifts towards firms which have a high wage bill share of the skill group in question. The second sum is the *within firms effect*, which captures changes in the wage bill share within each firm, weighted by the firm's average share of the total wage bill. The within component captures technological change within firms, the between component captures the product demand changes across firms, and the entry/exit components reflect the demographic changes in firm population. The education groups in our decompositions are Basic, Intermediate and High, as explained previously. The occupation groups are Abstract, Routine, and Services, following Acemoglu and Autor (2010).

Table 1 reports the decomposition of changes in the wage bill shares by the education groups for the period 2002-2008.[6] The within and total changes for the low and intermediate education groups are negative for this period. The respective changes for

[6] We present these decompositions here as background for the regressions below only for the 2002-2008 period, because the ICT variables in regressions are available only for the 2000's.

the highest educated are large positive. The entry component for the basic education group and the exit component for the highest educated are also positive. The between components for all education groups are minimal. These changes imply a rapid skill upgrading at the highest education level during the 2000's. This overwhelmingly occurs within firms. The patterns are loosely consistent with polarization in the sense that the intermediate education group loses shares, but in general these results show that the development has been "linear" with respect to education. Largest decline in shares occurs for the lowest educated and largest increase for the highest educated.

Table 1 here

Table 2 shows the decompositions of change in the wage bill shares by the occupation groups. In contrast to education, both within and between components are important for occupational changes and affect in the same direction, except for the service occupations. We find that in total the routine occupation share declines and the abstract and service occupation shares increase, so that the total change is clearly consistent both with the routinization hypothesis and with job polarization. The entry and exit effects are small in general, but the entering firms are mildly service intensive and less intensive in abstract occupations. The shifts in production between different firms, the between component, seems to be more important in explaining the polarization in the occupational shares than in the educational shares. The shifts in production towards service-intensive firms and away from the routine intensive firms contribute clearly to the polarized pattern of total change. This suggests that changes in product demand may have a role in explaining the increase in the service occupations. However, for the abstract and routine occupations the overwhelming majority of change occurs within existing firms, which is consistent with technological change being important in explaining the declining shares of routine occupations.

Table 2 here

4. Specifications and results from firm-level regressions

In order to examine the importance of the technology explanation for the shifts in the structure of labour demand we estimate, at the firm level, equations for the wage bill shares of the education groups (g=Low, Middle, High), and occupation groups (g=Abstract, Routine, Service), as follows:

$$\Delta SHR_{it}^{g} = c^{g} + \beta_1^{g} ICT_{it} + \beta_2^{g} \ln\mathrm{SIZE}_{it} + \beta_3^{g} X_{it} + u_{it}^{g}$$

where ICT denotes the adoption of new technology at each firm, lnSIZE is the log of firm size (employment), X denotes a vector of other control variables (two-digit industry indicators in all models, and the firm-level average education and age in some models). Berman, Bound and Griliches (1994) showed that this type of share equation can be derived from a short-run trans-log cost function to examine the relative demand for different labour groups. However, this specification omits capital intensity and relative wages that we control with use of two-digit industry indicators. They also account for all permanent differences between industries. Michaels, Natraj and Van Reenen (2010) derive similar equations from a three-input CES production function, which allows for ICT capital to substitute for the medium educated, and to complement for the highly educated. The polarization hypothesis implies that following the adoption of ICT (increase in ICT capital) the wage bill share of the Highly educated (skilled) workers increases ($\beta_1^{High} > 0$) and the share of the Middle educated (skilled) declines ($\beta_1^{Middle} < 0$), and the change in the share of lowest educated is ambiguous. We perform a factor analysis on a large number of ICT indicators at the firm level to create underlying ICT factors, which we use as explanatory variables in our regressions instead of ICT capital that is not available at the firm level.

We also perform similar regressions for occupational groups, i.e. Abstract, Routine, Service occupations. Analogously with the treatment of education, the polarization hypothesis now implies that technological change increases the demand and therefore the wage bill share of the Abstract occupations ($\beta_1^{Abstract} > 0$) and reduces the share of the Routine occupations ($\beta_1^{Routine} < 0$) and has an ambiguous effect on the Service occupations ($\beta_1^{Service} = ?$). The ambiguity for services is based on the assumption that service occupations are technologically independent of the ICT, and the demand effects are general equilibrium effects, which are not manifest at the firm level in those exact firms that adopt new technologies. The decompositions above on the other hand incorporate the demand effects also in all other firms. They indicated that the shifts in production between firms and the entry of new service intensive firms do play a role in explaining the rise in service occupations.

We return to possible endogeneity issues in this specification below, but start here with OLS results for educational shares in Table 3. Columns (1) include ICT factors, lagged level of the dependent (to control for the regression-towards-mean phenomenon), firm size and two-digit industries as explanatory variables for the change in wage bill share of each education group. Columns (2) add the average age of the firm's total labour force (and the average education years in similar regression for occupation shares in Table 4). Firm size controls for the fact that the adoption of ICT technologies correlates positively with the size of the firms. To the extent that firm size itself or omitted factors correlated with size (e.g. capital intensity or the quality of firm management) have effects on the structure of labour demand, omitting firm size would obscure the results for ICT factors.[7] Our results are based on comparing similar sized "ICT and non-ICT firms" (different levels of ICT factors). Similarly, we include the average age and education of employees as controls for possibly differential effects of SBTC on wage growth and employment development of different education and occupation groups. Our first-differenced equations already control for all firm fixed effects, but here we attempt to control for possibly differential effects of SBTC on wage growth (e.g. increasing skill premiums) or employment growth by education or occupation.

For education groups in Table 3 we find that the additional control for the average age of employees affects the results only marginally. We omit the average education years as an additional control for educational shares as there are clear problems of reverse correlation if it is included in educational regressions. In both columns (1) and (2) the

[7] Our results (not reported) are much less robust and less significant if firm size is not included in the equation.

pattern of results is supportive of SBTC, but not the routinization hypothesis. All the ICT factors are positively associated with the increase of high skilled wage bill and negatively associated with the share of basic education group. The regression coefficients are both statistically and numerically significant. We find that the wage bill share of the highly educated increases 1.3 %-points faster in firms that have one standard deviation (0.43) higher score for the EDI factor, 1.7 %-points faster for the Internet factor (s.d. 0.30), and 0.9 %-points faster for the SCM factor (s.d. 0.45). These effects are economically large compared to the total increase of 7.4% of the high-educated wage bill share over the 2002-2008 period as reported in Table 1. The coefficients for both intermediate and basic educated groups are respectively negative, but only the coefficients for the basic group are significant. The fact that the intermediate educated are not affected significantly negatively by ICT factors is at odds with the predictions from job polarization. It should be noted that the association between firm’s ICT factors and its employment structure prevails across firms within two-digit industries, as the industry effects are controlled in all our equations.

Table 3 here

Similar OLS results for occupational shares are reported in Table 4. The results in columns (1), without additional firm controls, are supportive of the routinization hypothesis related to most ICT factors. The factors for EDI and Internet are positively and significantly related to the change in the share of Abstract occupations, and negatively and significantly related to the routine share. The magnitude of the coefficients is again economically significant, compared to the mean of the dependent variable, or to the total change in shares in Table 2. A one standard deviation increase in the EDI factor increases the abstract occupation share by 1.5 %-points and reduces the routine occupation share by 1.1 %-points. Corresponding effects for the Internet factor are 1.8%-points and –1.2 %-points using the coefficients in column (2). Service share is independent of ICT at the firm level, which is consistent with the routinization hypothesis. The pattern is similar for the SCM factor with marginally (in)significant coefficients at the 10% level for the abstract and routine shares, but adding firm-level average age and education turns all SCM coefficients insignificant. Also EDI and Internet factors turn insignificant in the routine equation, but retain significance in the abstract share equation with these additional controls. The (unreported) coefficients for average education years are highly significant in these regressions. It is likely that average education suffers from reverse correlation problems also in these occupational equations, as increasing employment share of abstract occupations increases the average education of firm's employees. Excluding it, but retaining the average age as a regressor, yields results similar to those in columns (1).

Table 4 here

As already noted above, our estimating equations are essentially first-differenced versions of the levels equations for wage bill shares, so any endogeneity related to the unobserved firm fixed effects in the levels equations is eliminated from our results. Other sources of endogeneity bias, however, remain in our models. First, measurement error in explanatory variables causes the standard attenuation bias. Second, the lagged level of wage bill correlates by definition with the its change. Third, there is the possibility of reverse causality in ICT, i.e. shocks to the firm's wage bill shares causing firms to change investments in new technology (adoption of ICT). Regarding the lagged level of the dependent variable, we use the corresponding wage bill share as well as the values of the firm's average age and education years lagged one more year as instruments (i.e. year 2001 values as instruments for year 2002 values). As for the ICT it is difficult to find firm-level instruments for these, but we have performed estimations with potential instruments that we have access to, for example firm-level lagged values of task intensities reflecting firms potential to replace routine tasks with ICT or workers performing analytical tasks benefiting from ICT. However, these firm-level instruments proved not valid (i.e. endogenous). Therefore, we report only results from using three-digit industry values for these task intensities as instruments for firm-level ICT factors. Industry-level task intensities should be (more or less) exogenous with respect to idiosyncratic firm-level shocks to the wage bill.

The results from our instrumental variables estimations are presented in Tables 5 and 6 for the education groups and the occupation groups respectively. In columns (1) only the lagged level of the wage bill share is instrumented, while in columns (2) also the ICT factors are instrumented. The importance of the lagged level of the dependent variable in all models confirms regression-towards-mean phenomenon, but it also implies that the educational employment structure changes quite slowly even over the six-year period. The coefficients for the lagged dependent variable estimate (ρ^6-1), where ρ is the AR(1) parameter for the level of wage bill share. The IV estimates in column (1) of Table 5 imply that ρ^6 is about 0.8 for the highly and medium educated, and about 0.6 for the low education group, which imply high persistence in educational wage bill shares at the firm level.

Table 5 here

For the educational groups the coefficients for ICT factors remain very similar to the OLS results in columns (1). Although the coefficients are somewhat smaller than in the OLS equations, they are still economically large in magnitude. However, when ICT factors are also instrumented in columns (2), their coefficients increase in absolute value compared to OLS results as expected due to measurement error bias. But the coefficients obtain in general implausibly large values and only one coefficient remains significant. The standard errors of IV estimates are considerably inflated compared to the OLS standard errors. The IV results clearly suffer from the weak instrument problems which is confirmed by the low F-tests for the additional instruments in the first-stage regressions for the ICT factors. On the other hand, the instruments for the lagged level of wage bill have strong explanatory power in the first stage. The overidentifying restriction test passes in columns (1) clearly for the high education group, but only at 10 % for the basic education group. In the medium education equation the test is strongly significant which means that the instrument set includes some variables which are not valid instruments (i.e. not exogenous).

Table 6 here

The results for the occupational groups in Table 6 similarly retain the significance of ICT factors in columns (1), when only the lagged level of the dependent variable is instrumented, but suffer from weak instruments in columns (2), when also ICT factors are instrumented. In columns (1) the instruments pass the overidentifying restriction test for the abstract and routine occupation groups, but not for the service occupations.

In summary, our IV results confirm the high persistence in educational and occupational employment structure even when the lagged level of the dependent variable is instrumented to account for the reverse correlation and measurement error. But our attempts to instrument ICT factors are unsuccessful and suffer from problems in finding proper instruments for them. Therefore, we can only establish an association between ICT factors and wage bill shares, not causality.

5. Conclusions

Using the new Harmonized Wage Structure Statistics (HWSS) data of Statistics Finland, we first document the patterns of employment polarization in the Finnish private sector labour market. Our results show that there has been considerable employment polarization at the aggregate level. The structure of changes in the employment shares by initial occupational wage deciles is clearly U-shaped. Matching task measure for individuals at the occupational level, we find that the changes in aggregate occupational structure are consistent with the routinization hypothesis, i.e. aggregate average routine intensity of employment declines in a trend-like manner over the period 1995-2008.

The new feature of our paper is, however, to perform decompositions and regression analyses that test for the routinization hypothesis and job polarization at the firm level, instead of the aggregate, industry or local level as in prior studies. Using firm-level approach, we are able to study routinization and job polarization at the micro level, where the actual labour demand decisions are made.

Our firm-level decompositions for the changes in wage bill shares provide some indication of likely reasons behind these changes. The decompositions for educational groups indicate that changes in education shares are towards more educated groups in a “linear” fashion with respect to education level, which is consistent with standard SBTC model. The total change occurs overwhelmingly within firms, which is suggestive of a technological cause for these changes. As for changes in occupational shares, we find that the increase in abstract and the decline in routine occupations also occurs substantively within existing firms. However, our decompositions also show that for the service occupations the shifts in production towards service-intensive firms and the entry of new service-intensive firms is relatively more important than for other occupation groups. This shows that the changes in product demand may have a role in explaining the increase in service occupations, which produces the polarized pattern in the total changes for occupation groups, that is increasing abstract and service shares and declining routine share.

Furthermore, we examined the technology-based explanations for routinization and job polarization at the firm level by using firm-level indicators for ICT usage as explanatory variables in the firm-level regressions. We first perform a factor analysis on a large number of indicators for ICT adoption at the firm level to obtain factor scores for three ICT factors. We then use these factor scores as explanatory variables in regressions for changes in wage bill shares of different education and occupation groups. OLS regressions show that these ICT factors are associated with increases in the demand for high educated and reductions in the demand for low educated, while the intermediate education group is independent of ICT. In regressions for occupation groups, we find that ICT factors are associated with increases in abstract occupation shares and declines in routine occupation shares. These occupational patterns support the routinization hypothesis at the firm level. Since routinization is the main mechanism producing polarization, these results are also consistent with job polarization. The service occupation share is independent of ICT at the micro level (as hypothesised), and the increasing aggregate share of services relates to the demand effects, as noted above. A possible reason for the apparent contradiction in our results with respect to educational and occupational employment structure is that the assumed substitutability of middle educated in general for ICT technologies does not hold, but it does prevail for occupations classified as routine in our analysis.

Finally, our attempts to establish causality of these associations using instrumental variable (IV) methods proved unsuccessful. We find strong persistence in educational and occupational shares over the six-year period 2002-2008 studied, even when the lagged level of wage bill share is instrumented in the first differenced equations. In these IV regressions the pattern and significance of ICT factors remain intact. But when we attempt to account for possible endogeneity of ICT factors using either firm-level task measures as instruments, or corresponding three-digit industry instruments, we stumble upon weak instrument problems.

References

Acemoglu, D. (2002). 'Technical change, inequality, and the labor market'. *Journal of Economic Literature*, 40: 7-72.

Acemoglu, D. and Autor D. (2010). 'Skills, tasks and technologies: Implications for employment and earnings'. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics,* Volume 4B. Amsterdam: North-Holland, pp. 1043-1171.

Asplund, R., Barth E., Lundborg, P. and Nilsen, K. M. (2011). 'Polarization of the Nordic labour markets'. *Finnish Economic Papers*, 24: 87-110.

Autor, D. and Dorn, D. (2013). 'The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market'. *The American Economic Review,* 103:1553-1597.

Autor, D., Dorn, D. and Hanson G.H. (2013). 'The Geography of Trade and Technology Shocks in the United States'. *The American Economic Review,* 103:220-225.

Autor, D., Dorn, D. and Hanson G.H. (2013b). 'The China Syndrome: Local Labor Market Effects of Import Competition in the United States'. *The American Economic Review,* 103: 2121-2168.

Autor, D., Levy, F. and Murnane, R. J. (2003). 'The skill content of recent technological change: An empirical exploration'. *Quarterly Journal of Economics,* 118: 1279-1333.

Autor, D., Katz, L. and Kearney, M. (2006). 'The polarization of the U.S. labor market'. National Bureau of Economic Research, Working Paper No. 11986.

Autor, D., Katz, L. and Kearney, M. (2008). 'Trends in U.S. wage inequality: Re-assessing the revisionists'. *Review of Economics and Statistics,* 90: 300-323.

Berman, E., Bound, J. and Griliches, Z. (1994). ‘Changes in the demand for skilled labor within US manufacturing: Evidence from the annual survey of manufactures’. *Quarterly Journal of Economics*, 109: 367-397.

Bloom, N., Draka, M. and Van Reenen, J. (2011). ‘Trade Induced technical change? The impact of Chinese imports on innovation, IT and productivity’. Centre for Economic Performance, Discussion Paper No. 1000.

Bound, J. and Johnson, G. (1992). ‘Changes in the structure of wages in the 1980s: An evaluation of alternative explanations’. *The American Economic Review,* 83: 371-392.

Böckerman, P., Laaksonen, S, and Vainiomäki, J. (2013). ‘Is There Job Polarization at the Firm Level?’. Tampere Economic Working Papers, No. 91, October 2013.

Card, D. and DiNardo J. E. (2002). ‘Skill-biased technological change and rising wage inequality: Some problems and puzzles’. *Journal of Labor Economics,* 20: 733-782.

Goos, M. and Manning A. (2007). ‘Lousy and lovely jobs: The rising polarization of work in Britain’. *Review of Economics and Statistics,* 89: 118-133.

Goos, M., Manning, A. and Salomons, A. (2010). ‘Explaining job polarization in Europe: The roles of technology and globalization’. Centre for Economic Performance, Discussion Paper No. 1026.

Jenkins, S. P. (1995). ‘Did the middle class shrink during the 1980s? UK evidence from kernel density estimates’. *Economics Letters,* 49: 407-413.

Michaels, G., Natraj, A. and Van Reenen, J. (2010). ‘Has ICT polarized skill demand? Evidence from eleven countries over 25 years’. Centre for Economic Performance, Discussion Paper No. 987.

Mitrunen, M. (2013). Työmarkkinoiden polarisaatio Suomessa. Government Institute for Economic Research, VATT Mimeo 33.

Vainiomäki, J. (1999). 'Technology and skill upgrading: Results from linked worker-plant data for Finnish manufacturing'. In J. Haltiwanger, J. Lane, J. Spletzer, J. Theeuwes and K. Troske (eds.), *The Creation and Analysis of Employer-Employee Matched Data.* Amsterdam: Elsevier/North-Holland, pp. 115-145.

FIGURE 1

Employment Polarization.

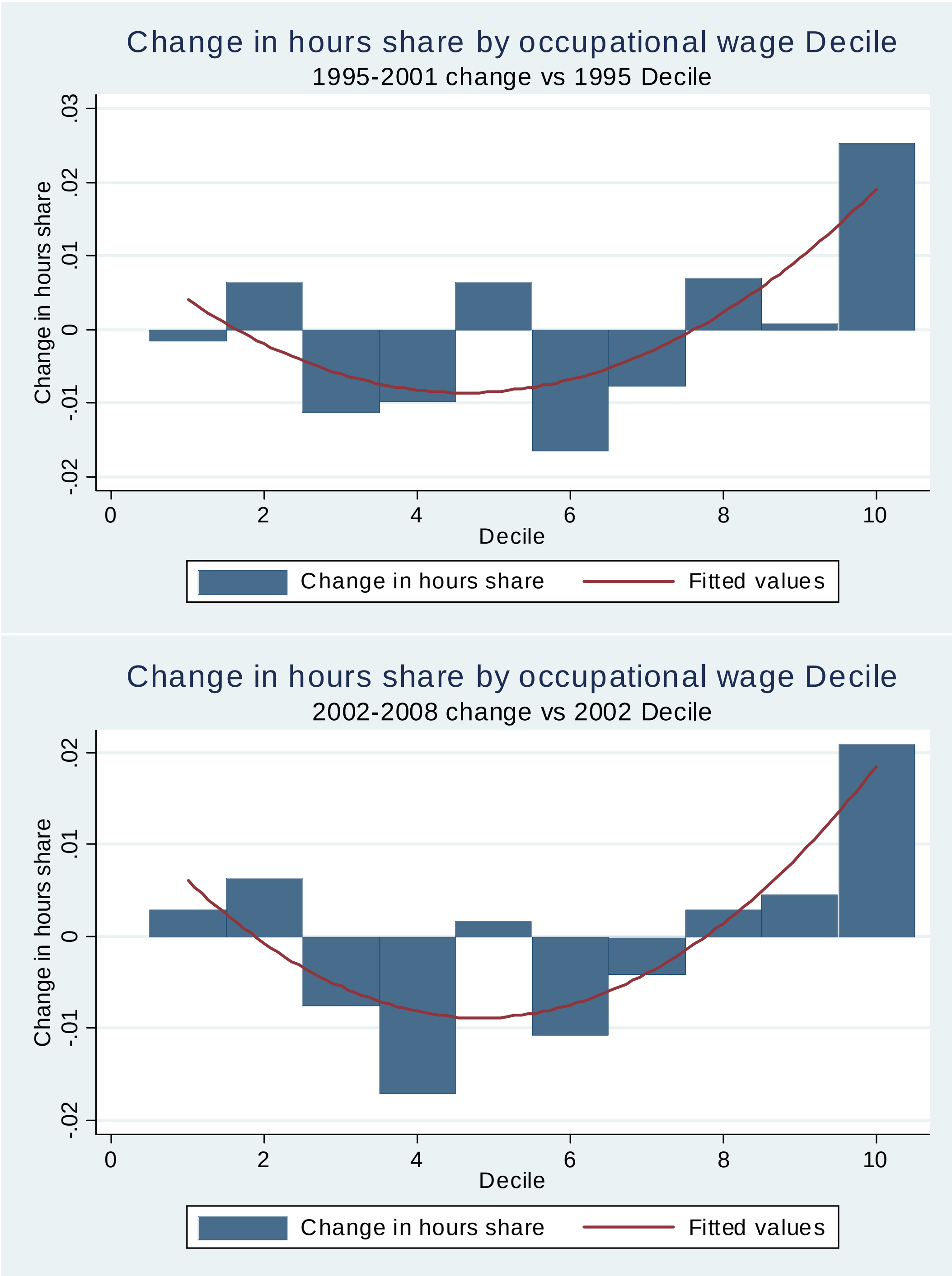
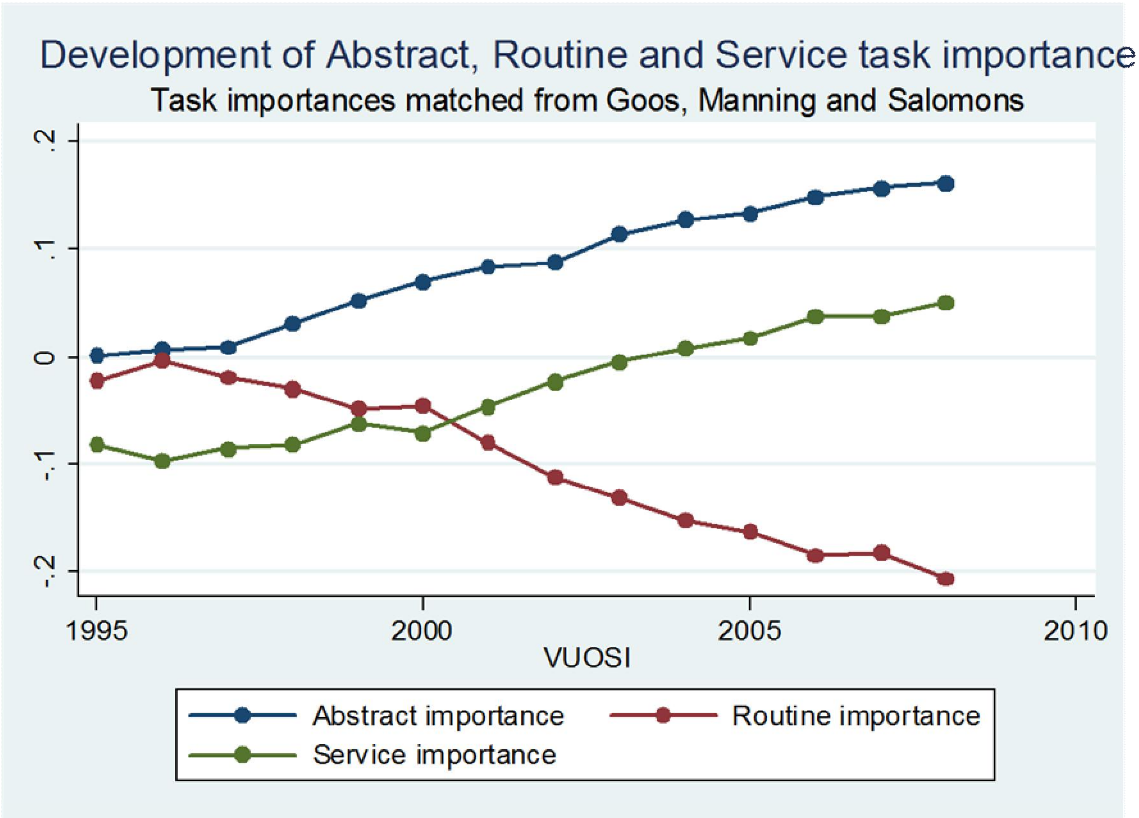


FIGURE 2.

Routinization Hypothesis.



Note: Hours weighted means of task variables across 3-digit occupations.

TABLE 1

Decompositions for Wage Bill Share by Education

Period	Education	Within	Between	Entry	Exit	Total	Share (2008)
2002-2008	Basic	-0.051	-0.001	0.012	-0.006	-0.046	0.138
2002-2008	Intermediate	-0.018	0.001	-0.004	-0.007	-0.028	0.571
2002-2008	High	0.068	0.000	-0.007	0.013	0.074	0.291

TABLE 2

Decompositions for Wage Bill Share by Occupation

Period	Occupation	Within	Between	Entry	Exit	Total	Share (2008)
2002-2008	Abstract	0.045	0.004	-0.004	0.000	0.045	0.464
2002-2008	Routine	-0.038	-0.019	-0.001	0.001	-0.057	0.450
2002-2008	Service	-0.007	0.015	0.005	-0.001	0.012	0.086

TABLE 3

OLS Regressions for the Change in Wage Bill Shares by Education Group

Variable	High		Medium		Basic	
ICT Factors	(1)	(2)	(1)	(2)	(1)	(2)
EDI	.031 (2.41)	.031 (2.41)	-.007 (-0.61)	-.007 (-0.58)	-.026 (-3.88)	-.027 (-3.95)
Internet	.056 (3.28)	.056 (3.28)	-.027 (-1.20)	-.027 (-1.22)	-.035 (-2.59)	-.035 (-2.61)
SCM	.019 (1.91)	.0194 (1.91)	-.012 (-0.98)	-.013 (-1.02)	-.014 (-2.24)	-.014 (-2.16)
Lagged dependent	-.233 (-5.80)	-.233 (-5.80)	-.306 (-5.88)	-.306 (-5.82)	-.436 (-8.54)	-.436 (-8.42)
lnSize	-.006 (-2.19)	-.006 (-2.18)	.004 (1.08)	.004 (1.12)	.005 (2.31)	.005 (2.33)
Other firm controls	No	Yes	No	Yes	No	Yes
Industry indicators	Yes	Yes	Yes	Yes	Yes	Yes
N	1543	1543	1543	1543	1543	1543
R Squared	0.245	0.245	0.301	0.304	0.514	0.521
F-test (industry) F(39, 1498)	63.49 (0.000)	45.47 (0.000)	8.67 (0.000)	8.77 (0.000)	16.13 (0.000)	13.93 (0.000)
Dependent Mean (weighted)	0.048		0.001		-0.049	

Notes: Weighted by firm size (employment). Robust t-values reported.
The dependent variable is a six-year difference over period 2002-2008.
Other firm controls in columns (2) include the average age of firm’s all employees.

TABLE 4

OLS Regressions for the Change in Wage Bill Shares by Occupation Group

Variable	Abstract		Routine		Service	
ICT Factors	(1)	(2)	(1)	(2)	(1)	(2)
EDI	0.056 (3.05)	.036 (2.72)	-0.046 (-2.65)	-.025 (-1.55)	-0.012 (-1.49)	-.010 (-1.34)
Internet	0.090 (3.03)	.059 (2.46)	-0.093 (-3.03)	-.040 (-1.45)	-0.003 (-0.24)	-.001 (-0.08)
SCM	0.026 (1.62)	.008 (0.62)	-0.029 (-1.69)	-.001 (-0.09)	-0.001 (-0.14)	.001 (0.22)
Lagged dependent	-0.232 (-4.91)	-.547 (-11.7)	-0.253 (-5.69)	-.458 (-8.92)	-0.278 (-5.38)	-.302 (-5.86)
lnSize	-0.016 (-3.48)	-.011 (-2.82)	0.014 (3.06)	.004 (1.01)	0.003 (1.93)	.003 (1.80)
Other firm controls	No	Yes	No	Yes	No	Yes
Industry indicators	Yes	Yes	Yes	Yes	Yes	Yes
N	1543	1543	1543	1543	1543	1543
R Squared	0.208	0.489	0.218	0.428	0.560	0.575
F-test (industry) F(39, 1498)	31.50 (0.000)	36.96 (0.000)	23.02 (0.000)	40.06 (0.000)	7.55 (0.000)	7.49 (0.000)
Dependent Mean (weighted)	0.024		-0.021		-0.004	

Notes: Weighted by firm size (employment). Robust t-values reported.
The dependent variable is a six-year difference over period 2002-2008.
Other firm controls in columns (2) include the average education years and age of firm’s all employees.

TABLE 5.

IV Regressions for the Change in Wage Bill Shares by Education Group

Variable	High		Medium		Basic	
ICT Factors	(1)	(2)	(1)	(2)	(1)	(2)
EDI	.034 (2.86)	.363 (1.19)	-.015 (-1.32)	-.326 (-1.52)	-.0244 (-3.68)	-.571 (-2.09)
Internet	.048 (2.85)	-.326 (-1.20)	-.012 (-0.52)	.079 (0.34)	-.036 (-2.46)	.294 (1.12)
SCM	.019 (1.87)	-.059 (-0.32)	-.004 (-0.31)	.067 (0.46)	-.016 (-2.16)	-.047 (-0.36)
Lagged dependent	-.220 (-4.68)	-.236 (-3.27)	-.237 (-3.93)	-.215 (-2.73)	-.386 (-5.74)	-.482 (-5.33)
lnSize	-.005 (-1.54)	-.008 (-0.26)	.001 (0.36)	.010 (0.43)	.004 (2.01)	.032 (1.61)
Other firm controls	No	No	No	No	No	No
Industry indicators	Yes	Yes	Yes	Yes	Yes	Yes
N	1317	1317	1317	1317	1317	1317
First stage F-tests: Lagged dependent EDI Internet SCM	 635.0	 405.2 3.68 4.62 4.36	 534.6	 275.7 3.69 4.36 4.55	 598.2	 294.7 3.61 4.25 4.70
Overidentifying restrictions	2.22 (0.33)	14.81 (0.005)	13.94 (0.001)	15.82 (0.003)	5.16 (0.08)	4.73 (0.32)
Dependent Mean (weighted)	0.048		0.001		-0.049	

Notes: Weighted by firm size (employment). Robust t-values reported.
The dependent variable is a six-year difference over period 2002-2008.
In columns (1) the level of lagged dependent variable in 2002 is instrumented with its value in previous year (2001) and the firm-level averages of education years and age in 2001. In columns (2) also the ICT factors are treated as endogenous, using three-digit industry values for task importance variables and for industry average education years and age as additional instruments in addition to those in column (1).

TABLE 6

IV Regressions for the Change in Wage Bill Shares by Occupation Group

Variable	Abstract		Routine		Service	
ICT Factors	(1)	(2)	(1)	(2)	(1)	(2)
EDI	.056 (3.14)	1.59 (3.60)	-.049 (-2.91)	-1.27 (-1.93)	-.012 (-1.37)	-.384 (-3.20)
Internet	.088 (2.81)	.397 (0.73)	-.090 (-2.66)	.219 (0.33)	-.005 (-0.40)	.226 (1.40)
SCM	.030 (1.79)	.516 (1.46)	-.026 (-1.50)	-.102 (-0.30)	-.005 (-1.05)	-.005 (-0.07)
Lagged dependent (level)	-.183 (-3.33)	-.292 (-2.78)	-.217 (-4.62)	-.268 (-3.33)	-.274 (-4.27)	-.078 (-0.46)
lnSize	-.016 (-3.31)	-.156 (-2.76)	.013 (2.59)	.083 (2.50)	.005 (2.85)	.020 (1.63)
Other firm controls	No	No	No	No	No	No
Industry indicators	Yes	Yes	Yes	Yes	Yes	Yes
N	1317	1317	1317	1317	1317	1317
First stage F-tests: Lagged dependent EDI Internet SCM	 478.4	 623.5 3.37 4.23 4.47	 557.6	 856.5 3.36 4.87 4.19	 17.4	 9.33 3.36 6.61 4.34
Overidentifying restrictions	1.35 (0.51)	3.01 (0.56)	1.33 (0.52)	2.75 (0.60)	16.5 (0.00)	4.78 (0.31)
Dependent Mean (weighted)	0.024		-0.021		-0.004	

Notes: Weighted by firm size (employment). Robust t-values reported.
The dependent variable is a six-year difference over period 2002-2008.
In columns (1) the level of lagged dependent variable in 2002 is instrumented with its value in previous year (2001) and the firm level averages of education years and age in 2001. In columns two also the ICT factors are treated as endogenous, using three-digit industry values for task importance variables and for industry average education years and age as additional instruments in addition to those in column (1).

Appendix 1.

Vaariables and definitions for terms used in the ICT survey (Source: Statistics Finland)

Variables

PC	Firm uses computers
INTER	Firm has Internet connection
WEB	Firm has a website
EPURCH	Firm orders through computer networks (websites or EDI)
ESALES	Firm sells through computer networks (websites or EDI; not email orders)
BROAD	Firm has broadband (ADSL, SDSL, cable modem; faster than ISDN)
MOB	Firm has mobile access to internet (laptop, mobile phone; 3G, 4G or slower)
ERP	Firm has Enterprise Resource Planning system (ERP program); 2006 onwards
SCMT	Sharing Supply Chain Management (SCM) data with suppliers using computer networks (inc. Internet); regular exchange of information on demand forecasts, inventories, production plans, delivery progress, product planning; available 2006-2009
SCMA	Sharing SCM data with customers; see above; available 2006-2009
CRMINF	Management and sharing of customer information with other business functions within firm; 2006 onwards
CRMANA	Firm analyses customer information for marketing purposes (price setting, promoting sales, choosing delivery channels); 2006 onwards
AUTTIED	Electronic data interchange used
AUTLASVA	Receiving e-invoices using electronic data interchange; available 2007-2009
AUTLASLA	Sending e-invoices using electronic data interchange; available 2007-2009
AUTKULJ	Using electronic data interchange in sending transport documents; available 2007-2009

Definitions of terms

Broadband

Broadband means a telecommunications connection with a capacity of at least 256 Kbps. In the statistics on the use of information technology in enterprises, broadband has in practice been defined through the type of technology used in the connection as either DSL (e.g. ADSL) or other broadband connection (faster than a traditional telephone modem or ISDN).

E-invoice

An e-invoice is an electronic invoice constructed according to a generally used message format, whose data can be handled and interpreted automatically. E-invoices are transmitted via a telecommunications service provider or a bank. E.g, Finvoice, eInvoice, TEAPPSXML, PostiXML.

E-mail invoice

An e-mail invoice is an invoice sent as a pdf-file attached to an e-mail.

EDI

EDI (Electronic Data Interchange) is a procedure by which information located in an enterprise’s data system is used to produce a specified data flow that is transmitted electronically to a receiving enterprise, where it is directly incorporated into the data system (e.g. order, payment order for invoice, price list or product catalogue).

EDI commerce

EDI commerce is electronic commerce that takes place between enterprises through the medium of EDI.

EDI invoice

An EDI invoice is an electronic invoice in machine code according to the EDI structure standards. EDI invoices are often sent via a telecommunications service provider.

Electronic invoice

An electronic invoice is an invoice transmitted in electronic form: an EDI invoice, an e-invoice, an e-mail invoice or some other electronic invoice. Payments entered by a customer into an online banking system or direct debit are not electronic invoices.

Homepage

A homepage here is defined as an enterprise’s own Internet homepages or its section in the homepages of a group. Homepages do not refer for example to publication of an enterprise’s contact details on various company and address lists.

Internet sales

Internet sales are communication between a person and a data system. Online shopping as defined here means an order placed by completing and sending a ready-made electronic form on the Internet and shopping in actual Internet shops. Orders placed by a standard email message are not defined as online shopping. Purchases made on an extranet subject to the same conditions are also counted as Internet sales.

Online shopping

Online shopping is the ordering of goods and services via a computer network, regardless of payment or delivery method.

Appendix 2.

Factor analysis of ICT variables

The variables in the ICT survey are (mostly) binary indicators. Factor analysis assumes that the observed variables are continuous (or at least ordinal), because they are modelled as linear combinations of continuous latent factors. One could proceed with a factor analysis for binary variables, which specifies a logistic link function between observed indicators and latent factors. Alternatively, one can continue to use ordinary factor analysis but base it on a tetrachoric correlation matrix. Tetrachoric correlations are estimates for the correlation coefficient of latent bivariate normal distribution based on observed binary variables. The results from factor analysis using tetrachoric correlations are similar to those from the binary factor analysis. The difference is mainly that tetrachoric correlations treat binary variables as incompletely observed underlying variables rather than observed items of latent factors.

We used tetrachoiric correlations for the 16 binary indicators for the usage of different aspects of ICT in the firms in the ICT survey. Using orthogonal rotation and the principal factors method yielded the following rotated factor loadings for three factors with eigenvalues greater than one.

Table A1. Rotated factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Uniqueness
Firm has website	0.1322	**0.8854**	-0.1564	0.1742
Firm has broadband	0.1354	**0.7020**	0.3243	0.3837
Firm has mobile access to internet	0.2011	**0.7382**	0.2517	0.3513
Firm orders through computer networks	0.2590	**0.5490**	0.3133	0.5333
Firm sells through computer networks	0.4405	**0.5793**	0.1155	0.4570
Firm has Enterprise Resource Planning	0.2516	**0.6405**	0.3845	0.3786
Shares SCM data with suppliers	0.3013	0.0714	**0.8988**	0.0963
Shares SCM data with customers	0.2640	0.2855	**0.8003**	0.2083
Shares customer information (within firm)	0.1774	**0.6967**	0.3841	0.3356
Analyses customer information for marketing	0.1907	**0.6689**	0.4096	0.3484
Receives e-invoices via electronic data exchange	**0.7500**	0.3795	0.0792	0.2872
Sending e-invoices via electronic data exchange	**0.7880**	0.4563	0.0414	0.1692
Electronic data interchange used	**0.9605**	0.1052	0.2008	0.0260
Receiving orders	**0.7872**	0.2495	0.2425	0.2593
To suppliers	**0.7729**	-0.1093	0.3617	0.2598
Sending transport documents	**0.7375**	0.1138	0.3758	0.3020

Source: ICT panel 2001-2009 (information refers to the end of previous year, i.e. 2000-2008). Variables are indicators that the firm has or utilizes the technology indicated by variable name.

Factor 1: Loads on using automated data exchange for sending and receiving invoices and orders and sending transport documents. We call this factor EDI (for electronic data interchange).

Factor 2: Loads on firm having a website and access to internet, selling and placing orders via computer networks (website or EDI). It also loads on firm having a special program (CRM) for sharing and analyzing customer information within firm, and firm having Enterprice Resource Planning (ERP). We call this a general Internet factor.

Factor 3: Loads on sharing supply chain management (SCM) data with suppliers or customers via computers (demand forecasts, inventory levels, production plans, deliveries, product planning information), so we call this factor SCM.