ISSN: 2281-1346

Department of Economics and Management

**DEM Working Paper Series**

# R&D, innovation and knowledge spillovers: An empirical reappraisal based on cross sectional dependence

Anna Bottasso
(Università di Genova)

Carolina Castagnetti
(Università di Pavia)

Maurizio Conti
(Università di Genova)

**# 45 (06-13)**

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

**June 2013**

# R&D, innovation and knowledge spillovers: An empirical reappraisal based on cross sectional dependence.

Anna Bottasso* Universitá di Genova
Carolina Castagnetti † Universitá di Pavia
Maurizio Conti ‡ Universitá di Genova

June 20, 2013

## Summary

Bottazzi and Peri (2007) show that the existence of a cointegrating relationship between the domestic stock of knowledge, domestic R&D and the international knowledge stock can be interpreted as a support of the semi-endogenous versus the endogenous growth models. We replicate their study in a wide sense by using three more countries, a more recent time period, a different measure of R&D resources and by estimating the cointegrating vector with an econometric methodology that is robust to cross sectional dependence. Our replication confirms Bottazzi and Peri's main results in favour of the semi-endogenous growth models but finds stronger spillover effects.

## 1 Introduction

In a recent paper, Bottazzi and Peri (2007), hereafter BP, propose a test to discriminate between the endogenous growth models à la Romer (1990) and Aghion and Howitt (1992) and the semi-endogenous growth models à la Jones (1995) and Segerstrom (1998). The test is based on the consideration that the former class of models are characterized by *strong scale effects*, whereby a country level of R&D resources is positively correlated to the growth rate of its technological knowledge; on the other side, semi-endogenous growth models display the so called *weak scale effects*, i.e. a positive correlation between the level of resources a country devotes to R&D and its level of technological knowledge.

In this note we assess the robustness of BP's results to the use of a larger set of countries, a slightly different time span, a different proxy for R&D activity and, most importantly, a different econometric technique that we argue is more appropriate to tackle the econometric issues involved in such an empirical framework: therefore this note can be considered as a replication in a wide sense of BP.

Turning to the theoretical framework that underpins BP's test, the authors start from a very simple neoclassical production function $y_t = BA_t^{\sigma} k_t^{\alpha}$, where $y_t$ and $k_t$ are output and physical capital per worker, respectively; $B$ is a term capturing the joint effects of (time-invariant) factors that might influence production efficiency (such as institutions, geography, etc.), $A$ is the country stock of knowledge, while $\sigma$ and $\alpha < 1$ are the elasticities of output with respect to the stock of knowledge and to the physical capital (per worker), respectively. Given the hypothesis of decreasing returns to physical capital, the growth model converges to a balanced growth path where GDP per worker grows only through the knowledge accumulation process, at a rate given by $g_y^* = \frac{\sigma}{1-\alpha} g_A^*$, where $g^*$denotes variables' steady state growth rate.

Moreover, BP assume the existence of a knowledge sector characterized by the production function $I_{it} = F(R\&D_{it}, A_{it}, AROW_{it})$ whereby the generation of new knowledge ( $I_{it}$) depends upon the

*Dipartimento di Economia e Metodi Quantitativi, Via Vivaldi 5, 16126, Genova, Italy.
†Dipartimento di Economia Politica e Metodi Quantitativi, Via San Felice 5, 27100 Pavia, Italy, castca@eco.unipv.it, tel.++390382986217, corresponding author
‡Dipartimento di Economia e Metodi Quantitativi, Via Vivaldi 5, 16126, Genova, Italy.

resources devoted to research and development ($R\&D_{it}$) in country $i$ as of time $t$ (measured as employment in the private R&D sector), the stock of knowledge ($A_{it}$) in the country (measured as the accumulated stock of ideas up to year *t-1* and available at the beginning of year $t$) and the stock of knowledge in the rest of the world ($AROW_{it}$) accumulated up to year *t-1*: while the effect of $R\&D_{it}$ on $I_{it}$ is assumed to be positive, the effect of both $A_{it}$ and $AROW_{it}$ could be either positive (*standing on the giants shoulders' effect*) or negative (*fishing out effect*).

BP proxy the stock of new knowledge ($I_{it}$) with the number of new patent applications filed at the USPTO and assume that the two variables are linked by the following relationship: $Pat_{it} = \varkappa_i I_{it}$, where $\varkappa_i$ stands for the country specific propensity to patent new ideas. By assuming that the knowledge production function is linear in logs and that the stock of knowledge $A_{it}$ evolves according to the law of motion $A_{it+1} = Pat_{it} + (1-\delta)A_{it}$, after some manipulations, BP (see their equations 4 and 5) derive the following equation:

$$\ln(g_{Ait} + \delta) - \ln \varkappa_i = (\phi - 1)\ln A_{it} + \lambda \ln R\&D_{it} + \xi \ln AROW_{it} \quad (1)$$

where $\delta$ is the depreciation rate, while $\lambda$, $\phi$ and $\xi$ represent the elasticity of patents with respect to R&D resources, the stock of domestic knowledge and the stock of knowledge in the rest of the world, respectively.

If the economy converges to a stochastic balanced growth path, then $\ln(g_{Ait} + \delta) - \ln \varkappa_i$ converges to a country-specific stationary stochastic process and therefore there must be a stationary long run relationship among the variables in the right hand side of equation (1). In other words, if the left hand side of (1) is stationary and $\ln A_{it}$, $lnR\&D_{it}$ and $lnAROW_{it}$ are non-stationary, there need to be a cointegrating relationship between them: the existence of such a cointegrating relationship supports the semi-endogenous growth models hypothesis.

After testing for the cointegration relationship, BP estimate the following long run empirical model:

$$\ln A_{it} = \mu \ln R\&D_{it} + \gamma \ln AROW_{it} + s_{it} \quad (2)$$

where $\mu = \lambda/(1-\phi)$; $\gamma = \xi/(1-\phi)$ and $s_{it}$ contains all deterministic and stochastic elements of the cointegrating relationship (e.g. the country-specific fixed effect $\varkappa_i$).

BP test for the stationarity of the rate of growth of knowledge creation using the Pesaran (2007) panel unit root test, which is robust to cross sectional correlation across countries, and could not reject the null hypothesis of non-stationarity. As an additional robustness check, they tested if the residuals from (2) are stationary using the LR-bar test statistics proposed by Larsson, Lyhagen, and Lothgren (2001) which again led them to conclude that $\ln A_{it}$, $lnR\&D_{it}$ and $lnAROW_{it}$.are indeed cointegrated with cointegration rank equal to one. Finally, authors estimate equation (2) for a set of 15 OECD countries observed over the period 1972-1999 by means of the Dynamic OLS (DOLS) approach proposed by Mark and Sul (2001). Empirical results suggest large, positive and statistically significant elasticities of both $R\&D_{it}$ and $AROW_{it}$, with the latter exhibiting higher values with respect to the knowledge elasticity of $R\&D_{it}$: such results are consistent with the *standing on the giants shoulders'* hypothesis.

# 2 Data

We carry out the empirical analysis on the same set of countries considered by BP (i.e. Australia, Canada, Germany, Denmark, Spain, Finland, France, Great Britain, Ireland, Italy, Japan, The Netherlands, Norway, Sweden and USA), plus Austria, Belgium and New Zealand. While BP focus on the period 1972-1999, we consider, for data availability reasons, the period 1981-2006. BP measure R&D resources with employment in the private R&D sector; in order to maximize the sample size we have used the Gross Domestic Expenditure in R&D (GERD), converted in constant PPP$.[1] The data for GERD are from Eurostat while, in the case of missing data, we have used information either from the OECD-STAN database or, for some countries, from Eurostat on BERD.[2] Only in the case of

[1]The results are robust to the use of Business Expenditure on R&D, BERD.

[2]We have proxied the missing GERD figure in year $t$ by assuming that the rate of growth of BERD between year *t-1* and $t$ was the same of GERD.

Sweden, Norway and New Zealand the amount of missing data was substantial and we had to rely on linear interpolation: for this reason, as a robustness check we have also performed our analysis after excluding these countries. As far as it concerns the derivation of the domestic stock of knowledge, we have closely followed BP by applying the perpetual inventory method to the number of citation-weighted patents[3] filed in each year at the USPTO; the base year for the stock of knowledge was computed as $A_{i,1976} = \frac{Pat_{i,1976}}{\overline{g_i}+\delta}$, where $Pat_{i,1976}$ is the number of citation-weighted patents in 1976, $\delta$ is the depreciation rate assumed equal to 0.1 and $\overline{g_i}$ is the average rate of growth of the number of citation-weighted patents over the period 1976-1981. Finally, $AROW_{it}$ was computed as in BP as: $AROW_{it} = \sum_{j\neq i} A_{it}$.

# 3 Methodological framework and empirical results

In Table 1 we have reported unit root tests for the variables $\ln A_{it}$, $lnR\&D_{it}$ and $lnAROW_{it}$. As in BP, we first report the Pesaran (2007) CADF test which is robust to the existence of cross sectional correlation arising from omitted variables, macroeconomic-wide shocks or spatial spillovers effects, among others. The CADF tests show that for each variable we can never reject the null hypothesis of non-stationarity (against the alternative of stationarity) at conventional confidence levels. To take into account the presence of cross sectional dependence in the data we have also carried out the PANIC analysis of Bai and Ng (2004) which allows each series to be composed of two components, namely a vector of common factors and an idiosyncratic error term. This approach allows for the possibility that a series is nonstationary if either the common factor, the idiosyncratic component or both are non-stationary. The pooleed test statistics for the hypothesis that the idiosyncratic components are non-stationary ($Z_e^\tau$ in Table 1) rejects the null of unit root for all variables and the PANIC analysis shows the presence of non stationarity in the unobserved common factors only. Results reported in the last column of Table 1 suggest that $\ln A_{it}$ and $lnR\&D_{it}$ present two independent stochastic trends, while in the case of $lnAROW_{it}$ we can not reject the hypothesis of four independent stochastic trends. Such results suggest that the non-stationarity is due to a pervasive source so that conventional methodologies adopted to estimate the cointegrating relationship, like those employed by BP, are not appropriate.

As in BP, we evaluate the stationarity of the growth rates of the three variables of interest by applying the Pesaran CADF test statistics: the large and negative test statistics reported in the bottom part of Table 1 suggest to reject, at conventional confidence levels, the null hypothesis of non-stationarity for all variables. In particular, the stationarity of $\Delta \ln A_{it}$ confirms BP results and supports the hypothesis of the presence of weak scale effects in the underlying growth model.

Table 1: Unit root tests

| | Pesaran CADF | Bai-Ng (2004) unit root tests | | | |
|---|---|---|---|---|---|
| Variable | CADF | # of factors | $Z_e^\tau$ | $MQ_f^\tau$ | # of $factors^{**}$ |
| lnA | -2.092 | 3 | 4.471 | 2.087 | 2 |
| lnAROW | -0.133 | 4 | 9.120 | -19.342 | 4 |
| lnR&D | -1.991 | 2 | 8.879 | -19.378 | 2 |
| $\Delta lnA$ | -3.008 | | | | |
| $\Delta lnAROW$ | -4.347 | | | | |
| $\Delta lnR\&D$ | -3.554 | | | | |

*Note: the suffices $\tau$ for the statistic $Z$ and $MQ_f$ indicate the intercept and linear trend case. The BIC3 criterion of Bai and Ng (2002) was used to estimate the number of unobserved common factors. $**$: number of factors estimated by the $MQ_f$ statistics. The CADF (CIPS) test statistics corresponds to the intercept and linear time trend case.*

[3] Citations have been corrected for truncation using information contained in the most recent release of the NBER patent database.

As an additional robustness check, we have tested whether the residuals of equation (2) are stationary. For this purpose BP used the battery of tests proposed by Pedroni (1999) as well as the LR-bar test of Larsson, Lyhagen, and Lothgren (2001). The results we have obtained for our sample are very close to those obtained by BP. Moreover, both the Larsson, Lyhagen, and Lothgren (2001) and the Larsson and Lyhagen (2000) testing procedures establish that there is a common cointegrating rank equal to 1 for all countries in the panel.

However, these approaches do not take into account the presence of cross sectional dependence. Banerjee, Marcellino, and Osbat (2004) have shown that neglecting a common factor structure in the data might have non negligible effects on the cointegration testing procedure. For this reason we have decided to follow the procedure outlined in Gengenbach, Palm, and Urbain (2006). Given that the tests of Bai and Ng (2004) suggest that non-stationarity in the three variables is due to a number of common stochastic trends, cointegration between them can occur only if the common factors of $\ln A_{it}$ cointegrate with those of $lnR\&D_{it}$ and $lnAROW_{it}$. In this case, Gengenbach, Palm, and Urbain (2006) propose to test for the null of no cointegration between the factors using the Johansen LR test. Empirical results reported in Table 2 confirm the conclusion of BP on the existence of a single cointegrating relationship.

Table 2: Gengenbach, Palm, and Urbain (2006) cointegration test

| Trace test statistic | Critical value | Cointegration rank |
|---|---|---|
| 251.75 | 222.21 | 0 |
| 177.27** | 182.82 | 1 |
| 134.92 | 146.76 | 2 |
| 94.86 | 114.90 | 3 |
| 64.78 | 87.31 | 4 |
| 40.36 | 62.99 | 5 |
| 23.08 | 42.44 | 6 |
| 10.97 | 25.32 | 7 |
| 4.12 | 12.25 | 8 |

*Note: All tests are perfomed at the 5% level. The results are from the 9 common factors extracted from the variables lnR&D, lnA and LnAROW using the approach proposed by Bai and Ng (2004).*

On the basis of the PANIC analysis, for the estimation of equation (2) we rely on the Bai, Kao, and Ng (2009) approach and apply the Continuously Updated and bias corrected Fully Modified (CUP-FM) model.[4] Such approach allows us to properly estimate a cointegrating vector in a panel data framework where non-stationarity rests on common factors.

Estimate results shown in the first column of Table 3 suggest that both R&D and the stock of knowledge in the rest of the world have positive and statistically significant coefficients: an increase in R&D expenditure by 1% is associated in the long run to an increase of the domestic stock of knowledge of about 0.55%; furthermore, an increase of 1% in the international stock of knowledge tends to increase the domestic stock of knowledge by about 0.56%. In particular, the positive coefficient of $lnAROW_{it}$ denotes the existence of large and significant spillover effects which in turn is evidence in favour of the *standing on the giants shoulders' effect* cited in the introduction. It is important to note that estimating the same regression with DOLS yields coefficients for $lnR\&D_{it}$ and $lnAROW_{it}$ of about 0.72 and 0.60, respectively, which are very similar to those reported in BP. In model II we insert country specific time trends and we find that the elasticities of $lnR\&D_{it}$ and the international stock of knowledge fall and increase, respectively. In Model III and IV we estimate the same regressions but without weighting patents with citations and results are barely altered. Finally, our main findings are confirmed when we drop Norway, Sweden and New Zealand from the sample.

[4]See Bottasso, Castagnetti, and Conti (2013) for one of the very few empirical applications.

Table 3: Estimates of the long run cointegration relationship. Cup-FM

| Dependent variable lnA | I: Basic | II: country trends | III: Un-weighted patents to build A | IV: Un-weighted patents to build A | V: omitting No, Swe, Nz |
|---|---|---|---|---|---|
| lnR&D | 0.547 | 0.282 | 0.579 | 0.293 | 0.476 |
| | (13.47) | (10.11) | (13.89) | (10.44) | (10.15) |
| lnAROW | 0.56 | 0.71 | 0.570 | (0.714) | 0.677 |
| | (12.82) | (35.89) | (13.02) | (35.42) | (10.24) |
| Country fixed effects | Yes | Yes | Yes | Yes | Yes |
| Country trends | No | Yes | No | Yes | Yes |
| Observations | 450 | 450 | 450 | 450 | 375 |
| N. of factors | 4 | 4 | 4 | 4 | 4 |

*Note: Number of countries:18 in Models (I-IV), 15 in Model V. Estimation method: CUP-FM of Bai, Kao, and Ng (2009). In Models I-II and V the stock of knowldege is estimated with citation-weighted patents, while row patents are used in Models III-IV. More details in the data section. Number of common factors estimated using the BIC3 criterion of Bai and Ng (2002). t statistics in parenthesis.*

# 4 Conclusion

In this note we provide a replication in a wide sense of Bottazzi and Peri (2007)'s work on the long run relationship between R&D activity, the domestic and the international stock of knowledge. Our results suggest that their findings do not hinge neither upon the sample nor on the adopted estimation methodology; moreover, even measuring R&D with a monetary measure instead of employees in the R&D sector does not seem to affect the analysis. However, our results suggest that the international stock of knowledge might have a stronger impact on domestic knowledge if compared to R&D expenditure: the magnitude of knowledge spillovers from the rest of the world as a driver of domestic knowledge increases when we take into account the existence of cross sectional correlation in the data.

# References

Aghion, P., and P. Howitt (1992): "A model of growth through creative distruction," *Econometrica*, 60, 323–51.

Bai, J., C. Kao, and S. Ng (2009): "Panel Cointegration with Global Stochastic Trends," *Journal of Econometrics*, 149, 82–99.

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

——— (2004): "A Panic Attack on Unit Root and Cointegration," *Econometrica*, 72, 1127–1177.

Banerjee, A., M. Marcellino, and C. Osbat (2004): "Some cautions on the use of panel methods for integrated series of macroeconomic data," *Econometrics Journal*, 7, 322–340.

Bottasso, A., C. Castagnetti, and M. Conti (2013): "And yet they Co-move! Public capital and productivity in OECD," *Journal of Policy Modeling*, forthcoming.

Bottazzi, L., and G. Peri (2007): "The international dynamics of R&D and innovation in the long run and in the short run," *Economic Journal*, 117, 486–511.

Gengenbach, C., F. Palm, and J. Urbain (2006): "Cointegration Testing in Panels with Common Factors," *Oxford Bulletin of Economics and Statistics*, 68, 683–719.

Jones, C. (1995): "Time series tests of endogenous growth models," *Quarterly Journal of Economics*, 110, 495–525.

Larsson, R., and J. Lyhagen (2000): "Testing for Common Cointegrating Rank in Dynamic Panels," Mimeo.

Larsson, R., J. Lyhagen, and M. Lothgren (2001): "Likelihood-based Cointegration Tests in Heterogenous Panels," *Econometrics Journal*, 4, 109–142.

Mark, M., and D. Sul (2001): "Nominal exchange rates and monetary fundamentals: evidence from a small post-Bretton Woods panel," *Journal of International Economics*, 53, 3354–3370.

Pedroni, P. (1999): "Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors," *Oxford Bulletin of Economics and Statistics*, 61, 653–670.

Pesaran, M. H. (2007): "A simple panel unit root test in the presence of cross-section dependence," *Journal of Applied Econometrics*, 22, 265–312.

Romer, P. M. (1990): "Endogenous technological change," *Journal of Political Economy*, 98, 543–59.

Segerstrom, P. (1998): "Endogenous growth without scale effects," *American Economic Review*, 80, 1077–91.