# Product variety, product quality, and evidence of Schumpeterian endogenous growth

# Data Appendix

Francesco Venturini

## 1 Data sources and methods

The analysis is performed on a panel sample of twenty US manufacturing industries for the period 1975-2003. Table A.1 lists industries according the ISIC Rev. 3 Classification, and groups them according to the OECD technology-based categories (OECD, 2005, Annex 1).

### Patent data

Patent data come NBER USPTO patent data files Edition 2006, available at the URL: https://sites.google.com/site/patentdataproject/Home (Release April, 9 2009). Attention is paid to patents granted between 1976 and 2006, issued to US assignees, for which application year, IPC technological class and claimed priorities were available (1,386,267 applications). For each application, we computed the amount of both forward citations received and backward citations made existing patents. Patents are classified according to their application year and ISIC industry categories. For the latter we employed the ISIC/IPC concordance table developed by Schmock *et al.* (2003).

A threefold adjustment was implemented on patent data. First, the number of counts is corrected for the truncation associated with the time lag existing between the application date and the grant date. This lag leads the number of observed applications for the period after 2001 to be underestimated with respect to the true distribution, Following Hall *et al.* (2005), industry applications for the period 2002-2003, denoted by $\widetilde{I}_{it}$, are corrected with a factor defined by the inverse of cumulative probabilities of the application-grant time lag, $L_s$ of patents applied since the early 1990s on:

$$C_s = (\sum_{s=0}^{1} L_{2003-s})^{-1}, \qquad I_{it} = C_t * \widetilde{I}_{it}.$$

After eliminating the highest value assumed by $C_s$, the employed factor amounts to 1.24 for 2002 and and 3.02 for 2003. Adjusted patent counts are thus defined as shown on the right-hand side of the previous expression. Second, we apply the fixed-effect procedure developed by Hall *et al.*, 2001 to control for the effect of citation truncation. Indeed, the number of forward cites of the most recent applications is downward biased as such patents have a shorter time window to be cited; for this reason, forward citations are scaled on citation mean of overall manufacturing. Third, since cites made to existing patents by earlier applications are underestimated due to the fact that they can refer to a relatively small number of applications, the sectoral amount of backward cites is re-estimated by multiplying the average backward cites by the corrected amount of applications. Note that estimates do not remarkably change when the original number of backward cites is used in place of the adjusted one.

$\dot{A}_{it}$ is defined as corrected patent applications, $A_{it}$ as their cumulative value obtained through the perpetual inventory method and geometrical depreciation:

$$K_{it} = I_{it} + (1-\delta)K_{it-1}, \quad K_{i0} = \frac{I_{i0}}{\delta + g_i}.$$

$\delta$ is assumed to be constant among sectors and over time, set to 15%. The initial value $K_{i0}$ is computed by means of Hall and Mairesse (1995)'s formula, where $I_{i0}$ is the amount of patent counts at 1975, $g_i$ the average annual rate of change of $I_{it}$ over the period 1975-1985.

### R&D data

R&D expenses come from National Science Foundation (NSF) for the period 1975-1986, and OECD ANBERD for years between 1987 and 2003. The latter source was integrated with NSF (based on SIC/NAICS categories) in presence of missing values. When both sources were uncompleted, missing values were linearly interpolated between available years. R&D expenditure is defined as the sum of federally- and privately-funded research expenses. Whenever they are not disclosed by NSF, publicly-funded R&D expenses were estimated by first interpolating the ratio between total and privately-funded R&D expenses for missing years and, then, applying the resulting coefficient to the private research expenditure (available on a regular base). In order to compute R&D capital, R&D expenses expressed at current prices were first converted into a constant price base using industry deflators for gross output (1995 dollars); then, the same procedure described above for patent counts, based on the perpetual inventory method and geometric depreciation, was applied to volume series for R&D expenditure.

### Industry accounts data

Gross output at 1995 constant prices is taken from EUKLEMS Industry Accounts (release March 2008).

## Panel unit roots and cointegration tests

The cointegration relationship described by eq. (2) is estimated through the Panel Ordinary Least Squares developed by Mark and Sul (2003), where the first-differences of two-period lags and leads of the explanatory variables are included as additional regressors to allow for the dynamic path around the long-run equilibrium and to account for endogeneity. Table A.1 reports the p-value associated with the tests of panel non-stationarity and panel cointegration for the data employed in the regressions reported in Table 2 of the main text. The former test has been proposed by Pesaran (2007) and checks the null hypothesis that all panel individuals contain unit roots against the alternative of heterogeneity. The long-run stationarity (cointegration) relationship between dependent variable and regressors is evaluated by means of the group mean variance ratio test, $VR_G$, of Westerlund (2005); the null hypothesis of this statistics is that there is no cointagration in the whole sample against an alternative of a non-vanishing fraction of panel individuals for which a cointegrated relationship exists.

Table A.1 Industry classification

| | *ISIC Rev. 3* | *Industry* | *OECD Technological classification* | *Label* | *Patent classes* (*Schmock* et al., *2003*) |
|---|---|---|---|---|---|
| 1 | 15-16 | Food , beverage and tobacco | Low-Tech | LT | 1-2 |
| 2 | 17-19 | Textile, leather and footwear | Low-Tech | LT | 3-5 |
| 3 | 21-22 | Pulp, paper, printing and publishing | Low-Tech | LT | 7-8 |
| 4 | 23 | Coke, refined petroleum and nuclear fuel | Medium- and Low-Tech | MLT | 9 |
| 5 | 24 | Chemicals and chemical products | Medium- and High-Tech | MHT | 10-12, 14-16 |
| 6 | 244 | Pharmaceuticals | High-Tech | HT | 13 |
| 7 | 25 | Rubber and plastics | Medium- and Low-Tech | MLT | 17 |
| 8 | 26 | Other non-metallic mineral | Medium- and Low-Tech | MLT | 18 |
| 9 | 27 | Basic metals | Medium- and Low-Tech | MLT | 19 |
| 10 | 28 | Fabricated metal products | Medium- and Low-Tech | MLT | 20 |
| 11 | 29 | Machinery, nec | Medium- and High-Tech | MHT | 21-27 |
| 12 | 30 | Office, accounting and computing machinery | High-Tech | HT | 28 |
| 13 | 31 | Electrical machinery and apparatus, nec | Medium- and High-Tech | MHT | 29-33 |
| 14 | 321 | Electronic valves and tubes | High-Tech | HT | 34 |
| 15 | 322-323 | Communication equipment | High-Tech | HT | 35-36 |
| 16 | 331-3 | Scientific instruments | High-Tech | HT | 37-39 |
| 17 | 334-5 | Other instruments | High-Tech | HT | 40-41 |
| 18 | 34 | Motor vehicles, trailers and semi-trailers | Medium- and High-Tech | MHT | 42 |
| 19 | 35 | Other transport equipment (incl. aerospace, etc.) | High-Tech | HT | 43 |
| 20 | 20, 36-37 | Wood, Furniture and Manufacturing nec | Low-Tech | LT | 6, 44 |

Table A.2 Panel unit roots and cointegration tests (p-values)

| | Row counts | | | | Forward citations $\times$ Counts | | | | Backward citations $\times$ Counts | | | | Claims $\times$ Counts | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
| Pesaran (2007) unit roots test (H0=unit roots) | | | | | | | | | | | | | | | | |
| $\ln \dot{A}$ | 0.99 | 0.95 | 0.99 | 0.86 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.85 | 0.82 | 0.68 | 0.53 | 0.36 | 0.85 |
| $\ln X$ | 1.00 | 1.00 | 1.00 | 0.96 | 1.00 | 1.00 | 1.00 | 0.96 | 1.00 | 1.00 | 1.00 | 0.96 | 1.00 | 1.00 | 1.00 | 0.96 |
| $\ln L$ | 0.99 | 1.00 | 0.49 | 0.41 | 0.99 | 1.00 | 0.49 | 0.41 | 0.99 | 1.00 | 0.49 | 0.41 | 0.99 | 1.00 | 0.49 | 0.41 |
| $\ln A$ | 1.00 | 0.96 | 1.00 | 1.00 | -0.55 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.96 | 1.00 | 0.06 | 0.16 | 0.40 | 0.46 |
| Westerlund (2005) cointegration test (H0=no cointegration) | | | | | | | | | | | | | | | | |
| $VR_G$ | 0.02 | 0.03 | 0.03 | 0.06 | 0.01 | 0.01 | 0.08 | 0.10 | 0.00 | 0.04 | 0.03 | 0.08 | 0.10 | 0.26 | 0.04 | 0.08 |

## References


Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER citations data file: Lessons, insights and methodological tools. *NBER Working papers No. 8498*.

— (2005). Market value and patent citations. *Rand Journal of Economics*, 36(1): 16–38.

Hall, B. H., & Mairesse, J. (1995). Exploring the relationship between R&D and productivity in French manufacturing firms. *Journal of Econometrics*, 65: 263–93.

Mark, N., Sul, D., 2003. Cointegration vector estimation by Panel DOLS and long-run money demand. Oxf Bull Econ Stat 65, 655–680.

OECD (2005). Science, Technology and Industry Scoreboard. Paris.

Pesaran, M. H., (2007). A simple panel unit test in the presence of cross section dependence. *J App Econom*, 22: 65–312.

Schmock, U., Laville, F., Patel, P., Frietsch, R. (2003). Linking Technology Areas to Industrial Sectors. Final Rapport for the European Commission, DG Research.

Westerlund, J. (2005). New simple tests for panel cointegration. *Econom Rev*, 24: 297–316.