EMPLOYMENT IN RESEARCH AND DEVELOPMENT ACTIVITY IN THE EU-27 COUNTRIES – MACRO PANEL DATA

Jarosław Wąsowicz

ORCID https://orcid.org/0000-0002-8809-1036
University of Economics in Katowice
Faculty of Spatial Economy and Regions in Transition
Department of Labour Market Analysis and Forecasting
1 Maja 50 Street, 40-287 Katowice, Poland
E-mail: jaroslaw.wasowicz@ue.katowice.pl

Abstract: *The paper focuses on explaining the relationship between employment in research and development activities and selected variables describing the low and high technology sector, taking into account individual effects in EU-27 countries, based on balanced macro panel data (n=405, 27 cross-sectional units, research period 2008-2022). The study used panel models with decomposition i.e. one-way error component regression model. The construction of one-way error models is based on the assumption that research units (groups of units) differ from each other in certain ways characteristics that remain constant over time for a specific entity. The hypothesis of the existence of an individual effect was verified. The model with intercept decomposition described share of employment in R&D better than the model without decomposition based on the same data. The applied panel models with intercept decomposition made it possible to estimate differences in share of employment in research and development activities resulting from the affiliation of the analyzed sector to a given EU-27 country.*

Key words: *employment in R&D, low-tech and high-tech sectors, patents, Knowledge Intensive Activities (KIA), macro panel data, error component regression model.*
JEL codes: *J2, O1, O33*.

1. Introduction

Digital technologies are reshaping EU labour markets, creating new jobs, and automating others while shifting demand towards more digitally intensive occupations (www1). Share of EU-27 adult workers whose tasks were replaced by new digital technologies in 2020-2021 is 14%. Share of adult workers who had to learn new digital technologies during the same period is 35%. Share of jobs demanding at least basic digital skills in the EU-27 in 2021 is 87% (www1; Da Roit & Iannuzzi, 2022; Leonard & Tyers, 2021; Hodder, 2020; Schlogl et al., 2021; Dorschel, 2022).

The analysis of employment changes in research and development activities, classified as one of the individual component classes of the high-tech sector, is justified in particular due to the dynamism of outsourcing and offshoring of R&D functions both within and outside the EU Member States. The servitization process taking place on a micro scale, i.e. in the activities of enterprises and, as a result, also in the national economies of the European Union, materializes in particular through the progressive fragmentation of the production process and the use of the effects of research and development work carried out by specialized external entities. As a result of such adjustment mechanisms in the economy, the specificity of the research and development sector takes on at least a dichotomous dimension, i.e. in some enterprises it still plays an integral role in the production process and for some, and this is a growing number of entities, it becomes an external service or a subcontract. The contemporary mechanisms of changes taking place in business models of enterprises described above are one of the important determinants of changes in the employment absorption capacity of individual economies in the European Union, thus contributing to changes in the state, dynamics and dispersion of employment in the group of its member states.

Research and development activity (R&D) as creative work undertaken systematically in order to increase the amount of knowledge, in particular knowledge about man, culture and society, and the use of this knowledge to develop new applications in the economy, is one of the basic classification categories (components) high technology sectors in the European Union countries. The innovative progress of enterprises may mean an improvement in their competitive position and the entire economy, and as a result, an increase in the ability to create new, highly efficient jobs in the knowledge-intensive activities.

The study took into account in particular employment in the high and low technology sectors as variables that could explain the diversity of the share of employment in research and development activities in EU countries. The source of a higher share of employment in research and development activities may result from the average technical advancement of a given economy and thus generating internal demand for knowledge-intensive services. The share of high-tech occupations in total employment indicates the technology intensity of a sector or of a whole economy.

The main aim of the paper is to explain the relationship between employment in research and development activities and selected variables describing the low and high technology sector, taking into account individual effects in EU-27 countries, based on balanced macro panel data.

2. Methodology and Data

Panel data analysis is used when information about individual cross-sectional units is compiled over a relatively long period of time, i.e. more than one period. As a result, the acquired statistical data are observed in at least two dimensions and are cross-sectional and temporal. One of the exemplifications of this type of data is a two-dimensional variable, dependent on time and space (Arellano & Honoré, 2000; Maddala, 2006; Hsiao, 2007; Kufel, 2007; Kunst, 2010), as shown below:

$$\boldsymbol{X_{it}} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1T} \\ X_{21} & X_{22} & \dots & X_{2T} \\ \dots & \dots & \dots & \dots \\ X_{N1} & X_{N2} & \dots & X_{NT} \end{bmatrix} \tag{1}$$

Statistical data from available databases, in which the unit of observation is a specific group (country, economic sectors, etc.), create a macro panel. Panel models are estimated on the basis of a special type of cross-sectional data in which the number of observed objects N exceeds, and in certain cases significantly exceeds, the number of points in time T (Dańska-Borsiak, 2009a; Kunst, 2010; Torres-Reyna, 2007). Macro panels are characterized by a relatively small number of cross-sectional observations (N ranges from a dozen to several dozen) and a longer sample period, compared to micro panels (Dańska-Borsiak, 2009b; Kunst, 2010), reaching the length of the time series T is in the order of several dozen observations. The panel data included in the analysis suggest that the studied objects, i.e. states, people, companies or countries, are heterogeneous. Panel data make it possible to control variables that cannot be observed or measured, for example those whose values change over time but not between the statistical objects. This means that such variables represent individual heterogeneity (Arellano & Honoré, 2000; Baltagi, 2005; Torres-Reyna, 2007; Kunst, 2010), thus representing one of the advantages of panel modeling.

The class of econometric models estimated on panel data, in which it is assumed that the evolution of the explained variable is influenced not only by a set of explanatory variables, but also by immeasurable, time-constant and object-specific factors - called group effects - is called the class of panel models. (Dańska-Borsiak, 2009b). Most often, these models are oriented towards cross-sectional analysis and their aim is to isolate differences between the studied objects, which are inextricably linked to factors specific to individual objects (Hsiao, 2007; Starzyńska & Grzelak, 2013). Panel data are characterized by a relatively high information value, greater volatility, less collinearity of variables, and provide a greater number of degrees of freedom (Baltagi, 2005; Hsiao, 2007; Kunst, 2010).

Econometric models estimated on panel data belong to the group of so-called models with decomposition (error component regression models). Panel models can be in the form of models with intercept decomposition (FEM - Fixed Effects Model) or models with random component decomposition (REM - Random Effects Model), and the decomposition can take into account only one factor (one-factor models) or two factors simultaneously (two-factor models) (Starzyńska & Grzelak, 2013). Models with fixed factors are often used in the case of long panels, i.e. with a small number of statistical units and a relatively long study period, which allow differences between units to be captured as differences in the intercepts (Maddala, 2006).

Naive analysis involves estimating a regression model taking into account all observations as if they were a set of cross-sectional data (Zwierzchowski, 2014). Therefore, the lack of individual effects is assumed, thus recognizing the homogeneity of the studied entities after taking into account differences in the available vector of observable variables and the lack of changes in the analysed phenomenon over time. With such assumptions, all observations can be treated as coming from a simple random sample and the ordinary least squares method can be used (Kunst, 2010; Zwierzchowski, 2014). In the panel model, observations are double indexed (Arellano & Honoré, 2000; Maddala, 2006; Hsiao, 2007; Kufel, 2007). The FEM model has the form:

$$y_{it} = \alpha + X_{it}\beta + u_i + \varepsilon_{it} \qquad i = 1, \dots, N;\ t = 1, \dots, T, \tag{2}$$

where:

i – given object,

t – time,

y_{it} – dependent variable,

α – scalar,

X_{it} – realizations of explanatory variables for the i-unit in t-period,

β – vector of dimension N of the model's structural parameters,

u_i – error containing an unobserved time-stable individual effect relating to the i-th unit of the panel,

ε_{it} – remainder disturbance.

In order to estimate the parameters of the FEM model (2), it is reduced to the following form:

$$y_{it} = \alpha + X_{it}\beta + v_{it} \quad v_{it} = u_i + \varepsilon_{it}, \tag{3}$$

where:

v_{it} – total random error consisting of the time-constant individual effect relating to the i-th unit of the panel (u_i),

ε_{it} – remainder disturbance.

Estimation of model parameters (3) using the ordinary least squares method is possible when individual effects do not occur and, consequently, the panel is treated as a set of cross-sectional data (Kufel, 2007). Model estimation is possible when the OLS estimator is consistent for the total error $E(v_{it}) = 0$, $Cov(v_{it}, x_{it}) = 0$, and for the remainder disturbance $E(\varepsilon_{it}) = 0$, $Cov(\varepsilon_{it}, x_{it}) = 0$ for $i = 1, \ldots, N$, $t = 1, \ldots, T$, and when there is no correlation between the individual effect u_i and explanatory variables x_{it} (Maddala, 2006). The null hypothesis about the existence of an individual effect is verified using the Breusch-Pagan test:

$H_0 : \sigma_u^2 = 0$ (the variance of the individual effects is zero),

$$H_1 : \sigma_u^2 \neq 0. \tag{4}$$

Rejecting the null hypothesis means a significant change in the variance when introducing individual effects, i.e. the justification for introducing individual effects into the model (4). In the case of an individual effect, two cases should be considered, i.e. fixed effects and random effects (5). The fixed effects model has the form:

$$y_{it} = X_{it}\beta + u_i + \varepsilon_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T, \tag{5}$$

where:

u_i – individual effect,

ε_{it} – remainder disturbance.

Individual effects may also take the form of random effects. In the case of a random effects model, we assume that the individual effects u_i are a random variable. Moreover, we know that they are not correlated with the remainder disturbances ε_{it}, $Cov(u_i, \varepsilon_{it}) = 0$, for $t = 1, \ldots, T$. The total random error, consisting of the individual effect (random effects) and remainder disturbances $v_{it} = u_i + \varepsilon_{it}$, is characterized by correlations in the same object. At the same time, it is assumed that there is no correlation for different objects. In this case, Generalized Least

Squares method should be used as the estimation method. The hypothesis about the existence of correlation between explanatory variables and random effects is verified using the Hausman test (6). The hypotheses have the following form:

$$\begin{aligned} H_0 &: Cov(u_i, x_{it}) = 0 \\ H_1 &: Cov(u_i, x_{it}) \neq 0 \end{aligned} \tag{6}$$

The lack of grounds to reject the null hypothesis indicates that both estimators are consistent – for fixed effects and random effects, with the random effects estimator being more effective. The rejection of the null hypothesis suggests the selection of fixed effects.

Taking into account the criteria of completeness and comparability of cross-sectional and time data, including changes in the statistical classification of economic activity in the European Union at the turn of 2007 and 2008, the research period 2008-2022 was adopted. Based on the substantive criterion, thirteen variables were initially classified into the macro panel data describing low and high-tech sectors in the EU-27: 1) employment in high-technology manufacturing and knowledge-intensive high-technology services, 2) employment in medium high-technology manufacturing, 3) employment in low and medium low-technology manufacturing, 4) employment in medium low-technology manufacturing, 5) employment in low-technology manufacturing, 6) employment in total knowledge-intensive services, 7) employment in knowledge-intensive high-technology services, 8) employment in knowledge-intensive activities, 9) employment in knowledge-intensive activities – business industries, 10) exports of high technology products as a share of total exports to all countries of the world, 11) total high-tech trade to all countries of the world, 12) share of research and development personnel and researchers, 13) patent applications, per million inhabitants, to the EPO (European Patent Office).

Statistical data (balanced macro panel) were extracted from the Eurostat database (www1). In order to determine the final set of variables, the procedure of eliminating statistically insignificant variables (p>0.05) was used within the considered specification of the panel model. On this basis, the final set of the following four variables was selected:

1. share of research and development personnel and researchers in total employment (*RD*_in_emp),
2. employment in high-technology manufacturing and knowledge-intensive high-technology services as a percentage of total employment (*HTEC*_emp_sectors),
3. employment in low-technology manufacturing as a percentage of total employment (*l_HTEC*_emp_man),

4. patent applications to the EPO, per million inhabitants, by country of applicants (*Pat*_app).

3. Results

The variable *RD*_in_emp – share of research and development personnel and researchers in total employment was defined as the dependent variable in the estimated panel model. The following were used as explanatory variables: employment in high-technology manufacturing and knowledge-intensive high-technology services (*HTEC*_emp_sectors), employment in low-technology manufacturing as a percentage of total employment (*l_HTEC*_emp_man), patent applications to the EPO, per million inhabitants, by country of applicants (*Pat*_app). In the one-way error component model used, the basis for the decomposition of the intercept/random component is the geographical location in the European Union (EU country). Models with the adopted specification were estimated in the following sequence: pooled regression without decomposition (see Tab. 1), one-way error component models: with intercept decomposition (FEM – Fixed Effects Model), model with random component decomposition (REM – Random Effects Model).

Tab. 1 Model Pooled OLS; dependent variable (Y): *RD*_in_emp, 405 observations, 27 cross-sectional units, time-series length = 15)

Variable	Coefficient	Standard error	*t*-ratio	*p*-value	
const	0.9023	0.1178	7.658	<0.0001	***
*HTEC*_emp_sectors	0.1005	0.0133	7.540	<0.0001	***
*l_HTEC*_emp_man	−0.0427	0.0113	−3.778	0.0002	***
*Pat*_app	0.0016	0.0001	10.73	<0.0001	***
Fit indices					
F(3, 401)		219.6089		<0.0001	
Adjusted R-squared		0.6216			
Joint significance of differing group means					
F(26, 375)		73.0867		<0.0001	
Breusch-Pagan test					
LM		1848.7		<0.0001	

Source: own estimation based on (www2).

All three explanatory variables included in the model are statistically significant ($p<0.05$). The pooled model, estimated using ordinary least squares (OLS), is a moderate fit to the empirical data. In order to determine whether the panel model can be estimated using OLS method, the hypothesis of the existence of an individual effect was verified. The total significance of inequality in group means is $F(26, 375) = 73.0867$ with p-value<0.0001 (see Tab. 1). A low p-value counts against the null hypothesis that the pooled OLS model is adequate. The null hypothesis should be rejected in favor of the H_1 hypothesis fixed effects alternative. As a result, the OLS estimation method cannot be used. The adoption of the alternative hypothesis (H_1) indicates the justification for introducing individual effects into the panel model (see Tab. 2).

Tab. 2 Fixed-effects model; dependent variable (Y): *RD*_in_emp, 405 observations, 27 cross-sectional units, time-series length = 15)

Variable	Coefficient	Standard error	*t*-ratio	*p*-value	
const	0.9738	0.1456	6.690	<0.0001	***
*m*HTEC_emp_sectors	0.1487	0.0148	10.02	<0.0001	***
*l*_HTEC_emp_man	−0.0683	0.0151	−4.518	<0.0001	***
*Pat*_app	0.0007	0.0002	3.193	0.0015	***
Fit indices					
LSDV F(29, 375)		194.4277		<0.0001	
LSDV R-squared		0.9377			
Joint test on named regressors					
F(3, 375)		87.8855		<0.0001	
Test for differing group intercepts (H_0: the groups have a common intercept)					
F(26, 375)		73.0867		<0.0001	

Source: own estimation based on (www2).

All three explanatory variables included in the model are statistically significant (p<0.05). The fixed-effects model fits the empirical data very well (R^2 = 0,94). The model with intercept decomposition described employment in R&D better than the model without decomposition based on the same data. The signs of the regression coefficients for the independent variables are consistent with economic theory. The higher the percentage of employment in high-technology manufacturing and knowledge-intensive high-technology services, the higher the share of research and development personnel and researchers in total employment. The higher employment in low-technology manufacturing, the lower the share of research and development personnel and researchers. The greater the number of patent applications to the EPO, per million inhabitants, in applicant country, the higher the share of research and development personnel and researchers in total employment. The statistics of the test for differentiation of the intercept indicate the justified use of a panel model with fixed effects.

The random effects model was estimated last. In order to verify the hypothesis about the lack of correlation between the explanatory variables and the random effect, the Hausman test was performed (Chi-square test statistic (3) = 19.117 with p-value = 0.0003). The result of the Hausman test confirms the need to reject the hypothesis about the consistency of both estimators i.e. fixed effects and random effects, and the choice of the fixed effects model.

4. Conclusions

The use of econometric panel models allowed to assess the relationship between the share of people employed in research and development activities and the share of employment in: high-technology manufacturing and knowledge-intensive high-technology services, low-technology manufacturing, patent applications to the EPO, per million inhabitants, by country of applicants. The applied panel models with intercept decomposition made it possible to estimate differences in employment in research and development activities resulting from the affiliation of the analyzed sector to a given EU-27 country. These differences are permanent. This means that the share of employment in research and development activities in the EU-27 depends not only on the explanatory variables mentioned above but also on the spatial unit, which is the analyzed country and its economic policy and other attributes.

References

Arellano, M. & Honoré, B. (2000). Panel Data Models: Some Recent Developments. *Working Paper*. 0016. CEMFI. November. 1-87. (ftp://www.cemfi.es/wp/00/0016.pdf).

Baltagi, B.H. (2005). Econometric Analysis of Panel Data. John Wiley & Sons Ltd. Chichester. 5-266.

Da Roit, B. & Iannuzzi, F. E. (2022). One of many roads to industry 4.0? Technology, policy, organisational adaptation and worker experience in 'Third Italy' SMEs. "*New Technology, Work and Employment*". Vol. 38. Issue 2. 252-271 (https://doi.org/10.1111/ntwe.12241).

Dańska-Borsiak, B. (2009a). Determinanty TFP według działów przemysłu w Polsce. Dynamiczna analiza panelowa. *Acta Universitatis Nicolai Copernici*. 389. 115-124.

Dańska-Borsiak, B. (2009b). Zastosowanie panelowych modeli dynamicznych w badaniach mikroekonomicznych i makroekonomicznych. *Przegląd statystyczny*. 2. 25-41.

Dorschel, R. (2022). Reconsidering digital labour: Bringing tech workers into the debate. "New Technology, Work and Employment". Vol. 37. Issue 2. 288-307 (https://doi.org/10.1111/ntwe.12225).

Hodder, A. (2020). New Technology, Work and Employment in the era of COVID-19: reflecting on legacies of research. *New Technology, Work and Employment*. Vol. 35. Issue 3. 262-275 (https://doi.org/10.1111/ntwe.12173).

Hsiao, C. (2007). Panel data analysis—advantages and challenges. *Invited Paper*. March 16th. 1-22. (https://pdfs.semanticscholar.org/843a/a44018e782840c07dc5f60d62a4ace740 746.pdf).

Kufel, T. (2007). Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL. PWN. Warszawa. 164-171.

Kunst, R.M. (2010). Econometric Methods for Panel Data. University of Vienna and Institute for Advanced Studies Vienna. 1-55. (http://homepage.univie.ac.at/robert.kunst/panpres.pdf).

Leonard, P. & Tyers, R. (2021). Engineering the revolution? Imagining the role of new digital technologies in infrastructure work futures. *New Technology, Work and Employment*. Vol. 38. Issue 2. 291-310. (https://doi.org/10.1111/ntwe.12226).

Maddala, G.S. (2006). Ekonometria. PWN. Warszawa. 655-672.

Schlogl, L., Weiss, E., & Prainsack B. (2021). Constructing the ‘Future of Work’: An analysis of the policy discourse. *New Technology, Work and Employment*. Vol. 36. Issue 3. 307-326 (https://doi.org/10.1111/ntwe.12202).

Starzyńska, W., Grzelak, M. (2013). Modele panelowe w analizach sektorowych na przykładzie działów przetwórstwa przemysłowego. 1-39. (https://stat.gov.pl/cps/rde/xbcr/lodz/ASSETS_ Konferencja_MRS_2013_Starzynska_W_Grzelak_M_Modele_panelowe_w_analizach_sektorowych.pdf).

Torres-Reyna, O. (2007). Panel Data Analysis Fixed and Random Effects using Stata (v. 4.2). *Data & Statistical Services*. Princeton University. December. 2-40. (https://www. princeton. edu/~otorres /Panel101.pdf).

Zwierzchowski, J. (2014). Wstęp do ekonometrii danych panelowych. 1-18. (https://www.e-sgh.pl/niezbednik/plik.php?id=27411519&pid=8610).

Online sources

(www1) https://www.cedefop.europa.eu/en/tools/skills-intelligence/trend-focus/digitalization-and-technology

(www2) https://ec.europa.eu/eurostat/data/database