Do Founders Matter?

Leadership Structure and New Firm Performance

By

Eszter Pázmándi

Submitted to

Central European University

Department of Economics

In partial fulfillment of the requirements for the degree of Master of Arts in Economic Policy in Global Markets

Supervisor: Gábor Békés

Budapest, Hungary

2017

Abstract

The overall objective of the thesis is to find enough evidence on the existence of causal relationship between leadership change in newly-founded enterprises and their performance. The used sample consists 13 516 Hungarian startups incorporated between 2005 and 2007. I test the effect of having the founder as CEO and examine the consequences of replacing the founder-CEO on different performance measures and survival prospects of firms. I carry out Difference-in-Difference regressions to see whether the shift to professional management has caused any change in the performance of firms after the transition. Short-and long-term survival chances of the startups are analyzed with linear probability models. I find that there are significant differences in performance measures and in long-term survival prospects among the examined startups, according to their leadership style. Corporations which went through a founder-CEO replacement show a positive and remarkable improvement both in their performances and survival probabilities. Despite that my results are statistically significant, I can not claim that there is an exclusive causal relationship between the replacement of founder-CEOs and performance improvement of startups, neither can exclude the possibility of a reverse causality.

Acknowledgements

I would like to express my gratitude to my supervisor, Professor Gábor Békés. He was always available for guidance and provided me valuable feedback throughout the whole thesis writing process.

Table of contents

List of tables

List of figures

1. Introduction

Why is it relevant to deal with the founders of startups from the aspect of firm performance? Why would they matter? In new ventures, the entrepreneur is the critical actor shaping the firm and its early decisions can have an "imprinting" effect, especially if they remain the Chief Executive Officers (CEO) for long enough (Boeker 1989, Wasserman 2006). Because entrepreneurs usually pursue opportunities they often lack key resources which are needed to build a successful organization (Wasserman 2006). That is where the academic field of management and leadership studies become applicable.

Entrepreneurship and small business research in management studies have become one of the fastest-growing academic literature, especially in the field of startups; newly-founded, fast-growing enterprises. The interest for this topic has also risen among decision-makers and politicians, as they started to see the potential in these types of businesses. Some of them see these types of businesses as the pool of potential voters, since between 70% and 95% of all firms are micro-enterprises in most of the countries worldwide. Others look at them as the solution for employment problems, because small-and medium-sized enterprises (SMEs) account for 60-70% of the employment in most countries (OECD 2016). However, one thing is sure, that the small business sector plays an important role in national economies, thus it is in everybody's interest that they become successful.

Typical factors that scholars have found as drivers of growth for startups: 1. industry; 2. governance; 3. firm characteristics; 4. management -particularly in the case of newly-founded firms: the impact of the founder-CEO (Dyer 2006). From these factors my thesis concentrates on the influence of management and leadership structure on firm performance. Good management is tightly linked to improved firm performance, measure in terms of productivity, profitability, growth and survival (Bloom et al. 2011). I define a founder-CEO as a founder

and an owner of a company, who takes the role of being a CEO. They are going to be in the center of my thesis research.

I employ a large and unique database that contains information on the universe of Hungarian firms, from which I selected companies according to their years of establishment. Displayed firms in the thesis were incorporated between 2005 and 2007. The accounting data runs until 2015, thus firms in the used database are between zero and ten years old. Consequently, all the 13 516 companies which are represented in the data can be defined as startups.

The main research question and the objective of the thesis focus on the exploration of whether there is any causal relationship between leadership structure and performance of newly-founded firms. I ask whether founder-CEOs have a causal effect on firm performance. More specifically, I am looking for answers for the following questions to find proof for causality: 1. Does startup performance can be associated with the type of its leadership structure?; 2. What is the magnitude of this association?; 3. Does this show causal relationship?. First, in order to discover the direction and the basic characteristics of the relationship I run OLS regressions on different firm performance measures, for different years in the data. After revealing that the duration of a founder-CEO plays a determinant role in the correlation, I run difference-in-difference regressions to see what happens to firms' performance after the replacement of the founder-CEO and to properly accommodate causality. As an additional feature, I also apply a linear probability model on survival probabilities of firms with founder-CEOs. I expect to find negative correlation between having the founder as CEO, but a positive change in performance after he or she had been replaced. While there are many international studies on this topic, in the Hungarian literature it is rare. The special characteristics of the used database (Appendix B.1) and the exclusivity of the data is what make most of the uniqueness of my thesis.

In the next chapter I discuss the previous findings on this topic and based on them I propose my main hypothesis. In chapter 3, 4 and 5, I explain how the final sample was developed, introduce

the main variables and describe the summary statistics. In chapter 6 I analyze the effect founder-CEO leadership from different aspects. Chapter 7 concludes.

2. Theoretical framework

In this section, first I narrow down the research topic then I try to summarize the main findings on the examined topic. First, from the definition of family businesses I derive why my thesis concentrates on startups rather than family firms. Then I introduce some studies and findings about the relationship between firm performance and having the founder both as the owner and the Chief Executive Officer of a company.

When talking about founder-CEO led companies the first thing that comes to mind is to look at performance of family businesses and compare their performance to non-family run enterprises. In family businesses the founder plays a major role in the life of the company and remains CEO for a very long time. From this aspect it would be reasonable to compare the financial performance of family businesses versus non-family businesses.

In the literature founder-led companies are often referred as family businesses, however privately owned and led firms are not the same as family businesses. The European Commission uses a rather strict definition of family business, to which I relate in my thesis.

According to this definition (Common European definition of a family business[1]), I cannot narrow down my topic to family businesses, because my dataset is not capable of providing information on the third point of the definition, which states that at least one of the representatives of the family is formally involved in the governance of the firm. It is quite uncertain whether there is some kind of family connection among the founders or among the forthcoming CEOs and founders.

[1] "1.The majority of decision-making rights are in the possession of the natural person(s) who established the firm, or in the possession of the natural person(s) who has/have acquired the share capital of the firm, or in the possession of their spouses, parents, child, or children's direct heirs.; 2. The majority of decision-making rights are indirect or direct.; 3. At least one representative of the family or kin is formally involved in the governance of the firm.; 4. Listed companies meet the definition of family enterprise if the person who established or acquired the firm (share capital) or their families or descendants possess 25 per cent of the decision-making rights mandated by their share capital." (European Commission Family Business, www.ec.europa.eu.).

That is why my thesis rather focuses on newly-founded firms, than family businesses and compare the performance of founder-led startups to professionally managed ones from the beginning.

2.2 Contradictory studies about the topic

Empirical studies in the separation of ownership from control started with the publication of Berle and Means' (1932) and Fama and Jensen's (1983) works. Since then, there have been several publications, which rather concern the effects of separation on firm performance. The role of CEOs as organizational leaders has also long been a popular topic for discussion and recently when evaluating success factors of -especially newly-founded -firms the importance of their owner- and leadership structure has got more emphasis.

Naturally, there are differing perspectives on the importance and the impact of CEOs on a firm's financial performance. There are the ones which argue that a CEO would have a determinant impact. These differing aspects are probably related to firm size, because in larger firms CEOs may have a smaller effect on outcomes mainly because of the complexity of the organization (Fama and Jensen 1983). Such observations provide evidence that the linkage between the role of the CEO and firm performance is more easily observed in small firms. Most of these studies have attempted to ascertain the impact and the contribution of a founder can make on firm performance. Basically, all the contradictory studies can be originated from the agency theory. The ones which found negative relationship emphasize the agency costs, while the others the agency benefits.

So, why is it reasonable to think that founder-led firms may perform better than their peers? Several studies report a positive effect of founder-CEO firms on firm performance (Adams, Almeida and Ferreira 2009; Fahlenbrach 2009). One possible reason for this discrepancy results

comes from the type of firms used in the analyses. The positive effect of founders is typically derived from the analysis of public US enterprises, which may have implemented standardized managerial practices.

Fama and Jensen (1983) noted that "*concentration of specific information relevant to decision implies that there are efficiency gains when the rights to manager and control decisions are combined in one or a few agents."* (Fama and Jensen 1983, 307). Small firm managers may be less constrained in their actions, thus could achieve a more significant influence on overall performance (Dalton and Kesner 1983). Begley and Boyd (1986) measured growth rate, profitability and return on investment for founder-operated firms, and found that founders have a significant and positive effect on firm performance. Additionally, the small organization tends to be noncomplex and characterized by concentrated leadership and decision-making control (Daily and Dalton 1992). Furthermore, there are studies which prove that due to different values and attitudes between professional managers and the organization, professional managers tend to behave differently than founders. This argument is especially important in the case of small and newly-founded firms. Professional managers' personal stakes in the business are not as strong, thus behave differently than founders (Daily and Dalton 1992).

Studies which emphasize the negative impact of a founder-CEO on firms' performance are focusing on the management skills of founders. A study by Bloom, Sadun and Van Reenen (2011) compares average management quality of companies broken down by various types of ownership. They found that founder owned enterprises with founder-CEOs are the worst managed ones, even when controlling for a host of other variables, such as age, size, industry, skills and so forth.

Bennett, Lawrence and Sudan (2015) using the World Management Survey prove that lower management skills of founder-CEOs are strongly associated with significant performance differentials. They propose two possible reasons for the managerial gap of founder-CEO firms:

1. informational problems preventing a clear understanding of the weakness of their firms' managerial practices; 2. non-pecuniary returns to potentially inefficient but power preserving practices.

From the aspect of my thesis it is also relevant to mention some findings on what happens to founder-CEO firms after they switch to professional management. In general, firms that experiencing founder-CEO replacement show significantly better performance results after the change of leadership (Bennett, Lawrence and Sudan 2015). Noam Wasserman (2006) while studying the "founder's dilemma" in startups, found that founders maximized the value of their firms by giving up the CEO position and board control.

Based on the characteristics of my dataset and the attributes of the firms in the sample, my main hypothesis is the following:

I. Founder-CEOs have a negative impact on newly established firms' performance.

Beyond that, to sufficiently prove my main concepts, three additional sub-hypotheses were set up, according to which the empirical analysis was conducted:

i. Founder-led startups have worse performance measures, on average;
ii. The replacement of the founder-CEO contributes to better firm performance after the change;
iii. Founder-CEOs and their replacement affect survival prospects of startups.

3. Data

The obtained multidimensional panel data contains companies over multiple time periods. The panel is unbalanced, not all companies are observed in all years. This unbalanced character of data mainly comes from the fact that I selected companies which were established between 2005 and 2007, not all observed companies may lived until the end of the chosen time period, and in addition not every company has started to submit their financial reports in their exact year of formation. The selected sample only contains CEO-led companies, to ensure that the measured effects are the ones which come from personal effects and the lack or the replacement of the CEO-founder indicates meaningful differences and variations.

The overall strategy when the final sample was created was to come up with constrains regarding firm size and performance to achieve the one of the research objectives: study financial performance and management structure of fast-growing small-and medium sized enterprises. I wanted to keep as many usable companies as possible and thus applied a conservative method when setting up the criteria regarding firm characteristics.

The thesis studies companies between 2005 and 2015. Companies were selected into the sample according to several eligibility criteria.

First, I chose companies which were established in either 2005, 2006 or 2007. My choice of time period based on several reasons. These mainly come from the features of the used database. In the original datasets there were companies which earliest year of foundation was 1907, however in the datasets of the owners and the leaders the earliest reported year was 1989 for the first owners or leaders of the companies. Consequently, I was not able to identify the founders of all the companies in the initial datasets, which caused gaps in my data. Moreover, the financial reports of the firms were available only from 2005 and until 2015. So, in overall the financial reports file's capacity determined my final choice on the analyzed time period.

After, I set up some eligibility criteria for firms to be selected into the final sample. These criteria were based on the main attributes of small-and medium-sized enterprises and startups: their annual net sales revenue and number of employees.

As a basic principle I applied the European Commission's regulations on SMEs. The EU recommendation 2003/361. determines the definition of a small-and medium-sized enterprise, which was also implemented in the Hungarian national law system. It sets the limits of net annual sales and number of employees for firms to be able to call themselves small-and medium-sized enterprises. The maximum net annual revenues cannot be more than 50 million euros (around 15.5 billion HUF, exchanged at the central rate) and SMEs cannot have more than 250 employees. The law does not restrain the minimum number of employees or the minimum amount of net annual sales. Since, my research is based on the performance of startups, ergo fast-growing companies, I did not exclude all the companies which had more than 50 million euros in revenues.

First, companies were evaluated according to their annual net sales. More specifically, their first-year annual net sales were evaluated after setting up a minimum limit of 15 million HUF of annual revenues. If throughout the whole operation period, a company has never reached this amount, it was excluded. Companies with very high first year revenues were also excluded from the sample. This meant, that 24 companies were dropped initially, which had unusually high first-year revenues (Appendix A.1 and A.2). Out of the 24 companies 18 had tax arrears from their year of formation. They were mainly in the Wholesale and retail trade, Manufacturing and Information and communication, which tend to have above average net annual sales. Most of these (17 out of the 24 companies) were foreign-owned and most of them operated in the Manufacturing and the Wholesale and retail industries, which does not belong to a typical SME sector.

Regarding the lower bound, one additional criterion was applied. The goal was to avoid the so-called "ghost companies" to get into the sample, because they cannot be called an enterprise or a startup regarding their traditional definition. The following approach was used: a company was dropped from the sample if after four years of operation it still had a zero or negative income. Moreover, there were companies with zero or negative income even before their year of caseation. In these cases the year of their death was modified to the first year when their revenues turned zero or below zero.

Furthermore, there were probably project-based companies which net annual sales were very volatile. They only had positive revenues in different years, while in the meantime they were operating with zero income. Accordingly, these were dropped if after their first three years, still did not have any positive income.

The similar rules were applied regarding the number of employees of a firm, which is also another important growth indicator. According to the EU Recommendation 2003/361, the maximum number of employees for an SME is no more than 250 employees. The regulation does not set up a lower bound regarding employment. Since employment is a growth indicator and I study the overall performance of startups, it was not allowed for me to keep companies which do not grow over the years. As a result, if after three years of operations the employment number of a company was still below two, it was dropped. In addition, not all companies with more than 250 employees were discarded. If a company had more than 250 employees only in its last three years, it was kept. On the other hand, if there are companies with more than 250 number of employees in their first three years of operations, these are probably not typical startups. Accordingly, four companies were dropped because of outstanding number of employees in the beginning. These had more than 250 employees in one of their first three years.

When selecting companies according to their industry categories, the main goal was to represent common small business sectors. So, the National Statistics Office's classification categories were used, thus companies in the Mining, Agricultural, Energy; Water and waste and the Financial sectors were dropped from the sample.

In sum, 13 516 companies were selected in the sample, from which 3 562 companies were established in 2005, 4 155 in 2006 and 5 799 in 2007 (Table 1).

4. Key variables

4.1 Outcome variables

To reveal whether having founder-CEOs affect firm performance and firm succession, I used three key independent variables to measure performance and size and two main dummy variables to determine survival probability of firms. Looking at the distribution of the performance variables, I found that all were highly skewed (Appendix A.3), thus I took the natural logarithm of them, except for the binary variables.

In the related literature, we can find several different examples to measure firms' financial improvement and firm size. Using Cochran and Wood (1984) typology, due to the limitations of the dataset and the characteristics of the chosen firms, accounting measures were taken into consideration as financial development indicators. More specifically, companies were evaluated according to their annual total net sales values, as one of the most important financial index of firms. I chose revenues, because that is the most easily interpretable and understandable financial performance measure. It was reported in good quality compared to other accounting measures in the data, the submitted profit-and-loss statements contain much fewer missing values for the revenues than for any other financial indicator. Thus, the probability of losing important information due to data transformation was mitigated this way. Additionally, the aggregated average annual revenues are also can be used as a total lifetime growth index of enterprises.

As the measure of firm size, I chose the yearly number of employees indicator, mainly because of the same reasons as with the total revenues. Furthermore, I determined the basic eligibility criteria for the firm size based on employment previously.

To determine survival rate of companies two dummy variables were applied according to firm age. Short-and long-run survival probabilities were estimated. Accordingly, the short-run

binary variable gave the value of one for every company that lived until 3 years. Long-run survival was determined in being alive at least for 8 years. When deciding about which years to consider during the evaluation of survival chances two major economic shocks were needed to be taken into consideration: the 2008 crisis and the introduction of the new Civil Code in 2015. Due to both shocks a huge proportion of companies went out of business in those years, which would have caused biased estimates. That is why not a given year, but instead firm ages were considered for companies which were established in different years.

4.2 The founder-CEO dummy

I use one main binary variable as the main independent variable in my regressions, to determine the type of the leadership structure of a firm, according to whether it is founder-led or not. Within this group one additional binary variable is used to differentiate between founder-led companies according to a change in their management. This binary takes the value of one when the founder-CEO was replaced any time during the examined time period and zero if the founder has remained the top manager for the entire time.

The dependent variable determines whether a firm has its founder as CEO or not. The applied dummies do not take into consideration the number of years for which a founder was CEO, only captures whether the change happened or not. The founderCEO dummy was created from the raw datasets (Appendix B). After the merge of the datasets (Appendix B.2.3), I looked at the owners and the CEOs of enterprises in their year foundation. If the starting date of the CEO matched with the starting date of the owner, then that particular owner was the founder and the CEO of the company and got the value of one for being the founder-CEO.

In order to differentiate among companies with their founders as CEOs two additional variables were created. The *duration* variable shows the number of years for which a founder led its

company and the *change* variable shows the year when the founder was replaced. These two variables were important when I was evaluating firm performance after the replacement of the founder-CEO.

Table 1 shows the proportion of companies according to their year of formation and their leadership structure, whether the founder was replaced or not. One can see that within the group of companies with founder-CEOs the proportion of companies with and without the replacement differs from each other according to the year of their formation. In sum 19.1 percent of the selected companies have replaced their founders from the leader position, while 13.4 percent of the companies kept their founders as the Chief Executive Officer of the firm. Comparing the proportions of companies within the group of founder-CEOs, we can see that the companies which were established in 2006 had the highest rate in the replacement of their founders. If we compare these companies according to the exact year of the replacement we can conclude that most of the replacements (more than 60 percent of the replacement among all companies according to their year of formation) happened after the 2008 crisis which is quite understandable. One would assume that younger companies due to the crisis would have a higher probability of replacing their leaders because of the crisis. But the sample shows the highest rate of replacements in 2009 were among the enterprises which were established in 2005.

Table 1: Number of companies according to their leadership structure

Year of formation	Companies w/o founder-CEOs	Companies w/ founder-CEOs		Total
		Companies w/ founder-CEOs replacement	Companies w/o founder-CEOs replacement	
2005	3266 (91.7%)	161 (4.5%)	135 (3.8%)	3562
2006	3059 (73.6%)	835 (20.1%)	261 (6.3%)	4155
2007	2793 (48.2%)	1596 (27.5%)	1410 (24.3%)	5799
Total	9118 (67.5%)	2592 (19.1%)	1806 (13.4%)	13516

Note: Companies are grouped by their leadership structure. Every row shows the number of companies according to their leadership structure and year of foundation. The percentages in brackets indicates the proportion of companies by their years of foundation.

4.3 Control variables

Without the ability to run experiments one can try to control for confounders in regressions to get closer to the causal effect even if no one can be sure about how close the coefficients are to true causal effect. When selecting the possible confounders the main goal was to sufficiently control for the effects of the various variables to truly isolate the impact of the founder on firm performance and to rule out alternative determinants of the performance of the sampled firms.

The used control variables can be divided into two categories. The ones which observe firm characteristics and the ones which observe leader-and ownership characteristics of firms. The applied control variables are dummies in most of the cases, thus can take the value of one or zero, based on whether the examined firm has that attribute or not. The other types of control variables are either factor or continuous.

The following confounders were used. Foreign ownership dummy to detect whether a company has ever had a foreign owner from 2005. A foreign leader dummy, which is also dated from 2005 and takes one if a company has ever had a foreign leader. A company ownership dummy

and an additional factor variable to indicate ownership type. If a company is owned by another the company ownership dummy equals to one. The ownership type factor variable represents the following ownership categories: privately held, company-owned or mixed (company and private) ownership. Firm age represents the age of the company by year. The number of owners and CEOs variables controls for firm size. It shows the number of owners or CEOs of every company according to years. The gender of the CEO or the founder-CEO is a binary, with the value of one if the CEO is a male and the value of zero if it is female. The age of the CEO and the founder was calculated from their years of birth. The industry controls were generated from the 4-digit Hungarian TEAOR codes and companies were classified into nine different categories according to their industry codes. Furthermore, year dummies were also used as well as year of foundation dummies, to differentiate between companies according to their years of establishments.

5. Descriptive statistics

First the general characteristics of the firms in the sample are shown, then the basic comparison of founder-managed versus professionally managed firms. The whole sample consists of 108 309 observations. Each observation represents a company in a year.

Table 2 represents the number of companies which are still alive in 2015. The table also shows the number of companies according to their leadership structure. One can see that 75.69 percent of companies are still alive in 2015, which means that from the initial 13 516 companies 10 230 of them did not go out of business in 2015. If we look at the proportion of companies according to their leadership structure we can see that founder-led companies have a lower probability of being alive in 2015. On average, 64.70 percent of them is still operating after their foundation, while this rate is 79.79 percent in the group of the professionally-managed companies.

Table 2: Number of companies still alive in 2015 according to their year of foundation

Year of formation		Number of companies	Number of companies in 2015
2005	Founder-led	296	174
	Professional leader	3 266	2 500
2006	Founder-led	1 096	729
	Professional leader	3 059	2 432
2007	Founder-led	3 006	2 068
	Professional leader	2 793	2 327
	Total	13 516	10 230

Note: Companies are shown by their leadership structure. The "Number of companies" column displays the number of incorporated firms by years indicated in the rows.

After looking at the proportions of surviving firms, I looked at the number of firms which went out of business according to their year of death. The data was retrieved from the raw dataset of basic information (Appendix B.1) and collects company exists from 2005 (Appendix A.4). It

shows that there was a big crisis in 2011 and in 2015 regarding the number of company exists. In that table all companies are considered, without regard to year of foundation. In 2011 and in 2015 one can see a big jump in the number of enterprise exists, compared to previous years' trend. If these numbers are compared to the number of company births, one can see that in 2012 there was a big drop and this decreasing trend kept continuing. Both changes are due to the change of the regulation environment for SMEs in Hungary. In 2011 amendments to the Hungarian Civil Code were introduced (2011 version of Act IV of 1959 on the Civil Code). As a result, the fee of the incorporation was increased by the government, the rules for establishment have become stricter and the incorporation procedure has been prolonged. Thanks to these changes, in the first round there was an "establishment wave" in the first quarter of 2012, but after that, the number of firm closures exceeded the number of newly founded enterprises (Hungarian Central Statistical Office 2016). In 2014, the regulations of the Civil Code were modified again, more specifically, a totally new Civil Code was introduced. The new Civil Code Act V. of 2013 modified the minimum amount of subscribed capital of limited liability companies. It was set to 3 million HUF from 500 thousand HUF. The new regulation had an enormous effect on operating enterprises since the limited liability company (kft.) is the most widely used form of business association in Hungary. Due to this change, a substantial number of enterprises needed to finish their operation because they were unable to satisfy the new requirement.

The table of the summary statistics of the control variables (Table 3) shows that, on average firms in the sample live up to 8 years. It is also shown that in general, firms mostly operate with one or two owners or CEOs, however there are companies with extreme number of owners or CEOs (these are mostly company-owned and foreign firms). Furthermore, one can also see that, on average, a founder-CEO leads its company for 5-6 years (5.71 years), and the quite high

value of the standard deviation of the duration of founder-CEOs (2.77) also suggests that the duration for which a founder leads its company varies within firm.

Table 3: Descriptive statistics of the control variables

Variables	Mean	Standard deviation	Median	Minimum	Maximum	10th percentile	90th percentile
Maximum firm age	8.34	1.47	8	2	10	6	10
Number of owners	2.42	2.43	2	1	156	1	4
Number of CEOs	1.48	0.76	1	1	27	1	2
Duration of a founder-CEO	5.71	2.77	6	1	11	2	9
Private ownership	0.95	0.19	0	0	1	1	1
Company ownership	0.16	0.36	0	0	1	0	1
Foreign leadership	0,02	0,12	0	0	1	0	0
Foreign ownership	0,10	0,30	0	0	1	0	1
Age of CEO	48.5	20.7	49	0	107	0	67
Gender	0.32	0.56	0	0	1	0	1

Notes: The table shows the descriptive statistics of the main control variables for the whole sample. The total number of observations is 108 309.

5.1 Comparison of firms with different leadership structure

If we compare companies according to their founder-CEO status, we can see significant differences. Professionally and founder-managed firms' performance measures show considerable variations and indicates an important distinction between the two types of firms.

Table 4 shows average performance of companies specified by their leadership categories between 2005 and 2015. According to the table, one can conclude that professionally-managed firms are tend to show larger financial performance measures. Annual net sales, earnings before income tax (EBIT) and equity values of firms with professional CEOs are significantly larger. Annual net sales are almost twice as large for these kinds of firms. If we look at long-term

liabilities of founder-managed firms, we can see that, on average these firms have more liabilities and their values vary more (their standard deviation is larger), which indicates a more volatile credit behavior among these firms.

Looking at the values of number of employees, which refers to firm size, one cannot see as big differences, on average, among the different categories. Founder-CEO firms have an average of 6.8 number of employees, while professionally managed firms have 8. This suggests that, on average even with the same number of employees (their standard deviations are also very close: 15.3 for founder firms and 16.7 for professional firms), the productivity of firms with a professional manager from the very beginning is better, they can produce more value with the same number of people. Thus, they capitalize their resources in a more efficient way.

Looking at some control variables of these firms, we can conclude that founder-managed firms have more CEOs. This is indicated by the mean number of CEOs for these firms and the maximum number of CEOs. The mean number of CEOs is 1.43 for firms with a professional manager and 1.59 for founder firms, while the maximum number of CEOs is 14 and 27, respectively. Regarding the maximum firm age variable and the dummy which shows whether the firm is still alive in 2015, one can deduce that non-founder companies tend to have better survival chances on the long-run. 85 percent of these companies are still alive in 2015, while only 75 percent of founder firms are still operating in 2015.

There is no difference between the two types of firms according to the gender of the CEO and whether they are foreign-owned or not. Majority of firms are male-led and not foreign-owned. The results of the CEO age variable let us assume that the older the CEO, the more proficient and well-run the managed firm is. Moreover, one can claim that professionally managed firms tend to have older leaders. As a matter of fact, several papers and articles have recently proved (Johnson 2013, Freedman 2012, Wadhwa 2013) that stories about younger entrepreneurs are just myths. The typical entrepreneur is more likely to be middle-aged. In fact, entrepreneurs

above their fifties are twice as successful as their younger peers (Wadhwa 2013). Average age of a CEO in the sample in a founder-managed firm is 42 years old, while in a professionally led company the average age is between 45 and 46 years. My results are in line with the previously mentioned findings.

If we look at the comparison of firms where the founder was replaced and where it was not, one can see that there are major differences among these firms regarding their performance measures. The mean total annual net sales of companies where the founder was replaced is 1.5 times larger, while their mean EBIT is 4.28 times larger. Regarding the financial performance of the companies where the founder was replaced, we can conclude that they outperform their peers with a stable founder-CEO. If these companies are compared to the ones, which were professionally-managed throughout the whole time, one can see that their mean values are closer to theirs. The probability of being alive in the end of the examined time period is the same for both. Their average maximum firm age is also the same, indicating that, that their survival probabilities are improved on the long-run, probably due to the change.

In sum, it can be declared that leadership and management structure matter during the assessment of firm performance. Moreover, the values of Table 7 suggest that startups, which went under a management change and replaced their founders, demonstrate larger efficiency. Their outcomes are closer to the better performing professionally-managed companies, however these results only show aggregated average performance of the examined firms. Therefore, in the next section, I will try to reveal further characteristics of the relationship between firm performance and management structure. I will try to make an attempt to discover the direction of this relationship, whether founder replacement contributes to better performance, or the other way, former better performance contributes to a change in the leadership.

Table 4: Summary statistics of firms by different leadership structure

	Statistics	Total annual net sales revenue (in mill HUF)	Number of employees	Equity (in mill HUF)	Long-term liabilities (in mill HUF)	EBIT (in mill HUF)	Number of CEOs	Maximum firm age	Alive in 2015	Gender	Age of the CEO	Foreign ownership
"Professionally-managed" firms	Mean	212.3	8	86.8	37.3	25.3	1.43	8.65	0.85	1.34	46.72	0.14
	Standard deviation	1 689	16.7	4530	2 260	2 130	0.70	1.32	0.35	0.31	19.65	0.35
	Median	44.6	3.50	6.6	0	0.9	1	9	1	1.50	49	0
	Minimum	0	0	-37 900	-34.4	-28 500	1	2	0	1.50	0	0
	Maximum	25 6380	347	72 8000	509 000	321 000	14	10	1	2	107	1
Companies without the replacement of their founder-CEO	Mean	130.5	6.24	2.2	9.4	2.9	1.61	6.96	0.59	1.34	43.54	0.11
	Standard deviation	591.7	11.48	234	81	60.6	0.94	1.92	0.49	0.30	21.98	0.32
	Median	30.46	3	3.45	0	0.4	1	8	1	1.50	48	0
	Minimum	0	0	-3 090	-1.46	-2 780	1	2	0	1	0	0
	Maximum	2 0237	175	9 500	2 990	2 120	27	10	1	2	99	1
Companies with the replacement of their founder-CEO	Mean	203.2	7.1	59.1	70.9	12.8	1.58	8.15	0.87	1.36	42.60	0.18
	Standard deviation	954.1	17.1	1 360	4 920	917	0.82	1.07	0.34	0.34	22.75	0.38
	Median	33.47	3	4.14	0	0.5	1	8	1	1.50	48	0
	Minimum	0	0	-23 100	-40	-12 200	1	3	0	1	0	0
	Maximum	45 723	371.80	175 000	728 000	134 000	9	10	1	2	93	1
Total	Mean	176.1	6.8	45.3	48	9.1	1.59	7.70	0.77	1.35	42.95	0.15
	Standard deviation	838.6	15.3	1 090	3 900	728	0.87	1.56	0.42	0.32	22.47	0.36
	Median	32.18	3	3.9	0	0.5	1	8	1	1.50	48	0
	Minimum	0	0	-23 100	-40	-12 200	1	2	0	1	0	0
	Maximum	45 723	371.80	175 000	728 000	134 000	27	10	1	2	99	1

Note: The table shows the summary statistics of firms in the sample by different leadership structure. Number of observation for "Professionally-managed firms": 72 762; for "Companies without the replacement of the founder-CEO" is 13 220; for the "Companies with the replacement of the founder-CEO" is 22 327. The total number of observations is 108 309.

6. Empirical analysis

6.1 Firm performance and founder-CEOs

To be able to answer the main research question, I looked for evidence whether there are significant performance differences between active and passive founder controlled firms. Summary statistics showed that, on average startups with different management structures have divergent outputs. Taking into consideration the overall objective of the thesis, to reveal whether ownership and management structure of startups can have an influential effect on firms' performance, I needed to test further attributes of this relationship in order to get closer to find causality.

First, I was interested in the direction and the strength of the relationship. More specifically, whether there is some kind of nonlinearity between financial performance and ownership structure as suggested by other studies (Thompson and Chen 2012; Maury 2005; Anderson et al. 2003). When comparing the total annual net sales of companies depending on whether they are managed by their founders or by a professional manager, I found that their relationship is rather linear and negatively correlated (Appendix A.5). Consequently, as a start, I decided to evaluate their relationship according to these results and apply linear regression models and examine the strength and correlation of the dependent and independent variables over the years.

6.1.1 Model

I checked whether the retention of the CEO title by the founder is correlated with firm performance and employment size in the sample. The summary statistics showed that there is a difference in the performance of founder-led versus non-founder led companies. To reveal more details about the strength and significance of this relationship, OLS regressions were run from

2011 until 2015 on firm performance measures on the left-hand side and the founder-CEO dummy as the main explanatory variable. The regressions were run with and without controls, to see if there are other factors that might affect the nature of the correlation.

The following OLS regressions was estimated for the different years:

$$E\left[Yperformance_{it} \mid founderCEO_{it}, Z_{it}\right] = \alpha + \beta founderCEO_{it} + \gamma Z_{it} \quad (1)$$

where the *Y* stands for the three used firm performance measures. β shows the expected value of the performance measure if the firm was led by its founder-CEO anytime during the examined time period. Z_{it} is the collection of the control variables. The following controls were applied: number of CEOs, number of owners, ownership type (as a company owner dummy), foreign ownership dummy, foreign leadership dummy, age of the CEO, gender of the CEO and year dummies were inserted to control for the years of foundation. The index *t* stands for the examined years (from 2011 to 2015).

When deciding about which years to test, the different years of foundation were taken into consideration. That is why not all the years of the time period are examined. Furthermore, I also had to pay attention to the two major shocks in 2008 and in 2015, which seriously hit the SME sector. So, it was useful to select the years after the crisis and also to leave some "recovery" time.

6.1.2 Results

The results are shown in Table 5. The table collects the results of the different β coefficients with and without controlling for confounders. The more detailed results of the regressions are not presented, because in the first place, I was interested in the direction and the size of the

relationship and to see how other factors influence the correlation. Firms were grouped according to their year of establishment. The results show a negative relationship, as expected, between having the founder as the leader of company and performance measures, although the coefficients are not significant in every observed case.

If one looks at the results in details, can see that, on average, coefficients increase year by year for every performance measure. Although there are differences among companies according to their years of establishment. For companies, which were established in 2005, only the results on the number of employees variable are significant. Their financial performance and productivity outcomes are negative but not significant. Furthermore, the trend of increasing coefficient size is not true in their cases as well. This can be explained by the fact these companies were hit the most by the negative consequences of the shock in 2015. (Majority of the caseated companies in 2015 were established in 2005.)

The comparison of the results according to whether controls were included or not brings major differences in the size of the outcome variables. This indicates that there are additional factors which also have an impact on the size and the direction of this relationship. Moreover, there are few cases when the results only become significant after the insertion of the controls. When, the possible control variables were selected, I found that adding the duration of a founder as a leader has intensified the relationship both in terms of coefficient size and in significance. This, and the growing size of the coefficient of the founder-CEO dummy year by year, even without the controls, let us assume that the duration of a certain leader can have a significant influence on the performance of startups.

Summarizing the results of the yearly OLS regressions, one can conclude that having a founder-CEO instead of a professional top manager modify average performance of startups. If new firms are led by their founders, their expected performance is worse than for firms with a professional CEO from the beginning. Although, the leadership structure of a new firm is not

solely explaining the worsening performance (this can be seen from the increasing size of the coefficients in the cases when controls were included in the regressions), but timing and the duration of stay still play a remarkable role in this relationship. These results formed the basis of the next section, where I will try to compare the performance of firms where the founder was replaced in the years after the change.

Table 5: Summarized results of OLS regressions on firm performance measures and the founderCEO dummy, by the years of foundation

Years of foundation	Years	Total net sales		Number of employees		Productivity	
2005	2011	-0.969***	-1.415**	-0.159**	-0.0816	-0.741**	-1.299**
		(0.335)	(0.718)	(0.0663)	(0.126)	(0.309)	(0.655)
	2012	-0.566*	-0.294	-0.170**	-0.129	-0.303	-0.0934
		(0.332)	(0.550)	(0.0712)	(0.146)	(0.302)	(0.500)
	2013	-0.460	-0.464	-0.192**	-0.158	-0.0708	-0.136
		(0.352)	(0.646)	(0.0792)	(0.149)	(0.308)	(0.600)
	2014	-0.0523	-0.274	-0.189**	-0.301*	0.391	0.449
		(0.349)	(0.696)	(0.0825)	(0.162)	(0.291)	(0.591)
	2015	-0.212	0.0528	-0.204**	-0.162	0.263	0.434
		(0.393)	(0.663)	(0.0860)	(0.171)	(0.334)	(0.576)
2006	2011	-0.809***	-1.153***	-0.107***	-0.132**	-0.659***	-0.970***
		(0.160)	(0.360)	(0.0358)	(0.0672)	(0.150)	(0.334)
	2012	-0.716***	-0.838**	-0.170**	-0.180**	-0.510***	-0.545
		(0.181)	(0.378)	(0.0712)	(0.0703)	(0.167)	(0.349)
	2013	-0.860***	-1.365***	-0.192**	-0.281***	-0.463***	-0.717**
		(0.201)	(0.424)	(0.0792)	(0.0801)	(0.170)	(0.362)
	2014	-1.257***	-1.553***	-0.189**	-0.232**	-0.819***	-0.982**
		(0.228)	(0.489)	(0.0825)	(0.0913)	(0.188)	(0.398)
	2015	-1.261***	-1.599***	-0.204**	-0.267***	-0.759***	-0.897**
		(0.248)	(0.519)	(0.0860)	(0.0945)	(0.203)	(0.423)
2007	2011	-0.670***	-0.925***	-0.107***	-0.0630	-0.588***	-0.845***
		(0.0937)	(0.185)	(0.0358)	(0.0435)	(0.0877)	(0.175)
	2012	-0.842***	-0.854***	-0.127***	-0.122***	-0.719***	-0.666***
		(0.113)	(0.197)	(0.0368)	(0.0440)	(0.106)	(0.183)
	2013	-0.690***	-0.894***	-0.176***	-0.142***	-0.397***	-0.560***
		(0.133)	(0.228)	(0.0412)	(0.0477)	(0.114)	(0.192)
	2014	-0.675***	-0.813***	-0.182***	-0.0957*	-0.383***	-0.569***
		(0.146)	(0.260)	(0.0444)	(0.0528)	(0.123)	(0.214)
	2015	-0.602***	-0.801***	-0.193***	-0.114**	-0.329***	-0.510**
		(0.151)	(0.277)	(0.0464)	(0.0556)	(0.126)	(0.226)
Controls		NO	YES	NO	YES	NO	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Results are restricted to only the coefficients of the founderCEO dummy. Each column reports the results of a firm level OLS regression of dependent variable noted at the top of each column in the years noted in the rows. Years of foundation are indicated in the very first column of the table. Controls are not included in every case: for every dependent variable, the first column represents the estimation without controls, the second with controls.

6.2 Founder-CEO replacement and subsequent firm performance

After proving that the duration of a founder remaining CEO has significance when examining firms' average performance, I was interested in how the replacement of a founder and the transition to a professional leadership structure might affect firm performance. The overall goal was to find evidence for the existence of causal relationship between founder-CEO replacement and firm performance change. Difference-in-difference estimation is a good method to compare average change of outcomes of two different groups before and after the examined intervention, which is the replacement of the founder in this case (Lechner 2010, Imbens and Wooldridge 2009).

6.2.1 Model

The Difference-in-Difference (DiD) approach is a research design for estimating causal effects. The design is usually based on comparing de facto four different groups of objects. Three of them are not affected by the examined intervention. Time is an important variable to distinguish between the groups. The group which experienced the intervention is called the treatment group, and the other one where things did not change is called the control group.

To study firm performance as independent variables the previously applied performance measures were used: total annual net sales, number of employees and productivity of firms, all in their natural logarithm forms.

To properly apply Difference-in-Differences, first some basic assumptions are needed to be determined. The parallel trends assumption states that without the treatment the change in the outcomes would be the same on average in the two compared groups (the treatment and the controlled groups). So, without the treatment the outcome variable would have changed in the same way in the treatment group, as it changed in the control group, on average. In the context

of my research, this means that I assume that, on average, the performance of the group of companies with the replacement of the founder, without the replacement, would have changed the same way as it changed in the group of companies without the replacement. Thus, this means that the companies which transferred to "professional leadership" have better performance measures only because of the change in the management. So basically, with the DiD analysis one can claim that the change in revenues or in the number of employees were only due to the change in the leadership of the observed startups. This way we can also eliminate reverse causality and the fact that there might be other unobserved factors, which impact the relationship and thus cause a change in the performance of the observed companies. As a result, control variables were also included in the equations, to sort out the effect of other factors (Wooldridge 2007).

During the estimation procedure, I only studied companies where the change happened between 2009 and 2013. I studied the effect of the interventions separately for firms according to the year of the change, thus the standard model for DiD was applied, to get a more sophisticated picture since, as it was showed before, there are significant differences among the years of the studied period from the aspect of SMEs. This way, I was able to concentrate more on the effect I was interested in, with the lower probability of having biased results due to these shocks. Furthermore, the estimated time period was two years after the intervention took place. Therefore, the "before period" consisted observations until the year of the change, and the "after period" started two years after the year of the change.

First, I estimated DiD regressions according to the different years of replacements. This way I was working with five different datasets. After that, since I am interested in the overall effect, a "final" DiD was run on the appended data, to get the weighted average of the separate results.

The following model was estimated for every scenario:

$$E[\Delta y_{it} | \Delta treated_{it}, Z_{it}] = \alpha + \beta \Delta treated_{it} + \gamma Z_{it} \qquad (2)$$

where Δy_i stands for the change in the observed performance measure. β shows the treatment effect, the estimated difference in the performance of a firm which replaced its founder two years after the change and Z_{it} is the collection of the different control variables applied in the regressions. The following control variables were inserted: number of owners, number of CEOs, industry dummies, year of foundation dummies, foreign ownership dummy, company ownership dummy, gender of the CEO, age of the CEO. Furthermore, in case of the estimation of log revenue changes I controlled for the underlying trend. If I did not insert it, it is possible that the intervention had no effect and the variable in question would have increased anyway. To include this trend I also controlled for the log of revenues two years before the change and the log revenue changes one or two years after the replacement. Based on previous papers' findings (Wasserman 2006, Ewens and Marx 2016) I expected a positive change, that treated companies would show a higher growth rate than the ones in the control group.

6.2.2. Results

The summarized results are shown in Table 6, 7 and 8. The results of the difference-in-difference analyses supports my initial assumption that companies after the replacement show improvement in their performance measures, thus the relationship between replacement and firm performance is positively related. The estimation results are separated by the usage of confounders. During the estimation of the effect on log total net sales change, besides the "basic" controls, additional confounders were applied. Lagged values of log revenues and future revenue growths were inserted to control for previous and subsequent years' trends.

The results show that confounders have an influence on the main relationship, including several controls magnifies the size of the coefficient of the aggregated *treated* dummy. Although when looking at individual cases one can see that the size of the coefficient changes varies within independent variables. In case of the number of employees variable the insertion of controls increases the inspected effect, but in case of the revenues and the productivity measures, they rather have reducing impact.

The aggregated results are positive and significant in case of revenue and employment size changes, but not significant in case of the productivity measure.

The largest effect can be seen in the change of the total annual net sales. The estimated coefficients are positive for all years, but only significant between 2009 and 2011. The same trend can be observed in the case of the other performance measures. The aggregated results show that revenues changed by 0.449 log units more 2 years after the replacement for companies where the management change took place, without including any controls. Thus, this coefficient demonstrated only the effect of the replacement on income changes in the following years of the replacement. This difference in revenues between treated and non-treated companies is increasing with the insertion of controls, which indicates that other individual firm characteristics contributes to further revenue growth.

In case of employment number one can also see a significant and positive change in the aggregated results. Without the controls, the size of the employment in startups, where a professional CEO was hired, changed 0.0187 log units two years after the replacement and in the forthcoming years, on average.

From 2012 most of the estimates are not significant. When I was estimating the log changes in these years I found a negative change in sales without controlling for trends and firm characteristics. This suggests that for companies where the founder was replaced in 2012, the

revenues have started to decrease from 2014. I assume that this negative trend can be explained by the fact that from 2012 the SME sector experienced two major shocks. Due to these shocks majority of the operating companies went out of business, or experienced bankruptcy. As a consequence, for the years after 2012 the growth in the income of startups has slowed down or turned into negative. Another scenario is also possible in this case, that companies which experienced a CEO replacement in those years were already in trouble, and founders left their companies because of the problems. Table 9 supports these theories.

Table 9: Number of company exists for the companies which replaced their founder in 2012 and 2013

Year of exit	Number of companies		Total
	Replacement in 2012	Replacement in 2013	
2012	39	-	39
2013	14	38	52
2014	11	16	27
2015	186	234	420

Note: Each row shows the number of companies for the years of 2012 and 2013. The numbers were retrieved from the individual datasets by the year of change, consists only the founder-CEO firms.

It demonstrates that most of these companies went out of business in 2015 or in the year when the change in the leadership took place. This means that these companies already had other problems independent of the current leadership structure. Probably that is why the founder left or needed a new leader to solve financial problems. Furthermore, I looked at the average 2-year log sales growth of these companies. I found a negative growth in revenues in both cases, which further supports my assumptions.

As an additional check, I ran the DiD regressions with interaction terms (Wooldridge 2012) by treatment years (Appendix A.7, A.8, A.9). The results support my initial hypothesis and the previously demonstrated estimations as well. This way also difference-in-difference regressions were estimated but this method takes into consideration previous years' performance even more. In case of the estimation of log revenue changes I also controlled for previous years' trend. The coefficients of the interaction terms are positive in all replacement years except for 2013. But this is in line with the main results and my above-mentioned theory. These regressions can also serve as a proof for the assumption, that startups with better performance more prone to replace their founders and hire a professional manager instead. Therefore we can not exclude reverse causality, the observed relationship is not unidirectional.

Moreover, as a part of checking the robustness of my results, I looked at graphically whether the common trend assumption holds. It is clearly important, because if it does not hold, then the estimated coefficients do not identify the effect correctly. Since the data covers only a short time period, it is not possible to check it econometrically. Nevertheless, I checked visually on graphs whether the treated and control groups move together. Such a graph for the treatment year of 2009 is shown in Appendix A.10. One can see that the values mostly moves together. The graph shows that until 2009, when the change took place, treated firms had worse performance on average, but slope of the lines is identical. More interestingly, it also shows that after the replacement the initial revenues of treated firms were worse, on average. This decline was probably because that after a management change firms need to adapt to the new situation. However, after 2011, the control firms' average income starts declining at a faster pace. Therefore, the graph rather serves as a proof for the fact that treated firms average revenues have improved two years after the change. On the other hand, it does not support my initial theory, that treated firms' average performance was better even before the intervention.

If that would be true, then the common trend assumption would not hold, which would be problematic from the aspect of causality. In sum, this graph rather supports that the common trend assumptions hold and the performance of the treated firms changed because of the replacement.

Concluding the results, one can see that, on average, total annual net sales and employment size of companies which went under a management change from 2009 to 2013 have improved in the following years after the transition. The aggregated coefficients are positive and significant even without the confounders. Although, there is not enough evidence to prove causal relationship between firm performance and founder replacement. The graph (Appendix A.10), showed that there might be an endogeneity problem, that maybe founders leave their firms because they are already in trouble. The cases of 2012 and 2013 brought up that issue. Furthermore, one can also claim that the regulation environment can have a serious impact on firm performance which cannot be handled only by hiring a professional manager.

In sum, firms with professional managers and without a founder-CEO perform better, on average, but there are definitely other factors which influence the direction and the size of this relationship, thus leadership structure cannot solely determine the growth of startups.

Table 6: DiD results on log revenue changes by the years of replacement

	Year of treatment (t). Dependent variable: Δlog(revenues)					
	2009	2010	2011	2012	2013	All
Treated companies without controls	0.662**	0.809***	0.698***	0.0851	0.0758	0.449***
	(2.26)	(0.264)	(0.261)	(0.257)	(0.241)	(0.0553)
Treated companies with controls	0.580**	0.690*	0.693**	0.272	0.0704	0.526***
	(2.74)	(0.358)	(0.334)	(0.325)	(0.342)	(0.0610)
Observations	1530	1260	1104	1200	1088	59594
# treated firms	512	398	335	250	288	1783
# control firms	1785	1680	1474	1262	1078	7279

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Each column reports the results of a firm level DiD regression on the treated dummy and log revenue changes by the year of the founder replacement. Treated companies are the ones with the founder replacement. Year dummies are included only in case of the "Treated companies with controls" estimation with other controls. "# treated firms" and "#control firms" rows show the number of companies with/without the replacement in the given year.

Table 7: DiD results on log employment size changes by the years of replacement

	Year of treatment (t). Dependent variable: Δlog(employmentsize)					
	2009	2010	2011	2012	2013	All
Treated companies without controls	0.125***	0.113**	0.241***	0.129**	0.0255	0.0187**
	(0.0372)	(0.0437)	(0.0548)	(0.0608)	(0.0418)	(0.00950)
Treated companies with controls	0.132***	0.100*	0.235***	0.114	-0.0199	0.0278***
	(0.0498)	(0.0596)	(0.0653)	(0.0749)	(0.0544)	(0.0105)
Observations	1 530	1260	1 104	1 200	1 088	59 594
# treated firms	512	398	335	250	288	1 783
# control firms	1 785	1 680	1 474	1 262	1 078	7 279

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Each column reports the results of a firm level DiD regression on the treated dummy and log employment number changes by the year of the founder replacement. Treated companies are the ones with the founder replacement. Year dummies are included only in case of the "Treated companies with controls" estimation with other controls. "# treated firms" and "#control firms" rows show the number of companies with/without the replacement in the given year.

Table 8: DiD results on log productivity changes by the years of replacement

	Year of treatment (t). Dependent variable: Δlog(productivity)					
	2009	2010	2011	2012	2013	All
Treated companies without controls	0.0593*	0.0435	0.145***	0.103*	0.0109	0.00820
	(0.0342)	(0.0418)	(0.0508)	(0.0553)	(0.0395)	(0.00886)
Treated companies with controls	0.0619	0.0288	0.135**	0.0884	-0.0244	0.00360
	(0.0462)	(0.0538)	(0.0574)	(0.0671)	(0.0512)	(0.00983)
Observations	1 530	1260	1 104	1 200	1 088	59 594
# treated	512	398	335	250	288	1 783
# control	1 785	1 680	1 474	1 262	1 078	7 279

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Each column reports the results of a firm level DiD regression on the treated dummy and log productivity changes by the year of the founder replacement. Treated companies are the ones with the founder replacement. Year dummies are included only in case of the "Treated companies with controls" estimation with other controls. "# treated firms" and "#control firms" rows show the number of companies with/without the replacement in the given year.

6.3 Firm survival and founder-CEOs

6.3.1 Model

To look at survival chances of firms with founder-CEOs and of firms with replaced founders, linear probability models were run on short-and on long-run survival prospects. I chose linear probability strategy because it has the advantage that the coefficients of the model can be easily interpreted as they measure the variable's contribution to the probability of being alive either on the short- or on the long-run. It is also suitable for my empirical estimation as it can handle dummy variables. On the other hand, it has some drawbacks: the predicted values can be greater than one or less than zero and sometimes the assumption of linearity cannot be maintained (Wooldridge, 2012).

Short-run survival was measured with a binary variable. Firms, which lived for at least 3 years, were marked with one, and zero otherwise. Long-run survival was evaluated according to whether a firm lived at least for 8 years. I was considering to look at firms' survival chances at the end of the period, in 2015, but very large number of firms went out of business in that year. Thus, I assumed that the results would be biased and reflect that rather than the exists because of the leadership structure.

The linear probability models were run on the initial founder-CEO dummy and then on the treatment group as well, to see whether the transition has brought any difference. However, there is potential endogeneity problem in these cases. It is possible that founders leave firms because they are already in the process of exiting, but my data is not capable of recording this kind of firm exits. Although, by taking into consideration zero sales when evaluating firm death and changing the year of the closure when the last positive revenue has appeared for a firm, might be a better indicator of death and thus can give us better prediction results.

The following linear probability models were applied for the above mentioned four different scenarios.

To predict short-run survival:

$$E[alive3yrs_{it} = 1|founderCEO_{it}] = \alpha + \beta founderCEO_{it} + \gamma Z_{it} + \delta C_{it} \quad (3)$$

$$E[alive3yrs_{it} = 1|treated_{it}] = \alpha + \beta treated_{it} + \gamma Z_{it} + \delta C_{it} \quad (4)$$

where the $alive3yrs_{it}$ variable is a dummy, which collects 3-year old companies in both equations (Equation 3 and Equation 4). Equation 3 shows the survival chances of firms with founder-CEOs, while Equation 4 shows the chances of companies, where the founder was replaced. Z indicates the used control variables and C stands for the company dummies according to their years of establishments. The used control variables were the industry characteristics, average number of CEOs, average number of owners, foreign ownership, foreign leadership of firms, the gender of founder-CEOs and the age of CEOs.

To predict long-run survival similar equations were applied:

$$E[alive8yrs_{it} = 1|founderCEO_{it}] = \alpha + \beta founderCEO_{it} + \gamma Z_{it} + \delta C_{it} \quad (5)$$

$$E[alive8yrs_{it} = 1|treated_{it}] = \alpha + \beta treated_{it} + \gamma Z_{it} + \delta C_{it} \quad (6)$$

where the only difference was in the dependent variable. The $alive8yrs_{it}$ dummy stands for companies which are still alive after 8 years of their foundation. β represents the survival chances of founder-CEO led and transitioned companies (Equation 5 and 6). Z_{it} stands for the same group of controls as before and C indicates the company-year dummies.

6.3.2 Results

Table 10 shows the results of firms' survival prospects on the short-run and on the long-term. I compared firms according to their years of establishments, because as previously discovered there are notable differences between firms according to their years of foundation. Thus, the tables organized according to this. In Column (4) one can see the weighted average of the three coefficients.

When evaluating survival chances of firms with founder-CEOs, I included an additional control variable, which shows the duration for which a company was operating without its founder as leader, but the results did not show significant differences in the two different scenarios. This means, that the duration of a founder-CEO does not remarkably determine firms' survival chances. In case of the analysis of the companies, where the founder was replaced the coefficients of the main independent variable slightly increased due to the insertion of this duration variable. This indicates that companies which replaced their founders earlier have better survival perspectives. Although, I decided to work with only the initial control variables, because of the not significant changes in coefficient size.

If we compare survival probabilities of firms with founder-CEOs on the short-and on the long-run, we can see that on the short-run having a founder-CEO does not have a negative effect on a firms' perspectives. However, despite the positive coefficients, we can see that even on the short-run very low proportions of founder-led enterprises survive. If we look at long-run survival of firms, we can see that survival probability is negatively correlated with leadership structure. When controls are included the size of the coefficients decreases on the short-run and increases on the long-run. Although, the difference is not outstanding. Thus, this let us assume that other firm characteristics can improve firms' chances on the long-run, but their overall influence is not enough to significantly boost startup survival.

The results of the group of companies which replaced their founders and switched to professional management, show a significant improvement in their probabilities. Within the group of founder-managed companies one can see a notable improvement for the ones, which transferred to professional leadership during their lifetime. The former negative coefficients for the long-run estimation became positive and we can see that much larger proportion of startups survive within this group. However, on the short-run the replacement of the founder show worsening survival chances, on average[2].

If the coefficients of the regressions with and without controls are compared, we cannot see any difference. So, the same as above can be concluded, that combining individual firm characteristics with leadership structure of enterprises does not improve survival prospects for firms.

In sum, the results of the linear probability regressions show the expected outcomes. Very few percent of founder-led companies survive on the short-run, while on the long-run they have even worse chances. This is indicated by the negative coefficients of the long-run regressions. While on the short-run 2.62 percent of founder-led companies survive, on the long-run this drops to -14.4 percent, which signals that more of them go out of business than stay alive, on average.

However, replacing the founder can significantly improve survival chances of founder-led companies, on the long-run. On average, 36.1 percent of transferred companies stay alive, which is an outstanding improvement comparing it to the previously negative coefficients.

In addition, time and the duration of the leadership play a significant role. Companies which were established earlier have worse survival chances, except for the ones which were

[2] This is in line with the results of the graph, which compares average performance of treated and non-treated companies. It showed that in the beginning these firms have worse average revenues.

established in 2007. These companies' survival chances are negative, even on the short-run. This can be explained by the two major economic shocks during the examined time period: the 2008 crisis, and in 2015 the introduction of the new Civil Code. Supposedly, these firms were affected the most.

Table 10: Short- and long-run survival probabilities according to leadership structure and year of foundation

	Variables	2005		2006		2007		All	
		Founder-CEO	Treated	Founder-CEO	Treated	Founder-CEO	Treated	Founder-CEO	Treated
short run	Founder-CEO dummy	0.110***	-0.0248**	0.0625***	0.0273**	-0.00976***	0.0368***	0.0262***	0.0305***
		(0.00885)	(0.0123)	(0.00594)	(0.0117)	(0.00342)	(0.00562)	(0.00295)	(0.00480)
	Founder-CEO dummy w/controls	0.104***	-0.0233**	0.0601***	0.0269**	-0.00887***	0.0377***	0.0243***	0.0311***
		(0.00891)	(0.0118)	(0.00589)	(0.0115)	(0.00339)	(0.00562)	(0.00292)	(0.00475)
long run	Founder-CEO dummy	-0.168***	0.398***	-0.129***	0.365***	-0.148***	0.356***	-0.144***	0.361***
		(0.0279)	(0.0512)	(0.0154)	(0.0337)	(0.0110)	(0.0160)	(0.00852)	(0.0139)
	Founder-CEO dummy w/controls	-0.137***	0.323***	-0.101***	0.329***	-0.115***	0.350***	-0.113***	0.345***
		(0.0258)	(0.0480)	(0.0150)	(0.0324)	(0.0110)	(0.0156)	(0.00837)	(0.0136)
	Year of foundation dummies	NO	NO	NO	NO	NO	NO	YES	YES
	Industry dummies	YES	YES	YES	YES	YES	YES	YES	YES
	Observations	3,562	296	4,155	1,096	5,799	3,006	13,516	4,398

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The columns show the results for the survival probability of firms by their year of foundation. Both short-and long-run probabilities are displayed. The estimations demonstrate the results with and without including controls. Last two columns show the aggregated probabilities with year of foundation dummies included. Industry dummies are always included. "Founder-CEO" indicates firms with founder-CEOs without the consideration of them being replaced. "Treated" columns indicate the firms with founder-CEOs which switched to professional management after a certain amount of time.

7.Conclusion and policy implications

The thesis analyzed the effect of having a founder-CEO on new firm performance. I used several methods to answer the main research question and to find evidence on the existence of a causal relationship between leadership structure and the change in performance measures. The results proved the significant effect of founders on their firms' prospects, both in terms of performance and survival, on the long-run. However, these findings are not powerful enough to prove the existence of a unidirectional causal relationship between performance change and founder replacement. The regressions revealed three main findings. First, firms where the founder-CEO was replaced demonstrate a positive change in their financial, size and productivity measures after the replacement. Second, at startups where the founder has remained CEO show worse survival prospects, but in case of hiring a professional their chances for survival improve significantly. Third, reverse causality can not be excluded, nor the existence of other unobserved factors, which lead to the varying performance of the observed businesses.

Although the conducted research can not yet establish causality but proves that founders matter and can arise a few initial policy implications. A skilled and well-educated entrepreneurial sector is essential to new firm development, but only if it is along with a stable and predictable regulation environment.

In view of the role of SMEs in economic restructuring, governments should promote entrepreneurship and facilitate firm start-up with the improvement of access to different types of financing opportunities and with the easing of regulation burdens. Regulatory and administrative burdens can be one of the greatest spur to entrepreneurship (OECD 2016). New firms are especially endangered in this case, by being more sensitive and vulnerable because of

the lack of a stable financial background. As a matter of fact, my thesis proves evidence for that. The change of the Civil Code severely affected the life of the Hungarian startups, caused disturbance and forced out numerous new firms out of the market.

Furthermore, management matters as well, especially good management. It makes difference in shaping national performance as well. Bloom, Sadun and Van Reenen (2012) has proved that the variation in management practices accounts for the productivity gap between Europe and the U.S. And what makes companies more likely to be good at it, are the people working for them. However, not just the quality of the employees but the quality of the top managers equally matters (Bloom et al. 2017). Consequently, on the short-run, creating incentives for continuing education of managers as well as employees and better information diffusion of best management practices may help. Whereas, on the long-run incremental changes are not enough, governments need to start investing more in education to effectively raise the value of human capital (Bloom, Lemos and Scur 2012).

Taken as a whole the results of my research suggest that managerial human capital matters and is important for the ability of startups to maintain growth and success, but the presented evidence is not enough to be able to confirm that worse than average startup performance is due to the lack of a professional management. Furthermore, a change in the leadership structure is related with better firm performance, but the direction of this relationship is unsure. The findings shown in my thesis are consistent with the emerging literature emphasizing the negative association between founders as managers and firm performance (Bloom, Sadun and Van Reenen 2011, Wasserman 2006). I extend this literature by providing additional evidence for Hungarian startups. As for future research, based on Bloom, Sadun and Van Reenen (2012) findings, I think the more detailed examination of leadership skills and management practices of Hungarian entrepreneurs would bring developments and would help to gain more insight on the characteristics of this relationship.

References

Act V of 2013 on the Civil Code effective as of 15 March 2014. Accessed May 29, 2017. net.jogtar.hu/ptk.

Anderson, Ronald C., and David M. Reeb. 2003. "Founding-Family Ownership and Firm Performance: Evidence from the S&P 500." *The Journal of Finance* 58(3): 1301-328.

Bennett, Manuel; Lawrence, Megan and Sadun, Raffaella. 2015. "Are Founder CEOs Good Managers?". Paper presented at the NBER/CRIW conference "Measuring Entrepreneurial Businesses: Current Knowledge and Challenges". Accessed May 23, 2017. http://www.nber.org/chapters/c13499.pdf.

Berle, Adolf and Means, Gardiner. 1932. *The Modern Corporation and Private Property*. Macmillan: London.

Bloom, Nicholas; Lemos, Renata and Scur, Daniela. 2012. Are Poor Management Practices Holding Back Middle-Income Countries -New evidence from the private and public sector in India. International Growth Center Policy Brief. Accessed May 28, 2017. http://www.theigc.org.

Bloom, Nicholas; Sadun, Raffaella and Van Reenen, John. 2015. Family Firms Need Professional Management. *Harvard Business Review*, March 25. Accessed May 31, 2017. https://hbr.org/2011/03/family-firms-need-professional.

Bloom, Nicholas et al. 2011. "Management practices across firms and countries". Harvard Business School. Working paper 12-052, December 19. Accessed May 31, 2017. http://www.hbs.edu/faculty/Publication%20Files/12-052.pdf.

Bloom, Nicholas et al. 2017. "What Drives Differences in Management?". NBER Working Paper Series. Working Paper 23300. Accessed May 31, 2017. http://www.nber.org/papers/w23300.pdf.

Boeker, Warren. 1989. "Strategic Change: The Effects of Founding and History." *The Academy of Management Journal* 32(3): 489-515.

Cheng, Jin and Thompson, Peter. 2012.: "New Firm Performance and the Replacement of Founder-CEOs". Copenhagen Business School and Emory University. Accessed May 13, 2017. http://funginstitute.berkeley.edu/wpcontent/uploads/2013/12/Peter_Thompson_turnover_final.pdf.

Cochran, Philip L., and Robert A. Wood. "Corporate Social Responsibility and Financial Performance." *The Academy of Management Journal* 27(1): 42-56.

EU recommendation 2003/361. Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises (Text with EEA relevance) (notified under document number C (2003) 1422).

Daily, Catherine.M. and Dalton, Dan. R. 1992. "Financial performance of founder-managed versus professionally managed small corporations". *Journal of Small Business Management* 30: 25-34.

Dalton, Dan R and Kesner, Idalene F. 1983. 'Inside/outside succession and organizational size: The pragmatics of executive replacement', *Academy of Management Journal* 26: 736–742. Accessed May 22, 2017. doi: 10.2307/255919.

Dyer, W. Gibb. 2006. „Examining the "Family Effect" on Firm Performance". *Family Business Review*, 19: 253–273.Accessed May 25, 2017. doi:10.1111/j.1741-6248.2006.00074.x.

Ewens, Michael and Marx, Matt. 2016. "Founder Replacement and Startup Performance". Accessed June 1, 2017. Available at SSRN: https://ssrn.com/abstract=2717124.

Fama, Eugene F., and Michael C. Jensen. "Separation of Ownership and Control." *The Journal of Law & Economics* 26 (2): 301-25.

Freedman, Marc. 2012. "Why Older Entrepreneurs Have an Edge". *Harvard Business review,* April 20. Accessed May 31, 2017. https://hbr.org/2012/04/why-older-entrepreneurs-have-a.

Frick, Walter. 2016."Proof That Good Managers Really Do Make a Difference". *Harvard Business Review*. April 11. Accessed May 31, 2017. https://hbr.org/2016/04/proof-that-good-managers-really-do-make-a-difference.

Hoffman, Reid (2013): "If, Why, and How Founders Should Hire a "Professional" CEO". *Reid Hoffman's personal website*, reidhoffman.org. January 21. Accessed June 1, 2017. http://www.reidhoffman.org/article/2169.

Hungarian Central Statistical Office. Accessed May 25, 2017. https://www.ksh.hu/.

Imbens, Guido W. and Jeffrey M. Wooldridge. 2009. "Recent developments in the econometrics of program evaluation". *Journal of Economic Literature* 47 (1): 5-86. Accessed June 3, 2017. doi:10.1257/jel.47.1.5.

Johnson, Whitney. 2013. "Entrepreneurs Get Better with Age". *Harvard Business Review*. June 23. Accessed May 30, 2017. https://hbr.org/2013/06/entrepreneurs-get-better-with.

Lechner, Micheal. 2010. "The Estimation of Causal Effects by Difference-in-Difference Methods". *Foundations and Trends in Econometrics* 4 (3):165–224. Accessed June 3, 2017. doi: 10.1561/0800000014.

Maury, Benjamin. 2005. "Family ownership and firm performance: Empirical evidence from Western European corporations". *Journal of Corporate Finance* 12:321– 341. Accessed May 25, 2017. doi:10.1016/j.jcorpfin.2005.02.002.

Monsen, R. Joseph. 1969. "Ownership and management: The effect of separation on performance". *Business Horizons* 12:45-52. Accessed May 15, 2017. https://doi.org/10.1016/0007-6813(69)90055-X.

OECD. 2016. "Entrepreneurship at a Glance 2016". *OECD Publishing*. Paris. Accessed June 4, 2017. doi: 10.1787/entrepreneur_aag-2016-en.

Ranjay Gulati, Ranjay and DeSantola, Alicia. 2016. “Startups Can’t Revolve Around Their Founders If They Want to Succeed”. *Harvard Business Review*, March 4. Accessed May 31, 2017. https://hbr.org/2016/03/startups-cant-revolve-around-their-founders-if-they-want-to-succeed.

Renée Adamsa, Renée and Almeidab, Heitor and Ferreirac, Daniel. 2009. “Understanding the relationship between founder–CEOs and firm performance”. *Journal of Empirical Finance*.16:136–150.

Schumpeter, A. Joseph. 1934. *The theory of economic development: an inquiry into profits, capital, credit, interest and the business cycle.* Campbridge, MA:Harvard Economic Studies. 46. Harvard College.

Wadhwa, Viviek. 2013. “The Truth About Entrepreneurs: Twice as Many Are Over 50 As Are Under 25”. *PBS Newshour.* April 26. Accessed May 31, 2017. http://www.pbs.org/newshour/rundown/the-truth-about-entrepreneurs-twi"ce-as-many-are-over-50-than-under-25/.

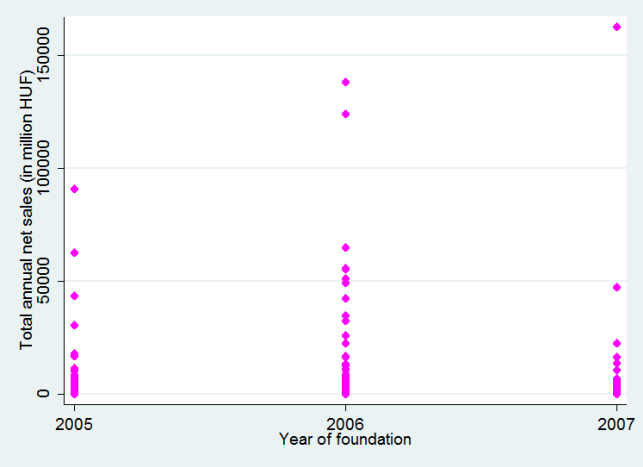
Wasserman, Noam (2006): "Rich versus King: The Entrepreneur's Dilemma." *Academy of Management Annual Meeting Proceedings.* Accessed May 24, 2017. http://www.hbs.edu/faculty/Pages/item.aspx?num=22821.

Wooldridge, Jeffrey M. 2007. “What’s New in Econometrics? Lecture 10 -Difference-in-Differences Estimation”.Lecture presented at the NBER Summer Institute. Accessed June 1, 2017. http://www.nber.org/WNE/lect_10_diffindiffs.pdf.

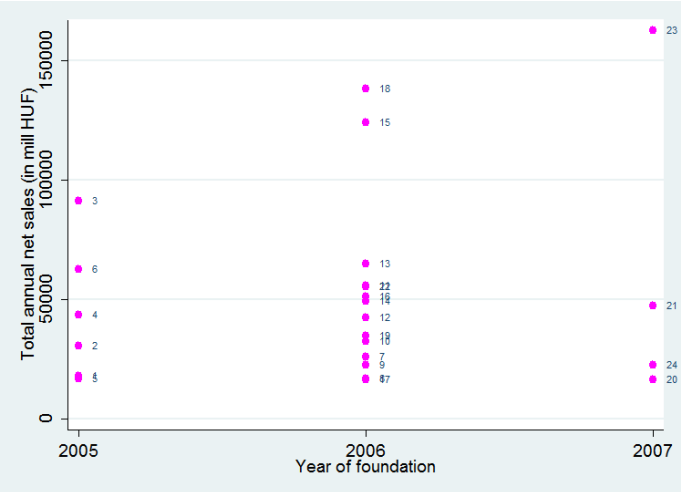
Wooldridge, Jeffrey M. 2012. *Introductory econometrics: A modern approach*. Cengage Learning. 5th edition.

Appendix A: Tables and figures

A.1: Figure of first-year total annual net sales of companies by their year of foundation



A.2: Figure of number of companies with extreme first year total annual net sales by their year of foundation



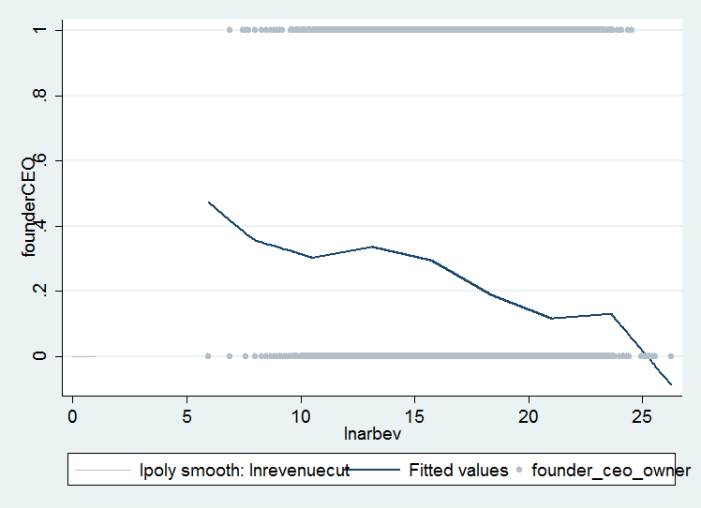
A.3: Summary statistics and distribution of the outcome variables

Statistics	Total annual net sales (in million HUF)	Number of employees
Mean	200.5	7.6
Standard deviation	1465.4	16.2
Minimum	0	0
5th percentile	0	1
10th percentile	2.85	1
Median	40.15	3.3
75th percentile	119.496	7
90th percentile	341.387	15
Maximum	256 380.5	371.8
Number of observations	108 309	108 309

A.4: Number of companies by their year of exit and birth

Year	Number of company exists	Percentage	Number of company births	Percentage
2005	757	0.38%	29 462	8.15%
2006	3 082	1.53%	27 692	7.66%
2007	5 682	2.81%	32 096	8.88%
2008	7 423	3.68%	41 318	11.44%
2009	9 803	4.86%	40 206	11.13%
2010	13 224	6.55%	40 578	11.23%
2011	15 041	7.45%	47 511	13.15%
2012	18 029	8.93%	29 665	8.21%
2013	23 427	11.61%	28 866	7.99%
2014	31 179	15.45%	23 744	6.57%
2015	39 546	19.59%	19 981	5.53%
2016	33 205	16.45%	199	0.06%
2017	1 455	0.72%	-	-
Total number of companies	201 853	100%	361 318	100%

A.5: Relationship between leadership structure and firms' financial performance



A.7: DiD regression results on log revenue changes with interactions by treatment years

	Year of treatment (t). Dependent variable: Δlog(revenues)				
	2009	2010	2011	2012	2013
Treatment	0.187	-0.117	-0.306***	-0.121	0.241**
	(0.202)	(0.136)	(0.113)	(0.109)	(0.113)
Post-treatment period	-2.850***	-2.088***	-1.546***	-1.174***	-0.744***
	(0.0912)	(0.0793)	(0.0853)	(0.103)	(0.175)
Treatment x Post-treatment period$_{t+2}$	0.607***	0.488***	0.785***	0.592**	0.408
	(0.221)	(0.187)	(0.185)	(0.234)	(0.338)
Observations	10,842	10,399	10,521	10,488	9,005
R-squared	0.362	0.349	0.325	0.262	0.053

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Results are shown for DiD regressions in different years of replacement. „Year of treatment" indicates the year of the founder replacement. The „Treatment x Post-treatment period" row shows the DiD estimate. Dependent variable is log revenue changes, independent variable is the dummy for the treated group, which is the group of companies with a replaced founder.

A.8: DiD regression results on log employment size changes with interactions by treatment year

	Year of treatment (t). Dependent variable: Δlog(employment)				
	2009	2010	2011	2012	2013
Treatment	-0.182***	-0.132***	-0.123***	-0.126***	-0.112***
	(0.0341)	(0.0318)	(0.0295)	(0.0290)	(0.0251)
Post-treatment period	0.457***	0.271***	0.153***	0.132***	0.136***
	(0.0204)	(0.0201)	(0.0218)	(0.0258)	(0.0359)
Treatment x Post-treatment period$_{t+2}$	0.368***	0.204***	0.327***	0.191**	0.408
	(0.0427)	(0.0458)	(0.0515)	(0.0671)	(0.338)
Observations	13,174	12,614	12,674	12,555	13,210
R-squared	0.073	0.024	0.012	0.005	0.003

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Results are shown for DiD regressions in different years of replacement. „Year of treatment" indicates the year of the founder replacement. The „Treatment x Post-treatment period" row shows the DiD estimate. Dependent variable is log changes in number of employment, independent variable is the dummy for the treated group, which is the group of companies with a replaced founder.

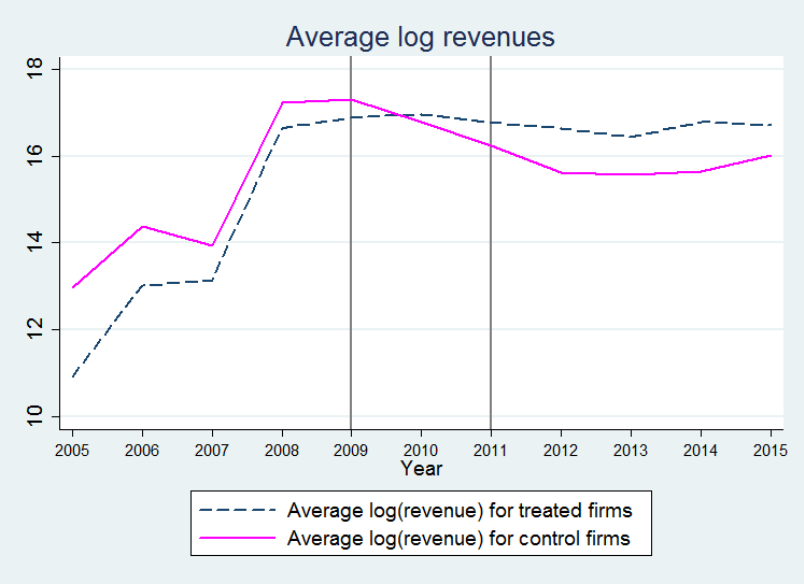
A.9: DiD regression results on log productivity changes with interactions by treatment year

	Year of treatment (t). Dependent variable: Δlog(porductivity)				
	2009	2010	2011	2012	2013
Treatment	0.418***	-0.113	-0.118	-0.0344	0.517***
	(0.113)	(0.113)	(0.140)	(0.128)	(0.0940)
Post-treatment period	-1.578***	-0.734***	-1.493***	-1.052***	-0.413***
	(0.0833)	(0.0898)	(0.100)	(0.118)	(0.150)
Treatment x Post-treatment period$_{t+2}$	0.383**	0.476**	0.496**	0.621**	0.167
	(0.150)	(0.188)	(0.216)	(0.264)	(0.293)
Observations	12,986	16,522	8,373	8,430	11,557
R-squared	0.238	0.170	0.268	0.202	0.202

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Results are shown for DiD regressions in different years of replacement. „Year of treatment" indicates the year of the founder replacement. The „Treatment x Post-treatment period" row shows the DiD estimate. Dependent variable is log productivity changes, independent variable is the dummy for the treated group, which is the group of companies with a replaced founder.

A.10: Average log revenues of treated and control groups of firms when the replacement year is 2009



Note: The graph shows how the average log total annual net sales for treated and non-treated companies changed between 2005 and 2015. The vertical line at 2009 indicates the year of the founder-CEO replacement, the second line shows the starting point of the period from which the analysis was conducted.

Appendix B: Data

B.1 Description of the raw data

The conducted research in my thesis is based on proprietary data from Bisnode D&B Hungary Ltd., an international digital business information provider. The company offers their support with their risk management, marketing and business information technology solutions to all kind of enterprises worldwide. Their IT based models are part of business' decision making support system.

The database provided by Bisnode contained several datasets which were connected with each other in several ways. It is called a relational database. A relational database has more than one table and the tables are linked using key or primary fields. To ensure uniqueness each table has one or several primary fields, which uniquely identifies every record of the table. In my case the main key field was the variable, which contained information on identification numbers of companies. Company ID numbers could be found in every dataset I was working from. Additional key fields were the identification numbers of owners and leaders, and code identifiers. These codes also served as ID numbers, but for different groups variables. Primary keys are can be used to reference to other tables. In my case, key fields were mainly referring to other tables and their contents. For example, the exact position of company leaders or different rows of financial reports were coded by numbers, which description could be found in additional datasets. These code datasets contained the exact name of a position or a financial report row. A primary key is called a simple key if it is a single column; it is called a composite key if it is made up of several columns. In most the cases I was working with composite key identifiers since my main datasets (the owners, leaders and the financial reports files) was made up from more than one key field. The relationship between datasets is a many-to-many, because all key fields were interconnected with other datasets and in most of the tables more than one identifier could be found. Thanks to this relationship it allows the user to access the data in an

almost unlimited number of ways , and to combine the tables and build blocks to create a complex database for my research. This kind of structure gave me the freedom to combine different datasets in the database and also enabled to create almost any kind of variable which I need, thus build a database with the most useful and insightful variables from the aspect of my research topic. In most of the cases this was the companyID and the codeID. In the owners and leaders files there were additional key fields, like the ID number of the person or the company which owns or manages the given company (Codd E. F. 1970)[3].

B.2 Structure of the database

The database I was working from was composed of 12 individual datasets. My final database contains information from all of these 12 files and my new variables were created by combining these together.

Out of the 12 files 2 contained the code identifiers. The codes for the financial reports data was stored in one file, while all the other identifier codes were in a different file. The identifier codes described the exact position of leaders and owners on the board and the employee number categories.

Two other files had information about the different features of the owners and the leaders of companies. One of them was built up from ID numbers of persons. In that I could find the basic socio-demographic characteristics of each private owner and leader of a company. It contained information on both the leaders' and the owners' nationality, gender, city of birth with exact postal code and birth date. The other file was made up from the individual identification

[3] Codd E. F. 1970. "A Relational Model of Data for Large Shared Data Banks", *Communications of the ACM.* 13:377–387, June 1970. Accessed May 15, 2017. http://www.morganslibrary.net/files/codd-1970.pdf .

numbers of owner companies. The variables defined the country and the city of origin of the linking company with exact postal codes, and its year of foundation.

The following eight files formed the main contents of my final database:

- *Basic information*

This file collected together all of the companies which could be found in either of the files. It had the most variables. They were mainly different binary variables referring to the ownership structure of companies or indicated what kind of legal procedures each of the company went under from the beginning of 2005. For example, the company_owner dummy variable took the value 1 if a company ever had another company as its owner from the year of 2005 and zero otherwise. Besides the company identifier, there was another key field, the codeID. This code ID number showed the current legal form of the company. Additional variables marked the year of formation, the year of cessation, the place of the current headquarter, the place of the current establishment (if had any) of each company. Moreover, there was another code identifier, which represented the name of the industry in which the individual company operates in currently.

- *Financial reports (balance sheets and income statements)*

The financial reports file had five main variables. The companyID, the exact start date of the reporting period, the exact end date of the reporting period, the report row code and the exact amount of each row in Hungarian forints. It had financial information of firms from 2005 until 2015. This file was merged with the code file which contained information on the row names of the submitted balance sheets and the profit and loss statements of companies.

- *List of headquarters and the list of establishments with the number of establishments (2 files)*

These files had the same structure regarding their variables except for one additional variable in the list of the establishments file. The variables were the companyID, the starting period of the validity of each headquarter or establishment, the end date of validity period, a dummy which showed whether the company uses that place as a HQ or as an establishment, the postal code of the city in which the HQ or the establishment place can be found and the name of the city. The additional variable in the establishments file showed the number of establishments for the indicated validity period.

- *Number of employees and categories of employees (2 files)*

These two files showed the number of employees of each firm. The categories of employees file had quarter-year information from 2006 until 2010 on employee number categories of firms. It also needed to be merged with the general codes file, because the exact description of these categories could be found in that file only.

The number of employees file has started in 2009 with monthly observations. It had three variables: the companyID, the date (year-month) when the company had that number of employees and the exact number of employees in a that month in a year.

- *Leaders*

It had eight variables. This was one of the most essential file for my research besides the owners and the financial reports files. The variables were the companyIDs, the beginning of the validity period of the company's leader, the end of the leadership of the leader, a validity dummy, the connecting person's ID, the connecting company's ID, a dummy which showed whether the leader was foreign or not, and a position code. Additional background information, like the socio-demographic characteristics of each leader could be found in the connecting person's and

in the connecting company's file. The leaders could be identified according to their ID number. The exact description on the position of each leader were located in the codes file.

- *Owners*

The owners file had nine main variables. The companyID, the beginning of the validity period of the company's owner, the end of the ownership of the owner, a validity dummy, the connecting person's ID, the connecting company's ID, a dummy which showed whether the owner was foreign or not, the owner's exact share in the company in percentages and a position codes. The codes gave further information about the proportions of the owners in the company, which was especially helpful in cases when the percent indicator was missing.

B.3 Details on data issues

In this section I describe the main issues which came up when I was dealing with creation of my final dataset. Since I did not download the data from any of the "official" firm databanks I think it is important to show the main steps and issues of the data cleaning process. I was working with raw data, which needed transformation to be able to work with it and retrieve the information that I needed. The overall goal was to create a company and year based panel data format from each of the files, where companyIDs and years uniquely identify the observations. This way after the merge it was able to avoid the typical problems which could occur when two datasets which contain different information are merged. It also eased the following data management steps after the merge (Chapman 2005)[4].

[4] Chapman, A. D.2005. "Principles and methods of data cleaning. – Primary Species and Species-Occurrence Data". version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. Accessed May 21, 2017. http://webcache.googleusercontent.com/search?q=cache:u5a300bNcokJ:www.gbif.org/system/files_force/gbif_resource/resource-80528/Principles%2520and%2520Methods%2520of%2520Data%2520Cleaning%2520-%2520ENGLISH.pdf%3Fdownload%3D1+&cd=16&hl=hu&ct=clnk&gl=hu .

The whole process of my data cleaning and error checking methods were centered around improving the quality of the raw data, make them 'fit for use' (Chapman 2005, 2) and try to decrease the probability of the occurrence of any kind of measurement errors or data distortion on the long-run.

During the procedure I was trying to follow some basic principle and created the data cleaning strategy according to them (Chapman 2005). I developed a policy and an overall strategy which will be followed throughout the whole process. This meant a conservative approach, in order to keep as many companies and observations in the dataset as possible. My goal was to cover as much firms in the final dataset as possible, so the results may have a larger significance and the model a larger effect size. Many data cleaning approaches simplify the problems by throwing away observations. Most of these approaches may lead to biased estimates or we can end up with larger standard errors due to reduced sample size (Gelman 2007).

On the other hand, I needed to get rid of messy observations as well, because of the danger of the classical measurement error and attenuation bias (Pischke 2007)[5]. However, for example data transformation also magnifies the measurement error problem, which was applied several times.

As I mentioned before the three main datasets I was working with were the financial reports, the leaders and the owners file.

The file about the balance sheets and profit and loss statements of companies contained 248 million rows all together. First, I determined my main financial variables, which I will need in the further work. I selected variables which are useful to measure firms' financial performance and can be applied when comparing financial performance of different firms. Thanks to this

[5] Pischke, S. 2007. "Lecture notes on measurement error". London School of Economics. Accessed May 21, 2017. http://econ.lse.ac.uk/staff/spischke/ec524/Merr_new.pdf.

selection procedure I was able to make the dataset smaller which facilitate my further work a lot. The variables which were retrieved are the following: total net sales and revenue; net income before taxes; operating costs; total assets; cash; common stock; stockholders' equity; current liabilities and long-term liabilities.

After extracting the variables the database needed to be reshaped to be able to create additional variables from it. It had a long panel format which was transformed to a wide panel. The kept rows of the balance sheet and the profit and loss statement became variables. A new *year* variable was created from the dates of the financial reports. However, there were companies which submitted more than one statement or balance sheet in a year. According to the Act C of 2000., Act on Accounting and the International Accounting Standards (IAS 2017)[6] the latest row was kept, because that contained the aggregated data. After the reshape, another rule was set up in order to filter the companies in the data. I looked at the annual net sales of companies and decided to keep only those companies which at least once in their lifetime had a 15 million or more annual net income from sales. This way I was left with 323 831 companies. These companies created the base of my database.

Since the structure of the leaders and the owners file were very similar I had to deal with almost the same issues in both cases when cleaning the datasets. First, the date variables needed to be handled to be able to create a wide panel data based on company-year observations. In the leaders file there were 2 842 817 company start and end date observations. The owners file had 3 413 499 rows. The file had that many rows because every different owner or leader of each company was put into a different row, moreover if the same leader experienced a position change or got back into the leadership after the replacement, he or she was put in as a new

[6] IAS 2015. International Accounting Standards. Accessed May 21, 2017. https://www.iasplus.com/en/standards/ias

observation row because the validity dates have changed. So basically, the exact dates of each leader or owner determined one observational row in the data.

B.3.1 Missing observations

To decide how to handle missing data, it is helpful to know why they are missing. The overall intuition is that we have substantial uncertainty about the missing values but by choosing an imputation strategy which we follow through, we can pretend that we know the true value with certainty.

In general, there are four types of missing data mechanisms (Gelman 2007).[7] In my case the most common mechanisms were when missing was random and when missing values depended on unobserved predictors. Unfortunately, we generally cannot be sure whether data really are missing at random, or whether the presence of missing values depends on unobserved predictors or the missing data themselves. The fundamental problem is that these values are unobserved thus we can never be sure about them and can not rule them out. We generally might make assumptions and insert values according to these assumptions or check for the possible reasons that what might be behind the missing values (Gelman 2007). So, rather than removing missing data from the files I tried to keep as many observations as possible, which can be advantageous for bias and precision, however this can also yield in different kind of bias and distortion. In general, imputation strategies are considered being conservative approaches because they rather lead to the underestimation of the true treatment effect, thus protect us from making the type one or the type two errors (Gelman 2007).

[7] A. Gelman and J. Hill. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Model. Cambridge University Press.

First, the missing values in the owners and leaders datasets needed to be handled. My assumption was that in the case of end dates a missing value indicates that the owner or the leader is still in place. This presumption was supported by the fact that in these rows the *validity* binary variable got the value one. Consequently, these missing values were replaced as 12 December 2016. In the case of the leaders we are talking about 877 070 missing end dates, from which 4 4 81 were marked as not valid, although I kept those observations as well. In the owners file 1 277 995 observations were missing.

The missing start dates of owners and leaders were dropped, because in those cases it was impossible to come up with a proper assumption, which would not distort the data and the represented information. For example, replacing the missing start dates with the start date of my further dataset would have been a mistake, because the duration of each leader or owner would be biased and these individuals would have become founders which would have a substantial effect on the overall results and probably would cause the overestimation of the final effect. That's why these observations were discarded.

Missing values occurred in the file of the financial reports as well. The main problem was when the total net sales values were missing since these were one of my main performance and success measurement variables besides the number of employees. In case of the missing and not positive values of the total net revenues I introduced a rule about how to handle these values. The decision was to first transform all negative, zero or missing sales into zero, because in general these values indicate that the firm did not have valuable income in that year. Furthermore, if I found that a company had non-positive income even after in its fourth year of operation, that company was dropped from the dataset. In addition, the year of death of a company was modified according to the first year when its revenues were zero or below zero for more than three years. In the financial reports file however, there were also companies which did not submit any statement in their year of formation. Their first statement was submitted

only after a certain amount of years. As a consequence, this made the final dataset even more unbalanced. In these case,s I found these kind of companies after their first year of submission had a positive balance until the end of their operation, thus these companies were kept.

When I was dealing with the number of employees files I also found some missing data for some years. Sometimes there was no available information on a company's number of employees throughout its whole lifetime. In addition, neither of the files were complete in the terms of the covered time period. One of them which had the employment number categories, had quarter yearly observations from 2006 until 2010. The other one, with the exact numbers on employment had monthly data from 2009 until 2016. These two were merged. So, there was no available information on the year of 2005. In order to solve the problem my presumption was that firms which were established in 2005, probably were in the same employee number category like one year after the establishment. For that reason, the missing values for 2005 were replaced with employee numbers from 2006. Additionally, the missing values for the whole time period of operation means that those companies had zero or one employees, which is possible in case of the so-called 'project-based companies'. So, these missing values were replaced as one, due to the reason that later these variables was transformed into natural logarithm.

Also, there were some years for some company with missing number of employees values. In these cases the applied general rule was that these were replaced with the values of the previous or following year, depending on which one was non-missing.

In the basic information file the date of the cessation of a company was also missing in several cases. However, this meant that particular company is still alive, thus these missing values were replaced as 2017. This was essential when firms' age was determined and it needed to decide whether that company is still operating or not. Beyond these, the year of closure does not play an essential role in the estimation process, that's why it could be replaced with 2017.

B.3.2 Handling the dates

Some observations were clearly misdated in both the leaders and the owners file. The most common errors were that the end dates or the start dates were dated later than 2017, the start dates have started later than the end dates (start dates were larger than the end dates). These observation s were dropped because there was no chance to figure out the right dates for them. When looking at the year of closure in the basic information file I found some additional strange things in the data which needed to be handled. In the case of 10 observations, which means four companies in this case, I found that the year of cessation was dated earlier that the year of formation of the company. My decision was to drop these firms. Another issue with the year of cessation came up when I compared start dates and years of closure. I observed that there were 103 observations where start date of a leader was dated later than the year of death of that company, which is technically impossible. Logically, these observations were dropped from the data. In the case of the leaders file we are talking about 1540 observations and 1694 observations in the owners file.

There were issues with the year of the cessation and the last date of the very last owner leader of companies. In some cases these were not equal, the last date was larger than the actual year of death of the company. This problem was solved by replacing the last date with the year in which the company died. I could not drop these observations because this was a quite frequent issue in both files. For example, in the owners file there were 353 993 observations, where the last date of the owner was larger than the year of closure of the firm.

In addition, there were cases when due to the change in the description of the leader, the same leader was marked as a new one, by replacing it in a new row and giving new start dates to him/her. This problem was solved with the merge of this rows, depending on the minimum and maximum values of their presented duration dates.

The same issue with the dates of owners came up just like with the managers' file. There were breaks in the dates of the same owners. As a result, the same owner was represented in several rows in the data instead of one row. However, the reason was not the same as with the leaders. Looking at the codes of owners, which indicates their type of ownership I found that after a "break" in the data the owner got the same code, which means that the change in the type of ownership did not cause the problem. Unfortunately, from the data I was not able to reveal the reason behind the different rows for the same owners. On the other hand, this meant that the owner's share in the company remained the same in the company, which will matter afterwards, when majority ownership will be determined and marked. This way, this issue did not need any additional treatment, so I applied the same procedure as with the leaders dataset.

B.3.3 Merging the datasets

There were two files which contained information about the number of employees of firms. There was a file about the type of staff number and another with the exact number of employees by enterprises. However, neither of them was complete in the sense of the time period they represented. These two were merged. In the first category file, there were no exact data on the number of people at a firm, only ranges were given: There were 16 categories (1;2; 3 – 4; 5 – 9; 10 – 19; 20 - 49 ; 50 – 99; 100 – 149; 150 – 199; 200 – 249; 250 – 299; 300 – 499; 500 – 999; 1000 – 1999; 2000 – 4999; 5000-). In these cases the midpoint of the bounds were calculated and then used as mean number of employees for each firm. Since this was not the most accurate way of determining the number of people at firms, I only kept these observations until 2008 because the second file, which had more accurate data on employment numbers, has started from 2009. In the second file I had a monthly data from 2009 until 2017, however I needed yearly company observations to be able to merge it with the others. Thus, I calculated a mean employment number for each firm in each year. Afterwards, I merged this with the

categories file and got an overall variable of mean number of employees for each firm in each year of their operation.

Afterwards, the leaders and owners were merged with the files which contained background information on the related leaders and owners and with the one, which had information about the names of their positions. The connecting persons and companies were identified according to the ID number in the main files. The supplementary file about the individuals had information on the gender, nationality and date of birth of the persons. The other file had information about the connecting companies' nationality, year of formation and the name of city of its headquarters'.

The codes file helped gaining more information about whether the leader is a CEO or a liquidator or another type of leader but without the rights of a CEO. Since, I was interested in the CEO-founders of a company, I only kept the leaders with a CEO title. In the case of owners file, the codes provided additional information about whether the owner is majority shareholder or not, thus I was able to recognize more majority shareholders because the majority of the data about ownership ratios were missing. As a result of these merges, several new variables could be created which were used as controls later in the applied models.

The final database was created by merging the number of employees, the owners, the leaders, the financial reports, and the basic information files together. Only those observations were kept where there was a perfect match.