SENIOR PROJECT
DEPARTMENT OF ECONOMICS

“The Effect of Networks on Firm-Level Innovation”

Daniel Wyrock
Spring 2016

Advisor: Dr. Francesco Renna

Abstract

Research over the past several years has begun to stress the importance of multiple factors in determining what drives innovation at the firm level. Using data from 344 publically traded U.S. manufacturing firms, this study investigates how the connectivity of the board of directors of a firm impacts the innovation of that firm. The results suggest that a 1 percent increase in the network of the board of directors (as proxied by the number of boards on which a member may sit) leads to a .15 percent increase in the number of patents output by the firm. Further research to investigate statistical significance of these results is recommended.

Table of Contents

I. Introduction

Entrepreneurs strive to use the most current and innovative ideas to create the next best widget, service or other offering that will enable some form of financial, social and or personal success. Huggins and Thompson (2015) show that innovation is a key driver in economic growth and suggests that one way to foster this growth and innovation is through networks. In general, a network can take on a different definition depending on the field and/or context. Some strands of the literature specify that an innovative network only encompasses "explicit arrangements, which do not include the informal information sharing arrangements... or the tacit links between firms..." (Love, 1999, p 3). Broadly, network capital includes all intentional relationships, either formal or informal, forged with the express interest of some economic benefit (Huggins and Thompson, 2015). Social capital networks (i.e. a type of network capital where the relationships are among friends, acquaintances and other informal connections) serve as valuable conduits through which knowledge can flow (Mazzola et al, 2016). Regardless of the definition of networks, the literature shows that firms that do network have higher levels of innovation (Freel and Harrison, 2006; Mazzola et al, 2016; Rogers, 2004; Tomlinson, 2010).

Beyond the formal networks of the firm (i.e. explicit agreements between the firm and others), there is evidence that the connectivity of a company's board members, in terms of the number of boards on which a member may sit (commonly known as board interlock), affects the value and innovativeness of the firm (Mazzola et al 2015; Omer et al, 2014). Even though the daily impact of a board is not always clear, they are ultimately held responsible when an organization is underperforming or has done something criminal (Adams et al, 2010). As part of this, the board of directors play an important role in upholding the integrity of the financial statements of a firm (Peasnell et al, 2005; Xie et al , 2002). Further, boards are in place to minimize the agency problem associated with managers of a firm having different and opposing financial incentives from the shareholders of the firm (Adams et al, 2010). One way in which boards can ensure the success and ongoing growth of the firm is to pursue strategies and agendas in which innovation is a key component. As mentioned above and described later on in the paper, evidence shows that having a more connected board leads to increased firm innovation and value.

Using the role of the board and its members' connections as a proxy of part of the firm's network, this paper aims to investigate the impact of networks on innovation. By applying the amount of board interlock of a board's members to the modified knowledge production function explored by Acs and Audtretsch (1989), this paper will add to the existing literature by examining the effect of a firm-level network measure on innovation in publicly traded manufacturing companies. Specifically, does an increase in the interlock of a board lead to an increase in the patent output of the firm.

II. Literature Review

Mark Rogers (2004) investigated how traditional factors (i.e. market structure), manufacturing status (i.e. if a firm is a manufacturer) and also inter-firm networks play a role in the innovation process. Using data on 8,911 Australian firms, the innovativeness of a firm within a given industry (measured as innovator or not) was estimated using probit regressions. The results showed that manufacturing firms that network (i.e. formal networking with other firms regarding innovation) may see higher innovation rates. Similar results were found for non-manufacturing firms showing that there was limited heterogeneity between manufacturing and non-manufacturing firms in the sample. By looking at the determinants of innovation, Rogers' research showed that external knowledge (i.e. in the form of networks) serves an important role in determining the self-reported innovativeness of a firm.

Mark S. Freel and Richard T. Harrison (2006) examined how small firm cooperation with external parties (i.e. other firms or individuals not employed at or by the firm) impacts the frequency and types of innovations output by the firm. Freel and Harrison used a modified knowledge production function to model innovative output of manufacturing and service firms in the UK. Specifically, innovation was split into two categories of either product or process innovation. From there, the likelihood of being a non-innovation versus novel innovation or incremental innovation versus novel innovation was calculated for each innovation category. The explanatory variables included a set of proxies for networks, (i.e. a dummy variable for innovative cooperation with customers, suppliers, competitors, universities or the public sector). Freel and Harrison found that in one example, “60 percent of ‘novel’ product innovators... were engaged in at least one innovation-related cooperative endeavor...”. As evident by the results, cooperation outside the firm leads to an increase in the frequency of innovation of the firm. Freel and Harrison caution, however, that although the results show that external resources increase innovation, internal resources are just as important. This is due to the fact that it is the job of the company itself to utilize the information and knowledge that is gained by cooperating with entities outside the firm to make an informed decision before investing in an innovative idea.

Building off the work of Freel and Harrison (2006), Philip Tomlinson (2010) investigated how co-operative ties of firms (i.e. co-operating with others when conducting R&D) impact their performance of creating innovations. The study examined the effect of co-operation (between the firm and its buyers, the firm and its suppliers, and the firm and its competitors) on both product and process innovation. In comparison to Freel and Harrison (2006), Tomlinson collected more descriptive data by using a five-point measure of co-operation (1 being no co-operation and 5 being a very high level of co-operation) as well as the number of innovations of the firm (split between product and process) for over 400 manufacturing firms in the UK in 2008. Using multivariate estimation on a standard knowledge production function through a hierarchical regression, Tomlinson found that inter-firm co-operation increases innovative performance. Further, the results show that there is heterogeneity between industries in terms of the types of co-operation and its impact on innovative performance. For example, Aerospace firms that co-operated with competitors saw an 18.6 percent increase in innovation compared to firms in Ceramics, IT & Software, Textiles and Healthcare which all saw a negative impact (with only Healthcare's result showing statistical significance). In all industries except IT & Software, co-operation with suppliers lead to a more than 20% increase in innovation. Building off of the work of Freel and Harrison (2006), Tomlinson again shows evidence that accessing external knowledge, in the form of cooperative ties with outside firms increase the innovativeness of the firm.

Thomas c. Omer, Majorie K. Shelley and Frances M. Tice (2014) investigated "the conditions in which well-connected directors and boards affect firm value." Omer et al (2014) argued that board members who are on multiple boards "have better access to information… [and to] other directors' experiences" which may "facilitate their monitoring and advising [of the firm]" (Omer et al, 2014, p 17). The study measured the impact on firm value (proxied using Tobin's Q) by four different measures of the board of directors' connectedness. These included (1) the number of direct connections between directors, (2) the eigenvector measure (the strength of indirect links through weighting of direct connections), (3) the closeness of directors (the shortest path between connected directors), and (4) the betweenness of a director (a measure of how much information and/or access to resources a director may have). Testing several models using estimations techniques that included ordinary least squares (OLS) and seemingly unrelated regressions (SUR) and tests for endogeneity, Omer et al (2014) found that an increase in the connectedness of a board (a composite of the four measures above) lead to an increase in the market capitalization of the firm in the following period. The study offered an example that moving from the 25th percentile of connectedness to the 75th percentile would lead to an 11.4 percent higher market capitalization in the following period. In line with the findings of the other studies above, Omer et al (2014) find that an increase in networking leads to better outcomes for the firm. In this specific study, the association between the network (connectedness) of the board resulted in an increase in firm value.

Using parts of the connectedness measure developed by Omer et al (2014), Erica Mazzola, Giovanni Perrone and Dzidzsio S. Kamuriwo (2016) investigate how the interaction between interlocking directorate networks and inter-firm networks affect new product development. Founded in Social Capital (SC) theory, Mazzola et al (2016) develop a theoretical model linking the prominence (i.e. having a large network with access to more information and resources) of the board members with the number of inter-firm connections of a firm. Prominent boards serve as a filter, avoiding redundant information and verifying the usefulness of knowledge acquired from inter-firm networks for use in new product development. Using a sample of 1,890 biopharmaceutical companies, Mazzola et al (2016) investigate how the network of the firm and its board affect the number of new product developments. The results show that an interlocking directorate which is highly connected leads to an increase in new product developments (i.e. innovation). Building off the theory in the study, Mazzola et al (2016) posit that the results of the study show how prominent boards help to focus information and resources which strengthen the firm's connections to other firms. Through filtering information and evaluating information for its economic usefulness, reliability and quality, the board of directors serve to increase the innovation of a firm.

III. Theoretical Model

The theoretical model is a modification of the standard knowledge production function developed by Acs and Audretsch (1989). Using a Cobb Douglas process, Acs and Audtresch (1989) model patents production (a proxy for innovation) as a function of 3 inputs: economically useful knowledge (i.e. knowledge that can be converted from an idea into a useful invention, service or other beneficial item) held by a firm, *K*; the firm size, *FS*; and the ease of appropriability (i.e. market structure and the extent of monopoly power), *A*. According to the original derivation by Acs and Audtresch (1989), the various components of the model can be formulated as follows:

$$P = aK^{\beta_1}FS^{\beta_2}A^{\beta_3}u$$

P is the total number of patents and is the proxy for innovative output. *K* represents the sources of knowledge employed by the firm, including Company R&D, Government funded R&D and external knowledge (measured as prior education and experience of the firm's labor force). *FS* is firm size as measured by the number of people employed by the firm. *A* is ease of appropriability (extent of monopoly power) and *u* is the error term. This research will build off the base of this function by removing appropriability (due to a lack of data to proxy for this) and adding in network. Network, as discussed in the literature, will be proxied by board interlocks (i.e. average number of outside boards that directors sit on).

IV. Data and Methodology

The data for this research is collected from three sources. First, the main variable of interest, networks, is proxied as board interlock and is gathered from data on directors of public companies from 1998-2006 from the ISS database (formerly RiskMetrics) provided by Wharton Research Data Services (WRDS)[1 2]. The data on the companies' information is gathered from the COMPUSTAT database also collected from Wharton Research Data Services (WRDS). These variables include the R&D of a company, the employment, the number of board members and the Standard Industry Codes used to construct the dummy variables for the types of manufacturing industry. The patent data is gathered from the National Bureau of Economic Research (NBER) which holds an identifier to match patent data of individual publically traded manufacturing firms to the COMPUSTAT data. This research uses both pooled ordinary least squares (OLS) and fixed effects estimation techniques. As the theoretical foundation of the econometric model uses a Cobb Douglas process, the natural log is taken for all the variables.

$$lnPatents_{it} = \beta_0 + \beta_1 BoardInterlock_{it} + \beta_2 lnR\&D_{it} + \beta_3 lnEmployees_{it} + \beta_4 lnBoardMembers_{it} + \varepsilon_i$$

lnPatents is the natural log of patents output by the firm from 1998-2006. Acs and Audrestch (1989) show that patents serve as a proxy for innovation.

lnBoardInterlock is the natural log of the average number of outside public company boards the directors of company i sit. patents. Serving as the proxy for network, this is expected to have a positive effect on patent output (Huggins and Thompson, 2015; Omer et al, 2014).

[1] Details on the dataset can be found in Coles et al, 2014
[2] Descriptive statistics can be found in Appendix under Table 2

lnR&D is the natural log of the amount of money spent on research and development by a firm in a given year. It is expected to have a positive sign as many firms seek to protect their innovations created by R&D through patents (Acs and Audretsch, 1989).

lnEmployees is the natural log of the number of people employed by the firm. This sign is expected to be positive as the larger the firms have access to more resources which lead to more innovation (Acs and Audretsch, 1989).

lnBoardMembers is the natural log of the number of directors sitting on the board. The sign can be expected to be negative as bigger groups lead to less marginal effort but can also be positive as larger boards may have higher total effort (Adams, et al, 2010).

The study covers 2,561 observations across 357 different public manufacturing companies spread over 9 years from 1998 to 2006. The average year of observations is 2002. Of the firms sampled, the directors sit on an average of .94 outside public boards. This is to say that the average director is nearly involved in at least one other board obligation. In terms of research, the average firm spends $216 million on research and development, yet the median amount spent is $26 million, showing that the largest of firms spend a significant amount more than most. On average, this R&D results in an average of around 31 patents per year. Similarly, the median patents output is 1, showing that those who patent do so at a higher rate than average firms. Additional descriptive statistics can be found in Table 1 of the appendix.

V. Results

Table 2[3] shows the results of the various regressions. The results of model 1 (a pooled OLS run on all R&D and patent data) shows every explanatory variable (excluding the intercept

[3] See Appendix

and the industry dummy variables) as statistically significant at the 1 percent level. Interpreting the results of the main variable of interest, the proxy for network, shows that a 1 percent increase in the participation of directors on other public boards leads to a .81 percent increase in the number of patents of the firm. In another way, doubling the average interlock of the board leads to an 81 percent increase in the number of patents output by the firm. This is in line with the theory that the director's network has an impact on how the board interacts with the firm and therefore the firm's decision to innovate. When looking at the difference between industries (as done by Tomlinson (2010)), there is heterogeneity in terms of the number of patents output by a firm. Compared to the food industry, manufactures saw a statistically significant increase in the number of patents output in the areas of Lumber (43 percent), Chemical (35 percent), Rubber (30 percent), Fabricated Metals (44.8 percent), Machinery (66.5 percent), Electronics (94 percent) and Instruments (37.3 percent) industries. Even with the variation between industries, the impact of the proxy for network is quite large.

In an effort to ensure the robustness of the study, a two-way fixed effects model is included to see if there is any unobserved heterogeneity between firms over time. Model 2 in table 2 shows the results of the regression. All of the variables, with the exception of R&D, are statistically significant at the 1 percent level. The variable of interest, *lnBoardInterlock,* shows that a 1 percent increase in the average number of outside boards on which the directors sit, leads to a .41 percent increase in the patent output (a more subdued result than the pooled OLS). The lack of significance on the co-efficient for R&D is a highly counterintuitive result.

In an effort to investigate the lack of significance on the co-efficient for R&D, any observations in the dataset that had zero patent output and zero R&D, zero R&D with positive patent output or zero patent output with positive R&D were removed. This adjustment captures only firms who spend on R&D and who patent. This reduces the sample to 1,252 observations covering 205 public manufacturing firms. The results of the pooled OLS (model 3) show R&D and the interlock measure to have a similar impact as in the unadjusted sample of model 1. All coefficients maintain some level of statistical significance. Due to the reduction in size of the sample, several industries lose observations and are their dummy variables are excluded from the model. The remaining industries show similar results as with the pooled OLS in model 1. The two-way fixed effects model shows quite different results using the adjusted sample. As shown in model 4, the R&D variable is now statistically significant at the 1 percent level and at a magnitude where a 1 percent increase in R&D leads to a .3 percent increase in patents output. The interlock variable, however, is now just beyond statistical significance. The magnitude of the effect of the board's interlock has reduced as well to a point where a 1 percent increase in interlock leads to a .16 percent increase in patent output. The employee's variable loses significance and magnitude. At a significance level of 1 percent, the F-Test confirms that the two-way fixed effects model is a better estimator than the pooled OLS (23.64>1).

VI. Conclusions

This study has researched the impact of a firm level network variable on innovation. Specifically, the impact of the network of the board of directors, proxied as board interlock (the number of other outside boards on which a director may sit) was studied against the patent output of the firm. Based off a two-way fixed effects model, the results of this study show that the network of the board of directors does not have a statistically significant impact on the patent output of the firm. This is counter to the network theory provided by Huggins and Thompson (2010) as well as the results of a study by Omer et al (2014) on interlocking directorates, inter-firm networks and new product development. Even though the results are not as statistically significant, there is still value to research this topic. One possible explanation for the loss in significance on the interlock variable could be due to the reduction of the sample size in the adjusted dataset. An improvement on this research could encompass using a larger dataset to investigate if the interlock variable could be statistically significant. Some limitations of this study included that the network variable was limited only to a simple measure of board interlock. To improve on this, researchers should use the study by Omer et al (2014) to develop a more comprehensive network measure that includes a weighting of the value of a director's connections, not just the quantity of connections. If applied to a large dataset such as the one used in this study, future research could continue to expand on the niche this study attempted to fill. Further limitations of this research include the simple measurement of innovation. Patents serve as a limited proxy for innovation as they only cover a limited set of innovations. Further, firms patent (or do not patent) at different rates and for different reasons which may not fully capture actual innovation of a firm.

References

Acs, Z. J., & Audretsch, D. B. (1989). Patents as a measure of innovative activity. Kyklos, 42(2), 171-180.

Adams, R. B., Hermalin, B. E., & Weisbach, M. S. (2010). The role of boards of directors in corporate governance: A conceptual framework and survey. Journal of Economic Literature, 48(1), 58-107.

Bessen, J., Hall, B. H. & Thoma, G. (2012). NBER patent data project. Retrieved from https://sites.google.com/site/patentdataproject/Home

Coles, J.L., Daniel, N.D., and Naveen, L., 2014. Co-opted Boards. Review of Financial Studies 27, 1751-1796.

Freel, M. S., & Harrison, R. T. (2006). Innovation and cooperation in the small firm sector: Evidence from 'northern britain'. Regional Studies, 40(4), 289-305.

Huggins, R., & Thompson, P. (2015). Entrepreneurship, innovation and regional growth: A network theory. Small Business Economics, 45(1), 103-128.

Love J. (1999) Patterns of Networking in the Innovation Process: A Comparative Study of the UK, Germany and Ireland. RP9913. Aston Business School, Aston University, Birmingham.

Mazzola, E., Perrone, G., & Kamuriwo, D. S. (2016). The interaction between inter-firm and interlocking directorate networks on firm's new product development outcomes Journal of Business Research, 69(2), 672-682.

Omer, T. C., Shelley, M. K., & Tice, F. M. (2014). Do well-connected directors affect firm value? Journal of Applied Finance, 24(2), 17-32.

Peasnell, K. V., Pope, P. F., & Young, S. (2005). Board monitoring and earnings management: Do outside directors influence abnormal accruals? Journal of Business Finance and Accounting, 32(7-8), 1311-1346.

Rogers, M. (2004). Networks, firm size and innovation. Small Business Economics, 22(2), 141-153.

Tomlinson, P. R. (2010). Co-operative ties and innovation: Some new evidence for UK manufacturing. Research Policy, 39(6), 762-775.

Wharton research data services, COMPUSTAT database. (2016). Retrieved from https://wrds-web.wharton.upenn.edu/wrds/

Xie, B., Davidson,Wallace N., I.,II, & DaDalt, P. J. (2003). Earnings management and corporate governance: The role of the board and the audit committee. Journal of Corporate Finance, 9(3), 295-316.

VIII. Appendix

Table 1 – Descriptive Statistics – Unadjusted Sample:

Table 1 : Descriptive statistics on the unadjusted sample

Variable	Number of Observations	Mean	Median	Standard Deviation	Minimum	Maximum
Board Interlock	2,561	0.94	0.86	0.668	0	7
Board Members	2,561	7.98	8.00	2.814	1	19
Year	2,561	2,002.28	2,002.00	2.546	1998	2006
R&D (millions)	2,561	216.43	26.20	783.364	0.00	12,183.00
Food (SIC 20)	2,561	0.07	-	0.263	0	1
Tobacco (21)	2,561	0.00	-	0.028	0	1
Textile (22)	2,561	0.00	-	0.059	0	1
Apparel (23)	2,561	0.02	-	0.129	0	1
Lumber (24)	2,561	0.02	-	0.145	0	1
Furniture (25)	2,561	0.02	-	0.142	0	1
Paper (26)	2,561	0.03	-	0.170	0	1
Printing (27)	2,561	0.03	-	0.158	0	1
Chemical (28)	2,561	0.15	-	0.361	0	1
Petroleum (29)	2,561	0.01	-	0.118	0	1
Rubber (30)	2,561	0.02	-	0.145	0	1
Leather (31)	2,561	0.01	-	0.116	0	1
Stone (32)	2,561	0.01	-	0.102	0	1
Metal (33)	2,561	0.03	-	0.174	0	1
Fabricated (34)	2,561	0.03	-	0.179	0	1
Machinery (35)	2,561	0.15	-	0.357	0	1
Electronic (36)	2,561	0.17	-	0.375	0	1
Transportation (37)	2,561	0.07	-	0.261	0	1
Instruments (38)	2,561	0.12	-	0.322	0	1
Misc (39)	2,561	0.02	-	0.129	0	1

Table 2 – Regression Results

Table 2: Number of Patents in relation to Network and other determinants				
Variable	**Parameter Estimate and Standard Error**			
	Dependent Variable: lnPatents			
	All R&D and Patents		**R&D and Patents >0**	
	(1) Pooled OLS	**(2) Two-Way Fixed Effects Model**	**(3) Pooled OLS**	**(4) Two-Way Fixed Effects Model**
Intercept	-1.639	-3.046	-1.062	-3.950
	0.222***	0.563***	0.298***	0.730***
lnBoardInterlock	0.902	0.411	0.629	0.159
	0.094***	0.094***	0.120***	0.104
lnR&D	0.320	0.035	0.501	0.306
	0.018***	0.037	0.031***	0.077***
lnEmployees	0.236	0.167	0.113	0.020
	0.025***	0.066***	0.034***	0.094
lnBoardMembers	-0.447	0.282	-0.318	0.252
	0.069***	0.078***	0.087***	0.087***
Tobacco	-1.284	-	-	-
	0.946			
Textile	-0.714	-	-	-
	0.456			
Apparel	-0.253	-	-	-
	0.226			
Lumber	0.430	-	1.049	-
	0.204***		0.454***	
Furniture	-0.445	-	0.078	-
	0.209***		0.311	
Paper	0.027	-	0.877	-
	0.181		0.296***	
Printing	0.051	-	-	-
	0.191			
Chemical	0.353	-	0.569	-
	0.131***		0.207***	
Petroleum	-0.499	-	0.755	-
	0.245		0.455*	
Table continues below…				

Rubber	0.300	-	0.707	-
	0.204***		0.341***	
Leather	0.560	-	-	-
	0.246			
Stone	0.428	-	0.853	-
	0.274		0.616	
Metal	-0.153	-	-0.310	-
	0.178		0.619	
Fabricated	0.448	-	0.709	-
	0.175***		0.281***	
Machinery	0.665	-	0.887	-
	0.128***		0.204***	
Electronic	0.940	-	1.251	-
	0.131***		0.208***	
Transportation	0.109	-	0.581	-
	0.143		0.224***	
Instruments	0.373	-	0.610	-
	0.137***		0.212***	
Misc	0.278	-	0.765	-
	0.229		0.302***	
Adjusted R-Squared	0.42	0.85	0.45	0.90
Number of Obs.	2561	344/9	1252	205/9
F-Value	-	19.51	-	23.64

* ,**, and *** denote significance at the 10%, 5%, and 1% confidence levels

Table 3 – Descriptive Statistics – Adjusted Sample:

Table 3 : Descriptive statistics on the adjusted sample

Variable	Number of Observations	Mean	Median	Standard Deviation	Minimum	Maximum
Board Interlock	1,252	1.11	1.09	0.682	0	4
Board Members	1,252	7.97	8.00	2.966	1	19
Year	1,252	2002	2002	2.267	1998	2006
R&D (millions)	1,252	369.71	66.50	1,019.260	0.90	12,183.00
Employees (thousands)	1,252	23.13	7.31	41.108	0.13	364.55
Patents	1,252	61.75	12.00	169.914	1.00	1,892.00
Food (SIC 20)	1,252	0.03	-	0.174	0	1
Tobacco (21)	1,252	-	-	-	0	0
Textile (22)	1,252	-	-	-	0	0
Apparel (23)	1,252	-	-	-	0	0
Lumber (24)	1,252	0.01	-	0.080	0	1
Furniture (25)	1,252	0.02	-	0.134	0	1
Paper (26)	1,252	0.02	-	0.143	0	1
Printing (27)	1,252	-	-	-	0	0
Chemical (28)	1,252	0.20	-	0.397	0	1
Petroleum (29)	1,252	0.01	-	0.080	0	1
Rubber (30)	1,252	0.01	-	0.116	0	1
Leather (31)	1,252	-	-	-	0	0
Stone (32)	1,252	0.00	-	0.056	0	1
Metal (33)	1,252	0.00	-	0.056	0	1
Fabricated (34)	1,252	0.02	-	0.155	0	1
Machinery (35)	1,252	0.21	-	0.404	0	1
Electronic (36)	1,252	0.23	-	0.422	0	1
Transportation (37)	1,252	0.08	-	0.270	0	1
Instruments (38)	1,252	0.14	-	0.349	0	1
Misc (39)	1,252	0.02	-	0.140	0	1

SAS Code

```
libname PERM 'C:\Users\djw103\Desktop\PATENT_PERM';
libname sas 'C:\Users\djw103\Desktop\SAS';

/*proc import datafile="E:\Semester 8\Senior Project\Innovation\Patents
Approach\Data\NBER_Berkeley\patsic06_mar09_ipc.dta"
out=PERM.Patents74_06_raw;*/
/*run;*/
data patents96_06;
set perm.patents74_06_raw;
if appyear<1996 then delete;
if pdpass=. then delete;
run;
proc sort data=patents96_06 nodupkey out=sort;
 by patent;
 run;
/*proc print data=mydata;*/
/*run;*/
/*proc import datafile="E:\Semester 8\Senior Project\Innovation\Patents
Approach\Data\NBER_Google.Sites\dynass.dta" out=PERM.dynass;*/
/*run;*/
data dynass;
set perm.dynass;
keep PDPASS gvkey1;
run;
/*proc means;*/
/*run;*/

proc sql;
create table patents96_06_combined as
select * from patents96_06, dynass
where dynass.pdpass=patents96_06.pdpass;
quit;
/*proc sort data=patents96_06_combined out=sort;*/
/* by gvkey1;*/
/* run;*/
/*proc sort data=patents96_06;*/
/*by pdpass;*/
/*run;*/
/**/
/*DATA merged; */
/*MERGE patents96_06 dynass; */
/*BY pdpass; */
/*run;*/

data merged;
set patents96_06_combined;
if patent = . then delete;
if gvkey1 = . then delete;
run;
proc datasets lib=work nolist nodetails;
  delete patents96_06_combined;
quit;
proc sort data=merged nodupkey;
by patent;
```

```
run;

proc sql;
create table counted as
select gvkey1, appyear, count(appyear) as
patents from merged
group by appyear, gvkey1;
quit;

/* Directors Data */
/* Renna note: 1 to n in proc merge*/

data company_info;
set sas.directors_legacy;
if DIRECTOR_DETAIL_ID = . then delete;
if OUTSIDE_PUBLIC_BOARDS = . then delete;
keep TICKER YEAR RT_ID CUSIP NAME DIRECTOR_DETAIL_ID FULLNAME CLASSIFICATION
INTERLOCKING OUTSIDE_PUBLIC_BOARDS;
run;
data compustat;
set sas.compustat_pull_recent;
run;
data company_info;
set company_info;
OUTSIDE_PUBLIC_BOARDS_new = OUTSIDE_PUBLIC_BOARDS+0;
interlocking_new=interlocking+0;
if classification='I' then independent=1;
else independent=0;
run;

proc means data=company_info noprint mean;
var OUTSIDE_PUBLIC_BOARDS_new independent interlocking_new;
class TICKER year ;
output out = avg_var mean(OUTSIDE_PUBLIC_BOARDS_new)=avg_boards
mean(independent)=independent_avg mean(interlocking_new)=interlocking_avg
N(OUTSIDE_PUBLIC_BOARDS_new)=number_obs;
run;

data avg_var;
set avg_var;
if TICKER = '' then delete;
if year = . then delete;
drop _type_ _Freq_;
run;

proc sort data=avg_var;
by TICKER year;
run;

proc sql;
create table regression as
select * from avg_var, compustat
where avg_var.TICKER=compustat.TIC and avg_var.year=compustat.FYEAR;
quit;
data regression;
set regression;
GVKEY1=GVKEY+0;
```

```
run;
/*proc sql;
create table regression_patent as
select * from regression, counted
where regression.GVKEY1=counted.gvkey1 and regression.FYEAR=counted.appyear;
quit;*/


proc sort data=regression;
 by gvkey1;
 run;
proc sort data=counted;
by gvkey1;
run;

data counted;
set counted;
rename appyear=fyear;
run;

DATA regression_patent;
MERGE regression counted;
BY gvkey1 fyear;
run;

data regression_patent;
set regression_patent;
if cusip = . then delete;
if patents = . then patents='0';
if xrd = . then xrd='0';
/*if capx = . then capx='0';*/
if emp = . then delete;
/*if revt = . then delete;*/
run;

data regression_patent;
set regression_patent;
SIC2=int(sic/100);
run;
/*data regression_patent;
set regression_patent;
if 20=<SIC2<40 then Manufacturing=1; else Manufacturing=0;
if 60=<SIC2<70 then FinInsReal=1; else FinInsReal=0;
if SIC2=80 then Health=1; else Health=0;
if SIC2=87 then Eng=1; else Eng=0;
run;
*/
data regression_patent;
set regression_patent;
if 20=SIC2 then Food=1; else Food=0;
if 21=SIC2 then Tobacco=1; else Tobacco=0;
if 22=SIC2 then Textile=1; else Textile=0;
if 23=SIC2 then Apparel=1; else Apparel=0;
if 24=SIC2 then Lumber=1; else Lumber=0;
if 25=SIC2 then Furniture=1; else Furniture=0;
if 26=SIC2 then Paper=1; else Paper=0;
if 27=SIC2 then Printing=1; else Printing=0;
```

```
if 28=SIC2 then Chemical=1; else Chemical=0;
if 29=SIC2 then Petroleum=1; else Petroleum=0;
if 30=SIC2 then Rubber=1; else Rubber=0;
if 31=SIC2 then Leather=1; else Leather=0;
if 32=SIC2 then Stone=1; else Stone=0;
if 33=SIC2 then Metal=1; else Metal=0;
if 34=SIC2 then Fabricated=1; else Fabricated=0;
if 35=SIC2 then Machinery=1; else Machinery=0;
if 36=SIC2 then Electronic=1; else Electronic=0;
if 37=SIC2 then Transportation=1; else Transportation=0;
if 38=SIC2 then Instruments=1; else Instruments=0;
if 39=SIC2 then Misc=1; else Misc=0;
run;
data regression_patent;
set regression_patent;
if 20=<SIC2<40 then Manufacturing=1; else Manufacturing=0;
/*if 60=<SIC2<70 then FinInsReal=1; else FinInsReal=0;*/
/*if SIC2=80 then Health=1; else Health=0;*/
/*if SIC2=87 then Eng=1; else Eng=0;*/
run;
data regression_patent;
set regression_patent;
if Manufacturing=0 then delete;
if Manufacturing=. then delete;
run;
proc sort data=regression_patent nodupkey;
by cusip year;
run;

data regression_clean;
set regression_patent;
if xrd = . then delete;
lnxrd=log(xrd +1);
lnpatents=log(patents+1);
/*lncapx=log(capx+1);*/
lnAvg_boards=log(Avg_boards +1);
lnEMP = log(emp*1000);
/*lnREVT=log(REVT);*/
lnnumber_obs=log(number_obs);
lninterlocking=log(interlocking_avg+1);
lnindependent=log(independent_avg+1);
run;
/*proc means data = regression_patent;*/
/*run;*/
/*proc corr data = regression_patent;*/
/*run;*/

ods pdf file = 'C:\Users\djw103\Desktop\20160723.pdf';
proc means data = regression_clean N Mean Median STD Min Max;
run;

/*Count the firms*/
proc sql;
create table counted_all as
select gvkey1, count(gvkey) as
Yearcount from regression_clean
group by gvkey1;
```

```
quit;

proc reg data=regression_clean;
/*model lnpatents = lnAvg_boards lnxrd lncapx lnemp lnnumber_obs
manufacturing lninterlocking lnindependent;*/
model lnpatents = lnAvg_boards lnxrd /*lncapx*/ lnemp lnnumber_obs  Tobacco
      Textile     Apparel     Lumber      Furniture   Paper Printing
      Chemical    Petroleum   Rubber      Leather     Stone Metal Fabricated
      Machinery Electronic    Transportation    Instruments Misc
/*manufacturing FinInsReal Health Eng*/;
model lnpatents = lnAvg_boards lnxrd /*lncapx*/ lnemp lnnumber_obs;
run;
quit;

proc datasets lib=work nolist nodetails;
  delete _doctmp:;
quit;


/*data test;*/
/*set regression_clean;*/
/*run;*/
/*proc sort;*/
/*by EMP;*/
/*run;*/


/* Prepare and run the two-way fixed effects model*/
proc means data=regression_clean noprint;
var gvkey1;
class gvkey1;
output out = cusip_count N=cusip_count;
run;
data cusip_count;
set cusip_count;
if gvkey1 = '' then delete;
drop _type_ _Freq_;
run;
proc sql;
create table regression_cusip_count as
select * from cusip_count, regression_clean
where cusip_count.gvkey1=regression_clean.gvkey1;
quit;
data regression_cusip_count;
set regression_cusip_count;
if cusip_count<2 then delete;
label cusip_count=cusip_count;
run;
proc panel data=regression_cusip_count plots=none;
id cusip year;
model lnpatents = lnAvg_boards lnxrd /*lncapx*/ lnemp lnnumber_obs /fixone;
model lnpatents = lnAvg_boards lnxrd /*lncapx*/ lnemp lnnumber_obs /fixtwo;
run;
quit;
proc datasets lib=work nolist nodetails;
  delete _doctmp:;
quit;
```

```
/*End Two-Fixed Effects model work, no lag*/

/*Start Two-Way Fixed Effects model, with lag*/
/*data regression_cusip_count2;*/
/*set regression_cusip_count;*/
/*laglnxrd1=LAG(lnxrd);*/
/*laglnxrd2=LAG(laglnxrd1);*/
/*laglnxrd3=LAG(laglnxrd2);*/
/*run;*/
/**/
/*data regression_cusip_count2;*/
/*set regression_cusip_count2;*/
/*if year<2001 then delete;*/
/*run;*/
/*proc means data=regression_cusip_count2 noprint;*/
/*var gvkey1;*/
/*class gvkey1;*/
/*output out = cusip_count2 N=cusip_count2;*/
/*run;*/
/*data cusip_count2;*/
/*set cusip_count2;*/
/*if gvkey1 = '' then delete;*/
/*drop _type_ _Freq_;*/
/*run;*/
/*proc sql;*/
/*create table regression_cusip_count2 as*/
/*select * from cusip_count2, regression_cusip_count2*/
/*where cusip_count2.gvkey1=regression_cusip_count2.gvkey1;*/
/*quit;*/
/*data regression_cusip_count2;*/
/*set regression_cusip_count2;*/
/*if cusip_count2<2 then delete;*/
/*label cusip_count2=cusip_count;*/
/*run;*/
/*proc panel data=regression_cusip_count2;*/
/*id cusip year;*/
/*model lnpatents = lnAvg_boards lncapx lnemp lnnumber_obs lnxrd laglnxrd1
laglnxrd2 laglnxrd3 /fixone;*/
/*model lnpatents = lnAvg_boards lncapx lnemp lnnumber_obs lnxrd laglnxrd1
laglnxrd2 laglnxrd3 /fixtwo;*/
/*run;*/
/*quit;*/
/*proc datasets lib=work nolist nodetails; */
/*  delete _doctmp:;*/
/*quit;*/
/*End Two-Way Fixed Effect Lag work*/


/*Start correct for negative R&D*/
data regression_cusip_count3;
set regression_cusip_count;
if xrd>0 and patents=0 then delete;
if patents>0 and xrd=0 then delete;
if xrd=0 and patents=0 then delete;
run;

proc means data=regression_cusip_count3 noprint;
```

```
var gvkey1;
class gvkey1;
output out = cusip_count3 N=cusip_count3;
run;
data cusip_count3;
set cusip_count3;
if gvkey1 = '' then delete;
drop _type_ _Freq_;
run;
proc sql;
create table regression_cusip_count4 as
select * from cusip_count3, regression_cusip_count3
where cusip_count3.gvkey1=regression_cusip_count3.gvkey1;
quit;
data regression_cusip_count4;
set regression_cusip_count4;
if cusip_count3<2 then delete;
label cusip_count3=cusip_count3;
run;

proc means data=regression_cusip_count4 N Mean Median STD Min Max;
run;

/*Count the firms*/
proc sql;
create table counted_all2 as
select gvkey1, count(gvkey) as
Yearcount from regression_cusip_count4
group by gvkey1;
quit;

proc reg data=regression_cusip_count4;
model lnpatents = lnAvg_boards lnxrd /*lncapx*/ lnemp lnnumber_obs  Tobacco
      Textile     Apparel     Lumber      Furniture    Paper Printing
      Chemical    Petroleum   Rubber      Leather      Stone Metal Fabricated
      Machinery Electronic    Transportation    Instruments Misc
/*manufacturing FinInsReal Health Eng*/;
model lnpatents = lnAvg_boards lnxrd /*lncapx*/ lnemp lnnumber_obs;
run;
quit;

proc panel data=regression_cusip_count4 plots=none;
id cusip year;
model lnpatents = lnAvg_boards /*lncapx*/ lnemp lnnumber_obs lnxrd
/*lninterlocking lnindependent*//fixone;
model lnpatents = lnAvg_boards /*lncapx*/ lnemp lnnumber_obs lnxrd
/*lninterlocking lnindependent*//fixtwo;
run;
quit;

proc datasets lib=work nolist nodetails;
  delete _doctmp:;
quit;
/*End correct for negative R&D*/


ods pdf close;
```