



Universidad
Carlos III de Madrid



This is a postprint version of the following published document:

Blazsek, S., Escribano, A. (2016). Patent propensity, R&D and market competition: Dynamic spillovers of innovation leaders and followers. *Journal of Econometrics*, v. 191, n. 1, pp. 145-163.
Available in: <https://doi.org/10.1016/j.jeconom.2015.10.005>

© Elsevier



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Patent propensity, R&D and market competition: Dynamic spillovers of innovation leaders and followers

Szabolcs Blazsek^a, Alvaro Escribano^{b,*}

^a School of Business, Universidad Francisco Marroquín, 6 calle final, zona 10, 01010, Ciudad de Guatemala, Guatemala

^b Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903, Getafe (Madrid), Spain

A B S T R A C T

In this article, dynamic interactions among stock return, Research and Development (R&D) investment, patent applications and patent propensity of firms are studied. Patent innovation leader and follower firms are identified with respect to their quality-adjusted knowledge stock. Significant and positive dynamic spillover effects are obtained in a panel vector autoregressive model. We find positive dynamic spillover effects from patent innovation leader to followers. We show that an increasing degree of competition enhances innovation and patent applications, which helps firms appropriating part of the benefits of their R&D investments.

JEL classification:

C15
C31
C32
C33
C41

Keywords:

Patent propensity
Competition
Innovation leaders
Panel vector autoregression
Simulated maximum likelihood

1. Introduction

The objective of this work is to learn about dynamic interactions (spillovers) between patent innovation leaders and followers, allowing for different degrees of propensity to patent for different firms. We use patent and firm-specific panel data of 4476 firms for all manufacturing and service industries of the United States (US) for period 1979–2000. Firms are classified into technological groups with respect to technological proximity. We focus on a specific technological cluster of 111 firms, which are mainly from the drugs market sector. We identify patent innovation leaders and followers of the cluster with respect to their quality-adjusted knowledge stock.

We consider different dynamic measures of innovation activity that may capture patented Research and Development (R&D) activity (i.e., publicly disclosed innovations) and non-patented

R&D activity (i.e., non-appropriated R&D or trade secrets). Patented and non-patented R&D activities are estimated by using a latent variable patent count panel data model. In this model, patent propensity is driven by a common latent factor, which represents unobservable common effects of the technological cluster. These common effects are, for example, degree of market competition, degree of deregulation, degree of enforcement of intellectual property rights and degree of information that is common knowledge. We study dynamic R&D spillovers among patent innovation leaders and followers by Panel Vector Autoregressive (PVAR) models. The econometric models applied involve variables that are observed by those firm managers who choose what proportion of firm R&D output to patent or keep secret, but the same variables are not included in the dataset available to the econometrician.

As noted by Boldrin and Levine (2008, p. 8): “we should protect not only the property rights of innovators but also the rights of those who have legitimately obtained a copy of the idea, directly or indirectly, from the original innovator. The former encourages innovation; the latter encourages the diffusion, adoption and improvement of innovations”. In this work, we show that more competition increases patent propensity and R&D investment in

* Corresponding author. Tel.: +34 916249854.

E-mail addresses: sblazsek@ufm.edu (S. Blazsek), alvaroe@eco.uc3m.es (A. Escribano).

order to enhance the absorptive capacity of firms (Escribano et al., 2009) thus more innovation and more patents are produced. This transmission mechanism has an extra dynamic multiplier effect through the spillovers between innovation leaders and followers. We find positive dynamic spillover effects from patent innovation leader to followers, but not the other way around. We also find support for the inverted-U relationship between market competition and innovation (Aghion et al., 2005). Increase in market competition conditions within the technological cluster are related to rise in patent propensity, and vice versa (feedback). This suggests that pharmaceutical firms reacted to the increasing level of market competition by patenting a significantly higher proportion of their innovation output after 1990, which at the same time expanded the diffusion of knowledge among competitors, enhancing innovation.

The remainder of this work is organized as follows. Section 2 reviews the literature. Section 3 presents the dataset, technological clustering of firms, and definitions of patent innovation leaders and followers. Section 4 describes the econometric models and presents empirical findings. Section 5 discusses our results about competition and patent propensity. Section 6 summarizes and concludes.

2. Literature review

2.1. Firm value and innovation activity

During recent decades, innovations protected by patents have played a key role in business strategies. This fact motivated a number of studies about the determinants of patents and impact of patents on innovation, firm value and competitive advantage.

Griliches (1981) constructs knowledge-based variables from lagged R&D expenses and number of patents. He finds significant positive association among market value, R&D expenditure and number of patents for a panel of large US firms. Lev and Sougiannis (1996) estimate the inter-temporal relation between R&D capital and stock return of US firms, and they show that R&D capital is associated with subsequent stock returns (see also Lev et al., 2005). Blundell et al. (1999) examine the relationship between surprise innovations and firm performance by using a dynamic count panel data model, and find a positive impact of innovation on market value of US firms. Chan et al. (2001) show a positive relationship between R&D capital to market value and abnormal future stock returns. They also find evidence of a delayed association of R&D activity and future excess stock returns, which could be due to a delayed reaction of the stock market or inadequate adjustment for risk (Chambers et al., 2002). Hall et al. (2005) investigate the association between knowledge stock and market value in the US for period 1963–1995. Their results show that, in addition to patent counts, patent citations also contain important information about stock market value.

2.2. Innovation leaders and followers

Technological improvement gives competitive advantage to innovator companies. Nevertheless, the non-rival nature of knowledge may create a business-stealing effect among competitors, as innovator's efforts may decrease the cost of competitor firms' subsequent innovations. Firms strategically decide to be R&D leaders or followers. Companies that introduce innovative products are R&D leaders, while other firms, who mimic products of innovation leaders, are followers. There is a large body of literature in economics and strategic management, which differentiates among

firms with respect to their R&D and patenting activity, in order to study the implications of the research intensity of firms on competitors' market value and innovations. Results in this body of literature suggest that R&D leaders have sustained future profitability.

Jaffe (1986) finds evidence of knowledge spillovers by using various indicators of R&D activity. He evidences that firms that invest in research in a sector where there is high R&D intensity, obtain more patents per dollar of R&D, higher accounting profit to R&D and higher market value to R&D, than firms in a sector with low R&D intensity. Nabseth and Ray (1974), Mansfield et al. (1981), Rogers (1983) and Pakes and Schankerman (1984) report that knowledge spills over gradually in a dynamic manner to other firms. Aghion et al. (2005) develop a model where competition discourages laggard firms, but encourages neck-and-neck firms to innovate. Due to the effect of competition on equilibrium industry structure, their model generates an inverted U-shaped relationship between innovation and competition. They also show that average technological distance between innovation leaders and followers increases with competition and the inverted-U is steeper when industries are more neck-and-neck. Lev et al. (2006) differentiate between R&D leaders and followers and compare stock market valuation of leaders and followers. They show that R&D leaders earn significant future excess returns, while R&D followers only earn average returns. Ciftci et al. (2011) find that R&D leaders obtain substantial risk-adjusted returns during the first four to five future years, following a new innovation. However, these excess returns converge to those of R&D followers, afterwards.

3. Data

The dataset includes $N = 4476$ firms from all manufacturing and service industries of the US economy for period 1979–2000 ($T = 22$ years). These firms published more than 500,000 patents during the sample period. We created the dataset based on the recommendations of Hall et al. (2001).

Data used in this work are from several sources. Patent data are from two sources: (a) National Bureau of Economic Research (NBER) US Patent Citations Data File and (b) MicroPatent LLC. The patent database includes the US Patent and Trademark Office (USPTO) patent number, application date, publication date, patent number of cited patents, three-digit US technological class and assignee name (company name, if the patent was assigned to a firm). Furthermore, annual stock return data were obtained from the Center for Research on Stock Prices (CRSP). Additional company-specific data have been obtained from Standard & Poor's Compustat data files. The firm-level dataset includes book value of equity, stock market value, Standard Industry Classification (SIC) code and R&D expenditure. Firm-specific accounting data were corrected for inflation by using the US Consumer Price Index (CPI) (source: US Department of Labor, Bureau of Labor Statistics).

In the remaining part of this section, we describe the technological clustering procedure as applied to US firms. Then, we present the definitions of patent innovation leaders and followers.

3.1. Technological proximity

We use a patent-based technological proximity measure in order to classify firms into technologically-similar groups. Technological clusters are formed as follows. For each firm, we assign technological categories based on their patent applications, with respect to the technological classification of Hall et al. (2001). These authors created 36 technological sub-categories

Table 1
Product-market-based industry classification in the technological cluster.

SIC industry name	SIC code	K	HV industry name	K
Pharmaceutical preparations	2834	47	Pharmaceuticals	92
Biological products (no-diagnostic substances)	2836	31	Non-manufacturing	10
In vitro and in vivo diagnostic substances	2835	7	Computers	4
Perfumes, cosmetics and other toiletry preparations	2844	3	Chemical products	2
Surgical and medical instruments, and apparatus	3841	3	Food and tobacco	2
Medicinal chemicals and botanical products	2833	2	Plastics and rubber products	1
Wholesale-drugs, proprietaries and druggists' sundries	5122	2		
Services-medical laboratories	8071	2		
Grain mill products	2040	1		
Beverages	2080	1		
Chemicals and allied products	2800	1		
Soap, detergents, perfumes and cosmetics	2840	1		
Paints, varnishes, lacquers, enamels and allied products	2851	1		
Agricultural chemicals	2870	1		
Plastics products	3089	1		
Electro-medical and electro-therapeutic apparatus	3845	1		
Wholesale-medical, dental and hospital equipment	5047	1		
Fire, marine and casualty insurance	6331	1		
Services-hospitals	8060	1		
Services-engineering, accounting, research and management	8700	1		
Services-commercial physical and biological research	8731	1		
Non-operating establishments	9995	1		
Total number of firms in the technological cluster		111		111

Notes: Standard Industry Classification (SIC). Number of firms (K). Hall and Vopel (HV, 1997) classification.

from the patent technological classification of USPTO that contains about 400 technological classes. We apply Ward's (1963) linkage clustering to perform technological clustering; motivated by Kuiper and Fisher (1975) and Jain et al. (1986). We use the angle distance measure to form technological clusters of firms which is purely directional therefore it is not affected by the degree of concentration of the research interests of firms (Jaffe, 1986). The technological clustering procedure creates a technology-related grouping of 16 clusters for the 4476 US companies.

We focus on a specific cluster of $N = 111$ companies. Table 1 shows the product-market industries of firms in the technological cluster, with respect to two classifications. First, according to SIC, the technological cluster includes 87 firms from the SIC283 drugs sector. Second, with respect to the modified SIC of Hall and Vopel (1997), 92 companies of the technological cluster are in the pharmaceutical sector. Table 1 also presents that the technological cluster includes companies from other product-market sectors. For example, it includes firms from the grain mill products (SIC2040), beverages (SIC2080), paints (SIC2851), plastics products (SIC3089) and electro-medical and electro-therapeutic apparatus (SIC3845) industries. Fig. 1(a) shows the evolution of total patent application count of firms in the technological cluster for period 1979–2000. The figure shows a significant growth of patent application counts for the sample period. The level of patent applications per year was about 600 patents in 1979, which grew to about 1300 patents per year in 2000.

3.2. Patent innovation leaders and followers

We define the permanent Innovation Leader (IL) firm, with respect to the absolute temporal dominance observed in the evolution of quality-adjusted knowledge stock, built up from the citations weighted annual patent counts. The evolution of knowledge stock of firm i is computed by $\sum_{s=1}^t FC_{is}P_{is}(1-\delta)^{t-s}$ for $t = 1, \dots, T$, where P_{is} denotes the number of successful patent applications for firm i and in period s (i.e., patent application count); FC_{is} is the number of citations received from subsequent patents (i.e., forward citations), corrected for sample truncation

bias, for firm i and in period s . We use FC_{is} to weight patent counts as Lanjouw and Schankerman (1999) and Hall et al. (2001) report that the number of forward patent citations is an appropriate measure of patent quality. More recent patents in the end of the sample have less chance to receive citations from later patents than earlier patents. This creates a sample truncation bias for the forward citations count. We correct for this bias with respect to the fixed effects approach suggested by Hall et al. (2001). Motivated by Hall (1993) and Hall et al. (2005), we use the annual depreciation rate $\delta = 15\%$ to account for decreasing value of past knowledge. The firm with the highest quality-adjusted knowledge stock in every year of the observation period is the permanent patent IL of the technological cluster. We define the dummy variable D_{it} ($i = IL$) that takes the value one in period t if firm i is the permanent IL, and zero otherwise. Other firms in the technological cluster are in the permanent patent Innovation Follower (IF) group. We define the dummy variable D_{it} ($i \in IF$) that takes the value one in period t if firm i is in the permanent IF group, and zero otherwise.

Table 2 shows 20 firms out of the 111 of the technological cluster for period 1979–2000. The table shows the average of each of the following variables computed over the sample period: (V1) patent application count P_{it} ; (V2) forward citations received count \tilde{FC}_{it} ; (V3) forward citations received count corrected for sample truncation bias FC_{it} ; (V4) knowledge stock $\sum_{s=1}^t FC_{is}P_{is}(1-\delta)^{t-s}$; (V5) log of R&D expenses r_{it} ; (V6) log of book value of equity BV_{it} ; (V7) log of market value of equity MV_{it} ; (V8) log R&D expenses to log sales r_{it}/s_{it} ; (V9) log R&D expenses to log market value r_{it}/MV_{it} . In the technological cluster, firms are ranked according to mean (V4). Table 2 presents the first 20 firms with respect to this ranking. The table shows that for seven out of the nine variables, Merck & Co., Inc. (Merck) is the leader.

Table 3 presents the evolution of knowledge stock for eight firms with the highest mean (V4) for period 1979–2000. The table shows that the quality-adjusted knowledge stock of Merck was permanently higher than that of other firms in every year. This result also supports our conclusion that Merck is the permanent patent IL of the technological cluster. In addition, Fig. 1(e) shows the evolution of log of number of patent applications and log of knowledge stock of IL, and cross-sectional mean of log of

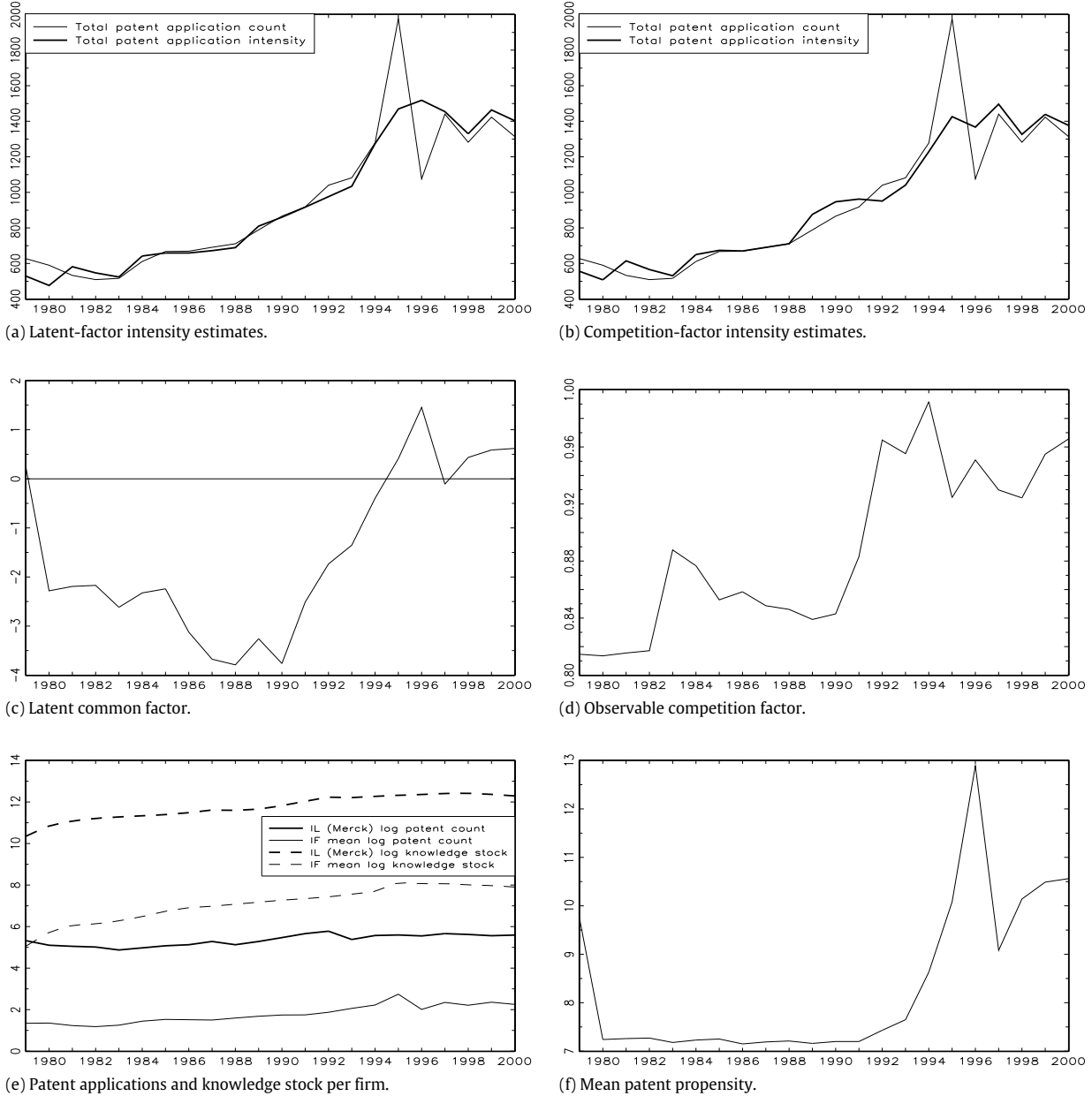


Fig. 1. Patent applications, patent intensity and patent propensity. *Notes:* (a) Evolution of total patent count $\sum_{i=1}^N P_{it}$ and total patent intensity $\sum_{i=1}^N \lambda_{it}$ for the patent count data model with latent common factor for $t = 1, \dots, T$; (b) evolution of total patent count $\sum_{i=1}^N P_{it}$ and total patent intensity $\sum_{i=1}^N \lambda_{it}$ for the patent count data model with observable competition factor for $t = 1, \dots, T$; (c) estimates of $E[\lambda_{it}^* | \mathcal{F}_t^0]$ for $t = 1, \dots, T$; (d) evolution of market competition CO_t for $t = 1, \dots, T$; (e) patent application count and knowledge stock per firm for IL and IF for $t = 1, \dots, T$; (f) evolution of mean patent propensity $(1/N) \sum_{i=1}^N P_{it}^*$ in percentage, for $t = 1, \dots, T$.

patent applications and log of knowledge stock of IFs for period 1979–2000. This figure also suggests the selection of Merck for permanent IL, since both variables of Merck are higher than the corresponding variables of IFs for every year for period 1979–2000.¹

¹ We also classify firms with respect to their knowledge stock to Group of Leaders (GL) and Group of Followers (GF). We form these groups based on the mean (V4) knowledge stock variable by using Ward's (1963) clustering method. GL includes Merck; Eli Lilly; Abbott Laboratories; Warner–Lambert; Pfizer; Bristol–Myers Squibb; American Home Products; Alza. Tables 2 and 3 present mean (V1) to (V9) and evolution of knowledge stock, respectively, for these firms. We estimate the model presented in Section 4.4 for both IL–IF and GL–GF clusters. We find similar results for both clustering thus in this work we only report the IL–IF results. Blazsek and Escribano (2014) present the GL–GF results.

4. Econometric models and empirical results

4.1. Benchmark innovation and market value model

Innovation activity has a positive impact on future cash flow and current value of firms, which motivates owners to promote R&D. As profits on innovative activity are realized during several years in the future, current accounting-based net profit is a rather noisy measure of R&D benefits. Pakes (1985) focuses on dynamic relationships among a number of successful patent applications, R&D expenditures and stock market value of firms. Pakes concludes that events that lead the market to revalue the firm are significantly correlated with unpredictable changes in both R&D and patent applications. Pakes' work avoids the problem of timing differential

Table 2

Patent innovations leadership classification of firms.

Firm name (SIC)	Cluster	(V1)	(V2)	(V3)	(V4)	(V5)	(V6)	(V7)	(V8)	(V9)	
1. Merck (2834)	IL	GL	217.6	1367.5	136.7	147232.4	12.39	8.47	13.20	0.82	1.16
2. Eli Lilly (2834)	IF	GL	116.0	613.6	58.6	43645.5	12.21	8.15	12.69	0.83	1.17
3. Abbott Lab. (2834)	IF	GL	97.5	720.8	73.9	40954.3	11.89	7.91	12.67	0.80	1.13
4. Warner-Lambert (2834)	IF	GL	81.7	656.2	61.3	31542.0	10.64	7.23	10.55	0.75	1.29
5. Pfizer (2834)	IF	GL	103.0	553.2	49.1	23373.0	12.21	8.32	12.79	0.81	1.16
6. Bristol-Myers (2834)	IF	GL	69.7	307.4	34.2	11509.1	12.11	8.27	12.95	0.80	1.14
7. Am. Home Prod. (2834)	IF	GL	52.8	330.7	30.7	8396.9	10.82	8.05	11.38	0.72	1.25
8. Alza (2834)	IF	GL	35.5	547.6	40.9	7683.0	8.24	5.28	9.92	0.78	1.01
9. Mallinckrodt (2835)	IF	GF	23.5	181.9	16.9	2007.8	9.25	6.92	9.82	0.68	1.11
10. Pharmacia & U. (2834)	IF	GF	21.4	45.9	8.5	1922.6	10.96	7.45	10.34	0.79	1.34
11. Church & Dwight (2840)	IF	GF	12.3	83.4	9.9	1537.8	7.96	5.10	9.74	0.66	0.89
12. NeoRx (2835)	IF	GF	6.5	68.4	7.8	500.5	7.25	4.12	7.92	1.07	0.96
13. Alliance Pharma. (2834)	IF	GF	4.2	69.9	6.9	369.8	7.42	4.33	8.83	1.06	0.87
14. Xoma (2836)	IF	GF	6.9	48.7	5.0	329.6	8.11	4.21	8.84	1.15	0.96
15. Enzon (2836)	IF	GF	4.2	47.1	6.0	235.9	6.95	4.13	8.79	1.08	0.83
16. Guilford Pharma. (2834)	IF	GF	3.5	23.0	4.9	216.5	6.75	4.20	6.98	0.98	1.01
17. Sugen (2836)	IF	GF	4.4	23.8	4.0	216.2	6.50	4.02	6.42	0.95	1.05
18. Inhale Therap. (2834)	IF	GF	2.5	31.3	7.0	169.3	6.33	4.17	7.15	0.92	0.97
19. Corvas (2836)	IF	GF	3.5	16.3	2.1	122.5	6.58	4.07	7.00	1.00	1.00
20. Molecular Bios. (2835)	IF	GF	2.0	57.5	4.8	90.5	6.97	4.23	7.97	1.02	0.93

Notes: Standard Industry Classification (SIC); Innovation Leader (IL); Innovation Follower (IF); Group of Leaders (GL); Group of Followers (GF). The table presents nine variables for 20 out of the 111 firms of the technological cluster. Average values for period 1979–2000 of the following variables are presented: (V1) patent application count; (V2) forward citations received count; (V3) forward citations received count corrected for sample truncation bias; (V4) knowledge stock; (V5) log R&D expenses; (V6) log book value of equity; (V7) log market value of equity; (V8) log R&D expenses to log sales; (V9) log R&D expenses to log stock market value.

Table 3

Knowledge stock of some firms from the technological cluster.

Year	Merck	Eli Lilly	Abbott Lab.	Warner-Lambert	Pfizer	Bristol-Myers	American Home P.	Alza
1979	31,215	4,269	4,467	571	4,804	778	1,087	1,049
1980	51,807	13,245	6,246	1,411	6,131	1,237	2,269	2,805
1981	65,132	21,253	6,629	3,498	7,800	1,679	3,108	3,014
1982	73,959	21,034	7,164	4,159	8,658	1,739	3,135	3,681
1983	79,546	22,916	7,026	8,318	9,752	2,061	3,422	3,815
1984	83,616	25,138	6,928	16,406	11,298	2,521	3,433	4,806
1985	89,006	25,415	6,517	30,040	16,978	3,292	3,961	5,523
1986	97,238	23,248	7,754	43,104	18,579	3,694	5,771	6,628
1987	110,944	20,924	9,491	46,594	20,046	4,022	7,457	7,599
1988	108,461	19,228	12,499	50,251	24,377	4,988	7,516	9,036
1989	115,519	18,659	22,947	51,266	24,662	5,125	8,781	8,867
1990	136,414	18,617	28,476	52,747	29,798	6,483	8,892	9,472
1991	168,611	18,200	41,039	48,972	29,716	6,086	12,374	10,247
1992	204,970	20,146	50,468	44,339	31,188	8,039	13,667	11,131
1993	201,721	33,182	59,326	43,195	29,225	13,129	12,963	10,693
1994	213,937	46,093	70,367	41,515	31,009	15,414	14,425	10,754
1995	224,626	125,948	103,236	42,818	31,387	22,760	14,590	11,092
1996	233,309	112,243	100,738	37,540	29,102	26,825	13,133	9,837
1997	246,212	108,158	100,803	34,863	31,735	28,880	11,966	10,861
1998	248,862	98,847	91,705	33,734	30,075	30,646	10,735	10,660
1999	235,728	86,997	83,547	31,169	42,459	32,501	10,286	9,403
2000	218,279	76,439	73,621	27,414	45,426	31,301	11,762	8,054

Notes: Knowledge stock is $\sum_{s=1}^t FC_{is} P_{is} (1 - \delta)^{t-s}$ for $t = 1, \dots, T$, where P_{is} denotes patent application count and FC_{is} represents forward patent citations, corrected for sample truncation bias.

of R&D expenses and the associated future cash flow to equity, since current stock prices are determined by a forward-looking perspective of investors.

The benchmark model of our empirical analysis is Pakes (1985), who studied a panel of 120 US firms for period 1968–1975. Pakes (1985) formulates the following system of three equations in order to measure dynamic and simultaneous interactions among stock return q_{it} ; log R&D expenses r_{it} ; log patent application count $\ln P_{it}$:

$$\begin{bmatrix} q_{it} \\ r_{it} \\ \ln P_{it} \end{bmatrix} = \begin{bmatrix} \epsilon_{it} + \eta_{1it} \\ \sum_{\tau=0}^{\infty} c_{2\tau} \epsilon_{it-\tau} \\ \sum_{\tau=0}^{\infty} c_{3\tau} \epsilon_{it-\tau} + \sum_{\tau=0}^{\infty} b_{3\tau} \eta_{3it-\tau} \end{bmatrix} \quad (4.1)$$

where $\eta_{1it} \sim N(0, \sigma_1^2)$, $\epsilon_{it} \sim N(0, \sigma_2^2)$ and $\eta_{3it} \sim N(0, \sigma_3^2)$ are independent. Eq. (4.1) specifies contemporaneous and dynamic

interactions among the endogenous variables, according to a restricted Vector Moving Average, VMA(∞) representation. Similar to Pakes (1985), in each equation of Eq. (4.1), we include time effects, a dummy variable to control for zero patent application count and firm size. Pakes (1985) is consistent with the empirical results of Fama (1970) and LeRoy and Porter (1981), since he uses the no arbitrage condition to model a one-period stock return as the sum of excess return and an uncorrelated error term. According to this condition, the noise term process does not allow investors who use publicly available information and simple trading rules to make excess returns on the market. An unexpected research-related event that shifts the value of the firm motivates managers to change the R&D program of the firm hence R&D expenses are determined by the weighted sum of current and past excess stock returns (ϵ_{it}). In this model, patent applications are influenced by current and past excess returns ϵ_{it} , and also by current and past values of an i.i.d.

Table 4
Parameter estimates and model diagnostics for Models 1 and 2.

Model 1				Model 2			
ζ matrix		Residual diagnostics		ζ^* matrix		Residual diagnostics	
γ_0	0.05 [*] (0.032)	χ^2 test	<i>p</i> -value	ζ_{11}^*	-0.06 ^{**} (0.029)	χ^2 test	<i>p</i> -value
		Var e_{1it}	0.12	ζ_{12}^*	0.05 ^{***} (0.016)	Var e_{1it}	0.12
		Var e_{2it}	0.18	ζ_{13}^*	-0.01(0.029)	Var e_{2it}	0.17
		Var e_{3it}	0.13	ζ_{21}^*	0.09 ^{***} (0.018)	Var e_{3it}	0.12
ζ_{22}	0.78 ^{***} (0.032)	LB test	<i>p</i> -value	ζ_{22}^*	0.77 ^{***} (0.032)	LB test	<i>p</i> -value
		LB e_{1it}	0.50	ζ_{23}^*	0.02(0.025)	LB e_{1it}	0.52
		LB e_{2it}	0.49	ζ_{31}^*	-0.01(0.033)	LB e_{2it}	0.50
		LB e_{3it}	0.33	ζ_{32}^*	0.09 ^{***} (0.023)	LB e_{3it}	0.33
ζ_{32}	0.05(0.032)	DH test	<i>p</i> -value	ζ_{33}^*	0.23 ^{***} (0.045)	DH test	<i>p</i> -value
ζ_{33}	0.23 ^{***} (0.045)	DH e_{1it}	0.12	$\rho(\zeta^*)$	0.780	DH e_{1it}	0.12
$\rho(\zeta)$	0.777	DH e_{2it}	0.12	Cholesky matrix, Ω^*		DH e_{2it}	0.12
Cholesky matrix, Ω		DH e_{3it}	0.11	σ_1^*	0.62 ^{***} (0.067)	DH e_{3it}	0.11
$\tilde{\sigma}_1$	0.63 ^{***} (0.067)	Model diagnostics		σ_2^*	0.57 ^{***} (0.036)	Model diagnostics	
$\tilde{\sigma}_2$	0.57 ^{***} (0.036)	LL	-7151	σ_3^*	0.74 ^{***} (0.025)	LL	-7132
σ_3	0.74 ^{***} (0.025)	AIC	15002	σ_{12}^*	0.04 ^{***} (0.011)	AIC	14978
$\tilde{\sigma}_{12}$	0.03 ^{***} (0.011)	BIC	17032	σ_{13}^*	-0.01(0.019)	BIC	17049
				σ_{23}^*	0.03 [*] (0.019)	LR	37.851
						LR <i>p</i> -value	0.000

Notes: Model 1 is $Y_{it} = a_i + \zeta Y_{it-1} + \Omega e_{it}$ with $Y_{it} = (q_{it}, r_{it}, \ln P_{it})'$. Model 2 is $Y_{it} = a_i + \zeta^* Y_{it-1} + \Omega^* e_{it}$ with $Y_{it} = (q_{it}, r_{it}, \ln P_{it})'$. Ljung-Box (LB); Doornik-Hansen (DH); Log Likelihood (LL); Akaike Information Criterion (AIC); Bayesian Information Criterion (BIC); Likelihood Ratio (LR); spectral radius $\rho(\cdot)$. Robust standard errors are reported in parentheses. We report average of *p*-values (where the average is computed for $i = 1, \dots, N$) for the χ^2 test with H_0 : variance of errors is one; LB test with H_0 : are independent; DH test with H_0 : errors have normal distribution. The LB test is performed for 5 lags.

^{*} Denotes parameter significance at the 10% level.
^{**} Denotes parameter significance at the 5% level.
^{***} Denotes parameter significance at the 1% level.

adjustment factor that represents patent propensity η_{3it} (Scherer, 1965a,b).

Pakes (1985) assumes that $\epsilon_{it} = \theta q_{it} + v_{it}$ so that v_{it} and q_{it} are orthogonal. Then, Eq. (4.1) in the VAR(∞) representation with contemporaneous relationships imposed can be written as

$$\begin{aligned} q_{it} &= \epsilon_{it} + \eta_{1it} \\ r_{it} &= c_{20}\theta q_{it} + \zeta_{22}(L)r_{it-1} + c_{20}v_{it} \\ \ln P_{it} &= \gamma_0 r_{it} + \zeta_{32}(L)r_{it-1} + \zeta_{33}(L)\ln P_{it-1} + \eta_{3it}. \end{aligned} \quad (4.2)$$

To obtain a VAR(1) in structural form, we assume that the coefficients in Eq. (4.2) satisfy that $\zeta_{m\tau} = \zeta_m^\tau$ for $m = 22, 32, 33$. Then, Eq. (4.2) can be written in a reduced form VAR(1) as

$$\begin{aligned} \begin{bmatrix} q_{it} \\ r_{it} \\ \ln P_{it} \end{bmatrix} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & \zeta_{22} & 0 \\ 0 & \gamma_0 \zeta_{22} + \zeta_{32} & \zeta_{33} \end{bmatrix} \begin{bmatrix} q_{it-1} \\ r_{it-1} \\ \ln P_{it-1} \end{bmatrix} \\ &+ \begin{bmatrix} 1 & 0 & 0 \\ c_{20}\theta & 1 & 0 \\ c_{20}\theta\gamma_0 & \gamma_0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{it} + \eta_{1it} \\ c_{20}v_{it} \\ \eta_{3it} \end{bmatrix}. \end{aligned} \quad (4.3)$$

We can also write Eq. (4.3) with a vector of standard normal i.i.d. error terms as follows:

$$\begin{aligned} \begin{bmatrix} q_{it} \\ r_{it} \\ \ln P_{it} \end{bmatrix} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & \zeta_{22} & 0 \\ 0 & \gamma_0 \zeta_{22} + \zeta_{32} & \zeta_{33} \end{bmatrix} \begin{bmatrix} q_{it-1} \\ r_{it-1} \\ \ln P_{it-1} \end{bmatrix} \\ &+ \begin{bmatrix} \tilde{\sigma}_1 & 0 & 0 \\ \tilde{\sigma}_{12} & \tilde{\sigma}_2 & 0 \\ \gamma_0 \tilde{\sigma}_{12} & \gamma_0 \tilde{\sigma}_2 & \sigma_3 \end{bmatrix} \begin{bmatrix} e_{1it} \\ e_{2it} \\ e_{3it} \end{bmatrix} \end{aligned} \quad (4.4)$$

where $(e_{1it}, e_{2it}, e_{3it})' \sim N(0_{3 \times 1}, I_3)$. Furthermore, $\tilde{\sigma}_1$, $\tilde{\sigma}_{12}$ and $\tilde{\sigma}_2$ can be expressed by the parameters of Eq. (4.3). Eq. (4.4), derived from Eq. (4.3) in Appendix A, is the restricted VAR(1) formulation of Pakes (1985). The diagonal elements of the lower triangular matrix associated to the error vector of Eq. (4.4) are positive thus the Cholesky decomposition of the covariance matrix of errors is unique and the covariance matrix of errors is positive definite.

Model 1. We extend the benchmark model of Pakes (1985) to a restricted PVAR(1) model by considering fixed effects a_i in Eq. (4.4):

$$\begin{aligned} \begin{bmatrix} q_{it} \\ r_{it} \\ \ln P_{it} \end{bmatrix} &= \begin{bmatrix} a_{q,i} \\ a_{r,i} \\ a_{p,i} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & \zeta_{22} & 0 \\ 0 & \gamma_0 \zeta_{22} + \zeta_{32} & \zeta_{33} \end{bmatrix}}_{\zeta} \begin{bmatrix} q_{it-1} \\ r_{it-1} \\ \ln P_{it-1} \end{bmatrix} \\ &+ \underbrace{\begin{bmatrix} \tilde{\sigma}_1 & 0 & 0 \\ \tilde{\sigma}_{12} & \tilde{\sigma}_2 & 0 \\ \gamma_0 \tilde{\sigma}_{12} & \gamma_0 \tilde{\sigma}_2 & \sigma_3 \end{bmatrix}}_{\Omega} \begin{bmatrix} e_{1it} \\ e_{2it} \\ e_{3it} \end{bmatrix} \end{aligned} \quad (4.5)$$

for $i = 1, \dots, N$ firms and $t = 1, \dots, T$ periods. The spectral radius of ζ , $\rho(\zeta)$, is less than one in order to have covariance stationary time series. We can also formulate Model 1 in a compact matrix notation. The endogenous variables of the three-dimensional PVAR(1) model are $Y_{it} = (q_{it}, r_{it}, \ln P_{it})'$. In the PVAR equation, fixed effects are denoted by $a_i = (a_{q,i}, a_{r,i}, a_{p,i})'$ and error terms are summarized by $e_{it} = (e_{1it}, e_{2it}, e_{3it})'$. Then, the model can be written as $Y_{it} = a_i + \zeta Y_{it-1} + \Omega e_{it}$. The Impulse Response Function (IRF) matrix Θ_j , derived in Appendix B, is $\Theta_j = \zeta^j \Omega$ for $j = 0, \dots, \infty$.

Model 2. The unrestricted PVAR(1) model with fixed effects is given by

$$\begin{aligned} \begin{bmatrix} q_{it} \\ r_{it} \\ \ln P_{it} \end{bmatrix} &= \begin{bmatrix} a_{q,i} \\ a_{r,i} \\ a_{p,i} \end{bmatrix} + \underbrace{\begin{bmatrix} \zeta_{11}^* & \zeta_{12}^* & \zeta_{13}^* \\ \zeta_{21}^* & \zeta_{22}^* & \zeta_{23}^* \\ \zeta_{31}^* & \zeta_{32}^* & \zeta_{33}^* \end{bmatrix}}_{\zeta^*} \begin{bmatrix} q_{it-1} \\ r_{it-1} \\ \ln P_{it-1} \end{bmatrix} \\ &+ \underbrace{\begin{bmatrix} \sigma_1^* & 0 & 0 \\ \sigma_{12}^* & \sigma_2^* & 0 \\ \sigma_{13}^* & \sigma_{23}^* & \sigma_3^* \end{bmatrix}}_{\Omega^*} \begin{bmatrix} e_{1it} \\ e_{2it} \\ e_{3it} \end{bmatrix} \end{aligned} \quad (4.6)$$

for $i = 1, \dots, N$ firms and $t = 1, \dots, T$ periods, where $\sigma_1^* > 0$, $\sigma_2^* > 0$, $\sigma_3^* > 0$, and the spectral radius of ζ^* , $\rho(\zeta^*)$, is less than one. We can formulate Model 2 in a matrix notation as $Y_{it} = a_i + \zeta^* Y_{it-1} + \Omega^* e_{it}$. The IRF matrix Θ_j is given by $\Theta_j = (\zeta^*)^j \Omega^*$ for $j = 0, \dots, \infty$; see Appendix B.

Estimation results. Models 1 and 2 are estimated by the Quasi-Maximum Likelihood (QML) method (Binder et al., 2005).

Table 4 presents parameter estimates, robust standard errors and diagnostic tests. Fig. 2 presents the off-diagonal elements of the IRF matrix for $j = 0, \dots, 20$ leads.

We summarize first the model diagnostic test results. For each firm i , for the residual time series of each endogenous variable (i.e., estimates of $e_{1it}, e_{2it}, e_{3it}$ with $t = 1, \dots, T$), we perform the following statistical tests: (a) χ^2 test with H_0 : variance of errors is one; (b) Ljung and Box (1978, LB) test with H_0 : errors are independent; (c) Doornik and Hansen (2008, DH) test with H_0 : errors have normal distribution. For each endogenous variable, Table 4 reports the average p -value, where the average is computed over $i = 1, \dots, N$. The table shows that on average we are not able to reject H_0 at the 10% level of significance. Furthermore, according to Table 4, the spectral radius of ζ is less than one for both models, i.e., we find covariance stationary time series.

Performance of Models 1 and 2 is evaluated by likelihood-based model selection metrics (Hamilton, 1994). Table 4 shows that according to the Likelihood Ratio (LR) test and Akaike Information Criterion (AIC), the more general Model 2 is preferred. Nevertheless, according to the Bayesian Information Criterion (BIC), the nested Model 1 is preferred. We summarize empirical results for both specifications.

In Table 4 (panel Model 1), the parameter estimates of the ζ matrix evidence significant Granger causality (Hamilton, 1994) of log R&D expenses on log patent application count since $\gamma_0\zeta_{22} + \zeta_{32}$ is significant. The IRF analysis of Model 1, presented in Fig. 2, suggests a positive dynamic impact of stock return shocks on log R&D leads (Θ_{21}), and also positive impact of stock return shocks on log patent application count leads (Θ_{31}). These results are similar to the findings of Pakes (1985, pp. 403–404). We can summarize the IRF results as follows:

$$\text{IRFs of Model 1: } q \xrightarrow{(+)} r \xrightarrow{(+)} \ln P$$

where \rightarrow indicates low IRF impact and \Rightarrow shows strong IRF impact. From the IRF diagram of Model 1, we can see that the transmission channel of returns to patents is through R&D investment (Pakes, 1985). Returns only affect R&D weakly and there is no feedback effect in this model.

In Table 4 (panel Model 2), the estimates of ζ^* show significant Granger causality of log R&D on stock return (ζ_{12}^*), stock return on log R&D (ζ_{21}^*) and log R&D on log patent application count (ζ_{32}^*). The IRF analysis of Model 2, presented in Fig. 2, provides the following evidence. First, stock return shocks have positive effects both on log R&D (Θ_{21}) and log patent application count (Θ_{31}). These effects are more significant for log R&D, which are observed contemporaneously and for all lags. For log patent counts, the positive effects start from the second lead. Second, log R&D expenditure shocks have significant positive impact on both stock return (Θ_{12}) and log patent application count (Θ_{32}), for all leads. We can summarize the IRF results as follows:

$$\begin{aligned} \text{IRFs of Model 2: } & q \xrightarrow{(+)} r \xrightarrow{(+)} \ln P \\ & q \xleftarrow{(+)} r \xleftarrow{(+)} \ln P \\ & q \xrightarrow{(+)} \ln P. \end{aligned}$$

The IRF diagram of Model 2 indicates that with a more flexible model (less restricted than Model 1), the connection between returns and R&D is stronger. Furthermore, there is a strong feedback from patent applications to R&D and a weak impact of R&D on the stock return of the firm, which are transmitted to patent applications. All these effects are positive.

4.2. Patented and non-patented R&D activity

We use the latent-factor patent count panel data model of Blazsek and Escribano (2010) to estimate patented and non-patented R&D activities of firms. This model involves a dynamic

specification for patent propensity. We model the conditional distribution of patent application count P_{it} of firm i in period t by the Poisson distribution with heterogeneous patent application intensity parameter $\lambda_{it} = P_{it}^o P_{it}^*$ (Hausman et al., 1984). In our model, P_{it}^o is the R&D output of firm i in period t and $P_{it}^* \in (0, 1)$ is the patent propensity of firm i in period t . P_{it}^* measures the proportion of patent applications, which result from an innovation of a given quality. P_{it}^* adjusts R&D activity for trade secrecy instead of revealing R&D information by patents (Kahn, 1962; Machlup, 1962) or the absorptive capacity of rents from R&D (Scherer, 1965b; Arora et al., 2008). If $P_{it}^* \simeq 1$, then $\lambda_{it} \simeq P_{it}^o$ and the total R&D activity will be submitted to the patent office. However, in general, only part of the R&D activity generates patents therefore $\lambda_{it} = P_{it}^o P_{it}^*$. If P_{it}^* is higher, then absorptive capacity of the firm will be higher or the portion of innovation productivity kept secret will be lower (Escribano et al., 2009). Furthermore, $(1 - P_{it}^*) \in (0, 1)$ is the propensity not to patent of firm i in period t , which is determined by the common latent factor of the technological cluster. $(1 - P_{it}^*)$ indicates the propensity of R&D investment decisions that do not generate patent applications. $(1 - P_{it}^*)$ represents non-appropriated R&D activity and adjusts R&D investment decisions for trade secrecy in order to avoid the disclosure of the information included in patent applications to competitors. We define the R&D output that does not generate patent applications (non-patented R&D activity) by $P_{it}^x = P_{it}^o(1 - P_{it}^*)$.

The conditional probability mass function of patent application count P_{it} is

$$f(P_{it} | \mathcal{F}_t) = \frac{\exp(-\lambda_{it})(\lambda_{it})^{P_{it}}}{P_{it}!} = \frac{\exp(-P_{it}^o P_{it}^*)(P_{it}^o P_{it}^*)^{P_{it}}}{P_{it}!} \quad (4.7)$$

for $i = 1, \dots, N$ firms and $t = 1, \dots, T$ periods (for period 1979–2000). The conditioning set is

$$\mathcal{F}_t = [(P_{i1}, r_{i1}, BV_{i1}, BC_{intra,i1}, BC_{inter,i1}, I_{i1}^*), \dots, (P_{it-1}, r_{it-1}, BV_{it-1}, BC_{intra,it-1}, BC_{inter,it-1}, I_{t-1}^*), (r_{it}, BV_{it}, BC_{intra,it}, BC_{inter,it}, I_t^*) : i = 1, \dots, N] \quad (4.8)$$

where r_{it} denotes log of R&D expenditure; BV_{it} is log of book value of equity; $BC_{intra,it}$ is the number of patent citations to past patents of other firms in the technological cluster (intra-cluster backward citations); $BC_{inter,it}$ is the number of backward citations to past patents of non-technologically-related firms, i.e., firms not in the technological cluster (inter-cluster backward citations); I_t^* is a common latent factor that drives patent propensity in the technological cluster.

Model 3. We specify the conditional expectation of P_{it} by a patent count panel data model, which is a combination of Poisson and probit models with latent common factor, as follows:

$$E(P_{it} | \mathcal{F}_t) = \lambda_{it} = P_{it}^o P_{it}^* \quad (4.9)$$

$$\begin{aligned} \ln P_{it}^o = & \mu_0 + \gamma_1 t + \gamma_2 t r_{it} + \gamma_3 r_{it}^2 + \gamma_4 BV_{it} \\ & + \gamma_5 P_{i1} + \sum_{k=0}^q \beta_k r_{it-k} + \sum_{k=0}^q \omega_k BC_{intra,it-k} r_{it} \\ & + \sum_{k=0}^q \phi_k BC_{inter,it-k} r_{it} + \sum_{k=1}^p \kappa_k \ln P_{it-k}^o \end{aligned} \quad (4.10)$$

$$\ln P_{it}^* = \ln \Phi(\mu_i + \sigma_i I_t^*) \quad (4.11)$$

$$I_t^* = \mu^* I_{t-1}^* + u_t \quad \text{with } u_t \sim N(0, 1) \text{ i.i.d.} \quad (4.12)$$

Eq. (4.10) specifies the log R&D activity component $\ln P_{it}^o$. We generate R&D activity, P_{it}^o , based on actual and past values of R&D expenditure adjusted for innovation quality (backward citations), past innovation outputs and the initial condition on patent applications. In Eq. (4.10), μ_0 is the constant parameter; γ_1 and γ_2 control for linear time trend; γ_3 captures non-linearities

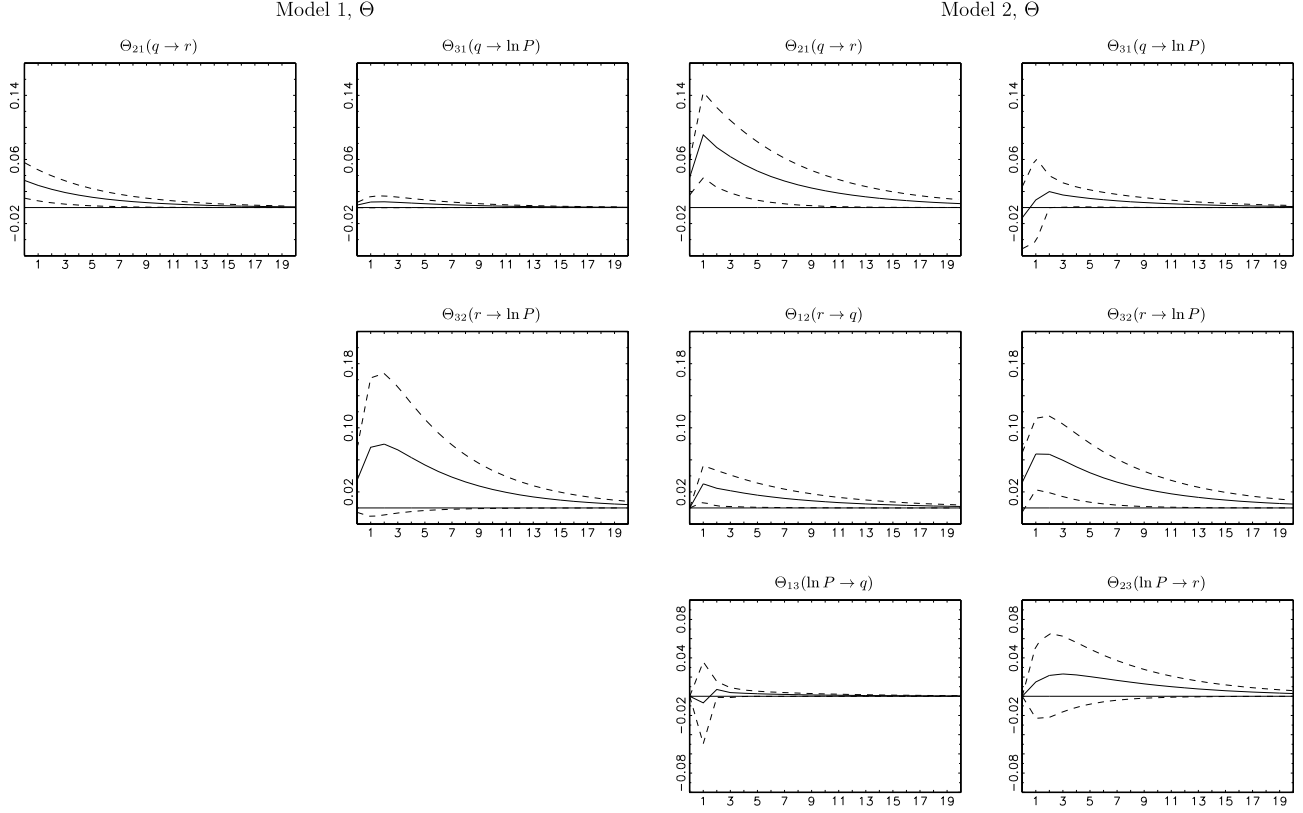


Fig. 2. Off-diagonal elements of the IRF matrix Θ_j for Models 1 and 2 for $j = 0, \dots, 20$ leads. *Notes:* The confidence band is defined by $\bar{\Theta}_j \pm 2\sigma(\Theta_j)$.

in log R&D expenditure; γ_4 measures the impact of firm size; γ_5 controls for initial conditions; β_k with $k = 0, \dots, q$ capture finite distributed lag effects of log R&D expenses; ω_k and ϕ_k with $k = 0, \dots, q$ control interactions of finite distributed lags of intra-cluster and inter-cluster backward citations, respectively, and current log R&D expenses; κ_k controls for AR dynamics (Blazsek and Escribano, 2010).

Eq. (4.11) specifies the patent propensity component P_{it}^* according to a probit model. In this equation, Φ is the cumulative distribution function of the standard normal distribution; μ_i is firm-specific fixed effect; $\sigma_i \in \mathbb{R}$ is a firm-specific scaling parameter for the common latent factor, l_t^* .

Eq. (4.12) specifies the common latent factor of the technological cluster l_t^* that affects all firms. l_t^* is an aggregate index of technological cluster characteristics in period t , which represents the degree of market competition, degree of deregulation, degree of enforcement of intellectual property rights and degree of information that is common knowledge. In Eq. (4.12), $|\mu^*| < 1$ measures the average persistence of the latent factor in the technological cluster. The constant term in this equation is restricted to zero and the dynamic parameter μ^* is assumed to be the same for all firms, since these are determined by common knowledge on technological cluster characteristics. These restrictions also help parameter identification. We extend the specification of Pakes (1985), since the propensity to patent time series $\{P_{it}^* : t = 1, \dots, T\}$ for firm i is serially correlated and driven by the dynamic latent common factor of the technological cluster. Nevertheless, P_{it}^* is firm-specific due to the level and scaling parameters μ_i and σ_i , respectively.

Estimation results. We estimate the latent-factor patent count data model by the Simulated ML (SML) method, applying the Efficient Importance Sampling (EIS) technique of Richard and Zhang (2007). Appendix C presents this method. Other applications of EIS are Liesenfeld and Richard (2003), Bauwens and Hautsch (2006) and Blazsek and Escribano (2010). EIS is a refinement of the

recursive importance sampling algorithm of Keane (1994), with respect to the construction of the source distribution of the draws in a way that gives greater efficiency. Robust standard errors of parameters are computed by the sandwich estimator (Davidson and MacKinnon, 2009).

Parameter estimates, robust standard errors and model diagnostic tests are presented in the first panel of Table 5. This table shows that the latent common factor is covariance stationary with $\mu^* = 0.91$. Fig. 1(a) presents the evolution of total patent application count $\sum_{i=1}^N P_{it}$ and total patent application intensity $\sum_{i=1}^N \lambda_{it}$ in the technological cluster for period 1979–2000. The figure shows that the latent-factor patent count panel data model fits well with patent application count time series, for most years. The only exceptions are 1995 and 1996, when significant differences are observed between patent counts and fitted patent intensity. These outliers may partly be due to the implementation of the General Agreement on Tariffs and Trade on 8 June 1995, which influenced effective pharmaceutical patent life (Grabowski and Vernon, 2000). The outliers in 1995 and 1996 may also be due to significant shifts in the level of market competition (see Section 5, which supports this idea).

We compute filtered estimates of the latent common factor by estimating $E[l_t^* | \mathcal{F}_t^o]$ for all t . In this expectation, we condition on the following observable information set:

$$\mathcal{F}_t^o = \left[(P_{i1}, r_{i1}, BV_{i1}, BC_{intra,i1}, BC_{inter,i1}), \dots, (P_{it-1}, r_{it-1}, BV_{it-1}, BC_{intra,it-1}, BC_{inter,it-1}), (r_{it}, BV_{it}, BC_{intra,it}, BC_{inter,it}) : i = 1, \dots, N \right]. \quad (4.13)$$

The estimation of $E[l_t^* | \mathcal{F}_t^o]$ involves the ratio of two high-dimensional integrals, each evaluated by the EIS technique. Appendix D presents the computation of $E[l_t^* | \mathcal{F}_t^o]$. The evolution of the latent common factor for period 1979–2000 is presented in Fig. 1(c). The figure shows that the unobservable common

Table 5

Parameter estimates and model diagnostics for Models 3 and 6.

	r_{it-k}		$BC_{intra,it-k}r_{it}$		$BC_{inter,it-k}r_{it}$		$BC_{inter,it-k}r_{it}$	Diagnostics	
Model 3: Poisson and probit models with latent common factor									
μ_0	0.50*** (0.090)	β_0	0.63*** (0.016)	ω_0	0.00** (0.000)	ϕ_0	0.00(0.000)	LL	-3741
$\gamma_1 t$	0.14*** (0.002)	β_1	-0.04*** (0.009)	ω_1	0.00(0.000)	ϕ_1	0.00(0.000)	AIC	8007
$\gamma_2 tr_{it}$	-0.01*** (0.001)	β_2	0.02*** (0.007)	ω_2	0.00(0.000)	ϕ_2	0.00(0.000)	BIC	8294
$\gamma_3 r_{it}^2$	-0.04*** (0.003)	β_3	-0.03*** (0.004)	ω_3	0.00(0.001)	ϕ_3	0.00(0.000)	LR	255.288
$\gamma_4 BV_{it}$	0.01*** (0.001)	β_4	-0.03*** (0.002)	ω_4	0.00(0.001)	ϕ_4	0.00(0.000)	LR p	0.000
$\gamma_5 P_{i1}$	0.05*** (0.001)	β_5	0.03*** (0.005)	ω_5	0.00(0.001)	ϕ_5	0.00(0.001)		
κ_1	0.00(0.000)	β_6	-0.01*** (0.002)	ω_6	0.00(0.004)	ϕ_6	0.00(0.001)		
μ^*	0.91*** (0.007)	β_7	-0.01*** (0.002)	ω_7	0.00(0.005)	ϕ_7	0.00(0.004)		
		β_8	-0.01*** (0.002)	ω_8	0.01(0.005)	ϕ_8	0.00(0.007)		
		β_9	-0.01(0.006)	ω_9	0.00(0.006)	ϕ_9	0.00(0.007)		
		β_{10}	0.02*** (0.008)	ω_{10}	0.01(0.007)	ϕ_{10}	-0.01(0.013)		
Model 6: Poisson and probit models with observable competition factor									
μ_0	0.83*** (0.080)	β_0	0.51*** (0.021)	ω_0	0.00 (0.000)	ϕ_0	0.00(0.000)	LL	-3868
$\gamma_1 t$	0.15*** (0.004)	β_1	-0.05*** (0.007)	ω_1	0.00(0.000)	ϕ_1	0.00(0.000)	AIC	8262
$\gamma_2 tr_{it}$	-0.01*** (0.001)	β_2	0.02*** (0.007)	ω_2	0.00(0.000)	ϕ_2	0.00(0.000)	BIC	8549
$\gamma_3 r_{it}^2$	-0.04*** (0.003)	β_3	-0.04*** (0.007)	ω_3	0.00(0.001)	ϕ_3	0.00(0.000)		
$\gamma_4 BV_{it}$	0.01*** (0.002)	β_4	-0.01*** (0.007)	ω_4	0.00(0.001)	ϕ_4	0.00(0.000)		
$\gamma_5 P_{i1}$	0.07*** (0.002)	β_5	0.02*** (0.008)	ω_5	0.00(0.001)	ϕ_5	0.00(0.001)		
κ_1	0.00(0.000)	β_6	-0.02*** (0.007)	ω_6	0.00(0.003)	ϕ_6	0.00(0.001)		
μ^*	0.84*** (0.126)	β_7	-0.02*** (0.008)	ω_7	0.00(0.005)	ϕ_7	0.00(0.005)		
μ_0^*	0.15(0.112)	β_8	-0.02*** (0.006)	ω_8	0.01(0.005)	ϕ_8	0.00(0.007)		
		β_9	-0.01*** (0.007)	ω_9	0.00(0.006)	ϕ_9	0.00(0.007)		
		β_{10}	0.03*** (0.007)	ω_{10}	0.01(0.007)	ϕ_{10}	-0.01(0.013)		

Notes: Model 3 is specified as $\lambda_{it} = P_{it}^0 P_{it}^*$ where $\ln P_{it}^0 = \mu_0 + \gamma_1 t + \gamma_2 tr_{it} + \gamma_3 r_{it}^2 + \gamma_4 BV_{it} + \gamma_5 P_{i1} + \sum_{k=0}^{10} \beta_k r_{it-k} + \sum_{k=0}^{10} \omega_k BC_{intra,it-k} r_{it} + \sum_{k=0}^{10} \phi_k BC_{inter,it-k} r_{it} + \kappa_1 \ln P_{it-1}$, $\ln P_{it}^* = \ln \Phi(\mu_i + \sigma_i I_{it}^*)$ and $I_{it}^* = \mu_i^* I_{t-1}^* + u_t$. Model 6 is $\lambda_{it} = P_{it}^0 P_{it}^*$ where $\ln P_{it}^0 = \mu_0 + \gamma_1 t + \gamma_2 tr_{it} + \gamma_3 r_{it}^2 + \gamma_4 BV_{it} + \gamma_5 P_{i1} + \sum_{k=0}^{10} \beta_k r_{it-k} + \sum_{k=0}^{10} \omega_k BC_{intra,it-k} r_{it} + \sum_{k=0}^{10} \phi_k BC_{inter,it-k} r_{it} + \kappa_1 \ln P_{it-1}$, $\ln P_{it}^* = \ln \Phi(\mu_i + \sigma_i CO_{it})$ and $CO_{it} = \mu_0^* + \mu^* CO_{t-1}^* + u_t$. Log Likelihood (LL); Akaike Information Criterion (AIC); Bayesian Information Criterion (BIC); Likelihood Ratio (LR); LR p is the p-value of the LR test. We use the non-nested LR test of [Vuong \(1989\)](#). We do not report the estimates of μ_i and σ_i . Robust standard errors are reported in parentheses.

* Denotes parameter significance at the 10% level.

** Denotes parameter significance at the 5% level.

*** Denotes parameter significance at the and 1% level.

factor decreases until 1990 and rapidly increases to a higher level afterwards.

[Fig. 1\(f\)](#) presents the evolution of mean patent propensity ($1/N \sum_{i=1}^N P_{it}^*$) in the technological cluster for period 1979–2000. The figure shows that during the 1990s, the percentage of patent applications increased from about 7.2% to above 10%. For 1996 and 1997, we see outliers for this variable but afterwards the level stabilizes above 10%.

[Fig. 3](#) presents first and second-order polynomial regression results about the determinants of patent propensity in the technological cluster. The figure shows fitted values of $(1/T) \sum_{i=1}^T \ln P_{it}^*$ for $i = 1, \dots, N$ firms, regressed on the mean of the variables (V1), (V2), (V5) and (V7) computed for period 1979–2000. The figure exhibits the estimates of the regression model fitted to mean $\ln P_{it}^*$ and the corresponding R-squared values, in order to inform about the explanatory power of each variable. [Fig. 3\(a\)](#) shows an inverted-U relationship for log of mean patent application count (V1) and mean log patent propensity. For most firms in the technological cluster, higher patent application count is associated with higher patent propensity level. Nevertheless, for some firms with a high number of patent applications, we see that patent propensity is relatively low. [Fig. 3\(b\)](#) shows a linear increasing relation between log mean citations received count (V2) and mean log patent propensity. Similar to [Fig. 3\(a\)](#), we see that some firms with a high number of citations received from subsequent patents, exhibit relatively low patent propensity. [Fig. 3\(c\)](#) shows an inverted-U relationship for mean log R&D expenses (V5) and mean log patent propensity. For most firms in the technological cluster, higher R&D expenses are associated with higher patent propensity. Nevertheless, some firms that exhibit high R&D expenditure have a relatively low patent propensity. [Fig. 3\(d\)](#) shows a positive linear relation between mean log stock market value and log patent

propensity. We also see that a high level of patent propensity is significantly associated with high firm value.

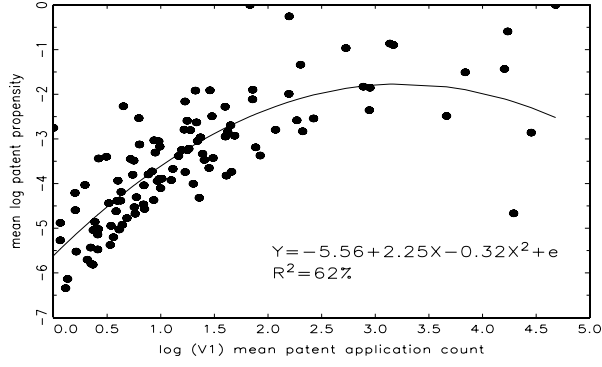
4.3. Extended innovation and market value model

We use the estimates of P_{it}^0 and P_{it}^* and extend the unrestricted three-dimensional innovation and market value Model 2, by proposing a four-dimensional dynamic model for stock return q_{it} ; log R&D expenses r_{it} ; log patent application count $\ln P_{it}$; log non-patented R&D activity $\ln P_{it}^x = P_{it}^0 (1 - P_{it}^*)$.

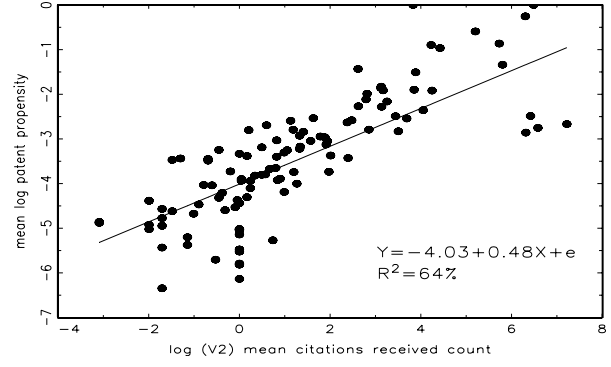
Model 4. The unrestricted PVAR(1) model with fixed effects is given by

$$\begin{bmatrix} q_{it} \\ r_{it} \\ \ln P_{it} \\ \ln P_{it}^x \end{bmatrix} = \begin{bmatrix} a_{q,i} \\ a_{r,i} \\ a_{p,i} \\ a_{x,i} \end{bmatrix} + \underbrace{\begin{bmatrix} \zeta_{11}^* & \zeta_{12}^* & \zeta_{13}^* & \zeta_{14}^* \\ \zeta_{21}^* & \zeta_{22}^* & \zeta_{23}^* & \zeta_{24}^* \\ \zeta_{31}^* & \zeta_{32}^* & \zeta_{33}^* & \zeta_{34}^* \\ \zeta_{41}^* & \zeta_{42}^* & \zeta_{43}^* & \zeta_{44}^* \end{bmatrix}}_{\zeta^*} \begin{bmatrix} q_{it-1} \\ r_{it-1} \\ \ln P_{it-1} \\ \ln P_{it-1}^x \end{bmatrix} + \underbrace{\begin{bmatrix} \sigma_{11}^* & 0 & 0 & 0 \\ \sigma_{12}^* & \sigma_2^* & 0 & 0 \\ \sigma_{13}^* & \sigma_{23}^* & \sigma_3^* & 0 \\ \sigma_{14}^* & \sigma_{24}^* & \sigma_{34}^* & \sigma_4^* \end{bmatrix}}_{\Omega^*} \begin{bmatrix} e_{1it} \\ e_{2it} \\ e_{3it} \\ e_{4it} \end{bmatrix} \quad (4.14)$$

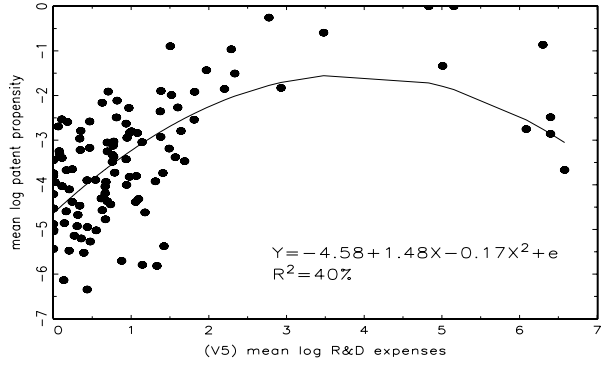
for $i = 1, \dots, N$ firms and $t = 1, \dots, T$ periods, where $(e_{1it}, e_{2it}, e_{3it}, e_{4it})' \sim N(0_{4 \times 1}, I_4)$, and the spectral radius of ζ^* , $\rho(\zeta^*)$, is less than one. The diagonal elements of Ω^* are positive thus the Cholesky decomposition of the covariance matrix of errors is unique and the covariance matrix of errors is positive definite. In each equation of Eq. (4.14), we include time effects, a zero patent application count dummy and firm size ([Pakes, 1985](#)). We consider the same variables in the PVAR model that measures spillovers



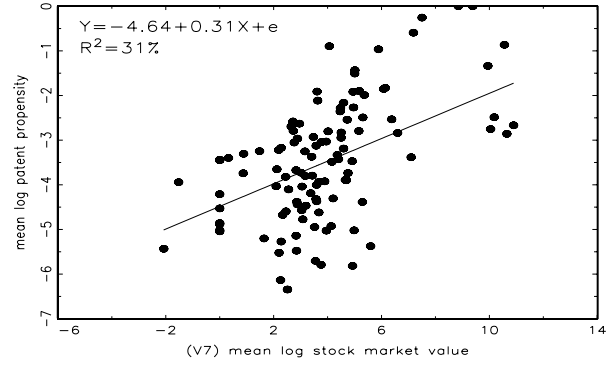
(a) Patent application count.



(b) Citations received count.



(c) R&D expenses.



(d) Stock market value of equity.

Fig. 3. Determinants of patent propensity. *Notes:* The figure shows the fitted values of $(1/T) \sum_{t=1}^T \ln P_{it}^*$ for $i = 1, \dots, N$. First and second order polynomial regressions are estimated by least squares. The definition of each explanatory variable is presented in Table 2.

in Section 4.4.² The endogenous variables of the four-dimensional PVAR(1) model are $Y_{it} = (q_{it}, r_{it}, \ln P_{it}, \ln P_{it}^{\times})'$. In the PVAR equation, fixed effects are denoted by $a_i = (a_{q,i}, a_{r,i}, a_{P,i}, a_{\times,i})'$ and error terms are summarized by $e_{it} = (e_{1it}, e_{2it}, e_{3it}, e_{4it})'$. We can formulate Model 4 in a compact matrix notation as $Y_{it} = a_i + \zeta^* Y_{it-1} + \Omega^* e_{it}$. The IRF matrix Θ_j , derived in Appendix B, is given by $\Theta_j = (\zeta^*)^j \Omega^*$ for $j = 0, \dots, \infty$.

Estimation results. Model 4 is estimated by the QML method. Table 6 (Model 4 panel) presents the parameter estimates, robust standard errors and model diagnostics. Fig. 4 presents the off-diagonal elements of the IRF matrix for $j = 0, \dots, 20$ leads. According to the residual diagnostic tests, on average, we are not able to reject model specification assumptions at the 10% level of significance. Furthermore, we find that Model 4 is covariance stationary.

The IFR figures of Model 4 show similar effects among q_{it} , r_{it} and $\ln P_{it}$ as in Model 2 (see Figs. 2 and 4). We focus on those IRFs that involve non-patented R&D activity $\ln P_{it}^{\times}$. Fig. 4 shows that stock return (Θ_{41}), log R&D expenses (Θ_{42}) and log patent count (Θ_{43}) shocks have a positive impact on log non-patented R&D activity. Furthermore, Fig. 4 also shows that the effects of log non-patented R&D activity shocks on other variables (Θ_{14} , Θ_{24} , and Θ_{34}) are less significant although the IRF is positive on average. We can

summarize the IRF results as follows:

$$\begin{aligned} \text{IRF of Model 4: } & q \xrightarrow{(+)} r \xrightarrow{(+)} \ln P \xleftrightarrow{(+)} \ln P^{\times} \xleftrightarrow{(+)} \text{latent common factor} \\ & q \xleftrightarrow{(+)} r \xleftrightarrow{(+)} \left\{ \begin{array}{l} \ln P \\ \ln P^{\times} \xleftrightarrow{(+)} \text{latent common factor} \end{array} \right. \\ & q \xrightarrow{(+)} \ln P \end{aligned}$$

where \rightarrow indicates low IRF impact and \Rightarrow shows strong IRF impact. Two-way arrows indicate feedback (low impact is denoted by \leftrightarrow and high impact is denoted by \Leftrightarrow). The flow diagram of Model 4 has similar transmission channels as Model 2, but it also includes a new component related to the common latent component, associated with the patent propensity. Patent propensity enhances the relation with R&D through the feedback with the R&D-related component of patent applications. Furthermore, patent propensity is highly and positively correlated with the latent common factor (see Section 5).

4.4. Extended model for patent innovation leaders and followers

We consider different interaction effects (spillovers) among stock return q_{it} , log R&D expenses r_{it} , log patent application count $\ln P_{it}$ and log non-patented R&D activity $\ln P_{it}^{\times}$ for patent innovation leaders and followers of the technological cluster. We formulate this model by using a compact matrix notation.

Model 5. The unrestricted PVAR(1) model for IL and IF with fixed effects, is given by

$$\begin{aligned} Y_{it} = & a_i + \zeta^* Y_{it-1} + \zeta_{IL}^* Y_{iL,t-1} D_{it} (i \in \text{IF}) \\ & + \zeta_{IF}^* \left(\sum_{k \in \text{IF}} Y_{kt-1} \right) D_{it} (i = \text{IL}) + \Omega^* e_{it} \end{aligned} \quad (4.15)$$

² We also estimate restricted four-dimensional PVAR model similar to that of Pakes (1985). Blazsek and Escribano (2014) report these results. As we obtain results for the restricted four-dimensional model that are similar to the unrestricted four-dimensional model (and moreover, as all likelihood-based model selection criteria support the unrestricted model), in this work we only report results for the unrestricted four-dimensional innovation and market value model.

Table 6
Parameter estimates and model diagnostics for Models 4 and 5.

Model 4				Model 5					
ζ^* matrix		Residual diagnostics		ζ^* matrix		ζ_{IL}^* matrix		Residual diagnostics	
ζ_{11}^*	-0.06** (0.029)	χ^2 test	p-value	ζ_{11}^*	-0.07** (0.033)	$\zeta_{IL,11}^*$	0.06(0.103)	χ^2 test	p-value
ζ_{12}^*	0.04*** (0.015)	Var e_{1it}	0.12	ζ_{12}^*	0.04(0.039)	$\zeta_{IL,12}^*$	-0.54*** (0.138)	Var e_{1it}	0.14
ζ_{13}^*	-0.01(0.030)	Var e_{2it}	0.17	ζ_{13}^*	-0.01(0.026)	$\zeta_{IL,13}^*$	-3.47(9.555)	Var e_{2it}	0.17
ζ_{14}^*	0.00(0.016)	Var e_{3it}	0.12	ζ_{14}^*	0.00(0.133)	$\zeta_{IL,14}^*$	0.87*** (0.197)	Var e_{3it}	0.17
ζ_{21}^*	0.09*** (0.018)	Var e_{4it}	0.10	ζ_{21}^*	0.08** (0.036)	$\zeta_{IL,21}^*$	-0.12(0.092)	Var e_{4it}	0.10
ζ_{22}^*	0.77*** (0.039)	LB test	p-value	ζ_{22}^*	0.75*** (0.021)	$\zeta_{IL,22}^*$	0.05(0.141)	LB test	p-value
ζ_{23}^*	0.02(0.025)	LB e_{1it}	0.51	ζ_{23}^*	0.06** (0.026)	$\zeta_{IL,23}^*$	-4.45(11.771)	LB e_{1it}	0.50
ζ_{24}^*	-0.01(0.120)	LB e_{2it}	0.49	ζ_{24}^*	0.07* (0.038)	$\zeta_{IL,24}^*$	1.12*** (0.193)	LB e_{2it}	0.49
ζ_{31}^*	-0.01(0.034)	LB e_{3it}	0.34	ζ_{31}^*	0.02(0.029)	$\zeta_{IL,31}^*$	-0.17* (0.100)	LB e_{3it}	0.33
ζ_{32}^*	0.09*** (0.024)	LB e_{4it}	0.14	ζ_{32}^*	0.09*** (0.035)	$\zeta_{IL,32}^*$	-0.10(0.100)	LB e_{4it}	0.25
ζ_{33}^*	0.23*** (0.045)	DH test	p-value	ζ_{33}^*	0.19*** (0.019)	$\zeta_{IL,33}^*$	4.20(11.215)	DH test	p-value
ζ_{34}^*	0.02(0.028)	DH e_{1it}	0.11	ζ_{34}^*	0.07(0.111)	$\zeta_{IL,34}^*$	-0.77*** (0.195)	DH e_{1it}	0.11
ζ_{41}^*	0.03*** (0.008)	DH e_{2it}	0.12	ζ_{41}^*	0.02(0.018)	$\zeta_{IL,41}^*$	-0.05* (0.032)	DH e_{2it}	0.12
ζ_{42}^*	0.00(0.016)	DH e_{3it}	0.12	ζ_{42}^*	-0.02** (0.008)	$\zeta_{IL,42}^*$	0.22*** (0.056)	DH e_{3it}	0.15
ζ_{43}^*	0.01(0.011)	DH e_{4it}	0.21	ζ_{43}^*	0.04*** (0.012)	$\zeta_{IL,43}^*$	-1.32(3.376)	DH e_{4it}	0.22
ζ_{44}^*	0.31*** (0.085)	Model diagnostics		ζ_{44}^*	0.41*** (0.012)	$\zeta_{IL,44}^*$	0.36*** (0.076)	Model diagnostics	
$\rho(\zeta^*)$	0.78	LL	-6956	$\rho(\zeta^*)$	0.763	ζ_{IF}^* matrix		LL	-6597
Cholesky matrix, Ω^*		AIC	14876	Cholesky matrix, Ω^*		$\zeta_{IF,11}^*$	0.00(0.253)	AIC	14221
σ_1^*	0.62*** (0.067)	BIC	17672	σ_1^*	0.62*** (0.005)	$\zeta_{IF,12}^*$	-0.01(0.337)	BIC	17202
σ_2^*	0.57*** (0.036)			σ_2^*	0.55*** (0.009)	$\zeta_{IF,13}^*$	0.00(0.140)	LR	718.925
σ_3^*	0.74*** (0.025)			σ_3^*	0.72*** (0.012)	$\zeta_{IF,14}^*$	0.00(0.711)	LR p-value	0.000
σ_4^*	0.22*** (0.016)			σ_4^*	0.21*** (0.002)	$\zeta_{IF,21}^*$	0.00(0.959)		
σ_{12}^*	0.04*** (0.011)			σ_{12}^*	0.02(0.028)	$\zeta_{IF,22}^*$	0.00(0.406)		
σ_{13}^*	-0.01(0.019)			σ_{13}^*	0.00(0.023)	$\zeta_{IF,23}^*$	0.00(0.252)		
σ_{23}^*	0.03* (0.018)			σ_{23}^*	0.06** (0.026)	$\zeta_{IF,24}^*$	0.01(1.255)		
σ_{14}^*	0.01*** (0.005)			σ_{14}^*	0.01(0.018)	$\zeta_{IF,31}^*$	0.00(1.180)		
σ_{24}^*	0.14*** (0.016)			σ_{24}^*	0.13*** (0.006)	$\zeta_{IF,32}^*$	0.00(1.363)		
σ_{34}^*	0.01** (0.004)			σ_{34}^*	0.01(0.013)	$\zeta_{IF,33}^*$	0.00(1.523)		
						$\zeta_{IF,34}^*$	0.01(2.708)		
						$\zeta_{IF,41}^*$	0.00(0.307)		
						$\zeta_{IF,42}^*$	0.00(0.104)		
						$\zeta_{IF,43}^*$	0.00(0.044)		
						$\zeta_{IF,44}^*$	0.01(0.339)		

Notes: Model 4 is $Y_{it} = a_i + \zeta^* Y_{it-1} + \Omega^* e_{it}$ with $Y_{it} = (q_{it}, r_{it}, \ln P_{it}, \ln P_{it}^X)'$. Model 5 is $Y_{it} = a_i + \zeta^* Y_{it-1} + \zeta_{IL}^* Y_{it-1} D_{it}(i \in IF) + \zeta_{IF}^* (\sum_{k \in IF} Y_{kt-1}) D_{it}(i = IL) + \Omega^* e_{it}$ with $Y_{it} = (q_{it}, r_{it}, \ln P_{it}, \ln P_{it}^X)'$. Ljung-Box (LB); Doornik-Hansen (DH); Log Likelihood (LL); Akaike Information Criterion (AIC); Bayesian Information Criterion (BIC); Likelihood Ratio (LR); spectral radius $\rho(\cdot)$. Robust standard errors are reported in parentheses. We report average of p-values (where the average is computed over $i = 1, \dots, N$) for the χ^2 test with H_0 : variance of errors is one; LB test with H_0 : errors are independent; DH test with H_0 : errors have normal distribution. The LB test is performed for 5 lags.

* Denotes parameter significance at the 10% level.

** Denotes parameter significance at the 5% level.

*** Denotes parameter significance at the 1% level.

for $i = 1, \dots, N$ firms and $t = 1, \dots, T$ periods, where ζ^* , ζ_{IL}^* and ζ_{IF}^* are unrestricted.³ The spectral radius of ζ^* is less than one. Ω^* is a lower triangular matrix with positive diagonal elements. Appendix B shows that the IRF matrix Θ_j is $\Theta_j = (\zeta^*)^j \Omega^*$ for $j = 0, \dots, \infty$, and the matrices of dynamic interaction multipliers are

$$\Gamma_j(IL \rightarrow IF) = (\text{effects of } Y_{iL,t-j} \text{ on } Y_{it} \text{ for } i \in IF) \\ = (\zeta^*)^j \zeta_{IL}^* \text{ for } j = 0, 1, 2, \dots, \infty \quad (4.16)$$

$$\Gamma_j(IF \rightarrow IL) = (\text{effects of } Y_{k,t-j} \text{ on } Y_{iL,t} \text{ for } k \in IF) \\ = (\zeta^*)^j \zeta_{IF}^* \text{ for } j = 0, 1, 2, \dots, \infty \quad (4.17)$$

Estimation results. Model 5 is estimated by the QML method. Table 6 (Model 5 panel) presents the parameter estimates, robust standard errors and diagnostic test results. Fig. 5 presents the off-diagonal elements of the IRF matrix for $j = 0, \dots, 20$ leads. Fig. 6

presents the matrix of dynamic interaction multipliers from IL to IFs for $j = 0, \dots, 20$ leads.

According to the residual diagnostic tests, on average we are not able to reject the model specification assumptions at the 10% level of significance. We find that Model 5 is covariance stationary. Moreover, Table 6 shows that Model 5 is superior to Model 4 with respect to the likelihood-based model performance metrics.

The IRFs of Model 5 show similar dynamic effects to the IRFs of Model 4. In the remaining part of this section, we focus on the dynamic effects among the patent innovation leader and its followers. Fig. 6 shows the dynamic interaction multipliers measuring spillovers from the IL (Merck) to IF companies, and provides the following results. First, we see positive spillover effects of all IL variables on IF stock return, starting from the first and second leads (see Γ_{11} , Γ_{12} , Γ_{13} , and Γ_{14}). The highest spillover effects are associated with the IL's log patent application count and log non-patented R&D activity on IF firms' stock returns. We can also see that IL's log R&D expenses have a contemporaneous negative effect on IFs' stock return. However, this effect changes to positive sign from the first lead (see Γ_{12}). Second, we see significant positive dynamic effects of IL's log non-patented R&D activity on IFs' log R&D expenses (Γ_{24}) and non-patented R&D activity (Γ_{44}). For both variables, Fig. 6 shows a clear spillover pattern, over several years. Third, Γ_{34} shows that IL's log non-patented R&D activity have negative contemporaneous effect on

³ We also estimate restricted four-dimensional PVAR model similar to that of Pakes (1985) with patent innovation leaders and followers. Blazsek and Escrignano (2014) report these results. As we obtain results for the restricted four-dimensional model that are similar to the unrestricted four-dimensional model (and moreover, as all likelihood-based model selection criteria support the unrestricted model), in this work we only report results for the unrestricted four-dimensional innovation and market value model with patent innovation leaders and followers.

Model 4, Θ

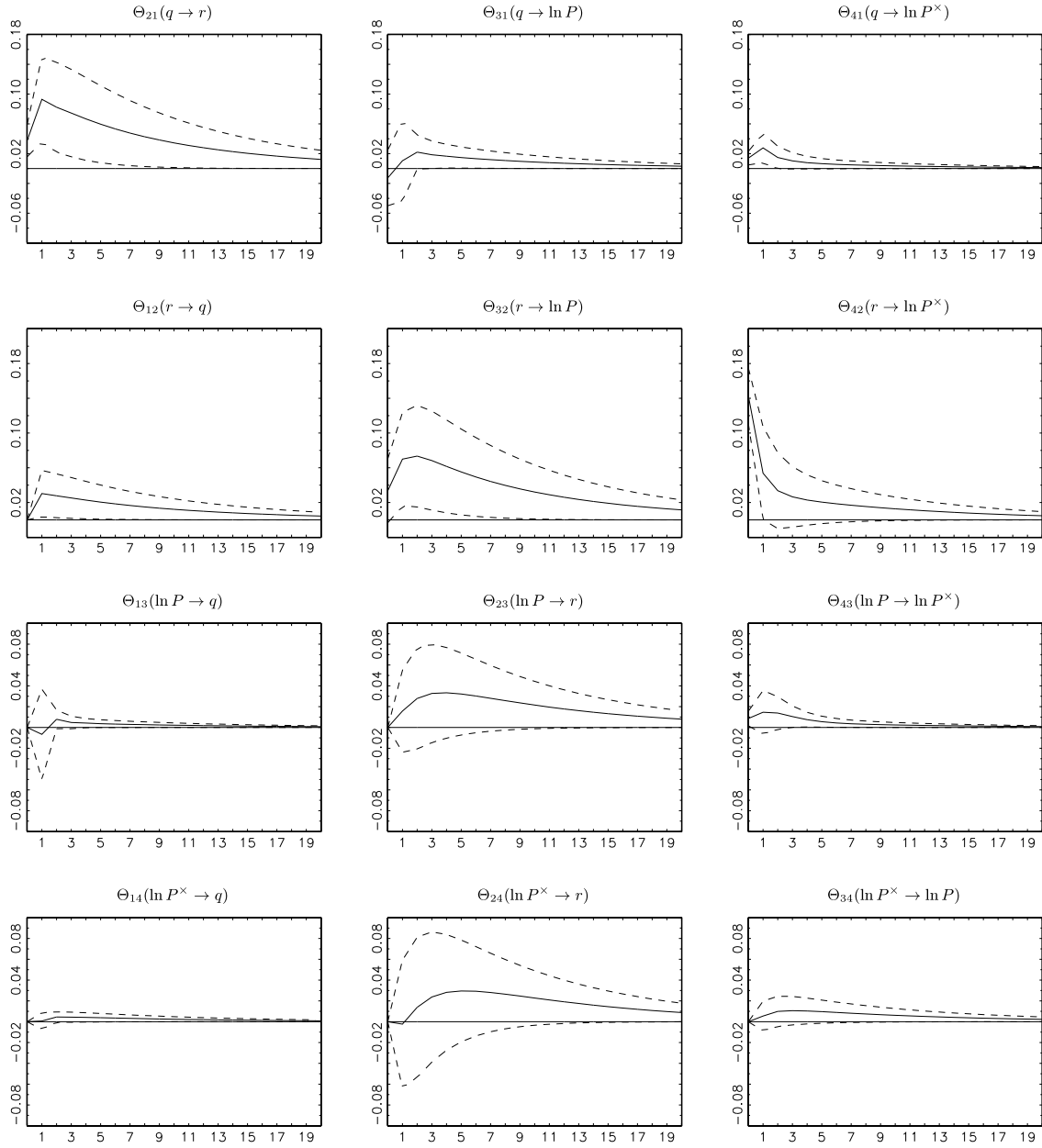


Fig. 4. Off-diagonal elements of the IRF matrix Θ_j for Model 4 for $j = 0, \dots, 20$ leads. Notes: The confidence band is defined by $\bar{\Theta}_j \pm 2\sigma(\Theta_j)$.

Dynamic multipliers of Model 5:

$$q_{IL} \xrightarrow{(+)} q_{IF}$$

$$r_{IL} \xrightarrow{(+)} \ln P_{IF} \xrightarrow{(+)} \ln P_{IF}^x \xrightarrow{(+)} \text{latent common factor}$$

$$\ln P_{IL} \xrightarrow{(+)} \ln P_{IF} \xrightarrow{(+)} \ln P_{IF}^x \xrightarrow{(+)} \text{latent common factor}$$

$$\ln P_{IL}^x \xrightarrow{(+)} r_{IF} \xrightarrow{(+)} \ln P_{IF} \xrightarrow{(+)} \ln P_{IF}^x \xrightarrow{(+)} \text{latent common factor.}$$

Box I.

IFs' log patent application count. Nevertheless, from the first lead this effect changes to positive sign for all subsequent years. Fourth, we observe positive dynamic effects of IL's log R&D expenses on IFs' non-patented R&D activity (see Γ_{42}). These findings suggest positive dynamic R&D spillovers from IL (Merck) to IF firms. For R&D (IL) and stock return (IF), and log non-patented R&D (IL) and log patent application (IF) count relations, we observe negative

contemporaneous effects from IL to IF firms that change to positive sign for all leading years. Moreover, the results also evidence that IF firms are more influenced by the log non-patented R&D activity of Merck than by its log patent application count, which emphasizes the importance of measuring the non-patented R&D activity. Dynamic interaction multiplier effects are summarized in [Box I](#).

We do not report the matrix of dynamic interaction multipliers from IF to IL firms, since none of the ζ_{IF}^* coefficients were found to be significantly different from zero (see Table 6). We also estimated the restricted Model 5 with $\zeta_{IF}^* = 0$ and performed the LR test to compare unrestricted and restricted versions of Model 5. We find that log-likelihood of these models does not differ significantly.

5. Competition and patent propensity

In this section, we discuss our results in the context of the competition and innovation model of Aghion et al. (2005). We assess the level of competition by the observable measure

$$\begin{aligned} CO_t &= 1 - \frac{1}{N} \sum_{i=1}^N LI_{it} \\ &= 1 - \frac{1}{N} \sum_{i=1}^N \frac{\text{operating profit}_{it} - \text{financial costs}_{it}}{\text{sales}_{it}} \end{aligned} \quad (5.1)$$

where LI_{it} is the Lerner index or price cost margin, which measures the degree of market power of firm i in period t (Nickell, 1996; Aghion et al., 2005). High values of this competition measure indicate competitive industry, while low values indicate market power.

5.1. Latent common factor versus observable competition factor

We study how the latent common factor, that drives patent propensity of firms, is related to the degree of competition in the technological cluster. First, we estimate an alternative patent count data model that considers CO_t as the common factor of patent propensity, replacing I_t^* . Second, we present some descriptive statistics comparing I_t^* and CO_t , for period 1979–2000, in order to provide additional arguments about the strong relationship between the latent common factor and CO_t .

Model 6. We specify the conditional expectation of P_{it} by a patent count data model that is a combination of Poisson and probit models with observable competition factor, as follows:

$$E(P_{it} | \mathcal{F}_t^o, CO_t) = \lambda_{it} = P_{it}^* P_{it}^o \quad (5.2)$$

$$\begin{aligned} \ln P_{it}^o &= \mu_0 + \gamma_1 t + \gamma_2 tr_{it} + \gamma_3 r_{it}^2 + \gamma_4 BV_{it} \\ &\quad + \gamma_5 P_{i1} + \sum_{k=0}^q \beta_k r_{it-k} + \sum_{k=0}^q \omega_k BC_{\text{intra},it-k} r_{it} \\ &\quad + \sum_{k=0}^q \phi_k BC_{\text{inter},it-k} r_{it} + \sum_{k=1}^p \kappa_k \ln P_{it-k}^o \end{aligned} \quad (5.3)$$

$$\ln P_{it}^* = \ln \Phi(\mu_i + \sigma_i CO_t) \quad (5.4)$$

$$CO_t = \mu_0^* + \mu^* CO_{t-1} + u_t \quad (5.5)$$

where u_t is an uncorrelated white noise process. The system of equations (5.2)–(5.4) is estimated by the ML method. The parameters of Eq. (5.5) are estimated by a separate AR(1) equation in order to compare the dynamics of the common factor between the two models. Parameter estimates, robust standard errors obtained by the sandwich estimator and model diagnostics are reported in the second panel of Table 5. By comparing the two patent application count data models of Table 5, we can see that estimation results are very similar. The same sign and similar level of significance are observed for each parameter. This result supports that the latent common factor is highly correlated with the degree of market competition.

The likelihood-based model performance metrics presented in Table 5 show that the latent-factor patent count data model dominates the patent count data model with observable

competition factor. This is due to the fact that the latent common factor captures additional common characteristics besides the degree of market competition (for example, degree of deregulation, degree of enforcement of intellectual property rights or degree of information that is common knowledge in the technological cluster). This finding supports the application of the latent-factor model to separate patented and non-patented R&D activity.

Additional arguments. We present the total intensity estimates $\sum_{i=1}^N \lambda_{it}$ for each model in Fig. 1(a)–(b). These figures show that both models provide similar approximation of total patent application count for period 1979–2000. Moreover, Fig. 1(c)–(d) also shows that the latent common factor and the level of market competition have a similar tendency for period 1979–2000, and they rise rapidly around 1990. This suggests that pharmaceutical firms reacted to the increasing level of market competition by patenting a significantly higher proportion of their innovation output after 1990.

5.2. Inverted-U relationship between competition and innovation

We study whether the inverted-U relationship between competition and innovation, implied by the theory of Aghion et al. (2005), exists in the pharmaceutical technological cluster. We use three measures of overall innovation in the technological cluster: mean patent application count $(1/N) \sum_{i=1}^N P_{it}$; mean R&D activity $(1/N) \sum_{i=1}^N P_{it}^o$; mean non-patented R&D activity $(1/N) \sum_{i=1}^N P_{it}^*$. Moreover, we also study the relationship between competition and mean patent propensity $(1/N) \sum_{i=1}^N P_{it}^*$. We estimate two models to competition and innovation data. First, Fig. 7 shows the estimates of a polynomial regression model that suggest an inverted-U relationship between competition CO_t and the innovation variables considered (Aghion et al., 2005). Second, as the figure shows two groups of data points, we also fit a logistic function. The fitted values indicate two clusters in Fig. 7. The cluster with lower degree of competition and lower level of innovation corresponds to data observed for period 1979–1991, while the cluster with higher degree of competition and higher level of innovation is for period 1992–2000. (Fig. 1(d) also shows that the degree of market competition CO_t in the technological cluster increases for the sample period, with a significant jump in the beginning of the 1990s.) Furthermore, Fig. 7(d) suggests that in the technological cluster, on average, as the level of competition increases, the proportion of R&D activity that generates patents also increases. The estimates of the logistic function in Fig. 7(d) suggest that pharmaceutical firms generate patents from a higher proportion of the R&D activity from the beginning of the 1990s. It is important to note that due to the small number of data points the results reported in this section are only suggestive.

5.3. Neck-and-neck firms in innovation activity

The dynamic effects reported in Fig. 6 indicate spillovers from the IL firm (Merck) to IFs. The estimation results show that Merck has rapid contemporaneous impact on followers, while IFs do not impact significantly Merck. The model of Aghion et al. (2005) implies that increasing market competition discourages laggard firms to innovate, while it encourages neck-and-neck firms to innovate. The positive dynamic spillover effects estimated in the drugs industry indicate that, in the technological cluster analyzed, firms on average are neck-and-neck in innovation activity for period 1979–2000. The spillover dynamics found are consistent with Nabseth and Ray (1974), Mansfield et al. (1981), Rogers (1983), Pakes and Schankerman (1984) and Jovanovic and MacDonald (1994), who report that innovation spills over gradually to competitors.

Model 5, Θ

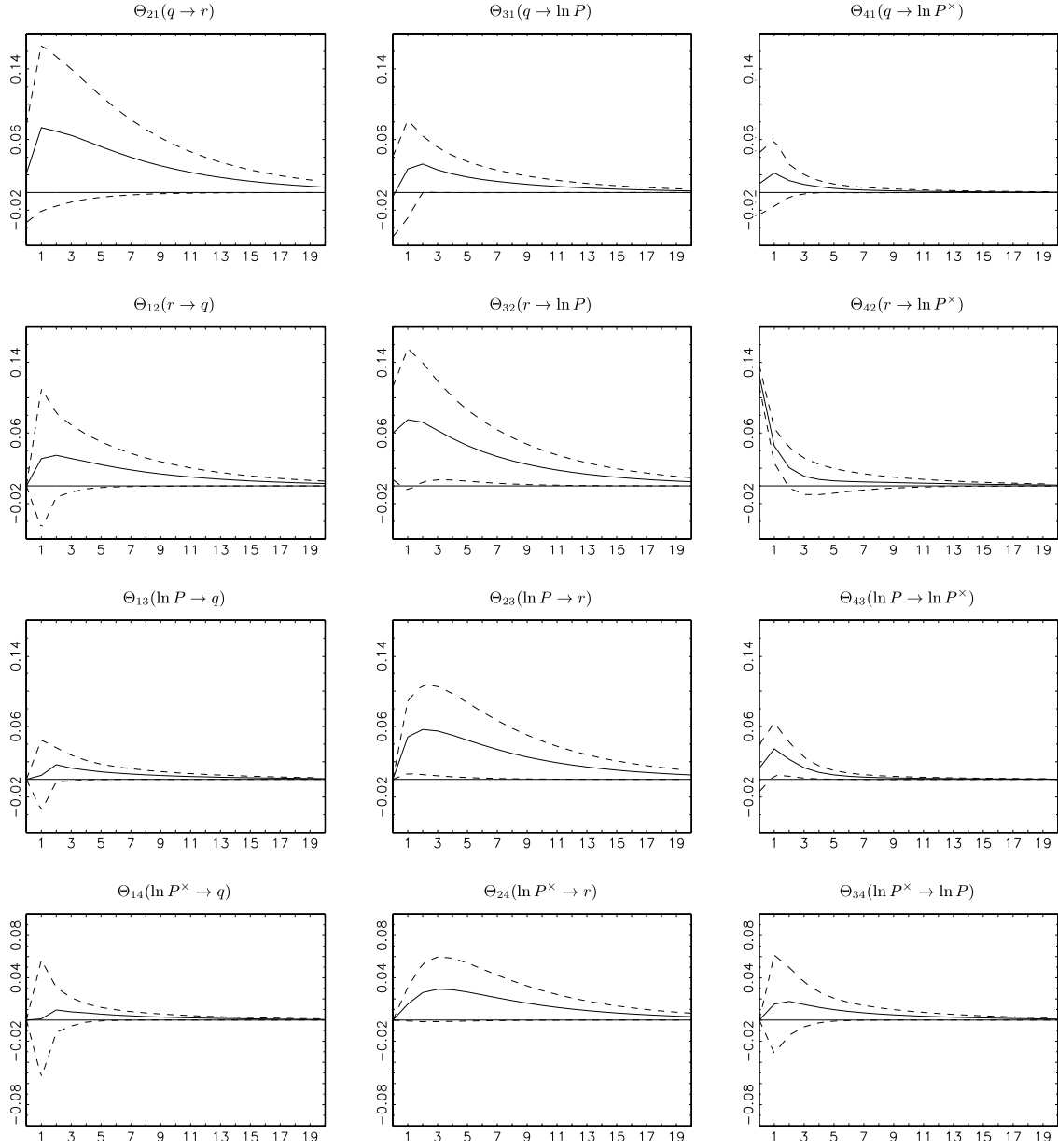


Fig. 5. Off-diagonal elements of the IRF matrix Θ_j for Model 5 for $j = 0, \dots, 20$ leads. *Notes:* The confidence band is defined by $\bar{\Theta}_j \pm 2\sigma(\Theta_j)$.

5.4. Competition and technological distance of leaders and followers

We study an implication of the theory of [Aghion et al. \(2005\)](#) for the pharmaceutical technological cluster, who suggest that the average technological distance between innovation leaders and followers increases with competition. [Aghion et al. \(2005\)](#) measure technological gap based on total factor productivity. In the present article, we use four alternative innovation-related variables in order to approximate technological distance between the IL and IFs of the technological cluster: (a) knowledge stock of firm i in period t , denoted by $KS_{it} = \sum_{s=1}^t FC_{is}P_{is}(1-\delta)^{t-s}$; (b) R&D activity P_{it}^o of firm i in period t ; (c) patent propensity P_{it}^* of firm i in period t ; (d) non-patented R&D activity P_{it}^\times of firm i in period t . By using the general innovation activity IA_{it} notation for variables (a)–(d), we approximate technological distance TD_t in the pharmaceutical technological cluster by the following measure (which is also used

by [\(Aghion et al., 2005\)](#)):

$$TD_t = \frac{1}{N} \sum_{i=1}^N \frac{IA_{iL,t} - IA_{it}}{IA_{iL,t}}. \quad (5.6)$$

Model 7. We estimate the linear regression model that controls for a non-linear time trend:

$$TD_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 CO_t + u_t \quad (5.7)$$

for $t = 1, \dots, T$. Robust OLS estimates with Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors ([Newey and West, 1987](#)) are presented in [Table 7](#). We find that competition has a significant positive effect on technological distance for all measures of innovation activity. This finding supports the implication of the model of [Aghion et al. \(2005\)](#).

Model 5, $\Gamma(\text{IL} \rightarrow \text{IF})$

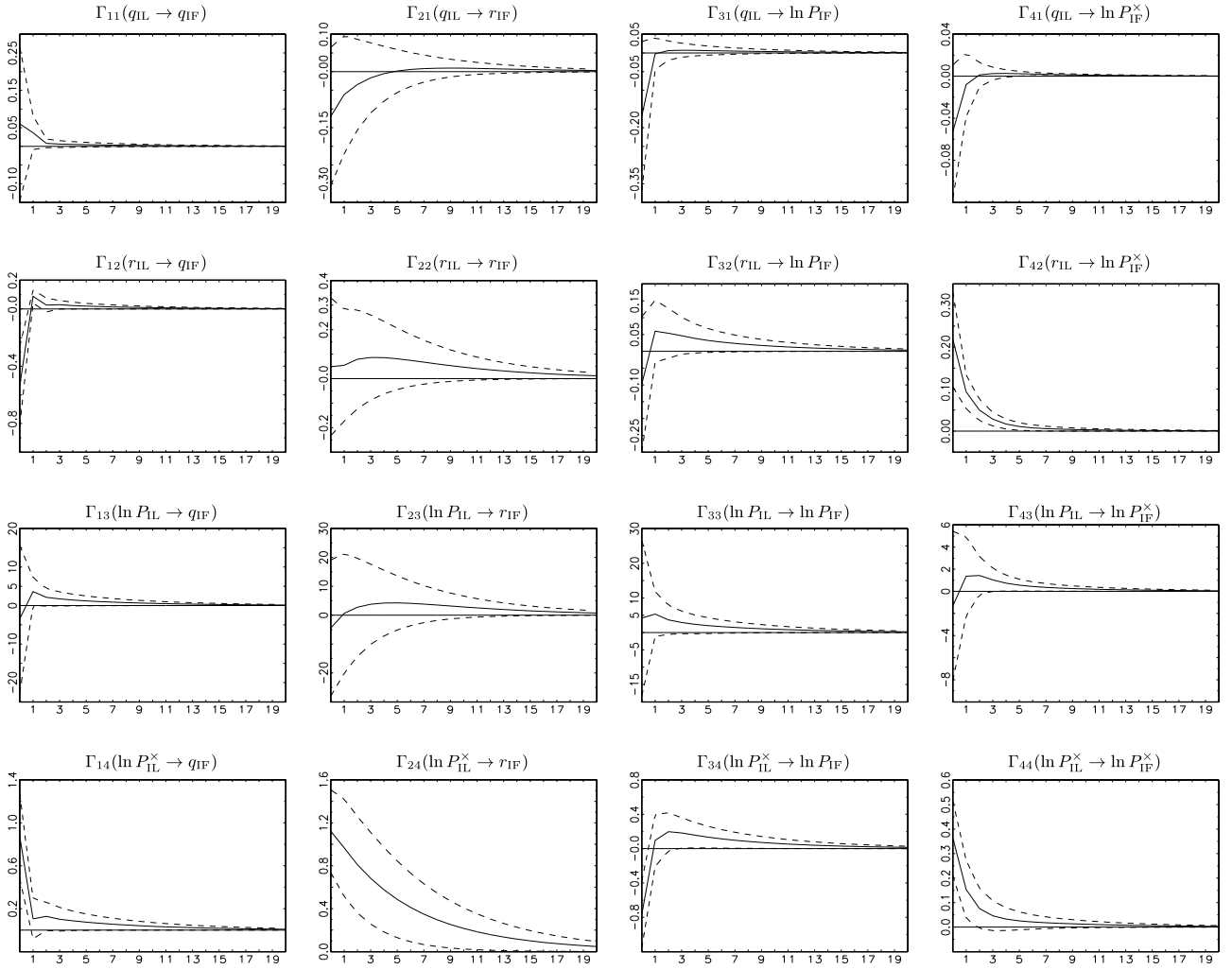


Fig. 6. Dynamic multiplier matrix $\Gamma_j(\text{IL} \rightarrow \text{IF})$ for Model 5 for $j = 0, \dots, 20$ leads. Notes: The confidence band is defined by $\bar{\Gamma}_j \pm 2\sigma(\Gamma_j)$.

Table 7

Parameter estimates and model diagnostics for Model 7.

$\text{TD}_t(\text{KS}_{it} : i = 1, \dots, N)$		$\text{TD}_t(P_{it}^0 : i = 1, \dots, N)$		$\text{TD}_t(P_{it}^* : i = 1, \dots, N)$		$\text{TD}_t(P_{it}^X : i = 1, \dots, N)$	
β_0	0.97*** (0.005)	β_0	0.92*** (0.019)	β_0	-0.10** (0.045)	β_0	0.93*** (0.016)
$\beta_1 t$	0.00*** (0.000)	$\beta_1 t$	0.00 (0.000)	$\beta_1 t$	0.00** (0.001)	$\beta_1 t$	0.00 (0.000)
$\beta_2 t^2$	0.00* (0.000)	$\beta_2 t^2$	0.00*** (0.000)	$\beta_2 t^2$	0.00** (0.000)	$\beta_2 t^2$	0.00*** (0.000)
$\beta_3 \text{CO}_t$	0.02*** (0.006)	$\beta_3 \text{CO}_t$	0.05** (0.023)	$\beta_3 \text{CO}_t$	0.22*** (0.057)	$\beta_3 \text{CO}_t$	0.05** (0.021)
Model diagnostics		Model diagnostics		Model diagnostics		Model diagnostics	
R-squared: 83%		R-squared: 89%		R-squared: 72%		R-squared: 88%	

Notes: Model 7 is $\text{TD}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \text{CO}_t + u_t$, where $\text{TD}_t(\text{IA}_{it} : i = 1, \dots, N) = (1/N) \sum_{i=1}^N (\text{IA}_{iL,t} - \text{IA}_{iF,t}) / \text{IA}_{iL,t}$. Technological distance TD_t is computed by different measures of innovation activity IA_{it} . Each panel shows estimation results for a different measure of IA_{it} . Robust standard errors are reported in parentheses (Newey and West, 1987).

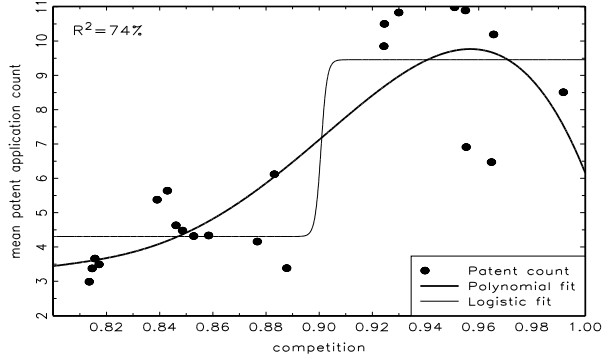
- * Denotes parameter significance at the 10% level.
- ** Denotes parameter significance at the 5% level.
- *** Denotes parameter significance at the 1% level.

6. Summary and conclusions

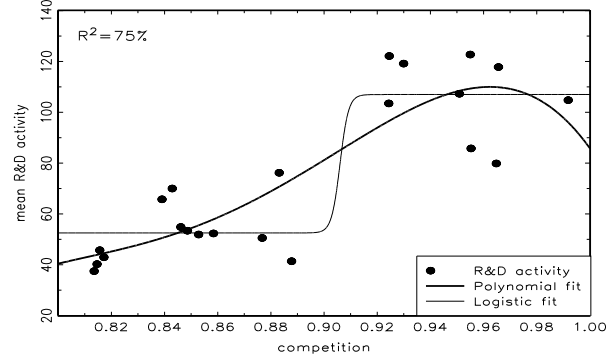
We have extended Pakes (1985) by considering different dynamic measures of innovation activity, that capture patented R&D activity (i.e., publicly disclosed innovations) and non-patented R&D activity (i.e., non-appropriated R&D or trade secrets), and innovation spillovers. Patented and non-patented R&D activities have been estimated by using a latent-factor patent count panel data model, which estimates different patent propensity for each firm. We use a large panel dataset of 4476

companies from all manufacturing and service industries of the US for period 1979–2000. Firms have been classified into different technological clusters, where each cluster includes technologically similar firms.

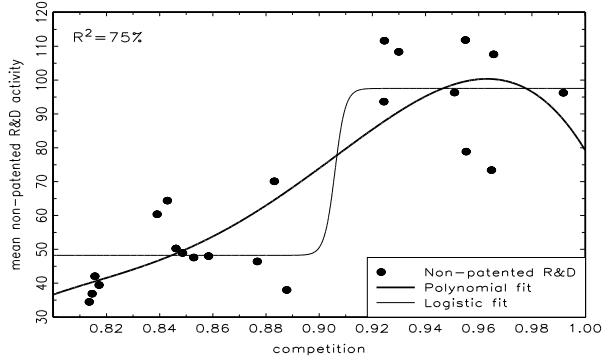
We have focused on a specific cluster of 111 firms that are mostly from the drugs market sector. In this technological cluster, the patent innovation leader (Merck) and followers have been identified. We have studied dynamic interactions between patent innovation leaders and patent innovation followers in the technological cluster, and allowed firms to have different



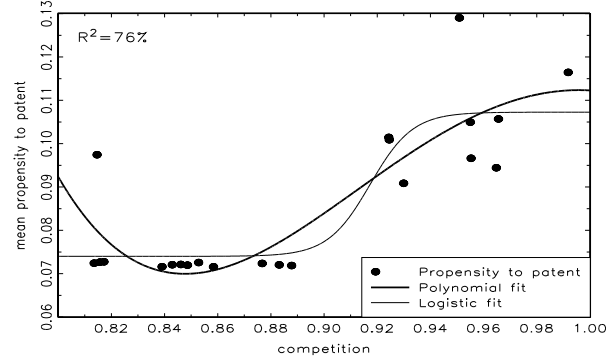
(a) Competition and mean patent application count.



(b) Competition and mean R&D activity.



(c) Competition and mean non-patented R&D activity.



(d) Competition and mean patent propensity.

Fig. 7. Competition and innovation in the pharmaceutical technological cluster. *Notes:* Competition is CO_t ; mean patent count is $(1/N) \sum_{i=1}^N P_{it}$; mean R&D activity is $(1/N) \sum_{i=1}^N P_{it}^0$; mean non-patented R&D activity is $(1/N) \sum_{i=1}^N P_{it}^0 (1 - P_{it}^*)$; mean patent propensity is $(1/N) \sum_{i=1}^N P_{it}^*$. Polynomial regression and logistic function least squares estimates of the dependent variable are presented for period $t = 1979, \dots, 2000$. The R-squares correspond to the polynomial regression model.

patent propensities. In the patent count panel data model, patent propensity is driven by the latent common factor that represents aggregate technological cluster characteristics and measures (for example, the degree of market competition, degree of deregulation, degree of enforcement of intellectual property rights and degree of information that is common knowledge). Given the estimates of patented and non-patented R&D activity, we have studied R&D spillover effects between patent innovation leader and followers by PVAR models.

Our PVAR estimates have supported the findings of [Pakes \(1985\)](#) about dynamic effects among stock return, R&D expenses and patent application counts of US firms. The extended PVAR models show that non-patented R&D activity is an important dynamic determinant of both patented and non-patented R&D activities in the cluster of technologically similar firms. We have found positive dynamic spillover effects from patent innovation leader to followers, but not the other way around. The positive spillover effects from the innovation leader to followers suggest that firms are neck-and-neck in innovation activity in the drugs technological cluster. We have found that in the beginning of the 1990s both competition and innovation increase significantly. This suggests that on average the escape competition effect dominates the Schumpeterian effect in the competition and innovation relationship in the drugs technological cluster. Market competition is the driver of the common latent factor that affects patent propensity of firms, showing that firms have patented a higher proportion of their innovation output as a reaction to the increasing level of competition in the technological cluster during the 1990s.

In summary, we find that competition is a driver of innovation, which affects positively the number of patent applications and the propensity to patent. The transmission channel between stock return and innovation output of firms is through R&D investment

decisions, as predicted by the theoretical framework of [Pakes \(1985\)](#).

Acknowledgements

We appreciate the comments received from the editor and the reviews of the associate editor, which helped us to significantly improve the previous versions of this work. We also thank the helpful discussions and comments received from Luc Bauwens, Angeles Carnero, Helmuth Chávez, Matthew Copley, Andrea Fosfuri, Marco Giarratana, Bronwyn Hall, Uwe Hassler, Jerry Hausman, Javier Hualde, Miklós Koren, Gábor Kőrösi and Martin Wagner. Previous versions of this study were presented at ETSERN Pamplona Meeting (Spain), June 2010; Summer Workshop in Economics of the Hungarian Academy of Sciences, MTA-KTI, Budapest, July 2010. The first author acknowledges and gives thanks for research funding from the School of Business of Universidad Francisco Marroquín, Guatemala. Funding from the Bank of Spain Excellence Program and the Ministry of Economy and Competitiveness of Spain (research project ECO2012-33427 and grant MDM2014-0431) is gratefully acknowledged by the second author.

Appendix A

We derive Eq. (4.4) from Eq. (4.3) as follows. Eq. (4.3) is

$$\begin{bmatrix} q_{it} \\ r_{it} \\ \ln P_{it} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \zeta_{22} & 0 \\ 0 & \gamma_0 \zeta_{22} + \zeta_{32} & \zeta_{33} \end{bmatrix} \begin{bmatrix} q_{it-1} \\ r_{it-1} \\ \ln P_{it-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ c_{20} \theta & 1 & 0 \\ c_{20} \theta \gamma_0 & \gamma_0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{it} + \eta_{1it} \\ c_{20} v_{it} \\ \eta_{3it} \end{bmatrix}. \quad (\text{A.1})$$

$$\begin{bmatrix} \epsilon_{it} + \eta_{1it} \\ c_{20}v_{it} \\ \eta_{3it} \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 + \sigma_2^2 & c_{20}[\theta^2\sigma_1^2 + (1-\theta)^2\sigma_2^2] & 0 \\ c_{20}[\theta^2\sigma_1^2 + (1-\theta)^2\sigma_2^2] & c_{20}^2[\theta^2\sigma_1^2 + (1-\theta)^2\sigma_2^2] & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} \right] \quad (\text{A.6})$$

Box II.

The first and second error terms are not orthogonal in this system since $\epsilon_{it} = \theta q_{it} + v_{it}$. First, we derive the covariance matrix of the errors in Eq. (A.1). The covariance between the first and second errors is

$$\begin{aligned} \text{Cov}(\epsilon_{it} + \eta_{1it}, c_{20}v_{it}) &= \text{Cov}(\theta q_{it} + v_{it} + \eta_{1it}, c_{20}v_{it}) \\ &= c_{20}\text{Var}(v_{it}) \end{aligned} \quad (\text{A.2})$$

where second equality uses the fact that η_{1it} , q_{it} , and v_{it} are orthogonal. We express v_{it} as

$$v_{it} = \epsilon_{it} - \theta q_{it} = \epsilon_{it} - \theta(\epsilon_{it} + \eta_{1it}) = (1-\theta)\epsilon_{it} - \theta\eta_{1it}. \quad (\text{A.3})$$

Taking the variance of this equation, we have

$$\begin{aligned} \text{Var}(v_{it}) &= (1-\theta)^2\text{Var}(\epsilon_{it}) + \theta^2\text{Var}(\eta_{1it}) \\ &= \theta^2\sigma_1^2 + (1-\theta)^2\sigma_2^2. \end{aligned} \quad (\text{A.4})$$

Therefore,

$$\text{Cov}(\epsilon_{it} + \eta_{1it}, c_{20}v_{it}) = c_{20}[\theta^2\sigma_1^2 + (1-\theta)^2\sigma_2^2]. \quad (\text{A.5})$$

The distribution of errors in Eq. (A.1) is presented in Box II.

We introduce the following notation:

$$\begin{pmatrix} \sigma_1^2 + \sigma_2^2 & c_{20}[\theta^2\sigma_1^2 + (1-\theta)^2\sigma_2^2] & 0 \\ c_{20}[\theta^2\sigma_1^2 + (1-\theta)^2\sigma_2^2] & c_{20}^2[\theta^2\sigma_1^2 + (1-\theta)^2\sigma_2^2] & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} \equiv \begin{pmatrix} \dot{\sigma}_1^2 & \dot{\sigma}_{12} & 0 \\ \dot{\sigma}_{12} & \dot{\sigma}_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}. \quad (\text{A.7})$$

The Cholesky matrix of the error covariance matrix is

$$\begin{pmatrix} \dot{\sigma}_1^2 & \dot{\sigma}_{12} & 0 \\ \dot{\sigma}_{12} & \dot{\sigma}_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}^{1/2} = \begin{pmatrix} \dot{\sigma}_1 & 0 & 0 \\ \dot{\sigma}_{12} & \sqrt{\dot{\sigma}_2^2 - \frac{\dot{\sigma}_{12}^2}{\dot{\sigma}_1^2}} & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}. \quad (\text{A.8})$$

Using this Cholesky matrix, we write Eq. (A.1) to make the error vector orthogonal:

$$\begin{bmatrix} q_{it} \\ r_{it} \\ \ln P_{it} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \zeta_{22} & 0 \\ 0 & \gamma_0\zeta_{22} + \zeta_{32} & \zeta_{33} \end{bmatrix} \begin{bmatrix} q_{it-1} \\ r_{it-1} \\ \ln P_{it-1} \end{bmatrix} + \begin{bmatrix} \dot{\sigma}_1 & 0 & 0 \\ c_{20}\theta\dot{\sigma}_1 + \frac{\dot{\sigma}_{12}}{\dot{\sigma}_1} & \sqrt{\dot{\sigma}_2^2 - \frac{\dot{\sigma}_{12}^2}{\dot{\sigma}_1^2}} & 0 \\ \gamma_0\left(c_{20}\theta\dot{\sigma}_1 + \frac{\dot{\sigma}_{12}}{\dot{\sigma}_1}\right) & \gamma_0\sqrt{\dot{\sigma}_2^2 - \frac{\dot{\sigma}_{12}^2}{\dot{\sigma}_1^2}} & \sigma_3 \end{bmatrix} \begin{bmatrix} e_{1it} \\ e_{2it} \\ e_{3it} \end{bmatrix} \quad (\text{A.9})$$

where $(e_{1it}, e_{2it}, e_{3it})' \sim N(0_{3 \times 1}, I_3)$. We simplify the notation in Eq. (A.9) by defining new parameters:

$$\begin{bmatrix} q_{it} \\ r_{it} \\ \ln P_{it} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \zeta_{22} & 0 \\ 0 & \gamma_0\zeta_{22} + \zeta_{32} & \zeta_{33} \end{bmatrix} \begin{bmatrix} q_{it-1} \\ r_{it-1} \\ \ln P_{it-1} \end{bmatrix} + \begin{bmatrix} \tilde{\sigma}_1 & 0 & 0 \\ \tilde{\sigma}_{12} & \tilde{\sigma}_2 & 0 \\ \gamma_0\tilde{\sigma}_{12} & \gamma_0\tilde{\sigma}_2 & \sigma_3 \end{bmatrix} \begin{bmatrix} e_{1it} \\ e_{2it} \\ e_{3it} \end{bmatrix} \quad (\text{A.10})$$

where $\tilde{\sigma}_1$, $\tilde{\sigma}_{12}$, and $\tilde{\sigma}_2$ can be expressed by the original parameters based on Eqs. (A.7) and (A.9).

Appendix B

IRF for Models 1, 2 and 4

We use the following compact matrix notation for Models 1, 2 and 4:

$$Y_{it} = a_i + \zeta Y_{it-1} + \Omega e_{it} \quad (\text{B.1})$$

where Y_{it} is a $K \times 1$ vector of the endogenous variables; a_i is a $K \times 1$ vector of fixed effects; Ω is a lower triangular matrix with positive elements in its diagonal; $e_{it} \sim N(0_{K \times 1}, I_K)$ is the vector of error terms. We write Eq. (B.1) as

$$(I_K - \zeta L)Y_{it} = a_i + \Omega e_{it} \quad (\text{B.2})$$

$$Y_{it} = (I_K - \zeta L)^{-1}a_i + (I_K - \zeta L)^{-1}\Omega e_{it} \quad (\text{B.3})$$

$$Y_{it} = \sum_{j=0}^{\infty} \zeta^j a_i + \sum_{j=0}^{\infty} \zeta^j \Omega e_{it-j}. \quad (\text{B.4})$$

Then, the IRF matrices are given by

$$\Theta_j = \frac{\partial Y_{it+j}}{\partial e_{it}} = \zeta^j \Omega \quad \text{for } j = 0, 1, 2, \dots, \infty. \quad (\text{B.5})$$

IRF for Model 5

We use the following compact matrix notation for Model 5:

$$\begin{aligned} Y_{it} &= a_i + \zeta Y_{it-1} + \zeta_{IL} Y_{IL,t-1} D_{it} (i \in \text{IF}) \\ &\quad + \zeta_{IF} \left(\sum_{k \in \text{IF}} Y_{k,t-1} \right) D_{it} (i = \text{IL}) + \Omega e_{it} \end{aligned} \quad (\text{B.6})$$

where Y_{it} is a 4×1 vector of the endogenous variables; a_i is a 4×1 vector of fixed effects; Ω is a lower triangular matrix with positive elements in its diagonal; $e_{it} \sim N(0_{4 \times 1}, I_4)$ is the vector of error terms. We write Eq. (B.6) as

$$\begin{aligned} Y_{it} &= (I_4 - \zeta L)^{-1}a_i + (I_4 - \zeta L)^{-1}\zeta_{IL}Y_{IL,t-1}D(i \in \text{IF}) \\ &\quad + (I_4 - \zeta L)^{-1} \left(\sum_{k \in \text{IF}} \zeta_{IF} Y_{k,t-1} \right) D(i = \text{IL}) + (I_4 - \zeta L)^{-1}\Omega e_{it} \end{aligned} \quad (\text{B.7})$$

$$\begin{aligned} Y_{it} &= \sum_{j=0}^{\infty} \zeta^j a_i + \left(\sum_{j=0}^{\infty} \zeta^j \zeta_{IL} Y_{IL,t-1-j} \right) D(i \in \text{IF}) \\ &\quad + \left(\sum_{j=0}^{\infty} \sum_{k \in \text{IF}} \zeta^j \zeta_{IF} Y_{k,t-1-j} \right) D(i = \text{IL}) + \sum_{j=0}^{\infty} \zeta^j \Omega e_{it-j}. \end{aligned} \quad (\text{B.8})$$

Then, we get the IRF matrices

$$\Theta_j = \zeta^j \Omega \quad \text{for } j = 0, 1, 2, \dots, \infty \quad (\text{B.9})$$

and the dynamic interaction multiplier matrices

$$\begin{aligned} \Gamma_j(\text{IL} \rightarrow \text{IF}) &= (\text{effects of } Y_{IL,t-j} \text{ on } Y_{it} \text{ for } i \in \text{IF}) \\ &= \zeta^j \zeta_{IL} \quad \text{for } j = 0, 1, 2, \dots, \infty \end{aligned} \quad (\text{B.10})$$

$$\begin{aligned} \Gamma_j(\text{IF} \rightarrow \text{IL}) &= (\text{effects of } Y_{k,t-j} \text{ on } Y_{IL,t} \text{ for } k \in \text{IF}) \\ &= \zeta^j \zeta_{IF} \quad \text{for } j = 0, 1, 2, \dots, \infty. \end{aligned} \quad (\text{B.11})$$

Appendix C

The Poisson-type patent count panel data model that involves the common competitive latent factor is estimated by the SML method (Gouriéroux and Monfort, 1991). The likelihood function is evaluated by using the EIS technique of Richard and Zhang (2007). The conditional density of l_t^* is given by

$$f^*(l_t^* | l_{t-1}^*) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{(l_t^* - \mu^* l_{t-1}^*)^2}{2} \right]. \quad (\text{C.1})$$

The likelihood of a realization $(P, L^*) \equiv (P_{it}, l_{it}^* : t = 1, \dots, T; i = 1, \dots, N)$ is

$$\prod_{i=1}^N \prod_{t=1}^T f(P_{it} | \mathcal{F}_t) f^*(l_{it}^* | l_{t-1}^*) = \prod_{i=1}^N \prod_{t=1}^T \left\{ \frac{\exp(-\lambda_{it}) (\lambda_{it})^{P_{it}}}{P_{it}!} \times \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{(l_{it}^* - \mu^* l_{t-1}^*)^2}{2} \right] \right\}. \quad (\text{C.2})$$

The likelihood of the patent counts observed is obtained by

$$\begin{aligned} \mathcal{L}(P | \mathcal{F}_T^0; \theta) &= \int_{\mathbb{R}^T} \prod_{i=1}^N \prod_{t=1}^T \frac{\exp(-\lambda_{it}) \lambda_{it}^{P_{it}}}{P_{it}!} \\ &\times \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{(l_{it}^* - \mu^* l_{t-1}^*)^2}{2} \right] dL^* \end{aligned} \quad (\text{C.3})$$

where θ summarizes all parameters. The same likelihood function with a compact notation is

$$\mathcal{L}(P | \mathcal{F}_T^0; \theta) = \int_{\mathbb{R}^T} \prod_{i=1}^N \prod_{t=1}^T g(P_{it}, l_{it}^* | \mathcal{F}_t; \theta) dL^* \quad (\text{C.4})$$

where g is the joint density of (P_{it}, l_{it}^*) . We introduce the auxiliary sampler m and include it in Eq. (C.4):

$$\begin{aligned} \mathcal{L}(P | \mathcal{F}_T^0; \theta, \theta^*) &= \int_{\mathbb{R}^T} \prod_{i=1}^N \prod_{t=1}^T \frac{g(P_{it}, l_{it}^* | \mathcal{F}_t; \theta)}{m(l_{it}^* | l_{t-1}^*; \theta_t^*)} \times m(l_{it}^* | l_{t-1}^*; \theta_t^*) dL^* \end{aligned} \quad (\text{C.5})$$

where $\theta^* = (\theta_1^*, \dots, \theta_T^*)$ denotes the parameters of the auxiliary sampler. Parameters of the auxiliary sampler are different in each period. However, the functional form of this sampler is constant over time. The *importance Monte Carlo estimate* of $\mathcal{L}(P | \mathcal{F}_T^0; \theta, \theta^*)$ for a given θ^* is

$$\hat{\mathcal{L}}_R(P | \mathcal{F}_T^0; \theta, \theta^*) = \frac{1}{R} \sum_{r=1}^R \prod_{i=1}^N \prod_{t=1}^T \frac{g(P_{it}, l_{it}^* | \mathcal{F}_t; \theta)}{m(l_{it}^* | l_{t-1}^*; \theta_t^*)} \quad (\text{C.6})$$

where $\{l_{tr}^* : t = 1, \dots, T\}$ denotes the r th trajectory of i.i.d. draws from m . Richard and Zhang (2007) suggest defining the auxiliary sampler m with its density kernel k :

$$k(l_t^* | l_{t-1}^*; \theta_t^*) = m(l_t^* | l_{t-1}^*; \theta_t^*) \chi(l_{t-1}^*; \theta_t^*) \quad (\text{C.7})$$

where

$$\chi(l_{t-1}^*; \theta_t^*) = \int_{\mathbb{R}} k(l_t^* | l_{t-1}^*; \theta_t^*) dl_t^* \quad (\text{C.8})$$

denotes the integrating constant associated to k . Richard and Zhang (2007) suggest choosing k as a kernel of the normal distribution. Moreover, similar to Bauwens and Hautsch (2006), we include f^* into the auxiliary sampler, m . Then, the normal density kernel is given by

$$\begin{aligned} k(l_t^* | l_{t-1}^*; \theta_t^*) &= \exp \left[\theta_{1t}^* l_t^* + \theta_{2t}^* (l_t^*)^2 \right] \\ &\times \exp \left[-\frac{(l_t^* - \mu^* l_{t-1}^*)^2}{2} \right] \end{aligned} \quad (\text{C.9})$$

where $\theta_t^* = (\theta_{1t}^*, \theta_{2t}^*)$ determines the conditional mean and variance of the auxiliary sampler for period t . The conditional mean μ_t and variance π_t^2 of the normal auxiliary sampler m are

$$\mu_t = \pi_t^2 (\theta_{1t}^* + \mu^* l_{t-1}^*) \quad (\text{C.10})$$

$$\pi_t^2 = \frac{1}{1 - 2\theta_{2t}^*}. \quad (\text{C.11})$$

A trajectory of $\{l_t^* : t = 1, \dots, T\}$ can be generated from the auxiliary sampler, as follows:

$$l_t^* = \mu_t + \pi_t \eta_t \quad (\text{C.12})$$

where $\eta_t \sim N(0, 1)$ are i.i.d. common random numbers. Richard and Zhang (2007) suggest using the same set of random numbers for every iteration of the maximum likelihood procedure (i.e., common random numbers). In the EIS method, parameters of the auxiliary sampler minimize the variance of the Monte Carlo estimator of the likelihood function:

$$\begin{aligned} \theta^* &= \arg \min_{\theta^*} \text{Var} \left[\hat{\mathcal{L}}_R(P | \mathcal{F}_T^0; \theta, \theta^*) \right] \\ &= \arg \min_{\theta^*} \text{Var} \left[\frac{1}{R} \sum_{r=1}^R \prod_{i=1}^N \prod_{t=1}^T \frac{g(P_{it}, l_{it}^* | \mathcal{F}_t; \theta)}{m(l_{it}^* | l_{t-1}^*; \theta_t^*)} \right]. \end{aligned} \quad (\text{C.13})$$

This variance is minimized by choosing such values for θ_t^* for which there is a good fit between g and m for every period t (see the last equality of Eq. (C.13)). To achieve this, Richard and Zhang (2007) suggest solving the minimization problem of Eq. (C.13) by estimating a recursive sequence of Ordinary Least Squares (OLS) problems, each being of the following form:

$$\begin{aligned} \ln g(P_{it}, l_{it}^* | \mathcal{F}_t; \theta) + \ln \chi(l_{t-1}^*; \hat{\theta}_{t+1}^*) \\ = \theta_{0t}^* + \theta_{1t}^* l_{it}^* + \theta_{2t}^* (l_{it}^*)^2 + u_{tr} \quad \text{with } r = 1, \dots, R \end{aligned} \quad (\text{C.14})$$

for $t = T, \dots, 1$, $\chi(l_{t-1}^*; \hat{\theta}_{t+1}^*) = 1$ and $\hat{\theta}_{t+1}^*$ is the OLS estimate of θ_{t+1}^* . These regressions are run backwards, from T to 1 and the sample size of each regression is equal to the number of trajectories drawn R . In our estimation we choose $R = 50$.

The right side of Eq. (C.14) includes the log kernel of the auxiliary sampler. Normal distribution is used for the auxiliary sampler since the log kernel of the normal distribution is a second order polynomial therefore its parameters can be estimated by OLS. The EIS technique involves the estimation of a large number of parameters of the auxiliary sampler in the SML procedure. The OLS method provides auxiliary sampler parameter estimates rapidly thus EIS based SML estimation is feasible in practice. We can summarize the EIS method as follows:

- Step 1: Draw $R = 50$ trajectories $\{l_{tr}^*\}_{t=1}^T$ from the distribution $N(\mu^* l_{t-1}^*, 1)$.
- Step 2: For each $t = T, \dots, 1$, estimate by OLS Eq. (C.14) to get the parameters of m .
- Step 3: Draw $R = 50$ trajectories $\{l_{tr}^*\}_{t=1}^T$ from the auxiliary samplers.
- Repeat Steps 2 and 3 five times.
- Step 4: Estimate the value of the likelihood function according to Eq. (C.6).

Appendix D

In order to approximate the value of l_t^* , we compute its filtered estimates $E[l_t^* | \mathcal{F}_t^o]$, where we condition on the observable information set

$$\mathcal{F}_t^o = [(P_{i1}, r_{i1}, BV_{i1}, BC_{intra,i1}, BC_{inter,i1}), \dots, (P_{it-1}, r_{it-1}, BV_{it-1}, BC_{intra,it-1}, BC_{inter,it-1}), (r_{it}, BV_{it}, BC_{intra,it}, BC_{inter,it}) : i = 1, \dots, N]. \quad (D.1)$$

The conditional expectation of l_t^* is

$$E[l_t^* | \mathcal{F}_t^o] = \int_{\mathbb{R}} l_t^* h(l_t^* | \mathcal{F}_t^o) dl_t^* \quad (D.2)$$

where h denotes the conditional density of l_t^* . We introduce $P_t = (P_{is} : i = 1, \dots, N; s = 1, \dots, t)$ and $L_t^* = (l_s^* : s = 1, \dots, t)$. We compute h as follows:

$$\begin{aligned} h(l_t^* | \mathcal{F}_t^o) &= \frac{\dot{g}(P_{t-1}, l_t^* | \mathcal{F}_t^o)}{f(P_{t-1} | \mathcal{F}_t^o)} \\ &= \frac{\int_{\mathbb{R}^{t-1}} f^*(l_t^* | l_{t-1}^*) \ddot{g}(P_{t-1}, L_{t-1}^* | \mathcal{F}_t) dL_{t-1}^*}{\int_{\mathbb{R}^{t-1}} \ddot{g}(P_{t-1}, L_{t-1}^* | \mathcal{F}_t) dL_{t-1}^*} \end{aligned} \quad (D.3)$$

where \dot{g} , f , \ddot{g} , and f^* are conditional density functions of the corresponding random variables or random vectors. By substituting Eq. (D.3) into Eq. (D.2) and using the fact that the denominator in Eq. (D.3) is not a function of l_t^* , we obtain

$$E[l_t^* | \mathcal{F}_t^o] = \frac{\int_{\mathbb{R}^t} l_t^* f^*(l_t^* | l_{t-1}^*) \ddot{g}(P_{t-1}, L_{t-1}^* | \mathcal{F}_t) dL_t^*}{\int_{\mathbb{R}^{t-1}} \ddot{g}(P_{t-1}, L_{t-1}^* | \mathcal{F}_t) dL_{t-1}^*} \quad (D.4)$$

where the joint density \ddot{g} is given by

$$\begin{aligned} \ddot{g}(P_{t-1}, L_{t-1}^* | \mathcal{F}_t) &= \prod_{i=1}^N \prod_{s=1}^{t-1} \frac{\exp(-\lambda_{is}) \lambda_{is}^{P_{is}}}{P_{is}!} \\ &\times \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(l_s^* - \mu^* l_{s-1}^*)^2}{2}\right]. \end{aligned} \quad (D.5)$$

The high-dimensional integrals in Eq. (D.4) are estimated by the EIS technique.

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R., Howitt, P., 2005. Competition and innovation: An inverted-U relationship. *Q. J. Econ.* 120, 701–728.

Arora, A., Ceccagnoli, M., Cohen, W.M., 2008. R&D and the patent premium. *Int. J. Ind. Org.* 26, 1153–1179.

Bauwens, L., Hautsch, N., 2006. Stochastic conditional intensity processes. *J. Financ. Econ.* 4, 450–493.

Binder, M., Hsiao, C., Pesaran, M.H., 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795–837.

Blazsek, S., Escribano, A., 2010. Knowledge spillovers in U.S. patents: A dynamic patent intensity model with secret common innovation factors. *J. Econometrics* 159, 14–32.

Blazsek, S., Escribano, A., 2014. Propensity to Patent, R&D and Market Competition: Dynamic Spillovers of Innovation Leaders and Followers. Working Paper 14–12, Economic Series, Universidad Carlos III de Madrid.

Blundell, R., Griffith, R., Van Reenen, J., 1999. Market share, market value and innovation in a panel of British manufacturing firms. *Rev. Econom. Stud.* 66, 529–554.

Boldrin, M., Levine, D.K., 2008. Against Intellectual Monopoly. Cambridge University Press, Cambridge, UK.

Chambers, D., Jennings, R., Thompson, R.B., 2002. Excess returns to R&D-intensive firms. *Rev. Account. Stud.* 7, 133–158.

Chan, L.K.C., Lakonishok, J., Sougiannis, T., 2001. The stock market valuation of research and development expenditures. *J. Finance* 56, 2431–2457.

Ciftci, M., Lev, B., Radhakrishnan, S., 2011. Is research and development mispriced or properly risk-adjusted? *J. Account. Audit. Finance* 26, 81–116.

Davidson, R., MacKinnon, J.G., 2009. *Econometric Theory and Methods*. Oxford University Press, New York.

Doornik, J.A., Hansen, H., 2008. An omnibus test for univariate and multivariate normality. *Oxf. Bull. Econ. Stat.* 70, 927–939.

Escribano, A., Fosfuri, A., Tribó, J.A., 2009. Managing external knowledge flows: The moderating role of absorptive capacity. *Res. Policy* 38, 96–105.

Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. *J. Finance* 25, 383–417.

Gouriéroux, C., Monfort, A., 1991. Simulation based inference in models with heterogeneity. *Ann. Econ. Stat.* 20–21, 69–107.

Grabowski, H.G., Vernon, J.M., 2000. Effective patent life in pharmaceuticals. *Int. J. Technol. Manage.* 19, 98–120.

Griliches, Z., 1981. Market value, R&D, and patents. *Econom. Lett.* 7, 183–187.

Hall, B.H., 1993. The stock market's valuation of research and development investment during the 1980s. *Amer. Econ. Rev.* 83, 259–264.

Hall, B.H., Vopel, K., 1997. Innovation, Market Share and Market Value. Working Paper, University of California at Berkeley and University of Mannheim. www.econ.berkeley.edu/~bhall/papers/HallVopel97.pdf.

Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2001. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper No. 8498.

Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2005. Market value and patent citations. *Rand J. Econ.* 36, 16–38.

Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Hausman, J., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* 52, 909–938.

Jaffe, A.B., 1986. Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. *Amer. Econ. Rev.* 76, 984–1001.

Jain, N.C., Indrayan, A., Goel, L.R., 1986. Monte Carlo comparison of six hierarchical clustering methods on random data. *Pattern Recognit.* 19, 95–99.

Jovanovic, B., MacDonald, G.M., 1994. Competitive diffusion. *J. Polit. Econ.* 102, 24–52.

Kahn, A.E., 1962. The role of patents. In: Miller, J.P. (Ed.), *Competition, Cartels and Their Regulation*. North-Holland, Amsterdam.

Keane, M.P., 1994. A computationally practical simulation estimator for panel data. *Econometrica* 62, 95–116.

Kuiper, F.K., Fisher, L., 1975. A Monte Carlo comparison of six clustering procedures. *Biometrics* 31, 777–783.

Lanjouw, J.O., Schankerman, M., 1999. The Quality of Ideas: Measuring Innovation with Multiple Indicators. NBER Working Paper No. 7345.

LeRoy, S.F., Porter, R.D., 1981. The present-value relation: Tests based on implied variance bounds. *Econometrica* 49, 555–574.

Lev, B., Radhakrishnan, S., Ciftci, M., 2006. The Stock Market Valuation of R&D Leaders. NYU Working Paper No. 2451/27475.

Lev, B., Sarath, B., Sougiannis, T., 2005. R&D-related reporting biases and their consequences. *Contemp. Account. Res.* 22, 977–1026.

Lev, B., Sougiannis, T., 1996. The capitalization, amortization and value-relevance of R&D. *J. Account. Econ.* 21, 107–138.

Liesenfeld, R., Richard, J.-F., 2003. Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *J. Empir. Finance* 10, 505–531.

Ljung, G., Box, G., 1978. On a measure of lack of fit in time-series models. *Biometrika* 65, 297–303.

Machlup, F., 1962. An economic review of the patent system. In: Miller, J.P. (Ed.), *Competition, Cartels and Their Regulation*. North-Holland, Amsterdam.

Mansfield, E., Schwartz, M., Wagner, S., 1981. Imitation costs and patents: An empirical study. *Econ. J.* 91, 907–918.

Nabseth, L., Ray, G.F., 1974. The Diffusion of New Industrial Processes: An International Study. Cambridge University Press, Cambridge, UK.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Nickell, S., 1996. Competition and corporate performance. *J. Polit. Econ.* 4, 724–746.

Pakes, A., 1985. On patents, R&D, and the stock market rate of return. *J. Polit. Econ.* 93, 390–409.

Pakes, A., Schankerman, M., 1984. The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. In: Griliches, Z. (Ed.), *R&D, Patents and Productivity*. University of Chicago Press, Chicago.

Richard, J.-F., Zhang, W., 2007. Efficient high-dimensional importance sampling. *J. Econometrics* 141, 1385–1411.

Rogers, E.M., 1983. *Diffusion of Innovations*. Free Press, New York.

Scherer, F.M., 1965a. Corporate inventive output, profits, and growth. *J. Polit. Econ.* 73, 290–297.

Scherer, F.M., 1965b. Firm size, market structure, opportunity, and the output of patented inventions. *Amer. Econ. Rev.* 55, 1097–1125.

Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 303–333.

Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.* 58, 236–244.