Discussion Paper No. 16-014

Attenuation Bias when Measuring Inventive Performance

Thomas Zwick and Katharina Frosch

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 16-014

Attenuation Bias when Measuring Inventive Performance

Thomas Zwick and Katharina Frosch

Attenuation Bias when Measuring Inventive Performance

Thomas Zwick[1], University of Würzburg and Centre for European Economic Research (ZEW), Mannheim

Katharina Frosch, Brandenburg University of Applied Sciences

February 2016

Abstract

Most previous results on determinants of inventive performance are biased because inventive performance is measured with error. This measurement error causes attenuation bias. More specifically, for example age and education as drivers of patenting success have biased coefficients and too high standard errors when inventive performance is measured in short observation periods. The reason for measurement errors in inventive performance is that patents are typically applied for in waves.

JEL Codes: C33, C52, O31

Key words: Measurement error, inventive performance, observation period

1. Introduction

An influential and growing literature measures inventive performance by counting the number, quality or value of patents of an inventor. To the best of our knowledge, almost all published papers on determinants of inventive productivity on the basis of patent applications employed very short measurement periods: four years (Walsh and Nagaoka, 2009 for the US), five years (Giuri et al., 2007; Sauermann and Cohen, 2010), six years (Walsh and Nagaoka, 2009 for Japan), nine years (Vänäänen, 2010), and ten years (Mariani and Romanelli, 2007).[2]

---

[1] Corresponding author. University of Würzburg, Sanderring 2, 97070 Würzburg, E-Mail: thomas.zwick@uni-wuerzburg.de. We thank Boris Hirsch, Karin Hoisl and an anonymous referee for useful comments, and Karin Hoisl for retrieving and cleaning the patent data and matching them to the survey data. We are grateful for financial support by the German Research Foundation (DFG, Grant ZW172/2-1).

[2] The only exceptions we are aware of are the papers by Hoisl (2007a, 2007b). She uses the patenting histories of 3049 German inventors at the European Patent Office between 1977 and 1999/2002 and calculates age – output relationships of inventors. Jung and Ejermo (2014) use the observation period 1985-2007 but do not

A short measurement period of patenting activity is prone to measurement error, however, because patents are not applied for continuously but in waves. Reasons for this pattern are for example strategic application behaviour and the splitting-up of one invention into several patents for example to create so-called "royalty stacking" where several patents apply to one new product (Lemley and Shapiro, 2007). In our data on the patenting history of German inventors for example, on average, inventors had no European patent in about 80% of the years during their career. This finding coincides with strong variance of patenting success over time found in other studies (Hoisl, 2007a, p. 627). We therefore expect an attenuation bias induced by measurement error in the dependent variable (Griliches and Mairesse, 1996) when the observation period of patenting behaviour is shorter than the entire (observable) career.

Let true individual inventive productivity measured over the entire career be y^*. The variable y measures true inventive productivity with error: $y = y^* + e$. The structural model is thus: $y^* = \boldsymbol{x}'\boldsymbol{\beta} + v$ with $\boldsymbol{x}$ a vector of explanatory variables, $\boldsymbol{\beta}$ a vector of regressors and v an error term. The data however show the relationship $y = \boldsymbol{x}'\boldsymbol{\beta} + v + e$. The first consequence of measurement error in y is that the composite error term is too large, which implies a larger covariance matrix for $\hat{\beta}_{OLS}$. The second consequence is that $\hat{\beta}_{OLS}$ is biased if e is correlated with $\boldsymbol{x}$ (Wooldridge, 2010, p. 77-78).

This paper demonstrates that the measurement of the effect of education level and age on inventive performance indeed suffers from both sources of attenuation bias on inventive performance when the observation period is shorter than the entire career. We chose the two exemplary explanatory variables because these are the most widely used exogeneous variables in the estimation of inventive performance literature.

2. Data

Our data have been collected in the course of a self-administered survey of German inventors mainly active in clean technology (CT) and mechanical elements (ME) and merged with the complete list of European patent (EP) applications of all inventors. The data are

---

calculate determinants of inventive performance. Gruber et al. (2013) use the patent history of inventors between 1977 and 2003. They explain technological recombination breadth instead of inventive performance.

therefore comparable with the papers mentioned above based on PatVal and other survey data linked with patent histories.

We identified all patent applications with priority dates between 2004 and 2008 assigned to the two fields that listed at least one inventor with home address in Germany. This resulted in 8,313 inventors in our basic sample. We received 1,700 responses (response rate 29.5 percent). We added all patent applications of the responding inventors between 1978 and 2010 using the PATSTAT database. Further details on the data generation and descriptive statistics can be found in Frosch et al. (2015).

3. Empirical strategy

Using our data, we replicate the estimation approaches of the papers mentioned above and vary the length of the observation period – we use the full career observation length, a ten year spell from 2000-2009 and a five year spell from 2005-2009.[3] We interpret changes in effect size and significance between the three observation lengths as attenuation bias. We can identify patent applications of our inventors for a maximum period of 33 years (1978-2010) and therefore assess inventive performance for most inventors for their entire career until 2010.

We use forward citations as measure of patent quality besides the number of patents. The binary variable $C_{bin, fract, j}$ indicates whether patent application *j* receives at least as many ($C_{bin, fract, j} = 1$) or less ($C_{bin, fract, j} = 0$) fractional citations (i.e. citations corrected for the number of co-inventors) than the average patent application in the same priority year *t* and technology field *k*:

$$C_{bin, fract, j} = \begin{cases} 1 \, if \, \dfrac{C_{j,k,t}}{n_j} \geq \overline{C}_{k_j t_j, fract} \\ 0 \, if \, \dfrac{C_{j,k,t}}{n_j} < \overline{C}_{k_j t_j, fract} \end{cases}$$

[3] Please note that in our sample about two thirds of the inventors only had inventions in the 2000-2009 period and one third only in the 2005-2009 period.

Inventive performance of inventor *i* equals the average of the above-average citation counts, i.e. the sum of the binary indicators for all patent applications *J* of inventor *i*, divided by the number of work years (Frosch et al., 2015):

$$C_{av,i} = \frac{\sum_{j=1}^{J} C_{bin,fract,j}}{EXP_i}$$

Birth cohort is measured by six dummy variables taking the value of one if the inventor was born before 1970, between 1966 - 1970, 1961 - 1965, 1956 - 1960, 1951 - 1955, 1946 - 1950, and zero otherwise. This implies that the oldest cohort is older than 60 years of age at the time of our survey.

The inventor´s highest formal educational level is measured by three dummy variables taking the value of one in case the inventor obtained a vocational education, conducted academic studies, or received a PhD, and zero otherwise.

In line with the literature, we also control for the characteristics of the inventors´ employers (patent applicants). We aggregate applicant characteristics over the entire career of inventors and compute applicant type shares. The latter are defined as the number of patent applications produced by the inventor while associated with an applicant institution of a certain type (research institution or university, private company, independent inventor) divided by the total number of patent applications filed by the inventor. The size of the patent portfolio, the number of employees of the applicant institutions, the patent status (pending, refused, withdrawn, and granted) and technology shares of patents are calculated similarly. Period shares capture the temporal distribution of the inventor's activity and help us to differentiate between time and cohort effects. The descriptive statistics of our sample are reported in Table 1.

We use the full observation periods for the calculation of all explanatory variables and therefore do not vary between observation lengths. The reason is that with our approach the measurement quality of age and education is not affected by the period length. Some explanatory variables such as applicant institution characteristics of patent portfolio would otherwise be measured with error in shorter periods. Using the entire observation periods for these explanatory variables reveals the pure attenuation bias induced by measurement errors of the dependent variable in shorter observation periods. Measurement errors in the explanatory variables might lead to an additional attenuation bias (Wooldridge, 2010).

4. Results

The average Gini-coefficient of all inventors for their patents during their career reveals a very unequal distribution of patents over time. On a possible range from 0 (completely equal distribution of patent citations over time) to 1 (completely unequal distribution of patent citations over time) it takes the value 0.84. This implies that patent with more than average citations are granted in waves and many inventors have a large share of years during their career without patents that have been cited more than average. When we take inventive performance data from the entire career, an academic education increases performance by about 13% and a PhD by 41% in comparison to vocational training. If we only consider maximally ten years of productive output (2000-2009), only PhD has a massive influence of over 145% and academic education has the same influence on productivity as vocational training. If we consider a period of five years (2000-2005), again the dummy for academic education has no significance and the PhD dummy has a much smaller value than in the nine year period. Longer observation periods hence lead to higher significant and more plausible estimates of the impact of education on inventive performance. Especially a PhD is – with an almost 50% higher number of patents per work year that are cited more than the average of the patents in the same field – a good predictor of a high inventive productivity during the career. When we look at the coefficients of our groups of birth cohorts, we also find that the size and sometimes even the sign of the coefficients changes. Birth cohort has hardly any impact in the shorter observation periods and the expected declining impact with age in the complete patenting history sample, compare Table 2. In the full career sample, we find that inventive productivity is about 5% lower for inventors who were older than 45 in the year 2010 in comparison to inventors who were younger than 40.

We repeat our estimations with the number of patents instead of their quality as dependent variable. Again, the coefficients of age and education in the shorter observation periods are smaller and less significant, compare Table 3. This means that attenuation bias also reduces the significance and size of the determinants of inventive productivity if we use a quantity measure instead of a quality measure.

We therefore replicate the comparably low and insignificant impact of age and education on inventive productivity when the dependent variable is measured by few years of patent history as in most papers in the literature (Mariani and Romanelli, 2007, Giuri et al., 2007; Walsh and Nagaoka, 2009, Sauermann and Cohen, 2010; Vänäänen, 2010). We also obtain a comparably stronger impact of age and education on the number of patents than the quality of patents as found for example by Mariani and Romanelli (2007).

When we use log age instead of age groups or fractional counts instead of whole patent counts in robustness tests, we get roughly the same results (also compare Frosch et al., 2015). Our results are also robust if we use average value and the maximum value of the patents invented by the inventor as dependent variables (Mariani and Romanelli, 2007) or if we include temporal concentration of patenting activity (Hoisl, 2007a). This means that our results are not driven by the choice of the dependent variable or the list of explanatory variables.

5. Conclusions

Our analysis shows that almost all papers on inventive productivity based on patent applications published in journals so far are affected by attenuation bias. More specifically, measurement errors for the dependent variable in short observation periods lead to biased coefficients and significance levels for age and education that are too low. Measurement errors are induced because patents are applied for in waves, for example as a consequence of several patents applied for from one invention. One reason for this behavior might be "royalty stacking". Patenting activities therefore should be measured on the basis of complete inventor careers data instead of shorter periods if possible.

Our paper also has implications for other measures of inventive productivity when these are not evenly distributed over time. An obvious example for other measures that might be affected by attenuation bias is publication activities of scientists and their citations. Van Ours (2009) for example notes that economists tend to have many years without publications and some years with a lot of publications.[4] Also citations of published work by scientists might strongly vary from year to year.

This short note only proofs on the basis of one linked employer-patent data set that extending the observation period dramatically changes the size and sometimes the sign of important explanatory variables in estimations on inventive productivity. It may be

[4] Levin and Stephan (1991) for example use a two years observation period to predict research productivity over the life cycle.

important to replicate published estimates on the determinants of inventive productivity that used short observation periods with longer patent data, citation or publication histories in order to make sure that the results are robust. Especially the impact of the variables that have a low or insignificant impact on inventive performance such as age and education might have been underestimated.

Based on the results of this paper, we recommend that future studies aiming to gain reliable insights into the drivers of inventive or scientific productivity should make use of performance measures that span as many periods as possible.

Literature

Frosch, K., D. Harhoff, K. Hoisl, C. Steinle and Zwick, T. (2015) Individual Determinants of Inventor Productivity: Report and Preliminary Results with Evidence from Linked Human Capital and Patent Data, ZEW Discussion Paper 015-001, Mannheim.

Giuri P., Mariani M., Brusoni S., Crespi G., Francoz D., Gambardella A., Garcia-Fontes W., Geuna A., Gonzales R., Harhoff D., Hoisl K., Lebas C., Luzzi A., Magazzini L., Nesta L., Nomaler O., Palomeras N., Patel P., Romanelli M., B. Verspagen (2007) Inventors and Invention Processes in Europe: Results from the PatVal-EU Survey. *Research Policy,* **36**(8), 1107-1127.

Griliches, Z., J. Mairesse (1995) Production Functions: The Search for Identification, NBER Working Paper 5067, Cambridge MA.

Gruber, M., D. Harhoff, K. Hoisl (2013) Knowledge Recombination across Technological Boundaries: Scientists versus Engineers. *Management Science* **59**(4), 837-851.

Hoisl, K. (2007a) Tracing Mobile Inventors – The Causality between Inventor Mobility and Inventor Productivity, *Research Policy* **36**(5), 619-636.

Hoisl, K. (2007b) *A Closer Look at Inventive Output – The Role of Age and Career Paths*, unpublished paper.

Jung, T. and Ejermo, O. (2014) Demographic Patterns and Trends in Patenting: Gender, Age, and Education of Inventors. *Technological Forecasting and Social Change* **86**, 110-124.

Lemley, M. and C. Shapiro (2007) Patent Holdup and Royalty Stacking, *Texas Law Review* **85**, 1991 – 2049.

Levin, S. and P. Stephan (1991) Research Productivity Over the Life Cycle: Evidence for Academic Scientists. *American Economic Review* **81**(1), 114-132.

Mariani, M. and M. Romanelli (2007) “Stacking“ and “Picking“ Inventions: The Patenting Behavior of European Inventors. *Research Policy* **36**(8), 1128-1142.

Ours, J. van (2009) Will you Still Need me: When I´m 64?. *De Economist* **157**(4), 441-460.

Sauerman, H., W. Cohen (2010) What Makes Them Tick? Employee Motives and Firm Innovation. *Management Science* **56**(12), 2134-2153.

Walsh, J. and S. Nagaoka (2009) Who Invents?: Evidence from the Japan-U.S. Inventor Survey, Cognitive Type and R&D Performance. *R&D Management* **29**(3), 247-254.

Wooldridge, J. (2010) *Econometric analysis of cross section and panel data*, 2nd edition, MIT Press, Cambridge MA.

Tables

Table 1: Variable descriptions and descriptive results (N = 1,593)

VARIABLES	Mean	Median	S.D.	Description
ln(patent quality+1)	0.06	0.03	0.09	ln of sum of dummy variables taking the value = 1 if a patent receives more (fractional) citations than average of all patents in same technology field and year, divided by number of working years plus one
Birth year after 1970	0.21	0	0.41	Dummy = 1 if inventor was born after 1970 (reference category)
Birth year 1966–1970	0.21	0	0.41	Dummy = 1 if inventor was born between 1966 and 1970
Birth year 1961–1965	0.14	0	0.35	Dummy = 1 if inventor was born between 1961 and 1965
Birth year 1956–1960	0.11	0	0.32	Dummy = 1 if inventor was born between 1956 and 1960
Birth year 1951–1955	0.16	0	0.37	Dummy = 1 if inventor was born between 1951 and 1955
Birth year 1946–1950	0.21	0	0.41	Dummy = 1 if inventor was born between 1946 and 1950
Vocational education	0.11	0	0.32	Dummy = 1 if highest level of education is vocational education in 2010 (reference category)
Academic education	0.59	1	0.49	Dummy = 1 if highest level of education is academic education
PhD	0.3	0	0.46	Dummy = 1 if highest level of education is PhD
Applicant size: 1 patent	0.04	0	0.17	Share of applicant firms with 1 patent (1978–2010)
Applicant size: 2–24 patents	0.17	0	0.34	Share of applicant firms with 2–24 patents
Applicant size: 25–249 patents	0.21	0	0.37	Share of applicant firms with 25–249 patents
Applicant size: 250–999 patents	0.13	0	0.3	Share of applicant firms with 250–999 patents
Applicant size: 1000–9999 patents	0.46	0.32	0.46	Share of applicant firms with more than 1,000 patents
Applicant type: private company	0.93	1	0.24	Share of applicant private companies
Applicant type: university/research institute	0.04	0	0.19	Share of applicant university/research institutes
Applicant type: individual	0.03	0	0.15	Share of individual applicants
Patent application pending	0.37	0.31	0.34	Share of pending patents

Patent application withdrawn	0.15	0	0.23	Share of withdrawn patents
Patent application refused	0.01	0	0.07	Share of refused patents
Patent application granted	0.47	0.5	0.34	Share of granted patents
Electrical machinery, apparatus, energy	0.12	0	0.27	Area share of patents at inventor level in the respective technological field
Electrical engineering	0.02	0	0.08	
Semiconductors	0.05	0	0.17	
Instruments	0.05	0	0.14	
Chemistry	0.03	0	0.14	
Materials/surface technology/chemical engineering/environmental technology	0.13	0	0.26	
Mechanical engineering	0.1	0	0.22	
Engines, pumps, turbines	0.15	0	0.3	
Mechanical elements	0.2	0	0.32	
Transport	0.1	0	0.23	
Other fields	0.04	0	0.15	
Share of patents 1975-1979	0	0	0.02	Period share of patents at inventor level in the respective period
1980-1984	0	0	0.03	
1985-1989	0.01	0	0.06	
1990-1994	0.03	0	0.08	
1995-1999	0.07	0	0.15	
2000-2004	0.25	0.15	0.29	
2005-2009	0.61	0.62	0.35	
2010-2012	0.02	0	0.07	

Table2: Regression models of patent quality with different observation periods

	(Model 1a) Full Career	**(Model 1b) 2000-2009**	**(Model 1c) 2005-2009**
VARIABLES	ln(patent quality + 1)	ln(patent quality+ 1)	ln(patent quality + 1)
Academic education	0.127**	0.741	0.271
	[0.055]	[0.449]	[0.298]
PhD degree	0.409***	1.458***	0.611**
	[0.086]	[0.500]	[0.308]
Birth year 1966-1970	-0.023**	-0.012	0.006
	[0.009]	[0.046]	[0.035]
Birth year 1961-1965	-0.040***	-0.011	-0.005
	[0.010]	[0.040]	[0.033]
Birth year 1956-1960	-0.043***	0.027	0.015
	[0.008]	[0.041]	[0.034]
Birth year 1951-1955	-0.062***	-0.013	0.035
	[0.009]	[0.040]	[0.041]
Birth year 1946-1950	-0.049***	0.053	0.068*
	[0.010]	[0.052]	[0.035]
Additional control variables[A]	Included	included	Included
Observations	1,593	1,593	1,473
R-squared	0.160	0.152	0.046

Notes:

Dependent variable: ln(patent quality +1), patent quality as measured by average above average citation counts

All time-varying explanatory variables are calculated between job entry and 2010 and divided by 10, besides dummies.

Reference categories: vocational education, birth year later than 1970.

[A] Applicant size and type shares, patent status shares, technical area shares, period shares, and constant.

Robust standard errors based on applicant clusters in brackets.

Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 3: Regression models of patent number with different observation periods

	(Model 2a) Full Career	(Model 2b) 2000-2009	(Model 2c) 2005-2009
VARIABLES	ln(average number of patents per work year+1)	ln(average number of patents per work year+1)	ln(average number of patents per work year+1)
Academic education	0.395**	0.530*	0.104
	[0.194]	[0.272]	[0.245]
PhD degree	1.452***	1.339***	0.545*
	[0.281]	[0.354]	[0.290]
Birth year 1966-1970	-0.138***	-0.132***	-0.099***
	[0.028]	[0.033]	[0.030]
Birth year 1961-1965	-0.188***	-0.109***	-0.060*
	[0.034]	[0.036]	[0.033]
Birth year 1956-1960	-0.245***	-0.122***	-0.088***
	[0.029]	[0.033]	[0.030]
Birth year 1951-1955	-0.303***	-0.133***	-0.058*
	[0.028]	[0.034]	[0.033]
Birth year 1946-1950	-0.329***	-0.092**	0.094**
	[0.038]	[0.044]	[0.043]
Additional control variables[A]	Included	included	Included
Observations	1,593	1,586	1,450
R-squared	0.264	0.169	0.124

Notes: see Table 2