Quantile regression for Panel data: An empirical approach for knowledge spillovers endogeneity

Aldieri, Luigi and Vinci, Concetto Paolo

Parthenope University of Naples, University of Salerno

January 2017

Quantile Regression for Panel Data:

An Empirical Approach for Knowledge Spillovers Endogeneity

Luigi ALDIERI•, Concetto Paolo VINCI •

Abstract

The aim of this paper is to investigate the extent to which knowledge spillovers effects are sensitive to different levels of innovation. We develop a theoretical model in which the core of spillover effect is showed and then we implement the empirical model to test for the results. In particular, we run the quantile regression for panel data estimator (Baker, Powell and Smith, 2016), to correct the bias stemming from the endogenous regressors in a panel data sample. The findings identify a significant heterogeneity of technology spillovers across quantiles: the highest value of spillovers is observed at the lowest quartile of innovation distribution. The results might be interpreted to provide some useful implications for industrial policy strategy.

JEL codes: O32, O33, C21

Keywords: Innovation; Spillovers; Quantile regression; Knowledge diffusion.

• Corresponding author: Department of Business and Economic Studies, University of Naples Parthenope, Italy, aldieri@uniparthenope.it, +390815474226.

•Department of economic and statistical sciences, University of Salerno, Italy, cpvinci@unisa.it.

1. Introduction

It is widely recognised that knowledge spillovers play a relevant role for firm competitiveness and economic growth. The idea of this work comes from the analysis of most empirical literature about this research topic (Jaffe, 1986; Griliches, 1992; Jaffe, Trajtenberg and Henderson, 1993; Verspagen, 1997; Capron and Cincera, 1998; Jaffe and Trajtenberg, 2002; Maurseth and Verspagen, 2002; Aldieri and Cincera, 2009; Bar and Leiponen, 2012; Malerba, Mancusi and Montobbio, 2013; Aldieri, 2013; Aldieri and Vinci, 2015 and 2016), where authors consider only the average effect of knowledge spillovers on the dependent variable, represented by innovation or productivity. Since we could assume that slope parameters may vary at various quantiles of the conditional distribution because of firms' heterogeneity, we implement a quantile regression model for panel data.

The results are very interesting in terms of policy implications of the industrial strategy. In particular, it is relevant to evaluate the innovation level of firms with respect to which we explore the spillovers effects. Indeed, the findings identify a significant heterogeneity of technology spillovers across quantiles: the highest value of spillovers is observed at the lowest quartile of innovation distribution.

The paper is structured as follows: Section 2 presents a theoretical model; in Section 3, we implement the quantile regression procedure and show the relative empirical findings; finally, Section 4 concludes.

2. Theoretical Model

This section is devoted to the analysis of the transmission of investment in R&D generated during the innovation process. We consider, in line with Acemoglu (1996), a simple non-overlapping generation model where each generation of two typologies of agents, living for two periods, is assumed to consist of a continuum of people normalized to unity and with a zero inter-temporal preference rate. People who collect physical capital, defined investors, while those who invest in R&D described as inventors. All of them are assumed to be risk-neutral.

At time *t=0* investors and inventors choose their investments respectively in physical and R&D capital, at time *t=1* knowledge production takes place in the form of a partnership of one inventor and one investor, agents will consume all their assets leaving no bequests. In their decisions inventors and investors will consider the stock of respectively R&D and physical capital and then die. The knowledge production takes the functional form:

$$P_{i,j,t} = Ah_{i,t}^{\alpha}k_{j,t}^{(1-\alpha)} \quad \text{with:} \quad 0 < \alpha < 1 \qquad (1)$$

where: A stands for a positive technology parameter, P the number of patents[1], h_i is the i.th inventor level of R&D, while k_j the physical capital level of the j.th investor.

Assuming randomness of the matching technology function in the sense that each inventor (investor) has the same probability of meeting an investor (inventor), and once a partnership has been formed it is too costly to break it up. The utility functions of the the two types of agents will be given by:

$$U_{i,t} = C_{i,t} - \frac{\lambda_i (1 + l_{i,t})^{(1+\gamma)}}{(1+\gamma)} R\,\&\,D_{t-1} \quad (2)$$

$$U_{j,t} = C_{j,t} - \frac{\theta_j (1 + e_{j,t})^{(1+\gamma)}}{(1+\gamma)} K_{t-1} \quad (3)$$

where: $\lambda_i(\theta_j)$ measures the disutility from the *R&D (K)* investments, $C_{i,t}$ and $C_{j,t}$ the consumption levels equal to the expected income of both the inventor an the investor, γ a taste positive parameter, $R\&D_{t-1}$ and K_{t-1} are the stock of *R&D* and *K* capital of the economy inherited from the previous generation defined respectively as follows:

$$R\,\&\,D_{t-1} = \int_0^1 h_{i,t-1}\, di \quad (4)$$

$$K_{t-1} = \int_0^1 k_{j,t-1}\, dj \quad (5).$$

Moreover, If the *i.th* inventor *R&D* capital level and the *J.th* investor *K* capital are given by:

$$h_{i,t} = (1-\delta)(1 + l_{i,t}) R\,\&\,D_{t-1} \quad (6)$$

$$k_{j,t} = (1-\delta)(1 + e_{i,t}) K_{t-1} \quad (7)^2$$

and that, following the standard mainstream on search models, the agents incomes derive from a bargaining process leading to a distribution rule according to given proportions β and $1-\beta$. From the f.o.c. of the maximization processes we may easily derive:

$$h_{i,t} = \left\{ \frac{A\beta\alpha(1-\delta)^{\gamma+1} R\,\&\,D_{t-1}^{\gamma} k_{j,t}^{\;(1-\alpha)}}{\lambda_i} \right\}^{\frac{1}{\gamma+1-\alpha}} \quad (8)$$

$$k_{j,t} = \left\{ \frac{A(1-\beta)(1-\alpha)(1-\delta)^{\gamma+1} K_{t-1}^{\gamma+1-\alpha} h_{i,t}^{\;\alpha}}{\lambda_i} \right\}^{\frac{1}{\gamma+\alpha}} \quad (9)$$

and state what follows:

Proposition: *Assuming* $\lambda_i = \lambda,\ \theta_j = \theta$ *:*

[1] We assume that patents' benefits are not unlimited because they depend on the law.

[2] δ is the capital depreciation rate

There exist positive externalities between R&D and physical capital. When a group of inventors (investors) increase investment in R&D (physical) capital, other agents, will respond, and the equilibrium rate of return of all subjects will improve. The positive externalities will be greater for lower levels of R&D (physical) capital.

The above result will imply that in case of different groups of inventors the positive externalities will have different magnitude depending on the initial level of *R&D* (Physical) capital implemented to generate patents. In order to measure quantitatively this magnitude, we introduce a quantile regression approach, whose estimates are presented in the following sections.

3. Empirical Strategy and results

Standard least squares method provides estimates based on the average effect of the independent variable on the average firm. In this analysis, we think that estimation of linear models by quantile regression is preferred. In this case, we may identify two advantages: quantile results are robust to outliers (Buchinsky, 1994) and quantile regression can describe the entire conditional distribution of the dependent variable (as discussed in Coad and Rao, 2011). Thus, we may assume that error terms are not identically distributed at all points of the conditional distribution and that slope parameters vary at different quantiles of the distribution, in line with the main result of theoretical model in the previous paragraph. According to Koenker and Bassett (1978), the quantile regression model is:

$$y_{it} = x_{it}^{'}\beta_{\theta} + u_{\theta it} \quad \text{with} \quad Quant_{\theta}(y_{it} / x_{it}) = x_{it}^{'}\beta_{\theta}$$

where *y* is the dependent variable, *x* is a vector of regressors, β is the vector of parameters to be estimated, *u* is a vector of residuals. $Quant_{\theta}(y_{it} / x_{it})$ identifies the θ^{th} conditional quantile of *y* given *x*.

In particular, we estimate a quantile regression model for panel data (QREGPD) with nonadditive fixed effects (Baker Powell, 2014 and 2016), maintaining the nonseparable disturbance term commonly associated with quantile estimation. The model is developed in an instrumental variable framework. The dataset used in the empirical analysis is the same as in Aldieri and Vinci (2015, 2016). The information on company profiles and financial statements comes from all EU R&D investment scoreboards editions issued every year until 2013 by the JRC-IPTS (European Commission, 2013). We select an unbalanced panel of 5951 observations for the period 2002-2010. For each firm, information is available for patents (P), the annual capital expenditures (C), the number of employees (L), annual R&D expenditures (R) and main industry sectors according to the Industrial Classification Benchmark (ICB) at the two digits level. OECD, REGPAT database, January 2012[3,4] is the second source of information used in this study. This database covers firms'

[3] See Maraut S., H. Dernis, C. Webb, V. Spieazia and D. Guellec (2008) for the methodology used for the construction of REGPAT.

[4] Please contact Helene.DERNIS@oecd.org to download REGPAT database.

patent applications to the European Patent Office (EPO) including patents published up to December 2011. By using the same dataset relative to international firms as in Aldieri and Vinci (2015), we apply the quantile regression technique to the following knowledge production function:

$$LnP_{it} = \beta_0 + \beta_1 \ln C_{it} + \beta_2 \ln L_{it} + \beta_3 K_{it} + \beta_4 TS_{it} + \beta_5 AbCap_{it} + u_{it} \quad (10)$$

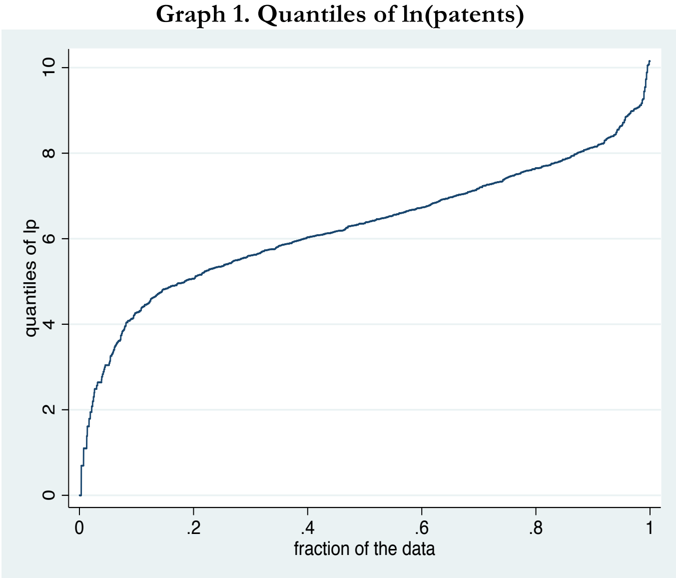
where P_{it} is the number of patents of firm i at time t, C_{it} is physical capital of firm I at time t, L_{it} is the number of employees, K_{it} represents the stock of knowledge capital computed by perpetual method on R&D expenditures, TS_{it} is total stock of spillovers computed as the weighted sum of other R&D capital stock. For the computation of spillovers, we implement the Jaffe (1986) procedure, based on the proximity between the technological vectors of firms (as in Aldieri, 2011 and 2013; Aldieri and Vinci, 2015 and 2016). All variables are measured in logarithmic terms. In Table 1, we display the descriptive statistics of the variables employed for the empirical estimation. Graph 1 shows the quantiles of dependent variable, ln(patents), and indicates symmetry.

Table 1. Summary statistics

Variable	Mean[a]	Std. Dev.
lnP	6.25	1.646
lnL	9.47	1.552
lnC	7.05	1.796
lnK	6.58	1.453
lnTS	12.89	0.428

Note: a) 5,951 observations;

Graph 1. Quantiles of ln(patents)



In table 2, we show the quantile regression with bootstrapped standard errors (Koenker, 2005) results. As expected, size, physical capital, R&D capital and knowledge spillovers affect positively the innovation procedure of firms at all quantiles of the conditional distribution. The findings identify a significant heterogeneity of technology spillovers across quantiles: the highest value of spillovers is observed at the lowest quartile of innovation distribution.

Table 2. Results: simultaneous quantile regression with bootstrapped standard errors

Dependent variable: ln P		Quantile: 10%	
		Estimate[a]	S.E.
lnL		-0.05	(0.036)
lnC		0.42***	(0.038)
lnK		0.24***	(0.039)
lnTS		2.14***	(0.146)
Pseudo-R^2	0.44		
		Quantile: 20%	
		Estimate[a]	S.E.
lnL		0.01	(0.026)
lnC		0.33***	(0.027)
lnK		0.29***	(0.027)
lnTS		1.44***	(0.089)
Pseudo-R^2	0.42		
		Quantile: 30%	
		Estimate[a]	S.E.
lnL		0.08***	(0.028)
lnC		0.25***	(0.034)
lnK		0.33***	(0.027)
lnTS		1.19***	(0.105)
Pseudo-R^2	0.42		

Notes: ***, **, * coefficient significant at the 1%. Country, time and industry dummies are included.

Dependent variable: ln P		Quantile: 40%	
		Estimate[a]	S.E.
lnL		0.08***	(0.023)
lnC		0.23***	(0.022)
lnK		0.37***	(0.018)
lnTS		1.06***	(0.038)
Pseudo-R^2	0.42		
		Quantile: 50%	
		Estimate[a]	S.E.
lnL		0.08***	(0.020)
lnC		0.22***	(0.021)
lnK		0.38***	(0.020)
lnTS		1.00***	(0.060)
Pseudo-R^2	0.43		
		Quantile: 60%	
		Estimate[a]	S.E.
lnL		0.12***	(0.025)
lnC		0.20***	(0.022)
lnK		0.35***	(0.022)
lnTS		0.95***	(0.067)
Pseudo-R^2	0.43		

Notes: ***, **, * coefficient significant at the 1%. Country, time and industry dummies are included.

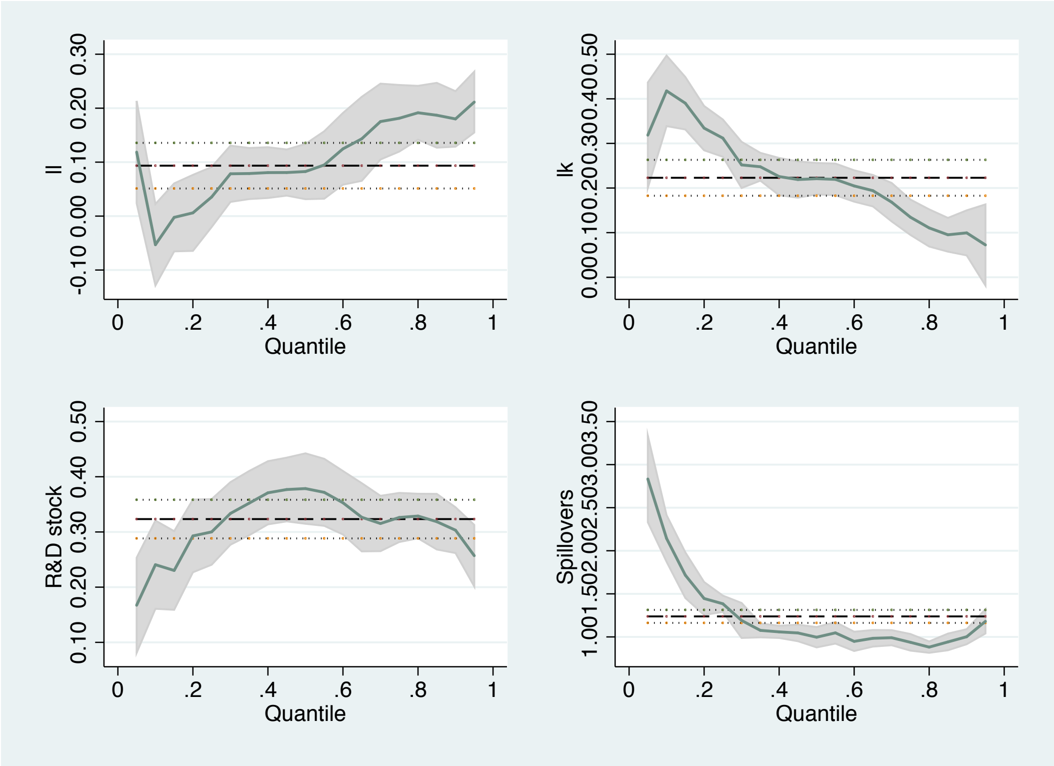
Dependent variable: ln P		Quantile: 70%	
		Estimate[a]	S.E.
lnL		0.18***	(0.028)
lnC		0.17***	(0.027)
lnK		0.32***	(0.022)
lnTS		0.99***	(0.053)
Pseudo-R^2	0.43		
		Quantile: 80%	
		Estimate[a]	S.E.
lnL		0.19***	(0.023)
lnC		0.11***	(0.024)
lnK		0.33***	(0.022)
lnTS		0.88***	(0.042)
Pseudo-R^2	0.43		
		Quantile: 90%	
		Estimate[a]	S.E.
lnL		0.18***	(0.023)
lnC		0.10***	(0.022)
lnK		0.30***	(0.021)
lnTS		1.00***	(0.048)
Pseudo-R^2	0.43		

Notes: ***, **, * coefficient significant at the 1%. Country, time and industry dummies are included.

In order to verify statistical variation of coefficients along innovation conditional distribution, we depict a graphic display of coefficients of interest. In Graph 2, we produce separate graphs for each regressor of the estimated coefficient plotted against the quantile q.

The horizontal lines are the OLS point estimates and confidence intervals (these do not vary with the quantile). The second plot shows that the coefficient on knowledge spillovers is positive, with a much larger effect at lower quantiles.

Graph 2. QR coefficients and confidence intervals as *quantile* varies from 0 to 1



Moreover, we also implement a Wald statistical test to control that coefficients on the spillovers variable have the same value:

Table 4. Wald Test of coefficient equality across different quantiles

F(8, 5921) = 12.20
Prob > F = 0.0000

The null hypothesis of coefficient equality is rejected at a level of 0.05, by confirming the graphical analysis.

As discussed in the empirical literature (Parente and Santos Silva, 2016), it is possible to identify intra-cluster correlation in event of data sampled from independent and identically distributed clusters. In our sample, the Parente-Santos Silva test demonstrates the intra-cluster correlation[5]. Thus, the consistency of quantile estimator could be questionable. For this reason, we develop a quantile regression with clustered standard errors (Machado, Parente and Santos Silva, 2011) in Table 5.

[5] Parente-Santos Silva test results for intra-cluster correlation can be provided from the authors upon request.

Table 5. Results: quantile regression with robust and clustered standard errors (QREG2)

Dependent variable: ln P		Quantile: 10%	
		Estimate[a]	S.E.
lnL		0.20*	(0.105)
lnC		0.15**	(0.079)
lnK		0.29***	(0.074)
lnTS		2.14***	(0.385)
Pseudo-R^2	0.43		
		Quantile: 20%	
		Estimate[a]	S.E.
lnL		0.20**	(0.080)
lnC		0.11	(0.094)
lnK		0.31***	(0.104)
lnTS		1.12***	(0.246)
Pseudo-R^2	0.45		
		Quantile: 30%	
		Estimate[a]	S.E.
lnL		0.17***	(0.063)
lnC		0.17***	(0.062)
lnK		0.31***	(0.060)
lnTS		0.94***	(0.083)
Pseudo-R^2	0.45		

Notes: ***, **, * coefficient significant at the 1%, 5%, 10% level respectively. Country, time and industry dummies are included.

Dependent variable: ln P		Quantile: 40%	
		Estimate[a]	S.E.
lnL		0.16**	(0.064)
lnC		0.22***	(0.075)
lnK		0.28***	(0.071)
lnTS		0.89***	(0.171)
Pseudo-R^2	0.45		
		Quantile: 50%	
		Estimate[a]	S.E.
lnL		0.12*	(0.067)
lnC		0.27***	(0.083)
lnK		0.26***	(0.059)
lnTS		0.85***	(0.096)
Pseudo-R^2	0.45		
		Quantile: 60%	
		Estimate[a]	S.E.
lnL		0.12*	(0.063)
lnC		0.25***	(0.080)
lnK		0.25***	(0.067)
lnTS		0.85***	(0.131)
Pseudo-R^2	0.45		

Notes: ***, **, * coefficient significant at the 1%, 5%, 10% level respectively. Country, time and industry dummies are included.

Dependent variable: ln P		Quantile: 70%	
		Estimate[a]	S.E.
lnL		0.15***	(0.052)
lnC		0.20***	(0.064)
lnK		0.23***	(0.058)
lnTS		0.84***	(0.113)
Pseudo-R^2	0.45		
		Quantile: 80%	
		Estimate[a]	S.E.
lnL		0.16***	(0.050)
lnC		0.20***	(0.059)
lnK		0.22***	(0.052)
lnTS		0.81***	(0.107)
Pseudo-R^2	0.45		
		Quantile: 90%	
		Estimate[a]	S.E.
lnL		0.12**	(0.058)
lnC		0.24***	(0.054)
lnK		0.20***	(0.036)
lnTS		0.92***	(0.096)
Pseudo-R^2	0.43		

Notes: ***, **, * coefficient significant at the 1%, 5%, 10% level respectively. Country, time and industry dummies are included.

As we may observe from Table 5, results of quantile regression concerning knowledge spillovers confirm the previous ones and thus, they are robust with respect to intra-cluster correlation.

Table 6. Results: quantile regression for panel data (QREGPD)

Dependent variable: ln P	Quantile: 10%	
	Estimate[a]	S.E.
lnL	0.19	(0.261)
lnC	0.25	(0.484)
lnK	0.09	(0.271)
lnTS	3.64***	(1.246)

	Quantile: 20%	
	Estimate[a]	S.E.
lnL	-1.35	(10.564)
lnC	-1.11	(10.413)
lnK	-2.41	(13.556)
lnTS	0.63	(4.203)

Notes: ***, **, * coefficient significant at the 1%. Country, time and industry dummies are included.

Instrumental variables: LnL(t-1); lnL(t-2); lnC(t-1); lnC(t-2); lnK(t-1); lnK(t-2); lnTS(t-1); lnTS(t-2).

Finally, the endogeneity issue in the implemented model could be questionable because of simultaneity of decision precesses. For this reason, we estimate a quantile regression for panel data (Baker, Powell and Smith, 2016; Powell, 2014 and 2016), where we use lagged explanatory variables as instruments in Table 6. It is sufficient to explore the results of the two quantiles (10% and 20%) to verify the robustness of the previous findings.

4. Policy implications and concluding remarks

In this paper, we investigate firms' innovation, measured by patents (OECD, REGPAT database, 2012), with a particular attention on the role of knowledge spillovers. In order to carry out our analysis, we apply different econometric techniques on an unbalanced panel of 5951 observations for the period 2002-2010. Firms' data relative to three economic areas (the USA, Japan and Europe) come from all EU R&D investment scoreboards editions issued until 2013 by the JRC-IPTS (European Commission, 2013).

In order to better analyse the relationship between firm's innovation and R&D spillovers, we study whether such relationship varies along the innovation distribution by applying quantiles regression techniques. In particular, we run a quantile regression model with bootstrapped standard errors (Koenker, 2005) and a quantile regression asymptotically valid under heteroskedasticity and intra-cluster correlation (Machado, Parente and Santos Silva, 2011; Parente and Santos Silva, 2016). Finally, we deal with the endogeneity of variables, by applying the quantile regression for panel data (Baker, Powell and Smith, 2016; Powell, 2014 and 2016).

Empirical results suggest that R&D spillovers positively affect firms' innovation. In particular, such effect has a bell shaped pattern along the innovation conditional distribution, with firms whose innovation ranges between values of about 10% and 20% exhibiting higher returns from R&D spillovers, while returns are found to be lower for successive quantiles.

Overall results suggest that R&D spillovers components help to explain heterogeneity in innovation capacity and might provide useful insights for the design of policy instruments aimed at favouring productivity improvements.

References

Acemoglu D. A. (1996). Microfoundation for social increasing returns in human capital accumulation. The Quarterly Journal of Economics, 111, 779-804.

Aldieri L. (2011). Technological and geographical proximity effects on knowledge spillovers: evidence from the US patent citations. Economics of Innovation and New Technology, 20, 597-607.

Aldieri L. (2013). Knowledge technological proximity: evidence from US and European patents. Economics of Innovation and New Technology, 22, 807-819.

Aldieri, L. and Cincera, M. (2009). Geographic and technological R&D spillovers within the triad: micro evidence from US patents. The Journal of Technology Transfer, 34, 196-211.

Aldieri L. and Vinci C. P. (2015). Knowledge migration: A cross-national analysis, De Economist, DOI 10.1007/s10645-015-9269-5;

Aldieri L. and Vinci C. P. (2016). Technological spillovers through a patent citation analysis, International Journal of Innovation Management, 20 (2), DOI: 10.1142/S1363919616500286;

Baker M., Powell D. and Smith T. A. (2016). QREGPD: Stata module to perform Quantile Regression for Panel Data. Statistical Software Components S458157, Boston College Department of Economics.

Bar, T., & Leiponen, A. (2012). A measure of technological distance. Economics Letters, 116, 457-459.

Buchinsky, M. (1994). Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression. Econometrica 62, 405-458.

Capron H. and M. Cincera, (1998), Exploring the spillover impact on productivity of world-wide manufacturing firms, Annales d'Economie et de Statistiques, 49-50: 565-587.

Chernozhukov V., Fernandez-Val I. and Kowalski A. E. (2015). Quantile regression with censoring and endogeneity. *Journal of Econometrics,* 186, 201-221.

Coad A. and Rao R. (2011). The firm-level employment effects of innovations in high-tech US manufacturing industries. Journal of Evolutionary Economics, 21, 255-283.

Cohen W. M and Levinthal D. A. (1989). Innovation and Learning: the two faces of R&D. The Economic Journal, 99, 569-596.

European Commission (2013). The 2013 EU Industrial R&D Investment Scoreboard, JRC Scientific and Technical Research series. http://iri.jrc.ec.europa.eu/scoreboard.html;

Griliches, Z. (1992). The search for R&D spillovers. The Scandinavian Journal of Economics, 94, S29-S47.

Jaffe A. B. (1986). Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value. American Economic Review, 76, 984–1001.

Jaffe, A. B., & Trajtenberg, M. (2002). Patents, citations, and innovations: A window on the knowledge economy. MIT press.

Jaffe, A., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. The Quarterly Journal of Economics, 108, 577-598.

Koenker, R. and Bassett, G., (1978). Regression Quantiles. Econometrica 46, 33-50.

Koenker R. (2005). Quantile regression. Cambridge University Press: New York.

Parente P.M.D.C and Santos Silva J.M.C. (2016). Quantile Regression with Clustered Data. *Journal of Econometric Methods*, 5(1), 1-15.

Machado J.A.F., Parente P.M.D.C and Santos Silva J.M.C. (2011). QREG2: Stata module to perform quantile regression with robust and clustered standard errors. Statistical Software Components S457369, Boston College Department of Economics, revised 07 Feb 2015.

Malerba, F., Mancusi, M. L., & Montobbio, F. (2013). Innovation, international R&D spillovers and the sectoral heterogeneity of knowledge flows. Review of World Economics, 149, 697-722.

Maurseth P. B. and Verspagen B. (2002). Knowledge spillovers in Europe: A patent citation analysis. Scandinavian Journal of Economics, 104, 531-545.

OECD, REGPAT database (accessed January 1, 2012). ftp://prese:Patents@ftp.oecd.org/ REGPAT_201301/.

Powell D. (2014). Did the Economic Stimulus Payments of 2008 Reduce Labor Supply? Evidence from Quantile Panel Data Estimation. RAND Labor and Population Working Paper 710-3.

Powell D. (2016). Quantile Regression with Nonadditive Fixed Effects. Available at: http://works.bepress.com/david_powell/1/

Powell D. (2016). Quantile Treatment Effects in the Presence of Covariates. RAND Corporation. Available at: http://works.bepress.com/david_powell/4/

Verspagen, B. (1997). Measuring intersectoral technology spillovers: estimates from the European and US patent office databases. Economic Systems Research, 9, 47-65.