Is Innovative Firm Behavior Correlated with Age and Gender Composition of the Workforce? Evidence from a New Type of Data for German Enterprises

Christian Pfeifer
Joachim Wagner

Is Innovative Firm Behavior Correlated with Age and Gender Composition of the Workforce? Evidence from a New Type of Data for German Enterprises

Christian Pfeifer
*Leuphana University Lueneburg
and IZA*

Joachim Wagner
*Leuphana University Lueneburg,
CESIS and IZA*

ABSTRACT

Is Innovative Firm Behavior Correlated with Age and Gender Composition of the Workforce? Evidence from a New Type of Data for German Enterprises[*]

This empirical research note documents the relationship between composition of a firm's workforce (with a special focus on age and gender) and its performance with respect to innovative activities (outlays and employment in research and development (R&D)) for a large representative sample of enterprises from manufacturing industries in Germany using unique newly available data. We find that firms with a higher share of older workers have significantly lower proportions of R&D outlays in total revenues and of R&D employment in total employment, whereas firms with a higher share of female employment seem to be more active in R&D.

JEL Classification: D22, D24, J21, J24, L25

Keywords: ageing, firm performance, gender, Germany, innovation, R&D

Corresponding author:

Joachim Wagner
Leuphana University Lueneburg
Institute of Economics
P.O. Box 2440
D-21314 Lueneburg
Germany
E-mail: wagner@leuphana.de

1. Introduction

Innovative activities are important for the performance of firms and the growth of the economy as a whole. This holds especially for firms from manufacturing industries in Germany. The manufacturing sector is known to be highly export intensive, and Germany is one of the leading actors on the world market for goods. International competitiveness of manufacturing firms depends to a high degree on the capability of firms to supply high-quality innovative products (see Wagner 2011). Therefore, research and development (R&D) activities that are important for continuous improvements of products and production processes are a decisive factor for the success of firms. That said, information on factors that are positively or negatively related to R&D activities of firms is important for both researchers and policy makers. This paper focuses on one aspect that is becoming more and more important not only for Germany but for other highly industrialized countries, too – the consequences of an ageing population and an ageing workforce for innovation activities of firms.

The effects of an aging workforce are usually analyzed in the context of productivity estimates that mostly indicate a negative effect of the oldest age groups on firm performance (see Pfeifer and Wagner (2012) for a review). In the context of innovations, we should also expect a negative correlation with age, because the literature about age-dependent skill formation suggests that age has a negative effect on human capital investments (e.g., due to decreasing amortization periods) and on cognitive skills such as intelligence, memory power, reasoning, creativity, and fluid problem-solving skills (e.g., Kaufman 2001; Pfeiffer and Reuß 2008), which are important for innovative activities. Previous studies, which use firm level data, find indeed that (old) age has a negative effect on firms' innovation activities (e.g., product

and process innovations, patents) (e.g., Schneider 2008 for Germany, Parotta et al. 2012 for Denmark). The demographic change makes it also necessary to activate female labor supply. In order to analyze the effect of female employment on innovative activities, we include in our regression analysis the gender composition of a firm's workforce in addition the age composition of the workforce. We extend the literature by using newly available high quality official data for Germany.

The remainder of the paper is structured as follows. The next section informs about the data and the estimation strategy. Section 3 presents the results of our fractional logit estimates. The paper concludes with a short summary in Section 4.

2. Data and Estimation Strategy

The empirical investigation uses data for enterprises[1] from manufacturing industries that come from two sources. The first source is the cost structure survey for enterprises in the manufacturing sector. This survey is carried out annually by the statistical offices as a representative random sample survey stratified according to the number of employees and the industries (see Fritsch et al. 2004). The sample covered by the cost structure survey represents all enterprises with at least 20 employees from manufacturing industries. About 45 percent of the enterprises with 20 to 499 employees and all enterprises with 500 and more employees are included in the sample.[2] While firms with 500 and more employees are covered by the cost structure survey in each year, the

---

[1] Data are for legal units (enterprises, or *Unternehmen*), not for local production units (establishments, or *Betriebe*). In this paper we use the term firm as a synonym for enterprise.

[2] For details see the quality report for the cost structure survey published by the Federal Statistical Office that is available on the web:
http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Qualitaetsberichte/VerarbeitendesGewerbeIndustrie/Kostenstruktur,property=file.pdf

sample of smaller firms is part of the survey for four years in a row only. This survey is the source for information on innovative behavior, firm size and industry affiliation:

We measure innovative firm behavior by two outcome variables. The first innovation variable is the *proportion of R&D outlays in total revenues*. The second innovation variable is the *share of workers in R&D in total employment*.

Firm size is measured by the number of people working in a firm. This measure is also included in squares in the empirical models to take care of non-linearity in the relation between firm size and innovative activities.

Industry affiliation of a firm is recorded at the two-digit level.

The second source of data is the Establishment History Panel (*Betriebs-Historik-Panel*).[3] Details aside, this data set is built from individual level information for employees covered by social security.[4] In a first step for each year from 1975 onwards information for all employees working in a local production unit (establishment) was aggregated, and this is the standard version of the Establishment History Panel. In this study we use a different version of the Establishment History Panel. Here for multi-establishment enterprises information from all establishments of the enterprise was aggregated in a second step. The result is a data set with detailed information about the characteristics of the employees (covered by social security) in each enterprise in a year. Information reported to the social security system includes, among others, the sex

---

[3] For an introduction to the Establishment History Panel see Spengler (2008); a detailed description of the current version is Hethey-Maier and Seth (2010).

[4] "All employees who are subject to at least one of the following compulsory insurances are liable to social security: health insurance, long-term care insurance, pension insurance, unemployment and accident insurance. However, not liable to social security and thus not included in the data are civil servants, conscripts, those doing alternative civilian service, self-employed, judges, scholars, students, pensioners, clergy and others." (Spengler 2008, p. 502)

of a person, the age and the qualification (educational level attained and vocational training concluded).

Share of employees from a certain age group is defined as the total number of employees (covered by social security) from the respective age group over the total number of employees (covered by social security) in an enterprise; the share is measured as a percentage.

Share of female employees is defined as the total number of females (covered by social security) over the total number of employees (covered by social security) in an enterprise; the share is measured as a percentage.

Share of medium qualified employees is defined as the total number of employees (covered by social security) with either the high-school diploma (*Abitur*) as the highest educational level attained or with vocational training concluded over the total number of employees (covered by social security) in an enterprise; the share is measured as a percentage.

Share of highly qualified employees is defined as the total number of employees with a polytech or university degree over the total number of employees (covered by social security) in an enterprise; the share is measured as a percentage.[5]

Share of part-time employees is defined as the total number of employees in part-time over the total number of employees (covered by social security) in an enterprise; the share is measured as a percentage.

---

[5] Note that this information on the diversity of the employees is not available in more detail; for example, the number of female employees aged 30 to 34 with a university degree is not available from the data (although it would be possible to compute this figure from the individual level information available).

The cost structure survey for enterprises in the manufacturing sector is conducted by the German statistical offices. The data can be accessed for scientific research via the Research Data Centres of the Federal Statistical Office and the Statistical Offices of the Federal States (see Zühlke et al. 2004). The Establishment History Panel is build from administrative data by the Research Data Centre of the Federal Employment Agency at the Institute for Employment Research. The data can be accessed via this Research Data Centre for scientific research (see Spengler 2008).

Linking these confidential firm level information across the borders of the data producers, however, is difficult. Details aside, it is technically not easy (but not impossible either) and it is legal only if the firm agreed in written form. The basic idea of the project *KombiFiD* (an acronym that stands for *Kombi*nierte *Fir*mendaten für *D*eutschland, or combined firm level data for Germany) that is in detail described on the web (see www.kombifid.de) is to ask a large sample of firms from all parts of the German economy to agree to match confidential micro data for these firms that are kept separately by three data producers (the Statistical Offices, the Federal Employment Agency, and the German Central Bank) in one data set. These matched data are made available for scientific research while strictly obeying the data protection law, i.e. without revealing micro level information to researchers outside the data producing agencies. In *KombiFiD* 54,960 firms were asked to agree in written form to merge firm level data from various surveys and administrative data for the reporting years 2003 to 2006. 30,944 firms replied and 16,571 agreed. These 16,571 firms are in the *KombiFiD Agreement Sample*.[6]

---

[6] Access to the data is easy and costless (for details, see www.kombifid.de). The data are of a high quality, and participation of the enterprises in the cost structure survey and in the delivery of information

The sample of enterprises used in the empirical investigation performed here consists of all firms from manufacturing industries in West Germany[7] in the *KombiFiD Agreement Sample* for which information from both data sources – the cost structure survey and the Establishment History Panel - could be linked in the *KombiFiD* project. Enterprises that do not have complete information for each year from 2003 to 2006 were dropped from the computations.[8] This leads to a balanced panel data set with 16,900 observations for 4,225 firms and 4 years.

Descriptive statistics for all variables and the pooled data are reported in Table 1. It is evident from these descriptive statistics that the variation of the variables over the four years covered is very small when compared to the variation between the firms in the sample. Therefore, the within firm variation of important dimensions of

---

on the employees covered by social security is mandated. Instead of opinions and "guesstimates" collected in surveys with voluntary participation, the *KombiFiD*-Data have reliably information on variables like revenues that are difficult to collect in interviews or questionnaires without mandatory participation. Therefore, the *KombiFiD*-Data should at least be carefully looked at by researchers from various fields in economics (including labor economics, industrial organization and international economics) interested in working with firm level data.

[7] The sample is limited to firms from West Germany. There are large differences between enterprises from West Germany and the former communist East Germany even many years after the unification in 1990. Therefore, an empirical study should be performed separately for both parts of Germany. The *KombiFiD Agreement Sample* for East German manufacturing firms, however, contains only a small number of firms, and this sample turned out to be not representative for the population of firms in a replication study that compares results based on the complete cost structure survey data and data from the *KombiFiD Agreement Sample* (see Wagner 2012).

[8] Firms with incomplete information for any variable in at least one year were dropped from all computations because there are, on the one hand, by construction no entries due to the fact that the firms taking part in the cost structure survey were sampled before the start of the survey in 2003. On the other hand, exits cannot be identified because firms with information in, say, 2003 but not in 2004 might have closed down – they might have, however, relocated out of manufacturing (or out of Germany) or they might have shrunk below the cut-off point relevant for the cost-structure survey.

diversity of the employees over time cannot be used in fixed effects models to sufficiently identify any relationship between changes in firm innovation proxies over time and diversity of employees.

Table 1: Descriptive statistics

		Standard deviation		
Variables	Pooled mean	Overall	Between	Within
Proportion of R&D outlays in total revenues	0.0111	0.0281	0.0264	0.0096
Share of R&D employment in total employment	0.0232	0.0527	0.0496	0.0178
Share of employees aged less than 30 years (%)	17.1608	9.3881	8.9765	2.7519
Share of employees aged 30 – 49 years (%)	57.5660	9.3267	8.7179	3.3166
Share of employees aged 50 years or older (%)	25.2534	10.0994	9.6630	2.9398
Share of female employees (%)	30.1249	20.7646	20.6295	2.3810
Share of medium qualified employees (%)	61.5195	20.5359	20.2429	3.4671
Share of highly qualified employees (%)	6.0263	7.8487	7.7044	1.5015
Share of part-time employees (%)	18.1817	14.6017	14.2332	3.2653
Firm size (number of employees)	429.6821	3649.5190	3644.7420	192.8708

Note: The data are from a balanced panel (4 years from 2003 to 2006) with a total of 16,900 yearly observations for 4,225 enterprises. For the definitions of the variables see text.

Our outcome variables of interest are the proportion of R&D outlays in total revenues and the share of workers in R&D in total employment as proxies for innovative firm behavior. Both variables are highly positively correlated in each year with a correlation coefficient that increases over time from 0.73 (in 2003) to 0.83 (in 2006). Both variables are censored, as their values are bound between zero and one. About 38 percent of our observations report a positive proportion of R&D outlays in total revenues and about 37 percent a positive share of workers in R&D in total employment as proxies for innovative firm behavior. In order to deal with fractions of this kind, we apply fractional logit regressions with robust standard errors clustered at the firm level, which have been proposed by Papke and Wooldridge (1996) and are normally used for this kind of dependent variables.

8

3.   Estimation Results

The results of the fractional logit regressions for the proportion of R&D outlays in total revenues are presented in Table 2. The first column contains the estimated coefficients, the second column their robust standard errors, and the third column the corresponding p-values for being significantly different from zero. In order to facilitate the interpretation, we have computed marginal effects (fourth column) and elasticities (fifth column) at the means of all covariates. As the mean proportion of R&D outlays in total revenues is only 1.11 percent, the estimated marginal effects look small at first glance. If we compute the elasticities, we can however see the economic importance of the workforce composition. A one percent higher share of employees 50 years or older decreases the proportion of R&D outlays in total revenues by 0.49 percent, which is statistical significant at high levels. The middle age group of employees 30-49 years has however no significant effect. Furthermore, the results reveal that a one percent higher female share increases the proportion of R&D outlays in total revenues significantly by about 0.28 percent. As one would expect for knowledge related outcomes, firms with a more qualified workforce have a higher proportion of R&D outlays in total revenues, whereas firms with a higher share of part-time employees have a lower proportion of R&D outlays in total revenues.

Table 2: Fractional logit regressions for the proportion of R&D outlays in total revenues

Explanatory variables	(1) Coefficient	(2) Standard error	(3) p-value	(4) Marginal effect at means	(5) Elasticity at means
Employees 30 – 49 years (%)	-0.001691	0.005566	0.7610	-0.000009	-0.0968
Employees 50 years or older (%)	-0.019652	0.004867	<0.0001	-0.000101	-0.4937
Female employees (%)	0.009466	0.001966	<0.0001	0.000049	0.2837
Medium qualified employees (%)	0.006972	0.001974	<0.0001	0.000036	0.4267
Highly qualified employees (%)	0.052009	0.003691	<0.0001	0.000267	0.3118
Part-time employees (%)	-0.014284	0.003293	<0.0001	-0.000073	-0.2584
Firm size (number of employees)	0.000051	0.000020	0.0110	2.61e-07	0.0217
Firm size (squared)	-3.74e-10	2.17e-10	0.0850	-1.92e-12	-0.0050
Constant	-8.205332	0.942781	<0.0001		

Note: Fractional logit regressions for proportion of R&D outlays in total revenues. All models include dummy variables for years and 2digit-level industries. Robust standard errors for coefficients clustered at the firm level. The data are from a balanced panel (4 years from 2003 to 2006) with a total of 16,900 yearly observations for 4,225 enterprises.

The results of the fractional logit regressions for the share of R&D employment in total employment are presented in Table 3. The mean share of R&D employment in total employment is only 2.32 percent so that the estimated marginal effects look again small at first glance. Thus, we focus on the elasticities. A one percent higher share of employees 50 years or older decreases the share of R&D employment in total employment significantly by 0.52 percent. The middle age group of employees 30-49 years has no significant effect. The results further reveal that a one percent higher female share increases the share of R&D employment in total employment significantly by about 0.16 percent. Firms with a more qualified workforce have a higher share of R&D employment in total employment, whereas firms with a higher share of part-time employees have a lower share of R&D employment in total employment.

Table 3: Fractional logit regressions for the share of R&D employment in total
employment

Explanatory variables	(1) Coefficient	(2) Standard error	(3) p-value	(4) Marginal effect at means	(5) Elasticity at means
Employees 30 – 49 years (%)	0.001420	0.004919	0.7730	0.000017	0.0808
Employees 50 years or older (%)	-0.020856	0.004195	<0.0001	-0.000243	-0.5205
Female employees (%)	0.005210	0.001854	0.0050	0.000061	0.1551
Medium qualified employees (%)	0.003722	0.001985	0.0610	0.000043	0.2263
Highly qualified employees (%)	0.056255	0.004145	<0.0001	0.000654	0.3350
Part-time employees (%)	-0.014297	0.002958	<0.0001	-0.000166	-0.2569
Firm size (number of employees)	0.000040	0.000017	0.0200	4.60e-07	0.0168
Firm size (squared)	-3.00e-10	1.89e-10	0.1120	-3.49e-12	-0.0040
Constant	-9.043343	0.949547	<0.0001		

Note: Fractional logit regressions for share of R&D employment in total employment. All models include dummy variables for years and 2digit-level industries. Robust standard errors for coefficients clustered at the firm level. The data are from a balanced panel (4 years from 2003 to 2006) with a total of 16,900 yearly observations for 4,225 enterprises.

4.   Conclusion

Overall, the impact of the workforce composition variables do not differ noteworthy between both proxies for innovative firm behavior in the fractional logit regressions. Our estimates for innovative firm behavior reveal on average a negative correlation with older employees and a positive correlation with female employment. The first finding indicates a potential problem for innovative firms (and economies) that are faced with an aging workforce due to the ongoing demographic change. The second finding shows however a potential strategy to solve this problem, namely to activate female labor supply, which is still not exhausted if we look at the lower employment rates and working hours of women compared to men - at least in Germany (e.g., Humpert and Pfeifer 2012).

References

Fritsch, Michael, Bernd Görzig, Ottmar Hennchen, and Andreas Stephan (2004), Cost Structure Surveys for Germany. Schmollers Jahrbuch / Journal of Applied Social Science Studies 124 (4), 557-566.

Hethey-Maier, Tanja, and Stefan Seth (2010), Das Betriebs-Historik-Panel (BHP) 1975 – 2008. Handbuch Version 1.0.2. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, FDZ-Datenreport 04/2010. Nürnberg. (Download: http://doku.iab.de/fdz/reporte/2010/DR_04-10.pdf, accessed 6.12.2011).

Humpert, Stephan, and Christian Pfeifer (2012), Explaining age and gender differences in employment rates: a labor supply side perspective. Journal for Labour Market Research (OnlineFirst, DOI 10.1007/s12651-012-0108-8).

Kaufman, Alan S. (2001), WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. Intelligence 29, 131-167.

Papke, Leslie E,. and Jeffrey M. Wooldridge (1996), Econometric methods for fractional response variables with an application to 401(K) plan participation rates. Journal of Applied Econometrics 11, 619-632.

Parrotta, Pierpaolo, Dario Pozzoli, and Mariola Pytlikova (2012), The nexus between labor diversity and firm's innovation. IZA Discussion Paper No. 6972.

Pfeifer, Christian, and Joachim Wagner (2012), Age and gender composition of the workforce, productivity and profits: Evidence from a new type of data for German enterprises. IZA Discussion Paper No. 6381.

Pfeiffer, Friedhelm, and Karsten Reuß (2008), Age-dependent skill formation and returns to education. Labour Economics 15, 631-646.

Schneider, Lutz (2008), Alterung und technologisches Innovationspotential: Eine Linked Employer-Employee Analyse. Zeitschrift für Bevölkerungswissenschaft 33 (1), 37-54.

Spengler, Anja (2008), The Establishment History Panel. Schmollers Jahrbuch / Journal of Applied Social Science Studies 128 (3), 501-509.

Wagner, Joachim (2011), Exports and Firm Characteristics in German Manufacturing Industries: New Evidence from Representative Panel Data. Applied Economics Quarterly 57 (2), 107-143.

Wagner, Joachim (2012), The Quality of the KombiFiD-Sample of Enterprises from Manufacturing Industries: Evidence from a Replication Study. Schmollers Jahrbuch / Journal of Applied Social Science Studies 132 (3), 379-392.

Zühlke, Sylvia, Markus Zwick, Sebastian Scharnhorst, and Thomas Wende (2004), The research data centres of the Federal Statistical Office and the statistical offices of the Länder. Schmollers Jahrbuch / Journal of Applied Social Science Studies 124 (4), 567-578.