To Schumpeter or not to Schumpeter, That is the Endogenous Growth Question: An Empirical Approach to Post-Soviet Countries

by

Jacquelyn Draper

B.A. (Hons), University of Manitoba, 2018

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Economics
Faculty of Arts and Social Sciences

Approval

Name:	**Jacquelyn Draper**
Degree:	**Master of Arts (Economics)**
Title:	**To Schumpeter or not to Schumpeter, That is the Endogenous Growth Question: An Empirical Approach to Post-Soviet Countries**
Examining Committee:	**Chair:** Alex Karaivanov Professor **Fernando Aragon** Senior Supervisor Associate Professor **Kenneth Kasa** Supervisor Professor **Kevin Schnepel** External Examiner Assistant Professor Department of Economics
Date Defended:	**April 14, 2020**

Abstract

This paper examines three empirical predictions of the Schumpeterian endogenous growth theory (1) the positive relation between innovation and growth rate, (2) the inverted U-shape relationship between competition and innovation, and (3) the positive relation of industry R&D spread on R&D investment. I examine the case of 10 Post-Soviet countries over the period 1996-2016. My analysis considers both aggregate and firm-level data while using a panel data methodology with fixed effects approach. My findings provide evidence of a positive relation of R&D investment, a measure of innovation, with growth. Meanwhile, there also appears to be a "stepping on toes" effect on growth with the R&D share, a measure of innovation commonly used in the theoretical literature. I find no evidence of an inverted-U relation at the industry-level. Instead, I find a positive relation between the industry spread of R&D investment and the firm's own investment. This finding suggests catching-up decisions by laggard firms and continued investment by lead firms.

Keywords: Macroeconomic Growth; Schumpeterian Theory; Patents; Competition; R&D Investment; R&D Industry Spread; Inverted-U

Dedication

To my father, Kevin, you are my biggest fan and I could never thank you enough for your love and support. You have always believed in my dreams and encouraged me along the way even if times seemed bleak. Thank you for raising me to believe that a girl can do anything a boy can and that hard work will see me through. You taught me to reach for the stars, trust in my abilities, and that I could do anything I put my mind to. This paper is proof of all the strength and hard-work that you have instilled in me.

Acknowledgements

I would like to acknowledge both of my supervisors, Professor Fernando Aragon and Professor Kenneth Kasa. Throughout this process I have benefited tremendously from their knowledge, encouragement, and invaluable time spent advising me.

I would like to thank the administration staff in the Department of Economics. To Ms. Lisa Agosti, thank you for answering my numerous questions, guiding me through the administrative procedures, and encouraging me throughout this entire process. I would like to acknowledge Professor Alex Karaivanov for Chairing my thesis defence, and Professor Kevin Schnepel for filling the role of External Examiner.

Contents

Approval **ii**

Abstract **iii**

Dedication **iv**

Acknowledgements **v**

Table of Contents **vi**

List of Tables **viii**

List of Figures **ix**

1 Introduction **1**

2 Literature Review **4**

3 Data Description **7**
- 3.1 Country Level Data . . . 7
- 3.2 Firm Level Data . . . 9
  - 3.2.1 R&D Variable . . . 9
  - 3.2.2 Measure of Competition . . . 9
  - 3.2.3 Measure of Spread . . . 10

4 Methodology **12**
- 4.1 Country Panel . . . 12
- 4.2 Industry Panel . . . 14

5 Results **16**
- 5.1 Aggregate Level Results . . . 16
- 5.2 Industry Level Results . . . 20
- 5.3 Robustness Checks . . . 23
  - 5.3.1 Aggregate Level . . . 23

5.3.2 Industry Level 27

6 Conclusion **33**

References **35**

Appendix A Additional Tables and Figures **37**

A.1 Country Level 37

A.2 Industry Level 48

List of Tables

Table 3.1 Country Level Averages for Innovation Variables 8
Table 3.2 Variable of Interest Averages by Industry Group 11

Table 4.1 Independent Variable Summary Statistics 13
Table 4.2 Independent Variable Summary Statistics 15

Table 5.1 Main Results 19
Table 5.2 Competition and Spread Effects on R&D 22
Table 5.3 Investment and Trade Variables 24
Table 5.4 4 Year Averages 26
Table 5.5 Alternative Variables 28
Table 5.6 Control and Fixed Effects Robustness 30
Table 5.7 Sample Robustness Check 32

Table A.1 Cross-Correlation Table 41
Table A.2 Trend Variable Inclusion 46
Table A.3 Alternative Innovation Measures 47
Table A.4 Firm Observations 49

List of Figures

Figure 5.1 Average Correlation of *log* Innovation & GDP per Capita 17
Figure 5.2 Average Correlation of R&D Share & GDP per Capita 17
Figure 5.3 *ln* R&D and Competition 20
Figure 5.4 *ln* R&D and R&D Spread 20

Figure A.1 Hausman Test Results - FE vs. RE Model 37
Figure A.2 Test for Heteroskedasticity Within the Panel Errors 38
Figure A.3 Woolridge Test for Serial Correlation 38
Figure A.4 Pesaran Test for Cross-Sectional Independence 38
Figure A.5 Bayesian Model Averaging Output 39
Figure A.6 High Filing Countries 42
Figure A.7 Medium Filing Countries 43
Figure A.8 Low Filing Countries 43
Figure A.9 Correlation of *log* Innovation and R&D Shares - Averages 44
Figure A.10 Growth Rate By Country 45
Figure A.11 Kernel Density of *ln*R&D 48
Figure A.12 Kernel Density of Competition 48

Chapter 1

Introduction

Since the Industrial Revolution, innovations have constantly evolved producing technologies that have reshaped industries to be more productive while, at the same time, economic growth has increased exponentially. Within macroeconomic theory, there is a wealth of literature searching to answer the question as to how macroeconomic growth (increases in Gross Domestic Product (GDP)) occurs. From Solow (1956) to Lucas (1988) to Romer (1990), many economists have sought to model economic growth in order to better explain the influencing factors as nations converge towards a steady state.

This paper focuses on providing empirical evidence in support of the Schumpeterian endogenous growth theory specifically for the case of Post-Soviet countries. These countries provide a unique opportunity to explore the relationship of innovation on growth in countries whose human capital investment remained high even in periods of low growth. Furthermore, Post-Soviet economies underwent periods of state ownership or investment which resulted in monopoly (un-levelled) markets and thus, with the fall of the Soviet union, competition increased as small private firms were able to develop. This provides an interesting case to analyse how the microeconomic foundations of the Schumpeterian theory hold as industries progress to be more competitive (levelled).

Schumpeterian endogenous growth theory, although quite extensive in its ability to explain growth, predicts many different effects that can offset each other. Theory predicts a positive relation between innovation and growth. However, the number of skilled labour employed in R&D, a measure of innovation commonly used in Schumpeterian modelling, can have both positive and negative effects on growth. The industry level theory predicts an inverted-U shape relation between competition and R&D investment, and a positive relationship of industry technology spread.

To test these predictions, I employ a panel data methodology on data from 10 Post-Soviet countries over the period 1996-2016. I use a Bayesian Model Averaging approach (Moral-Benito, 2012) for panel data to determine additional control variables for the aggregate regressions. The fixed effects for the growth analysis are at the country level with specific time dummy variables to control for major shocks.[1] For the industry regressions, I used industry, country, and year fixed effects, as well as cluster robust standard errors at the industry level.

I find a positive relationship between R&D innovation and growth that is statistically significant, however, there is also a negative statistically significant result for the R&D share of researchers that counteracts any positive correlation of innovation. This finding describes a "stepping on toes effect" where more researchers could deter patent production due to many researchers exploring the same idea without knowing it. At the industry level, the inverted-U shape relation is not present even with the alternative competition measure. The main results from the industry regression show that the magnitude of the industry R&D spread positively relates to R&D investment, and using Distance to the Frontier (DTF) spread as an alternative provides the same results. Both aggregate and industry level results are robust to alternative variables, additional controls, and sampling changes.

Schumpeterian literature has documented mixed results when analysing the competition inverted-U, and this paper further concurs with papers unable to find strong evidence of this relationship (Polder and Veldhuizen (2012) and Tingvall and Poldahl (2006)). I expand on the existing literature by providing evidence that spreads within industries are the driving force for R&D innovation, and also by using more than one country for my analysis. The use of multiple countries will provide a broader approach to analysing how the theory holds with a group of countries. Furthermore, I use the industry spread of R&D, instead of the commonly used DTF spread, to see if the lack of conclusive findings in the literature when it comes to industry spread is due to the type of measure being used.

The main contribution of this paper is to document the positive relation between R&D innovation and the growth rate. The literature remains mostly theoretical when analysing the effects of innovation on growth. Therefore, by evaluating the Post-Soviet case, I am able to begin to fill the gap in the empirical side of the literature, while also exploring countries whose structure of high human capital but low growth makes them a unique study for Schumpeterian theory. Lastly, this paper contributes to the existing literature by merging both the aggregate and industry level analysis into one cohesive paper so that the Schumpeterian theory can be fully analysed from the micro foundations and industrial organization to the endogenous growth theory.

This paper is organized as follows: Chapter 2 explores the current literature on Schumpeterian growth and competition. Chapter 3 describes the datasets and defines main variables used for both the country level regressions (section 3.1) and industry level regressions (section 3.2). Chapter 4 explains the methodology followed for each analysis, country (section 4.1) and industry (section4.2) respectively, whilst also providing definitions for ad-

[1] Yearly fixed effects were too costly to the degrees of freedom and caused overestimation of the model due to the small sample size with robust standard errors.

ditional control variables. Chapter 5 presents initial graphs, main results and robustness regressions. Lastly, Chapter 6 concludes and provides future research extensions.

Chapter 2

Literature Review

Innovation and economic growth have been intertwined throughout history as systems of industry and trade have evolved. The industrial revolution catapulted the world into new economic prosperity as innovation and industry became rampant, leading to exponential increases in GDP per capita. However, it was not until economist Joseph Schumpeter (1942) began to describe the gale of creative destruction that the idea of an endogenous growth theory was introduced. Endogenous growth literature has attempted to explain the link between growth theory and variables within the production function that is optimally chosen to increase technological progress and GDP growth respectively. Schumpeter (1942) explains that during the process of creative destruction, competitive firms seek to gain monopoly rents through innovation. New technology makes the older technology obsolete. By seeking monopoly rents, old technology is destroyed, adoption of the new technology as the norm occurs, and that process is then repeated. It was not until much later that Schumpeter's words were presented in an economic model.

Schumpeterian growth theory is less widespread than alternative models such as that of Romer (1990) and Lucas (1988). The Romer model and Schumpeterian model both rely on innovation, however, Romer (1990) applies a horizontal integration approach for technology. The horizontal integration process leads to more diverse substitutable products being created within an industry, whereas the vertical integration process described by Schumpeterian theory, leads to more advanced products in the industry that effectively replace and "destroy" the older products. Lucas (1988) focuses on human capital investment as increased educational attainment leads to increased total capital, as he saw physical and human capital derive from separate technologies. Why do these differences matter? Well, Schumpeterian growth theory delves into macroeconomics, industrial organization, and microeconomic foundations, providing economists with the opportunity to analyse growth theory on multiple levels.

Aghion and Howitt (1992) formalized creative destruction in a theoretical model emphasizing that the share of skilled labour employed in R&D is chosen by firms to increase profitability over their competitors. The optimal choice of labour in R&D relies on the value of profits gained from innovation, and also the speed in which that innovation will become obsolete. Firm structure and entrepreneurship models also began to emerge as Schumpeterian theory strived to present a well rounded growth theory encompassing both macro and microeconomics (Aghion, Akcigit, & Howitt, 2015). From Schumpeterian theory there are several testable predictions: (1) there is a positive relationship between innovation and growth, (2) there is an inverted-U shape relation between competition and R&D, and finally, (3) industry R&D spread has a positive relationship on R&D.

Various Schumpeterian growth models have come to the same conclusion that innovation has a positive effect on the growth of GDP per capita. Aghion and Howitt (1992) provide the foundational model for creative destruction and growth using the R&D share for innovation determination. Since then, the literature has expanded to address R&D input growth (Howitt, 1999), product innovation (Lentz & Mortensen, 2008), scale effects in endogenous growth (Segerstrom, 1998), patent races (Zeira, 2011) and patent protection (O'Donoghue & Zweimüller, 2004).[1] All of which come to the conclusion that innovation has a positive effect on growth while addressing various theoretical questions. This paper will help to fill the literature gap on the empirical side by evaluating how R&D variables affect the growth rate for Post-Soviet countries. Not only will this begin to bridge the gap between theoretical models and empirical models, but also expand on current patent literature by analysing how patents and trademarks result in increased growth rates.[2]

The use of patents and trademarks as a proxy for innovation at the aggregate growth level allows for continuity when transitioning into the industry dynamics of Schumpeterian theory. At the industry level, intangible assets encapsulate patents and other research expenditures such as trademarks and technological software. Using this as a dependent variable allows for many different types of innovation expenditure to be captured across various industries. McGrattan and Prescott (2010) use intangible assets to analyse a relationship between technology capital and the US current account, while Corrado, Hulten, and Sichel (2009) have used intangible assets when evaluating macroeconomic growth. Through the use of patents and trademarks for the aggregate innovation variable of interest, as well as, the dependent variable for the industry regression, this paper is able to connect the different literature of growth theory into one empirical analysis.

The second and third predictions being tested in this paper stem from the industrial organization and firm dynamic side of Schumpeterian literature. Incentives to gain a short-term monopoly in order to best the competition is the foundation for Schumpeterian theory

[1] Jones (1995) was quick to criticise Aghion and Howitt (1992) by noting that in developed countries, GDP growth seems to slow while the share of employees in R&D grows at a much higher rate. Responses to this critique have included scale effects (Howitt, 1999), and different researching effects such as "standing on shoulders" and "stepping on toes"(Segerstrom, 1998). These additions, although theoretically being able to explain the increase in the R&D share, do not present empirical evidence to counteract the critiques of Jones (1995).

[2] Theoretical models includes those of Aghion and Howitt (1992), Howitt (1999), Segerstrom (1998), and more within the literature but not cited within this paper.

at the firm level. This then implies that higher competition would increase the incentive to innovate under neck-in-neck type industries, known as the escape effect (Aghion et al., 2015). However, there is an alternative effect for un-levelled industries in which Schumpeterian competition leads to an inverted-U shape as competition in the market increases (Futia, 1980). Furthermore, the use of technology gap (or spread) as an interaction term with competition, is commonly applied as a control in the inverted-U literature.[3]

The prediction of an inverted-U relationship was first empirically tested by Aghion et al. (2005) who use UK firm data to analyse the relationship of industry competition and innovation investment. They find evidence of this inverted-U using the Lerner index as the measure of competition. Polder and Veldhuizen (2012) use firms in the Netherlands to test for an inverted-U with the Price Cost Margin (PCM) and Profit Elasticity (PE), finding mixed results as the prediction holds for the PE measure but not for the PCM. The inverted-U is again not conclusively found by Tingvall and Poldahl (2006) who use both the Herfindahl index and PCM. The Herfindahl results show a statistically significant inverted-U but again the PCM does not. The support for Aghion et al. (2005) is dependent on the competition indicator used and which countries are being evaluated (Peroni & Ferreira, 2012). This paper will empirically test the Lerner (PCM) measure of competition, along with the Herfindahl index in order to see if the inverted-U shape is indeed sensitive to these measures of competition and is present for Post-Soviet economies.

The final prediction of Schumpeterian growth theory is that the industries technology spread has an increasing effect on R&D. This prediction explains the Schumpeterian effect. As firms innovate to catch-up to the leaders, they invest less in R&D as less profits are obtained the closer they get to the frontier (Aghion et al., 2015). I use the industry spread of R&D investment as a proxy for technology spread and include the DTF spread as an alternative measure in order to compare with existing literature. When the DTF spread is included in the competition model there are positive effects on R&D investment (Polder and Veldhuizen (2012); Tingvall and Poldahl (2006)), which suggest that firms approaching the frontier reduce their R&D investment. However, Peroni and Ferreira (2012) do not find conclusive evidence of the positive relationship between spread and R&D as the DTF variable results are unstable. The inclusion of an interaction term between DTF spread and competition has inconclusive results as there are findings of positive effects (Aghion et al., 2005) and negative effects (Polder & Veldhuizen, 2012). The empirical literature does not have absolute evidence for how the spread of R&D effects innovation. My use of industry R&D investment spread will hopefully bring about a more stable result to predict how R&D spreads effect R&D investment.

[3]See Aghion, Bloom, Blundell, Griffith, and Howitt (2005); Polder and Veldhuizen (2012);Tingvall and Poldahl (2006); Peroni and Ferreira (2012).

Chapter 3

Data Description

The data for both the firm and aggregate analysis focuses on 10 Post-Soviet countries. These countries include Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, and Slovenia. These developing economies in Central and Eastern Europe reflect economies which were formally independent from the former Soviet Republic yet were structurally influenced by Russia up until the fall of the Soviet Union.

3.1 Country Level Data

The aggregate country level data was obtained from the World Databank (Worldbank, 2019a) and provides the country aggregates for GDP, education enrolment, number of researchers, research expenditure, gross capital formation, and net exports. Missing research data was collected from the World Databank source UNESCO institute for statistics (UNESCO, 2019). The patent and trademark data from World Databank is derived from the World Intellectual Property Organization Database.

The panel covers the 10 countries stated above and spans 1996-2016. The panel is mostly balanced with around 200 observations per country.[1] Due to the small sample of countries and longer time period, there are several econometric concerns that will be discussed in section 4.1.

The main dependent variable is the growth of per capita GDP. Following Mankiw, Phelps, and Romer (1995) and Hassan, Sanchez, and Yu (2011), I define this variable as follows:

$$GROWTH_{it} = log(\frac{Y}{N})_{it} - log(\frac{Y}{N})_{it-1}, \tag{3.1}$$

where Y_{it} is the real GDP and N_{it} is the population in time t. The main variable of interest is innovation (both physical and non-physical). Innovation can be measured through

[1] For the 4 year average models the data goes until 2015 to maintain five 4 year groupings, see table 5.7. The data begins in 1996 with the exception of Estonia for which data begins in 1998.Patent and trademark data is missing for Slovenia whose intellectual property data stops after 2011 due to internal auditing of the patent system from 2012-2017.

residential filings of patents and trademarks, which from this point on, will be referred to as INOVR. Using patents as a form of innovation proxy stems from the work of Zeira (2011) and O'Donoghue and Zweimüller (2004). The logarithm of INOVR provides the main variable of interest for the country level regressions and is measured as follows:

$$log(INOVR)_{it} = log(ResidentialPatents_{it} + ResidentialTrademarks_{it}) \tag{3.2}$$

R&D share is an alternative measure to INOVR, and follows the models of Aghion and Howitt (1992) and Segerstrom (1998). The R&D share is the share of researchers employed in the economy. Using number of researchers per million people, a variable was created to give the total researchers in the population. From there, total researchers was divided by the number of employed persons.

$$RDShare_{it} = \frac{TotalRD_{it}}{Employed_{it}} \tag{3.3}$$

In order to analyse other forms of R&D that may influence growth, R&D expenditure (RDE) can be used as another alternative measure for the effects of R&D on growth. R&D expenditure is measured as total R&D expenditure in each country's economy as a percentage of GDP. Table 3.1 provides each country's average for each innovation measure as well as the growth rate variable.

Table 3.1: Country Level Averages for Innovation Variables

	Growth	R&D Share	RDE	Innovation Stock	*log*INOVR (Innovation)
Bulgaria	0.036	0.006	0.549	4689.524	8.405
Czech Republic	0.023	0.009	1.334	8719.905	9.068
Estonia	0.039	0.008	1.190	1087.105	6.968
Hungary	0.025	0.007	0.9997	4300.333	8.353
Latvia	0 .05	0.006	0.505	1377.286	7.214
Lithuania	0.052	0.008	0.753	2009	7.589
Poland	0 .04	0.005	0.695	15998.52	9.675
Romania	0.035	0.004	0.456	8712.952	9.021
Slovakia	0.038	0.007	0.684	2543.19	7.83
Slovenia	0 .023	0.012	1.716	1554.625	7.321

Notes: Growth is the growth rate variable measured through the log difference of GDP per capita for each country-year. R&D Share is the number of researchers divided by the number of employed people for each country-year. RDE is research expenditure which is an aggregate measure as a percent of the countries GDP. Innovation stock is the number of patent and trademarks filed by residents and finally *log* Innovation is the logarithm of the number of filings.

3.2 Firm Level Data

The firm-level data was collected from the COMPUSTAT database, totalling an initial 17,629 observations for 1,493 firms from the aforementioned countries.[2] The date range initially spanned from 1993 to 2018, however, after dropping missing data the final sample consists of 1,083 firms spanning 1996 to 2018 with 7,376 observations. The panel is unbalanced due to firm entry and exit and there are 63 different industries represented by the firms. The firms were then aggregated by industry and country in order to provide industry level variables. The final outcome is 2,673 observations for 63 industries across 10 countries over the years 1996 to 2018. The number of observations and means for key variables per 4 digit industry group is reported in table 3.2.

All firm data is converted into US dollars, as the Euro was not introduced until 1999. The exchange data is from the Worldbank Global Economic Monitor (GEM) database (Worldbank, 2019b) and follows an annual exchange rate obtained from monthly averages by the respective country's official authority. Because firms have different fiscal year-ends, each annual exchange rate needed to be collected for each possible fiscal year-end month so that the correct exchange rate could be used for the individual firm.

3.2.1 R&D Variable

The dependent variable, *ln* R&D, is an industry sum of the intangible to total assets ratio in order to analyse changes to the industry over time.[3]

$$lnRD_{jct} = ln[\Sigma(\frac{(intangibles_{ijct} - goodwill_{ijct})}{totalassets_{ijct}})] \tag{3.4}$$

Innovation includes intangibles such as trademarks, patents, copyrights and so forth, but excludes goodwill. Furthermore, to account for outliers the data was trimmed through winsoring at both the top and bottom 1 percent of the RD variable. After the variable was summed by industry-country-year the natural logarithm was taken to produce the final dependent variable being used for the regressions.

3.2.2 Measure of Competition

The main measure of competition, C_{jct}, is one minus the industry Lerner index following Aghion et al. (2005). The firm Lerner index was initially calculated, followed by

[2]Financial institutions were dropped from the data as the reporting variables differed from the industrial format and the focus of this paper is to evaluate the R&D changes for industrial firms, not financial innovations.

[3]The natural logarithm is used to measure the proportion of the investment variable instead of the level. Appendix Figure A.11 shows that the *ln*R&D variable is normally distributed.

the average being calculated for each industry as shown in equation (4) and (5).

$$L_{ijt} = \frac{Revenue_{ijct} - VariableCost_{ijct}}{Sales_{ijct}} \tag{3.5}$$

$$L_{jct} = \frac{1}{N_{jct}} \Sigma L_{ijct} \tag{3.6}$$

Since the Lerner index represents the level of monopoly power in the industry, 1-L_{jct} will provide the level of competition in the industry.[4]

$$C_{jct} = 1 - L_{jct} \tag{3.7}$$

The higher the industry Lerner index the lower the competition level between firms. Following Aghion et al. (2005), the competition variable C_{jt} will also be squared in order to provide evidence, if present, of an inverted-U shape between competition and R&D. If this relationship exists, the competition variable will provide a positive statistically significant coefficient. The coefficient on the squared variable will be negative. Although the Lerner index provides a theoretical measure of competition, to check for a robust competition measure the Herfindahl index will be used.

3.2.3 Measure of Spread

SPREAD$_{jct}$: this term measures the spread of R&D investment in each industry for each country and year. It is calculated by subtracting the lowest proportion of R&D investment in the industry from the highest proportion. Following that calculation, the variable is divided again by the industry leader in investment. Spread thus shows the variation in R&D investment for each industry-country-year and will be used to analyse whether there is solely catch-up by laggard firms, or also continued investment by lead firms.

$$SPREAD_{jct} = \frac{RDmax_{ijct} - RDmin_{ijct}}{RDmax_{ijct}} \tag{3.8}$$

An alternative measure of spread used to check the robustness of the results is the Distance to the Frontier (DTF) spread. This variable measures the average industry spread of total factor productivity from the frontier firm in each industry-country-year (see section 4.2). Further control variables and all alternative measures are presented in the methodology section (4.2).

[4] See Figure A.12 for distribution of the variable.

Table 3.2: Variable of Interest Averages by Industry Group

GICS Industry Groups	*ln*R&D Mean	Competition Mean	R&D Spread Mean
Automobiles & Components	-5.491	0.143	0.144
Capital Goods	-4.55	.169	.169
Commercial & Professional Services	-5.136	0.191	0.455
Consumer Durables & Apparel	-3.527	0.259	0.333
Consumer Services	-4.542	0.206	0.490
Diversified Financials	-4.123	0.260	0.536
Energy	-4.82	0.23	0.23
Food & Staples Retailing	-4.376	0.135	0.392
Food, Beverages & Tobacco	-3.977	0.105	0.413
Healthcare Equipment & Services	-4.667	0.153	0.401
Household & Personal Products	-4.295	0.211	0.42
Insurance	-2.907	0.152	0.577
Materials	-3.032	0.251	0.763
Media & Entertainment	-4.903	0.132	0.45
Pharmaceuticals, Biotech. & Life Sciences	-2.016	0.231	0.516
Real Estate	-3.662	0.228	0.39
Retailing	-3.965	0.112	0.335
Software & Services	-2.042	0.294	0.467
Technology Hardware & Services	-3.445	0.162	0.5
Telecommunication Services	-2.842	0.15	0.17
Transportation	-5.067	0.211	0.246
Utilities	-4.506	0.116	0.505
N	2673	2673	2673

Notes: *ln*R&D is the measure of R&D investment and is the main dependent variable. It is the proportion of intangible assets adjusted for goodwill to total assets summed to the industry level for each industry-country-year. Competition is 1-Lerner, where Lerner is the Lerner index measuring monopoly power in a industry. R&D Spread is the distance between the top R&D investor and the lowest R&D investor in a industry-country-year, divided by the top investor.

Chapter 4

Methodology

4.1 Country Panel

This paper employs a panel data methodology to empirically test if the positive prediction of innovation affecting the growth of per capita GDP is present. The model reflects similar growth theory papers including Frankel and Rose (2002)who use the Mankiw, Romer, and Weil (1992) equation for their growth analysis with trade. Because the Frankel and Rose (2002) regression is cross-sectional, some adjustments have been made to reflect panel data growth theory.

I use panel data with country fixed effects and time specific dummy variables. The standard errors are corrected using the Driscoll-Kraay fixed effects approach. This approach provides more conservative standard errors that correct for heteroskedasticity, serial correlation, and cross-sectional dependence (Hoechle, 2007).

I select additional covariates using a Bayesian Model Averaging Approach (BMA) to panel data (Moral-Benito, 2012). The BMA approach allows additional macroeconomic and development variables to be fitted to the model and asses whether they should be included. This helps to avoid omitted variable bias when dealing with macroeconomic growth, which can be influenced by many factors. This approach suggests including average precipitation and life expectancy as additional controls.[1]

The dependent variable reflects the change in the growth rate of GDP per capita to see the effects of the independent variables on the growth path over time. The adjusted empirical model is as follows:

$$\begin{aligned} GROWTH_{it} = \alpha_i + \beta_1 logINOVR_{it} + \beta_2 OPEN_{it} + \beta_3 logPOP_{it} \\ + \beta_4 FINCRISIS_t + \beta_5 PRE2000_t + \beta_6 Z_{it} + \beta_7 log(Y)_{i0} + \epsilon_{it} \end{aligned} \tag{4.1}$$

[1] The posterior inclusion probability for 20 additional variables was used to determine if any should be included within the controls (see Figure A.5 for BMA output). Using a PIP of 0.75 as a baseline, only average precipitation and life expectancy had strong enough inclusion probabilities to be used. Only life expectancy was included in the models as average precipitation was omitted due to collinearity.

The dependent variable $GROWTH_{it}$ is the growth rate of GDP and therefore the coefficients reflect an increase or decrease in the rate of growth and not the level. The main variable of interest is *log* $INOVR_{it}$ (see section 3.1). There is country fixed effects used, and in place of time fixed effects, two time dummy variable controls are used for the major time series shocks.[2] The rest of the independent variables are: initial GDP per capita, openness, *log* population, tertiary enrolment, investment, gross domestic savings (GDS), European Union (EU) dummy, life expectancy, inflation, *log* R&D share, *log* R&D expenditure, FDI inflows, business R&D expenditure, government R&D expenditure, higher education R&D expenditure, and net exports. Descriptions for each independent variable are presented in Appendix A1.[3]

Table 4.1 presents the summary statistics for the independent variables. The mean, standard deviation, and number of observations, provide an initial description of the main variables of interest along with the control and robustness check variables. Dummy variables are not included as their values only take on 0 or 1 so their mean and standard deviations would not provide much information.

Table 4.1: Independent Variable Summary Statistics

	Mean	Standard Deviation
Initial GDP_{i0}	9.301	0.461
log$INOVR_{it}$	8.176	0.886
log RDS_{it}	-5.004	0.348
log RDE_{it}	-0.245	0.484
$Investment_{it}$	0.238	0.052
$Inflation_{it}$	11.994	74.475
$OPEN_{it}$	1.07	0.366
NX_{it}	-0.016	0.066
$SCHOOL_{it}$	56.262	17.577
log$Population_{it}$	15.633	1.019
GDS_{it}	0.227	0.054
Life $Expectancy_{it}$	74.011	2.623
GERD - BE_{it}	0.44	0.374
GERD - HE_{it}	0.222	0.148
GERD - GOV_{it}	0.216	0.08
FDI $Inflows_{it}$	5.303	7.095
Observations	208	208

Notes: INOVR is residential patent and trademarks. RDS and RDE are R&D share and expenditure. OPEN is exports plus imports as a share of GDP. NX is net exports. SCHOOL is tertiary school enrolment rates. GDS is gross domestic savings. GERD is R&D expenditure by business enterprises (BE), higher education (HE), and the government (GOV).

[2]Over estimation will occur if year FE are included. See Figure A.10 for growth variable plot and Table A.2 for inclusion of a trend variable

[3]Appendix Table A.1 provides a correlation matrix for key variables.

4.2 Industry Panel

The industry level panel also follows a fixed effects panel methodology. The fixed effects are at the industry, country and year levels. The logarithmic R&D term as the dependent variable can be interpreted as the percent change in the respective industries summed proportion of intangible assets.

$$\begin{aligned} ln(RD_{jct}) = \alpha_j + \lambda_c + \gamma_t + \beta_1 C_{jct} + \beta_2 C_{jct}^2 + \beta_3 SPREAD_{jct} \\ + \beta_4 lnSIZE_{jct} + \beta_5 N_{jct} + \beta 6 CI_{jct} + \epsilon_{jct} \end{aligned} \tag{4.2}$$

The competition variable and its square term are the main independent variables of interest in this regression. Alternative measures used for robustness are:

- lnRDI$_{jct}$ is the natural logarithm of R&D intensity. This is an alternative dependent variable used to check robustness of how competition and spread affect an alternative measure of R&D. RDI is calculated by intangible assets divided by value of sales.

- The Herfindahl-Hirschman Index (HHI) is the sum of squared market shares in the industry for each year. The market shares are represented by the sales of firm i in industry j at time t, that is:

$$HHI_{jct} = \Sigma(\frac{Sales_{ijct}}{\Sigma Sales_{ijct}})^2, \tag{4.3}$$

  the higher values of HHI indicate more concentration and less competition in the industry. Since a high HHI indicates less competition, there should be a negative coefficient for HHI indicating that R&D decreases with less competition.

- DTF Spread$_{jct}$: An alternative measure that can be used for technology spread in an industry is a comparison of total factor productivity (TFP) for the firm versus the most productive firms in that industry. This measure is called distance to the frontier (DTF). More productive firms in this case are assumed to have more efficient technology employed in their production process. This is calculated following Polder and Veldhuizen (2012):

$$DTF_{ijct} = \frac{TFP_{Fjct} - TFP_{ijct}}{TFP_{Fjct}} \tag{4.4}$$

$$DTF_{jct} = \frac{1}{N_{jct}} \Sigma DTF_{ijct}; \tag{4.5}$$

  Where TFP$_{Fjct}$ ref*l*ects the frontier firm for the industry. DTF$_{ict}$ is the distance to the frontier for the individual firm and DTF$_{jct}$ is the average distance to the frontier for the industry.

Additional control variables include firm size, number of firms in the industry (N), capital intensity (CI), return on assets (ROA), gross profit margin (GPM), and deferred taxes and investment credits (DTIC). Descriptions for each of these variables can be found in Appendix A2.

Table 4.2 provides the summary statistics for the right-hand side variables. The mean, standard deviation and number of observations provide an initial description of the main variables of interest along with the control and robustness variables to be used.

Table 4.2: Independent Variable Summary Statistics

	Mean	Standard Deviation
Competition$_{jct}$	0.177	0.11
Competition $^2{}_{jct}$	0.043	0 .061
HHI $_{jct}$	0.796	0 .272
HHI $^2{}_{jct}$	0.708	0.369
R&D Spread$_{jct}$	0.407	0.456
DTF Spread$_{jct}$	0.164	0.24
Capital Intensity$_{jct}$	1.961	3.262
ln Firm Size$_{jct}$	4.717	1.915
No. of Firms$_{jct}$	2.791	3.797
ROA$_{jct}$	-0.635	0.633
DTIC$_{jct}$	0.006	0.179
GPM $_{jct}$	0.396	0.239
Observations	2673	2673

Notes: HHI is the Herfindahl Hirschman Index defined in the above section. DTF Spread reflects the spread within each industry-country-year for the average firm to be at the technology frontier. ROA is the return on assets measured by net income divided by total assets. Gross profit margin (GPM) is the gross profit divided by revenue.

The initial regressions for the industry level variables will be conducted with the above regression as a base model. For robustness, the alternative measure of competition (HHI variable), the alternative measure of spread (DTF Spread), and an alternative measure for R&D, denoted RDI for R&D intensity, will be used to compare if alternative measure choices hold robust results (See table 5.5). Table 5.6 addressed removing and adding fixed effects and control variables as a further robustness check. Lastly, Table 5.7 addresses sampling robustness by changing the sample through various methods. The results from all regressions are presented in section 5.

Chapter 5

Results

5.1 Aggregate Level Results

Figures 5.1 and 5.2 show the correlation for the R&D variables of interest and the logarithm of GDP per capita. Table 5.1 presents the main estimation results for equation (4.1). These results show a positive relationship of innovation and a negative relationship of the R&D share on the growth rate.[1] Robustness of these results can be found in Tables 5.3 and 5.4.

The correlation story with innovation is not a clear one according to Figure 5.1. There is a slight negative correlation between GDP per capita and the logarithm of innovation. This would suggest that countries with higher amounts of innovation filings have a lower GDP per capita. The Czech Republic (CZE) stands apart from the trend as there is a high amount of innovation as well as a high level of GDP per capita. Since the average was used for these plots there could be a time component as to how innovation filings changed over the years (see Appendix Figures A.6-8 for detailed filing trends). The correlations between the R&D share presents a much stronger story for how R&D leads to growth (Figure 5.2). There is a strong positive correlation between the R&D share and GDP per capita. Countries with higher shares of researchers have a higher GDP per capita. This correlation would seem to suggest that Aghion and Howitt (1992) and Segerstrom (1998) were correct in modelling the R&D share as an important determinant in economic growth. The relationship between *log* Innovation and R&D shares can be seen in the Appendix (Figure A.9).

[1] Appendix Table A.3 compares the different R&D variables from the literature (Innovation, R&D share and R&D expenditures). R&D expenditures is not statistically significant and so it is excluded from the growth analysis.

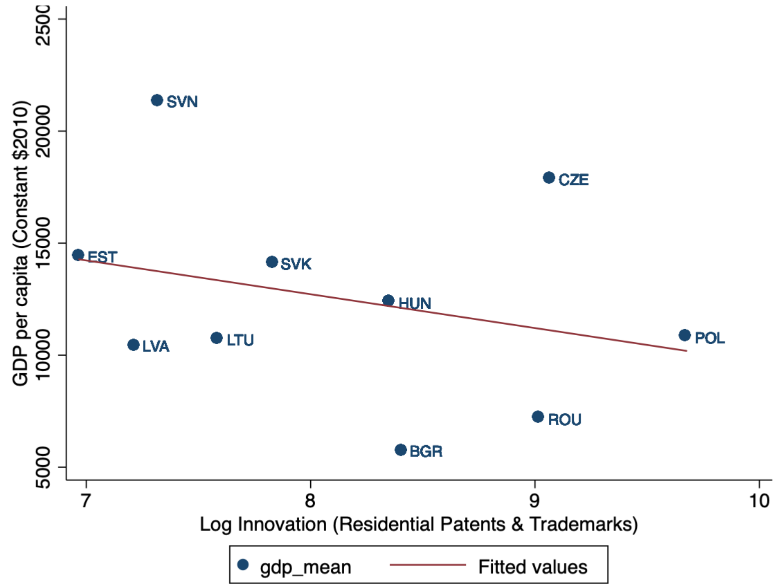

Figure 5.1: Average Correlation of *log* Innovation & GDP per Capita

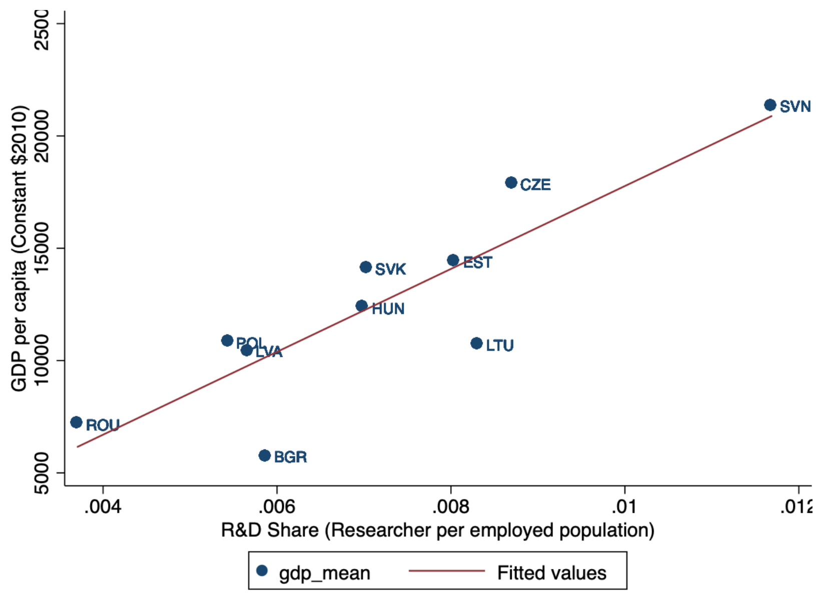

Figure 5.2: Average Correlation of R&D Share & GDP per Capita

I begin with the baseline estimates from equation (4.1). The main variable of interest is *log* Innovation as it pertains to the amount of ideas created in an economy and how that affects growth. A secondary variable of interest is *log* R&D Share as this reflects the proportion of people focusing on research and ideas in the economy. The results of column (4) show that innovation has a positive relationship with the growth rate of approximately 0.028 percent. This finding could reflect the "standing on shoulders" effect (Segerstrom,

1998) that ideas build off each other and produce more efficient innovations contributing to growth as time goes on. The *log* R&D Share, on the other hand, has a negative coefficient of approximately 0.031 percent. This could be capturing what is termed "stepping on toes" as there could be many researchers all researching the same idea without knowing it, thus the R&D share could increase while innovations stagnate. This finding could confirm why Jones (1995) found that growth for developed nations slowed while R&D shares increased. The stronger relationship of the two is the negative coefficient for *log* R&D share suggesting that a high R&D share could reflect an inefficient research sector. Both variables of R&D are statistically significant across the models which suggests that they should be considered by policymakers looking to increase their growth rates.

Columns (1)-(4) in Table 5.1 check the robustness of my results to a different set of controls. Column (1) shows the estimates of *log* Innovation with only the initial GDP, country fixed effects, and the dummy variables controlling for time specific effects (Financial Crisis and Pre-2000's). The time specific effects control for all major shocks and have the expected negative correlation with growth along with statistical significance.[2] Column (2) includes the variables suggested in Frankel and Rose (2002) which pertain to the neoclassical models. These controls lower the positive coefficient that *log* Innovation has from 0.07 percent to 0.02 percent but it is still statistically significant at the 10 percent level.

When additional controls in column (3) are added, the positive impact of *log* Innovation increases to approximately 0.03 percent and is statistically significant at the 1 percent level. The additional controls are gross domestic savings, the European Union dummy variable, and life expectancy.[3] Since Slovenia has missing patent and trademark data, column (5) excludes it from the regressions to see if that has any significant effect on the *log* Innovation coefficient. Excluding Slovenia causes the coefficient of *log* Innovation to increase from 0.028 percent to 0.034 percent, and also causes the negative coefficient of *log* R&D share to strengthen from 0.031 percent to 0.041 percent. The results remain strongly significant even with one less country.

[2]See Figure A.9 for time series graph of the growth variable and major shocks.There are tumultuous growth rate before 2000 presumably from Post-Soviet countries transitioning to capitalist economies. Furthermore, there are dips right before 2000 (Y2K could be an explanation), and during the financial crisis of 2008. In Tables 5.2 and 5.3 the time specific dummy variables are used but not reported as their coefficients are consistent with those reported in Table 5.1.

[3]Gross domestic savings is a fundamental aspect to the Solow-Swan model(Solow, 1956) and any other AK styled endogenous growth model. The EU dummy variable is a logical control to add as joining the EU provided increased markets for these Post-Soviet economies as well as the spread of ideas. Life expectancy was included due to the high posterior inclusion probability from the BMA analysis (see Figure A.5).

Table 5.1: Main Results

Dependent: Growth	(1)	(2)	(3)	(4)	(5)
Initial GDP	-0.0804***	-0.181***	-0.197***	-0.202***	-0.218***
	(0.0175)	(0.0520)	(0.0595)	(0.0593)	(0.0573)
log Innovation	0.0702***	0.0193*	0.0322***	0.0275**	0.0344***
	(0.0162)	(0.00856)	(0.00948)	(0.00897)	(0.00894)
Investment		0.488***	0.437***	0.431***	0.425***
		(0.118)	(0.0998)	(0.0940)	(0.106)
Openness		0.0768**	0.0752**	0.0834**	0.0750**
		(0.0309)	(0.0267)	(0.0273)	(0.0276)
log Population		-0.149	-0.245*	-0.219*	-0.207*
		(0.0845)	(0.112)	(0.0986)	(0.0941)
Tertiary Enrolment		0.0004	0.0003	0.0004	0.0006
		(0.000301)	(0.0003)	(0.0003)	(0.0003)
Gross Domestic Savings			0.235**	0.261**	0.319***
			(0.0808)	(0.0824)	(0.0919)
log R&D Share				-0.0311**	-0.0406**
				(0.0137)	(0.0146)
Financial Crisis	-0.0525***	-0.0425***	-0.0407***	-0.0397***	-0.0410***
	(0.0111)	(0.00904)	(0.00906)	(0.00924)	(0.00958)
Pre-2000's	-0.0262**	-0.0189*	-0.0211**	-0.0188*	-0.0195*
	(0.00842)	(0.00871)	(0.00818)	(0.00844)	(0.00888)
Constant	0.223	3.689*	5.578**	4.995**	4.734**
	(0.148)	(1.711)	(2.157)	(1.856)	(1.767)
N	203	203	203	203	187
Within R^2	0.413	0.591	0.624	0.633	0.643

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Driscoll-Kraay standard errors in parentheses. MA(2) process determined by the regression model. Dependent variable growth is the log difference of GDP per capita. All regression have both country fixed effects and time specific dummy controls (Financial Crisis and Pre-2000's). Columns (2)-(5) contains the inflation rate, while columns (3)-(5) also contain controls for life expectancy and European Union dummy, not included in the table. Column (5) excludes Slovenia in the regression to check the robustness of the results from data challenges.

5.2 Industry Level Results

Figures 5.3 and 5.4 begin the industry analysis by showing the plotted relationships between the dependent variable, *ln*R&D, and the two key independent variables of interest, competition (1-Lerner index) and industry R&D spread. The plots show an inverted-U shape for competition and an exponentially increasing function for R&D spread. Table 5.2 presents the main estimation results for equation (4.2). The findings do not show an inverted-U shape between competition and R&D, however there is a positive relation from the industry spread of R&D investment on R&D. This would suggest that the laggard firms are investing heavily to catch-up, while at the same time lead firms are investing to maintain monopoly positions. Robustness checks can be found in Tables 5.5-5.7.

Figure 5.3 shows the relationship between the *ln* R&D variable and competition. Using a fractional polynomial best fit line allows for flexibility in the relationship shown. There appears to be an inverted-U shape however most of the data is clustered towards a lower level of competition signalling monopoly or less competitive industries.[4] Figure 5.4 shows the relationship between the *ln* R&D variable and the R&D spread in the industry. The scatter plot again uses a fractional polynomial line to allow flexibility and non-linearities. Also, the data points at the frontier (spread=0) were excluded from the plot since the relationship of interest is investment behaviour as the spread in an industry is increasing. There appears to be an exponential patterned increase in R&D investment as the spread increases in the industry. Having these initial patterns visualised, the main results and robustness regressions will provide further evidence as to whether these patterns hold.

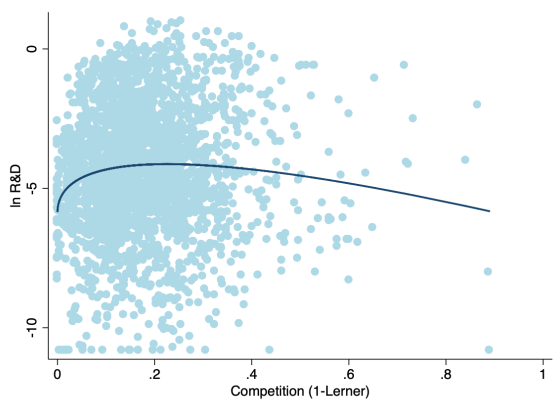

Figure 5.3: *ln* R&D and Competition

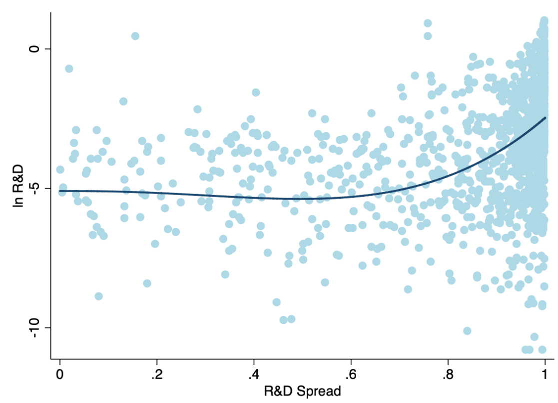

Figure 5.4: *ln* R&D and R&D Spread

[4]The lower competition levels could be due to the Post-Soviet sample as many of the countries had state run industries which would have held monopoly positions in the industry. These positions could have persisted even after the fall of the Soviet Union and transition to capitalism.

The question as to whether competition or the industry R&D spread, are the driving forces of innovation investment is the focus of equation (4.2). The results from the initial regressions specifically test if there is any evidence of an inverted-U or if R&D spread has a positive or negative relationship with R&D investment (see Table 5.2). The inverted-U evidence is not present, which is concurrent with Polder and Veldhuizen (2012) findings when using the Lerner index method of competition measure. The competition and competition squared terms have the correct signs (positive for competition and negative for competition squared) however they do not hold statistical significance.

The Table 5.2 results for R&D spread provide statistically significant evidence that it may have a stronger correlation with *ln* R&D than the theorised inverted-U. Column (2) shows that R&D spread has a positive relationship with the *ln* R&D of approximately 1.3 percent. When competition and its squared term are added to the regression the positive relation increases to 1.32 percent. The coefficient of spread further increases to 1.9 percent when the interaction between spread and competition is included in the regression. This positive coefficient of spread could be describing a perpetual cycle of catch-up for laggard firms as they invest to reach the frontier, while at the same time, the firm at the frontier is investing to avoid losing monopoly rents. The variable of interest appears to be whether the industry has a large or small spread of innovation investment, and will be the focus of Tables 5.5 to 5.7.

The interaction term between R&D spread and competition explains how firm investment behaviour changes if both R&D spread and competition increase. The R&D spread and competition interaction term has negative statistically significant result of 3.58 percent. This result shows evidence in support of Aghion et al. (2005), that if competition increases in an un-levelled industry (increasing spread) then the effects will be negative as the Schumpeterian effect dominates.[5] The other control variables are important for microeconomic and Schumpeterian theory. Firm size and the number of firms in the industry help to control for firm entry and exit conditions. Increased firm size has a 0.179 percent positive relation with R&D investment which is logical as larger firms tend to stay in the market longer and tend to have available funds for R&D investment. The number of firms in the industry also has a positive correlation on R&D. As the number of firms increases, R&D investment will increase by approximately 16 percent. This signals that firms entering the market invest in R&D as soon as they enter in order to compete. Further controls are added in the robustness checks (Table 5.5 to Table 5.7).

[5] The Schumpeterian effect describes that as laggard firms invest to catch-up in an un-levelled industry the spread decreases and the competition increases. This in turn has diminishing returns to the laggard firms as they attempt to reach the frontier making it less appealing as they approach.

Table 5.2: Competition and Spread Effects on R&D

Dependent: *ln* R&D	(1)	(2)	(3)	(4)
Competition_{jct}	2.723		2.285	3.515
	(2.207)		(2.152)	(2.104)
$\text{Competition}^2{}_{jct}$	-2.082		-0.550	-1.618
	(3.444)		(3.387)	(3.217)
$\text{R\&D Spread}_{jct}$		1.296***	1.323***	1.915***
		(0.220)	(0.227)	(0.366)
Spread_{jct} $\text{*Competition}_{jct}$				-3.581**
				(1.604)
$\text{Capital Intensity}_{jct}$	-0.0974***	-0.0849***	-0.0985***	-0.0967***
	(0.0289)	(0.0263)	(0.0256)	(0.0268)
ln Size_{jct}	0.206**	0.131*	0.181**	0.179**
	(0.0788)	(0.0664)	(0.0698)	(0.0673)
$\text{No. of Firms}_{jct}$	0.205***	0.159***	0.157***	0.160***
	(0.0443)	(0.0347)	(0.0359)	(0.0354)
Constant	-6.173***	-5.831***	-6.429***	-6.598***
	(0.517)	(0.310)	(0.469)	(0.454)
N	2237	2237	2237	2237
Adj. R^2	0.506	0.540	0.545	0.549

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors in parentheses and clustered at the industry level. The regressions contain year, industry, and country fixed effects with interaction terms not reported. Competition is measured by (1-Lerner Index) in the industry-country-year. R&D spread is measured by (RDmax_{jct}-RDmin_{jct})/RDmax_{jct}, where RDmax is the (national) industry leader in proportional R&D asset investment and RDmin is the firm in the (national) industry with the lowest proportion of R&D asset investment. The remaining variables listed are controls included to address factors that could influence R&D investment. Control variables are aggregated to the industry level as the average firm for industry-country-year except number of firms reflects the count of each unique firm GVKEY in each industry-country-year. Capital Intensity is defined as capital expense to the firm divided by sales. *Ln* of Firm Size is the natural logarithm of the average firms total assets in a given industry and country. Number of firms controls for the number of firms in each industry per country for the sample.

5.3 Robustness Checks

5.3.1 Aggregate Level

To check the robustness of the results in Table 5.1, I analysed if additional investment variables or an alternative trade variable would cause inconsistencies. Table 5.3 presents the results of including Foreign Direct Investment (FDI) inflows, *log* R&D expenditure broken into three categories, and the share of net exports as an alternative to openness. Both *log* Innovation and *log* R&D share are robust to the inclusion of additional investment variables (columns (1)-(3)) and remain statistically significant. FDI inflows are not statistically significant, however with its inclusion the coefficient of the R&D variables of interest strengthen slightly. The *log* R&D expenditure variable is broken into three different sources of expenditure: business enterprises (business), higher education institutions (Higher Ed.), and the government. All categories of the R&D expenditure investment variables are statistically insignificant and do not significantly change the coefficients for *log* Innovation or *log* R&D share.

Table 5.3 presents interesting results when the share of net exports is used instead of openness. Not only does the positive relationship *log* Innovation have on the growth rate increase to 0.038 percent, but the *log* R&D share variable becomes statistically insignificant with a smaller negative relation with growth. This would suggest that the variable selected for trade matters when analysing how R&D shares affect growth. When net exports are used instead of openness, the trade variable could be negative as countries could have higher imports than exports. If the nation is an import dominant country then their local researchers may be inefficient and thus imports are cheaper or higher quality than the nationally produced items. Comparatively, by definition, the openness variable is always positive as it adds both imports and exports together as a way to measure total trade occurring as a share of GDP. Import focused nations and using net exports could explain why R&D shares become statistically insignificant while *log* Innovation remains significant and has a larger positive estimate. Increased innovations could lead to more exports for the nation and therefore lead to economic growth. Column (5) includes the additional investment variables but there are no significant changes that occur. The positive coefficient of *log* Innovation remain robust at approximately 0.03 percent with openness and 0.036 when net exports is used along with the additional investment variables. The "stepping on toes" effect appears to be sensitive to trade variable definitions but remains robust to additional investment terms.

Table 5.3: Investment and Trade Variables

Dependent: Growth	(1)	(2)	(3)	(4)	(5)
Initial GDP	-0.201***	-0.203***	-0.218***	-0.190**	-0.219**
	(0.0591)	(0.0594)	(0.0607)	(0.0745)	(0.0749)
log Innovation	0.0285***	0.0304**	0.0287**	0.0376***	0.0358***
	(0.00859)	(0.0105)	(0.0107)	(0.00706)	(0.00799)
log R&D Share	-0.0322**	-0.0343**	-0.0326**	-0.0102	-0.0145
	(0.0137)	(0.0137)	(0.0136)	(0.0153)	(0.0142)
Investment	0.437***	0.434***	0.441***	0.547**	0.523***
	(0.0967)	(0.0957)	(0.0970)	(0.173)	(0.161)
Gross Domestic	0.258**	0.260**	0.273**	0.195	0.237
Savings	(0.0816)	(0.0827)	(0.0901)	(0.275)	(0.268)
European Union	0.0149	0.0152	0.0170	0.0290*	0.0298*
	(0.00995)	(0.0101)	(0.0110)	(0.0156)	(0.0155)
Openness	0.0864**	0.0864**	0.0783**		
	(0.0287)	(0.0289)	(0.0292)		
Net Exports				0.190	0.153
				(0.225)	(0.220)
FDI Inflow	-0.000219	-0.000218	-0.000113		0.000282
	(0.000250)	(0.000249)	(0.000238)		(0.000191)
log R&D Expend		0.00210	0.00198		0.00366
(Business)		(0.00372)	(0.00432)		(0.00343)
log R&D Expend			0.0138		0.0216
(Higher Ed.)			(0.0102)		(0.0124)
log R&D Expend			-0.00691		-0.00875
(Government)			(0.00637)		(0.00568)
Constant	4.899**	4.807**	5.382**	4.859*	5.666*
	(1.843)	(1.773)	(1.800)	(2.560)	(2.568)
N	203	203	203	203	203
Within R^2	0.635	0.635	0.642	0.585	0.603

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Driscoll-Kraay standard errors are in parentheses. MA(2) process determined by the regression model. The dependent variable growth is the log difference of GDP per capita. All regression include country fixed effects and time specific dummy controls (financial crisis and pre-2000's). Columns (1)-(3) add alternative investment controls while column (4)-(5) includes the share of net exports instead of openness. Controls included but not presented are: inflation rate, life expectancy, and *log* population.

A common issue with panel regressions in macroeconomics is reverse causality or statistical insignificance when looking at short-run effects. To address this, Table 5.4 reports the results when five 4-year period averages were used instead of a continuous year to year regression. The periods are arranged as follows: 1996-1999, 2000-2003, 2004-2007, 2008-2011, 2012-2015.[6] When looking at the effects in the long-run, *log* Innovation has a larger positive relationship on the growth rate of approximately 0.14 percent when openness is used and 0.135 when net exports is used. The negative relationship of *log* R&D share also increases to approximately 0.148 percent when openness is used and 0.168 percent when net exports is used alternatively. In the long run both "standing on shoulders" and "stepping on toes" are once again present and statistically significant. This confirms that increasing the number of researchers in the economy might not be as beneficial as one thinks and may actually lead to lower growth rates. Instead, policymakers should try to incentivise innovation filings like that of patents and trademarks in order to increase the growth rate in the long-run. A further finding to note is that in column (6) when Slovenia is again excluded the positive estimate of *log* Innovation is greater than the negative estimate of *log* R&D shares at 0.228 percent and 0.205, respectively.

Table 5.4 also provides evidence that tertiary enrolment is statistically significant in the long-run, whereas in the short-run that variable was included but was not significant. This would suggest that education takes time to cycle through the economy which is logical if tertiary schooling takes several years to complete. The coefficient for tertiary enrolment is relatively small compared to the other variables only contributing to a positive increase of approximately 0.003 percent. Furthermore, life expectancy at birth also becomes statistically significant in the long-run and provides a positive relationship of 0.022 percent in column (3) and 0.026 percent in column (4). Something to note is that with the 4-year averages approach both trade variables and investment become statistically insignificant. Whether this is due to their effects being dominantly short term, or a sampling issue could be an area of future research.

[6] The variables were averaged across these 4 years, however the growth rate variable was the logarithm of GDP per capita in the last year of the period, less the initial GDP per capita (being the last GDP term from the previous period).

Table 5.4: 4 Year Averages

Dependent: Growth	(1)	(2)	(3)	(4)	(5)
Initial GDP	-0.447***	-0.561***	-0.690***	-0.726***	-0.904***
	(0.0648)	(0.0740)	(0.0162)	(0.0178)	(0.0196)
log Innovation	0.204**	0.131*	0.140*	0.135**	0.228***
	(0.0749)	(0.0598)	(0.0620)	(0.0515)	(0.0436)
log R&D Share			-0.148***	-0.168***	-0.205***
			(0.0298)	(0.0129)	(0.0121)
Investment		-0.114	-0.177	-0.833	-0.133
		(0.328)	(0.383)	(0.639)	(0.208)
Gross Domestic Savings			0.496**	0.901***	1.041***
			(0.154)	(0.175)	(0.178)
Openness		-0.0598	-0.0550		-0.0632
		(0.0530)	(0.0738)		(0.0396)
Net Exports				-0.781	
				(0.451)	
*log*Population		-0.308*	-0.430*	-0.549**	-0.342***
		(0.152)	(0.199)	(0.229)	(0.0512)
Tertiary Enrolment		0.0031***	0.0025***	0.003***	0.0033***
		(0.0005)	(0.0002)	(0.0002)	(0.0002)
Life Expectancy			0.0216*	0.0255**	0.0403***
			(0.0116)	(0.0108)	(0.00580)
Constant	2.710*	9.089***	9.693***	11.50***	7.684***
	(1.213)	(2.655)	(2.249)	(2.814)	(0.502)
N	49	49	49	49	45
Within R^2	0.701	0.772	0.817	0.819	0.898

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Driscoll-Kraay standard errors in parentheses. MA(2) process determined by the regression model. Dependent variable GROWTH is the log difference of GDP per capita. All regressions have country fixed effects and time specific dummy controls (Financial crisis and Pre-2000's). Year averages are: 1996-1999, 2000-2003, 2004-2007, 2008-2011, and 2012-2015. For the dependent variable growth, the log difference in GDP per capita is the last year in the period minus the last year from the previous period (initial GDP variable). Columns (2)-(5) contain inflation and European dummy as a control, not included in the table. Column (5) excludes Slovenia from the regression.

5.3.2 Industry Level

The first robustness test is to see whether alternative measures for competition and R&D spread report similar findings (see Table 5.5). Columns(1)-(3) keep the *ln* R&D as the dependent variable but alternative measures of competition and spread are used. Columns (4)-(6) use *ln* R&D intensity as the alternative dependent measure of R&D investment.[7] The use of HHI as the measure of competition does not support evidence of an inverted-U shape. Similarly to the Lerner index measure, HHI has the correct signs for an inverted-U but lacks statistical significance. R&D spread is statistically significant and has a positive coefficient in column (1) of 0.71 percent when HHI is used compared to the Lerner index measure of competition. Column (4) shows a slight decrease in the observed positive relation to 0.66 percent however this is still statistically significant even with both competition and dependent variable alternatives.

The alternative measure for R&D spread, DTF spread, has a positive and statistically significant relationship with *ln* R&D. The positive correlation is approximately 3.2 percent when using the Lerner competition measure, and the correlation decreases to 1.96 percent when the HHI measure is employed. Under the alternative R&D variable, *ln* R&D intensity, the positive coefficient for DTF spread is 2.87 percent decreasing to 1.504 percent with the Lerner measure and HHI measure, respectively. These results would suggest that even with alternative competition measures and definitions of R&D productivity, the spread of technology is still a driving force for investment. This result could describe that both catch-up and stay ahead type mentalities are being used by firms within the industry. It is important to note that in column (5) competition becomes statistically significant at the 1 percent level when alternative spread and R&D investment variables are used. This provides evidence against the inverted-U shape and actually promotes classical theory that competition has a positive upward sloping relationship with R&D (Polder & Veldhuizen, 2012).

[7] See section 4.2 for alternative variable explanations and equations.

Table 5.5: Alternative Variables

Dependent:	*ln* R&D			*ln* R&D Intensity		
	(1)	(2)	(3)	(4)	(5)	(6)
$Competition_{jct}$		2.700			6.740***	
		(2.020)			(2.410)	
$Competition^2{}_{jct}$		-1.293			-6.585	
		(3.307)			(4.101)	
HHI_{jct}	-2.597		-2.761	-3.199		-3.055
	(1.963)		(1.892)	(2.051)		(1.995)
$HHI^2{}_{jct}$	0.532		0.652	0.914		0.723
	(1.307)		(1.280)	(1.376)		(1.363)
R&D $Spread_{jct}$	0.709***			0.660***		
	(0.223)			(0.225)		
DTF $Spread_{jct}$		3.178***	1.959***		2.865***	1.504**
		(0.604)	(0.601)		(0.666)	(0.661)
Capital $Intensity_{jct}$	-0.0839***	-0.0919***	-0.0798***	0.0448	0.0299	0.0481
	(0.0267)	(0.0259)	(0.0271)	(0.0367)	(0.0348)	(0.0382)
ln $Size_{jct}$	0.146**	0.198***	0.153**	0.175**	0.260***	0.182**
	(0.0654)	(0.0703)	(0.0658)	(0.0718)	(0.0731)	(0.0732)
No. of $Firms_{jct}$	0.103***	0.0990***	0.0625*	0.119***	0.125***	0.0899**
	(0.0347)	(0.0362)	(0.0347)	(0.0369)	(0.0404)	(0.0397)
Constant	-3.825***	-6.379***	-3.738***	-3.773***	-7.173***	-3.690***
	(0.832)	(0.456)	(0.808)	(0.877)	(0.480)	(0.859)
N	2237	2237	2237	2237	2237	2237
Adj. R^2	0.554	0.548	0.559	0.534	0.534	0.535

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors in parentheses and clustered at the industry level. All regressions use industry, country and year fixed effects as well as the interaction effects, not reported. Competition is measured by (1-Lerner Index) in the industry-country-year. R&D spread is measured by ($RDmax_{jct}$-$RDmin_{jct}$)/$RDmax_{jct}$, where RDmax is the (national) industry leader in proportional R&D asset investment and RDmin is the firm in the (national) industry with the lowest proportion of R&D asset investment. Distance to the Frontier (DTF) Spread is measured by the average of (DTF top_{jct}-DTF_{jct})/DTF top_{jct} for the industry-country-year, where DTF top is the top firm's total factor productivity in the industry denoting "at the frontier". HHI is the Herfindahl Hirschman Index for the industry. The remaining variables listed are controls included to address factors that could influence R&D investment. Control variables are aggregated to the industry level as the average firm for industry-country-year and explained in detail in section 4.2.

Table 5.6 presents the estimates for how changes in controls and fixed effects influence the robustness of the main results. The results remain robust to removing fixed effects completely (1), adding back year and industry level fixed effects (2), and then finally adding country fixed effects as well (3). Competition becomes statistically significant at the 10 percent level when all fixed effects are removed and when all controls are added. Additional controls include the interaction term between competition and spread, return on assets, gross profit margin, and deferred tax and investment credits (DTIC). Not surprisingly, DTIC has a statistically significant and positive coefficient as deferred taxes and investment credits would incentivise firms to invest in innovation thereby avoiding high taxes. The positive estimates for competition when statistically significant are 3.03 percent (1) and 4.07 percent (5). These results are not robust when the fixed effects are added back into the model or when all controls are removed, and thus I would be cautious in declaring that competition has any positive causation for *ln* R&D.

The R&D spread variable remains robust to both fixed effect changes and control variable changes. When all fixed effects are removed the positive relationship of R&D spread and R&D investment is 1.213 percent. That correlation decreases to 1.041 percent when time and industry effects are included, and then increases to 1.323 percent when all fixed effects are returned. When all controls are removed the positive estimate of R&D spread is approximately 1.8 percent. This increase is predictable as some of the controls hold explanatory power in the regression and thus their removal would cause that explanatory power to go to the remaining variables in the regression. When the industry and additional controls are all added, the positive relation for spread changes slightly to 1.76 percent. The conclusion from removing and adding both fixed effects and control variables is that the R&D spread variable positive results are robust while competition continues to show no signs of an inverted-U.

Table 5.6: Control and Fixed Effects Robustness

Dependent: *ln* R&D	(1)	(2)	(3)	(4)	(5)
Competition_{jct}	3.034*	1.664	2.285	2.122	4.074*
	(1.647)	(1.778)	(2.152)	(2.057)	(2.106)
$\text{Competition}^2{}_{jct}$	-1.398	-0.877	-0.550	-3.283	-2.046
	(2.982)	(3.183)	(3.387)	(3.178)	(2.680)
R&D Spread_{jct}	1.213***	1.041***	1.323***	1.825***	1.761***
	(0.235)	(0.244)	(0.227)	(0.233)	(0.369)
Capital Intensity_{jct}	-0.0799**	-0.115***	-0.0985***		-0.0905***
	(0.0320)	(0.0344)	(0.0256)		(0.0276)
ln Size_{jct}	0.179**	0.267***	0.181**		0.162**
	(0.0691)	(0.0573)	(0.0698)		(0.0713)
No. of Firms_{jct}	0.218***	0.232***	0.157***		0.158***
	(0.0351)	(0.0365)	(0.0359)		(0.0345)
SPREAD*Competition_{jct}					-2.986*
					(1.601)
Return on Assets_{jct}					0.0189
					(0.180)
Gross Profit Margin_{jct}					0.300
					(0.512)
DTIC_{jct}					0.441***
					(0.0706)
Constant	-6.577***	-6.791***	-6.429***	-5.351***	-6.648***
	(0.392)	(0.410)	(0.469)	(0.259)	(0.568)
Year Fixed Effects	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Industry Fixed Effects	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Country Fixed Effects	*No*	*No*	*Yes*	*Yes*	*Yes*
N	2673	2259	2237	2237	2123
Adj. R^2	0.327	0.476	0.545	0.477	0.556

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors in parentheses and clustered at the industry level. Competition is measured by (1-Lerner Index) in the industry-country-year. R&D spread is measured by (RDmax_{jct}-RDmin_{jct})/RDmax_{jct}, where RDmax is the (national) industry leader in proportional R&D asset investment and RDmin is the lowest proportion of R&D asset investment. Columns (1)-(3) reduce and then add back fixed effects. Column (4) takes away all control variables leaving only the main variables. Column (5) adds the following control variables: Deferred Taxes and Investment Credit (DTIC) reflects the balance sheet data for each firm converted to dollars per hundred thousand; Return on Assets is the net income divided by the total assets; and Gross Profit Margin is the gross profit of the firm divided by the revenue.

The final group of robustness checks focuses on changes to the sample used. Table 5.7 provides six different sampling changes to check for the robustness of the variables to the data. Since the countries being analysed are all post-soviet economies, assessing whether the results remain robust to the time period after joining the European Union (EU) is an important factor. Column (1) presents the results for only including the years after each country achieved ascension into the EU. The post-EU results are robust for R&D spread as it maintains its positive effect of 1.35 percent at the 1 percent significance level. This would suggest that even after joining a trade union, where ideas, capital, and investment can move more freely, the level of variation within the industry still matters significantly for if firms chose to invest in R&D.

Columns (3), (4) and (6) provide very similar results for R&D spread with each estimation resulting in a positive statistically significant effect of between 1.34 percent to 1.39 percent. Excluding small sample sized countries in column (3) does not have a large impact on the spread coefficient. The most notably affected variables are *ln* size and competition squared. The squared competition term is only affected by having the sign flip to being positive when the small sample countries are dropped. The firm size variable is more significantly impacted by a reduced coefficient and going from a 5 percent level of significance to a 10 percent level of significance. Column (4) with the untrimmed data does not hold any significant changes from column (3) except for the slight decrease in spread from 1.39 percent to 1.34 percent, and the squared competition term once again becomes negative. Column (6) represents the service-type industries and thus many of the controls gauged towards manufacturing-type industries lose statistical significance. The spread variable, however, is robust in its results compared to the other models.

One of the more notable changes in Table 5.7 for the spread of R&D coefficient is in column (2). The positive relation decreases more than double to only 0.52 percent. This significant decrease in the coefficient along with a decrease in statistical significance to the 10 percent level would suggest that Poland matters quite a bit for these results. This is not surprising though as many of the observations come from Poland (see Appendix Table A.4) in the total sample of firms and thus future research could include a more balanced data sample. The last notable change in R&D spread occurs when only manufacturing type industries are included in the regression (5). The positive estimate decreases to approximately 1.03 percent and is statistically significant at the 5 percent level. Splitting the data into manufacturing provides evidence as to which type of industries are more affected by these variables. When comparing manufacturing industries to service industries, it appears that service industries are more affected by a wider spread in the industry R&D. This could be explained by the fact that telecommunications and media industries would be considered services yet they highly rely on technological efficiency and innovation. More research into individual industries could prove beneficial for explaining how spread affects these dynamics.

Table 5.7: Sample Robustness Check

Dependent: *ln* R&D	Post-EU (1)	Exclude Poland (2)	Exclude Small (3)	Non-Winsor (4)	Manu. (5)	Service (6)
Competition_{jct}	1.408	3.823	1.263	2.462	3.097	1.258
	(2.109)	(3.378)	(2.281)	(2.148)	(3.843)	(2.765)
$\text{Competition}^2{}_{jct}$	0.730	-1.758	0.144	-0.956	-0.662	0.562
	(3.331)	(6.431)	(3.485)	(3.382)	(5.583)	(4.299)
R&D Spread_{jct}	1.350***	0.520*	1.390***	1.339***	1.032**	1.339***
	(0.250)	(0.291)	(0.252)	(0.229)	(0.407)	(0.271)
Capital	-0.0988***	-0.0711	-0.0931***	-0.103***	0.0129	-0.121***
Intensity_{jct}	(0.0266)	(0.0554)	(0.0266)	(0.0265)	(0.0651)	(0.0215)
DTIC_{jct}	-203.2**	0.443***	0.428***	0.401***	0.387***	-55.60
	(76.74)	(0.0536)	(0.0672)	(0.0646)	(0.0719)	(154.8)
ln Size_{jct}	0.190**	0.276***	0.123*	0.173**	0.246***	0.136
	(0.0817)	(0.0863)	(0.0731)	(0.0705)	(0.0778)	(0.105)
No. of Firms_{jct}	0.153***	0.517***	0.158***	0.158***	0.305***	0.113***
	(0.0346)	(0.115)	(0.0400)	(0.0361)	(0.0564)	(0.0323)
Constant	-6.275***	-7.691***	-6.033***	-6.410***	-7.440***	-5.799***
	(0.507)	(0.607)	(0.507)	(0.467)	(0.606)	(0.677)
N	1894	1431	1990	2237	960	1228
Adj. R^2	0.562	0.454	0.555	0.544	0.490	0.588

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Standard errors in parentheses and clustered at the industry level. The regressions contain year, industry, and country fixed effects as well as the interaction effects, not reported. Competition is measured by (1-Lerner Index) in the industry-country-year. R&D spread is measured by (RDmax_{jct}-RDmin_{jct})/RDmax_{jct}, where RDmax is the (national) industry leader in proportional R&D asset investment and RDmin is the firm in the (national) industry with the lowest proportion of R&D asset investment. The remaining variables listed are controls included to address factors that could influence R&D investment. Control variables are aggregated to the industry level as the average firm for industry-country-year except number of firms reflects the count of each unique firm GVKEY in each industry-country-year. Capital Intensity is defined as capital expense to the firm divided by sales. Deferred Taxes and Investment Credit (DTIC) reflects the balance sheet data for each firm converted to dollars per hundred thousand. *Ln* of Firm Size is the natural logarithm of the average firms total assets in a given industry and country. Number of firms controls for the number of firms in each industry per country for the sample. Column (1) is the regression for the years after each country joined the EU. Column (2) is all countries excluding Poland, (3) includes Poland but excludes small sample size countries - Estonia and Slovakia. Column(4) uses the untrimmed (non-windsored) data. Columns (5)-(6) are dividing the data in to manufacturing type industries (5), and service type industries (6).

Chapter 6

Conclusion

Post-Soviet economies have provided an interesting case in which to analyse Schumpeterian growth and competition theories in both a macroeconomic and micro foundational manner. Due to their maintained emphasis on education during the Soviet regime, these satellite states provide a unique opportunity to analyse the relationship between R&D and the growth rate without major human capital changes. These countries also provided a unique opportunity to see whether the inverted-U shape relation between competition and R&D investment holds in the un-levelled industries and to test how the industry R&D investment spread matters for firm R&D investment decisions.

The results of equation (4.1) provide evidence that innovation through patent and trademark filings has a positive 0.028 percent correlation with growth in the short-run, and 0.14 percent in the long-run.[1] There is also a negative relationship present when including the R&D share variable used by Aghion and Howitt (1992) and Segerstrom (1998) in their theoretical models. This negative relation of the *log* R&D share is approximately 0.031 percent in the short-run, and 0.148 percent in the long-run. These two different relationships suggest that both a "standing on shoulders" positive effect and "stepping on toes" negative effect are occurring. These results would provide evidence that policymakers looking to increase the growth rate should provide incentives for innovation filings while also attempting to decrease the R&D share of researchers in the economy.

Since there was evidence that R&D matters for growth, the dynamics that lead to R&D investment at the industry-level provide further insight into how firm investment decisions are made. An inverted-U shape is not present in the results as neither the competition variable nor its squared term is statistically significant. This concurs with the findings of Tingvall and Poldahl (2006) and Polder and Veldhuizen (2012)(when the PCM is used). Occasionally throughout the robustness checks, the competition variable became statistically significant with a positive coefficient which could suggest that classical theory is correct. However, this result is not robust and I would be leery to state this as evidence

[1] All main result percentages are for when Slovenia is included in the regressions.

without further research. Further research with alternative competition measures and various subsamples of levelled and un-levelled industries could provide some clarity for the inability to find the inverted-U relation.

The main results of equation (4.2) suggest that as the industry spread of R&D increases it has a positive relation of approximately 1.3 percent with R&D investment. This suggests that laggard firms are investing to catch up to the frontier while leading firms are investing to maintain a competitive edge. Even using an alternative spread variable presents similar results showing that a wider spread in an industry has a positive relation with R&D investment. With the positive results of the spread variable, policymakers should not attempt to increase competition in oligopoly or diverse R&D industries. Instead, they should incentivise R&D investment to promote catch-up attempts from laggard firms and continued innovation investment by frontier firms.

From the findings in this paper, Schumpeterian growth theory should be considered by policymakers looking to increase their country's growth rate. Furthermore, policymakers should evaluate how industry policies affect the spread of R&D as it appears to have a greater impact than competition on R&D investment. Future research into how Schumpeterian growth holds for different income level countries throughout the world economy would strengthen these findings and enhance the empirical growth literature. At the industry level, more research into firm investment dynamics and how competition affects R&D would provide greater insight into how these two variables are connected. In conclusion, this paper is able to provide a first step empirical analysis into Schumpeterian growth theory and start to fill the gaps in the literature. The aggregate results show evidence in favour of Schumpeterian growth, while the industry level results are less conclusive. Further research would hopefully provide more conclusive results for the industry analysis while also strengthening the aggregate results.

References

Aghion, P., Akcigit, U., & Howitt, P. (2015). The schumpeterian growth paradigm. *economics*, *7*(1), 557–575.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The quarterly journal of economics*, *120*(2), 701–728.

Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, *60*(2), 323–351.

Corrado, C., Hulten, C., & Sichel, D. (2009). Intangible capital and us economic growth. *Review of income and wealth*, *55*(3), 661–685.

De Luca, G., & Magnus, J. R. (2011). Bayesian model averaging and weighted-average least squares: Equivariance, stability, and numerical issues. *The Stata Journal*, *11*(4), 518–544.

Dutz, M. A., & Vagliasindi, M. (2000). Competition policy implementation in transition economies: an empirical assessment. *European economic review*, *44*(4-6), 762–772.

Fosu, S. (2013). Capital structure, product market competition and firm performance: Evidence from south africa. *The quarterly review of economics and finance*, *53*(2), 140–151.

Frankel, J., & Rose, A. (2002). An estimate of the effect of common currencies on trade and income. *The quarterly journal of economics*, *117*(2), 437–466.

Futia, C. A. (1980). Schumpeterian competition. *The Quarterly Journal of Economics*, *94*(4), 675–695.

Hassan, M. K., Sanchez, B., & Yu, J.-S. (2011). Financial development and economic growth: New evidence from panel data. *The Quarterly Review of economics and finance*, *51*(1), 88–104.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The stata journal*, *7*(3), 281–312.

Howitt, P. (1999). Steady endogenous growth with population and r. & d. inputs growing. *Journal of Political Economy*, *107*(4), 715–730.

Jones, C. I. (1995). Time series tests of endogenous growth models. *The Quarterly Journal of Economics*, *110*(2), 495–525.

Lentz, R., & Mortensen, D. T. (2008). An empirical model of growth through product innovation. *Econometrica*, *76*(6), 1317–1373.

Lucas, R. (1988). On the mechanics of economic development. *Journal of monetary economics*.

Mankiw, N. G., Phelps, E. S., & Romer, P. M. (1995). The growth of nations. *Brookings papers on economic activity*, *1995*(1), 275–326.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, *107*(2), 407–437.

McGrattan, E. R., & Prescott, E. C. (2010). Technology capital and the us current account. *American Economic Review*, *100*(4), 1493–1522.

Moral-Benito, E. (2012). Determinants of economic growth: a bayesian panel data approach. *Review of Economics and Statistics*, *94*(2), 566–579.

O'Donoghue, T., & Zweimüller, J. (2004). Patents in a model of endogenous growth. *Journal of Economic Growth*, *9*(1), 81–123.

Peroni, C., & Ferreira, I. S. G. (2012). Competition and innovation in luxembourg. *Journal of Industry, Competition and Trade*, *12*(1), 93–117.

Polder, M., & Veldhuizen, E. (2012). Innovation and competition in the netherlands: Testing the inverted-u for industries and firms. *Journal of Industry, Competition and Trade*, *12*(1), 67–91.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, *98*(5, Part 2), S71–S102.

Schumpeter, J. A. (1942). *Socialism, capitalism and democracy.* Harper and Brothers.

Segerstrom, P. S. (1998). Endogenous growth without scale effects. *American Economic Review*, 1290–1310.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics*, *70*(1), 65–94.

Tingvall, P. G., & Poldahl, A. (2006). Is there really an inverted u-shaped relation between competition and r&d? *Economics of Innovation and New Technology*, *15*(2), 101–118.

UNESCO. (2019). Science, technology and innovation.

Worldbank. (2019a). World development indicators.

Worldbank. (2019b). World global economic monitor.

Zeira, J. (2011). Innovations, patent races and endogenous growth. *Journal of Economic Growth*, *16*(2), 135–156.

Appendix A

Additional Tables and Figures

A.1 Country Level

Appendix A.1 provides additional information on variable selection, model testing, and innovation filing history for the growth rate panel regressions.

Below (Figure A.1) is the Hausman test results for identifying if fixed or random effects is the appropriate methodology to use. The null hypothesis (random effects) is rejected and therefore there is correlation between the unique errors and regressors of the model. Thus, I will use a fixed effects model with both country and year fixed effects.

```
                    b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test:  Ho:  difference in coefficients not systematic

              chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =       57.21
            Prob>chi2 =      0.0000
```

Figure A.1: Hausman Test Results - FE vs. RE Model

To check for homoskedasticity or heteroskedasticity in the errors a Modified Wald Test (MWT) was used. Figure A.2 presents the STATA output from the test. The null hypothesis of homoskedastic errors was rejected so in order to account for the heteroskedasticity robust standard errors should be used. With panel data there are several ways for standard errors to be robust to heteroskedasticity. The most common practice to correct for heteroskedasticity and serial correlation is clustering. For my panel regressions there is also cross-sectional dependence which requires Driscoll-Kraay standard errors instead of clustered standard errors (See Figure A.4).

```
Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (10)  =      28.63
Prob>chi2 =      0.0014
```

Figure A.2: Test for Heteroskedasticity Within the Panel Errors

The Woolridge test presented in Figure A.3 checks the data for serial correlation. It is important to check if there is autocorrelation with macroeconomic data as the time period tends to be longer. The test for autocorrelation checks to see if there is serial correlation in the idiosyncratic errors. The null hypothesis is rejected at the 1 percent level that there is no auto-correlation and therefore robust standard errors can be used to correct for this. Again clustering is the common practice, but due to the small sample size and the additional problem of cross-sectional dependence, Driscoll-Kraay standard errors should be considered.

```
Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,       9) =     38.013
           Prob > F =      0.0002

.
```

Figure A.3: Woolridge Test for Serial Correlation

The Pesaran test for cross-sectional independence is an important test for ensuring correct standard errors are used. The null hypothesis is that there is cross-sectional **independence**. Since the null is strongly rejected, Driscoll-Kraay standard errors for a fixed effects model are used. These standard errors have a MA(q) process and the maximum lag chosen by the model is 2. Figure A.4 presents the output from the Pesaran test.

```
Pesaran's test of cross sectional independence =     6.081, Pr = 0.0000

Average absolute value of the off-diagonal elements =     0.278
```

Figure A.4: Pesaran Test for Cross-Sectional Independence

Bayesian Model Averaging results for additional controls. 1,048,576 models were evaluated to provide the coefficient, standard errors, posterior inclusion probability (PIP) and standard deviation. The posterior inclusion probability is the key focus for a BMA model selection according to Moral-Benito (2012). I only included variables with a PIP over 0.8

as it strongly suggests that the variable should be included in the regression. De Luca and Magnus (2011) created the "bma" code in stata and note that a PIP of 0.5 would be robustly correlated with the outcome. Since my data sample is small and additional variables may overestimate the model, I restrict my inclusion criteria to being over 0.8 PIP. Average precipitation, although highly suggested for inclusion, was omitted due to collinearity thus it was not included in the regressions.

GROWTH	Coef.	Std. Err.	t	pip	[1-Std. Err. Bands]	
_cons	1.829459	.8882672	2.06	1.00	.9411917	2.717726
log_lag_gdp	-.0952424	.0249007	-3.82	1.00	-.120143	-.0703417
log_INOVR	.0253505	.012176	2.08	1.00	.0131744	.0375265
log_pop	-.0318339	.0130497	-2.44	1.00	-.0448835	-.0187842
invest	.2692644	.0904278	2.98	1.00	.1788366	.3596921
GDS	.2535295	.7846877	0.32	1.00	-.5311582	1.038217
inflation_pri~s	-.0003339	.0002425	-1.38	1.00	-.0005764	-.0000913
open	.0768965	.0249616	3.08	1.00	.051935	.1018581
fin_crisis	-.051683	.0070951	-7.28	1.00	-.058778	-.0445879
pre2000	-.0136711	.0087753	-1.56	1.00	-.0224464	-.0048958
EU	.0084475	.0129896	0.65	1.00	-.0045421	.0214371
school_tertiary	.0001761	.0002997	0.59	1.00	-.0001236	.0004759
school_pri~y	-.0000462	.0002405	-0.19	0.08	-.0002867	.0001944
school_sec~y	.0001244	.0003803	0.33	0.14	-.0002559	.0005048
consump	.1187402	.7894015	0.15	0.29	-.6706614	.9081417
govt	.3321994	.7469223	0.44	0.44	-.4147229	1.079122
Urban_pop	-.0000334	.0003368	-0.10	0.07	-.0003701	.0003034
unemploy_p~F	-.0003039	.0007262	-0.42	0.20	-.0010301	.0004223
gfc_format~n	2.82e-13	6.25e-13	0.45	0.22	-3.43e-13	9.07e-13
pop_growth	-.0081619	.0088747	-0.92	0.53	-.0170366	.0007129
patent_non~s	-6.10e-08	8.33e-07	-0.07	0.06	-8.94e-07	7.72e-07
RD_total	3.83e-08	2.78e-07	0.14	0.08	-2.39e-07	3.16e-07
RD_expend	.0004427	.0033261	0.13	0.06	-.0028835	.0037688
in_VA	1.15e-12	8.18e-13	1.40	0.76	3.27e-13	1.96e-12
manu_VA	1.29e-13	6.84e-13	0.19	0.12	-5.55e-13	8.12e-13
tradema~nres	1.55e-06	2.83e-06	0.55	0.30	-1.27e-06	4.38e-06
lifeexpect	-.0147446	.0035265	-4.18	1.00	-.0182711	-.0112181
Average_pr~p	.0001741	.0000427	4.07	0.99	.0001313	.0002168
Mort_inf	-.002546	.0049992	-0.51	0.40	-.0075453	.0024532
Mort_5	-.0015861	.0040008	-0.40	0.36	-.0055869	.0024147
Pop_65	5.35e-10	4.10e-09	0.13	0.09	-3.56e-09	4.63e-09
trade	-1.73e-13	1.88e-13	-0.92	0.54	-3.61e-13	1.47e-14

Figure A.5: Bayesian Model Averaging Output

Appendix section A.1 further presents independent variable descriptions, figures and tables providing additional information, and robustness checks for the aggregate regressions.

Additional independent variable descriptions are presented below:

- OPEN_{it} is a variable representing the level of trade openness in an economy. This variable is created by adding together imports and exports and dividing them by the real GDP. Imports and exports are added together to reflect the total amount of trade occurring for each country in a given year.
- *log* POP_{it} is the logarithm of the total population for each country-year.
- FINCRISIS_{t} is a dummy variable taking on the value of 1 during the years 2008, 2009 and 2010 in order to account for the 2008 financial crisis.
- PRE2000_{t} is a dummy variable taking on the value of 1 for the years 1996-2000 to control for the tumultuous economic period after the Soviet Union fell and countries adjusted to capitalist markets.
- Z_{it} is a vector of further control variables:
  - SCHOOL_{it} is the tertiary school enrolment rate. This schooling variable was included to capture the human capital element of the model and control for level of education within the economy.
  - Investment_{it} is gross capital formation divided by the real GDP.
  - GDS_{it} is Gross Domestic Savings which is measured by real GDP less final consumption expenditure as a share of the GDP.
  - EU_{it} is a dummy variable taking on a value of 1 for the time period after the post-soviet countries individually entered the European Union.
  - Life Expectancy_{it} is the average life expectancy from birth measured in years.
  - Inflation_{it} is the inflation rate measured in prices.
  - *log* R&D Share_{it} is the logarithm of R&D Share (see section 3.1) measuring the number of researches per employed persons.
  - *log* R&D Expend_{it} is the logarithm of R&D expenditures as a percent of GDP.
  - FDI Inflows_{it} is the net inflow of foreign direct investment for each country.
  - GERD BE_{it} is the research expenditure as a percent of GDP performed by business enterprises.
  - GERD GOV_{it} is the research expenditure as a percent of GDP performed by the government.
  - GERD HE_{it} is the research expenditure as a percent of GDP performed by higher education institutions such as universities.
  - NX_{it} is net exports as a share of GDP which is an alternative measure to trade openness.
- $ln(Y)_{i0}$ is the initial GDP per capita for the time period. Since the first model regression is annual, the initial GDP per capita reflects the GDP from the previous year as the starting base of the current year.

Table A.1: Cross-Correlation Table

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1.GROWTH	1.000											
2. Initial GDP	-0.225	1.000										
	(0.001)											
3. *log*INOVR	-0.043	-0.122	1.000									
	(0.546)	(0.082)										
4. *log* RDS	-0.179	0.661	-0.411	1.000								
	(0.010)	(0.000)	(0.000)									
5. *log* RDE	-0.234	0.799	-0.115	0.805	1.000							
	(0.001)	(0.000)	(0.102)	(0.000)								
6. Investment	0.251	0.370	-0.032	-0.037	0.150	1.000						
	(0.000)	(0.000)	(0.647)	(0.599)	(0.030)							
7. GERD-BE	-0.245	0.670	-0.051	0.707	0.881	0.118	1.000					
	(0.000)	(0.000)	(0.467)	(0.000)	(0.000)	(0.090)						
8. GERD-HE	-0.055	0.563	-0.343	0.542	0.586	0.060	0.284	1.000				
	(0.431)	(0.000)	(0.000)	(0.000)	(0.000)	(0.386)	(0.000)					
9. GERD-GOV	-0.138	0.269	0.162	0.447	0.492	0.011	0.412	-0.003	1.000			
	(0.046)	(0.000)	(0.021)	(0.000)	(0.000)	(0.870)	(0.000)	(0.968)				
10. Openness	-0.079	0.610	-0.342	0.636	0.526	0.101	0.412	0.536	0.070	1.000		
	(0.257)	(0.000)	(0.000)	(0.000)	(0.000)	(0.148)	(0.000)	(0.000)	(0.313)			
11. Net Exports	-0.354	0.138	0.079	0.335	0.367	-0.602	0.372	0.147	0.236	0.268	1.000	
	(0.000)	(0.046)	(0.264)	(0.000)	(0.000)	(0.000)	(0.000)	(0.034)	(0.001)	(0.000)		
12. Tertiary Enrol.	-0.071	0.501	-0.178	0.446	0.392	0.031	0.284	0.576	0.038	0.500	0.010	1.000
	(0.310)	(0.000)	(0.011)	(0.000)	(0.000)	(0.659)	(0.000)	(0.000)	(0.582)	(0.000)	(0.886)	

GROWTH reflects the log difference in GDP per capita (dependent variable). INOVR is the innovation variable measured by residential filing of patents and trademarks. RDS is the R&D employment share. RDE is the R&D expenditure as a percent of GDP. GERD-BE, GERD-HE, GERD-GOV represent the R&D expenditures as a percent of GDP for business enterprises (BE), higher education (HE), and the government (GOV).

Figures A.6-A.8 present the time series by filing group in order to better understand how the main independent variable, *log* Innovation, changes over time. Intellectual property filings are not always consistent over time or across countries. In order to evaluate how filing trends have evolved, each country was assigned to a high, medium or low filing group. If the average number of filings for patents and trademarks was greater than 8000 then the country was considered high filing, between 2500-4000 medium filing, and below 2500 low filing.

The high filing countries maintain a much more stable plot of patents and trademarks compared to the middle and low filing groups. There appears to be a slight upward trend for Poland and Romania, while the Czech Republic is relatively constant. The medium and low filing groups show higher variance in their time series. The medium filing group seems fairly robust to shocks, however, Bulgaria experienced a steep increase in innovation filings up until around 2006 when the filings sharply decline. There is a slight increasing trend for these countries but it is not very significant. The low filing countries not only experience much more instability in filings, but they also diverge after the 2008 financial crisis. Estonia and Latvia have more of a constant trend, whereas Lithuania and Slovenia have an obvious increasing trend.

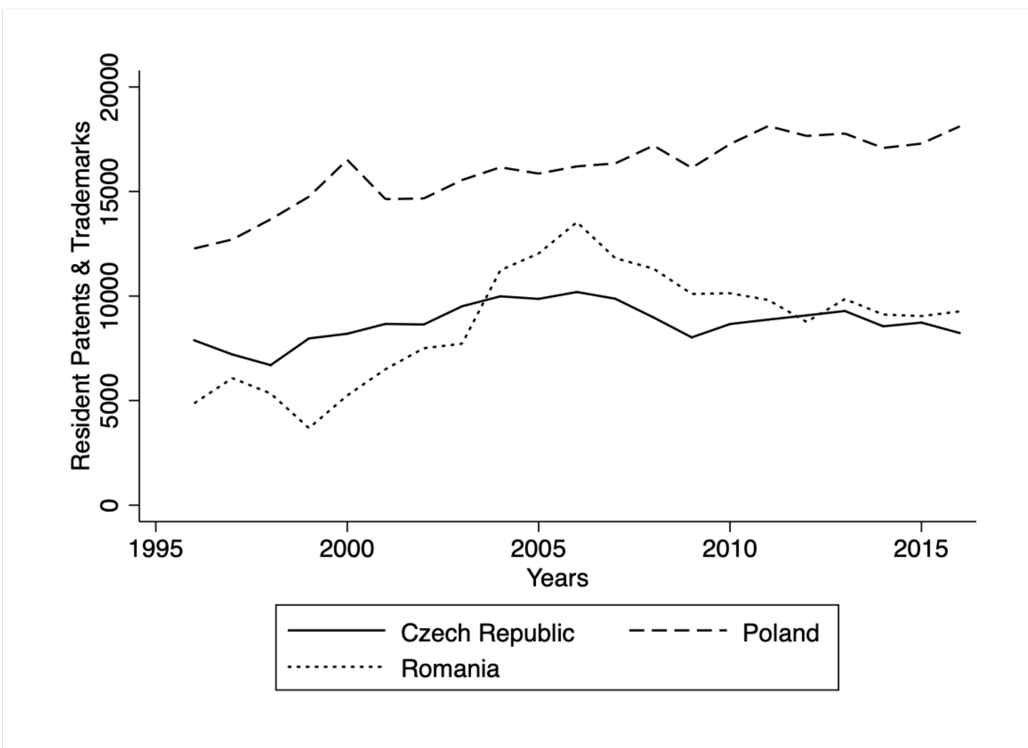

Figure A.6: High Filing Countries

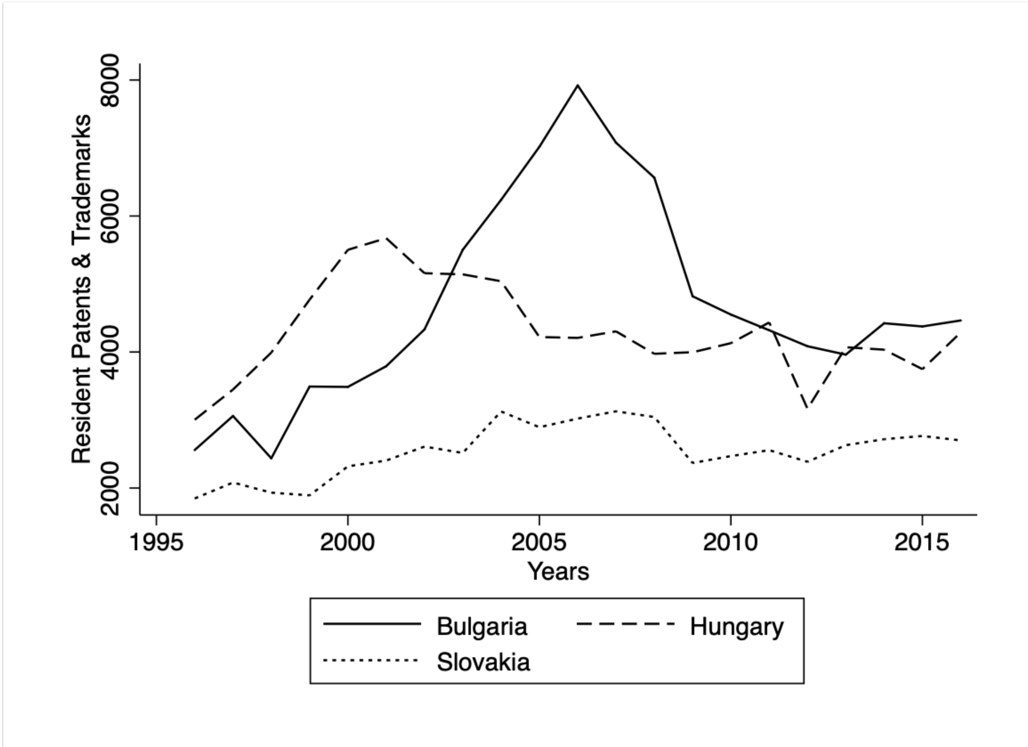

Figure A.7: Medium Filing Countries

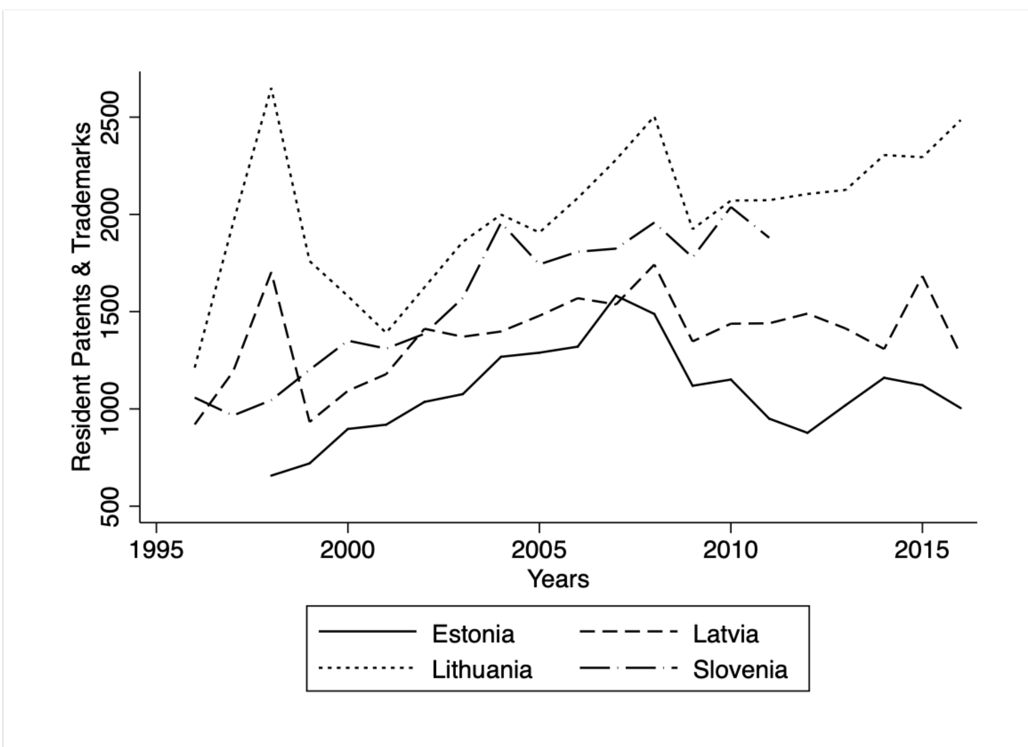

Figure A.8: Low Filing Countries

Figure A.9 presents a scatter plot showing the correlation between *log* Innovation and R&D share. Each country's marker is weighted by their working population in order to see if larger potential work forces have any pattern with the R&D share. There is a clear negative correlation between *log* Innovation and R&D share. This concurs with the findings in Tables 5.1-5.3. It appears that increased shares of researchers results in a lower number of innovation filings. This would describe "stepping on toes" and an inefficient researching sector.

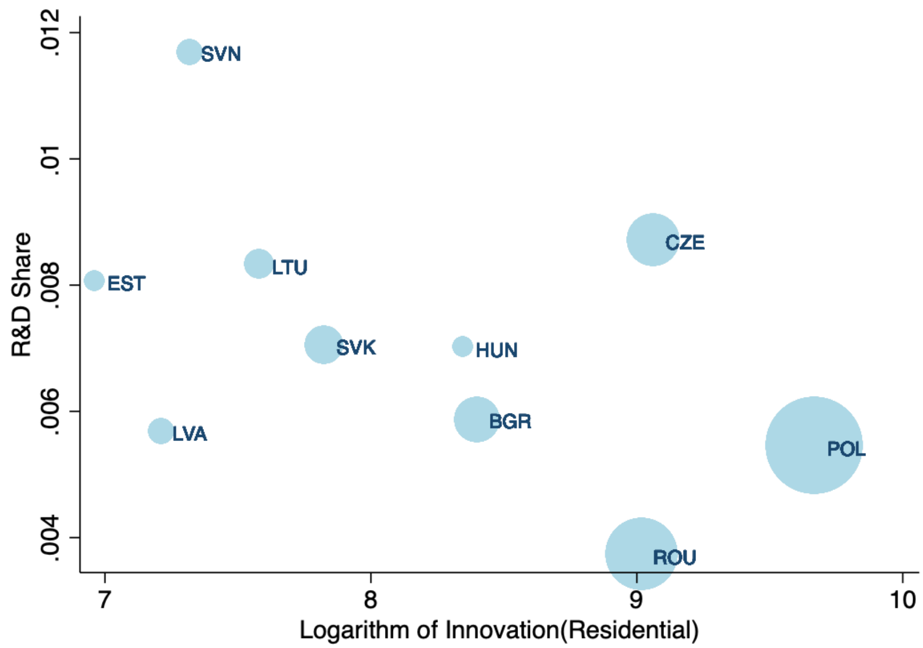

Figure A.9: Correlation of *log* Innovation and R&D Shares - Averages

Since there is limited degrees of freedom with a sample this small, I decided to plot the dependent variable growth and control for any major shocks occurring throughout time. By controlling for these shocks, I add time specific controls but do not include as many variables as time fixed effects. The major variations in growth occur pre-2000 which is logical as this would be a tumultuous time for these Post-Soviet countries experiencing transition. Furthermore, before 2000 there was the concern of Y2K which could have had an impact on the growth rate. The second time specific control is for the 2008 financial crisis. This crisis is evident in Figure A.10 as there is a massive shock that begins around 2008 and ends around 2010. There does not appear to be any linear time trend for the growth rate, however for confirmation, regressions in Table A.2 include a linear time trend.

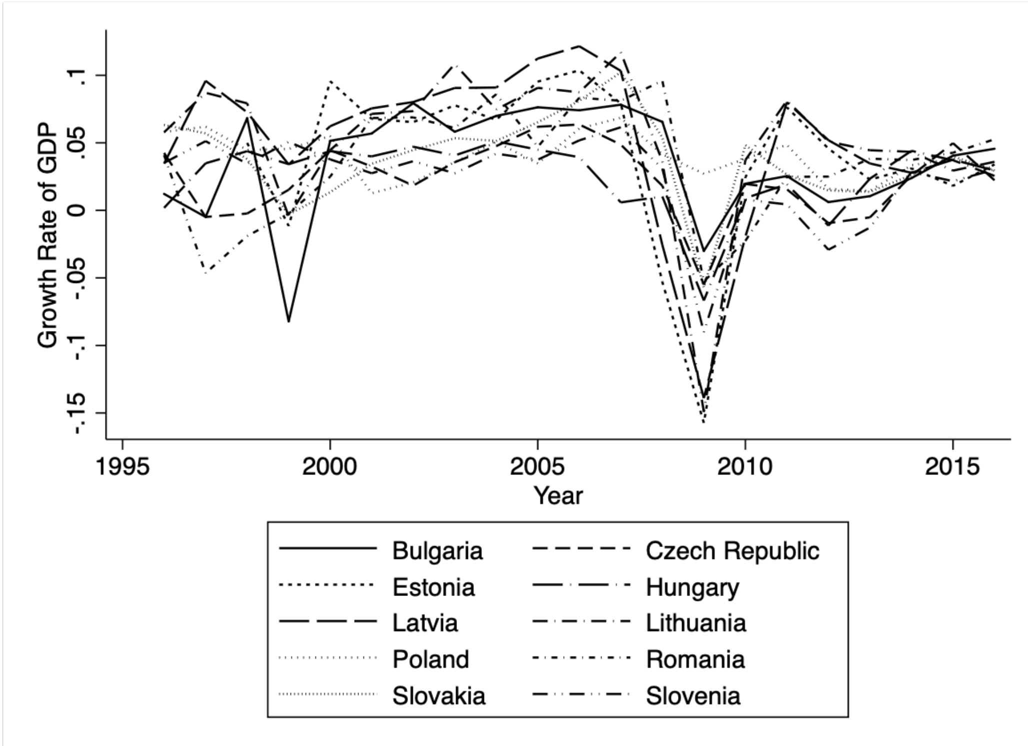

Figure A.10: Growth Rate By Country

Table A.2 shows that including a linear trend into the regression does not change the results in any significant way. The trend itself confirms what Figure A. 10 displays, that the trend is not statistically significant from 0. Therefore, in the results section, trend is not included and only time specified effects are significant.

Table A.3 shows the regression results when alternative R&D measures are used. The R&D share and R&D expenditure variables were alternatives to *log* Innovation. Matching the regression results, both *log* Innovation and *log* R&D share are statistically significant. The *log* R&D expenditure variable does not appear to hold any significance when it comes to analysing the growth rate.

Table A.2: Trend Variable Inclusion

Dependent: Growth	(1)	(2)	(3)	(4)	(5)
Initial GDP	-0.108*	-0.200**	-0.274**	-0.285**	-0.294***
	(0.0575)	(0.0824)	(0.0907)	(0.0911)	(0.0869)
log Innovation	0.0738***	0.0193*	0.0342***	0.0289***	0.0360***
	(0.0195)	(0.00879)	(0.00795)	(0.00770)	(0.00729)
Financial Crisis	-0.0503***	-0.0412***	-0.0355***	-0.0340***	-0.0354***
	(0.0106)	(0.00983)	(0.00931)	(0.00948)	(0.00979)
Pre-2000	-0.0235**	-0.0178*	-0.0177*	-0.0147	-0.0154
	(0.00922)	(0.00886)	(0.00834)	(0.00866)	(0.00893)
Trend	0.00130	0.00127	0.00712	0.00764	0.00712
	(0.00228)	(0.00267)	(0.00400)	(0.00418)	(0.00408)
Investment		0.503***	0.474***	0.471***	0.458***
		(0.143)	(0.132)	(0.128)	(0.135)
Openness		0.0669**	0.0402*	0.0474*	0.0415*
		(0.0229)	(0.0202)	(0.0210)	(0.0219)
*log*Population		-0.154	-0.259**	-0.229**	-0.216**
		(0.0905)	(0.106)	(0.0918)	(0.0889)
European Union			0.0154*	0.0164	0.0202*
			(0.00836)	(0.00907)	(0.01000)
Gross Domestic Savings			0.250**	0.282***	0.331***
			(0.0773)	(0.0813)	(0.0887)
log R&D Share				-0.0367**	-0.0447**
				(0.0147)	(0.0153)
Constant	0.437	3.929*	7.046**	6.467**	6.086**
	(0.429)	(2.070)	(2.595)	(2.271)	(2.186)
N	203	203	203	203	187
Within R^2	0.417	0.592	0.649	0.662	0.669

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Driscoll-Kraay standard errors are in parentheses. MA(2) process determined by the regression model. The dependent variable growth is the log difference of GDP per capita. All regression include country fixed effects and time specific dummy controls (financial crisis and pre-2000's). Controls not reported are life expectancy, inflation, and tertiary enrolment.

Table A.3: Alternative Innovation Measures

Dependent: Growth	(1)	(2)	(3)	(4)	(5)
Initial GDP	-0.197***	-0.183**	-0.193***	-0.194**	-0.202***
	(0.0595)	(0.0592)	(0.0579)	(0.0605)	(0.0593)
log Innovation	0.0322***			0.0290**	0.0275**
	(0.00948)			(0.00930)	(0.00897)
log R&D Expenditure		-0.0174		-0.0115	
		(0.0121)		(0.0119)	
log R&D Share			-0.0337*		-0.0311**
			(0.0151)		(0.0137)
Investment	0.437***	0.510***	0.487***	0.446***	0.431***
	(0.0998)	(0.0900)	(0.0880)	(0.0935)	(0.0940)
Openness	0.0752**	0.0800**	0.0857**	0.0770**	0.0834**
	(0.0267)	(0.0252)	(0.0267)	(0.0258)	(0.0273)
log Population	-0.245*	-0.198	-0.171	-0.242*	-0.219*
	(0.112)	(0.109)	(0.0934)	(0.114)	(0.0986)
European Union	0.0138	0.0109	0.0131	0.0131	0.0146
	(0.00911)	(0.00748)	(0.00888)	(0.00881)	(0.00990)
Financial Crisis	-0.0407***	-0.0415***	-0.0401***	-0.0411***	-0.0397***
	(0.00906)	(0.0101)	(0.00995)	(0.00936)	(0.00924)
Pre-2000	-0.0211**	-0.0246**	-0.0241**	-0.0198**	-0.0188*
	(0.00818)	(0.00888)	(0.00903)	(0.00805)	(0.00844)
Gross Domestic Savings	0.235**	0.206**	0.227**	0.239**	0.261**
	(0.0808)	(0.0676)	(0.0733)	(0.0773)	(0.0824)
Constant	5.578**	4.775**	4.275**	5.454**	4.995**
	(2.157)	(2.102)	(1.765)	(2.178)	(1.856)
N	203	208	208	203	203
Within R^2	0.624	0.616	0.622	0.626	0.633

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Driscoll-Kraay standard errors are in parentheses. MA(2) process determined by the regression model. The dependent variable growth is the log difference of GDP per capita. All regression include country fixed effects and time specific dummy controls (financial crisis and pre-2000's). Controls not reported are life expectancy, inflation, and tertiary enrolment.

A.2 Industry Level

This appendix presents additional information for the industry level data. Figures A.11 and A.12 present kernel density plots of the dependent variable and competition measure. Table A.4 provides the number of observations per country at the firm level and lastly, the variable descriptions for the control terms are presented.

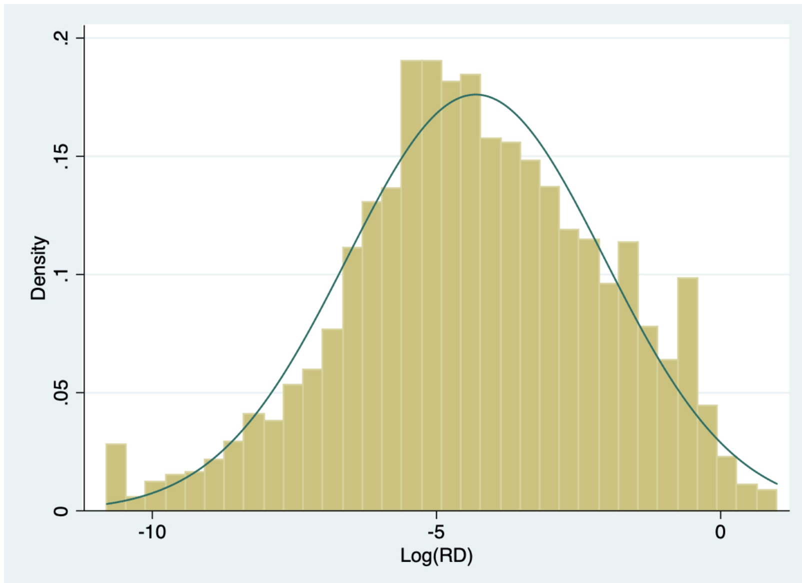

Figure A.11: Kernel Density of lnR&D

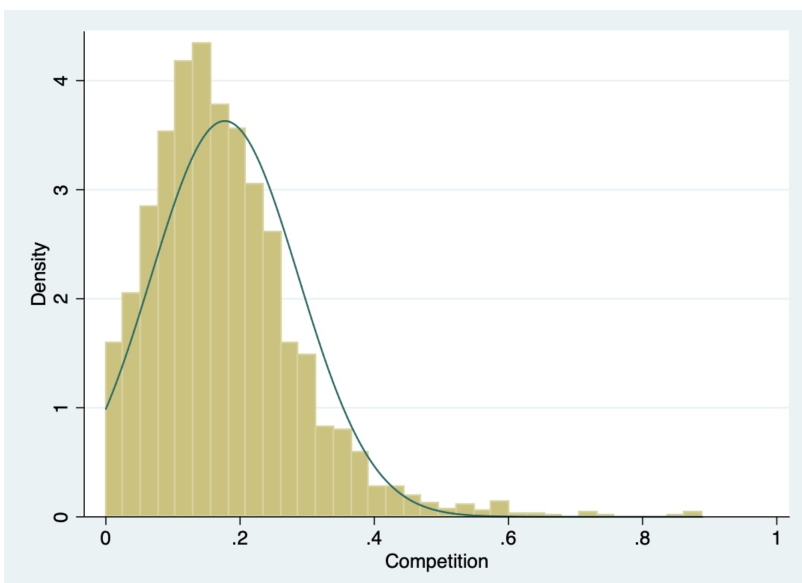

Figure A.12: Kernel Density of Competition

Table A.4: Firm Observations

Country	Total	Post-EU
Bulgaria	458	405
Czech Republic	267	146
Estonia	165	127
Hungary	243	185
Lithuania	349	340
Latvia	205	192
Poland	4636	4409
Romania	692	620
Slovakia	81	63
Slovenia	280	233
N	7376	6720

Table of firm observations for each country

Independent Variables Description:

- SIZE_{jct} is the average size of the firm in the industry. This is calculated by finding the average firm's total assets in each industry and then taking the logarithm (Fosu, 2013). This would represent the average firm's size so industries with larger firms should have more resources available to devote to R&D.
- N_{jct} reflects the number of firms in each industry for each country in the dataset. This variable is used as a control and to also avoid endogeneity as the number of firms in the industry could affect the level of R&D as well as competition.
- CI_{jct} is the measure of capital intensity for each industry. This is measured by dividing the total assets by the dollar value of sales for each firm and then averaging for the industry in each country and year.
- $ln\text{RDI}_{jct}$ is the natural logarithm of R&D intensity. This is an alternative dependent variable used to check robustness of how competition and spread affect an alternative measure of R&D. RDI is calculated by intangible assets divided by value of sales.
- ROA_{jct} reflects the return on assets of a firm. This variable controls for asset performance for the average firm in each industry-country-year. This control will help to determine if the return on total assets incentivised further investment. ROA is calculated by dividing net income of the firm by total assets (Fosu, 2013).
- GPM_{jct} is Gross Profit Margin, measured by gross profit divided by revenue.
- DTIC_{jct} is the balance sheet reported value of deferred taxes and investment credits. This control was originally in millions but converted to per 100,000.