Identification Strategies for Models of Innovation, R&D and Productivity

Juana Sanchez *

Abstract
Increasing productivity is a major priority for the U.S. and other advanced nations. Because of this, the link between innovation input (R&D), innovation output (process, product and logistic innovation) and productivity is a relevant policy issue. This presentation is about the capability of currently available survey data such as the Business Research and Development and Innovation Survey (BRDIS) to establish a causal relationship that is useful for making predictions about the consequences of changing innovation inputs. The observational nature of BRDIS makes it necessary to recognize potential selection, simultaneity and, more generally, endogeneity biases before any links between R&D, innovation and productivity can be established. Lacking random assignment policy evaluations and a causal interpretation of the findings, a new identification strategy based on structural equation models is presented in this paper that tries to address all those problems.

Key Words: Structural equation models, SUR, graphical models, R&D,

1. Introduction

The goal of this paper is to validate well known theoretical causal hypotheses about drivers of R&D investment using a micro-econometric model. The following two research questions are addressed: (i) do those drivers help predict innovation outcomes and the percentage of sales due to innovation, and if so how can we measure their effect? (ii) what percentage of innovation sales accrue via the effect of those drivers on R&D as compared to their direct effect on innovation sales? Because the causal assumptions made are grounded on well known theories, the building blocks of the model are defensible and consistent with the current state of knowledge. In particular, the micro-econometric model under investigation is intended to provide convincing evidence for causal claims about the effect of demand, appropriability of returns, collaboration, technological opportunity, firm characteristics (such as size, market power, diversification, legal form of organization), use of external knowledge [65], market structure and financial constraints on the decision of firms to invest in R&D and on innovation sales. Separating out the causal effect of R&D and the direct effect of those drivers on innovation is necessary in order to account in the analysis for firms that innovate without R&D, but may be driven instead by collaboration, intellectual property transfers (IP) and costless incentives. This in turn is useful for making predictions about the consequences of changing circumstances or policies relevant for both types of firms such as for example open innovation policies that affect intellectual property exchange; the separation would tell us what would happen in alternative (or counterfactual) worlds[2, 3, 49].

Private Research and Development outlays (R&D) are investments of firms on activities intended to innovate. Successful implementation, i.e., commercialization, is necessary for the innovative effort to end up being an innovation and for sales revenues due to innovation to ensue [53, 10] . The innovativeness of the U.S., however, is measured almost exclusively

*Department of Statistics, University of California Los Angeles. Email: jsanchez@stat.ucla.edu. The research in this paper is conducted while the author is Special Sworn Status Researcher of the U.S. Census Bureau at the Center for Economic Studies. Research results and conclusions expressed are those of the author and do not necessarily reflect the views of the Census Bureau. The results have been screened to insure that no confidential data are revealed.

Table 1: Percentage of firms reporting innovation, by R&D status. Source: BRDIS 2008-2011. Note that these are row percentages. The numbers in the last row reflect only the sample that reports innovation. Sample size: 159,000.

R&D status	Goods	Services	Process	Logistics	Support
No R&D 120,500	7,600 (6%)	6,600 (5%)	6,200 (5%)	3,800 (3%)	8000 (7%)
R&D 38,500	22300 (58%)	10500 (27%)	14100 (37%)	6500 (17%)	11400 (30%)
159,000	29900	17100	20300	10300	19400

by its R&D (i.e., by its innovative activities effort) [12, 7, 66]. By that metric, the U.S. was 10th in the world in 2011, with somewhat over 2.8% of GDP[23].

The recognition that private innovative activity, measured by private R&D investment, is a fundamental contributor to the economic growth and competitiveness of the U.S. [22, 4, 55] has led national science and innovation policy under the umbrella of the America COMPETES Act to aim to bolster corporate research by making R&D tax credits permanent [9, 54, 43, 44, 11, 64, 36, 12]. R&D tax credits have been the most important private innovation policy of the last two decades, despite evidence that R&D and total economic growth at the firm level do not always correlate [36, 12]. Finding channels for alternative private innovation policies that do not lean on tax credits and R&D but could nonetheless foster innovation in those firms that innovate without R&D is not easy. To find them, it is first necessary to know how well R&D predicts a firm‘s actual innovation and innovation sales, and what are alternative mechanisms leading to innovation, which are not well understood yet [24, 33, 34]. Causal inference in this complex context is hard for three reasons: first, the state of knowledge prevailing in this field is not always consensual because it has been obtained through a variety of data sets, periods and methods, leading to multiple interpretations, as indicated by the comprehensive survey of Cohen [12]; second, although there are long time series records of R&D data, the production of data on U.S. innovation and innovation sales at a large scale is still in its infancy, making large scale panel data analyses of U.S. innovation not feasible yet; and third, innovation strategies of firms have many dimensions, such as for example intramural, extramural, research, development, commercialization, goods and services or process. In particular, as services increasingly dominate the global economy, firms aim to introduce service innovations and gain competitive advantage [20]. While small scale studies of particular industries have provided valuable knowledge [60], the importance of innovation for U.S. competitiveness [55] calls for causal inference using large and diverse populations of business establishments. Using the same data throughout the preliminary analysis described in this paper helps focus the interpretation on the methodology and not on the data.

This paper contains the results of preliminary research on the determinants of innovation sales (with or without R&D). Firms in the U.S. are classified by their innovation strategies, with groups differing by drivers of innovation such as demand, appropriability of returns, technological opportunity, R&D and economic sector [62]. That descriptive confirmation of causal assumptions grounded in theory needs is strengthened with causal inference intended to determine the size of the causal effects [61]. Regression approaches to causal inference with seemingly unrelated regression models (SUR) for Research, Development, innovation sales and productivity indicate that each of those outcomes is affected by different groups of drivers, as should be expected, given the differences in characteris-

Table 2: Percentage of goods and services innovative U.S. firms that have new-to-market or new to firm innovation, by R&D status and economic sector. Source: BRDIS 2009-2011. Weighted. Sample size: 47000.

R&D Status	Sector	New-to-market	New-to-firm	Sample size
No R&D	Service	52%	67%	4700
No R&D	Manufacturing	54%	68%	4600
R&D	Service	71%	64%	6500
R&D	Manufacturing	70%	73%	11100

Table 3: Percent of sales due to "new-to-market (ntm)" and "new-to-firm" (ntf) innovations and sales due to the usual line of business. Weighted. Source: BRDIS 2009-2011.

R&D Status	Sector	N	Variable	Average % sales	sd
No R&D	Service	4700	% sales due to ntm	13	29
			% sales due to ntf	14	27
No R&D	Manufacturing	4600	% sales due to ntm	10	23
			% sales due to ntf	12	23
R&D	Service	6500	% sales due to ntm	24	39
			% sales due to ntf	16	29
R&D	Manufacturing	11100	% sales due to ntm	13	24
			% sales due to ntf	12	22

tics among them. Separately, multinomial choice models support the hypothesis that the type of innovation pursued by firms is also associated with some of those drivers. The Structural equation models (SEM) then help identify which factors are relevant when all the relations are simultaneously taken into account and the separation of direct and indirect effects of the drivers on innovation sales is made. The preliminary results were obtained using microdata produced by the Business Research and Development and Innovation Survey (BRDIS) 2008-2011, which contains data on innovation by firms that do and do not conduct R&D. BRDIS is an annual mandatory survey of approximately 40,000 for-profit non-farm firms (with at least 5 employees) representing about 2 million firms in the U.S. each year [68, 8, 10]. The paper ends with a discussion of possible directions for future research, in particular, the examination of the causal hypotheses already examined by means of Bayesian Graphical Models (also known as Directed Acyclic Graphs or Dags).

2. Theoretical context and research hypotheses

Causal claims on the effect of drivers of innovation have been usually framed in a knowledge (K) production function and the contribution of innovation to productivity in an output production function [31, 30]. The knowledge production function approach assumes that the production of new knowledge depends on current and past investment in new knowledge (e.g. current and past R&D expenditures) and on other factors such as knowledge flows from outside the firm. The augmented Cobb-Douglas production function for innovation sales that drives this paper's hypotheses has the following form [36]:

$$Y = AL^{\alpha}C^{\beta}[K]^{\lambda}[K^0]^{\phi}e^{\mu}.$$

Table 4: Dual innovation status of U.S. firms, BRDIS 2008-2011 (weighted percentage). A star indicates that the manufacturing sector has higher share of that category than the non-manufacturing sector.

Status	**Percent of all firms**
Dual innovation	13%($*$)
Goods innovation/no services innovation	22% ($*$)
No innovation in goods or services	60%
Services innovation/no goods innovation	5%

Y is innovation sales, L is labor decomposed into specialized R&D labor and other labor and the two Ks have a direct and indirect component. The direct component measures the direct effect of drivers of innovation on innovation sales (Y) and the indirect one the influence on Y due to the effect of those drivers on R&D. By separating the effects, we allow for the R&D and non-R&D firms to be represented by the same function. The variable A captures other drivers of innovation such as firm's age, firm's legal form, demand, appropriability of returns, market structure and other exogenous factors that may have an impact on innovation sales.

A stylized fact is that innovation depends on how much is spent on R&D and economic sector [46]. BRDIS firms that do and do not do R&D provide a counterfactual scenario with a quasi-treatment group (R&D) and a quasi-control group (no R&D). The data confirms that: (i) there are firms that innovate without investing in R&D [8, 26, 27, 40, 14, 68, 62] (Table 1) and 6% of the latter were involved in goods innovation [10]; (ii) there are firms investing in R&D that do not innovate [60] (Table 1); (iii) the proportion of sales due to innovations is not largest in the manufacturing sector which is the most prominent investor in R&D and the recipient of the largest proportion of R&D tax credit incentives (Table 3). Open innovation policies could be targeted to firms that innovate without R&D outlays [24, 33, 34]. The manufacturing sector in the tables and elsewhere in the paper comprises companies with NAICS 31-33. R&D active companies are companies that either perform or fund R&D research.

Table 2 describes the percent of firms with at least one type of innovation that have new-to-market and new-to-firm goods and services innovations. Innovative enterprises are firms that actively create new knowledge, they have new-to-market innovations [1, 36, 29]. New-to-market innovation is sequential [56, 58] and requires both research (R) and development (D), whereas new-to-the-firm innovations may only require D outlays. As we can see in Table 3, sales revenue due to innovativeness in goods and services, do not necessarily follow the R&D outlays. The sales revenue from R&D outlays on radical innovations is higher in the non-manufacturing sector, despite the fact that manufacturing firms introduce more product and process innovations (22% versus 8%) [8], are the major investors in R&D and receive the largest share of R&D tax credits.

Table 4, which focuses only in goods and services innovations, reveals that a significant share of U.S. firms (13%) implement both goods and services innovation, and more so in the manufacturing sector than in the service sector.

2.1 Drivers of innovation in the literature. Research hypotheses

Regarding the effect of the following factors on R&D [12], the literature has presented mixed results. This paper hypothesizes that the following factors are drivers of innovation, not only through R&D but also indirectly via other mechanisms yet to be found:

Table 5: Factor Analysis of innovation strategies of firms, based on BRDIS data, years 2008-2011.Rotated Factor Pattern. Varimax rotation. N=159000.

Variable	Factor1	Factor2	Factor3	Factor4
New to firm innovation	0.12	0.11	**0.80**	0.25
new-to-market innovation	0.21	0.12	**0.80**	0.16
Technical process innovation or not	0.23	0.05	0.33	**0.75**
Nontechnological process innovation	0.14	0.06	0.13	**0.89**
In house RD	**0.66**	0.22	0.43	0.16
firm applies, receives or gets license money from patents.	**0.69**	0.21	0.18	0.05
Paid by firm, performed by others	0.19	**0.95**	0.14	0.06
RD expenses in equipment	**0.61**	0.26	0.37	0.08
Purchased external knowledge	0.17	**0.95**	0.12	0.05
Design patents are important or very important	**0.78**	0.04	0.03	0.14
Copyright is important or very important	**0.76**	0.05	0.05	0.17
Variance explained	24.3	18.2	16.4	14

appropriability of returns obtained through patents, copyrights and trademarks [22, 28, 42, 35]. The surge in patenting in the last two decades has not translated into a significant effect on economic performance, and only a small fraction of innovative firms relies on patents to protect their inventions[18, 13, 16, 35, 51]. Trademarks, trade secrets, and copyrights are more important forms of protection, and a potential driver of innovation [14].

technological opportunity (closeness to science, technology within the industry, maturity of the industry) affects research differently than development;

demand conditions , how much the market is expecting the innovation affects innovation directly;

industry structure has had a mixed impact that will continue with the BRDIS data set [12];

firm characteristics such as market power, monopoly, size, legal form of organization affect innovation directly;

intellectual property transactions and research collaboration schemes are more relevant now than in the past, and may result in innovation without R&D.

formal collaboration with other companies [48, 37] leads to higher productivity depending on the type of R&D and management of technology.

cash flow has a mixed effect on innovation [50].

in-licensing or out-sourcing and other forms of strategic alliances such as cross-licensing mediate effects of market structure and other factors on innovation [13, 63, 32].

3. Research Results

Guided by the theoretical causal assumptions described above, the empirical validation that follows is done under different perspectives and multivariate methodologies, seeking replicability of the results. This section summarizes the results obtained.

3.1 Innovation strategies of firms

Exploratory and confirmatory factor analysis identifies latent innovation strategies of firms that could be seen as different dimensions of the innovation process. Cluster analysis to group firms into categories with specific ratings in terms of their innovation strategies scores [25] result in groups well defined by variables traditionally used in R&D academic research [62].

The results of the factor analysis presented in Table 5 are based on unweighted data, principal component analysis and varimax rotation method [62]. An item is said to load (be given a lot of weight) on a given component if the factor loading is .50 or greater for that component, and is less than .50 for the other. The magnitude of the factor loading represents how correlated a firm‘s response for that item and the underlying, latent dimension is. Using these criteria, four items are found to load on the first component, which is subsequently labelled the intramural, capital intensive, intellectual protection intensive R&D strategy. Two items load on the second component, which is labelled the extramural, intensive use of external knowledge R&D strategy. Two items load on the third component, which is labeled the goods and services innovation strategy. Finally, two factors load on the fourth component, which is labeled the process innovation strategy. Confirmatory factor analysis suggests that this is a good fit to the data.

Firms present more than one innovation strategy. Factor scores are assigned to each firm. A factor score is an average that weights each item according to how salient it is to the concept being measured. They range from approximately -3 to 3. A firm with a factor score of 3 in factor 1 is strongly characterized by the mode of innovation represented by factor 1, that is, it performs above average in relation to factor 1.The proximity of firms to other firms based on the factor scores yielded 4 main groups obtained with cluster analysis [62] [19] [41]. We can see preliminary results in Table6. The values are unweighted, thus they represent only the sample.

The average and standard deviation of the factor scores in each cluster, the average size and sales of the firms in each cluster, and the average values of many other relevant variables can be seen in Table 6, which shows that innovation strategies present themselves in groups of firms with distinctive economic characteristics. Cluster 1 contains the smallest (on average) firms that are intramural and goods and service innovators, over 70% R&D active, relying on trade secrets for intellectual property protection and the highest percentage of sales from innovations. On the other hand, cluster 4 has the largest, intramural and heavy R&D companies that process innovate, make heavy use of external knowledge and other forms of intellectual property protection and are second in innovation sales. Cluster 2 clearly represents the companies that are below average in all innovation strategies. Groups 1, 3, and 4 have approximately the same proportion of manufacturing firms.

3.2 Strategies for the validation of causal claims

Clarifying the economic mechanisms that bring the innovation phenomenon and the innovation sales into existence requires more than descriptive analysis. Knowing the process that generates innovation sales can indicate points for policy intervention. There has been one main evidentiary strategy for supporting studies‘ causal claims: mimic an experiment

Table 6: Cluster analysis based on BRDIS data 2008-2011. Factor scores are standard deviations above (+) or below average (−). Dichotomous variables are in proportions of firms with the characteristic. Sales are in thousand dollars. Source of data: BRDIS 2008-2011.

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4
N	14000	117400	18000	9600
Average Sales	211466	270601	561078	1259655
Average employment	540	778.83	1438	2111
Factor1	**0.53**	−0.16	0.29	**0.70**
Factor2	−0.57	−0.18	−0.40	**3.77**
Factor3	**2.19**	−0.31	0.08	0.43
Factor4	−0.34	−0.32	**2.31**	0.17
New to firm innovation	**0.72**	0.00	0.28	0.43
new-to-market innovation	**0.66**	0.00	0.25	0.44
Technical process innovation	0.33	0.01	**0.79**	**0.37**
Nontechnological process innovation	0.13	0.00	**0.82**	0.28
In house RD (intramural)	**0.75**	0.12	**0.57**	**0.91**
Firm applies, receives or gets License money from patents.	0.32	0.05	0.22	**0.54**
R&D paid by firm, performed by others (extramural)	0.05	0.00	0.01	**1**
RD expenses in equipment.	0.45	0.06	0.33	**0.69**
Purchased external knowledge.	NA	0.00	0.00	**0.95**
Design patents are important or very important.	0.32	0.09	0.28	**0.46**
Copyright is important or very important.	0.46	0.13	0.42	**0.58**
Utility patents	0.41	0.09	0.33	**0.65**
IP transfer	0.21	0.04	0.33	**0.65**
Formal collaboration expenditures	0.08	0.00	0.03	**0.22**
Total World RD	**0.72**	0.11	**0.55**	**1**
RD paid by others, own performed	0.13	0.03	0.10	**0.23**
Weight of trade secrets	**0.64**	0.17	0.60	**0.78**
Weight of mask works	0.06	0.03	0.08	0.11
Variety of industries		Most		least
Percent in manufacturing	54%	40%	57%	58%
Average innovation sales	1st	last	2nd	2nd

by attempting to hold other explanatory variables constant through stratification, matching, or regression, in order to strip an observed association of all spurious components. This approach has dominated Economics for decades and has been and will continue to be a major tenet of any research. A look at the long list of papers on innovation and R&D testifies to that [12].

3.2.1 Multinomial models of the decision to innovate

The first question addressed is: does R&D have any effect on the type of innovation obtained by the firm. To have innovation sales a firm must first innovate. This is a rather stochastic event that happens if the right decisions are made, and if chance helps as well. Thus it differs from the decision to invest in R&D, which is made only by the firm. When there are more than two categorical responses without order, the logit model is also called generalized logit model, discrete choice model or multinomial model [38, 57, 47]. This

Table 7: Four Seemingly Unrelated Regressions. Dependent variables: Research part of R&D, Development part of R&D, Innovation sales and Productivity. Results include coefficients and standardized coefficients (in parenthesis). Starred are significant at 5% level. Weighted. Sample size used for computations=2200 (rounded) (those observations with complete data). The data gets reduced to those firms that conduct both Research and Development, have sales due to innovation new-to-market or new to firm, and have non-missing value of productivity. System weighted R-square=0.67. Source of data: BRDIS 2008-2011.

Lagged Independent Variable	Research Intensity	Development Intensity	Innovation sales	Productivity.
lag sales		0.42(0.52)*		0.57(1.26)*
lag sales square		$-0.03(-0.48)$*		$-0.04(-1.28)$*
lag IP trade		0.07(0.02)*	0.18(0.07)*	
lag capital		$-0.14(-0.24)$*		$-0.05(-0.17)$
lag capital square		0.01(0.21)*		0.005(0.26)*
lag demand	0.10(0.03)*			
lag high tech	0.16(0.04)*			
lag market appropriability	0.46(0.02)			
lag log R intensity	0.76(0.79)*		0.06(0.07)*	
lag log D intensity		0.8(0.90)*	0.14(0.17)*	
lag sales	$-0.13(-0.07)$	$-0.16(-0.09)$*	$-0.21(-0.15)$*	0.59(0.64)*
group			0.21(0.08)*	
lag innovation			0.25(0.05)*	
lag market power				0.34(0.96)*
lag markpower sq				0.03(1)*

type of model has been used to model firm's innovation strategies choices, complementarities in innovation, and innovation and competition [15, 21, 5]. The results are usually presented in the form of odds ratios.

A weighted generalized logistic regression (or multinomial) model used to investigate what characteristics of firms lead them to be in any of the categories seen in Table 4 [61] revealed the following: Manufacturing firms that hire R&D workers and engage in informal IP trading and do intramural R&D are more likely to be dual innovators other things constant. Tacit knowledge embedded in the R&D workers is eight and a half more likely to result in dual innovation. Firm size and productivity had odds ratios very close to one, indicating that although significant they are equally likely to result in any of the types of innovation.

3.2.2 Systems of equations with SUR models

Some authors have shown that Research (R) and Development (D) have different determinants [6]. Seemingly unrelated regression models (SUR), which fit separate regression models to each of the dependent variables, but assuming that there are correlations between the residuals of these models, is an appropriate method to answer this question [39]. Table

7 shows results for four equations for dependent variables log Research intensity, log Development intensity, log innovation sales and log productivity. All independent variables are lagged to avoid reverse causality and endogeneity that could arise due to heterogeneity of omitted variables [52].The model controls for 11 industry fixed effects and year (not shown, but coefficient estimates reveal that there is significant heterogeneity across firms). The lagged independent variables include lagged values of the R and D and innovation sales variables. Table 7 only contains variables that were significant for at least one of the dependent variables. Other variables such as process and logistic innovation were not significant. The most significant finding is that blocks of the independent variables affect one dependent variable but not the others. All four dependent variables are inelastic with respect to the independent ones. The effect sizes are small.

We see in Table 7 that past innovation in goods and services, past log Research and log Development and past IP exchange is associated with higher logged innovation sales. Those lagged variables, however, do not affect productivity (sales per employee). The latter is only affected by lagged firm size, lagged sales, lagged capital (a nonlinear effect) and lagged market power. The only variable that appears to depend the most on the traditional determinants of R&D is log research. The coefficients are elasticities. In parenthesis are the Beta coefficients, which allow us to see which variable has higher effect on the dependent variable. The system wide R square is 67%. The correlation between research and development is 0.26, research and productivity 0.21, development and productivity 0.26. Only firms that participated in the survey more than one year and that had values of all the variables could be included. One drawback of this model is that there is no correlation between the residuals of the innovation equation and the other equations.The other dependent variables are correlated however. SUR also does not allow the estimation of indirect effects in a way that would allow to establish strong causal conclusions. It is possible that the weak strength in the direct associations observed is compensated with a stronger indirect association.

3.2.3 Structural Equation Models

The estimation of functional forms like those described earlier in section 3.2.2 are useful lend statistical support to some of the causal hypotheses posed by economic theory regarding development, research, innovation sales or productivity. This is consistent with the descriptive findings of section 2. Correlation of errors of the four SUR equations ameliorates the endogeneity bias. But stronger causal assumptions can not be brought together and their implication or plausibility can not be imposed on the model to test their compatibility with the data. The two strongest types of causal assumptions are: (1) imposing zero coefficients and (2) imposing zero covariances. These are addressed with structural equation models (SEM).

SEM is a causal inference engine that expresses relationships among a system of variables that can be either observed variables (manifest variables) or unobserved hypothetical variables (latent variables). SEM presents an opportunity to specify causal mechanisms as a way to provide empirical support to causal claims. It takes in two inputs, qualitative causal assumptions and empirical data, and derives two logical consequences of these inputs: quantitative causal conclusions and statistical measures of fit for the testable implications of the assumptions. Failure to fit the data casts doubt on the strong causal assumptions of zero coefficients or zero covariances and guides the researcher to diagnose, or repair the structural mis-specifications. Fitting the data does not prove the causal assumptions, but it makes them tentatively more plausible. Any such positive results need to be replicated and to withstand the criticisms of researchers who suggest other models for the same data [67].

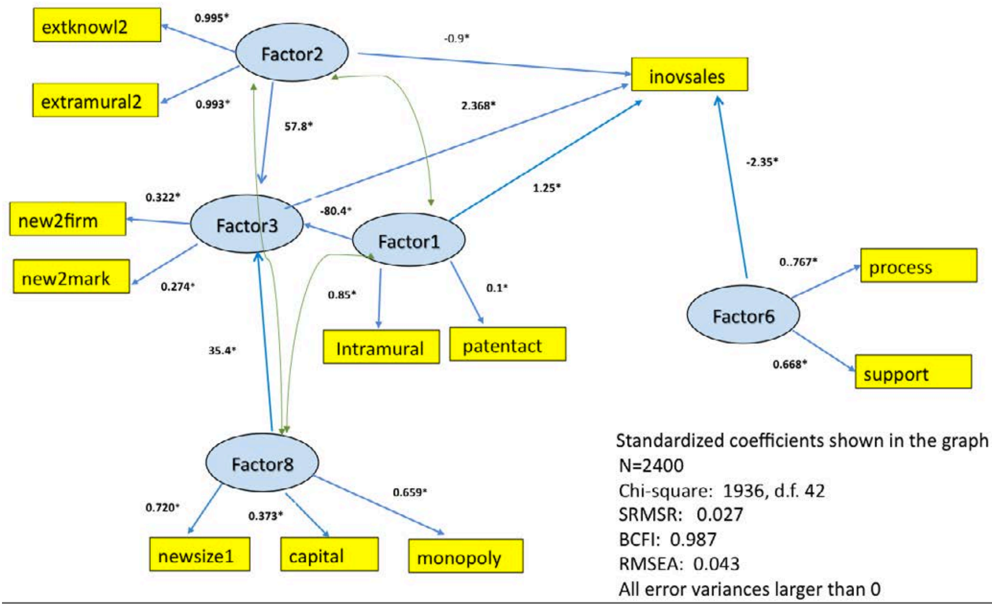

Figure 1: Structural Equation Model. Preliminary results. Source of data: BRDIS 2008-2011

Allowing multiple directions of causality constitutes a conceptual improvement over the single-equation and even instrumental variables and sample selection models. SEM also allows the estimation of direct and indirect effects [59].

The data ended up supporting the four initial factors considered earlier in this paper(now numbered 1, 2, 3, 6) and an additional factor, 8, representing firm characteristics. The structural part of the model is added to the first model described above by incorporating causal relations of the drivers of innovation with innovation sales, economic sector and other variables, as well as causal relations among the factors. The structural equation model supported by the data is that seen in Figure 1 .

Shown in Figure 1 are the standardized coefficients, which measure the effect in standard deviation units and show which variable has the strongest effect. The variables in boxes are observed, the ones in ovals are latent factors. These are variables that are in the theory tested but are not observed as indicated in Section 2. One directional arrows represent effect, two directional ones (in green color) simply association due to unknowns. For example, the correlations between the latent factors, are practically and statisticially significant for factor2 and factor1 (0.921), factor8 and factor1 (0.77), an factor8 and factor2 (0.46). The numbers are not shown, but the green arrows indicate that they are significant. The Chi-square statistic is misleading for such large data set. But the other model fit statistics such as BCFI (larger than 95%), SRMSR (smaller than 0.05), RMSEA (smaller than 0.08) are saying that the data support the causal assumptions made for identifiability.

To conduct the structural equation modeling we restrict the sample to those firms that had positive R&D expenditures or reported some type of innovation. This guarantees to have firms that have only R&D but not innovation and firms that have R&D and innovation. But firms that do not conduct R&D will not be represented because there is no R&D data for these firms. This will have to be palliated in the future with imputation of missing data.

The effects seen in Figure 1 are total effects. Table 3.2.3 identifies the direct and indirect effects in paths where they are significant. The intramural dimension has a positive direct effect on innovation sales but a very strong negative indirect effect via factor3. On the other hand, the extramural dimension has a negative direct effect on innovation sales, but

Table 8: Structural equation modeling (N=24000, rounded). Method of estimation is maximum likelihood. Standardized direct and indirect effects. Preliminary results. Source of data: BRDIS 2008-2011.

Variables	Total	Direct	Indirect
factor1 on inovsales	−189.228	1.253	−190.481
factor2 on inovsales	135.952	−0.900	136.852
factor8 on inovsales	83.837	0	83.837

a very strong positive effect through innovation in goods and services. The characteristics of the firm represented by factor 8 has only an indirect effect on innovation sales. The first result is counterintuive. Intramural research having negative effect on innovation in goods and services dimension has not been supported by the research described earlier. Future research will investigate this issue further. The results presented in Figure 1 are preliminary. Research in progress using the Longitudinal Business Database, and imputation will shed more light on these issues.

An advantage of structural equation models is that it forces us to think about the covariance structure, as each model represents one. Thinking in terms of the covariance structure helps see the hypotheses that we are making regarding the covariances and variances. Another advantage we can account for the error in variables, via the variables error term. But using cross-sectional data with structural equation models leads to the same problem found in single multivariate models with cross-sectional data. Temporal mismatch. Lagged values of the explanatory variables must be used. Versions of the model presented here will be tested using lagged variables and incorporating more bidirectionality among factors and heterogeneity among firms by incorporating industry controls.

4. Conclusions and directions of future research

The preliminary results presented in this paper lead to the conclusion that the innovation process is complex and that looking beyond R&D and into the actual percentage of innovation sales due to innovation helps disentangle some of the complexity. The analysis reveals that some of the drivers of R&D investment traditionally considered in the literature can be considered drivers of innovation sales as well. However, when drivers of innovation are examined in contexts that do not consider the simultaneity of the variables involved, the conclusions seem to be very different from the conclusions reached in simultaneous equation models. Moreover, the identification made with SEM is an identification of causal associations, not causal effects [45]. If we fix a factor at a particular value, the joint probability of all variables does not include a probability term for that factor, since probability of a given value is 1. To represents the causal effect, the joint probability of an outcome under all possible values of the parent variable is needed. With Directed Acyclic Graphs (DAGs) we can obtain causal effects and that is the main reason why new directions of research being considered include use Directed Acyclic Graphs (DAGs) or Bayesian networks to help identify what the structural model should look like and then build the SEM model and confirm with it the relationship hypothesized [17].

References

[1] F. Aiello and P. Cardamone. R&D spillovers and productivity growth: Evidence from italian manufacturing microdata. *Applied Economics Letters*, 12(10):625–631, 2005.

[2] Joshua D. Angrist and J-Steffen Pischke. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30, 2010.

[3] Joshua D. Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics. An Empiricist‘s Companion*. Princeton University Press, 2009.

[4] Daniele Archibugi and Jonathan Michie. Technology and innovation: An introduction. *Cambridge Journal of Economics*, 19:1–4, 1995.

[5] Corinne Autant-Bernard, James P. Lesage, and Olivier Parent. Firm innovation strategies: A spatial cohort multinomial probit approach. *Annales d'Economie et de Statistique*, (87-88):63–80, 2007.

[6] Andres Barge-Gil and Alberto Lopez. R&D determinants: Accounting for the differences between research and development. *Research Policy*, 43:1634–1648, 2014.

[7] National Science Board. *Science and Engineering Indicators*. Arlington VA: National Science Foundation (NSB 14-01), 2014.

[8] M. Boroush. NSF releases new statistics on business innovation. *Infobrief NSF 11-300*, 2010. Arlington,VA:National Science Foundation, Division of Science Resources Statistics.

[9] Barry Bozeman and Albert N. Link. Toward an assessment of impacts from U.S technology and innovation policies. *Science & Public Policy (SPP)*, 42(3):359–376, 2015.

[10] Census Bureau/NCSES. Business research and development and innovation survey (BRDIS), 2010. http://www.census.gov/econhelp/brdis.

[11] Fulvio Castellacci and Christine Mee Lie. Do the effects of R&D tax credits vary across industries? *Research Policy*, 44:819–832, 2015.

[12] W. M. Cohen. Fifty years of empirical studies of innovative activity and performance. pages 129–213, 2010. in Hall, B. H. and Rosenberg, N.(ed.), *Economics of Innovation*. North Holland.

[13] W.M Cohen. Patents and antitrust: Differing shades of meaning. *Research Policy*, 31:1349–1367, 2002.

[14] National Research Council. *Communicating Science and Engineering Data in the Information Age*. 2012.

[15] Bruno Crepon and Emmanuel Duguet. Research and development, competition and innovation pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity. *Journal of Econometrics*, 79(2):355–378, 1997.

[16] C. Crepon, E. Duguet, and J. Mairesse. Research, innovation and productivity: an econometric analysis at the firm level. *Economics of Innovation and New Technology*, 7:115–158, 1998.

[17] Subrata Das. *Computational Business Analytics*. Chapman and Hall/CRC, 2014.

[18] Pelin Demirel and Mariana Mazzucato. Innovation and firm growth: Is R&D worth it? *Industry and Innovation*, 19(1):45–62, 2012.

[19] G. Der and B. S. Everitt. *A Handbook of Statistical Analyses using SAS. Third Edition*. CRC Press, 2008.

[20] Thomas Dotzel, Venkatesh Sharkar, and Leonard L. Berry. Service innovativeness and firm value. *Journal of Marketing Research*, L:259–276, 2013.

[21] M'hand Fares. Using a multinomial probit to test for complementarity. *Applied Economics Letters*, 21(1-3):180–184, 2014.

[22] Lucia Foster and Cheril Grim. Characteristics of the top R&D performing companies in the U.S.: Evidence from the survey of industrial R&D. 2010. Census Bureau, Center for Economic Studies Discussion Paper CES-WP-10-33.

[23] National Science Foundation. *Science and Engineering Indicators 2014*. http://www.nsf.gov/statistics/seind14/index.cfm/chapter-4/c4s2.htm.

[24] Christopher Freeman and Luc Soete. Developing science, technology and innovation indicators: What we can learn from the past. *Research Policy*, 38(4):583–589, 2009.

[25] Marion Frenz and Ray Lambert. Exploring non-technological and mixed modes of innovation across countries. In OECD, editor, *Innovation in Firms, a Microeconomic Perspective*, pages 69–109. OECD Publications, 2009.

[26] Fred Gault. *Innovation Strategies for a Global Economy. Development, Implementation, Measurement and Management.* Edward Elgard, 2010.

[27] Fred Gault. Social impacts of the development of science, technology and innovation indicators. *UNU-MERIT Working Papers ISSN 1871-9872*, 2011.

[28] S. Gopalakrishnan. A review of innovation research in economics, sociology and technology management. *Omega, International Journal of Management Science*, 25(1):15–28, 1997.

[29] Rachel Griffith, Stephen Redding, and John Van Reenen. Mapping the two faces of R&D: Productivity growth in a panel of OECD industries. *Review of Economics and Statistics*, 86(4):883–895, 2004.

[30] Zvi Grilliches. Issues in assessing the contribution of R&D to productivity growth. *Bell Journal of Economics*, 10:92–116, 1979.

[31] Zvi Grilliches. Productivity, R&D and the data constraint. *The American Economic Review*, 84(1):1–23, 1994.

[32] P.C. Grindley and D. J. Teece. Managing intellectual capital: Licensing and cross-licensing in semiconductors and electronics. *IEEE Engineering Management Review*, 26:92–111, 1998.

[33] H. Grupp and M.E. Mogee. Indicators for national science and technology policy: how robust are composite indicators? *Research Policy*, 33(9):1373 – 1384, 2004.

[34] H. Grupp and T. Schubert. Review and new evidence on composite indicators for evaluating national performance. *Research Policy*, 39:67–78, 2010.

[35] B. Hall, C. Helmers, M. Rogers, and V. Sena. The choice between formal and informal intellectual property protection: a review, 2012. NBER Working Paper Series.

[36] Bronwyn H. Hall, Jacques Mairesse, and Pierre Mohnen. Measuring the returns to R&D. Technical report, 2010. in Hall, B. H. and Rosenberg, N.(ed.), *Economics of Innovation*. North Holland.

[37] Jonathan Haskel and Gavin Wallis. Public support for innovation, intangible investment and productivity growth in the UK market sector. *Economics Letters*, 119(2):195–198, 2013.

[38] W. D.Jr. Hosmer and S. Lemeshow. *Applied logistic regression, 2nd. ed.* John Wiley and Sons, Inc., 2000.

[39] G. Imbens. Instrumental variables: An econometrician's perspective. *Statistical Science*, (3):323–358.

[40] J.E. Jankowski, L. Carlson, L. Gibson, P. Hough, R. Lee, R. Shackelford, and R. Wolfe. The redesign of the national science foundation's business R&D and innovation survey: What worked and what did not. *American Economic Association Proceedings*, 2010.

[41] Leonard Kaufman and Peter J. Rousseeuw. *Finding Groups in Data*. Wiley, 1990.

[42] W.R. Kerr and S. Fu. The industry R&D survey: Patent database link project. *Center for Economic Studies, Census Bureau, working paper 06-28*, 2006.

[43] Anne Marie Knott. R&D returns causality: Absorptive capacity or organizational iq. *Management Science*, 54(12):2054–2067, 2008.

[44] Anne Marie Knott. The trillion dollar R&D fix. *Harvard Business Review*, 2012.

[45] S. L. Lauritzen. Causal inference from graphical models. 2001. In Barndorff-Nielsen, O. E., Cox, D. R., and Kluppelberg, C., editors, Complex Stochastic Systems, pages 63107 Chapman and Hall/CRC Press, London/Boca Raton.

[46] F. Malerba. How innovation differ across sectors and industries. 2005. In Fagerberg, J., Mowery, D.C., Nelson, R.R. (Eds.), The Oxford Handbook ofInnovation. Oxford University Press, Oxford.

[47] D. McFadden. *Conditional Logit Analysis of Qualitative Choice Behavior*. 1974. in P.Zarembka, ed. *Frontiers in Econometrics*, New York: Academic Press.

[48] G. Medda, C. Piga, and D. S. Siegel. University R&D and firm productivity: Evidence from italy. *Journal of Technology Transfer*, 30(1-2):199–205, 2005.

[49] Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference. Methods and Principles for Social Science Research*. Cambridge, 2nd edition, 2015.

[50] Pierluigi Murro. The determinants of innovation: what is the role of risk? *Manchester School*, 81:293–323, 2012.

[51] T. Nicholas. Did R&D firms use to patent? evidence from the first innovation surveys. *The Journal of Economic History*, 71(4):1029–1056, 2011.

[52] Junichi Nishimura. R&D portfolios and pharmaceutical licensing. *Research Policy*, 43(7):1250 – 1263, 2014.

[53] OECD/Eurostat. *Oslo Manual*. OECD Publishing, 2005. /content/book/9789264013100-en.

[54] U.S. Department of Commerce. The competitiveness and innovative capacity of the united states, 2012.

[55] M.E. Porter. The competitive advantage of nations. *Harvard Business Review*, pages 73–93, 1990.

[56] G Zaltman R. Duncan and J. Holbek. *Innovations and Organizations*. Wiley, New York, 1973.

[57] G. Roberts, J.N. Rao, and S. Kumar. Logistic regression analysis of sample survey data. *Biometrika*, 74:1–12, 1987.

[58] A. Robertson. Innovation management. *Management Decision Monograph*, 12:6–16, 1974.

[59] James M. Robins and Sander Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155, 1992.

[60] Juana Sanchez. The missing variable in the economic analysis of technological change. *Journal of Economics (MVEA)*, 17:19–24, 1991.

[61] Juana Sanchez. Innovation output choices and characteristics of firms in the U.S. *Center for Economic Studies, U.S. Census Bureau. Working Paper 14-42*, 2014.

[62] Juana Sanchez. Non-technological and mixed modes of innovation in the U.S. evidence from the business research and development and innovation survey 2008-2011. *Center for Economic Studies, U.S. Census Bureau. Working Paper 14-35*, 2014.

[63] C. Shapiro. Navigating the patent thicket: Cross-licenses, patent pools, and standard setting. Jaffe, A.B., Lerner, J. and Stern,S. *Innovation Policy and the Economy* Volume 1: 119-150.

[64] Richard M. H. Suen. Research policy and U.S. economic growth. *University of Connecticut, Department of Economics, Working papers: 2013-18, 2013, 87 pp.*, page 87, 2013.

[65] D.J. Teece. Firm organization, industrial structure, and technological innovation. *Journal of Economic Behavior and Organization*, 31:193–224, 1996.

[66] OECD. *OECD Compendium of Productivity Indicators, 2015*. OECD Publishing, Paris, France, 2015. http://dx.doi.org/10.1787/pdtvy-2015-en.

[67] Xiaolu Wang and Michael E. Sobel. New perspectives on causal mediation analysis. 2013. in Handbook of Causal Analysis for Social Research. Stephen E. Morgan Editor.

[68] R.M. Wolfe. U.S. businesses report 2008 worldwide R&D expense of $330 billion: Findings from new NSF survey. *Infobrief NSF 10-322*, 2010. Arlington,VA:National Science Foundation, Division of Science Resources Statistics.