



Università degli Studi di Ferrara

DIPARTIMENTO DI ECONOMIA, ISTITUZIONI, TERRITORIO

Corso Ercole I D'Este n.44, 44100 Ferrara

Quaderni del Dipartimento

n.14/2004

Settembre 2004

**Ricerca industriale e pubblicazioni dello scienziato accademico:
complementi o sostituti ?**

Un modello di scelta razionale – Problemi ed esempi di verifica empirica.

di *Roberto Iorio*

Settembre 2004

**Ricerca industriale e pubblicazioni dello scienziato accademico:
complementi o sostituti ?**

Un modello di scelta razionale – Problemi ed esempi di verifica empirica.

di Roberto Iorio*

Abstract

Il presente lavoro si inserisce nel dibattito che negli ultimi anni si è fatto particolarmente vivo sull'efficacia e i possibili effetti negativi di una crescente collaborazione tra università e industria. Un tema particolarmente dibattuto è se un'intensa attività di collaborazione col mondo industriale abbia dei riflessi positivi o negativi sull'attività di pubblicazione del docente, riducendone la quantità o la qualità o spingendolo a dilazionare la pubblicazione, per ragioni di tutela della segretezza, imposta da accordi intercorsi tra scienziato e impresa, o imposta dalla normativa sui brevetti. Il presente articolo sintetizza la letteratura recente sul rapporto tra università ed industria e quindi, più specificamente, la relazione tra brevetti e pubblicazioni; propone quindi un semplice modello per l'analisi del comportamento dello scienziato accademico e cerca di razionalizzare le relazioni esistenti tra input e output della ricerca, che possa rendere conto delle indagini empiriche fin qui compiute e fare da base per stime ulteriori. Si analizza quindi un campione di medie dimensioni con dati *cross section* e viene quindi proposto un esempio (analizzando un piccolo campione di soggetti) di come si possa procedere, con dati *panel*, alla valutazione empirica del tipo di rapporto esistente tra pubblicazioni e brevetti a livello individuale, nell'attesa di disporre di dati più ampi che possano condurre a risultati maggiormente probanti.

Desidero ringraziare Francesco Lissoni, Stefano Breschi, Margherita Balconi e Fabio Montobbio per aver consentito l'accesso ai dati del database EPO-INV-DOC e ai dati sulle pubblicazioni dei docenti raccolti da loro e dai loro collaboratori. A Fabio Montobbio va un ringraziamento anche per i commenti preziosi effettuati in occasione della presentazione di questo articolo durante il I Seminario Dottorandi SIEPI (Società Italiana di Economia e Politica Industriale), tenutosi a Ferrara il 17 e 18 giugno 2004.

Ringrazio inoltre Lucio Poma e Patrizio Bianchi per la fiducia e il sostegno intellettuale per le mie ricerche di dottorato, Francesco Galassi ed Maria Elena Bontempi per i suggerimenti teorici ed econometrici.

*Dottorando in Economia presso l'Università di Ferrara, Dipartimento di Economia, Istituzioni, Territorio.
E-mail: iorio@economia.unife.it

1. Introduzione

Le relazioni tra attività accademica ed attività industriale dei professori universitari, i fattori predisponenti o deterrenti all'accostamento degli scienziati accademici alla realtà industriale e i possibili effetti della ricerca industriale sulle attività accademiche sono temi esplorati negli ultimi anni da un'ampia letteratura. L'incremento di ricerche in questo campo è stato stimolato dal crescente rapporto tra università ed industria, soprattutto nel settore della biotecnologia e soprattutto negli Stati Uniti, mentre in Europa ha costituito a lungo tema di discussione il cosiddetto "paradosso europeo", secondo il quale l'Europa sarebbe a pari livello degli Stati Uniti dal punto di vista della capacità scientifica, ma non riuscirebbe a tradurre la sua capacità in prodotti commercialmente altrettanto validi di quelli d'oltreoceano; di qui la teorizzazione della necessità di un più stretto contatto tra accademia e mondo produttivo.¹

Questo tema presenta spunti di grande interesse: studiare le collaborazioni tra università e industria comporta l'analisi delle differenti norme di comportamento nei due ambiti, il modo in cui queste si vanno evolvendo ed i possibili rischi dell'avvicinamento di un'istituzione, l'università, volta alla diffusione della conoscenza con un'altra, l'industria, volta invece all' "appropriazione" della conoscenza. Il problema è dunque non solo di natura sociologica, in quanto coinvolge regole di comportamento, norme e valori condivisi, ma anche l'analisi economica ha qualcosa da dire nel momento in cui studia gli incentivi dei singoli e valuta il problema della natura del bene-scienza e del bene-tecnologia alla luce delle categorie concettuali dei beni pubblici, quasi-pubblici e privati (si veda in questo senso il filone di ricerca della "New Economics of Science" inaugurato da Dasgupta-David, 1994, che cerca di "inglobare" le categorie sociologiche mertoniane nell'analisi economica, oppure gli studi sulle reti di innovatori, accademiche e industriali, di Balconi-Breschi-Lissoni, 2002).

Tuttavia, nonostante le ampie prospettive di ricerca che questo tema permette, la vasta letteratura a cui si accennava sopra è in buona parte di tipo empirico, basata sovente su interviste dirette agli interessati², mentre solo di recente stanno emergendo studi che tentino di mettere in relazione analisi empiriche sul tema con un quadro teorico rigoroso. Soprattutto scarseggiano studi che riferiscano le analisi empiriche del comportamento degli scienziati di fronte alla crescente realtà delle collaborazioni con l'industria ad un'analisi concettuale degli incentivi e delle motivazioni degli stessi.

Un tentativo interessante di analizzare il comportamento dello scienziato secondo i criteri della razionalità economica, nell'accezione schumpeteriana (scienziato imprenditore come imprenditore schumpeteriano) è stato compiuto da Bonaccorsi (2000). In realtà egli stesso avverte che l'incertezza che caratterizza l'attività dello scienziato rende difficile, se non impossibile, "modellizzarne" il comportamento, in quanto essa può essere così elevata da non potere essere neppure trattata in un modello probabilistico: vi sono ambiti di ricerca in cui non è possibile assegnare una probabilità all'eventualità di una scoperta e dunque non è possibile effettuare un calcolo razionale sull'opportunità di intraprendere un nuovo filone di ricerca. Tuttavia l'attività di collaborazione industriale presenta senz'altro caratteristiche di "scienza normale", essendo rivolta alla soluzione di problemi tecnologici e non ad affrontare "massimi sistemi" scientifici. In questo senso il problema evidenziato da Bonaccorsi, l'incertezza, sembra riguardarla solo in misura

¹ Per un'esposizione semplice ma esauriente dei reciproci effetti della collaborazione tra università e industria si veda Cohen, Florida, Randazzese e Walsh (1998). Per una parziale rassegna di temi di *policy* e di indagini empiriche su questa tematica si veda il paragrafo sulle relazioni tra università e industria in Bianchi, Iorio, Labory e Malagoli (2002). Si veda invece Bonaccorsi (2000) per una posizione critica nei confronti del "paradosso europeo": secondo questo autore non solo l'Europa è dietro agli Stati Uniti dal punto di vista dell'innovazione tecnologica, ma anche l'affermazione che l'Europa sia vicina agli USA dal punto di vista della ricerca scientifica è smentita da analisi bibliometriche.

² Si vedano ad esempio Meyer, Kraemer e Schmoch (1998) e Schibany e Schartinger (2001).

limitata: i problemi affrontati nell'ambito della ricerca industriale non solo tali da impedire l'assegnazione di una qualunque probabilità all'esito di una ricerca.

Semmai i limiti di un'indagine basata su ipotesi di razionalità sono di altra natura, squisitamente metodologica.

Un'analisi del comportamento dei soggetti fondata unicamente su un'ipotesi di comportamento razionale, basato sulla massimizzazione dell'utilità, è andata incontro ad una tale mole di critiche, sul piano non solo teorico ma anche empirico (si pensi solo all'opera di Simon), che non è più possibile ignorarle e non si può impostare un'analisi sui ben noti presupposti neoclassici senza avvertire che si tratta di una semplificazione, persino grossolana, che ha la sua unica e tuttavia persistente utilità nella semplicità concettuale e nella "trattabilità" dei risultati. Il peso di questi vantaggi ha prevalso rispetto ai limiti, a tal punto che l'analisi dei comportamenti degli individui visti come soggetti massimizzanti, anziché andare incontro ad una serie di ragionevoli limitazioni ed approfondimenti in direzione di una più realistica e organica visione del comportamento umano, si espande invece a considerare ambiti sempre nuovi, che vengono quindi "conquistati" all'analisi neoclassica.

La "modellizzazione" dei fenomeni economici rischia dunque di semplificare i meccanismi di scelta, a meno di non considerare esplicitamente tra le motivazioni degli agenti i fattori istituzionali e sociali, come cercano appunto di fare Dasgupta e David, sottolineando i differenti incentivi che muovono scienziati accademici e ricercatori che lavorano nel mondo industriale, istituendo quindi una chiara dicotomia tra i due tipi di comportamento.

In questo modo l'analisi "razionale" diventa maggiormente plausibile, perchè consente di delimitare l'oggetto di osservazione e di isolare alcune relazioni causali, con la possibilità di sottoporle a verifica empirica col rigore della metodologia statistica.

E' ciò che si tenterà di fare in questo articolo rispetto al problema della relazione tra pubblicazioni e ricerca industriale. Si cercherà, cioè, di proporre un contesto teorico entro cui verificare se è fondato il timore che un crescente impegno dei docenti universitari nell'attività di ricerca industriale possa ridurre la quantità e la qualità delle pubblicazioni dagli stessi effettuate.

D'altro canto il problema relativo alla scelta dello scienziato accademico è solo una parte, e forse nemmeno la più rilevante, nell'analisi della relazione tra pubblicazioni e brevetti. Il problema teorico fondamentale risiede non tanto nel determinare quali siano le scelte degli scienziati, quanto l'impatto che tali scelte possono avere sugli output della ricerca.

Poiché questa è ancora in buona parte una *black box*, alla fine il problema è di natura essenzialmente empirica.

Questo tipo di analisi si rifà dunque al filone di indagini, che è appunto per lo più di natura empirica, sulle possibili conseguenze negative per l'università della più frequente attività di collaborazione con l'industria.

Balconi, Breschi e Lissoni (2003) nella loro rassegna di indagini su tali *unintended consequences* concludono comunque che

l'esperienza dimostra che sacrifici marginali dei valori universitari (universalismo, disinteresse, comunitarismo e libertà accademica, secondo la tradizione mertoniana) che risultano dalle relazioni con l'industria sono un prezzo accettabile da pagare per conseguire gli aumenti della conoscenza e della possibilità di formazione di giovani ricercatori che risultano dall'infusione di fondi industriali nei laboratori in cui lavorano docenti di talento (cit., pag. 7).

Sul tema in particolare delle relazioni tra pubblicazioni e ricerca industriale Zucker, Darby e Armstrong (1998) riferiscono che gli scienziati accademici "di vertice" con legami con l'industria pubblicano ad un tasso più alto (prima, durante e dopo i rapporti di collaborazione) di coloro che non hanno mai avuto questi legami. Ma quest'indagine riguarda appunto solo gli scienziati di vertice ed è relativa ad una realtà come quella americana con molte caratteristiche che la differenziano da quella italiana, su cui si tenterà di condurre un'indagine empirica.

Su questo argomento l'attenzione degli studiosi va particolarmente concentrandosi in tempi recenti. Un recente studio empirico di Stephan, Gurmu, Sumell e Black (2002) mette in evidenza una relazione positiva tra pubblicazioni e brevetti. Si tratta però di uno studio condotto su dati *cross-section*, la qual cosa, come si vedrà meglio in seguito, può portare ad una significativa distorsione nell'interpretazione dei risultati: un coefficiente di regressione positivo può significare semplicemente che chi pubblica di più brevetta di più *ceteris paribus*, ma ciò può essere dovuto ad una più alta domanda industriale verso i docenti più eminenti, mentre non esclude un effetto negativo dell'attività di ricerca industriale sulle pubblicazioni dei singoli docenti.

Infatti uno studio condotto su dati *panel* da Agrawal e Henderson (2002) conduce ad una relazione incerta tra pubblicazioni e brevetti.

Un effetto che appare invece pressoché certo è, come riportano Geuna e Nesta (2003) nella loro *survey* sull'argomento, è che il brevetto ritardi la pubblicazione. Ciò è dovuto alla legislazione stessa sui brevetti, per cui non è possibile brevettare se il risultato è stato reso in qualunque modo pubblico e dunque anche tramite una pubblicazione.

Tuttavia nell'analisi del loro ampio *panel* Calderini e Franzoni (2004) riscontrano che il numero di pubblicazioni, nonché la qualità degli articoli, misurata tramite le citazioni, aumenta in media nell'anno in cui si brevetta ed aumenta anche nell'anno successivo al brevetto. Se si considera che le norme sui brevetti impongono di non aver effettuato pubblicazioni prima del brevetto, mentre naturalmente è possibile pubblicare *dopo*, la relazione positiva tra pubblicazioni e brevetti nello stesso momento appare sorprendente, mentre era prevedibile il secondo risultato. In ogni caso le conclusioni dei due autori sembrano un tantino affrettate, nel momento in cui deducono dalle relazioni da loro stimate una sicura presenza di complementarità tra ricerca di base e applicata ed un rovesciamento della prospettiva della "New economics of science", secondo la quale la scelta tra ricerca accademica e ricerca industriale si muove inevitabilmente su un piano di alternativa e significa scegliere tra diffondere i risultati della ricerca oppure secretarli e/o appropriarsene tramite il brevetto.

Breschi, Lissoni e Montobbio (2004) prendono in considerazione sia il problema della riduzione *tout court* delle pubblicazioni (visto come indicatore di una sostituzione della ricerca di base con la ricerca applicata), sia quello del rinvio delle pubblicazioni e cercano di considerare i diversi effetti che può provocare un aumento dell'attività di ricerca applicata, come il reperimento di fondi aggiuntivi e l'effetto di segnalazione per il singolo docente che risulti inventore o titolare di un brevetto. I loro risultati sono meno ottimistici degli studi precedentemente citati, non potendo escludere una riduzione delle pubblicazioni a seguito del conseguimento di un brevetto.

Altri studi considerano l'effetto dell'attività di ricerca brevettabile sul complesso delle attività accademiche, come Jensen e Thursby (2003), sottolineando con il loro modello il possibile rischio che la ricerca applicata vada a discapito dell'attività accademica, senza tuttavia sottoporre il tema ad una verifica empirica.

Concettualmente, il rapporto tra pubblicazioni e ricerca industriale si può configurare nei termini seguenti.

Si può ipotizzare che sia le pubblicazioni, sia la ricerca industriale siano desiderate dal professore universitario, in quanto le prime producono prestigio accademico e possibili avanzamenti di carriera, la seconda remunerazione diretta; pubblicazioni e remunerazioni da ricerca industriale sono quindi due argomenti della funzione di utilità del docente.

Il vincolo del docente è rappresentato dal tempo, ovviamente limitato. Egli si trova dunque a dover allocare il suo tempo tra ricerca accademica e ricerca industriale. La riduzione della prima a vantaggio della seconda può comportare una riduzione delle pubblicazioni, in virtù del contenuto presumibilmente più applicato della ricerca industriale e soprattutto dei vincoli alle pubblicazioni sovente imposti dall'industria. In questo senso pubblicazioni e ricerca industriale sono due "beni sostituti".

D'altro canto, però, sono facilmente prevedibili effetti di complementarità tra i due "beni", nel senso che i risultati della ricerca industriale possono fornire spunti per la pubblicazione di articoli: non è raro il fenomeno della pubblicazione congiunta tra ricercatori universitari e industriali, come pure la ricerca industriale può fornire spunti importanti per la ricerca accademica. Inoltre la collaborazione con l'industria può aumentare la produttività accademica dello scienziato grazie al supporto finanziario alla ricerca (Balconi, Breschi e Lissoni, 2003)

Ma se questi sono gli effetti possibili della ricerca industriale sulle pubblicazioni, è altresì possibile prevedere una relazione causale inversa tra i due "beni": le imprese naturalmente desiderano collaborare con i docenti più capaci (Balconi, Breschi e Lissoni, 2003); ebbene, proprio le pubblicazioni rappresentano un fondamentale indicatore della qualità del docente accademico; esse svolgono dunque un *effetto di segnalazione*.

Le relazioni tra attività di collaborazione industriale ed attività di pubblicazione sono dunque complesse e, secondo il modello presentato, non è possibile prevedere in sede teorica quale effetto prevalga. E' chiara, però, l'importanza pratica della questione e dunque la necessità di analizzarla empiricamente: un'eventuale *trade-off* tra pubblicazioni e ricerca industriale potrebbe mettere in discussione tutte le politiche e tutti gli auspici volte ad incentivare un rapporto più stretto tra università e industria: se davvero ci fosse, non è detto che questo *trade-off* sarebbe "un prezzo accettabile da pagare".

E' possibile formalizzare questo quadro concettuale è quindi derivarne delle relazioni empiriche stimabili. Naturalmente non è possibile ottenere dei dati relativi al tempo dedicato alla ricerca industriale. I dati "oggettivi" disponibili sono quelli relativi ai brevetti. E' necessario però chiarire le semplificazioni che si sono implicitamente effettuate. In primo luogo, se è chiaro che un brevetto non deriva da ricerca fondamentale, per il requisito richiesto di applicabilità, tuttavia non tutta la ricerca applicata si traduce in brevetti, sia perché l'esito della ricerca è sempre di natura probabilistica, sia perché vi sono altri metodi di appropriazione dei risultati oltre alla brevettazione.

Inoltre fin qui e poi in seguito si è usato il termine di ricerca industriale come sinonimo di ricerca applicata: anche qui, se è improbabile che un'impresa collabori con un docente per effettuare della ricerca di base, è tuttavia possibile che lo scienziato conduca ricerca applicata non in collaborazione con l'industria. D'altronde i confini tra ricerca di base e ricerca applicata si fanno sempre più sfumati.

Queste ragioni sono alla base della problematicità dell'operazione che di solito, più o meno esplicitamente, si compie, che è quella di utilizzare i brevetti come *proxy* della ricerca applicata condotta, vedendo quindi nell'eventuale relazione positiva tra brevetti e pubblicazioni una complementarità tra ricerca di base e applicata.

La struttura dell'articolo è la seguente.

Nel par. 2 si espone in un modello estremamente semplificato il meccanismo di scelta dello scienziato accademico e il modo in cui egli alloca il tempo tra ricerca accademica e ricerca industriale.

Il par. 3 illustra invece il rapporto tra pubblicazioni e brevetti, cercando di individuare alcuni dei complessi meccanismi che possono legare queste due realtà. In particolar modo l'indagine è volta ad individuare le cause possibili di un coefficiente positivo nella relazione tra brevetti e pubblicazioni, che viene riscontrato in molte indagini empiriche, sottolineando che esso può risiedere anche in ragioni diverse da una complementarità tra forme diverse di ricerca.

Il par. 4 espone i risultati di due indagini empiriche: il primo condotto su un campione di medie dimensioni (180 soggetti) con dati *cross section*, il secondo condotto su un piccolo campione di dati longitudinali (24 soggetti per 2 periodi, per un totale di 48 osservazioni), evidenziando il confronto tra stime *cross section* e stime *panel*. Date le dimensioni molto ridotte di questo

campione, tali stime hanno valore esemplificativo, piuttosto che poter essere oggetto di vera inferenza statistica.

Il par.5 presenta una sintesi dei problemi analizzati e dei risultati esposti.

2. L'allocazione del tempo tra forme diverse di ricerca

Nell'introduzione si è esposto il quadro concettuale entro cui si considera il tema delle relazioni tra pubblicazioni e ricerca industriale. Le relazioni causali esposte vengono dunque presentate in maniera più rigorosa, facendo comunque astrazione da alcuni aspetti che pure potrebbero essere rilevanti.

Le ipotesi del modello che sviluppiamo sono le seguenti :

- Il professore universitario trae la sua utilità dalle pubblicazioni scientifiche (che gli garantiscono progressi di carriera e prestigio accademico) e dai guadagni derivanti dalla ricerca extra-universitaria.
- La quantità e qualità delle pubblicazioni sono proporzionali alla quantità di ricerca accademica e di ricerca industriale effettuata.
- Il numero di brevetti è proporzionale alla quantità di ricerca industriale effettuata. Questa relazione e quella precedente possono essere intese in termini probabilistici.
- Si suppone che tutta la ricerca di base sia ricerca accademica e che tutta la ricerca applicata venga fatta in collaborazione con l'industria.
- Il professore universitario deve allocare il suo tempo tra ricerca accademica e ricerca industriale (si suppone dunque, per semplicità, che sia costante il tempo da dedicare alla ricerca nel suo complesso e quindi che la scelta ottima tra tempo per la ricerca e tempo per altre attività sia uguale per tutti).
- Si suppone che l'attività di ricerca industriale sia remunerata, ovvero che la ricerca industriale abbia un prezzo. Per semplicità si suppone che il prezzo sia costante per ogni unità di tempo di ricerca industriale. Il guadagno derivante dalla ricerca industriale sarà dato perciò dal prodotto del prezzo unitario per le unità di tempo di ricerca industriale.
- Il professore universitario dunque massimizza la sua funzione di utilità sotto il vincolo del tempo. Ne deriva un'offerta di ricerca industriale in funzione del prezzo della ricerca stessa.
- L'industria ha dal canto suo una domanda di ricerca industriale affidata a professori universitari. Si suppone che questa domanda sia inversamente proporzionale al prezzo e direttamente proporzionale alle pubblicazioni ed al tempo di ricerca industriale effettuato in precedenza. Si suppone dunque che le pubblicazioni precedenti svolgano un ruolo di segnalazione dell'abilità del docente e che la ricerca industriale precedente segnali una capacità e una volontà del docente di collaborare con l'industria. Per semplicità si suppone che la domanda di ricerca industriale sia influenzata dalle pubblicazioni e dalla ricerca industriale effettuati nel solo periodo appena precedente (ma si accennerà anche alla possibilità di adottare un modello con un'ipotesi meno restrittiva in tal senso).
- Si suppone che esista una soluzione di equilibrio per cui la domanda di ricerca industriale eguaglia l'offerta di ricerca.

Il problema del professore universitario è dunque, in un dato momento t:

$$\text{Max } U_t = U(P_t, G_t) \quad (1)$$

Dati:

$$P_t = P(RA_t, RI_t) \quad (2)$$

$$G_t = v \cdot RI_t \quad (3)$$

$$RA_t + RI_t = T \quad (4)$$

$$RI_t^d = RI^d(v_t, RI_{t-1}, P_{t-1}) \quad (5)$$

$$RI_t^d = RI_t^s \quad (6)$$

Dove:

t = periodo considerato

U = utilità

P = pubblicazioni

G = guadagni da ricerca industriale

RA = tempo dedicato alla ricerca accademica

RI = tempo dedicato alla ricerca industriale

v = prezzo di un'unità di tempo di ricerca industriale

T = tempo dedicato alla ricerca (accademica + industriale)

RI^d = domanda di ricerca industriale da parte dell'industria

RI^s = offerta di ricerca industriale da parte del docente

Il problema è dunque di massimizzazione vincolata: il docente massimizza la (1) sostituendo le variabili obiettivo secondo la (2) e la (3), dunque sceglie il livello ottimo di RA_t e RI_t dato il vincolo (4). Ne deriva un'offerta di ricerca industriale, funzione del prezzo della ricerca industriale. Quindi, data la domanda di ricerca industriale (5), la soluzione è quella di equilibrio data dalla (6).

Formalmente il problema si risolve nel modo seguente.

Si ha il lagrangiano (si omette l'indice t per comodità di scrittura)

$$\Lambda = U(P, G) + \lambda (T - RA - RI)$$

quindi, sostituendo le variabili secondo la (2) e la (3) si ha:

$$\Lambda = U(P(RA, RI), v \cdot RI) + \lambda (T - RA - RI)$$

Le condizioni di ottimo sono dunque :

$$\partial \Lambda / \partial RA = 0$$

$$\partial \Lambda / \partial RI = 0$$

$$\partial \Lambda / \partial \lambda = 0$$

da cui si ha:

$$\partial U / \partial P \cdot \partial P / \partial RA - \lambda = 0 \quad (7)$$

$$\partial U / \partial P \cdot \partial P / \partial RI + \partial U / \partial G \cdot v - \lambda = 0 \quad (8)$$

$$RA + RI = T \quad (9)$$

Dalla (7) e dalla (8) si ha :

$$\partial U / \partial P \cdot \partial P / \partial RA = \partial U / \partial P \cdot \partial P / \partial RI + \partial U / \partial G \cdot v$$

da cui, dividendo per $\partial U / \partial P$,

$$\partial P / \partial RA - \partial P / \partial RI = (\partial U / \partial G) / (\partial U / \partial P) \cdot v \quad (9a)$$

che si può scrivere anche:

$$v = (\partial P/\partial RA - \partial P/\partial RI) \cdot (\partial U/\partial P)/(\partial U/\partial G) \quad (9b)$$

Questa è dunque la curva di offerta di ricerca industriale del docente.

Data la domanda di ricerca industriale, la (5), a sua volta funzione di v , uguagliando domanda e offerta ((9b) = (5)), considerando la (4), si ottengono le soluzioni ottime RI_t^* , RA_t^* , v_t^* , da cui si ottengono a loro volta, sostituendo nella (2) e nella (1), P_t^* e U_t^* .

Il modello può avere diverse formulazioni alternative.

Ad esempio si può ipotizzare che vi sia una quota di ricerca applicata non industriale; in tal modo il problema dell'incontro di domanda e offerta di ricerca si porrebbe solo per la quota di ricerca applicata svolta in collaborazione con l'industria, mentre la ricerca applicata non industriale sarebbe svincolata da questo aspetto. Ciò complicherebbe il modello, perché si avrebbero allora tre forme di ricerca (di base, applicata accademica e applicata industriale): i vantaggi in termini di capacità esplicativa sarebbero modesti in confronto agli svantaggi in termini di complicazione. Dal punto di vista empirico, la presenza anche di una quota di ricerca industriale lascia intatti i rischi di endogenità³.

Si può poi ipotizzare che il brevetto generi di per sé (a prescindere dalla remunerazione per la ricerca condotta) una remunerazione per le licenze, oltre a comportare dei costi. In realtà non sembra che finora le licenze siano state particolarmente profittevoli per i docenti.

Si può altresì ipotizzare esplicitamente che l'attività di ricerca industriale incrementi le risorse e quindi generi un maggiore potenziale di ricerca. In realtà sovente l'attività di ricerca industriale è stata intrapresa per sopperire alla riduzione di fondi pubblici, lasciando quindi le risorse complessive sostanzialmente invariate⁴.

Si può ipotizzare che il vincolo temporale dello scienziato accademico includa almeno un'altra voce, quale potrebbe essere il tempo dedicato all'insegnamento: l'incremento della ricerca applicata potrebbe ripercuotersi quasi interamente sull'insegnamento anziché sull'attività di ricerca di base. Tuttavia, ai fini delle stime empiriche, questo allargamento ha senso solo se si dispone di dati relativi all'insegnamento, altrimenti ogni relazione tra pubblicazione e brevetti diventa ancor più difficile da interpretare.

Di particolare interesse è una variazione del modello che consideri oggetto di contrattazione tra ricercatore e industria, oltre al tempo da dedicare alla ricerca, non il prezzo ma il grado di "disclosure", cioè la quota di nuova conoscenza che lo scienziato può pubblicare (questa ipotesi è formulata da Carayol, 2001): la tutela industriale può andare infatti dalla totale segretezza, ad una segretezza "parziale", qual è quella imposta dal brevetto, alla segretezza totale. Ai fini empirici, in ogni caso, questa variante cambia poco, perché la quota di "disclosure" verrà comunque considerata, ma non potrà che esserlo come variabile non osservabile; vincolandola alla domanda industriale essa finisce per dipendere in ogni momento dai valori passati e può essere dunque fonte di endogenità; tuttavia il fatto che in ogni caso esistono altre fonti di endogenità rende ininfluente il considerarla come variabile esogena.

3. La relazione tra pubblicazioni e brevetti

3.1 La relazione "teorica"

La quantità di pubblicazioni ottenuta dal docente una volta risolto il problema di massimizzazione sarà:

$$P_t^* = P(RA_t^*, RI_t^*),$$

³ Il problema dell'endogenità tra pubblicazioni e brevetti verrà illustrato in seguito.

⁴ Per l'effetto di variazioni del finanziamento pubblico all'università, in una prospettiva storica, si veda Geuna (2001).

secondo la (2).

Non esistono studi empirici precedenti che possano darci informazioni sulla specificazione di questa “funzione di pubblicazione”.

Le ipotesi che è tuttavia lecito fare sono:

$$\partial P / \partial RA > 0 \quad (10)$$

$$\partial P / \partial RI > 0 \quad (11)$$

$$\partial^2 P / \partial RA \partial RI \geq 0 \quad (12)$$

L'ipotesi (10) è la più ovvia: la fonte principale delle pubblicazioni accademiche è la ricerca accademica. Anche se naturalmente il legame tra ricerca e pubblicazioni non è di tipo deterministico, appare del tutto ragionevole ipotizzare che un aumento del tempo dedicato alla ricerca accademica comporti *in media* un aumento del numero delle pubblicazioni.

L'ipotesi (11) si basa sulla constatazione che esistono riviste che pubblicano articoli di ricerca applicata (Geuna e Nesta, 2003), qual è tipicamente quella che si fa in collaborazione con le imprese. Ma soprattutto la ricerca industriale è fonte di finanziamenti per la ricerca: disponendo di nuove attrezzature, laboratori, ecc. si può condurre anche una migliore ricerca di base, e quindi pubblicare di più. Inoltre alcuni questionari confermano che la ricerca industriale fornisce spunti importanti anche per la ricerca di base (Balconi, Borghini e Moisello, 2003). Di qui anche la giustificazione per l'ipotesi (12) sulla derivata seconda mista: per un dato livello di ricerca accademica, un'ora aggiuntiva di ricerca industriale aumenta ragionevolmente lo stock complessivo di conoscenze del professore, con un effetto certamente non negativo sulle pubblicazioni (ma verosimilmente positivo).

Alla luce di queste ipotesi possiamo stabilire cosa accade alle pubblicazioni in corrispondenza di valori diversi della scelta ottima (RA_t^*, RI_t^*).

Si ha che:

$$dP/dRI = (\partial P / \partial RA) * (\partial RA / \partial RI) + \partial P / \partial RI \quad (13)$$

Dato il vincolo temporale (4), si ha:

$$\partial RA / \partial RI = -1 \quad (14)$$

Da cui la (13) diventa:

$$dP/dRI = \partial P / \partial RI - \partial P / \partial RA \quad (15)$$

In base alle ipotesi fatte, nessuna conclusione si può trarre sul segno di questa derivata.

Si noti l'importanza della positività del segno della (12): esso vuol dire in sostanza che esiste un'interazione positiva tra le due forme di ricerca e che la produttività di una dipende dal livello dell'altra. In assenza di tale interazione la produttività di ciascuna forma di ricerca è costante o dipende dal livello di quella stessa forma di ricerca. In tal caso è presumibile che aumentando la ricerca industriale e diminuendo la ricerca accademica diminuisca il numero delle pubblicazioni, essendo la pubblicazione il fine ultimo della seconda, mentre non è il fine unico della ricerca industriale, ed inoltre possono essere imposte delle restrizioni sulla divulgazione dei risultati di ricerca industriale. Ma se esiste un'interazione positiva tra RA e RI, allora, aumentando la ricerca industriale, la ricerca accademica, pur diminuendo, diventa più produttiva, e questa maggiore produttività può più che compensare l'effetto negativo dato dalla sua riduzione.

Un esempio di “funzione di pubblicazione” potrebbe essere:

$$P = \beta_0 + \beta_1 RA + \beta_2 RI + \beta_3 RA RI \quad (16)$$

essendo tutti i coefficienti positivi.

Anche se si ipotizza $\beta_1 > \beta_2$, la presenza del termine di interazione (e dunque la derivata seconda mista positiva) fa sì che vi sia un intervallo in cui $dP/dRI > 0$. Tale funzione risulta essere crescente per valori bassi di RI, quindi decrescente (ad un certo punto l'effetto dato dall'interazione non riesce più a compensare la perdita data dalla riduzione del tempo di ricerca accademica). L'effetto medio può comunque essere positivo (se una relazione di questo tipo, quadratica, si stima con una relazione lineare si valuta sostanzialmente l'effetto medio).

Viceversa, se $\beta_3 = 0$, si ha $dP/dRI < 0$, qualunque sia il livello di RI⁵.

Per quanto riguarda i brevetti, si può ipotizzare che essi dipendano esclusivamente, ed in maniera positiva, dal tempo dedicato alla ricerca industriale. Naturalmente è ragionevole ipotizzare che si tratti di una funzione probabilistica o comunque discreta.

Avremo dunque:

$$\Pr(BR) = f(RI) \quad (17)$$

con

$$\partial \Pr(BR) / \partial RI > 0 \quad (18)$$

e quindi :

$$\partial BR / \partial RI \geq 0 \quad (19)$$

Se si pongono in relazione pubblicazioni e brevetti si avrà dunque:

$$dP/dBR = (dP/dRI) / (dBR/dRI) \quad (20)$$

Assumendo come positivo il segno del denominatore, evidentemente il segno dell'espressione dipende dal segno del numeratore⁶.

Dunque, confrontando due individui con un numero diverso di brevetti, l'individuo con un numero maggiore di brevetti avrà effettuato un maggior numero di ore di ricerca industriale; se l'effetto di un'ora aggiuntiva di ricerca industriale, a scapito di un'ora aggiuntiva di ricerca accademica, è negativo sulle pubblicazioni, allora l'individuo con più brevetti avrà meno pubblicazioni; viceversa se sostituire un'ora di ricerca accademica con un'ora di ricerca industriale provoca un effetto positivo sulle pubblicazioni.

Pertanto, in questo schema concettuale, una relazione media positiva tra brevetti e pubblicazioni è determinata da un impatto positivo sulle pubblicazioni della sostituzione della

⁵ Un aspetto del "modello" che è alla base di questa derivazione che può essere considerato poco realistico è la somma costante del tempo dedicato alle due forme di ricerca. E' infatti possibile scomporre in più voci il tempo del docente, includendo ad esempio le attività didattiche (che include non solo il tempo dedicato alle lezioni, ma anche orari di ricevimento, assistenza alle tesi, ecc.) o si può arrivare a comprendere l'intero "tempo" del docente, ivi incluso il suo tempo libero. Qualunque sia la variante considerata, non è più certo che un aumento di tempo dedicato alla ricerca industriale sia perfettamente compensato da una riduzione del tempo dedicato alla ricerca accademica. Adirittura può darsi un aumento contemporaneo delle due forme di ricerca. Dunque, se rimane vera la (13), non è più vera la (14) e di conseguenza la (15). In ogni caso il segno di dP/dRI rimane incerto. Infatti, se col vincolo $RA+RI=T$, è certo che all'aumentare della ricerca industriale la ricerca accademica diminuisca e dunque vi può essere un effetto positivo sulle pubblicazioni solo se la produttività della ricerca industriale è (almeno per certi intervalli) superiore a quella della ricerca accademica, se il vincolo si allenta, l'effetto positivo di un aumento della ricerca industriale può essere dovuto anche al fatto che può contemporaneamente aumentare (o rimanere costante o ridursi di poco) il tempo dedicato alla ricerca accademica.

Il modello "corretto" in merito all'allocazione temporale non è veramente rilevante perché ciò che si vuole dimostrare in quest'articolo, è che, qualunque sia l'impatto della ricerca industriale sulle pubblicazioni, la semplice relazione tra brevetti e pubblicazioni non è in grado di misurarlo.

⁶ Se il denominatore è pari a zero, il rapporto tende ha il limite all'infinito: la relazione tra pubblicazioni (y) e brevetti (x) è una retta verticale: le pubblicazioni variano, per effetto del variare di RI, ma RI non ha effetto su BR, che rimangono costanti.

ricerca accademica con la ricerca industriale, almeno per alcuni valori dei tempi di ricerca. Ciò può essere ragionevolmente spiegato dalla presenza di effetti di interazione tra le due forme di ricerca.

3.2 Esiti possibili dell'analisi empirica

In base alla relazione “teorica” tra brevetti e pubblicazioni, è possibile vedere il risultato che si ottiene volendo impostare una regressione lineare tra le due variabili. Dopo il modello più semplice, che riflette l'impostazione generale, priva di eterogeneità individuali e di fattori casuali, considereremo alcuni “fattori di disturbo” che possono allontanare la situazione reale da quella appena delineata e che quindi possono condizionare la stima dei parametri di una regressione.

a) Il modello più semplice

L'idea alla base del “modello” è che il numero di pubblicazioni dipenda direttamente dalla distribuzione del tempo del ricercatore tra ricerca accademica e industriale; poiché la loro somma è costante, naturalmente basta porre la dipendenza da una delle due forme di ricerca.

A sua volta il numero di brevetti è in relazione diretta col tempo dedicato alla ricerca industriale.

Dunque, a parità di accordi sulla “disclosure” (quanta parte di conoscenza applicata brevettare e quale pubblicare) c'è una relazione esclusivamente indiretta tra pubblicazioni e numero di brevetti, che dipende dalle due relazioni sottostanti (tra pubblicazioni e ricerca industriale e tra brevetti e ricerca industriale). Dunque i brevetti possono essere considerati una *proxy* del tempo dedicato alla ricerca industriale.

Considerando delle semplici relazioni lineari, avremo⁷:

$$P_i = \gamma_0 + \gamma_1 RI_i + \varepsilon_i \quad (21)$$

$$BR_i = \delta_0 + \delta_1 RI_i \quad (22)$$

da cui, sostituendo, si ottiene:

$$P_i = (\gamma_0 - \gamma_1 \delta_0 / \delta_1) + (\gamma_1 / \delta_1) BR_i + \varepsilon_i \quad (23)$$

Se si stima dunque la relazione:

$$P_i = \beta_0 + \beta_1 BR_i + \varepsilon_i \quad (24)$$

Si avrà:

$$E(b_1) = \beta_1 = \gamma_1 / \delta_1 \quad (25)$$

essendo b_1 è la stima di β_1 .

Questa applicazione econometrica con funzioni lineari riflette dunque quanto asserito su una forma generica delle funzioni: l'effetto dei brevetti sulle pubblicazioni è dato dal rapporto tra l'effetto del tempo di ricerca industriale sulle pubblicazioni e l'effetto del tempo di ricerca industriale sui brevetti.

Dando dunque per certo il segno positivo di δ_1 , il segno di $E(b_1)$ dipenderà dal segno di γ_1 , cioè dall'effetto medio dell'incremento di un'unità di tempo di ricerca industriale sulle pubblicazioni.⁸

⁷ In tutto l'articolo per semplicità verrà omessa nelle equazioni la matrice delle altre variabili rilevanti (settore tecnologico, età, ecc.)

Tuttavia ciò è vero soltanto ad una prima approssimazione. Cioè è possibile dimostrare che, stimando la (24) si può avere un $E(b_1)$ positivo anche in presenza di un effetto negativo di un aumento del tempo dedicato alla ricerca industriale sulle pubblicazioni. In altre parole vi possono essere spiegazioni di una relazione positiva tra pubblicazioni e brevetti diverse da un effetto di complementarità tra le forme di ricerca.

b) Caratteristiche individuali non osservabili

Una prima spiegazione può essere rappresentata dalla presenza di caratteristiche individuali non osservabili. Considerando due individui, essi possono ottenere un numero diverso di brevetti pur svolgendo il medesimo numero di ore di ricerca industriale (o addirittura può avere più brevetti chi fa meno ricerca industriale), a causa di differenti “abilità” o di altre caratteristiche non osservabili (come possono essere differenti accordi sulla “disclosure” dei risultati: due individui, pur avendo svolto lo stesso numero di ore di ricerca industriale ed avendo ottenuto gli stessi risultati in termini di nuova conoscenza, hanno tuttavia un numero differente di brevetti, perché diversi sono i loro accordi con le imprese committenti).

Le relazioni che si ipotizzano come “vere” sono dunque:

$$P_i = \gamma_0 + \gamma_1 R_i + \varepsilon_i \quad (21)$$

$$BR_i = \delta_0 + \delta_1 R_i + B_i \quad (26)$$

essendo B_i caratteristiche individuali non osservabili.

In tal caso si otterrà:

$$P_i = (\gamma_0 - \gamma_1 \delta_0 / \delta_1) + (\gamma_1 / \delta_1) BR_i - (\gamma_1 / \delta_1) B_i + \varepsilon_i \quad (27)$$

essendo B_i inosservabili, può essere stimata esclusivamente la relazione:

$$P_i = \beta_0 + \beta_1 BR_i + \varepsilon_i \quad (28)$$

caratterizzata dunque dall’omissione di una variabile rilevante rispetto al modello “vero”. Si avrà pertanto:

$$E(b_1) = \gamma_1 / \delta_1 - \gamma_1 / \delta_1 \cdot [\text{Cov}(BR_i, B_i) / \text{Var}(BR_i)] \quad (29)$$

Se esiste una covarianza positiva tra brevetti e caratteristiche individuali non osservabili, allora il segno di questa espressione può essere positivo anche se γ_1 è negativo.

In termini intuitivi, ciò si può tradurre in questo modo: supponiamo che la relazione tra pubblicazioni e tempo dedicato alla ricerca industriale sia negativa: chi fa più ricerca industriale ha meno pubblicazioni. Se chi ottiene un numero maggiore di brevetti lo ha fatto non grazie ad un tempo superiore dedicato alla ricerca industriale, ma grazie ad “abilità” maggiore, allora chi detiene più brevetti può aver svolto anche meno ore di ricerca industriale e avere al tempo stesso più pubblicazioni.

⁸ Se esiste un’interazione tra le due forme di ricerca, come abbiamo visto, la relazione tra tempi di ricerca e pubblicazione e dunque tra brevetti e pubblicazioni, può essere non lineare. Avendo il solo termine di primo grado si stima il solo effetto medio, che può essere positivo o negativo.

c) *Correlazione positiva tra caratteristiche individuali non osservabili*

Questo risultato è ancora più forte se si fa dipendere anche il numero di pubblicazioni dalle abilità individuali non osservabili e si ipotizza una relazione positiva tra le abilità nella pubblicazioni e le abilità nell'ottenere brevetti.

Le relazioni che si ipotizzano come "vere" sono dunque:

$$P_i = \gamma_0 + \gamma_1 R_i + A_i + \varepsilon_i \quad (30)$$

$$BR_i = \delta_0 + \delta_1 R_i + B_i \quad (26)$$

essendo A_i caratteristiche individuali non osservabili.

Ipotizziamo che esista una relazione tra i due tipi di caratteristiche individuali. Sia ad esempio:

$$A_i = B_i + v_i \quad (31)$$

Sostituendo si otterrà:

$$P_i = (\gamma_0 - \gamma_1 \delta_0 / \delta_1) + (\gamma_1 / \delta_1) BR_i - (\gamma_1 / \delta_1 - 1) B_i + v_i + \varepsilon_i \quad (32)$$

Per cui, stimando la relazione

$$P_i = \beta_0 + \beta_1 BR_i + \varepsilon_i \quad (24)$$

si avrà:

$$E(b_1) = \gamma_1 / \delta_1 - (\gamma_1 / \delta_1 - 1) \cdot [\text{Cov}(BR_i, B_i) / \text{Var}(BR_i)] \quad (33)$$

Questo valore è maggiore di quello espresso dalla (29)

d) *Eterogeneità individuali e endogenità derivante della domanda di ricerca industriale*

Qual è l'evidenza empirica di tali termini di eterogeneità individuale?

Dall'osservazione dei dati su pubblicazioni emerge una forte variabilità del loro numero tra i diversi individui, una variabilità che non può trovare una completa spiegazione con differenti caratteristiche osservabili degli individui (allocazione dei tempi di ricerca, dipartimento, età, posizione accademica, ecc.). E' dunque evidente il peso di caratteristiche individuali non osservabili nella "funzione di pubblicazione" Questa è in fondo una formulazione alternativa del mertoniano "effetto san Matteo" (Merton, 1988), per cui chi pubblica più oggi pubblicherà di più in futuro, con la differenza che qui questo effetto è spiegato esclusivamente da capacità reali.

Minore è invece la variabilità dei dati sui brevetti. Quindi sul piano empirico non vi sono ragioni stringenti per introdurre un termine di eterogeneità individuale nelle "funzione di brevetti"

Ebbene, è facile verificare che, se le due equazioni "vere" sono:

$$P_i = \gamma_0 + \gamma_1 R_i + A_i + \varepsilon_i \quad (34)$$

$$BR_i = \delta_0 + \delta_1 R_i \quad (22)$$

si ottiene:

$$P_i = (\gamma_0 - \gamma_1 \delta_0 / \delta_1) + (\gamma_1 / \delta_1) BR_i + A_i + \varepsilon_i \quad (35)$$

Stimando la consueta equazione di regressione:

$$P_i = \beta_0 + \beta_1 BR_i + \varepsilon_i \quad (24)$$

vi è sì l'omissione di una variabile, tuttavia non c'è alcuna ragione di supporre questa variabile correlata col numero di brevetti, e quindi la sua omissione non è rilevante nella stima di β_1

Si ha cioè:

$$E(b_1) = \gamma_1/\delta_1 \quad (25)$$

Tuttavia, anche in assenza di un termine esplicito di eterogeneità individuale non osservabile nella “funzione di brevettazione”, le ipotesi fatte sulla domanda di ricerca industriale comportano di per sé l'esistenza di un *bias* nella stima dell'effetto dei brevetti sulle pubblicazioni. Ovvero la stima di quest'effetto non può comunque ridursi al rapporto tra produttività della ricerca industriale rispetto alle pubblicazioni e produttività della ricerca industriale rispetto ai brevetti.

In altre parole, le ipotesi fatte sulla domanda di ricerca industriale comportano la possibilità che la relazione tra pubblicazioni e brevetti sia positiva, anche se l'effetto della ricerca industriale sulle pubblicazioni è negativa.

Infatti, le ipotesi fatte sulla domanda di ricerca industriale comportano che

$$RI_{it} = RI(P_{i,t-1}, BR_{i,t-1}) \quad (36)$$

ma

$$P_{i,t-1} = P(A_i) \quad (37)$$

$$BR_{i,t-1} = P(A_i) \quad (38)$$

da cui:

$$RI_i = RI(A_i) \quad (39)$$

e quindi:

$$BR_i = BR(A_i) \quad (40)$$

Dunque, essendo “vere” la (34) e la (22) “vera” la (35), ma data anche la (40) quando si va a stimare la consueta equazione di regressione:

$$P_i = \beta_0 + \beta_1 BR_i + \varepsilon_i \quad (24)$$

si sta omettendo una variabile rilevante (A_i) e la variabile indipendente inclusa è correlata alla variabile esclusa. Pertanto si avrà:

$$E(b_1) = \gamma_1/\delta_1 + [\text{Cov}(BR_i, A_i)/\text{Var}(BR_i)] \quad (41)$$

E' dunque possibile ottenere un b_1 positivo anche in presenza di un effetto negativo sulle pubblicazioni di un aumento della ricerca industriale.⁹

Detto in termini intuitivi: supponiamo che la ricerca industriale abbia un impatto negativo sulle pubblicazioni (cioè: sostituire ricerca accademica con ricerca industriale ha un impatto negativo sulle pubblicazioni). Consideriamo due individui con tutte le caratteristiche identiche ma di “abilità” diversa, che partono con un pari livello di ricerca industriale: il più abile pubblicherà di più, dunque avrà una maggiore domanda di ricerca industriale, quindi farà più brevetti. A questo punto il più abile, pur facendo più ricerca industriale, potrebbe continuare a pubblicare di più in ragione della sua abilità: dunque colui che brevetta di più farà comunque più pubblicazioni, anche se l'impatto *ceteris paribus* della ricerca industriale sulle pubblicazioni è negativo.

⁹ Non è tuttavia scontato che la covarianza tra BR_i e A_i sia positiva: i soggetti più “abili” potrebbero svolgere così tanta ricerca industriale da far abbassare il loro numero di pubblicazioni

e) *Stime "unbiased" con dati panel*

A questo punto va tuttavia fatta un'osservazione molto importante: queste caratteristiche individuali, variabili da individuo a individuo ma per ogni individuo costanti nel tempo risultano variabili non osservabili solo se si dispone di dati *cross section* (rilevazioni su più individui in uno stesso momento temporale).

Se si dispone invece di dati *panel* (più individui in più momenti) è possibile esplicitare questi effetti individuali e dunque stimarli tramite un modello *fixed effect* o *random effect*.¹⁰

Se le equazioni "vere" sono:

$$P_{it} = \gamma_0 + \gamma_1 RI_{it} + A_i + \varepsilon_{it} \quad (42)$$

$$BR_{it} = \delta_0 + \delta_1 RI_{it} \quad (22)$$

si ottiene:

$$P_i = (\gamma_0 - \gamma_1 \delta_0 / \delta_1) + (\gamma_1 / \delta_1) BR_{it} - (\gamma_1 / \delta_1) A_i + \varepsilon_{it} \quad (44)$$

Si può dunque stimare l'equazione di regressione:

$$P_{it} = \beta_0 + \beta_1 BR_{it} + \sum_{j=1, n-1} \beta_2 D_{ji} + \varepsilon_{it} \quad (45a)$$

essendo n il numero degli individui.

Oppure:

$$P_{it} = \beta_0 + \beta_1 BR_{it} + U_i + \varepsilon_{it} \quad (45b)$$

La (45a) illustra un modello *fixed effect*, la (45b) un modello *random effect*.

Come è noto, se c'è correlazione tra variabile esplicativa (numero di brevetti) ed effetti individuali (A_i) è da preferirsi il modello *fixed effect* (stimatore LSDV), perché lo stimatore del modello *random effect* (stimatore GLS) darebbe stime *biased*. Viceversa, in assenza di tale correlazione, il modello *random effect* è preferibile, perché lo stimatore LSDV darebbe stime inefficienti, ancorché corrette.

Le ipotesi avanzate nei paragrafi precedenti ci permettono di dire qualcosa in merito alla correlazione tra variabile esplicativa e caratteristiche individuali: se sono veri i modelli illustrati nel paragrafo b) o nel paragrafo c) (presenza di effetti individuali nella "funzione di brevettazione", correlati o meno con gli effetti individuali della "funzione di pubblicazione") c'è correlazione tra regressore (numero di brevetti) ed effetti individuali (A_i). In tal caso solo il modello *fixed effect* dà stime corrette.

Se invece ci si trova nella situazione descritta nel paragrafo a) oppure nella prima parte del paragrafo d) (assenza di effetti individuali oppure effetti individuali solo nella "funzione di pubblicazione", senza ipotesi sulla domanda di ricerca industriale) la correlazione non esiste e il modello *random effect* è preferibile¹¹.

In ogni caso, sia che tale correlazione esista, sia che non esista, scegliendo lo stimatore opportuno è possibile ottenere stime *unbiased*, cioè:

$$E(b_1) = \gamma_1 / \delta_1 \quad (25)$$

Ci si aspetta dunque che il segno di b_1 rifletta la produttività della ricerca industriale in termini di pubblicazioni, senza includere ulteriori effetti, come abbiamo visto può accadere (e se le

¹⁰ Qui si fa riferimento al solo modello di regressione lineare e non ad altri tipi di modelli, come la regressione di Poisson, la negative binomial regression, ecc., che potrebbero essere indicate data la frequente presenza nella variabile dipendente (le pubblicazioni) di molte osservazioni pari a 0.

¹¹ Se è vera l'ipotesi di assenza di correlazione tra caratteristiche individuali non osservabili e numero di brevetti una stime OLS su dati *cross section* che aggregano i differenti periodi in cui è suddiviso l'ipotetico database *panel* risultano corrette ed efficienti.

ipotesi sulla formazioni della domanda di ricerca industriale sono corrette accade inevitabilmente) con dati *cross-section*.

Si osservi che, se gli effetti individuali sono significativi, stimando la stessa equazione di regressione sullo stesso insieme di dati con il metodo OLS *pooled* (che tratta quindi i dati come *cross section*) o con il metodo dei LSDV (stime *panel* con effetti fissi) si dovrebbero ottenere stime significativamente diverse.

f) *Fonti del "bias" con dati panel.*

Tuttavia anche disponendo di dati *panel* il discorso non può concludersi qui.

Il problema nasce dalle ipotesi fatte sulla domanda di ricerca industriale: con la struttura ipotizzata della (36) (ricerca industriale al tempo t dipendente da pubblicazioni e ricerca industriale effettuate al tempo t-1) si verifica quanto visto nel paragrafo d) e quindi, essendo vera la (40), c'è correlazione tra regressore (numero di brevetti) ed effetti individuali (A_i). Ciò esclude l'utilizzo del modello *random effect*.

Ma inoltre, fatta l'ipotesi della (36), non risulta corretto neanche lo stimatore LSDV del modello *fixed effect*, a causa della correlazione della variabile esplicativa con i residui. Infatti, dovendo stimare la (45a), lo stimatore *within* stima.

$$P_{it} - P_i = \beta_1 (BRI_{it} - BRI_i) + (\varepsilon_{it} - \varepsilon_i) \quad (46)$$

in cui P_i , BRI_i , ε_i sono le medie nel tempo per ciascun individuo delle rispettive variabili. Ora, data la (36), si ha correlazione tra BRI_{it} e $P_{i,t-1}$ e quindi tra BRI_{it} e ε_i .

Ne conseguono stime *biased*. Non va tuttavia dimenticato che la correlazione tra esplicative e residui viene meno per t grande.

Il vantaggio di avere più momenti temporali, e per di più piccoli, nel *database* deriva anche da un'altra considerazione: se i momenti temporali sono ampi e quindi pochi, si possono sovrapporre più unità di tempo. Si parla qui di unità di tempo come concetto astratto, non coincidente con una precisa unità di tempo effettiva. L'unità di tempo teorica può essere definita come quello spazio temporale per cui la ricerca industriale svolta in quel momento non è influenzata dalle pubblicazioni e dai brevetti svolte nello stesso momento.

Al momento concreto della stima bisogna fare riferimento ad una effettiva unità di tempo, ad esempio l'anno. Ebbene, non può escludersi che la quantità di ricerca industriale svolta in quell'anno sia influenzata dalle pubblicazioni svolte in quel medesimo anno (la ricerca industriale svolta alla fine dell'anno può dipendere dalle pubblicazioni effettuate all'inizio). Ciò genera endogenità ed è quindi una fonte di "bias".¹²

¹² Se l'unità di tempo effettiva è più ampia di quella teorica è come se si stessero sommando le pubblicazioni di due o più unità di tempo. Si ha cioè:

$$P_0 = \gamma_0 + \gamma_1 RI_0 + u_0$$

$$P_1 = \gamma_0 + \gamma_1 RI_1 + u_1$$

Immaginiamo, per semplificare, che t=0 sia il primo "anno" di ricerca per il docente. In t=1, la quantità di ricerca industriale svolta dal docente dipende, in virtù della domanda dell'industria, dalle pubblicazioni e dai brevetti svolti in t=0. Si ha cioè:

$$RI_1 = RI_0 + \eta P_0(u_0) + \theta RI_0 + \varepsilon_1$$

Da cui:

$$P_0 + P_1 = 2\gamma_0 + 2\gamma_1(RI_0 + RI_1(u_0)) + (u_0 + u_1)$$

Consideriamo la più semplice delle relazioni tra brevetti e ricerca industriale, cioè la relazione deterministica, senza neanche la costante:

$$BR_0 = \delta_1 RI_0$$

Se gli intervalli di tempo considerati sono piccoli questo rischio si riduce (l'unità di tempo empirica tende a coincidere con quella teorica).

E' dunque vantaggioso per più motivi suddividere il periodo analizzato in molti sottoperiodi: aumenta il numero di osservazioni, dando quindi consistenza allo stimatore LSDV (la correlazione tra esplicativa e residui viene meno asintoticamente) e si riduce il rischio di sovrapposizione di piani temporali, riducendo quindi il rischio della medesima correlazione.

Il *bias* derivante dalla correlazione tra esplicative e residui sussiste fatta l'ipotesi (36) sulla domanda di ricerca industriale.

Tuttavia, anche nell'ipotesi che tale domanda non dipenda dalle pubblicazioni e/o dai brevetti passati, non sembra in ogni caso possibile sfuggire al problema del *bias* nella stima dell'effetto dei brevetti sulle pubblicazioni.

E' infatti evidente che il conseguimento di un brevetto è un risultato fortemente legato a componenti casuali. Vi possono essere ricerche che, per quanto prolungate ed intense, non portano a risultati significativi, e viceversa.

Appare dunque certamente corretto far dipendere il numero di brevetti conseguiti non solo dalla quantità di ricerca industriale svolta ma anche da un termine casuale.

Le relazioni "vere" sarebbero dunque:

$$P_{it} = \gamma_0 + \gamma_1 RI_{it} + A_i + \varepsilon_{it} \quad (42)$$

$$BR_{it} = \delta_0 + \delta_1 RI_{it} + u_{it} \quad (47)$$

In tal caso si otterrà:

$$P_{it} = (\gamma_0 - \gamma_1 \delta_0 / \delta_1) + (\gamma_1 / \delta_1) BR_{it} + A_i - [(\gamma_1 / \delta_1) u_{it} + \varepsilon_{it}] \quad (48)$$

Se si stima:

$$P_{it} = \beta_0 + \beta_1 BR_{it} + \sum_{j=1, n-1} \beta_2 D_{ji} + \varepsilon_{it} \quad (45)$$

si verifica una correlazione tra BR_{it} e i residui (essendo BR_{it} non indipendente da u_{it}), da cui si ottiene:

$$E(b_1) = \gamma_1 / \delta_1 - \gamma_1 / \delta_1 \cdot [\text{Cov}(BR_{it}, u_{it}) / \text{Var}(BR_{it})] \quad (49)$$

Se dunque la covarianza tra brevetti ed errori nella "funzione di brevettazione" è positiva, è possibile una relazione positiva tra pubblicazioni e brevetti anche in caso di un effetto negativo sulle pubblicazioni di un aumento della ricerca industriale.

In termini intuitivi questo vuol dire: dati due individui che effettuano due livelli diversi di ricerca industriale, può accadere che colui che effettua meno ricerca industriale sia più "fortunato" e brevetti di più dell'altro; inoltre, fatta l'ipotesi di effetto negativo della ricerca industriale sui

$$BR_1 = \delta_1 RI_1$$

Si ha:

$$P_0 + P_1 = 2\gamma_0 + \gamma_1 / \delta_1 (BR_0 + BR_1(u_0)) + (u_0 + u_1)$$

Stimando:

$$\Sigma P_{it} = \alpha + \beta \Sigma BR_{it} + v_{it}$$

si ha correlazione tra la variabile esplicativa e il termine di errore e conseguente *bias* nella stima di β , da cui $E(b) \neq \gamma_1 / \delta_1$. Ciò vale anche in assenza di caratteristiche individuali non osservabili (che infatti abbiamo ommesso in questo esempio). Dunque in una cross-section che consideri un arco temporale vasto, al *bias* dovuto all'impossibilità di controllare per le eterogeneità individuali non osservabili, si somma questa fonte di *bias*.

Pertanto, anche se le caratteristiche individuali non osservabili fossero assenti o ininfluenti le stime risulterebbero comunque *biased*.

In un'analisi *panel* si può controllare per le eterogeneità individuali e si può sfuggire all'endogenità di natura "temporale" solo considerando periodi adeguatamente piccoli.

brevetti, colui che fa meno ricerca industriale avrà più pubblicazioni: dunque l'individuo con più brevetti avrà anche più pubblicazioni.

Questo effetto è ancora più forte se esiste una correlazione positiva tra gli errori della "funzione di brevettazione" e gli errori della "funzione di pubblicazione": dal punto di vista statistico il confronto tra questa situazione e quella illustrata precedentemente è analogo al confronto tra la situazioni illustrate ai punti b) e c).

La logica di una correlazione positiva tra questi due termini di errore può essere così giustificata: la ricerca industriale genera nuova "conoscenza applicata", parte della quale viene pubblicata, parte viene brevettata: è l'intera produzione di conoscenza applicata a dipendere da un termine casuale: gli errori per pubblicazioni e brevetti dipendono per intero o in parte da questo errore.

Nel termine di errore può essere fatto rientrare concettualmente anche la quota di conoscenza applicata che può essere pubblicata e quella che può essere brevettata: si tratta infatti di un termine che varia da individuo a individuo ed anche da momento a momento e non è comunque esplicitamente osservabile. Anche questo termine può essere dunque alla base di una relazione positiva tra brevetti e pubblicazioni: se, facendo più ricerca industriale, quindi producendo più conoscenza applicata, aumenta sia la quota di conoscenza applicata che può essere brevettata, sia la quota che può essere pubblicata (a scapito della quota che rimane segreta), ci sarà un aumento dei brevetti e può esserci anche un aumento delle pubblicazioni (anche se *ceteris paribus* l'aumento di ricerca industriale facesse calare le pubblicazioni).

g) *Il ritardo nelle pubblicazioni*

L'analisi fin qui condotta ha fatto astrazione da un problema che sempre più spesso viene preso in considerazione come più probabile effetto negativo dell'attività di brevettazione svolta da docenti universitari: il ritardo nella pubblicazione di risultati innovativi, quando questi possono essere oggetto di un brevetto. Infatti, perché una nuova conoscenza possa essere brevettata è necessario che la conoscenza abbia le caratteristiche della novità e dunque non sia stata pubblicata. Dopo il deposito della domanda di brevetto è naturalmente possibile pubblicare ciò che era stato "rimandato".

Sebbene il fenomeno del ritardo nella pubblicazione non possa essere considerato grave come la riduzione *tout-court* delle pubblicazioni, si tratta tuttavia di una minaccia alla libera circolazione del sapere e potrebbe rappresentare un freno allo sviluppo.

Come detto, nell'analisi fin qui condotta si è fatto astrazione da questo aspetto, considerando come unità di tempo quella in cui la domanda di ricerca industriale non è influenzata dal comportamento precedente (in termini di pubblicazioni e brevetti) dello scienziato o comunque quella in cui non esiste endogenità tra pubblicazioni e brevetti.

Volendo invece condurre un'indagine sul ritardo nella pubblicazioni bisogna distinguere tra un momento precedente ed uno successivo alla domanda di brevetto.

Per condurre più efficacemente questa indagine è necessario esplicitare il fatto che la ricerca industriale produca conoscenza applicata (AK) e che di questa una parte (q) viene pubblicata prima del brevetto, una parte (r) viene appunto brevettata ed una parte (s) viene tenuta segreta.

Si ha dunque:

$$q + r + s = 1 \quad (50)$$

La conoscenza oggetto di un brevetto viene quindi pubblicata nel momento successivo.

Sia A è il numero di pubblicazioni derivanti da ricerca applicata (che nella nostra ipotesi semplificatrice coincide con la ricerca industriale), e B è il numero di pubblicazioni derivanti da ricerca di base (che nella nostra ipotesi semplificatrice coincide con la ricerca applicata).

Consideriamo una sequenza ideale di 3 momenti: t-1, t e t+1; un brevetto avviene nel momento t.

Si avrà:

$$P_{it-1} = B_{it-1} + A_{it-1} \quad (51)$$

dove

$$A_{it-1} = q AK_{it-1} \quad (52)$$

Poichè in questo momento non ci sono brevetti, viene esclusa dalla pubblicazione solo l'eventuale conoscenza "segreta", da cui:

$$q = 1 - s \quad (53)$$

da cui:

$$P_{it-1} = B_{it-1} + (1-s) AK_{it-1} \quad (54)$$

Nel periodo successivo, in cui invece c'è uno o più brevetti, si avrà:

$$q = 1 - r - s \quad (55)$$

e dunque:

$$P_{it} = B_{it} + (1-r-s) AK_{it} \quad (56)$$

e quindi:

$$BR_{it} = f(rAK_{it}) \quad (57)$$

La conoscenza generata da tale brevetto sarà poi pubblicata nel momento successivo, da cui:

$$P_{it+1} = B_{it+1} + A_{it+1} + rAK_{it} \quad (58)$$

Nel momento successivo nuovamente non vi sono brevetti e dunque si ha

$$q = 1 - s \quad (59)$$

da cui:

$$P_{it+1} = B_{it+1} + (1-s) AK_{it+1} + rAK_{it} \quad (60)$$

Se la relazione tra brevetti e ricerca industriale è deterministica (si hanno più brevetti se e solo se si fa più ricerca industriale), se s rimane costante, se sostituendo ricerca accademica con ricerca industriale si ottiene una quantità di conoscenza (di base più applicata: B + AK) non superiore e se tutti i risultati brevettabili vengono poi pubblicati, è facile ricavare che:

$$P_{it} < P_{it-1} \quad (61)$$

$$P_{it} < P_{it+1} \quad (62)$$

Se si aggiunge l'ipotesi che al momento t+1 si torna a fare la stessa ricerca industriale del momento t-1, si ricava:¹³

¹³ Qui il problema è generato dalla natura discreta del numero dei brevetti. A rigore, la "funzione di brevettazione" è continua, e quindi un numero pari di brevetti corrisponde ad un pari numero di ore di ricerca industriale. Ma nella realtà il numero di brevetti è discreto, per cui si può immaginare una funzione aggiuntiva che faccia corrispondere il numero "vero" di brevetti a degli intervalli della funzione "teorica". In tal modo è possibile che un pari numero di brevetti corrisponda ad un numero diverso di ore di ricerca industriale.

$$P_{it} < P_{it-1} < P_{it+1} \quad (63)$$

Cioè il livello di pubblicazioni scende nel periodo in cui si fa una ricerca che viene poi brevettata e si innalza dopo il brevetto, anche oltre il livello iniziale.

Questo risultato, che è abbastanza intuitivo sul piano logico, è tuttavia legato ad una serie di ipotesi, che non è detto che si verifichino. Tutti i casi di violazione di queste ipotesi possono essere ricondotte alle analisi compiute in precedenza.

Si noti che la dilazione nelle pubblicazioni comporta un aggiustamento del modello di scelta dello scienziato. Egli infatti sa già che parte dei risultati ottenuti oggi saranno sfruttati in parte, come pubblicazioni, domani e teoricamente può conoscere la parte dei risultati ottenuti vanno dilazionati al momento successivo; dunque la massimizzazione è almeno su un orizzonte biperiodale.

4. Esempi di analisi empirica

4.1 Il database utilizzato

Vediamo dunque il *database* di cui si dispone per l'analisi.

Dei docenti del CESPRI, istituto facente capo all'Università Bocconi di Milano, hanno raccolto i dati relativi agli inventori e ai titolari dei brevetti depositati presso lo European Patent Office (EPO) dal 1978 al primo semestre del 1999: questo è il *database* EPO-INV. Degli inventori è stata identificata l'eventuale istituzione di appartenenza, isolando in particolare quegli inventori che risultano essere docenti universitari, generando quindi il *database* EPO-INV-DOC.

E' poi in corso la costruzione di un altro database relativo alle pubblicazioni dei docenti-inventori e di un campione di docenti non-inventori.

A nostra disposizione sono state messe parti di questi due database: i dati sulle pubblicazioni relativi ai docenti di cinque classi tecnologiche e la parte del *database* EPO-INV-DOC relativa alle università di Bologna e Ferrara. Cioè, abbiamo a disposizione i nomi di tutti i docenti di questi due atenei che risultano inventori di un brevetto depositato presso l'EPO dal 1978 a metà del 1999.

Le informazioni relative alle pubblicazioni di tali docenti sono state tratte invece dal *database* dell'Institute for Scientific Information (ISI), in particolare nella sezione relativa alla Science Citation Index.

4.2 Le indagini effettuate

La relazione che si vuole stimare è quella esistente tra brevetti (variabile indipendente) e pubblicazioni (variabile dipendente) con l'eventuale inserimento di altre variabili rilevanti come variabili "di controllo" (che risultino pertanto correlate non solo con la dipendente ma anche con la indipendente).

I database a disposizione sono due. Il primo ha dati esclusivamente *cross-section*, per cui la relazione stimata è:

$$P_i = \beta_0 + \beta_1 BR_i + \varepsilon_i \quad (24)$$

Probabilmente la formulazione migliore di una tale funzione aggiuntiva è di tipo dicotomico, per cui i brevetti sono zero o uno: infatti in una data unità di tempo un numero di brevetti superiore ad uno difficilmente è dato da più ricerche diverse, ma più verosimilmente è generato dal porre più brevetti su una stessa scoperta.

Il secondo database è costituito da dati longitudinali. Esso consente dunque di controllare per le caratteristiche individuali non osservabili. Si è pertanto confrontata la stima della regressione precedente con la stima della seguente equazione di regressione:

$$P_{it} = \beta_0 + \beta_1 BR_{it} + \sum_{j=1, n-1} \beta_2 D_{ij} + \varepsilon_{it} \quad (45a)$$

Ovvero:

$$P_{it} = \beta_0 + \beta_1 BR_{it} + u_i + \varepsilon_{it} \quad (45b)$$

Inoltre con i dati del secondo *database* si è sottoposta a verifica l'ipotesi che la domanda di ricerca industriale e quindi il numero di brevetti dipenda dal numero di pubblicazioni e di brevetti effettuati in precedenza. La limitatezza dei dati non ha consentito di controllare per gli effetti individuali. Si è dunque stimata la regressione:

$$BR_{it} = \beta_0 + \beta_1 P_{i,t-1} + \beta_2 BR_{i,t-1} + \varepsilon_{it} \quad (64)$$

4.3.A. Analisi di dati cross-section su un campione di medie dimensioni.

La prima analisi condotta è su dati *cross-section* relativi a 180 docenti, appartenenti a 5 settori tecnologici: due nell'ambito della chimica (C07X, Chimica Farmaceutica e C08X, Chimica Farmaceutico-Tecnologico-Applicata), due nell'ambito della bioingegneria (I26A, Bioingegneria meccanica e I26B, Bioingegneria Chimica) ed uno di fisiologia (E04B, Fisiologia Umana). Di questi docenti sono state rilevate le pubblicazioni nell'arco temporale suddetto. Tra i docenti appartenenti a questi settori tecnologici, 100 sono coloro che risultano titolari e/o inventori di almeno un brevetto; quindi all'interno degli stessi settori tecnologici sono stati selezionati altri 80 docenti con caratteristiche simili (per età, università, ecc.) ai 100 inventori. Il dato a disposizione da chi scrive sui brevetti è unicamente se ciascuno di questi 180 docenti sia stato inventore o meno (non si ha cioè il numero dei brevetti).

Il campione individuato risulta così composto:

Tabella 1A – Suddivisione del settore per classe tecnologica

		SETTORE			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	C07X	92	51,1	51,1	51,1
	C08X	55	30,6	30,6	81,7
	I26A	14	7,8	7,8	89,4
	I26B	4	2,2	2,2	91,7
	E04B	15	8,3	8,3	100,0
	Total	180	100,0	100,0	

La distribuzione di inventori e non inventori per settore tecnologico è la seguente:

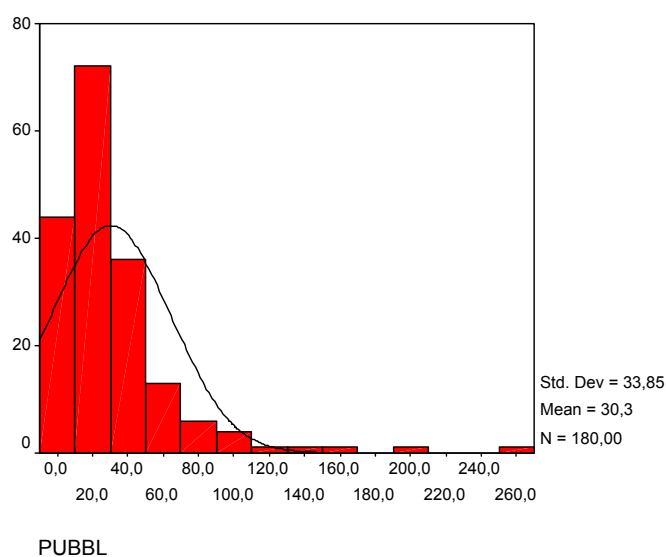
Tavola 2A - Distribuzione di inventori e non inventori per classe tecnologica

DISCIP_5 * BREVETTO Crosstabulation

Count		BREVETTO		Total
		non inventori	inventori	
DISCIP_5	CO7X	39	53	92
	CO8X	25	30	55
	I26A	6	8	14
	I26B	2	2	4
	EO4B	8	7	15
Total		80	100	180

Come è noto, la distribuzione delle pubblicazioni nella popolazione dei docenti tende ad essere fortemente asimmetrica (Allison e Stewart, 1974): pochi docenti hanno un elevato numero di pubblicazioni, la maggioranza dei docenti ha un numero molto inferiori. L'istogramma delle pubblicazioni in questo campione conferma in pieno questa aspettativa.

Tabella 3A – Istogramma delle pubblicazioni



Per valutare la relazione tra pubblicazioni e brevetti è naturalmente necessario “controllare” per quelle variabili che sono in grado di influenzare le pubblicazioni e risultano avere una correlazione coi brevetti. Appare necessario controllare per la classe tecnologica, come dimostra il differente numero medio di pubblicazioni nelle diverse classi tecnologiche.

Tabella 4A – Media delle pubblicazioni per classe tecnologica**PUBBL * SETTORE**

PUBBL

SETTORE	Mean	N	Std. Deviation
CO7X	29,2935	92	25,6342
CO8X	31,4727	55	32,5223
I26A	12,3571	14	9,8341
I26B	21,5000	4	8,1854
EO4B	51,2667	15	73,2525
Total	30,3000	180	33,8477

Raggruppando le classi tecnologiche affini in un unico ambito disciplinare si ha:

Tabella 5A – Media delle pubblicazioni per ambito disciplinare**PUBBL * DISCIP_3**

PUBBL

DISCIP_3	Mean	N	Std. Deviation
chimica	30,1088	147	28,3177
bioingegneria	14,3889	18	10,0536
fisiologia	51,2667	15	73,2525
Total	30,3000	180	33,8477

Il dato più semplice e immediato per confrontare le pubblicazioni di docenti inventori e non inventori è naturalmente il confronto delle medie per i due gruppi.

Un'altra variabile rilevata è il grado accademico dei docenti (distinti tra professore ordinario, professore associato e ricercatore universitario). La tabella seguente mostra la distribuzione di inventori e non inventori per i tre gradi accademici.

Tabella 6A – Distribuzione di inventori e non inventori per grado accademico**GRADO_AC * BREVETTO Crosstabulation**

Count

	GRADO_AC	BREVETTO		Total
		non inventori	inventori	
	ordinario	31	45	76
	associato	29	36	65
	ricercatore	20	19	39
Total		80	100	180

Si nota una più forte propensione dei professori ordinari a brevettare.

Vediamo ora invece se esiste una relazione tra grado accademico e numero di pubblicazioni, condizione necessaria per inserire questa variabile come variabile di controllo.

Tabella 7A – Media delle pubblicazioni per grado accademico

Report

PUBBL			
GRADO_AC	Mean	N	Std. Deviation
ordinario	35,6842	76	29,6995
associato	25,0308	65	37,1761
ricercatore	28,5897	39	34,9230
Total	30,3000	180	33,8477

Anche nel numero di pubblicazioni c'è una superiorità dei professori ordinari, che può naturalmente essere dovuta ad un effetto dell'età. Sorprende in questo senso la media più alta dei ricercatori rispetto agli associati

L'indicazione statistica più immediata della relazione tra pubblicazioni e brevetti è data dal confronto delle medie delle pubblicazioni tra docenti inventori e non inventori.

Tabella 8A – Media delle pubblicazioni per docenti inventori e non inventori

Group Statistics

BREVETTO		N	Mean	Std. Deviation	Std. Error Mean
PUBBL	non inventori	80	24,1750	30,7899	3,4424
	inventori	100	35,2000	35,5019	3,5502

Si nota la media considerevolmente più alta per i docenti inventori. Si può sottoporre a test statistico questa differenza delle medie, in modo da verificare se tale differenza può essere estesa alla popolazione. Si rifiuta l'ipotesi di uguaglianza delle varianze. Il relativo t-test sulle medie porta a rifiutare al 5% di significatività l'ipotesi di uguaglianza nel numero di pubblicazioni tra inventori e non inventori.

La relazione diretta tra pubblicazioni e brevetti è dunque positiva. Impostiamo dunque la regressione, avendo il numero di pubblicazioni come variabile dipendente, la *dummy* inventore/non inventore come esplicativa, assieme alle variabili di controllo relative alla classe tecnologica e al grado accademico. I risultati sono esposti nella tabella seguente, seguita dalle statistiche relative all'R-quadro ed alla tavola dell'ANOVA

Tabelle 9A – Risultati della regressione lineare su dati *cross-section*

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	53,831	9,679		5,561	,000
	BREVETTO	11,309	4,913	,166	2,302	,023
	CO7X	-25,268	9,202	-,374	-2,746	,007
	CO8X	-23,245	9,611	-,317	-2,418	,017
	I26A	-41,367	12,177	-,328	-3,397	,001
	I26B	-33,084	18,407	-,144	-1,797	,074
	PA	-11,051	5,548	-,157	-1,992	,048
	RU	-8,554	6,522	-,104	-1,312	,191

a. Dependent Variable: PUBBL

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,330 ^a	,109	,072	32,5987

a. Predictors: (Constant), RU, I26A, I26B, BREVETTO, CO8X, PA, CO7X

b. Dependent Variable: PUBBL

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22293,489	7	3184,784	2,997	,005
	Residual	182780,311	172	1062,676		
	Total	205073,800	179			

a. Predictors: (Constant), RU, I26A, I26B, BREVETTO, CO8X, PA, CO7X

b. Dependent Variable: PUBBL

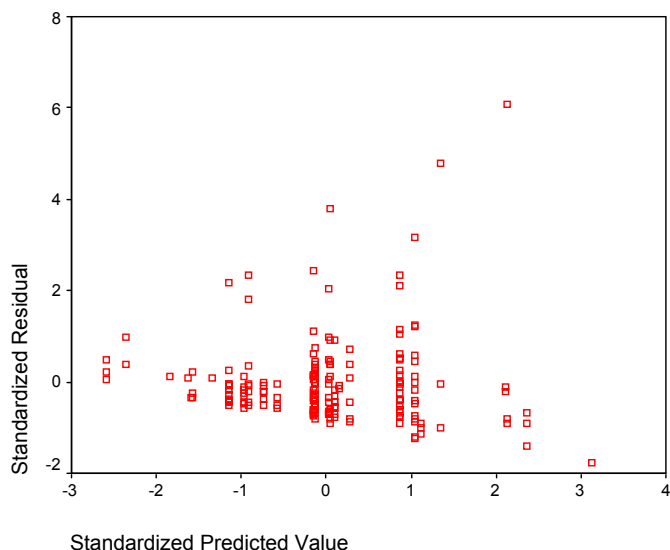
Il coefficiente della *dummy* è dunque positivo e significativo al 5%. Anche le *dummy* relative all'ambito disciplinare e al grado accademico risultano significative.

Ricordiamo il significato del coefficiente della variabile relativa ai brevetti: in dati *cross-section* non si può controllare per le caratteristiche individuali non osservabili: se il numero di pubblicazioni dipende anche da un'abilità individuale non osservabile e se il numero di brevetti effettuati dipende dalle pubblicazioni passate, allora la variabile relativa ai brevetti risulta correlata con tali effetti individuali. Inoltre un arco temporale così lungo non può essere un'unica "unità temporale", nel senso definito al paragrafo precedente: l'aver effettuato o meno un brevetto in questo periodo è in relazione con le pubblicazioni effettuate in questo stesso periodo (se sono corrette le ipotesi sulla domanda di ricerca industriale); dunque c'è endogenità, ovvero correlazione tra la variabile relativa ai brevetti e i residui. Di conseguenza il parametro della variabile relativa ai

brevetti soffre di un *bias* e non dice se per ciascun individuo fare più brevetti comporta avere più pubblicazioni.

La presenza di tale correlazione può essere alla base dell'eteroschedasticità dei residui, di cui si mostra il grafico.

Tabella 10 A – “Scatter” dei residui standardizzati e dei valori stimati standardizzati



Il test di Cook - Weisberg conferma la significatività all'1% dell'eteroschedasticità dei residui.

In conclusione, dato il segno positivo del coefficiente della variabile relativa ai brevetti, non è possibile comprendere se l'averne un brevetto comporta un effetto benefico sulle pubblicazioni o se chi brevetta ha più pubblicazioni di chi non lo fa perché si tratta di individui più abili e quindi anche più richiesti dall'industria per collaborazioni di ricerca, o magari semplicemente di individui più “fortunati”, che in una fase della loro carriera accademica hanno trovato risultati brevettabili, pur non svolgendo ricerca industriale in quantità più alta della media.

In ogni caso non è risultato trascurabile il verificare che, anche a parità di settore tecnologico, di grado accademico, a brevettare non siano i docenti che pubblicano di meno.

4.3.B Analisi di un piccolo campione con dati longitudinali

A fini esplorativi ed esemplificativi, si è costruito un piccolo “sottocampione” a partire dai due *database* su pubblicazioni e brevetti. Si sono considerati due dipartimenti dell'Università di Bologna in cui è stato particolarmente elevato, in proporzione sul totale, il numero di docenti che hanno brevettato: si tratta del Dipartimento di Chimica applicata e scienza dei materiali e del Dipartimento di Chimica industriale e dei materiali. All'interno di questi due dipartimenti, sono stati selezionati i professori ordinari e gli associati. Si può dunque dire che la popolazione a cui si fa riferimento è quella dei professori ordinari e associati, che si può presumere avere un comportamento differente, rispetto al tema in analisi, dalla popolazione dei ricercatori (l'attività accademica, per questi ultimi, non può essere considerata “a regime” e quindi sia il rapporto con il mondo industriale sia la continuità nell'attività di pubblicazione può considerarsi ancora in una fase iniziale e perciò soffrire di “discontinuità”). Se invece il riferimento è al personale docente dell'università nel suo complesso (inclusi i ricercatori) l'esclusione dal campione dei ricercatori naturalmente distorce il campione stesso però, nel nostro caso, elimina un problema. Infatti, avendo a disposizione unicamente la composizione attuale dei dipartimenti suddetti, c'è il problema della

discrepanza temporale fra il momento a cui l'analisi fa riferimento (1999) e appunto il momento in cui si hanno informazioni sulla composizione della popolazione (2003). In tal modo c'è un duplice rischio: escludere dal campione dei soggetti che facevano parte della popolazione in analisi ed includerne altri che non ne facevano parte; l'inclusione del campione dei soli professori ordinari e associati, escludendo i ricercatori, riduce al minimo il secondo tipo di rischio (è ben difficile che un individuo che oggi è professore non facesse parte del corpo docente quattro anni fa).

Il numero dei professori, associati e ordinari, nei due dipartimenti e dunque il numero di soggetti inclusi nel campione è pari a 24.

Valgono dunque tutti i *caveat* del caso, derivanti dalla ridotta dimensione del campione (che si pone sulla soglia di definizione del *large sample*), e dai limiti prima enunciati: la non casualità nella scelta del campione, sia da un punto di vista geografico, sia dal punto di vista dell'appartenenza settoriale.

Per la costruzione di *dati panel* è necessario disporre di dati relativi a più soggetti ed a più momenti temporali. Dalla descrizione del *database* abbiamo visto che si dispone per il momento di dati relativi a 24 professori universitari, divisi in 2 dipartimenti e in 2 gradi accademici. L'arco di tempo per cui si dispone di dati relativi ad entrambe le variabili (pubblicazioni e brevetti) è 1990-1998.

La costruzione del *panel* è *in fieri*, in particolare per quanto riguarda le pubblicazioni. Per il momento si dispone soltanto di una suddivisione temporale "asimmetrica", con due periodi, cioè, di durata disuguale e quindi, relativamente al primo dei due sottoperiodi, anche troppo vasta: per il momento, infatti, si hanno dati relativi ai sottoperiodi 1990-1995 e 1996-1997.

Ci sono dunque diverse ragioni che rendono il campione utilizzato in queste prime indagini insoddisfacente dal punto di vista della rappresentatività e del rigore statistico. Pertanto i risultati che si presentano in seguito sono da considerarsi puramente indicativi ed hanno il principale scopo di mostrare quali siano le indagini che si intendono condurre quando la raccolta dei dati avrà raggiunto uno stato soddisfacente.

Prima di procedere all'analisi dei risultati ottenuti ci sembra opportuno effettuare alcune considerazioni.

Per quanto riguarda le pubblicazioni, nel modello teorico si è sempre volutamente mantenuta un'ambivalenza di accento su quantità e qualità. Il vincolo temporale può infatti portare, dedicando meno tempo alla ricerca accademica, a diminuire non il numero *tout-court* delle pubblicazioni, ma la loro qualità. Similmente, se le pubblicazioni costituiscono un segnale della qualità del docente, più che il loro numero dovrebbe contare la loro qualità. In effetti se il riferimento è al numero di pubblicazioni internazionali, in qualche modo il riferimento è contemporaneamente sia alla qualità che alla quantità, in quanto solo le pubblicazioni migliori riescono ad arrivare alla pubblicazione su riviste internazionali. Tuttavia si è comunque creato un indicatore di qualità, basato sulle citazioni ricevute dagli articoli: la media delle citazioni ottenute dai 5 articoli più citati di ciascun docente; si è altresì considerato il numero di citazioni ottenute dall'articolo più citato di ciascun docente.¹⁴

Un grosso problema riguarda la dimensione temporale delle stime.

Il modello fa riferimento in astratto a dei periodi, senza precisarne la durata; un'analisi empirica ha invece bisogno di una precisa dimensione temporale. Solo un'estesa osservazione empirica, unita possibilmente ad interviste dirette, può aiutare a far luce sulle dimensioni temporali.

Inoltre, il modello teorico fa riferimento al tempo dedicato alla ricerca industriale, mentre i dati sono relativi al momento di deposito del brevetto: non solo i due momenti non coincidono, ma anche la distanza temporale tra i due momenti non è costante. Anche quando si fa riferimento alle pubblicazioni, in realtà la visuale è duplice: da un lato esse sono viste come segnalazione di abilità e quindi conta effettivamente il momento in cui l'articolo è pubblicato; dall'altro, nel momento in cui

¹⁴ Va comunque detto che la correlazione tra queste tre variabili, e soprattutto tra le prime due, è molto elevata, significativa all'1%.

si ipotizza un impatto della ricerca industriale e accademica sulle pubblicazioni, si fa riferimento in realtà al momento di stesura dell'articolo/monografia, che può anche essere distante da quello di pubblicazione.

Il modello ci dice dunque che:

-la ricerca industriale svolta nel periodo t è in relazione con le pubblicazioni e la ricerca industriale svolta fino al periodo t ;

-le pubblicazioni effettuate nel periodo t sono in relazione con la ricerca industriale svolta nel periodo t .

Il primo problema empirico è: quanto è ampio il periodo t ? E quanto il lasso temporale "fino al periodo t "?

Data una risposta, anche convenzionale, a questo problema, ci sono i problemi ulteriori; ad esempio supponiamo che si voglia stimare la relazione tra brevetti e pubblicazioni effettuate nell'arco di tempo 1990-95. Si può supporre che la ricerca industriale e la stesura dell'articolo avvengano nel medesimo anno del deposito del brevetto e della pubblicazione dell'articolo. Dunque si prenderanno in considerazione brevetti e pubblicazioni nell'arco temporale 1990-95; una fonte di errore potrebbe essere data da un'invenzione industriale avvenuta di fatto nel 1995 ma il cui brevetto viene depositato nel 1996, mentre l'articolo scaturito da questa ricerca viene pubblicato nel 1995: nella stima comparirà la pubblicazione ma non la ricerca industriale da cui essa deriva. Oppure un brevetto depositato nel 1990 è in realtà relativo ad una ricerca industriale effettuata nel 1989 che produce un articolo nello stesso anno: nella stima verrà esclusa la pubblicazione ma non la ricerca industriale.

Oppure: si vuole stimare la relazione tra la ricerca industriale avvenuta nel 1996 e le pubblicazioni e la ricerca industriale avvenute fino a quel momento. Per farlo si considereranno i brevetti depositati nel 1996 e le pubblicazioni e i brevetti depositati fino al 1995. Ma una collaborazione di ricerca industriale avvenuta nel 1996 potrebbe generare un brevetto depositato nel 1997: nella stima compariranno dunque le pubblicazioni e i brevetti determinanti nel far nascere il rapporto di collaborazione, ma non il rapporto che è nato.

Vi è inoltre un problema che è non solo empirico ma anche teorico: nel modello si ipotizza che:

a) la ricerca industriale in un dato periodo dipenda dalle pubblicazioni e dalla ricerca industriale del solo periodo precedente.

In realtà altre ipotesi sono ammissibili:

b) la ricerca industriale effettuata in un dato periodo dipende da tutte le pubblicazioni e da tutta la ricerca industriale effettuate fino a quel periodo;

c) la ricerca industriale effettuata in un dato periodo dipende da tutte le pubblicazioni e da tutta la ricerca industriale effettuate fino a quel periodo, ma con un peso decrescente all'aumentare della distanza temporale (le pubblicazioni e le collaborazioni industriali più recenti "pesano" di più delle pubblicazioni e delle ricerche passate).

Le tre ipotesi coincidono se il modello è biperiodale (periodo presente e periodo passato), non coincidono se il modello è, più realisticamente, multiperiodale.

Un modello multiperiodale in cui si facciano le ipotesi b) e c) genera equazioni più complesse rispetto a quelle che abbiamo individuato, ma non cambiano i suoi risultati essenziali.

La natura del *database* dell'ISI sulle pubblicazioni, più accurato riguardo alle pubblicazioni più recenti, ci fa adottare naturalmente l'ipotesi c) anche se le equazioni stimate sono quelle derivanti dall'ipotesi a).

Il tipo di problemi empirici che si ha di fronte fanno comprendere bene che gli errori saranno tanto più ridotti quanto più ampi sono i periodi considerati. Infatti gli errori possibili riguardano soprattutto gli estremi del periodo (come l'esempio visto: considerato l'arco di tempo '90-'95 "sfugge" all'osservazione una collaborazione effettuata nel '95 che produce un brevetto depositato nel '96; oppure viene inserita impropriamente una ricerca industriale effettuata nell'89

ma non la pubblicazione che da essa deriva; è improbabile, invece, che “sfugga” una collaborazione effettuata in un anno intermedio o che negli anni intermedi siano inserite ricerche effettuate al di fuori del periodo: una collaborazione industriale effettuata nel '93 difficilmente produrrà un brevetto oltre il '95, oppure un brevetto depositato nel '93 difficilmente si riferirà ad una ricerca effettuata nell'89).

Il problema è che, se l'unità di tempo “vera” è un periodo limitato considerare un arco di tempo vasto, che risulta quindi essere la sommatoria di più periodi, porta a stime in cui c'è una correlazione tra variabili esplicative ed errori, causando quindi stime OLS *biased*¹⁵.

Nell'indagine sulle relazioni tra brevetti e pubblicazioni partiamo da dati di statistica descrittiva. Vediamo innanzitutto quanti docenti risultano inventori in tutto l'arco temporale considerato.

Tabella 1B- Frequenze degli inventori

BREV9097

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	15	62.5	62.5	62.5
	1.00	3	12.5	12.5	75.0
	2.00	2	8.3	8.3	83.3
	3.00	1	4.2	4.2	87.5
	4.00	1	4.2	4.2	91.7
	5.00	2	8.3	8.3	100.0
Total		24	100.0	100.0	

Dalla Tabella 1B si evince che, dei 24 soggetti considerati, 15 non risultano mai inventori nell'arco temporale 90-97; degli altri 9 che risultano inventori almeno una volta è interessante notare che ben 6 risultano inventori più di una volta: già questo dato fornisce un'indicazione sul fatto che chi collabora con l'industria tende poi a rifarlo.

Considerare l'intero arco temporale 1990-97 come un unico periodo vuole dire avere a disposizione dei dati *cross-section*, relativi cioè a più individui in un solo momento temporale.

Con questi dati a disposizione possiamo rispondere alla domanda posta dall'equazione (24): docenti che brevettano di più hanno anche più pubblicazioni o c'è invece una relazione negativa o nulla?

Conduciamo questa analisi dapprima ad un livello elementare distinguendo semplicemente tra chi risulta inventore e chi no nel periodo considerato, a prescindere dal numero di invenzioni.

Creiamo dunque una variabile *dummy* che distingua chi risulta inventore di almeno un brevetto da chi non risulta inventore di nessun brevetto (variabile BR9097_D) e confrontiamo le medie delle pubblicazioni (variabile PUB_9097) e degli “indici di qualità” (rispettivamente CIT5_97 e CIT1_97) delle pubblicazioni tra i due gruppi.

¹⁵ Si veda in proposito il par. 3.2.f) e relativa nota per la definizione di unità di tempo in questo contesto e per il possibile *bias* derivante da endogenità.

Tabella 2B - Quantità e qualità delle pubblicazioni di inventori e non inventori**Group Statistics**

	BR9097_D	N	Mean	Std. Deviation	Std. Error Mean
PUB_9097	inventori	9	19,1111	21,6590	7,2197
	non inventori	15	10,1333	8,8388	2,2822
CIT5_97	inventori	9	22,0667	20,7709	6,9236
	non inventori	15	14,4400	19,3461	4,9951
CIT1_97	inventori	9	36,6667	34,1980	11,3993
	non inventori	15	34,7333	42,7559	11,0395

Si vede che gli “inventori” hanno una media di numero di pubblicazioni e di citazioni più alta rispetto ai non inventori, anche se la differenza è minima per quel che riguarda la citazione più elevata.

Tuttavia questa differenza può essere dovuta esclusivamente al campionamento e non rispecchiare una differenza reale nella popolazione. A tale scopo è necessario effettuare un test di significatività delle differenze delle medie. Più precisamente si sottopone a test-t l’ipotesi che nella popolazione in realtà i due gruppi (inventori e non inventori) abbiano media uguale. Ebbene, anche effettuando il test in cui è respinta l’ipotesi di varianza costante nei due gruppi (ipotesi opportunamente testata), l’ipotesi nulla di uguaglianza delle medie non può essere rigettata nemmeno ponendo una soglia di significatività del 10% per nessuna delle due variabili.

Nella precedente indagine abbiamo verificato se vi sono delle caratteristiche osservabili dei docenti che influenzano la quantità di collaborazioni industriali. Le caratteristiche che possiamo “osservare” con il *database* a nostra disposizione sono il grado accademico (professore ordinario o associato) e il dipartimento (chimica applicata e scienza dei materiali - chimica industriale e dei materiali).

Dunque la relazione tra brevetti e pubblicazioni andrebbe condotta a parità di queste caratteristiche, il che, in termini di analisi di medie, vuol dire condurre l’analisi in ciascun dipartimento e per ciascun grado accademico; bisogna verificare, cioè, se in ciascun dipartimento e per ciascun grado esiste una differenza nella media delle pubblicazioni degli inventori e dei non-inventori. Ovviamente la suddivisione ha senso se risulta esserci una correlazione tra queste caratteristiche e il numero sia di pubblicazioni che di invenzioni, cioè se esiste una differenza significativa nella quantità e qualità delle pubblicazioni e nel numero degli inventori nei due dipartimenti e per i due gradi accademici.

Iniziamo l’analisi dal dipartimento.

La tabella 3B mostra la media delle pubblicazioni, delle prime 5 citazioni e della prima citazione nei due dipartimenti: si vede che la media è considerevolmente più alta nel Dipartimento di Chimica industriale e dei materiali rispetto al Dipartimento di Chimica applicata e scienza dei materiali. Il t-test sulle medie (non mostrato) ci dice che questa differenza delle medie è significativa per la popolazione al 10% per quel che riguarda le pubblicazioni, al 5% per entrambe le variabili relative alle citazioni.

Tabella 3B - Quantità e qualità delle pubblicazioni per dipartimento**Group Statistics**

DIPARTIM		N	Mean	Std. Deviation	Std. Error Mean
PUB_9097	chimica applicata	17	7,9412	4,3080	1,0448
	chimica industriale	7	27,0000	23,1012	8,7314
CIT5_97	chimica applicata	17	9,3412	6,2140	1,5071
	chimica industriale	7	36,6286	28,0137	10,5882
CIT1_97	chimica applicata	17	19,6471	10,5472	2,5581
	chimica industriale	7	73,8571	55,7059	21,0548

Vediamo ora cosa accade per quel che riguarda la distinzione per grado accademico. La tabella 4B mostra la media delle pubblicazioni e delle prime citazioni per i professori ordinari e associati. Si nota che le medie sono molto vicine e i t-test confermano che non esistono differenze in media tra la popolazione.

Tabella 4B - Quantità e qualità delle pubblicazioni per grado accademico**Group Statistics**

GRADO_AC		N	Mean	Std. Deviation	Std. Error Mean
PUB_9097	ordinario	14	14,9286	18,4160	4,9219
	associato	10	11,5000	9,5598	3,0231
CIT5_97	ordinario	14	16,5429	19,8741	5,3116
	associato	10	18,3600	20,7214	6,5527
CIT1_97	ordinario	14	36,5714	43,4648	11,6165
	associato	10	33,9000	33,9426	10,7336

Sembra dunque di poter concludere che la variabile dipartimento ha influenza su quantità e qualità delle pubblicazioni; non così per il grado accademico.

Analizziamo ora l'altra relazione, quella tra queste due variabili e le invenzioni.

La tabella 5B mostra la suddivisione degli "inventori" nei due dipartimenti: nel Dipartimento di Chimica industriale 3 professori su 7 (il 42%) sono stati inventori almeno una volta (la media di invenzioni a testa è 1,7), nel Dipartimento di Chimica applicata 6 su 17 (il 35%) sono risultati inventori (per una media di 0,7 invenzioni a testa). Si nota dunque che nel dipartimento dove è più alta la quantità e la qualità delle pubblicazioni (chimica industriale e dei materiali) è più alta la propensione alla collaborazione industriale.

Tabella 5B - Frequenze delle invenzioni per dipartimento**BREV9097 * DIPARTIM Crosstabulation**

Count

	DIPARTIM		Total
	chimica industriale	chimica applicata	
BREV9097 ,00	4	11	15
1,00		3	3
2,00	1	1	2
3,00		1	1
4,00		1	1
5,00	2		2
Total	7	17	24

La tabella 6B mostra lo stesso tipo di indagine condotto per grado accademico. La differenza è modesta: tra gli ordinari 9 su 14 (il 64,2%) non risultano mai inventori (il numero medio di invenzioni a testa tra gli ordinari è 0,9), tra gli associati 6 su 10 (il 60%) non sono mai inventori (media di 1,1 invenzioni a testa).

Tabella 6B - Frequenze delle invenzioni per grado accademico**BREV9097 * GRADO_AC Crosstabulation**

Count

	GRADO_AC		Total
	associato	ordinario	
BREV9097 ,00	6	9	15
1,00	1	2	3
2,00	1	1	2
3,00	1		1
4,00		1	1
5,00	1	1	2
Total	10	14	24

In sintesi: il grado accademico non sembra avere influenza né su quantità e qualità delle pubblicazioni né sul numero di “invenzioni”, mentre sembra significativa l’influenza del dipartimento sia sulle pubblicazioni sia sulle invenzioni.

Ne consegue che, volendo verificare se c’è differenza in media nelle pubblicazioni tra inventori e non-inventori bisogna “controllare” per il dipartimento, cioè condurre questa analisi in ciascuno dei dipartimenti.

Dunque ripetiamo l’analisi condotta in principio (confronto tra la media delle pubblicazioni degli “inventori” e dei “non inventori”) in ciascuno dei due dipartimenti.

I risultati nei due dipartimenti sono abbastanza difforni. La tabella 7B mostra ciò che accade nel dipartimento più grande, quello di Chimica applicata e scienza dei materiali: qui avviene quanto accade a livello aggregato, cioè gli “inventori” hanno valori medi più alti per tutte e tre le variabili relative alle pubblicazioni. Il t-test in questo caso ci porta a rifiutare l’ipotesi nulla (con significatività rispettivamente del 10% e del 5%) che le medie siano uguali per la popolazione, relativamente sia al numero totale di pubblicazioni sia alla media delle prime 5 pubblicazioni.

Nel dipartimento più piccolo, quello di Chimica industriale e dei materiali (si veda la tabella 8B), invece, i risultati sono dissimili per le tre variabili (superiorità del numero di pubblicazioni per gli inventori, media quasi identica per le prime 5 citazioni, superiorità dei non inventori per l’articolo più citato). Il t-test suggerisce che nella popolazione non ci sono differenze in media per nessuna delle tre variabili.

Tabella 7B - Dipartimento di chimica applicata e scienza dei materiali
Quantità e qualità delle pubblicazioni di inventori e non inventori

Group Statistics

	BR9097_D	N	Mean	Std. Deviation	Std. Error Mean
PUB_9097	inventori	6	10,3333	4,7610	1,9437
	non inventori	11	6,6364	3,6131	1,0894
CIT5_97	inventori	6	14,7667	5,4235	2,2141
	non inventori	11	6,3818	4,4474	1,3410
CIT1_97	inventori	6	26,0000	3,2863	1,3416
	non inventori	11	16,1818	11,6260	3,5054

Tabella 8B - Dipartimento di chimica industriale e dei materiali
Quantità e qualità delle pubblicazioni per inventori e non inventori

Group Statistics

	BR9097_D	N	Mean	Std. Deviation	Std. Error Mean
PUB_9097	inventori	3	36,6667	33,5609	19,3764
	non inventori	4	19,7500	12,3659	6,1830
CIT5_97	inventori	3	36,6667	34,2429	19,7701
	non inventori	4	36,6000	28,0680	14,0340
CIT1_97	inventori	3	58,0000	60,2246	34,7707
	non inventori	4	85,7500	57,8641	28,9320

Naturalmente lo strumento sintetico per analizzare la relazione tra collaborazioni industriali e pubblicazioni, “controllando” per il dipartimento e il grado è rappresentato dalla regressione lineare.

Il confronto tra pubblicazioni degli inventori e dei non-inventori, illustrato nella tabella 2B, è ugualmente ottenibile tramite una regressione avente come variabile dipendente una variabile relativa alle pubblicazioni e come indipendente una variabile *dummy* avente valore 0 per coloro che non sono mai risultati inventori di un brevetto e valore 1 per coloro che lo siano stati almeno una volta. Il segno di quest’ultima variabile risulterà positivo ma non significativo (gli “inventori” pubblicano di più ma non si può inferire questo effetto per la popolazione); si noti naturalmente che il valore del test t è identico a quello ottenuto col test sull’uguaglianza delle medie. Vediamo in

dettaglio nella tabella 9B. Naturalmente un discorso analogo vale per le variabili “qualitative” sulle pubblicazioni.

Tabella 9B - Regressione 1: effetto dei brevetti sulle pubblicazioni 1990-97
(stime OLS su dati cross-section)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	10,133	3,832		2,644	,015	2,186	18,08
	BR9097_D	8,978	6,258	,292	1,435	,165	-4,001	21,95

a. Dependent Variable: PUB_9097

N. Osservazioni: 24

R-Squared: 0,086

Adj. R-Squared: 0,044

Se questa è una semplice replica dell’analisi della differenza delle medie, la regressione consente, appunto, di “controllare” per la variabile “dipartimento” e la variabile “grado accademico”. In realtà, come già aveva fatto sospettare l’analisi delle medie, quest’ultima variabile risulta non significativa, mentre lo è la variabile relativa al dipartimento: si nota infatti che la sua introduzione aumenta molto l’R-quadro (mentre aggiungendo ancora la variabile relativa al grado accademico l’R-quadro corretto scende), modifica, riducendolo, il valore del coefficiente della “dummy” delle invenzioni, ma ne aumenta la significatività, sia pure non fino a superare la “soglia” minima standard del 10% Vediamo il risultato di questa regressione nella tabella 10 B.

Tabella 10 B - Regressione 2: effetto dei brevetti sulle pubblicazioni 1990-97
(stime OLS su dati cross-section)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	23,680	5,144		4,603	,000	12,983	34,37
	BR9097_D	7,746	5,188	,252	1,493	,150	-3,042	18,53
	DIPARTIM	-18,473	5,526	-,565	-3,343	,003	-29,964	-6,98

a. Dependent Variable: PUB_9097

N. Osservazioni: 24

F (2, 21) = 7,093

Prob.>F = 0,004

R-Squared: 0,449

Adj. R-Squared: 0,396

Anche per le variabili che indicano la qualità delle pubblicazioni, inserire il “controllo” per il dipartimento non porta alla significatività della variabile relativa alle invenzioni: non si può dire che per la popolazione gli inventori producano articoli più citati rispetto ai non inventori.

Questo tipo di analisi perde però un’informazione fondamentale: distinguendo solo tra inventori e non-inventori non distingue tra questi chi ha inventato di più e chi meno; invece dalla tabella 1 si evince che ben 1/4 dei soggetti del campione (6 su 24) risultano inventori per più di un brevetto. Ovviamente, se il numero di brevetti vuole essere una *proxy* della quantità di ricerca industriale condotta, bisogna avvertire che più brevetti possono scaturire da una stessa ricerca. Tuttavia rimane la considerazione che “appiattare” gli inventori assegnando a tutti lo stesso valore rischia di essere fuorviante.

Pertanto l’operazione più corretta è regredire la variabile relativa alle pubblicazione sul numero di invenzioni (che diventa quindi una variabile continua), controllando per il dipartimento.

I risultati di questa regressione sono esposti nella tabella 11 B.

Tabella 11 B - Regressione 3: effetto dei brevetti sulle pubblicazioni 1990-97
(stime OLS su dati cross-section)

		Coefficients ^a						
		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	22,193	5,315		4,175	,000	11,139	33,24
	BREV9097	2,804	1,597	,303	1,756	,094	-,518	6,12
	DIPARTIM	-16,231	5,648	-,496	-2,874	,009	-27,976	-4,48

a. Dependent Variable: PUB_9097

N. Osservazioni: 24

F (2, 21) = 7,739

Prob.>F = 0,003

R-Squared: 0,424

Adj. R-Squared: 0,369

Si noti il differente risultato rispetto alla regressione precedente: la variabile relativa ai brevetti ha coefficiente sempre positivo ma qui è significativo al 10%: l’esito è che, a parità di dipartimento, gli individui che risultano inventori di più brevetti sono anche coloro che pubblicano di più.

L’effetto sulla quantità delle pubblicazioni è pari a quello sulla loro qualità: conducendo la stessa regressione avente la media dei 5 articoli più citati come dipendente vede un coefficiente per la variabile relativa alle invenzioni positivo ma con livello di significatività più basso.

I risultati esposti nella tabella 11 B vanno però considerati con molta cautela. Infatti un’analisi dei residui ci dice che essi sono affetti sia da eteroschedasticità sia da non-normalità (non si accettano al 5% le ipotesi di omoschedasticità e di normalità dei residui). Ciò comporta forti dubbi sugli *standard error* ottenuti dei parametri e quindi sulla loro significatività. Il problema della non-normalità si può aggirare con le proprietà asintotiche delle stime OLS, supponendo quindi che il campione qui considerato sia un *large sample*, il che è un po’ forzato data la numerosità di 24. Il problema dell’eteroschedasticità si aggira con stime GLS “robuste” rispetto a questo problema: ebbene, tali stime danno per la variabile relativa alle invenzioni un coefficiente positivo ma non

significativo (p-value: 0.364; rimane invece significativo il coefficiente relativo alla variabile dipartimento). Rimane dunque incerta la risposta alla domanda che ci si è posti: a parità di caratteristiche osservabili, i docenti con più pubblicazioni sono anche inventori di più brevetti? La risposta è positiva relativamente ai dati del campione, ma l'inferenza riguardo alla popolazione è incerta.

E' possibile condurre questo tipo di analisi, però soltanto ritornando alla divisione "secca" tra inventori e non inventori, utilizzando tecniche statistiche non parametriche, che non facciano cioè assunzioni sulla forma della distribuzione della popolazione. Ci riferiamo in particolare alla *combinazione non parametrica di test dipendenti*.¹⁶

Si è dunque sottoposta a test l'ipotesi che il numero medio di pubblicazioni degli "inventori" sia pari a quello dei non inventori contro l'ipotesi di differenza in media; ebbene, l'ipotesi di uguaglianza della medie non può essere rifiutata se la soglia di significatività è del 10%, anche se da questa soglia non si è moto distanti (i p-value che si ottengono con prove ripetute si collocano attorno a 0,16 avendo come test parziale il t-test e attorno a 0,17 avendo come test parziale il "difference of means").

Sottoponendo a identico test l'analoga ipotesi riguardante le prime cinque pubblicazioni, si rimane invece parecchio lontani dalla soglia di significatività del 10% (p-value attorno a 0,4 con entrambi i test parziali) e dunque decisamente non si può rifiutare l'ipotesi di uguaglianza delle medie.

I risultati ottenuti con questa tecnica sono dunque molto simili a quelli ottenuti coi test t: entrambi ci inducono a non rifiutare l'ipotesi di uguaglianza delle medie delle pubblicazioni e prime cinque citazioni tra "inventori" e "non inventori", con forte evidenza per le citazioni, con minore sicurezza nel caso delle pubblicazioni.

Questo tipo di tecnica consente poi di effettuare direttamente un'analisi che, nel campo delle tecniche parametriche è possibile soltanto tramite una regressione, cioè confrontare la media di due gruppi tenendo conto di ulteriori variabili che differenziano il campione.

Conducendo un'analisi che tenga conto della differenziazione per dipartimento non si può rifiutare al 10% l'ipotesi di uguaglianza delle medie di pubblicazioni tra inventori e non inventori (il p-value si colloca, a seconda delle differenti funzioni di combinazione e dei differenti test parziali, tra 0,12 e 0,18). Si conferma invece la differenza in media nel Dipartimento di Chimica applicata e scienza dei materiali e l'uguaglianza in media nel Dipartimento di Chimica industriale e dei materiali.

In sintesi, le osservazioni campionarie indicano che i docenti che maggiormente collaborano con l'industria sono in media quelli che hanno più pubblicazioni e i loro articoli sono mediamente più citati, anche a parità di caratteristiche osservabili rilevanti, quali il dipartimento di appartenenza.

Tuttavia, applicando i criteri rigorosi dell'induzione statistica, non sembra che si possano inferire con soddisfacente grado di probabilità questi risultati rispetto alla popolazione.

Ricordiamo che, a causa della possibile presenza di caratteristiche individuali non osservabili, è una cosa diversa domandarsi se, mediamente, per un singolo docente fare più ricerca industriale comporta avere più o meno pubblicazioni. Per rispondere a questa domanda è necessario stimare la (45) e per farlo è necessario disporre di *time series* o di *dati panel*. A questo scopo, come detto, si è suddiviso l'arco di tempo considerato (1990-97) in due sottoperiodi: 1990-95 e 96-97. Una delle ipotesi alla base del modello è che i parametri siano costanti nel tempo: è un'ipotesi in parte logica in parte operativa che, se violata, porta a conseguenze piuttosto diverse da quelle illustrate.

Disponendo di due sottoperiodi e di 24 soggetti si hanno quindi 48 osservazioni: vediamo cosa accade se si conduce una regressione cosiddetta *pooled*, in cui si trattano cioè tutte le

¹⁶ Il pacchetto statistico utilizzato per effettuare queste stime è NPC.

osservazioni come se si trattasse di 48 individui distinti. Rispetto alla cross-section che considera il periodo 90-97 come un tutt'uno si ha una riduzione dello *standard error* della variabile relativa ai brevetti (qui definita *brev_2per*, mentre la dipendente è definita *pubb_2per*) ed un coefficiente, positivo, che sfiora la soglia di significatività del 5%. I risultati sono esposti nella tabella 12 B. Anche qui però va precisato che ci sono sia problemi di eteroschedasticità sia di non-normalità dei residui e che stime GLS “robuste” danno un coefficiente della variabile brevetti positivo ma non significativo (p-value: 0.203; rimane invece significativo al 5% il coefficiente della variabile dipartimento).

Tabella 12 B - Regressione 4: effetto dei brevetti sulle pubblicazioni 1990-97
(stime OLS “pooled” su dati panel: sottoperiodi 1990-95 e 1996-97)

pubb_2per	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
brev_2per	1.894086	.9661895	1.96	0.056	-0.05192	3.840091
dipart	-8.574411	2.080934	-4.12	0.000	-12.76563	-4.383194
cons	11.8765	1.89342	6.27	0.000	8.06295	15.69004

N. Osservazioni: 48

F(2, 45) = 13.01

Prob>F = 0.0000

R-squared = 0.3665

Adj R-squared = 0.3383

L'interesse di questa stima sta nella possibilità di effettuare un confronto immediato e diretto con quella che tiene conto degli effetti individuali non osservabili. In questa, cioè, le 48 osservazioni non vengono, per così dire, poste sullo stesso piano ma si considerano le “particolarità” individuali, si ipotizza cioè che per tutti gli individui valgano le stesse relazioni tra le variabili ma che ciascun individuo si trovi ad un’ “altezza” diversa, cioè, appunto secondo quanto ipotizzato dal nostro modello, individui aventi lo stessa quantità di ricerca industriale, appartenenti allo stesso dipartimento possono avere livelli diversi di pubblicazioni a causa della diversa “abilità” individuale.

Quella che si effettua è dunque la stima della (45), dove il coefficiente della variabile relativa ai brevetti di cui si è inventori dice per ciascun professore l'effetto di un aumento della ricerca industriale. Se si ipotizza che questo aumento può essere indotto da misure di *policy* o da un mutamento di clima culturale che modifichino dunque le preferenze dei singoli, si comprende l'importanza di questa stima.

Dal punto di vista della stima econometrica rimane il problema se utilizzare un modello “a effetti fissi” (stimatore LSDV) o uno “a effetti casuali” (stimatore GLS): in quest'ultimo gli scostamenti individuali sono supposti distribuirsi in maniera casuale.

In caso di non correlazione tra regressori ed effetti individuali lo stimatore GLS è preferibile, in quanto dà stime corrette ed efficienti (mentre lo stimatore LSDV dà stime corrette ma inefficienti). Viceversa, se c'è correlazione tra effetti individuali e regressori è senz'altro da preferirsi lo stimatore LSDV, che dà stime corrette ed efficienti, mentre le stime GLS sono *biased*.

Questi stimatori vanno a loro volta confrontati con lo stimatore OLS *pooled*, preferibile quando le caratteristiche individuali non sono significative.

Oltre all'esistenza ed alla correlazione degli effetti individuali altri possono essere i criteri di scelta. Alcune ragioni fanno propendere, almeno con il campione di cui si dispone, per la stima “random effect”: è noto, infatti, che il modello “random effect” è preferibile quando la dimensione

temporale è ridotta rispetto al numero di individui (come è in questo caso, essendovi solo 2 periodi) e se l'obiettivo è fare inferenza sulla popolazione, piuttosto che fare inferenza condizionata ai soli individui del campione: l'obiettivo di questo studio è chiaramente il primo. Inoltre la stima "a effetti fissi" tiene conto della sola variabilità "within", cioè per ciascun individuo, tra un periodo e l'altro, mentre non tiene in considerazione la variabilità "between", cioè tra individui, che però, nel campione considerato, è la fonte di variabilità largamente preponderante: infatti la capacità esplicativa della stima "a effetti fissi" risulta molto bassa¹⁷.

Il test di Hausman dà rigore statistico a queste considerazioni: sotto l'ipotesi nulla di non correlazione tra effetti casuali e regressori del modello, la stima "random effect" risulta preferibile perché quella "fixed effect" risulta meno efficiente. Ebbene, nel campione considerato tale ipotesi nulla non può essere rifiutata ($\text{Prob} > \chi^2 = 0.1222$).

Naturalmente anche l'ipotesi che un modello con effetti individuali sia preferibile ad uno privo di questi effetti va sottoposta a verifica. Una volta accertato che il modello a effetti casuali è preferibile rispetto a quello a effetti fissi, il test statistico appropriato per verificare la preferibilità del modello "random effect" rispetto a quello "pooled" è il test di Breusch-Pagan; ebbene, l'ipotesi nulla per cui gli effetti individuali collassano nel modello "pooled" è rifiutata ($\text{Prob} > \chi^2 = 0.0006$).¹⁸

Il risultato della stima con effetti casuali è esposto nella tabella 13 B.

¹⁷ Si riporta qui il risultato della stima "a effetti fissi". Si osservi il valore della stima puntuale del coefficiente della variabile brevetti, molto prossimo a zero e si noti la significatività della stima F sugli effetti individuali: è rifiutata l'ipotesi nulla che questi siano irrilevanti. Tuttavia si noti il basso valore di F per la stima complessiva e il basso impatto sulla variabilità complessiva della variabilità "within", che è quella stimata da questo tipo di regressione. La variabile relativa al dipartimento è naturalmente omessa in quanto "time invariant".

pubb_2per	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
brev_2per	.1666667	.8210692	0.20	0.841	-1.531844	1.865178
_cons	6.666667	.6158019	10.83	0.000	5.392783	7.94055

sigma_u	7.5301728					
sigma_e	3.1799872					
rho	.84865383	(fraction of variance due to u_i)				

F test that all u_i=0:	F(23, 23) =	9.82	Prob > F = 0.0000			
N. Osservazioni:	48					
F(1,23)	= 0.04					
Prob > F	= 0.8409					
R-sq: within	= 0.0018					
between	= 0.1979					
overall	= 0.1274					
corr(u_i, Xb)	= 0.3532					

¹⁸ Si osserva inoltre che il valore di "theta" nella stima "random effect" è pari a 0.626. Si ricorda che se theta è pari a 0 la stima GLS (random effect) è pari alla stima LSDV (fixed effect), se theta è pari a 1 la stima GLS (random effect) è pari alla stima OLS (pooled). Un valore lontano dagli estremi come questo sembra dunque confermare l'ipotesi che lo stimatore GLS sia quello preferibile.

**Tabella 13 B - Regressione 5: effetto individuale dei brevetti sulle pubblicazioni 1990-97
(stime con effetti casuali su dati panel: sottoperiodi: 1990-95 e 1996-97)**

pubb_2per	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
brev_2per	0.7179557	0.739551	0.97	0.332	-0.73154	2.167449
dipart	-9.167417	2.766404	-3.31	0.001	-14.58947	-3.745365
_cons	12.88461	2.392532	5.39	0.000	8.19533	17.57389
-----+						
sigma_u	5.5916823					
sigma_e	3.1799872					
rho	.75561849 (fraction of variance due to u_i)					

N. Osservazioni: 48
Wald chi2 (2) = 13.03
Prob > chi2 = 0.0015

R-sq: within = 0.0018
between = 0.3783
overall = 0.3461

Si osserva che la stima puntuale del coefficiente della variabile relativa alle invenzioni industriali è positivo, ma l'intervallo di confidenza al 95% include valori negativi. Ciò vuol dire che non può essere rifiutata l'ipotesi che nella popolazione un incremento della ricerca industriale provochi un effetto negativo sul numero di pubblicazioni internazionali di ciascun docente.

Se questo risultato fosse confermato da stime più estese e quindi più attendibili, le implicazioni di *policy* sarebbero interessanti: un mutamento di clima culturale o specifici incentivi che portino i docenti universitari ad una scelta di maggior collaborazione con l'industria potrebbe avrebbe effetti negativi sull'attività "propria" e tipica del docente universitario che è quella della pubblicazione scientifica.¹⁹ In tal caso i benefici di un aumento della collaborazione andrebbero confrontati con questo tipo di perdita.

Comunque anche nel caso delle stime "a effetti casuali" va rilevata la distribuzione non normale dei residui, che, data la ridotta numerosità del campione", getta qualche ombra

¹⁹ La validità di queste stime necessita di un altro fortissimo *caveat*, derivante da quanto osservato nel paragrafo 3.2.f): le stime del coefficiente della variabile relativa ai brevetti sono *unbiased* solo se si considera gli archi di tempo considerati come la vere unità temporali e non somma di più periodi. Questo vuol dire che la collaborazione industriale negli archi di tempo considerati (1990-95 e 96-97) sono influenzata solo dalle pubblicazioni e dalla ricerca industriale precedenti a quei periodi, cioè rispettivamente anteriormente al 1990 e dal '90 al '95. E' evidente che, ammesso che sia vero che la ricerca industriale presente è influenzata da pubblicazioni e ricerca industriale passate, la distanza temporale non può essere troppo grande: è assai inverosimile che un incarico di ricerca industriale nel 1994 sia influenzato dalle pubblicazioni fatte fino al 1989 e non da quelle fatte dal 1990 al 1994. Se questa ipotesi "eroica" non è vera c'è correlazione tra la variabile delle invenzioni e gli errori e quindi queste stime sono *biased*.

Per questo motivo, per evitare questo problema, è necessario costruire un *database* con brevi intervalli temporali: in tal modo sarà realistico ipotizzare che la ricerca industriale avvenuta in questo intervallo non sia influenzata da pubblicazioni e ricerca industriale avvenute in questo intervallo.

E' superfluo ricordare che questa fonte di correlazione tra esplicative e residui non è presente, quale che sia l'arco temporale considerato, se non sussiste la relazione tra ricerca industriale presente e pubblicazioni e ricerca industriale passate. L'assenza di questa relazione rende anche non correlate la abilità individuali nella pubblicazione con i residui.

sull'attendibilità dell'inferenza sui parametri e sottolinea la necessità di estendere l'indagine campionaria.

Soprattutto va ricordato che, in base a quanto affermato nel par.3.2.f), vi sono ragioni logiche per sospettare che anche le stime *panel* possano essere affette da un *bias* derivante dall'estrema difficoltà di sfuggire all'endogenità di pubblicazioni e brevetti, dovuta alla correlazione tra questi ultimi e i residui.

In ogni caso, l'esito del confronto tra i tre stimatori considerati ("fixed effect", "random effect", OLS "pooled") ci dice che, nel determinare il numero di pubblicazioni, gli effetti individuali sono sì rilevanti, tuttavia essi non sono significativamente correlati con il numero di brevetti, il che mette in discussione le ipotesi fatte sulla formazione della domanda di ricerca industriale e quindi sul numero dei brevetti, secondo cui questi dipenderebbero dalle pubblicazioni passate e dal numero passato di brevetti. (ugualmente non trova riscontri l'ipotesi che anche il numero di brevetti dipenda da un'inosservabile abilità individuale).

Proviamo dunque a sottoporre direttamente a verifica empirica questa ipotesi.

Stimiamo dunque la seguente equazione di regressione:

$$BR_{it} = \beta_0 + \beta_1 P_{i,t-1} + \beta_2 BR_{i,t-1} + \varepsilon_{it} \quad (64)$$

Se l'arco di tempo considerato per è il 1990-97, perché tali stime siano corrette è necessario che tale arco di tempo vada considerato come un'unica unità temporale, cioè in pratica che i brevetti avvenuti in questo intervallo di tempo siano in relazione solo con i brevetti e le pubblicazioni avvenuti anteriormente a questo intervallo temporale. Dunque, per coerenza con il modello e con le stime precedenti, la stima andrebbe effettuata regredendo le invenzioni del periodo 1990-97 sulle pubblicazioni precedenti al 1990. Purtroppo non disponiamo di dati relativi alle pubblicazioni antecedenti al 1990. Ma seguendo la stessa logica e considerando l'ipotizzata costanza dei parametri nel tempo, si possono regredire le invenzioni successive al 1997 sulle pubblicazioni e le invenzioni relative all'arco temporale 1990-97.

Effettuando una stima *cross-section* di questo tipo -e dunque trascurando gli effetti individuali non osservabili che generano eterogeneità nella scelta della quantità di collaborazione industriale (differenti funzioni di utilità)-²⁰ con le invenzioni del 1998 (variabile BREV_98) regredite sulle pubblicazioni (PUB_9097) e le invenzioni dal '90 al '97 (BREV9097) si ottiene un coefficiente negativo e significativo al 5% per le pubblicazioni, positivo e significativo al 5% per le invenzioni.

Questa regressione, mostrata nella tabella 14 B, ci dice che in un dato periodo t sono gli individui con meno pubblicazioni e più collaborazioni industriali precedenti a collaborare maggiormente con l'industria. Anche in questo caso, però, i residui sono viziati da

²⁰ La presenza di tali eterogeneità individuali non osservabili genera un *bias* in una stima con dati *cross-section*. Anche in questo tipo di indagine vi può essere un *bias* anche con dati *panel* derivante dalla sovrapposizione di più "unità di tempo", che genera endogenità. Si ha infatti:

$$RI_0 = RI_0 + \varepsilon_0$$

Essendo RI_0 la quota non stocastica della ricerca industriale condotta in $t=0$.

Dunque, in virtù della domanda dell'industria, si ha:

$$RI_1 = RI_0 + \eta P_0 (RI_0(\varepsilon_0)) + \theta RI_0(\varepsilon_0) + \varepsilon_1$$

$$RI_2 = RI_0 + \eta P_1 (RI_1(\varepsilon_1, \varepsilon_0)) + \theta RI_1(\varepsilon_1, \varepsilon_0) + \varepsilon_2$$

$$RI_0 + RI_1 + RI_2 = 3 RI_0 + \eta [P_0(\varepsilon_0) + P_1(\varepsilon_1, \varepsilon_0)] + \theta [RI_0(\varepsilon_0) + RI_1(\varepsilon_1, \varepsilon_0)] + \varepsilon_0 + \varepsilon_1 + \varepsilon_2$$

Se

$$BR_t = \delta_1 RI_t$$

Si ha:

$$\Sigma RI = \alpha + \eta \Sigma P + \theta / \delta_1 \Sigma BR + \Sigma \varepsilon$$

c'è dunque correlazione tra le variabili esplicative ΣP e ΣBR con il termine di errore: la stima OLS di η e θ / δ_1 sarà "biased".

eteroschedasticità e non normalità. Stime “robuste”, comunque, non fanno salire il livello di significatività al di sopra del 10%.²¹

Tabella 14 B - Regressione 6: effetto di pubblicazioni e brevetti 1990-97 sui brevetti del 1998 (stime OLS su dati cross section)

		Coefficients ^a				
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,124	,077		1,595	,126
	PUB_9097	-1,02E-02	,004	-,456	-2,395	,026
	BR_9097	,139	,039	,673	3,533	,002

a. Dependent Variable: BREV_98

N. Osservazioni: 24

F (2, 21) = 6,664

Prob.>F = 0,006

R-Squared: 0,388

Adj. R-Squared: 0,330

Poiché la variabile dipendente assume solo valori 0 e 1 (ci sono 3 soggetti titolari di una invenzione nel 1998 e tutti gli altri con nessuna) sarebbe forse più corretta una *logit* o una *probit*. I risultati che si ottengono con queste analisi sono analoghi alla regressione lineare per quanto riguarda il segno dei coefficienti, ma il coefficiente della variabile relativa alle invenzioni precedenti non risulta significativo.

Ma il vero problema che affligge questa stima è un altro: i 3 soggetti che risultano inventori nel 1998 sono in realtà co-inventori. Due di essi erano ricercatori al momento dell'invenzione, quindi con poche pubblicazioni ma con esperienza di ricerca industriale alle spalle; l'altro era ordinario, dunque con alto numero di pubblicazioni e molte collaborazioni industriali precedenti. Si tratta dunque di un *team* di ricerca in cui il docente di maggiore prestigio ha *portato con sé* due collaboratori. Ebbene, la logica del modello è che esista una domanda di ricerca industriale per ogni soggetto, dipendente dalle pubblicazioni e dalla ricerca industriale di ciascun soggetto; è invece immaginabile che l'industria desideri un ricercatore di vertice che abbia molte pubblicazioni ed esperienza di ricerca industriale; se questi poi svolge la ricerca con dei collaboratori giovani e meno prestigiosi l'industria non ha problemi ad accettare la loro collaborazione, ma non forma una domanda direttamente su di loro.

Dunque nell'analisi delle determinanti della ricerca industriale bisogna considerare le collaborazioni tra docenti e quindi l'appartenenza a reti di ricerca; il modello presentato si può arricchire considerando che sono consueti scambi tra reti di pubblicazioni e reti di ricerca industriale.

Il rilievo delle collaborazioni tra docenti e il suo impatto sulla stima delle determinanti della ricerca industriale è confermato da quanto accade effettuando la stessa stima di sopra relativamente ai brevetti del 1996 “regrediti” sulle pubblicazioni e i brevetti dal 1990 al '95. Nel 1996 gli

²¹ Una regressione con dei “pesi” dati alle pubblicazioni e alle invenzioni precedenti, in modo che i dati più recenti pesino di più di quelli più remoti, danno risultati del tutto analoghi a quelli della regressione “non pesata”, anzi l'R-quadro risulta più basso (0,259 l'R-quadro corretto).

“inventori” sono 6 e uno di essi è titolare di 3 invenzioni. Non vi sono co-invenzioni, dunque viene meno il problema precedente: ognuno “fronteggia” una propria domanda industriale. Ebbene, il segno del coefficiente delle invenzioni precedenti si inverte ed è significativo: chi ha fatto più pubblicazioni in passato ha più collaborazioni oggi; il segno delle collaborazioni industriali precedente rimane positivo ma non è significativo. Vediamo i risultati nella tabella 17. Aggiungiamo che in questo caso i residui sono omoschedastici (accettata l’ipotesi nulla nel test di Cook-Weisberg) ma non normali. La stessa stima effettuata con la *logit* e con la *probit* (distinguendo dunque solo tra inventori e non inventori nel 1996) dà coefficienti con gli stessi segni della regressione lineare, ma nessuno è significativo.

Tabella 15 B - Regressione 7: effetto di pubblicazioni e brevetti 1990-95 sui brevetti del 1996 (stime OLS su dati cross section)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-,188	,138		-1,362	,188
	PUB_9095	6,028E-02	,012	,703	4,882	,000
	BR_9095	,115	,086	,192	1,336	,196

a. Dependent Variable: BREV_96

N. Osservazioni: 24

F (2, 21) = 14,420

Prob.>F = 0,000

R-Squared: 0,579

Adj. R-Squared: 0,539

L’impressione è dunque che nella regressione 6 c’era una variabile rilevante (la collaborazione tra professori) che è stata omessa, da cui una distorsione nelle stime, che sono dunque diverse da quelle della regressione 7, in cui la variabile “collaborazioni tra professori” non è presente nei dati, quindi le stime non soffrono di omissione di variabile rilevante. Questa idea, di omissione di una variabile rilevante nella regressione 6, sembra essere confermata dal suo R-quadro notevolmente più basso.

Si noti che in entrambi i casi né la variabile relativa al dipartimento né quella relativa al grado accademico risultano significative.

Se si tenta di fare un’analisi *panel* considerando due periodi, in modo da stimare l’effetto per un singolo docente di una maggiore attività di pubblicazione e di ricerca industriale passata (il “puro” effetto della domanda industriale, a parità cioè di preferenze individuali e di altre caratteristiche individuali non osservabili)²², le stime si scontrano con questi problemi di variabili omesse e di limitatezza delle informazioni, generando risultati che sembrano troppo poco attendibili e che dunque non ci sembra il caso di esporre.

²² L’equazione da stimare per controllare per le caratteristiche individuali non osservabili è infatti:

$$BR_{it} = \beta_0 + \beta_1 P_{i,t-1} + \beta_2 RI_{i,t-1} + \sum_{j=1, n-1} \beta_3 D_{ji} + \epsilon_{it}$$

L'imperativo è dunque arricchire le informazioni, in modo da costruire un *database panel* dettagliato, tenendo conto di altri fenomeni importanti rimasti esclusi da questa analisi, come la già citata presenza di reti di collaborazione, nelle pubblicazioni e nella ricerca industriale, la presenza di possibili fenomeni ulteriori di cumulatività, come nelle pubblicazioni (chi ha pubblicato molto in passato su riviste internazionali ed è dunque considerato autore prestigioso, a parità di valore della ricerca ha più probabilità di pubblicare di un autore poco conosciuto), l'influenza sulla nascita di contratti di collaborazione industriale di altre variabili "osservabili" oltre quelle qui rilevate, come l'appartenenza a centri di ricerca, ecc. Vanno inoltre meglio indagati i meccanismi che sottostanno a quella che abbiamo chiamato la produttività della ricerca industriale in termini di pubblicazione, da fenomeni positivi in questo senso come le pubblicazioni congiunte tra scienziati accademici e industria e fenomeni negativi come la restrizione o i ritardi imposti alle pubblicazioni.

5. Conclusioni

Questo articolo affronta il problema delle collaborazioni di ricerca tra università e industria e soprattutto delle possibili conseguenze che un incremento di tali collaborazioni -una realtà che è effettivamente in crescita- può avere sull'attività "tipica" dell'università che è, oltre all'insegnamento, l'avanzamento della conoscenza scientifica e la sua diffusione tramite la pubblicazione di libri e di articoli.

Viene qui proposto un semplice modello di scelta razionale dello scienziato accademico e dell'industria basato sull'incontro di domanda e ricerca industriale per docenti universitari.

Soprattutto si cerca di razionalizzare il rapporto esistente tra output e input della ricerca, cioè tra pubblicazioni e brevetti da un lato e tempi di ricerca di base e applicata dall'altro. L'obiettivo è quello di mettere in luce i possibili rapporti sia di sostituibilità che di complementarità tra brevetti e pubblicazioni.

Infatti in sede teorica è stato sottolineato da più parti il possibile impatto negativo dell'attività di ricerca industriale sull'attività di pubblicazione, mentre diverse analisi empiriche hanno riscontrato una relazione positiva o almeno non chiaramente negativa. L'analisi di alcune delle possibili, complesse relazioni esistenti tra output ed input della ricerca consente di conciliare questi due aspetti apparentemente contraddittori, mostrando come sia possibile riscontare una relazione positiva tra pubblicazioni e brevetti, pur in presenza di effetti di sostituzione tra forme di ricerca.

Nella seconda parte dell'articolo si tentano delle verifiche empiriche in merito alla relazione tra pubblicazione e brevetti.

La raccolta di dati su cui effettuare stime attendibili è ancora largamente incompleta, pertanto le stime finora compiute e di cui si presentano i risultati hanno soprattutto un valore di esemplificazione metodologica, mentre i loro risultati vanno considerati con la massima cautela.

In ogni caso, nel più vasto campione considerato, i docenti che collaborano con l'industria hanno una media di pubblicazioni più alta rispetto a coloro che non collaborano.

Lo stesso risultato si riscontra nel campione più piccolo, ma nel momento in cui si passa ad analizzare i dati secondo la tecnica *panel*, quindi considerando gli effetti a livello di singoli individui, la direzione della relazione diventa incerta.

Una possibile spiegazione di questa riduzione o addirittura inversione della relazione positiva tra brevetti e pubblicazioni nel momento in cui si passa al livello individuale può risiedere nel fatto che pubblicazioni e brevetti passati rappresentino un effetto segnalazione e quindi siano alla base di una maggiore domanda di ricerca applicata da parte dell'industria. Tuttavia nel piccolo campione considerato non si è trovata una verifica diretta di tale effetto. L'incertezza di questa relazione sembra dipendere da fenomeni ancora poco considerati in questo contesto, quali l'esistenza di *team* di collaborazione tra docenti sia nell'attività di pubblicazione che in quella di

ricerca industriale, *team* in cui si può supporre che i componenti di maggiore prestigio rappresentino il punto di riferimento per l'industria ed abbiano quindi la capacità di "portare con sé" i loro collaboratori.

In ogni caso molti aspetti della complessa relazione tra pubblicazioni e brevetti sembrano difficilmente indagabili, perché non si manifestano in variabili direttamente osservabili. Sembrerebbe pertanto particolarmente sensato accompagnare e integrare l'indagine statistico-econometrica con delle interviste, che rendano esplicite relazioni altrimenti oscure e puramente ipotetiche.

Il cantiere di studio teorico ed empirico su questi temi è dunque aperto: il lavoro da fare è molto, la speranza è che produca risultati di un qualche interesse.

Bibliografia

- Agrawal, A., Henderson, R. (2002), "Putting Patents in Context: Exploring Knowledge Transfer from MIT", *Management Science* 48, 48-60.
- Allison P., Stewart J. (1974), "Productivity Differences Among Scientists: Evidence for Accumulative Advantage", *American Sociological Review*, 39 (August), 596-606.
- Balconi, M., Borghini, S., Moisello, A. (2003), "Ivory Tower vs. Spanning University: il caso dell'Università di Pavia" in "Il sistema pubblico della ricerca in Italia", a cura di Andrea Bonaccorsi, FrancoAngeli.
- Balconi M., Breschi S., Lissoni F. (2002), "Networks of Inventors and the Location of University Research: an Exploration of Italian Data", CESPRI Working Paper, March.
- Balconi M., Breschi S., Lissoni F. (2003), "Il trasferimento di conoscenze tecnologiche dall'università all'industria in Italia: nuova evidenza sui brevetti di paternità dei docenti", in "Il sistema della ricerca pubblica in Italia", a cura di Andrea Bonaccorsi, FrancoAngeli.
- Breschi S., Lissoni F., Montobbio F. (2004), "Open Science and University Patenting: a Bibliometric Analysis of the Italian Case", paper presentato alla 10a conferenza internazionale della Schumpeter Society, Milano 9-12 giugno 2004.
- Bianchi P., Iorio R., Labory S., Malagoli N. (2002), "EU Policies for Innovation and Knowledge Diffusion", Quaderno del Dipartimento di Economia Istituzioni Territorio dell'Università degli Studi di Ferrara, n.17, maggio.
- Bonaccorsi A. (2000), "La scienza come impresa", FrancoAngeli, Milano.
- Calderini M, Franzoni C. (2004), "Who is patenting in public research?" And to what detriment? An empirical analysis of the relationship between scientific careers and patent applications", paper presentato all'International Workshop on Public Research Institutions, International Business and Technological and Economic Catch Up in Developing Regions, Catania, 19-20 marzo 2004.
- Carayol N. (2003), "Objectives, agreement and matching in science-industry collaborations: reassembling the pieces of the puzzle", *Research Policy*, 32, 887-908.
- Cohen, W. M., Florida R., Randazzese L., Walsh J. (1998), "Industry and the Academy: Uneasy Partners in the Cause of Technology Advance, in Roger Noll, ed., *Challenge to the Research University*. Washington, DC: Brookings Institution.
- Dasgupta P., David P. (1994), "Toward a New Economics of Science", *Policy Research*, vol.23, 487-521.
- Geuna A. (2001), "The Changing Rationale for European University Funding: Are There Negative Unintended Consequences?", *Journal of Economic Issues*, 3 (September), 607-631.

- Geuna, A, Nesta L. (2003), "University Patenting and its Effect on Academic Research", SPRU.Electronic Working Paper Series no.99.
- Jensen R, Thursby, M (2003), "The Academic Effect of Patentable Research", mimeo.
- Merton R.K. (1988), "The Matthew Effect in Science, II. Cumulative Advantage and the Symbolism of Intellectual Property", *ISIS* 79, 606-623.
- Meyer-Krahmer, F., Schmoch U. (1998), "Science-based Technologies: University-Industry Interaction in Four Fields", *Research Policy*, 27, 835-851.
- Schibany A., Schartinger D. (2001), "Interactions between Universities and Enterprises in Austria: an Empirical Analysis at the Micro Sector Levels" in OECD, *Innovative Networks. Cooperation in National Innovation Systems*.
- Stephan, P. E., Gurmu, S, Sumell, A.J., Black, G. (2002): "Patenting and Publishing: Substitutes or Complements for University Faculty?", mimeo.
- Zucker P., Darby M., Armstrong J. (1998), "Geographically Localized Knowledge: Spillovers or Markets?", *Economic Inquiry*, XXXVI, January, 65-86.