# What do firms know? What do they produce? A new look at the relationship between patenting profiles and patterns of product diversification

Giovanni Dosi [1] Marco Grazzi [2] Daniele Moschella [1]

[2]DSE, University of Bologna

[1]Institute of Economics, Scuola Superiore Sant'Anna

Scuola Sup. Sant'Anna - Pisa
December 2nd, 2015

# What do firms know? What do they produce? Some established evidence

- Importance of product and technological diversification for:
  - firm performance through growth opportunities, economies of scope, risk diversification (Hirsch and Lev, 1971; Montgomery, 1994; Bottazzi et al., 2001; Garcia-Vega, 2006)
  - accumulation of capabilities (Dosi, 1988; Pavitt, 1998)
- Relationship between technological knowledge and product portfolios:
  - Evidence on large firms: technological scope greater than product scope (Patel and Pavitt, 1997; Gambardella and Torrisi, 1998; Brusoni et al., 2001)
- Characteristics of diversification process:
  - Evidence of path-dependent and coherent processes (Teece et al., 1994; Bottazzi et al., 2001; Breschi et al., 2003; Bottazzi and Pirino, 2010)

# What do firms know? What do they produce? Our contribution

- Bringing together information on patents and products we analyze the following characteristics of technological knowledge and product portfolios:
  - Scope of both technological knowledge and product portfolios
  - Relevance of specific technological knowledge in backing products
  - Scaling relation between the size of the firm and diversification
  - Coherence and diversification
- Results are well in tune with a capability-based theory of the firm
- We resort to Lybbert and Zolas (2014) for the matching of IPC to industrial sectors and products
  - Improvement over Schmoch et al. (2003)

# Outline of the talk

1. Introduction
2. Dataset description
3. Characteristics of patenting firms
4. Technological and product diversification: some stylized facts
5. Scaling and diversification
6. Diversification and firm coherence
7. Conclusions

# Data sources

1. Amadeus
   - Provides information on patent applications for more than 20,000 Italian firms (including the IPC classification code, the application date, and whether the patent has been granted or not. Bureau Van Dijk patent-firm match)
2. Archivio Statistico delle Imprese Attive (ASIA)
   - Census of all operating businesses: age, employment, total turnover, geographical location and main activity of the firm, 1998-2006
3. Statistiche del Commercio Estero (COE) Custom data
   - Transactions level data: export values and quantity of the firm for HS6 product-country destination pairs
   - All cross-border transactions at the firm-product-country level, 2000-2007

Resorting to Lybbert and Zolas (2014), we link IPC codes to 125 4-digits ISIC codes (Rev. 3), and SH6 codes to 145 4-digits ISIC codes. Firms in our sample patent in 118 different technological fields and produce 138 different products.

# Constructing the dataset

- We focus on granted patents that have been applied to USPTO or EPO offices (in Amadeus: 49,803 patents owned by 7,311 firms)
- Step 1: AMADEUS $\rightarrow$ ASIA. We match around 85% of firms in AMADEUS and 90% of their patent applications to 2006 ASIA archive
- Step 2: AMADEUS-ASIA $\rightarrow$ COE. We match around 90% of patent applications and 70% of firms to 2006 COE
  - High export propensity of patenting firms (see Dosi et al., 2015)
  - Around 70% of firms that patent and do not export are active in non-manufacturing sectors

# Patents and firms, by period of application and patent office

Table 1:

| Period | Total | | USPTO | | EPO | |
|---|---|---|---|---|---|---|
| | PATENTS | FIRMS | PATENTS | FIRMS | PATENTS | FIRMS |
| 1949-1978 | 1,086 | 187 | 1,086 | 187 | | |
| 1979-1995 | 8,055 | 1,426 | 3,929 | 863 | 4,126 | 1,168 |
| **1996-2006** | **21,305** | **2,946** | **9,817** | **1,647** | **11,488** | **2,499** |
| 2007-2014 | 9,340 | 1,948 | 4,871 | 1,006 | 4,469 | 1,550 |
| 1949-2014 | 39,786 | 4,411 | 19,703 | 2,586 | 20,083 | 3,709 |

*Note.* Number of USPTO and EPO granted patents owned by Italian firms. The period refers to the application date. Data from AMADEUS, ASIA, and COE.

⇒ In **boldface**, the period considered in this work

# Distribution of firms and patents by number of patents

Table 2:

| | No. of patents | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10+ | Obs. |
| % OF FIRMS | 41.68 | 19.14 | 9.37 | 6.45 | 4.41 | 3.29 | 1.66 | 1.60 | 1.15 | 11.24 | 2,946 |
| % OF PATENTS | 5.73 | 5.26 | 3.86 | 3.54 | 3.03 | 2.71 | 1.60 | 1.75 | 1.43 | 71.08 | 21,441 (*) |

*Source.* Amadeus, ASIA, and COE, 1996-2006.
(*) There a few co-patentees.

⇒ The typical firm owns just one patent (41.68%)
⇒ Large patentees (10+) account disproportionately for the whole stock of patents (71.08%)

# Size, age, and product scope: patenting *vs.* non-patenting firms

$$X_f = \alpha + \beta D_{PAT_f} + \epsilon_f$$

Table 3:

| | (1) $X_f =$ $log(exports_f)$ | (2) $X_f =$ $log(age_f)$ | (3) $X_f =$ $log(age_f)$ | (4) $X_f =$ $log(\#products_f)$ | (5) $X_f =$ $log(\#products_f)$ |
|---|---|---|---|---|---|
| $D_{PAT_f}$ | 3.092*** | 0.320*** | 0.189*** | 1.003*** | 0.371*** |
| | (0.047) | (0.016) | (0.017) | (0.015) | (0.012) |
| $log(exports_f)$ | | | 0.042*** | | 0.205*** |
| | | | (0.001) | | (0.001) |
| $N$ | 139,360 | 139,360 | 139,360 | 139,360 | 139,360 |
| adj. $R^2$ | 0.195 | 0.060 | 0.074 | 0.122 | 0.484 |
| Sector dummies | Yes | Yes | Yes | Yes | Yes |

*Note.* Standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$\Rightarrow$ Patenting firms are older and bigger than non patenting firms, both in terms of total exports and in terms of product scope

# Price and quantity: patenting *vs.* non-patenting firms

$$\log(exports_{fpc}) = \log(quantity_{fpc}) + \log(unitvalue_{fpc})$$
$$X_{fpc} = \alpha + \beta D_{PAT_i} + d_{pc} + \varepsilon_{fpc}$$

Table 4:

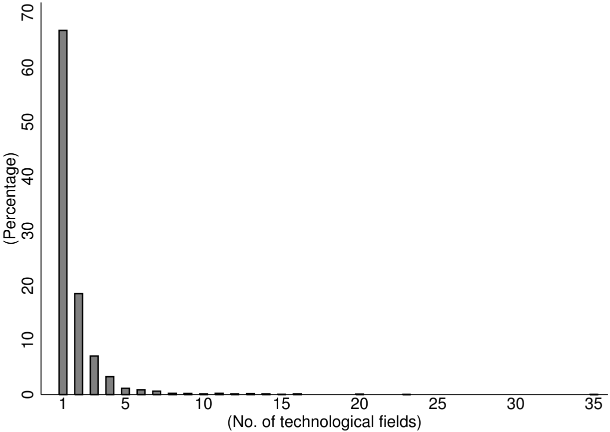
| | (1) $X_{fpc} =$ $log(exports_{fpc})$ | (2) $X_{fpc} =$ $log(quantity_{fpc})$ | (3) $X_{fpc} =$ $log(unitvalue_{fpc})$ |
|---|---|---|---|
| $D_{PAT_f}$ | 0.287*** | 0.172*** | 0.115*** |
| | (0.021) | (0.031) | (0.023) |
| $N$ | 1,286,689 | 1,286,689 | 1,284,150 |
| adj. $R^2$ | 0.233 | 0.251 | 0.430 |
| Country-Product FE | Yes | Yes | Yes |

*Note.* Standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$\Rightarrow$ Within product-country pair, patenting firms have both higher export quantities and prices than non-patenting firms

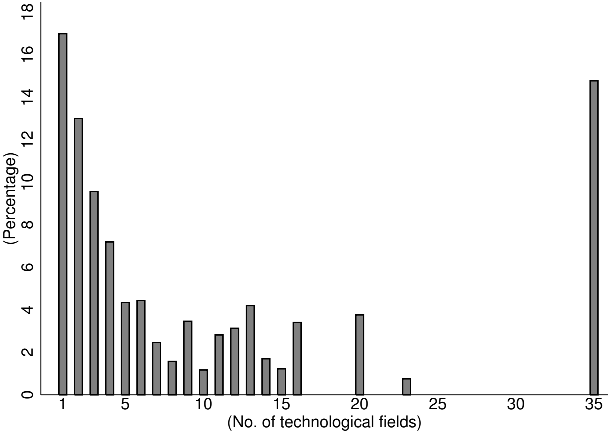
# Distribution of firms by number of technological fields

Figure 1:



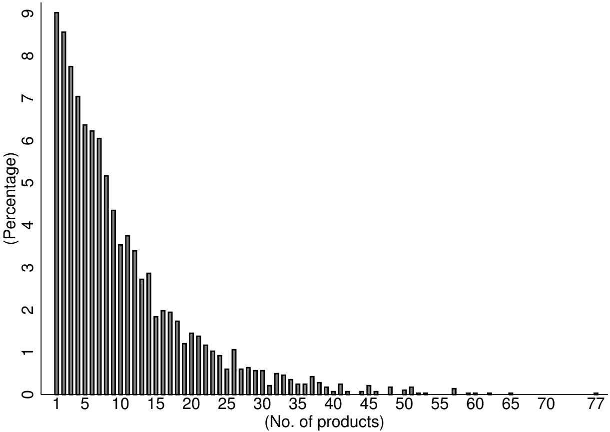
# Distribution of firms' patents by number of tech. fields

Figure 2:



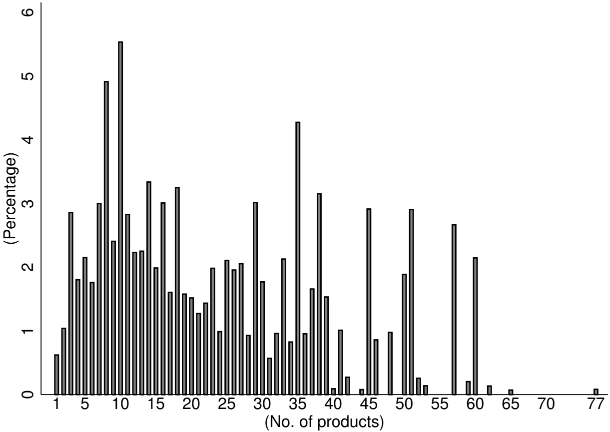
# Distribution of firms by number of products

Figure 3:



# Distribution of firms' export value by number of products

Figure 4:



# Joint distribution of firms by # of tech. fields and products

Table 5:

| #Technological fields | # Products 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10+ | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 208 | 194 | 168 | 156 | 136 | 130 | 115 | 95 | 81 | 603 | 1,886 |
| | (11.03) | (10.29) | (8.91) | (8.27) | (7.21) | (6.89) | (6.10) | (5.04) | (4.29) | (31.97) | (100.00) |
| 2 | 35 | 38 | 38 | 31 | 32 | 27 | 40 | 31 | 23 | 228 | 523 |
| | (6.69) | (7.27) | (7.27) | (5.93) | (6.12) | (5.16) | (7.65) | (5.93) | (4.40) | (43.59) | (100.00) |
| 3 | 5 | 6 | 7 | 9 | 6 | 15 | 10 | 13 | 10 | 119 | 200 |
| | (2.50) | (3.00) | (3.50) | (4.50) | (3.00) | (7.50) | (5.00) | (6.50) | (5.00) | (59.50) | (100.00) |
| 4 | 4 | 4 | 3 | 1 | 3 | 1 | 4 | 3 | 6 | 64 | 93 |
| | (4.30) | (4.30) | (3.23) | (1.08) | (3.23) | (1.08) | (4.30) | (3.23) | (6.45) | (68.82) | (100.00) |
| 5 | 2 | 0 | 0 | 1 | 0 | 2 | 0 | 1 | 3 | 24 | 33 |
| | (6.06) | (0.00) | (0.00) | (3.03) | (0.00) | (6.06) | (0.00) | (3.03) | (9.09) | (72.73) | (100.00) |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 23 | 25 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (4.00) | (4.00) | (0.00) | (92.00) | (100.00) |
| 7 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 16 | 18 |
| | (0.00) | (0.00) | (0.00) | (5.56) | (0.00) | (0.00) | (0.00) | (5.56) | (0.00) | (88.89) | (100.00) |
| 8 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 4 | 7 |
| | (0.00) | (0.00) | (28.57) | (0.00) | (14.29) | (0.00) | (0.00) | (0.00) | (0.00) | (57.14) | (100.00) |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 5 | 6 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (16.67) | (0.00) | (0.00) | (83.33) | (100.00) |
| 10+ | 1 | 0 | 1 | 0 | 2 | 1 | 0 | 1 | 0 | 29 | 35 |
| | (2.86) | (0.00) | (2.86) | (0.00) | (5.71) | (2.86) | (0.00) | (2.86) | (0.00) | (82.86) | (100.00) |
| Total | 255 | 242 | 219 | 199 | 180 | 176 | 171 | 146 | 123 | 1,115 | 2,826 |
| | (9.02) | (8.56) | (7.75) | (7.04) | (6.37) | (6.23) | (6.05) | (5.17) | (4.35) | (39.46) | (100.00) |

*Note.* Absolute and percentage (in brackets) frequencies. Total number of firms (2826) is different from Table 1 and 2 (2946) because for some patents IPC is not available.

# Product rank and firm knowledge

Table 6: Matching between technological fields and products

| Product rank | # Products | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 (745) | 2 (566) | 3 (880) | 4 (564) | 5 (816) | 6 (501) | 7 (652) | 8 (1,107) | 9 (561) | 10+ (13,941) |
| 1 | 17.32 | 21.38 | 17.27 | 24.82 | 19.98 | 19.56 | 13.50 | 36.04 | 29.23 | 23.61 |
| 2 | | 4.95 | 4.89 | 10.99 | 10.17 | 6.79 | 4.60 | 2.17 | 3.74 | 6.17 |
| 3 | | | 6.14 | 1.06 | 1.96 | 2.79 | 6.60 | 2.44 | 1.25 | 3.92 |
| 4 | | | | 2.49 | 2.21 | 4.59 | 1.23 | 3.52 | 2.14 | 1.74 |
| 5 | | | | | 1.03 | 2.40 | 2.45 | 3.99 | 0.18 | 1.46 |
| 6 | | | | | | 0.60 | 0.31 | 0.45 | 0.71 | 1.75 |
| 7 | | | | | | | 2.15 | 0.90 | 3.92 | 0.91 |
| 8 | | | | | | | | 0.45 | 1.25 | 3.72 |
| 9 | | | | | | | | | 1.07 | 1.66 |
| 10+ | | | | | | | | | | 0.81 |

*Note.* Each cell reports the percentage matching between technological fields and product categories across the relevant set of firm-products. In parentheses, the number of patents for the relevant set of firm-products.

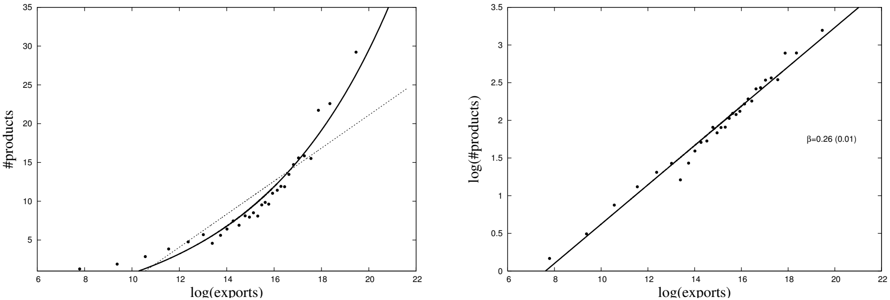
⇒ The matching is much higher for the products that, within each firm, account for most of its export

# Summary of results

- Firms active in more than one IPC are quite rare, rarer than multiproduct firms
- Firms tend to exploit innovative knowledge more related to their main products
- Firms that diversify their activities across different products or different technologies account disproportionately for the total exports and the total patents
    - This evidence hints at some underlying relation between size and firm diversification: bigger firms in terms of patents or exports are also those which explore more opportunities in the product and technological space

# Exports (size) and number of products

Figure 5:



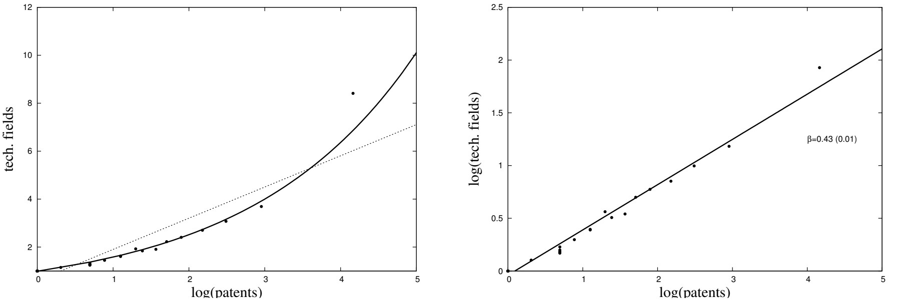

(a) #products against log(exports). Linear (dotted line) and exponential (solid line) fit

(b) log(#products) against log(exports). Linear fit

$\Rightarrow$ The relationship is well captured by a log-linear function, with a strongly signficant slope $\beta = 0.26$

# Exports (size) and IPC

Figure 6:




(a) #tech. fields against log(patents). Linear (dotted line) and exponential (solid line) fit

(b) log(#tech. fields) against log(patents). Linear fit

$\Rightarrow$ Also this relationship is well captured by a log-linear function, with a strongly signficant slope $\beta = 0.43$

# Scaling properties: implications

- Log-linear relationships: diversification patterns can be described by a stochastic branching process (Bottazzi and Secchi, 2006)
- $\beta < 1$: a large firm is less diversified than a collection of small (single product) firms which add up to the same size of the large one. A large firm is more risky than a collection of smaller firms
- If diversification is a competence-driven process, one should expect it to be associated with non-decreasing levels of firm coherence

# Measuring firm coherence: relatedness

We adopt a "survivor" measure of relatedness (Teece et al., 1994): if a selection mechanism is at work and if relatedness confers some advantage, then related activities will appear with higher frequency within the same (surviving) firm.

- A co-occurrences matrix $C$, whose generic cell $J_{ij}$ is equal to the total number of firms active in both activities:

$$J_{ij} = \sum_k C_{ik} C_{jk} \tag{1}$$

  with $C_{ik} = 1$ if firm $k$ is active in product (or IPC) $i$ and 0 otherwise, so that $J_{ij}$ is the number of firms active both in $i$ and $j$

- The p-value of the generic cell of the observed $J_{ij}$:

$$p_{ij}(J, H) = Prob[\tilde{J}_{ij} \leq J_{ij} | H] \tag{2}$$

  where $\tilde{J}_{ij}$ is the value of the relative cell under the null hypothesis $H$; $p_{ij}$ can be greater than, equal, or less than 0.5, implying that activities $i$ and $j$ are positively, not related, or negatively related.

# Measuring firm coherence: null hypothesis

- Standard null hypothesis: the total number of firms active in a given sector (or product or patent class) is fixed and equal to the one observed in the actual data
  - the probability to obtain a given value of $\tilde{J}_{ij}$ is distributed according to a hypergeometric random variable
  - the implied distribution of firm scope converges to a binomial
- Alternative null hypothesis (Bottazzi and Pirino, 2010): both firms scope and the number of firms per activities are fixed and correspond to the observed ones
  - Deriving the implied distribution using Monte Carlo tecniques

# Measuring firm coherence: WAR

- Weighted average relatedness measures the inverse of the average distance from a firm activity to *all* other activities:

$$WAR_k(H) = \frac{1}{n} \sum_i C_{ik} \left( \frac{\sum_{j \neq i} p_{ij}(H) w_{jk}}{\sum_{j \neq i} w_{jk}} \right) \tag{3}$$

  where $n$ is total number of products (technological fields) in which a firms is active and $w_{jk}$ the weight of product (technological field) $j$ with respect to firm $k$. We weight products with export share and technological fields with patent count

- As firms diversify in new products and technological fields, $WAR$ is expected, on average, to increase.

# Measuring firm coherence: WARN

- Weighted average relatedness of neighbors measures the inverse of the average distance from a firm activity *to its neighbor activity* :

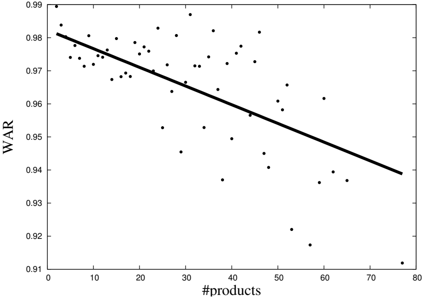
$$WARN_k(H) = \frac{1}{n} \sum_i C_{ik} \left( \frac{\sum_{j \neq i} p_{ij}(H) m_{ij}^k w_{jk}}{\sum_{j \neq i} m_{ij}^k w_{jk}} \right) \tag{4}$$

where $m_{ij}^k = 1$ if the pair $ij$ is in the maximum spanning tree of firm $k$, defined as the graph with $n-1$ links such that the sum of the relatedness measures on each link is largest

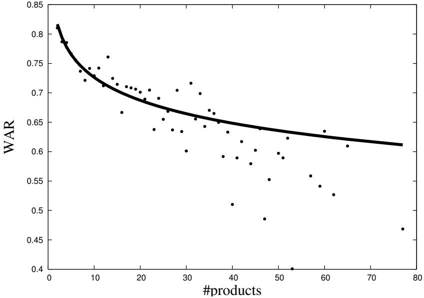
- If firms add products (technologies) that are related to some portion of existing products (technologies), the *WARN* measure should be roughly constant across levels of diversification

# Product WAR as function of #products

Figure 7:




(a) Standard null hypothesis. Linear fit

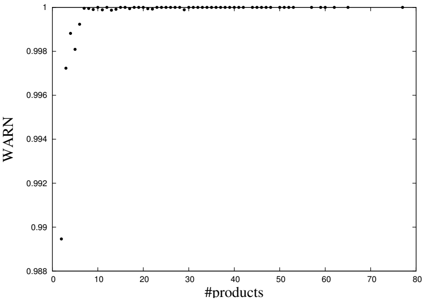



(b) Alternative null hypothesis. Log-linear fit

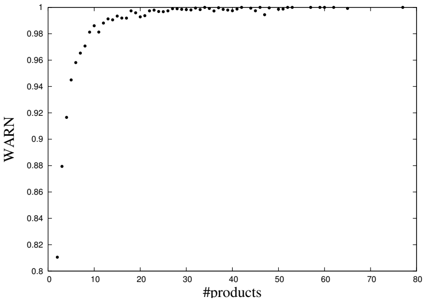
$\Rightarrow$ As expected, as firms increase their product scope, the coherence across all its activities decrease. Non-linearly under the alternative null hypothesis

# Product WARN as funtion of #products

Figure 8:




(a) Standard null hypothesis

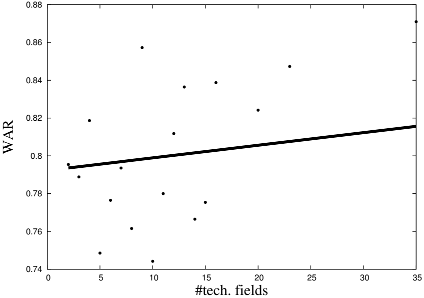



(b) Alternative null hypothesis

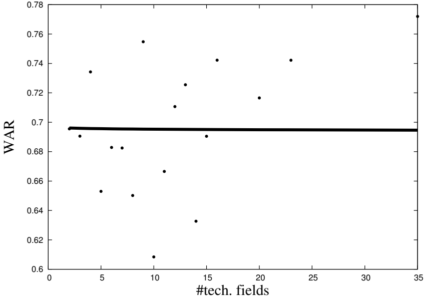
$\Rightarrow$ As firms introduce new products the coherence between neighboring activities slightly increase for relatively low levels of diversification, and stay constant for sufficiently diversified firms

# Technological WAR as funtion of #tech. fields

Figure 9:




(a) Standard null hypothesis. Linear fit

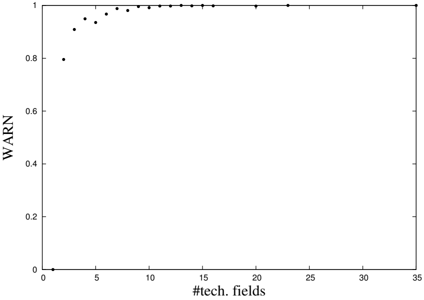



(b) Alternative null hypothesis. Log-linear fit

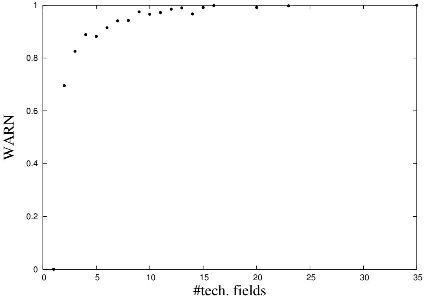
⇒ The analysis of relatedness measure on patents suffer for lower degree of technological diversification as opposed to product

# Technological WARN as funtion of #tech. fields

Figure 10:




(a) Standard null hypothesis




(b) Alternative null hypothesis

# Conclusions

- Firms are more diversified in terms of products than in terms of technologies. "Pavitt" firms, who know more than they make, are quite rare, mostly limited to the few large patentees
- The patterns of diversification are consistent with a branching process whereby knowledge on production and innovation, so to speak, "spurs out" from what the firm already does and knows
- The coherence in the directions of diversification reinforced the point: where a firm stands in term of pre-existing capabilities shapes to a good extent where it will go
- Results are consistent with capabilities-based and evolutionary views of the firm: these theories show how opportunities of diversification are shaped by technological imperatives and path-dependent learning dynamics within the firm (Teece et al., 1994; Dosi et al., 2000; Winter, 2003)

Thank you!