THE BONN JOURNAL OF ECONOMICS

Volume VI(2) December 2017

DEPARTMENT OF ECONOMICS

UNIVERSITY OF BONN

Legal Information

Publisher	The Bonn Journal of Economics Adenauerallee 24-42 53113 Bonn http://bje.uni-bonn.de bje@uni-bonn.de
V. i. S. d. P.	JProf. Tobias Berg University of Bonn
Place of Publication	Bonn, Germany
ISSN	2195-3449

THE BONN JOURNAL OF ECONOMICS

FOUNDED IN 2012

VOLUME 6
ISSUE 2

DEPARTMENT OF ECONOMICS
UNIVERSITY OF BONN

THE BONN
JOURNAL OF ECONOMICS

Contents

The Impact of Demographic Change on Inflation Dynamics: An Empirical Analysis

Christopher Lancier*

1 Abstract

In recent years inflation has been damping around the zero line and the issue of demographic change towards ever increasing life expectancy and diminishing fertility rates is receiving an increasing degree of public and scientific attention. This work tries to contribute to the rather scarce existing research on the connection between demography and inflation by characterizing and comparing the impact different age cohorts exert on inflation dynamics, each for a sample of advanced and developing countries. The analysis is conducted via a fixed effects regression model with a focus on robustness across different time periods and control variables. In fact, the results indicate substantial differences in virtue between advanced and developing countries regarding the impact of demographic change on inflation dynamics.

2 Introduction

The world today is heading towards unknown territory regarding the consequences of yet unseen demographic developments. These developments manifest in a decrease in mortality and fertility coupled with increasing life expectancy, which

*Christopher received his B. Sc. from the University of Bonn in March 2017. Afterwards he began his Master studies at the University of Freiburg.

alter the age structure of countries. And even though a certain consensus can be attributed over the nature of these developments, their direct influences on economic activity have in contrast just been partially explored by research.

In context of the rapid pace, in which the population in Japan is aging ever since the 1970's [1], it was the first country to draw attention on inflation based consequences of an aging and in the Japanese case also rapidly declining population. One of the first theoretical assessments was conducted by Derek Anderson (2014) who used the IMF's Global Integrated Fiscal and Monetary Model to conclude, that aging in Japan exerts remarkable deflationary pressure mainly through a decline in growth and land prices. Even though, as of today the demographic transmission from the young to the old age cohort is most advanced in Japan, the Euro Area is also experiencing a profound era of low inflation. One of the reasons for this was described by Imam (2014) who discovered that the shift to an older society also weakens the effectiveness of monetary policy. One of the first empirical studies directly targeting the effect of demographic changes on inflation was conducted by Thomas Lindh (2000) who used the age structure of 20 OECD countries to forecast inflation trends and concluded, that especially net saving age cohorts like the middle aged, dampen inflation whereas retirees in the early stage of retirement increase inflation. Another empirical investigation of OECD countries conducted by Jong-Won Yoon (2014) found that aging and at the same time shrinking societies, a fate most likely to capture countries like Japan and Germany in the future, are affected by substantial deflationary pressure. Building up on this empirical approach Mikael Juselius (2015) conducted a similar but more extensive empirical analysis with a focus on the robustness of their findings. In contrast to Jong-Won Yoon (2014) they found, that while larger young and old age cohorts are correlated with higher inflation, only an increase in the working age cohort is associated with deflationary pressure.

The thin empirical evidence on the impact of demographic changes on inflation dynamics is surprising, as the potential influences could severely exacerbate the inflation targeting of central banks. This is even truer in a time of interest rates at the lower limits, through which the arsenal of monetary authorities to react upon unanticipated shocks to inflation is somewhat limited. Furthermore the existing empirical evidence, as small in quantity as it is, is focused solely on

[1] see United Nations (2015)

advanced economies with no recognition of developing countries. One reason for that might be the imminence of demographic changes particularly regarding advanced economies besides the differences in available data sources. Being aware of this hurdle, this work will nevertheless attempt to investigate not only advanced economies, but will also conduct empirical analysis on a sample of emerging countries to investigate potential differences in virtue of demographic changes on inflation dynamics.

In this sense, this work will continue with a brief description of the demographic developments in the past in section 3, followed by a brief assessment of the theoretical channels by which demography can alter inflation dynamics in section 4. Then the model used in the regression is introduced in section 5. The actual regression results will be presented in section 6 before finishing with a concluding remark in section 7.

3 Demographic Developments

After World War II high-income countries[2] like central Europe experienced high fertility from 1950 to 1965 which started to decline quite rapidly afterwards, dropping from 3 children per woman to just over 2 in 1980 as can be seen in Figure 1. With time, the combination of this temporarily high fertility with a rapid decline to follow, lead to a significant drop in dependency ratios[3], just when the high fertility generations entered the labor force at the beginning of 1980. This phenomenon is known to be a demographic dividend. This dividend led to a growth friendly environment for the economy as the work force was relatively large in comparison to the non working force.

Nevertheless, as fertility remained low and even declined further to just over 1.8 in 2015, the downside of the past baby boom is becoming increasingly imminent as the once large working age cohort is now retiring. This led to an increase in the dependency ratios ever since 2007. While this is the case for high-income

[2]It has to be noted, that the classification in high and low income regarding the status of development of countries for this section refers to the definitions by the United Nations which capture a larger sample than the sample used in this work

[3]The dependency ratio is defined as the ratio of young and old individuals not participating in the workforce to middle aged individuals in the work force

economies, quite the opposite is true for low-income economies. There fertility remained constant up to 1985. Even though it also started to decline from 1990 on, it is on average with 5 children per woman almost 2.5 times higher than in high-income economies. Fertility might therefore be the factor that makes these demographic developments to be of more concern for Europe than for other high-income regions like North America. This is due to the fact that Europe is also projected to suffer from severe decline in population[4]. While in almost every region the growth rate of population is projected to decline, Europe might be the first region to permanently experience negative growth rates. All of these developments led to the situation, where the age distribution in the population had diverged significantly between the low and high income economies. In Europe and North America a labor force of about 50% of the population is accompanied by about equal shares for old and young with accelerating momentum towards the old age cohort[5]. In Africa in opposition, a rather small workforce of one third of the population is accompanied by a twice as large young age cohort. As contrary as those developments might be, both pose similar even though not exactly equal challenges for high and low income countries regarding the future ahead.

4 Theoretical Implications

This sections briefly discusses some of the possible channels of effect between demography and inflation.

4.1 Land, Consumption and Labor Channel

A shift of a society's age structure towards the older age cohorts is likely to alter the consumption preferences. Older people might prefer smaller houses, which would exert downward pressure on house and land prices[6] and could have second round effects exerting further deflationary pressure. Another possible channel affects the price of labor, as labor force participation likely shrinks with growing old age cohorts resulting in upward pressure on wages. It is also feasible to

[4] See Figure 2 in the Appendix
[5] See Figure 3 in the Appendix
[6] In fact a drop in house prices was found by Takats (2012)

believe that consumption patterns of an older society will relocate towards higher expenditures for health care and related sections, while probably demanding less supply of transportation, entertainment and given low fertility levels, also education. This trade off makes the effect of altering consumption patterns on overall inflation unclear.

4.2 Saving Channel

Another channel through which demographic developments might affect inflation dynamics is by different preferences in saving behavior. With increasing life expectancy especially working age people could tend to save a substantially larger part of their income.[7] As with increasing age also the probability to reacquire potential losses from assets held diminishes, it is likely to assume an increase in risk aversion for older age cohorts. This could lead to a shift from risky to safer assets like treasury bonds. Both effects would put downward pressure on interest rates and could therefor increase overall inflation.

4.3 Fiscal Channel

Yet another channel through which aging of a society might affect inflation dynamics is on the fiscal level. As for a given level of government debt, the combination of a reduced income tax base with increasing expenditures for health care and pensions would alter the government account balance negatively. That in turn might lead to lower inflation expectations, as high present and projected account deficits could put unfavorable pressure on refinancing conditions of the government and would make fiscal consolidation necessary in the future. Yet this leaves the possibility that overall government spending might just temporarily increase as also spending related to younger age cohorts, especially education but also potentially military expenditures, become less important.

[7] Indications for such behavior was found by Ray C. Fair (1991)

4.4 Policy Channel

One other considerable channel through which aging of a society can impact inflation dynamics is through the shift in objective and weakening of the effectiveness of monetary policy. According to Bullard (2012) older age cohorts would prefer high interest rates coupled with low inflation, in contrast to younger cohorts who prefer lower interest rates and higher inflation as they tend to be debtors instead of creditors like the older cohorts. This might alter its objective on how to secure sustainable economic activities in a country. These wealth considerations might also make old age cohorts less sensitive on changes in the interest rate which in turn would dampen interest rate policy effectiveness in aging societies and might substantially increase the effort necessary to achieve a given inflation target[8].

5 The Fixed Effects Model

This work seeks to investigate the link between demographic developments and inflation dynamics for a sample of 29 advanced economies and 38 countries[9] considered as developing economies[10] over a time horizon of 55 years from 1960 to 2015. This investigation is conducted through a panel regression of the form:

$$\pi_{i,t} = \alpha_i + \delta_t + \beta Demo_{i,t} + \gamma Z_{i,t} + \varepsilon_{i,t} \tag{1}$$

Where π is the inflation rate, *Demo* is a set of demographic variables, Z is a set of control variables that will gradually be implemented and ε corresponds to the error term of the regression. The subscripts i and t refer to the different countries and years respectively. Additionally the variables α and δ are indicator variables of the form:

$$\alpha_i, \delta_t = \begin{cases} 1 \text{ if Country i, Time t} \\ 0 \text{ otherwise.} \end{cases}$$

These dummy variables incorporate the so called fixed effects and lead to the reference of this kind of models as Least Squares Dummy Variable (LSDV) models.

[8] as found by Imam (2014)

[9] For the list of countries see Tables 6 & 7 in the Appendix

[10] The classifications are closely related to that of the IMF, see IMF (2016)

Implementation of the fixed effects on the country level allows controlling for time invariant unobservable heterogeneity across countries such as geographical, cultural or institutional factors. Similar logic applies to time fixed effects, which capture time varying unobserved factors that influence every country equally.

To capture demographic characteristics this work uses the relative size of the different age cohorts of a society[11].

As it is extremely difficult to disentangle different influences on a macroeconomic level in order to interpret them causally, this work will not try to give an on-point estimation of the causal effect changes in the age structure exert on inflation. Instead the focus will be on whether there exists a stable relationship between age structure and inflation and in what direction this relationship possibly goes.

6 Empirical Findings

This Section will provide presentation of the results obtained by regression to analyze the potential link between demography and inflation. First the results for the sample of advanced economies are presented in the basic model which will then gradually be tested for robustness through inclusion of different time periods and controls. Similar pattern will then be applied to a sample of developing countries.

6.1 Advanced Economies

Basic Specification

To get a first sense for the connection between demography and inflation, the first specification includes only the dependency ratio. As can be seen in Table 1, the dependency ratio appears to be highly correlated with inflation (Model 1). Though, when controlling for time fixed effects (Model 2), the coefficient on the dependency ratio drops largely, rendering it statistically insignificant. Therefore

[11] The different age cohorts are defined as following: Age 0-24 for the young age cohort, age 25-65 for the working age cohort and 65+ are for the old age cohort.

this first specification leaves us ambiguous whether there is a stable relationship between demography and inflation. As the dependency ratio implicitly assumes identical effects for the young and the old age cohorts, the analysis will proceed to allow for more flexible effects by including all three age cohorts separately (Model 3 & 4). This assumption seems to be supported, as both coefficients on the young and the old age cohorts are indistinguishably different from zero while only the working age cohort seems to negatively affect inflation. Interestingly, adding time fixed effects and therefore accounting for general shocks to inflation but also partially trend movements, renders all age cohorts to negatively affect inflation in similar size while also being highly significant on the 1% level. This indicates, that the old and the young age cohorts are positively correlated with factors affecting the long run variation in inflation, which is partially removed by the time dummies. To account for omitted variables biasing the results, the next specifications also incorporates the output gap[12] to control for business cycle effects and the real interest rate[13], which are theoretically both closely related to inflation (Model 5 & 6).

While both coefficients for the output gap and the real interest rate show expected signs according to theory, they also somewhat alter the effect of the young and the working age cohorts. The former is now exerting significant inflationary pressure while the latter turns insignificant. Including time fixed effects has the same impact as before. There is one drawback in this specification though, as due to the lack of interest rate data the number of used observations more than halves. Therefor it is left out to avoid sample bias[14]. The specification in Model 7, controlling for the output gap and not including time fixed effects yields the benchmark model to which several robustness checks will be addressed.

This first brief investigation of the link between demography and inflation revealed constant negative pressure exerted by working age cohorts, while coefficients on the young and the old seem to be very unstable, especially upon inclusion of time fixed effects.

[12] Defined as the deviation from a Hodrick Prescott filtered Trend of real GDP

[13] Although it has to be noted that, for the real interest rate being the ex post real interest rate leaves some endogeneity concerns, as every movement in the inflation rate which is not accompanied by the nominal interest rate will systematically overestimate its coefficient.

[14] Missing observations are closely related to time, with more missing values further back in time. For further discussion about possible limitations see the online appendix

Different Time Periods

To further investigate the relationship of demographic developments on inflation dynamics, the benchmark specification developed in the previous section will be tested across different time periods. This is to see whether the relationship changed in context of an increasing adaption of inflation targeting of monetary policy in the 1990's. To account for that we will investigate the behavior of the benchmark specification in three periods: From 1960 to 1990, from 1985 to 2015 and from 1985 to 2006. The last period is used to evaluate a possible influence of the financial crisis in 2007 which lead to a sharp decline in inflation until today.

Interestingly there is quite some difference between the time periods. Especially striking is the effect of the old age cohorts in the pre 1990 period (Table 2, Model 9 & 10). Where it previously exerted mainly small and insignificant deflationary pressure, it is now connected to large inflationary pressure while other cohorts remained basically unchanged. This significance diminishes though upon inclusion of time fixed effects leaving only the working age cohort to exert substantial and significant deflationary pressure. What is also interesting to see, is the fact that in both periods after 1985 all age cohorts exert significant deflationary pressure (Model 11 to 14). This deflationary pressure furthermore seems to arise most pronounced for the young age cohort, even though this influence seems to weaken after 2006, as the coefficient in the period up to 2015 is significantly less negative than in the sample excluding the period following 2006. This is somewhat surprising.

Overall the findings show, that the influence of demographics did alter over time. This might indicate that the impact on inflation is connected to the relative size of the age cohorts, especially for the old and the young as their share relatively to it size vary a lot more over time than the working age cohort. Only the deflationary pressure of the working cohort remained stable.

Additional Controls

As the influence of omitted variables on our findings is like a phantom menace for the interpretation of our result, this subsection will try to implement several further control variables that might interfere with our attained results. In a first step two demographic controls, life expectancy and population growth are implemented in Table 3 to account for additional demographic characteristics (Model 15 & 16). Both turn out to be insignificant and also do not alter any coefficients on the age cohorts.

The next step includes two variables controlling for labor characteristics in the form of labor productivity per hour and the overall hours worked (Model 17 & 18). While both variables are significant and together also show theory conform signs, they do not alter the coefficients on the age cohorts substantially. In the last two specifications further controls for monetary and fiscal characteristics are added in the form of growth of the M2 aggregate, the real interest rate and the fiscal account balance (Model 19 & 20). The latter two though only for a period from 1985 onward due to lack of data in the earlier periods. While all of controls except for the account balance are significant and with the expected sign, they also do not substantially alter the coefficients on the age cohorts. The not existing sensitivity of the age cohorts upon controlling for the real interest rate could indicate that central banks did not consider demographic influences on inflation.

This leave the benchmark specification we proposed largely unchanged throughout the analysis.

6.2 Developing Countries

Basic Specification

When comparing the inflation development between developing and advanced countries, with the former experiencing much higher inflation over the whole period, one would expect to see at least larger coefficients in the respective direction for the developing economies.

While Table 4 shows that this is not true for the first specification (Model 1 & 2), it is in fact true when specifying the different age cohorts separately (Model 3 & 4). What is striking to see, is the increased inflationary pressure of the young age cohort . The same is true for deflationary pressure of the old age cohort.

In addition to that, including time fixed effects does not crowd out the effects which remain highly significant. The inclusion of the output gap also does not significantly change the estimates, in fact it tends to increase the coefficients even further in the respective directions (Model 5 & 6).

Establishing the benchmark model (Model 5) for the sample of developing countries reveals substantially different effects compared to the sample of advanced countries. Especially older age cohorts seem to exert substantially higher deflationary pressure in developing countries. One possible explanation might stem from increased old age poverty in those countries as state run pension systems tend to be less relevant in developing countries.

Different Time Periods

At first glance, the alteration induced by different time periods looks similar to that of the sample of advanced countries. It is once again striking to see the switch of signs from negative to positive for the old age cohort in the pre 1990 period. It is though also not robust to time fixed effects. (Table 5, Model 7 & 8).

Furthermore while the behavior of the coefficients, comparing the two post 1985 samples, is similar to previous findings, the change of the coefficient on the old age cohort though is enormous (Model 9–12). In fact, the magnitude of the negative coefficient in the period up to 2006 on the old age cohort leaves large concerns of some kind of omitted variable bias. One other difference between the samples is constituted in the working age cohort that, in contrast to the advanced sample, constantly switches signs and remains insignificant unless controling for time fixed effects.

Additional Controls

The robustness regarding different controls for the sample of developing countries posed very difficult as many of the previously used control variables were largely missing in this sample. Due to that fact, it is only possible to include other demographic variables like life expectancy and population growth and the real interest rate and account balance, the latter though again only for a sub period from 1985 on.

Controlling for life expectancy does only slightly alter the coefficients for all age cohorts negatively, though rendering the coefficient on the working age cohort to be insignificant (Model 13). The same effect is discovered by controlling for population growth, although even less pronounced (Model 14). Also controlling for real interest rates and fiscal account balance does not alter the coefficients in a substantial way in comparison to the regression on the same period in the previous section (Model 15).

All in all it seems feasible to believe, that the effects demographic developments exert on inflation dynamics are substantially different for developing and advanced economies. Especially profound is the difference for the working and the old age cohorts. Where the effect of the old age cohorts is very unclear and insignificant for the advanced economies, the old age cohort exerts substantial deflationary pressure in the developing countries. The contrary is true for the working age cohort, being constantly deflationary in the advanced sample but highly unstable in the developing sample.

7 Conclusion

This work attempted to analyze the impact of demographic changes in terms of altering age structures on inflation dynamics through analysis of a panel data set for a sample of advanced and developing economies. The results indicate substantial differences between these two samples. For advanced economies, only the impact of the working age cohort is stable throughout robustness checks including different time periods and additional control variables, indicating deflationary pressure associated with an increase in this age cohort. In sharp contrast to this, obtained results from the sample of developing countries indicate substantial impact of the young and old age cohorts. There, the younger age cohorts are associated with inflationary pressure while older age cohorts exert profound deflationary pressure.

The findings for the advanced economies are in accordance with theoretical considerations that working age cohorts tend to be deflationary through their increased propensity to save their income. The same is true for young age cohorts in the developing economies who tend to consume more than they produce and thus act inflationary. In addition to that, the deflationary impact of older age cohorts in developing countries could, to some extent, stem from less pronounced pension systems preventing old age cohorts from increased consumption spending in retirement. Furthermore the analysis regarding different time periods indicated, that the influences of the different age cohorts are far from constant over time and likely to be sensitive to various economic, social and cultural developments. This renders projections of the change in inflation dynamics due to upcoming demographic developments a difficult task.

Even though the overall impact is differing between advanced and developing countries and therefore unclear, demographics will very likely pose a crucial factor for future inflation dynamics. This sets needs for further research, especially on the channels through which demographic characteristics affect inflation dynamics. In addition to that, static statistical models like the one utilized in this work can hardly fully characterize the highly dynamic relationship between demographics and inflation. Therefore further research should implement statistically more suitable dynamic methods to more precisely investigate the relationship between demographics and inflation. Especially the impact of the older age cohorts demands more investigation arising out of the ongoing shift towards these older age cohorts, to ensure central banks can appropriately react to such developments.

Online Appendix to: The Impact of Demographic Change on Inflation Dynamics: An Empirical Analysis by Christopher Lancier

Data Sources

Demographic Variables

UN population prospect 2015

Economic Data including Inflation

WDI Database of the World Bank
http://data.worldbank.org/data-catalog/world-development-indicators

Some additional control variables

Conference Board's Total Economic Database
https://www.conference-board.org/data/economydatabase/index.cfm?id=27762

Tables

Table 1.1: Basic Regression for Advanced Countries

	Dependent variable:							
	Inflation							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep.Ratio	0.442***	0.058						
	(0.067)	(0.066)						
Young			0.431	−0.739**	0.976***	−0.015	0.823***	−0.343
			(0.348)	(0.332)	(0.249)	(0.184)	(0.274)	(0.239)
Working			−0.546***	−0.696***	−0.230	−0.324***	−0.479***	−0.412***
			(0.146)	(0.139)	(0.158)	(0.125)	(0.135)	(0.097)
Old			−0.018	−0.639***	0.110	−0.264*	0.229	−0.292***
			(0.317)	(0.193)	(0.237)	(0.135)	(0.290)	(0.106)
Y.gap					0.183*	0.152**	0.217*	0.077
					(0.097)	(0.063)	(0.117)	(0.090)
Real.Interest					−0.422***	−0.353***		
					(0.096)	(0.083)		
Country Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1,422	1,422	1,422	1,422	667	667	1,078	1,078
Adjusted R^2	0.104	0.331	0.161	0.370	0.373	0.650	0.229	0.504
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 1.2: Different Time Periods for Advanced Countries

	Dependent variable:					
	Inflation					
	(9)	(10)	(11)	(12)	(13)	(14)
Young	0.509***	−0.042	−0.476	−0.782	−0.995***	−1.912***
	(0.175)	(0.273)	(0.326)	(0.498)	(0.268)	(0.547)
Working	−1.096***	−0.997***	−0.912***	−0.705***	−1.251***	−0.840***
	(0.189)	(0.236)	(0.182)	(0.147)	(0.148)	(0.140)
Old	3.063***	1.072	−0.770***	−0.452***	−1.053***	−0.561**
	(0.531)	(0.679)	(0.179)	(0.156)	(0.152)	(0.238)
Y.gap	0.130	0.019	0.174***	0.123	0.221***	0.190**
	(0.159)	(0.151)	(0.047)	(0.083)	(0.068)	(0.078)
Country Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Effects	No	Yes	No	Yes	No	Yes
Time Period	1960-1990	1960-1990	1985-2015	1985-2015	1985-2006	1985-2006
Observations	578	578	615	615	435	435
Adjusted R^2	0.241	0.428	0.266	0.342	0.250	0.391
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 1.3: Different Controls for Advanced Countries

	Dependent variable:					
	Inflation					
	(15)	(16)	(17)	(18)	(19)	(20)
Young	0.729***	0.751***	0.776***	0.592**	0.769***	−0.601**
	(0.256)	(0.289)	(0.245)	(0.273)	(0.252)	(0.257)
Working	−0.591***	−0.485***	−0.479***	−0.630***	−0.348***	−0.191**
	(0.147)	(0.141)	(0.120)	(0.159)	(0.110)	(0.081)
Old	0.121	0.140	−0.011	−0.226	0.286	−0.575***
	(0.216)	(0.286)	(0.237)	(0.302)	(0.233)	(0.215)
Life.Expectancy	0.128					
	(0.157)					
Pop.Growth		14.530				
		(56.428)				
HrsWorked				−14.268***		
				(4.346)		
LP.Hrs			0.042***	−0.055**		
			(0.014)	(0.023)		
M2Growth					0.124**	
					(0.057)	
Real.Interest						−0.112**
						(0.057)
Acc.Balance						−0.010
						(0.060)
Y.gap	0.235**	0.263**	0.178*	0.208**	0.182	0.270***
	(0.116)	(0.102)	(0.091)	(0.084)	(0.169)	(0.056)
Country Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Effects	No	No	No	No	No	Yes
Time Period	1960-2015	1960-2015	1960-2015	1960-2015	1960-2015	1985-2015
Observations	1,054	1,060	1,058	1,058	503	408
Adjusted R^2	0.223	0.238	0.291	0.332	0.221	0.455
Note:					*p<0.1; **p<0.05; ***p<0.01	

Table 1.4: Basic Regression for Developing Countries

	Dependent variable:					
	Inflation					
	(1)	(2)	(3)	(4)	(5)	(6)
Dep.Ratio	0.427***	−0.109				
	(0.072)	(0.070)				
Young			2.364***	1.505***	3.119***	1.620***
			(0.483)	(0.283)	(0.561)	(0.607)
Working			0.390**	0.930***	0.428**	0.950***
			(0.161)	(0.170)	(0.208)	(0.248)
Old			−4.351***	−3.493***	−4.629***	−4.607***
			(1.169)	(0.933)	(1.543)	(1.156)
Y.gap					−0.007	−0.165
					(0.153)	(0.169)
Country Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Effects	No	Yes	No	Yes	No	Yes
Observations	1,647	1,647	1,647	1,647	1,189	1,189
Adjusted R^2	0.014	0.166	0.091	0.213	0.128	0.237
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 1.5: Different Time Periods and Controls for Developing Countries

	Dependent variable:								
	Inflation								
	Time Periods						Different Controls		
	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Young	3.955***	−0.047	−0.646	2.227***	−1.371**	1.842*	2.594***	2.832***	1.918***
	(0.418)	(0.756)	(0.935)	(0.687)	(0.640)	(0.940)	(0.401)	(0.782)	(0.478)
Working	−0.214	−0.084	−0.358	2.313***	0.812	3.554***	0.229	0.345	2.893***
	(0.455)	(0.520)	(0.461)	(0.486)	(0.494)	(0.541)	(0.248)	(0.348)	(0.517)
Old	9.887***	4.159	−7.392**	−5.303**	−18.302***	−12.921***	−5.725***	−5.151***	−4.862**
	(2.935)	(3.245)	(3.473)	(2.616)	(2.645)	(2.479)	(1.529)	(1.607)	(1.992)
Life.Expectancy							0.341*		
							(0.188)		
Pop.Growth								−126.090	
								(241.249)	
Real.Interest									−0.153
									(0.102)
Acc.Balance									0.084
									(0.169)
Y.gap	0.098	0.005	−0.106	−0.309	−0.232	−0.458*	0.004	0.025	−0.345
	(0.165)	(0.160)	(0.161)	(0.231)	(0.191)	(0.236)	(0.147)	(0.150)	(0.239)
Country Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Effects	No	Yes	No	Yes	No	Yes	No	No	Yes
Time Period	1960-1990	1960-1990	1985-2015	1985-2015	1985-2006	1985-2006	1960-2015	1960-2015	1985-2015
Observations	604	604	706	706	492	492	1,166	1,171	593
Adjusted R^2	0.137	0.213	0.158	0.232	0.144	0.191	0.130	0.129	0.260
Note:									*p<0.1; **p<0.05; ***p<0.01

Table 1.6: List of Sample Advanced Countries

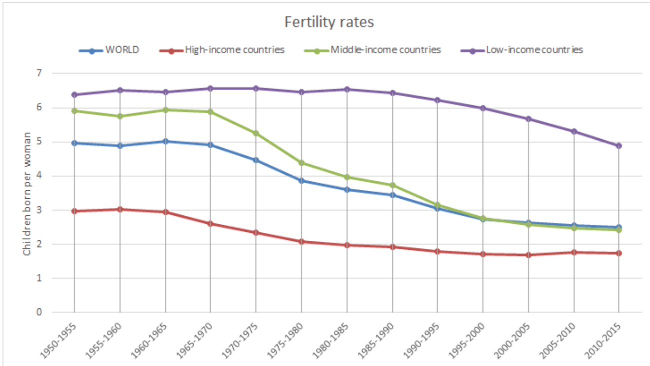
Australia	Austria	Belgium
Canada	Czech Republic	Denmark
Finland	France	Germany
Greece	Hungary	Ireland
Israel	Italy	Japan
Luxembourg	Netherlands	New Zealand
Norway	Poland	Portugal
Slovakia	Slovenia	South Korea
Spain	Sweden	Switzerland
United Kingdom	United States	

Table 1.7: List of Sample Developing Countries

Argentina	Bangladesh	Brazil
Bulgaria	Chile	China
Colombia	Costa Rica	Croatia
Dominican Republic	Ecuador	Egypt
Estonia	Iceland	India
Indonesia	Iran	Kenia
Latvia	Malaysia	Mexico
Morocco	Nigeria	Pakistan
Peru	Philippines	Qatar
Romania	Russia	Senegal
South Africa	Thailand	Tunisia
Turkey	Ukraine	United Arab Emirates
Uruguay	Venezuela	

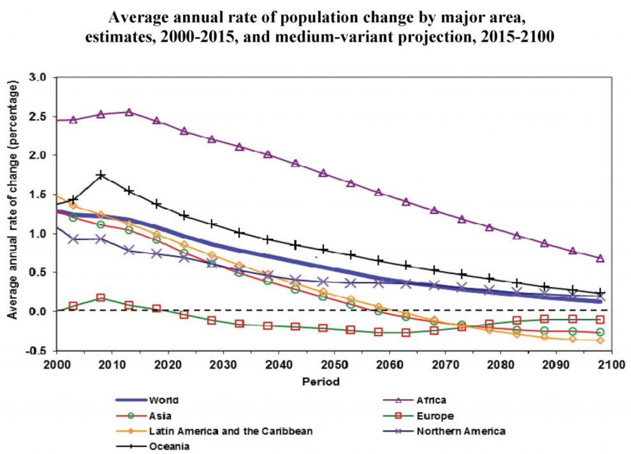
Figures

Figure 1.1: Fertility Rates



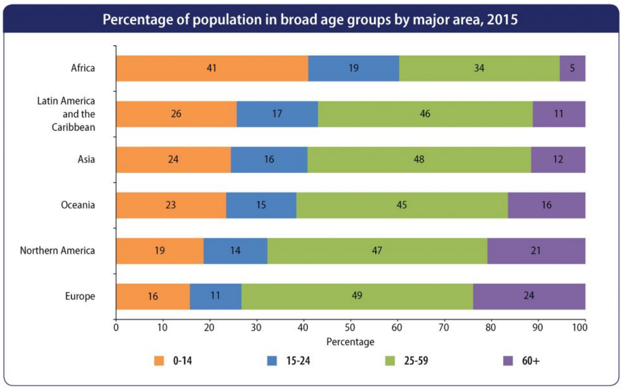
Source: U.N. World Population Prospects: The 2015 Revision

Figure 1.2



Source: U.N. World Population Prospects: The 2015 Revision

Figure 1.3



Source: U.N. World Population Prospects: The 2015 Revision

Summary of Variables for Sample of Advanced Countries

Statistic	N	Mean	St. Dev.	Min	Max
Inflation	1,422	5.544	6.642	−4.480	78.310
Acc.Balance	982	−0.218	4.675	−14.652	16.187
Life.Expectancy	1,591	74.738	4.243	53.001	83.588
Real.Interest	890	4.225	3.732	−12.226	16.612
HrsWorked	1,463	1.855	0.263	1.362	3.042
LP.Hrs	1,463	36.281	17.548	1	93
Young	1,624	14.819	2.267	9.377	22.217
Working	1,624	50.650	4.241	36.194	59.388
Old	1,624	10.279	2.718	2.758	21.367
Dep.Ratio	1,624	49.897	5.425	38.168	65.651
Y.gap	1,118	−0.020	2.447	−12.590	9.461

Summary of Variables for Sample of Developing Countries

Statistic	N	Mean	St. Dev.	Min	Max
Inflation	1,647	12.948	16.398	−7.634	99.877
Acc.Balance	1,338	−1.852	5.624	−25.549	33.185
Life.Expectancy	2,090	65.111	8.936	37.183	82.917
Real.Interest	1,015	6.484	14.433	−91.724	93.937
Young	2,128	18.068	2.477	9.708	26.273
Working	2,128	42.347	8.257	27.768	72.223
Old	2,128	4.965	3.142	0.558	16.156
Dep.Ratio	2,128	56.187	10.270	18.996	78.849
Y.gap	1,326	−0.167	3.836	−26.968	9.976

Figure 1.4: Advanced Countries

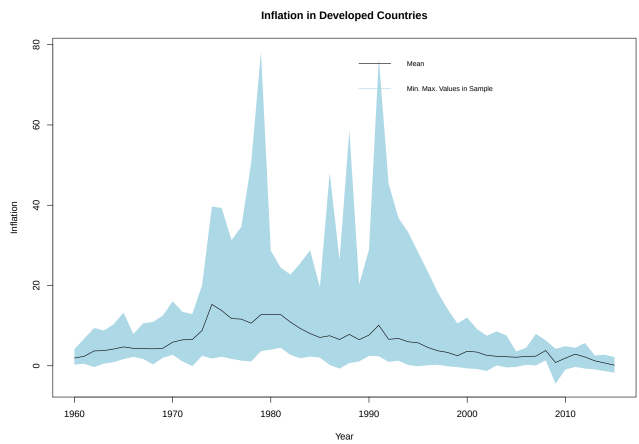
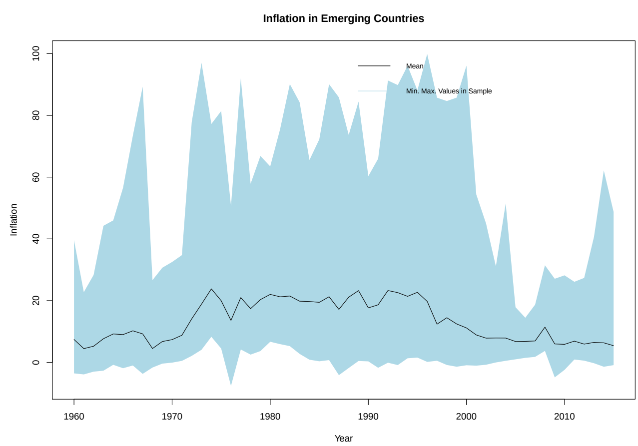
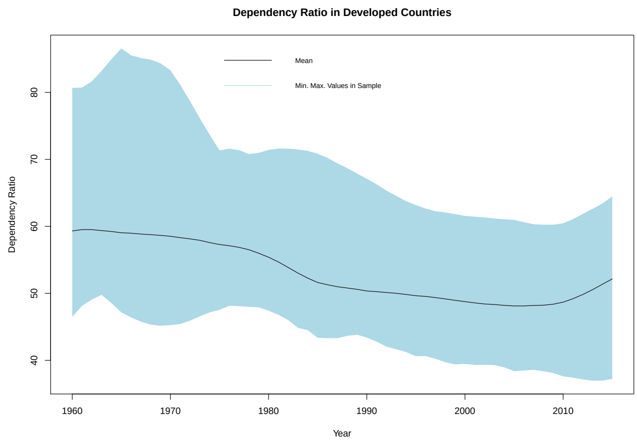


Figure 1.5: [Developing Countries

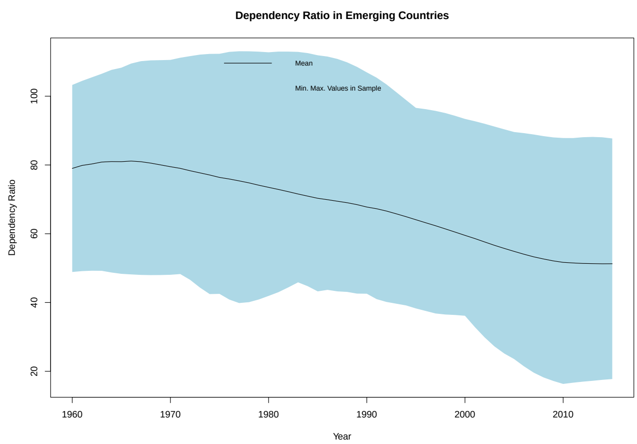


Inflation Description of Sample

Dependency Ratio Advanced Sample



Dependency Ratio Developing Sample



Possible Limitations

Investigating economical relationships through statistical analysis is in theory the way to proceed in order to go from cause to effect. Yet these theoretical assessments often rely on assumptions that are in many cases violated in some kind in a practical environment. This section will provide a brief discussion of the weaknesses induced by the selected model and sample specifications.

Unbalanced Data

One general problem in panel analysis is the existence of missing values in the sample. While samples that have observations for every individual over every time period are considered to be balanced, this is rarely the case in practice. So if there are missing observations over some time periods for some observations, such a panel is referred to as a unbalanced panel. If the nature of missing observations is due to randomness they should not interfere with the analysis. If though there is a certain pattern that induces missing observations, it cannot be ruled out that this biases the outcomes of the analysis. For the used sample over advanced and developing countries this could be true, as the probability of a missing value for a given country is closely related to the time of the observation. With a lot more missing values in earlier time periods. According to Wooldridge (2010) this poses a problem to the chosen fixed effect setting only if the pattern behind missing values may be correlated to unobserved factors in the error term of our regression, leading to inconsistent estimates of the standard errors. As this cannot be ruled out and hardly be accounted for, it needs to be kept in mind when interpreting the results. Fortunately the crucial data of interest for our analysis, in particular the data on Inflation and the demographic variables are just slightly affected by missing values. While demographic data is entirely balanced across countries and time, inflation suffers from 12.5% missing values of inflation in the sample of advanced countries and 22.6% in the sample of emerging countries. Missing values are though a limiting factor when it comes to testing the robustness of findings regarding different control variables, as many of possible controls suffer from many missing values. Crucial control variables in most models with inflation as dependent variable, like the real interest rate and broad money growth suffer from severe lack of observations across both samples with more than half missing values. This limits inclusion of these variables to certain time periods.

Endogeneity

Another concern regarding the model specification lies in the problem of feasible endogeneity concerns for our demographic variables of interest regarding inflation. While demographics may have an effect on inflation, which we seek to investigate, inflation in turn will also have certain influence on demographics, which would result in possible overestimation of the coefficients on demographics. When considering the theoretical nature behind this endogeneity[15], that inflation will immediately affect fertility and migration which will in turn alter the age structure, we can address this problem partially by omitting the youngest age cohort with age 0-14 as they are most likely affected by changes in fertility. Furthermore, this makes an exact on point estimation of the impact of demographic characteristics on inflation and incredibly difficult task, as it can hardly ever be entirely ruled out that some kind of omitted variable bias or endogeneity is existing in the model specification.

Cross Sectional and Serial Correlation

What could also interfere with our results is the nature of inflation. This stems from the fact that inflation is largely autocorrelated, meaning that inflation in period t is closely correlated to inflation in period $t-1$. This autocorrelation will therefore most likely also be existent in the error terms of our regression, if it is not possible to explicitly specify the source of the correlation in the model. This would lead to wrong estimation of the standard errors in the regression. Another concern in similar manner stems from cross sectional dependence. Just as serial autocorrelation relates to correlation across time, Cross sectional dependence relates to a correlation across countries. As for example price adjustments in a border region will very likely also affect prices on the other side of the border.[16] In fact the cross sectional dependency Lagrange multiplier test developed by

[15] As also stated by Lindh and Malmberg (2000)
[16] See Baltagi (2014) p. 473

Breusch and Pagan indicated such dependency. In order to address these concerns, robust standard errors after Driscoll and Kraay [17] are implemented over the entire analysis, which account for autocorrelation and cross sectional dependence to some extent.

[17] See Dricoll and Kraay (1998)

Automatic Stabilization in a Small Open Economy - (How) does the Exchange Rate Regime matter?

Teresa Schildmann*

1 Abstract

In this article I address the question whether a country in a currency union can, ceteris paribus, expect more business cycle stabilization from systematic fiscal policy than a country outside, given monetary policy seems to be less responsive. To this end, I build on the heterogeneous agent DSGE model of McKay (2016) and modify the monetary policy rule such that it approximates long-run price dynamics of a country facing a currency peg. The subsequent experiments of reducing the scope of automatic fiscal stabilizers however show that fixed exchange rates are in general not more supportive to fiscal stabilization policies.

2 Introduction

With a prolonged slump in many European countries which only recently seemed to have come to a halt, and being coupled with an exhaustion of conventional monetary policy tools, attention has in the last years - in the scholarly debate,

*Teresa received her Master's degree from the University of Bonn in 2017 and subsequently started her studies towards the PhD at the Graduate School of Economics, Finance, and Management (GSEFM) in Frankfurt. The present article refers to her Master's thesis under supervision of Prof. Keith Kuester which has been submitted in May 2017.

as well as in policy advisory or operative institutions such as central banks - been shifted to assigning a greater role for business cycle stabilization to fiscal policy.

In some cases, e.g. at the zero lower bound (ZLB) on the nominal interest rate, economic theory supports the view that fiscal policy is not only called for, but also *most effective* in circumstances when monetary policy is constrained (see influential contributions by Eggertsson (2011) and Christiano (2011) on the government spending multiplier).

The second prominent case of constrained monetary policy is the impossibility of independent interest rate and exchange rate adjustments to counteract shocks faced by a country being member of a currency union. In this situation, the 'surrogate' hypothesis of fiscal policy being a substitute for the missing self-oriented monetary policy and re-balancing due to exchange rate adjustments is particularly tempting.

Studies on discretionary fiscal spending (i.e. in the form of government spending shocks), however, show mixed evidence: In the traditional Mundell-Fleming-view an exchange rate peg prevents the domestic currency from appreciating after a government spending shock, such that net exports are not crowded out, the trade balance stays unchanged and overall demand in the small open economy increases, in contrast to the flexible exchange rate system where the effect is zero. Empirical structural vector-autoregression (SVAR)-analyses by Ilzetzki (2013) and Born (2013) also document that the government spending multiplier is larger in countries with a fixed exchange rate than under a floating exchange rate regime. However, more recent theoretical studies have difficulties rationalizing these findings: When agents are intertemporally optimizing, what matters for their consumption-savings decision is the long-term real interest rate. Contrary to the case of a zero nominal interest rate (ZLB) as depicted above, where inflationary pressure induced by a government spending shock is completely accommodated by monetary policy such that real interest rates fall and private consumption spending is encouraged, dynamics are different in a currency union: As prices in a small open economy with fixed exchange rates are eventually pinned down by 'foreign' ones due to purchasing power parity (PPP), Corsetti (2013) find that the initial price increase following a domestic government spending shock must be met by a falling price level over time, leading to a rise in long-term real interest rates on impact and thus a *decline* in private expenditures and economic activity.

Farhi (2016) argue along the line that precisely because the exchange rate is fixed, an increase in government demand triggers an increase in domestic prices and thus a loss of competitiveness vis-à-vis the rest of the world. The terms of trade (price of imports relative to the price of exports) deteriorate which depresses overall consumption in the small open economy.

Summing up, besides the missing consensus, there is another short-coming of the literature which makes further work on this topic necessary: All studies cited above have focused on government spending shocks, however, especially in a currency union where these discretionary fiscal policy shocks may foster macroeconomic volatility and financial instability which contribute to cross-country imbalances and sovereign debt problems, casting an eye over *systematic* fiscal policies might be useful. These are less prone to suffer from drawbacks such as information, decision, and implementation lags, as well as a high risk of excessive debt accumulation if spending is not adjusted symmetrically over the cycle.

I will therefore focus on the workings and effectiveness of 'automatic fiscal stabilizers'[1] in stabilizing, i.e. minimizing the scope of business cycles of an open economy, depending on the prevalent exchange rate regime.

In this regard, the inclusion of an extensive fiscal branch and heterogeneous agents will be important, giving rise to a variety of fiscal transmission mechanisms. I will thus employ the framework of the heterogeneous agent incomplete markets New Keynesian business cycle model developed by McKay (2016) which is calibrated to the U.S. and through modifying its monetary policy rule from inflation to price level targeting (PLT) I achieve a closed economy approximation to a small open economy in a currency union.

This framework then allows me to assess whether a small country in a currency union could, ceteris paribus, expect more business cycle stabilization from fiscal policy than a country outside.

[1] This term is assigned to fiscal policies which have not primarily been implemented for the sake of short-term business-cycle stabilization, but rather for purposes related to social security or redistribution. There is therefore a 'natural' or 'automatic' movement of these government expenditure and revenue categories over the business cycle as income and employment status of citizens varies.

My results show that there is no evidence of an enhanced role of automatic fiscal stabilizers in this setting. While PLT leads to a more stable economy already, however, the interaction with fiscal measures as they are currently in place in the U.S. economy does not further reinforce this stabilizing effect. It also becomes clear that results hinge on the specific instrument as the effect is slightly reversed when considering the stabilization properties of progressive taxes, as well as on assumptions regarding the degree of nominal frictions in the economy.

The paper proceeds as follows: The next section describes the model setup, the quantitative fiscal experiment and its results are presented in Section 4. After a sensitivity check in Section 5, Section 6 discusses the findings and concludes.

3 The Model

By nature of the design of tax and social insurance systems, considering instrument as well as household heterogeneity and incomplete markets is almost obligatory in an analysis of automatic stabilizers. The model of McKay (2016) well combines this with a New Keynesian model setup featuring **nominal rigidities** in the form of Calvo (1983)-staggered price setting and monopolistic competition in the intermediate goods market, such that demand matters for aggregate dynamics.

I thus decided to build my model economy on theirs. Due to lack of space, for a description of the full model setup and equilibrium equations I refer the reader to McKay (2016) and will in the following only briefly sketch its main ingredients.

Household heterogeneity arises because a fraction of 80% of all households are ex ante less patient (the 'impatient' households) than the other 20% (the representative 'patient' household), modeled by a lower discount factor $\hat{\beta} < \beta$. Moreover, they face labor market status risk in the form of exogenous transitions between the three states employed, unemployed, and needy, and similar skill-level risk which they both cannot self-insure against because there is only one asset class (risk-less government bonds) available. Consequentially, there is an ex-post endogenously arising, non-degenerate income and wealth distribution which enables the analysis of not only intertemporal consumption smoothing motives as in classical permanent-income representative agent models, but moreover an assessment of channels such as precautionary savings or redistribution. These will most likely be of great concern, as automatic stabilizers themselves are designed mostly for equality and social insurance purposes in the first place.

The **fiscal side** is also deliberately complex and consists of flat consumption, property and capital taxes, progressive income taxes, as well as transfers to the unemployed and needy households, and (wasteful) government consumption. Government solvency is assured at all times by two fiscal feedback rules from public debt to government purchases and lump-sum transfers to the 'patient' households. By construction, the model therefore lets the researcher set aside issues of fiscal dominance and sovereign default as these fiscal rules ensure that the public budget constraint is respected at all times and the government obtains solvency for all possible levels of the price level (passive fiscal policy à Leeper (1991)). As the budget also varies along the business cycle, it constitutes an automatic stabilizer itself.

The **production side** consists of competitively producing final goods firms which buy intermediate inputs supplied by monopolistically competitive intermediate goods producers which in turn hire labor from the households and rent capital from a representative capital-producing firm. Profits of all firms are rebated back only to the patient households in the form of dividends.

There are **three aggregate shocks** - technology, monetary, and mark-up shocks - which are driving the business cycle. In order to allow for time-varying labor market risk, depending on the state of the economy, the transition matrix between employment states is modeled so as to depend on a linear combination of these shocks. Idiosyncratic income shocks are therefore transitory but persistent and business-cycle dependent.

Finally, **monetary policy** in this cashless-limit economy (no money trading frictions, following Woodford (1998)) is conducted by the central bank setting an interest rate target.

3.1 PLT as an Approximation to Fixed Exchange Rates

The attentive reader must have noticed that by now, there was not a single aspect of an open economy, so how is it possible to analyze different exchange rate regimes in this framework without introducing export and import sectors and trade explicitly?

The answer builds on Farhi (2016), Faia (2008) and Corsetti (2015): There is an 'open-closed economy isomorphism' which posits that the only crucial difference between a closed and an open economy model is the specification of the monetary policy conduct.

Moreover, in the long-run, the price level in a small open economy with a fixed exchange rate regime (nominal exchange rate $e_t = 1 \ \ \forall t$) is pinned down by the price level of 'foreign' (the rest of the world) due to purchasing power parity (PPP), which is also equal to the domestic steady state level in the absence of foreign shocks. Therefore, after e.g. a positive domestic shock which leads to increased inflation, PPP requires that the price level eventually returns back to its initial (i.e. the foreign) steady state value.

This price level behavior can exactly be modeled with the help of price level targeting (PLT): Under this monetary policy rule, the central bank counteracts all deviations from the steady state price level harshly because - in contrast to a Taylor Rule - it does not follow the notion to let 'bygones be bygones', but has a built-in 'error-correction' mechanism: Thus, while in a closed economy under a Taylor rule the price level never reverts back to its previous level after a shock of finite duration, PLT ensures that the interest rate increases so much as to let the price level fall asymptotically towards its long-run value, ensuring stationarity of the price level as depicted by Corsetti (2013).

PLT therefore basically mimics the long-run terms of trade (difference between domestic and foreign prices) adjustment needed in a credible exchange rate peg. A higher-than-steady-state price level (here, only arising due to sticky prices, not tradable versus non-tradable goods-producing sectors) therefore corresponds to a loss of competitiveness of the home economy. If interest rates do not adjust, the exchange rate in real terms would need to deteriorate (appreciate) under a peg, with contractionary effects on aggregate demand as it makes foreign goods cheaper and domestic goods more expensive, thereby reducing net exports. This finally drives down GDP until inflation in the domestic country has fallen to the world-level again, which is equivalent to domestic firms regaining competitiveness. In my closed economy example, these adjustments do not take place along the export-import margin, but are approximated by the consumption behavior of domestic consumers.

Therefore, in the following, I will employ the closed economy model depicted above but approximate it being in a credible exchange rate peg by modifying the monetary policy rule to a price-level-targeting one.

The baseline simple Taylor rule (TR) in which the nominal interest rate is reacting to contemporary inflation

$$I_t = \bar{I} + \phi_\pi \underbrace{\log(P_t/P_{t-1})}_{(\pi_t - 1)} - \varepsilon_t \tag{1}$$

is therefore changed into the price-level targeting (PLT) rule

$$I_t = \bar{I} + \phi_P \log(P_t) - \varepsilon_t. \tag{2}$$

Subtracting I_{t-1} yields

$$I_t = I_{t-1} + \phi_P \underbrace{\log(P_t/P_{t-1})}_{(\pi_t - 1)} - [\varepsilon_t + \varepsilon_{t-1}], \tag{3}$$

which is more convenient to implement as one does not need to track prices, but again only the inflation rate.[2]

Although the line of comparison in the subsequent analysis thus at first sight just seems to be along monetary policy conduct - Taylor rule inflation targeting (TR) versus price level targeting (PLT) -, I am actually able to extract channels working in the open economy and exchange rate regime dimension and therefore contrast an economy under flexible exchange rates (TR) with one under a peg (PLT).

[2]The reason to exclude any output smoothing term in any of the rules is driven by considerations of simplicity and of rendering the monetary policy rule not fully optimal in order to yield more room for fiscal stabilization, a strategy also employed by McKay (2016).

3.2 Calibration

The model's calibration is targeted to the U.S. economy over the time period 1960-2011 for most variables. For details I refer the reader to Table 2.1 in the appendix and Section 3.2 of McKay (2016).

In my baseline calibration of the PLT-rule, I have assigned the same parameter value to the price level which McKay (2016) assign to inflation, namely $\phi_\pi = \phi_P = 1.55$ in order to keep comparison across monetary policy regimes as tractable as possible.[3]

Table 2.2 in the appendix displays central moments of the model under either the original Taylor rule (TR) or my Price Level Targeting (PLT) monetary policy conduct and compares it to the corresponding U.S. time series used to calibrate the model. Note that the parameters have been chosen under the baseline Taylor rule specification (to mimic the standard deviation of log output, unemployment, and inflation) and thus fit the data a little less well under PLT monetary policy. A general difference one can detect is that, ceteris paribus, price-level targeting smooths the evolution of all macroeconomic aggregates in the model, most notably of inflation whose standard deviation declines from 0.6 to 0.14 percent; it thus has a stabilizing effect of its own.

4 The Experiment: The Impact of Automatic Stabilizers

After having estimated the model with all fiscal stabilizers in place, separately for each monetary policy rule, now, the fiscal experiment of interest is conducted: Following McKay (2016) it consists of a replacement of the progressive income tax by a flat tax, a cut in all proportional taxes by 10% and a cut in unemployment and poverty benefits of 80%. Moreover, the fiscal adjustment parameters in the fiscal rules are reduced proportionally so that the budget deficit variance falls by 10%.

[3] Note that under PLT, unlike in an inflation-targeting Taylor rule, the coefficient of the price level may be well below unity - and therefore violating the Taylor principle -, without jeopardizing determinacy of the equilibrium (see Michael (2003)). Robustness checks of mine not presented in this article confirm that results are not sensitive to this specific parameterization.

It is then possible to assess the effect of this experiment - basically the removal of automatic fiscal stabilizers - on the volatility of major economic aggregates either in an open economy with flexible exchange rates (Taylor Rule specification) or an interest rate peg (Price-Level Targeting specification).

The indicator of interest is the Smyth (1966)-stabilization-coefficient, defined as

$$S_{var} := \frac{V'}{V} - 1$$

where V is the ergodic variance of a selected macroeconomic variable in the baseline calibration with all stabilizers being active, and V' the variance in the counterfactual when stabilizers are 'turned off' according to the experiment described above. The measure thus provides a quantification of the fraction by which the variance of aggregate activity would increase if one removed all automatic fiscal stabilizers. I define a similar coefficient, S_{mean}, in order to illustrate the change in the first moment of a variable.

Table 2.3 in the appendix provides the Smyth (1966)-coefficients for five variables deemed important for characterizing an economy's business cycle (output, working hours, consumption, investment, and inflation) and allows for two dimensions of comparison: Firstly, the effect of fiscal policy on the mean and variance of the respective variable (given the monetary policy specification), and secondly, the comparison across economies with different monetary policy conduct (i.e. exchange rate system).

To begin, the removal of all automatic fiscal stabilizers according to the experiment would lead to *higher* average economic activity (positive S_{mean}). This result is not only qualitatively but also quantitatively identical across monetary policy specifications.

Economically, one can imagine that e.g. eliminating the progressiveness of income taxes increases the returns from working for the high-income households and thus increases labor supply and production on average. Lowering capital and property taxes stimulates investment and thus positively affects the aggregate capital stock. Finally, lowering the amount of transfers payed to impatient households diminishes government outlays and therefore the lump-sum tax on the patient household needed to finance these. Income of the latter thus rises and alongside their savings and investment. The falling income of the poor and needy is counteracted by a rising labor supply which again stimulates production, while the decline in overall consumption is cushioned by the increased goods demand of the wealthy, patient households.

Built-in fiscal stabilizers thus actually seem to *dampen average* economic activity robustly across monetary policy specifications. But do they at least decrease the scope of business cycles, i.e. the *variability* of major aggregates?

The answer seems to be 'no', confirming the results of McKay (2016): *Lowering* all automatic stabilizers would actually lead to a *more stable* economy in terms of the variance of output, hours, investment(, and inflation), as indicated by negative values of S_{var}; only consumption would be more volatile, with its variance increasing about 10.5% and 12.3% under PLT and TR, respectively(, and inflation under PLT).

In order to try to grasp the dynamics behind this change, it is useful to keep in mind that there can only be a substantial effect on the volatility of variables if stabilizers affect agents' income differently at different points of the business cycle, because only in this case do they significantly affect their intertemporal decisions.

The positive link between fiscal policy and consumption stability is most likely driven by the impatient households: in recessionary periods when more households of this type are laid off, with drastically reduced transfers their incomes would experience a substantial decline. Anticipating this, impatient households try to self-insure already in good times by engaging in precautionary savings to meet this heightened income risk. However, they cannot compensate for the whole transfer amount and therefore, consumption of the impatient households will vary considerably more in an economy without fiscal stabilizers.

For the mass of patient households, in turn, as they have to finance the transfer payments to impatient households especially during recessions, their income is substantially more time-varying when automatic stabilizers are in place, leading to lower consumption and investment variability when they would be removed. Concerning labor supply, a wealth effect inherent to the utility specification discourages labor supply of impatient households relatively more in recessions because of the income increase poor working households experience under fiscal stabilizers. They will therefore adjust their labor supply procyclically, increasing working hour variability and thus output volatility relatively to the scenario when automatic fiscal stabilizers are turned off.

Turning to the comparison across monetary policy - i.e. exchange rate - regimes, except for the case of inflation variability, automatic stabilizers seem to have led to a *relatively more unstable economy under PLT* than under TR.[4]

The reason for this negative interaction most likely lies in the property that PLT counteracts inflationary pressures more harshly than inflation targeting. This implies that *countercyclical* fiscal policies leading to increased demand in recessions which in the presence of nominal rigidities raises output and finally - through rising labor demand and marginal costs - prices, will be hampered in their effectiveness.

The inflationary effect of *procyclical* fiscal policies, by reversion of the argument, will be dampened, leading to relatively higher real rates in booms and thus to less volatility and more business cycle stabilization under PLT than under TR inflation targeting.

In order to give further support to this hypothesis and to shed light on the channels at work, in the following I present the effects of removing only one stabilizer at a time on business cycle dynamics of the economy: Firstly, I assess the effects of an experiment of replacing progressive taxes by flat taxes (while making sure steady-state government revenues are unaffected), and secondly, a cut in transfer payments to unemployed and needy households by 80% (with the resulting government surpluses being rebated to the patient households in a lump-sum fashion).

Experiment 1: Progressive Taxation and PLT Table 2.4 in the appendix provides the Smyth (1966)-coefficients of this experiment. One can see that with respect to the business cycle stabilization properties of progressive taxes, compared to the experiment of lowering *all* stabilizers as analysed above, results are qualitatively reversed as well as quantitatively of a lower magnitude. For the comparison across monetary policy regimes this implies that progressive taxes seem to lead to a slightly more stable economy under price level than under

[4]This can be seen in Table 2.3 as the coefficients of S_{var} under PLT are lower than under TR monetary policy conduct, meaning the removal of all stabilizers leads to a relatively lower variance of the respective variables.

inflation targeting as indicated by higher values of S_{var} in the first case (except for investment), although the difference is not very pronounced.

In order to reconcile these findings, one needs to be aware of the interaction between progressive income taxes and monetary policy to determine the effective response of real interest rates to inflation. As the after-tax nominal interest rate $I_t^{\tau} := (1 - \tau^x(x_{t+1}))I_t$ under progressive taxation is both lower than the before-tax interest rate and less responsive (it falls by less in recessions when income x_{t+1} is low and falls by relatively more when incomes are high), also real interest rate fluctuations are cushioned (the real rate is less elastic to inflation/interest rate changes). Inflationary tendencies in boom periods are therefore less moderated and vice-versa in recessionary periods. PLT, however, is a way of rendering real interest rates more responsive. As it is less accommodative, in response to positive shocks exerting inflationary pressure it induces already a higher pre-tax interest rate response than under inflation targeting, depressing consumption and investment in the high-activity episode and thereby stabilizing macroeconomic aggregates over the cycle.

Experiment 2: Transfers and PLT As the results in Table 2.5 show, while a cut in transfer payments renders the economy more volatile in the TR-case (except for consumption), the effect under PLT is small with signs being reversed.

With sticky prices, the transfers paid to the impatient households in a recession raise output because of the high marginal propensity to consume of this subgroup of the population. The need to finance these requires a drop in government spending and rising lump-sum taxes on the patient households by the fiscal adjustment rules. As the fall in government spending is again crowding in private consumption, the price level may rise when firms adjust their production. By the above-mentioned channels, the inflation-sensitive price-level targeting rule might in response increase, leading to a rise in real interest rates which encourages households to save, at the expense of consumption and investment. Transfers therefore ultimately lead to a less stable economy in terms of investment, hours, and output volatility.

In the Taylor rule economy, with a lower interest rate response and less adverse incentive effects, the intratemporal reaction of patient households seems to dominate, who increase their labor supply in recessions due to their lower wealth to finance the transfer payments, thereby stabilizing labor supply and output.

In summary, the relative ranking of TR and PLT regarding which monetary policy renders fiscal measures more effective, is depending on the fiscal policy considered: While progressive taxation seems to be slightly more stabilizing under PLT, for transfers the opposite is true. This might be rationalized by differential price dynamics induced by the different policies, leading for instance to rising real interest rates in economic downturns under transfers, counteracting their stimulative effect.

It also becomes clear that the stabilizer mix considered in the full fiscal experiment even leads to a more volatile economy under PLT than the policies examined in isolation.

5 Sensitivity: The Degree of Nominal Rigidity

Following Erceg (2012) who assign first-order effects to the degree of nominal rigidities in determining the impact of a government spending cut on the economy in cases of extremely constrained monetary policy, I also take a brief look at how the effectiveness of automatic stabilizers on smoothing the business cycle might change if the degree of price stickiness is being increased: Previously, the model set-up entailed a Calvo (1983) price stickiness parameter of 0.286, calibrated to match an average price duration of 3.5 quarters, which is now adjusted to 0.167, implying that price adjustments only take place every six quarters.[5]

Results of the experiment of reducing all automatic stabilizers seem to substantially depend on this parameter, as the Smyth (1966)-coefficients in Table 2.6 demonstrate: The change in the variance of major economic aggregates is quantitatively substantial and leads to a sign reversion compared to the baseline case with more flexible prices. Now, automatic stabilizers seem to contribute to macroeconomic stability ($S_{var} > 0$), with the effect being substantial under TR inflation targeting.

[5] This is for instructive purposes, because empirically the baseline value is more supported, given that e.g. Klenow (2011) finds evidence that prices in the U.S. adjust roughly every year.

One possible explanation I can think of to reconcile these findings with my previous results is that, due to slower adjustments in inflation and the nominal interest rate, also the real rate fluctuates too little to induce sufficient buffer-stock savings behavior in good times, leading to less self-insurance possibilities for impatient households and too much consumption and investment volatility. Automatic stabilizers can counteract this development by providing insurance for the impatient households facing income risk while lowering the income of the patient households who are showing more procyclical behavior, thereby smoothing their investment behavior and labor supply. The relatively lower effectiveness of automatic stabilizers under PLT, as pointed out above, could again be a result of overshooting price level stabilization, e.g. by counteracting inflationary tendencies induced by transfer payments in recessions.

In general, the results thus point to a higher effectiveness of fiscal policy in currency unions when prices are stickier, albeit not nearly as pronounced as under flexible exchange rates.

6 Conclusion

With this work, I wanted to shed light on the question whether the exchange rate system is an important determinant of the ability of automatic stabilizers to smooth the national business cycle and cushion domestic shocks.

To conclude and answer the question I have posed in the title of this work, yes, I can confirm that there generally is an effect of the exchange rate system. The direction and quantitative scope is however highly dependent on the fiscal instrument considered and the institutional environment of product markets.

I have taken a long-run perspective in positing that the main difference between flexible and fixed exchange rates is the fact that the latter economy is ultimately tied to the foreign price level, which can be approximated by price-level targeting. In this regard I confirm recent findings that fixed exchange rates display less, not more accommodative behavior. This is in contrast to the case of a binding zero lower bound (ZLB) and underlines once more the need for a distinct analysis of fiscal policy in each of the monetary policy-constrained settings. The actually higher responsiveness (instead of non-responsiveness under ZLB) of nominal interest rates leads to more movement in real rates. The induced behavioral responses might harm the workings of automatic stabilizers especially in cases where they cause counter-cyclical price shifts (i.e. upward price level shifts in recessions).

The findings therefore support recent theoretical work on discretionary fiscal policy (e.g. Farhi (2016), Corsetti (2013) or Corsetti (2016)) and contrast the Mundell-Fleming view which would posit that fiscal policy is more effective in an exchange rate peg.

It shows that the stabilizer mix as present in the U.S. in the last decades would have proven even more destabilizing on the business cycle if the country had been part of a currency union.

While qualitative implications might probably hold, the exact quantitative results will most likely not be transferable to a different country as fiscal policies as well as institutions and market dynamics are quite country-specific. Countries in the European Monetary Union (EMU) for example, for which the question is also practically relevant, are known to face a higher degree of nominal rigidities than the U.S. which, following my sensitivity check, could be slightly beneficial to the workings of automatic stabilizers. Moreover, further research could integrate *endogenous* labor market dynamics, creating novel channels through which fiscal policy transmits to the aggregate economy via affecting participation and job creation decisions.

Lastly, it would also be interesting to see if and how limitations to the debt and deficit ratios e.g. of countries in the EMU in the form of the Stability and Growth Pact (SGP) influence the workings of automatic stabilizers and the possibilities for households to self-insure in the open economy setting.

This list of possible extensions is surely long and the assessment of the workings of automatic stabilizers in different economic environments is expected to be a promising field of research.

With my analysis I hope to have shed light on at least some aspects of these.

Online Appendix to: Automatic Stabilization in a Small Open Economy - (How) does the Exchange Rate Regime matter? by Teresa Schildmann

Calibration

Table 2.1: Calibration of the Parameters

Symbol	Parameter	Value	Target (Source)
Panel A. Tax bases and rates			
τ^c	Tax rate on consumption	0.0535	Avg. revenue from sales taxes
β	Discount factor of stock owners	0.989	Consumption-income ratio = 0.689 (NIPA)
τ^p	Tax rate on property	0.00258	Avg. revenue from property taxes
α	Coefficient on labor in production	0.296	Capital income share = 0.36 (NIPA)
τ^k	Tax rate on corporate income	0.35	Statutory rate
υ	Deduction of capital costs	0.68	Avg. revenue from corporate income tax
ξ	Fixed costs of production	0.575 0	Corporate profits / GDP = 9.13% (NIPA)
μ	Desired gross markup	1.1	Avg. U.S. markup (Basu, Fernald, 1997)
Panel B. Government outlays and debt			
$\bar{T}^u$	Unemployment benefits	0.144	Avg. outlays on unemp. benefits
$\bar{s}^u/\bar{T}^u$	Max. UI benefit / avg. income	0.66	Typical state law (BLS, 2008)
$\bar{T}^s$	Safety-net transfers	0.151	Avg. outlays on safety-net benefits
G/Y	Steady-state purchases / output	0.145	Avg. outlays on purchases
γ^T	Fiscal adjustment speed (tax)	-1.6	St. dev. of deficit/GDP = 0.0093 (NIPA)
γ^G	Fiscal adjustment speed (spending)	-1.28	St. dev. of spending = 0.0126 (NIPA)
B/Y	Steady-state debt / output	1.7	Avg. interest expenses
Panel C. Income and wealth distribution			
ν	Non-participants / stock owners	4	
β^h	Discount factor of households	0.979	Wealth of top 20% by wealth
$\bar{s}$	Skill level of stock owners	3.72	Income of top 20% by wealth (SCF)
Panel D. Business-cycle parameters			
θ	Calvo price stickiness	0.286	Avg. price spell duration = 3.5 (Klenow, Malin, 2011)
ψ_1	Labor supply	21.6	Avg. hours worked = 0.31 (Cooley, Prescott, 1995)
ψ_2	Labor supply	2	Frisch elasticity = 1/2 (Chetty, 2011)
δ	Depreciation rate	0.0114	Annual depreciation expenses / GDP = 0.046 (NIPA)
ζ	Adjustment costs for investment	6	St. dev. of I = 0.053 (NIPA)
ρ_z	Autocorrelation productivity shock	0.75	Autocorrel. of log GDP = 0.864 (NIPA)
σ_z	St. dev. of productivity shock	0.00294	St. dev. of log GDP = 1.539 (NIPA)
ρ_m	Autocorrelation monetary shock	0.62	Largest AR for inflation = 0.85 (Pivetta, Reis, 2006)
σ_m	St. dev. of monetary shock	0.00353	Share of output variance due to shock = 0.25
ρ_p	Autocorrelation markup shock	0.85	
σ_p	St. dev. of markup shock	0.0251	Share of output variance due to shock = 0.25
ϕ_π	Interest-rate rule on inflation	1.55	St. dev. of inflation = 0.638 (NIPA)

Note: Table corresponds to Table II in McKay and Reis (2016).

Table 2.2: Model and Data Moments, Baseline with All Stabilizers

	TR	PLT	Data
Stdev output (×100)	1.54	1.43	1.54
Stdev inflation (×100)	0.60	0.14	0.64
Stdev hours (×100)	2.42	2.29	1.89
Stdev investment (×100)	5.18	4.72	5.30
Stdev unemployment (×100)	0.92	0.92	0.94
Stdev SNAP ratio (×100)	1.96	1.96	2.05
Corr output - unemployment	-0.77	-0.74	-0.62
Largest AR root of inflation	0.68	0.54	0.85

Note: output, hours, consumption and investment are in logs and HP filtered with smoothing parameter 1600. SNAP is short for Supplemental Nutrition Assistance Program. **McKay2016** *use this variable to calibrate the share of needy households.*

Experiment Results

Table 2.3: The Effect of All Stabilizers on the Business Cycle

	TR		PLT	
	S_{var}	S_{mean}	S_{var}	S_{mean}
Output	-0.0229	0.0567	-0.0422	0.0567
Hours	-0.0296	0.0344	-0.0325	0.0344
Consumption	0.1232	0.0603	0.1046	0.0603
Investment	-0.0421	0.0174	-0.1420	0.0174
Inflation	-0.2828	0.0001	0.1051	0.0000

Note: Proportional change caused by cutting all stabilizers.

Table 2.4: The Effect of Progressive Taxes on the Business Cycle

	TR		PLT	
	S_{var}	S_{mean}	S_{var}	S_{mean}
Output	0.0023	0.0446	0.0048	0.0446
Hours	-0.0147	0.0388	0.0159	0.0389
Consumption	-0.0665	0.0507	-0.0258	0.0508
Investment	0.2135	-0.0051	0.1046	-0.0051
Inflation	-0.3893	0.0002	0.0727	0.0000

Note: Proportional change caused by cutting/removing the respective stabilizer.

Table 2.5: The Effect of Transfers on the Business Cycle

	TR		PLT	
	S_{var}	S_{mean}	S_{var}	S_{mean}
Output	0.0603	-0.0004	-0.0165	-0.0004
Hours	0.0944	-0.0098	-0.0190	-0.0098
Consumption	-0.0133	-0.0004	0.0330	-0.0005
Investment	0.2729	0.0001	-0.1038	-0.0000
Inflation	-0.1914	0.0000	-0.0126	0.0000

Note: Proportional change caused by cutting/removing the respective stabilizer.

Sensitivity Analysis

Table 2.6: The Effect of All Stabilizers on the Business Cycle, Stickier Prices

	TR		PLT	
	S_{var}	S_{mean}	S_{var}	S_{mean}
Output	0.2129	0.0568	0.0164	0.0567
Hours	0.1305	0.0344	0.0176	0.0344
Consumption	0.3126	0.0605	0.1160	0.0603
Investment	0.3952	0.0174	0.0987	0.0174
Inflation	-0.1700	0.0001	0.2329	-0.0000

Note: Proportional change caused by cutting all stabilizers.

An Early And Lasting Advantage? – Long Term Effects Of The School Starting Age

Arnim Seidlitz*

1 Abstract

This paper investigates long term effects of the school starting age. It is based on the Swiss TREE-panel which consists out of PISA data for about 6000 students and a follow-up survey in nine waves. The outcome variables are PISA test scores, the decisions to continue schooling until the university entrance diploma and to attend a college and the events of obtaining the university entrance diploma and a college degree. I estimate a general effect of the school starting age for a subgroup of my sample using a fixed-effects regression and an isolated effect of the age position compared to the peers with an IV-regression. The estimations for PISA test scores, the decision to continue schooling and finally obtaining the university entrance diploma show a negative effect of being older at school entrance. These outcomes are at odds with the previous literature and may result from a sample selection issue. However, the estimations for the effect of being older compared to the peers at school does not face this problem. I find a significant negative effect of a higher age position on the likelihood to continue schooling but no effect on the likelihood to graduate with a university entrance diploma.

*Arnim received his M.Sc. from the University of Bonn in September 2017. Afterwards he began his Ph.D. at Humboldt University Berlin.

2 Introduction

Following a first intuition, one may think the school starting age is not a really important issue. The age of children when they enter school differs for most cases by a few months, at most up to year. However, there is convincing evidence for a positive effect of being older on early test scores in school (**hamori**; **bedard**; **datar**; **elder**; **pena**; **robertson**).

In this master thesis, I try to answer the question whether this early advantage translates into differences in long term outcomes. In particular, I estimate the effect of the school starting age on test scores in high school and educational choices as well as attainments afterwards.

Once a difference in skills has been established, it is unclear whether that difference increases, stabilizes or diminishes (**cunha**). The evidence so far is mixed. **black**; **nam** do not find a significant effect on the probability of obtaining a university diploma, whereas **zweimueller**, **pena** and **fredriksson** are able to show a positive age effect at least for certain subgroups in Austria, Mexico and Sweden, respectively.

Generally, the empirical estimation of age effects faces the problem of endogeneity. In most countries, children possess a school starting age assigned by the legislator which is determined by her date of birth and a cutoff date. This assigned school starting age is most commonly treated as exogenous and empirically in many cases assigned and actual school starting will coincide. Yet, most systems contain exceptions and possibilities allowing families and school authorities to deviate from the assigned school starting age. As those deviations – retaining, so-called "redshirting" or entering the school voluntary a year earlier – are presumably non-random, the actual school starting age is ultimately an endogenous variable. In my main analysis, I restrict my sample to a subsample containing only observations for which I argue that the assigned and the actual school starting age are equivalent.

I estimate further the influence of the age position relatively to the peers in ninth grade. That question is of course related to the question of the school starting age as the age position is driven by the age at school entrance and it may contribute to the debate about the concrete channels of the age effect. Especially age effects on test scores could originate from different sources, the age relative to the student's peers, the age at school start or the age of the day when the test is taken. Most databases are unsuitable to separate those effects and in many cases, researches rather regress on a combined effect. The papers which are able to destinguish the effects find that the positive effect indeed arises from age at test while the direct effect of being among the oldest at school entry is found to be negative. The relative age compared to the peers is also found to be negative. However, both negative effects are offset by the age-at-test-effect such that the combined effect is positive (**black**; **crawford**; **pena**).

My paper is based on the Swiss TREE-panel which provides me with data of approximately 6000 ninth graders in the school year 1999/2000. For them, the period from their mid-teenage until their late 20s is closely observed. The data include detailed information on the students' test scores in ninth grade, their educational choices and their final educational achievements.

As a main advantage of my data, it is possible to separate outcomes of obtaining a degree and starting the education to get it. This distinction is not trivial and potentially important. Once, a subgroup is identified to have a lower probability to graduate with a university degree for example, this outcome does not provide any information why those people have a lower probability to get the formal attainment. A higher share of university drop-outs or of those who never start attending a college are equally possible. For a policy reform aiming to increase the share of graduates this difference might be essential. To my current knowledge no other paper investigates both. Finally, I seem to be the first who investigates the effects of the school starting age for Switzerland. I would attribute this lack of evidence to the complex situation with cutoff dates on state level and the challenge of working with legal texts in German, French and Italian[1].

The remainder of this text is organized as follows. In the third and fourth section, I explain the data base and my empirical strategy. The fifth section contains the results from my estimations and the sixth section a discussion about their interpretation.

[1] The different school entry laws for the relevant period in the 26 states can be found in **cesdoc**.

3 Data

This paper is based on the data of the first PISA wave in 2000[2] and a follow up survey from Switzerland. A subsample of the test takers was chosen to participate in the so called TREE-panel[3] and those who replied were surveyed in nine waves so far. The last one was conducted in 2014 when the participants were about 29 years of age (**tree**). Information is available on parental backgrounds and reading, math and science test scores in ninth grade[4] and for the years to follow on educational choices and attainments.

To construct my main independent variable, I use the month and year of birth, which are reported in the PISA-data base and the school cutoff dates. In Switzerland those cutoff dates are subject to state laws, see table 3.1.

My data do not contain the actual school entry. However, from birth dates and the respective cutoff date I am able to construct the assigned year of school entry and according to this, the grade the student would attend in the school year 1999/2000, when the PISA test was taken, in absence of grade retention or skipping.

I restricted my sample to those observations that explicitly reported to be in grade nine and whose expected grade does not deviate by more than one year from that. The sample is thus based on those students who should have started school in 1991 according to their birth dates and consequently entered grade nine in 1999 when they always proceeded from one grade to the next. I will refer to them henceforth as "compliers". They are roughly two thirds of my observations. Further, I included those observations whose age does not deviate by more than one year from that, meaning those cases who would be in grade eight or ten, respectively, if they complied to the school entry laws and were always advanced to the next grade. Together with the "compliers", they form the "extended sample".

[2]The Programme for International Student Assessment (PISA) was set up by the Organisation for Economic Co-operation and Development (OECD) and firstly conducted in 2000 (**pisa**).

[3]The Swiss panel study TREE (Transitions from Education to Employment) is a social science data infrastructure mainly funded by the Swiss National Science Foundation (SNF) and located at the University of Bern (**tree**).

[4]Unlike most other countries, Switzerland did not draw a sample of 15 years old student but a sample of ninth-graders regardless of their age (**pisa**).

3.1 Outcome Variables

The information on test scores is not complete for all observations. Some test takers solved tasks in all three subjects, the majority in reading and either math or science and a third group was only tested in reading[5]. I normalized for each subject the test scores to have mean zero and a standard deviation of one and formed a variable for quantitative skills capturing the scores in math and science. For students having scores in both subjects, the average was taken. For the others, I used the available test score in math or in science. I then normalized the quantitative scores again.

As other outcome-variable, educational choices are of interest. One outcome is whether a student ever chose to start such an academic upper secondary education – continued schooling at an institution where it is possible to obtain the university entrance diploma. Analogously, for the academic tertiary education, the criterion is whether a person was ever enrolled at a university. Observations will be considered as not choosing to start the upper secondary and the tertiary education if they have not done so by their 23rd and their 27th birthday, respectively.

Further, I regress on educational attainments and that for different points in time. Precisely, I use the outcome of holding the academic upper secondary certificate at the 19th, 21st and 23rd birthday and a university certificate at the 23rd, 25th and 27th birthday as depending variable.

3.2 Independent and control variables

The main independent variable in this analysis is the month of birth relative to the cutoff date, *bmrc*. I centered that variable around zero such that $bmrc \in [-5.5; 6.5]$. For a state with the cutoff at December 31st (or January 1st), it would be equal to - 5.5 for children born in December, as they are the youngest, - 4.5 for those born in November and finally 5.5 for the oldest group, January-born[6] (**fredriksson**).

[5] In PISA 2000 each participant was confronted with one out of nine different test booklets each consisting of four tasks. Those booklets contained either four reading tasks, three reading tasks and one math or one science task or two reading and both, a math and a sciences task. Further, there were different possible tasks for each of the three subjects (**pisa**).

[6] Due to the shift in cutoff dates in some states, there are few observation who would be twelve month older than their youngest peers. Consequently, for those cases *bmrc* equals 6.5. The

As discussed above, it is also interesting to learn more about the concrete nature of the age effect. I try to contribute to this discussion with a variable which captures the relative age effect. This variable is the age position compared to the peers in ninth grade, $agepos9$. It is based on the age rank which is one for the youngest student in school, two for the second youngest and for a school with n observed students n for the oldest one. I divided that rank by the number of observed students for the school. This fraction can be interpreted as the age percentile within the school. It will be for example 0.75 if three quarters of the peers at school are younger or have the same age. To center that variable around zero as well, I subtracted the school average from each observation, hence $agepos9 \in (-0.5; 0.5)$.

As control variables, I use some information from PISA-data, that are the information on gender, the country of birth of the participants and their parents, maternal education, the language spoken at home, the composition of the household and the school type visited in ninth grade. For the distribution of those variables and a more detailed description see table 3.2.

3.3 Selectivity

Unfortunately, there are reasons to fear a selectivity bias.

Figure 3.1 shows the distribution of birth months relative to the cutoff date. As one would expect, the extended sample comes close an identical distribution. However, for the compliers we see a different pattern. For them, the share of students born in months directly before the cutoff and consequently having a low value for *bmrc*, seems to be relatively low, figure 3.1. Of course, these descriptives do not allow any causal interpretation but the discrepancy would be in line with the idea that children born close to the cutoff are either significantly more often redshirted by school entrance or have a higher probability of being retained.

To check for selectivity-issues, I also performed a regression of the control variables on a dummy for being a complier. The t-statistics of this regression are displayed in the right-hand column of table 3.2. Indeed, some of the differences between the compliers and the extended sample are estimated to be significant.

relative high number of states shifting their cutoffs in the early 90s can be explained by a shift in the begin of the school year in those states in 1989.

Apart from gender, these variables are known to be proxies for the socioeconomic background of the student. We need therefore to presume that the compliers are positively selected according to their socioeconomic status. This will be crucial for the interpretation of the regression results.

4 Empirical Strategy

As described above, the estimation of age effects in school faces generally the problem of endogeneity. My strategy to avoid this endogeneity issue is based on the variable birth month relative to the cutoff, *bmrc*, which is constructed using month of birth and the cutoff of the respective state and not perfectly correlated to the month of birth as cutoffs vary on state level.

One main shortcoming of my data is that they do not contain the information on the year of school entrance. I try to resolve this problem by restricting my analysis to those students who should be in ninth grade according to their date of birth, the "compliers". For them it seems justifiable to assume a common year of school entrance. The birth month relative to the cutoff are then perfectly collinear to the actual school starting age. Individual deviation from that may exists but they are rare and there is no reason to fear that they add a systematic bias.

To estimate the effect of the school starting age, I use the following fixed effects model.

$$y_{ism} = \alpha_0 + \beta_{bmrc} bmrc_{ism} + X'_{ism}\gamma + \delta_s + \lambda_m + \varepsilon_{ism} \quad (1)$$

In this regression, y_{ism} is the outcome variable for student i of school s born in month m. α_0 is a constant, X_{ism} is a vector of controls containing the variables discussed in section three, δ_s and λ_m are fixed effects for the school and the month of birth[7], respectively and ε_{ism} is the error term.

[7] There is a literature arguing that the month of birth influences outcomes through differences in parental characteristics (**buckles**; **cascio**). It is hence advisable to control for a direct month of birth effect when working with a sample which is confronted with multiple cutoffs (**bedard**; **elder**), as I do, see table 3.1.

Despite using the fixed effects, the standard errors are further clustered on school level. I thus allow for within-school correlations and heteroskedasticity which cannot be fully captured by fixed effects (**cameron**).

The second part of my empirical analysis is targeted on the effect of the relative age towards the peers in school. Following a first intuition, one could think of an equation which is analogous to (1) and only substitute $agepos9$ for *bmrc* but that would involve accepting the age position as exogenously given.

To avoid the problem of endogeneity of the age position, I use *bmrc* as an instrument for $agepos9$. Students who are born directly before the cutoff will be relatively younger regardless of any endogenous influences like redshirting, grade skipping or retention.

I will hence use an instrument variable approach based on a two-stage least square estimation. The model stays the same as above. I use the same controls, fixed effects and standard errors. Its second stage with fitted values of the first regression, $age\hat{pos}9_{ism}$, can be displayed as:

$$y_{ism} = \alpha_{2nd} + \beta_{2nd} age\hat{pos}9_{ism} + X'_{ism}\gamma_{2nd} + \delta_s + \lambda_m + \omega_{ism} \quad (2)$$

5 Results

Table 3.3 shows the estimates for the effect on PISA test scores. All effects are given in percentages of a standard deviation. Columns one, two, four and five display the effect of the school starting age measured in *bmrc* for the compliers on reading and quantitative skills.

Surprisingly, the age effect of being a month older at school entrance is significantly negative. The effects of my regressions seem to be rather small, each additional month of age decreases on average the scores by about one percent of a standard deviation. On an individual level nevertheless, these differences could still matter as parents are potentially confronted with the decision to increase their child's school starting age not by one but by twelve months.

The columns three and six show that the point estimates for the relative age effect estimated by the IV-regression, are positive. In the literature this effect was shown to be negative (**elder**; **pena**). However, the estimated coefficients are far from being significant.

In table 3.4, we see the estimates for the regression on educational choices. The left-hand panel shows the result for having entered an upper secondary education meaning ever being enrolled in a school where it is possible to obtain the general university entrance diploma. The right-hand panel contains the results for the analogous regressions on the tertiary education. Here, the outcome is a dummy on having ever been enrolled at a university whereas I censored the outcome of starting the two education forms on the 23rd and 27th birthday respectively.

For the upper secondary education, the effect of age at school entrance is again significantly negative. Being a month older decreases the probability of ever attending the upper school by slightly less than one percentage point.

As shown in the third column, the relative age has a significant negative effect as well. This effect is quite large in size, but it is also a bit harder to interpret. Recall, the variable for the relative age effect, $agepos9$, is constructed to have a magnitude of one meaning that the estimated 22 percentage points display already the maximum difference, the difference between the youngest and the oldest student in that grade.

Concerning a university education, the main variables of interest are close to zero and insignificant.

Table 3.5 displays the regressions on the outcome of holding the university entrance diploma at the 19th, 21st and 23rd birthday. Unsurprisingly, both age effects are strongest among the 19-years old. Students who were older at school entrance or those who belonged to the oldest in ninth grade are thus less likely to be among the youngest school graduates. For this age group, the negative age effect seems even to be the most important channel compared to the influence of the covariates.

All differences between the regressions on holding the certificate on the 21st or 23rd birthday are insignificant. The negative effect of the school starting age is still significant, but it is estimated to be only about half of a percentage point in absolute value. In general, my estimates in column two and three almost replicate those in the regression on the decision to start the upper secondary education in column two of table 3.4.

The relative age effect on the upper secondary certificate is only significant for the first point in time. Later it is still negative but insignificant. For the decision to start the upper secondary education, the effect was significantly negative. Students which have a high age position in ninth grade are thus less likely to continue schooling but there is no significant difference in the probability of finally holding the university entrance diploma.

The regressions for obtaining a university diploma are shown in table 3.6. Only for the outcome of holding the degree on the 23rd birthday, there is a significant effect of the school starting age. It is less than one percentage point but highly significant. However, the point estimate of being a month younger at school entrance is approximately a third of the effect on holding the upper secondary degree at age 19. The people graduating very young from high school are thus only partly able to translate this advantage into an early success at the university. The effect of the age position in ninth grade is close to zero and always insignificant.

In the full version of this paper, I also included a heterogeneity analysis concerning gender, migration background, maternal education and single parent status. In the literature, there is evidence for the existence of those effects (**suziedelyte**; **hamori**; **elder**). However, in my data I could not find any hint for a heterogeneous effect on any outcome.

6 Discussion

Probably most important for the interpretation of the results for the school starting age remains the selectivity concerns. As argued in section three, the compliers are presumably positively select.

On the other hand, it seems still reasonable to assume that the event of birth is exogenous. It should hence be the case that there is no sorting into the complier period or into certain months in general. One would thus suspect that there was an initial state in which firstly all calendar months and consequently all birth months relative to cutoff had approximately an equal share on the number of compliers and secondly demographic characteristics were equally distributed between compliers and non-compliers. However, in our random sample of ninth graders neither seems to be the case. The share of birth months directly before the cutoff is remarkably lower, as shown in figure 3.1 and different control variables vary significantly between compliers and non-compliers, see table 3.2.

The compositional distortions could be attributed to a finding of **elder** who found children who are younger at school start to have a higher probably of being retained. The PISA test takers needed to pass many stages of potential sorting until they made it to ninth grade. It could be that predominately younger weak students got deselected by grade retention. Consequently, the share for low values of *bmrc* on the total number of compliers would be lower as it is shown in figure 3.1 and the remaining compliers who are born directly before the cutoff would be positively selected. Since the positive selected group is located at the bottom of the *bmrc* distribution, this bias would decrease my estimates and could thus be an explanation for the negative sign of my estimates. Due to the potential bias – the previous literature found positive (**hamori**; **bedard**; **elder**) or no significant effect (**robertson**; **pena**) on test scores for teenagers – I do not want to put too much emphasis on these results.

I would see my addition to the literature rather in my approach to regard the educational choices as an important outcome. The distinction between not obtaining a degree and never starting the education is important. It can help to answer the question why people fail to achieve certain educational levels and is therefore also relevant for policy makers who seek to increase the share of graduated.

While my estimates for the school starting age are at odds with virtually all previous research and a selection bias seems to be plausible, the estimation for the relative age in ninth grade is in line with the literature and may complement it. The estimate for the decision to continue schooling for the upper secondary is significantly negative. To my extend of knowledge, this outcome has not been investigated before and the result would be plausible if we believed in the negative effects on test scores in that age group found by **pena**; **elder**.

Students who belong to the youngest in their high school cohort are thus more likely to proceed their schooling career until the final university entrance diploma. This positive effect of being younger has been attributed to positive effects through older peers and a lower probability to engage in risky behavior.

Interestingly, the estimates for the relative age position afterwards are apart for holding the school leaving certificate at the 19th birthday insignificant. This would suggest that those who did not start the upper secondary education because of their high age position would not have on average a lower probability to finally succeed in it. One could thus conclude on an inefficiency in the Swiss system as older students are prevented from starting the upper secondary education despite not having averagely worse chances to succeed in it.

Online Appendix to: An Early And Lasting Advantage? – Long Term Effects Of The School Starting Age by Arnim Seidlitz

Appendix

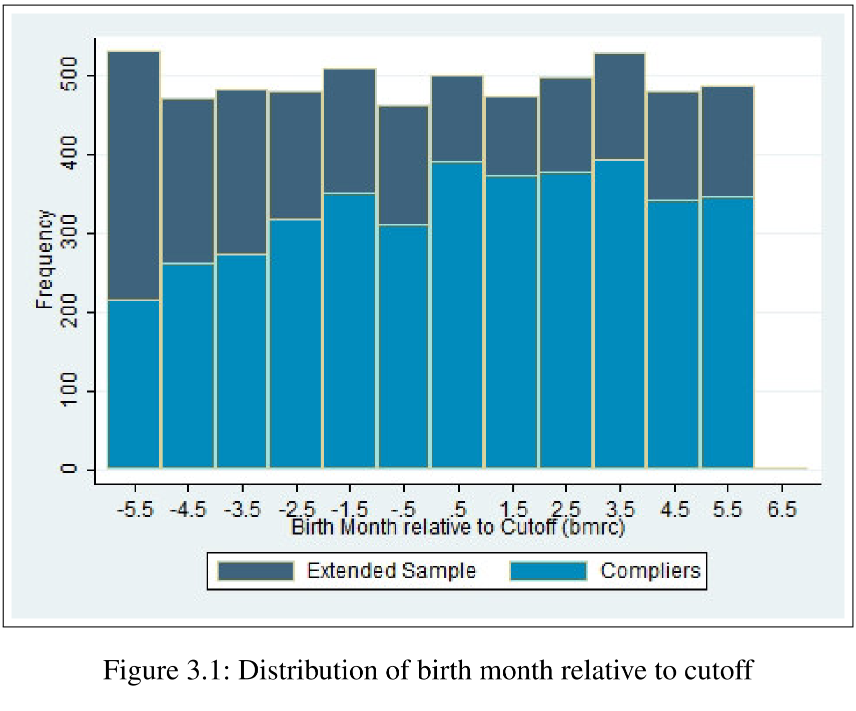


Figure 3.1: Distribution of birth month relative to cutoff

Table 3.1: Cutoff Dates in Swiss states

State (Canton)	Acronym	Cutoff Date	Number of Observations			
			Compliers		Extended Sample	
Aargau	AG	April 30th	169	(4.26)	283	(4.79)
Appenzell Ausserrhoden	AR	January 1st	14	(0.35)	22	(0.37)
Basel-Landschaft	BL	May 1st[1]	96	(2.42)	135	(2.28)
Basel-Stadt	BS	May 1st	47	(1.19)	75	(1.27)
Bern	BE	May 1st	462	(11.65)	664	(11.23)
Fribourg	FR	July 31st	231	(5.82)	396	(6.70)
Genève	GE	June 30th	528	(13.31)	845	(14.29)
Glarus	GL	April 30th	2	(0.05)	7	(0.12)
Graubünden	GR	December 31st	80	(2.02)	92	(1.56)
Jura	JU	June 1st[2]	147	(3.71)	201	(3.40)
Luzern	LU	May 1st	64	(1.61)	99	(1.67)
Neuchatel	NE	August 31st	193	(4.87)	268	(4.53)
Nidwalden	NW	April 30th	13	(0.33)	22	(0.37)
Obwalden	OW	December 31st	53	(1.34)	61	(1.03)
St. Gallen	SG	April 30th	365	(9.20)	600	(10.15)
Schaffhausen	SH	May 1st[1]	23	(0.58)	41	(0.69)
Schwyz	SZ	April 30th	32	(0.81)	55	(0.93)
Solothurn	SO	February 28th	49	(1.24)	76	(1.29)
Thurgau	TG	April 30th[3]	51	(1.29)	90	(1.52)
Ticino	TI	December 31st	546	(13.77)	725	(12.26)
Wallis	VS	September 30th	320	(8.07)	435	(7.36)
Vaud	VD	June 30th	233	(5.87)	354	(5.99)
Zug	ZG	April 30th	28	(0.71)	43	(0.73)
Zürich	ZH	April 30th[4]	220	(5.55)	325	(5.50)
Total Number of Observations			3966	(100)	5914	(100)

Cutoff dates for the school year 1991/1992 (**cesdoc**). Shares on the total number of observations in parenthesis.
1. BL and SH uses April 1st as cutoff for the school year 1989/1990 and May 1st from 1990/1991 on.
2. JU use August 1st as cutoff for the school year 1990/1991 and June 1st from 1991/1992 on.
3. TG uses March 31st as cutoff for the school year 1990/1991 and April 30th from 1991/1992 on.
4. ZH uses March 31st as cutoff for the school year 1989/1990 and April 30th from 1990/1991 on
The two states with the least number of inhabitants, Appenzell Innerrhoden and Uri, are not part of the sample.

Table 3.2: Control Variables

Variable	Explanation		Number of Observations				
			Compliers		Extended Sample		t-statistics[1]
not born in Switzerland		0	3,586	(90.42)	5,138	(86.88)	
		1	380	(9.58)	776	(13.12)	-5.97
migration background	participant herself or one of the	0	2,583	(65.13)	3,657	(61.84)	
	parents not born in Switzerland	1	1,383	(34.87)	2,257	(38.16)	-1.78
mother holds an upper	any ISCED 3 certificate	0	1,419	(35.78)	2,269	(38.37)	
secondary certificate	or higher	1	2,547	(64.22)	3,645	(61.63)	1.75
mother holds a	any ISCED 5 certificate	0	3,232	(81.49)	4,836	(81.77)	
tertiary certificate	or higher	1	734	(18.51)	1,078	(18.23)	-0.72
raised by single parents	0 if participant reports to	0	3,144	(79.27)	4,620	(78.12)	
	live with both parents, 1 else	1	822	(20.73)	1,294	(21.88)	-3.14
other language	other than the language at school	0	3,504	(88.35)	5,048	(85.36)	
	is reported to be main language at home	1	462	(11.65)	866	(14.64)	-2.38
hightrack	attends highest possible track	0	2,120	(53.47)	3,290	(55.65)	
	of the respective state	1	1,845	(46.53)	2,622	(44.35)	3.17
male		0	2,248	(56.68)	3,242	(54.82)	
		1	1,718	(43.32)	2,672	(45.18)	-3.76

Shares on the respective total number of observations in parenthesis.
1. t-statistics for a fixed-effects regression of the control variables on a dummy for being a complier. Fixed effects for schools and birth months are included. Standard errors are clustered on school level.

Table 3.3: PISA test scores

	Reading Skills			Quantative Skills		
	(1)	(2)	(3)	(4)	(5)	(6)
	Compliers	Compliers	IV, extended Sample	Compliers	Compliers	IV, extended Sample
school starting age	-0.012**	-0.010*		-0.010*	-0.009*	
	[-0.020,-0.005]	[-0.018,-0.002]		[-0.018,-0.002]	[-0.017,-0.001]	
relative age effect			0.029			0.195
			[-0.247,0.305]			[-0.111,0.501]
male	-0.226***	-0.225***	-0.224***	0.266***	0.267***	0.264***
	[-0.280,-0.171]	[-0.280,-0.170]	[-0.274,-0.175]	[0.210,0.321]	[0.212,0.323]	[0.213,0.315]
highest track	0.816***	0.816***	0.833***	0.801***	0.802***	0.834***
	[0.719,0.913]	[0.720,0.912]	[0.740,0.926]	[0.697,0.906]	[0.698,0.906]	[0.739,0.929]
not born in CH	-0.097	-0.096	-0.175***	-0.119*	-0.114	-0.243***
	[-0.202,0.007]	[-0.200,0.009]	[-0.257,-0.093]	[-0.239,-0.000]	[-0.233,0.005]	[-0.331,-0.155]
migration background	-0.069*	-0.071*	-0.094***	-0.136***	-0.135***	-0.156***
	[-0.128,-0.011]	[-0.130,-0.012]	[-0.145,-0.044]	[-0.196,-0.075]	[-0.196,-0.075]	[-0.207,-0.106]
other language speaker	-0.294***	-0.291***	-0.291***	-0.261***	-0.260***	-0.259***
	[-0.393,-0.196]	[-0.390,-0.193]	[-0.365,-0.216]	[-0.373,-0.148]	[-0.372,-0.147]	[-0.341,-0.176]
single parents	-0.056*	-0.055*	-0.075**	-0.032	-0.030	-0.079**
	[-0.111,-0.002]	[-0.110,-0.001]	[-0.121,-0.028]	[-0.094,0.030]	[-0.091,0.032]	[-0.131,-0.028]
maternal education:						
upper secondary certificate	0.254***	0.254***	0.273***	0.226***	0.225***	0.267***
	[0.203,0.304]	[0.202,0.305]	[0.227,0.320]	[0.173,0.278]	[0.172,0.278]	[0.219,0.315]
tertiary certificate	-0.063	-0.062	-0.041	-0.011	-0.013	0.009
	[-0.128,0.003]	[-0.127,0.003]	[-0.097,0.014]	[-0.087,0.065]	[-0.089,0.064]	[-0.053,0.071]
birth month FE	No	Yes	Yes	No	Yes	Yes
school FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3959	3959	5906	3509	3509	5244
R^2	0.513	0.514	0.512	0.502	0.503	0.499

95% confidence intervals in brackets
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
The dependent variable of these regression are the reading and the quantitative PISA test scores respectively. The coefficients can be interpreted in percentages of a standard deviation. The standard errors are clustered on school level in all regressions.

Table 3.4: Educational Choices

	Academic upper secondary education			Academic tertiary education		
	(1)	(2)	(3)	(4)	(5)	(6)
	Compliers	Compliers	IV, extended Sample	Compliers	Compliers	IV, extended Sample
school starting age	-0.008***	-0.008***		-0.005	-0.003	
	[-0.012,-0.004]	[-0.012,-0.004]		[-0.010,0.000]	[-0.009,0.002]	
relative age effect			-0.217**			-0.022
			[-0.360,-0.074]			[-0.225,0.182]
male	-0.059***	-0.058***	-0.059***	-0.005	-0.004	0.000
	[-0.088,-0.029]	[-0.088,-0.028]	[-0.087,-0.031]	[-0.045,0.034]	[-0.044,0.036]	[-0.031,0.031]
highest track	0.617***	0.619***	0.601***	0.426***	0.429***	0.492***
	[0.547,0.686]	[0.549,0.689]	[0.537,0.666]	[0.346,0.507]	[0.350,0.508]	[0.425,0.559]
not born in CH	-0.018	-0.017	0.023	0.011	0.010	-0.026
	[-0.075,0.039]	[-0.074,0.039]	[-0.019,0.065]	[-0.058,0.079]	[-0.059,0.079]	[-0.086,0.034]
migration background	0.035	0.036	0.045**	0.034	0.035	0.044*
	[-0.003,0.072]	[-0.002,0.073]	[0.015,0.075]	[-0.004,0.073]	[-0.004,0.074]	[0.010,0.077]
other language speaker	-0.007	-0.007	-0.017	0.028	0.031	-0.014
	[-0.056,0.042]	[-0.056,0.043]	[-0.050,0.016]	[-0.035,0.092]	[-0.031,0.094]	[-0.065,0.036]
single parents	0.002	0.002	-0.001	-0.044	-0.045	-0.052*
	[-0.038,0.042]	[-0.038,0.042]	[-0.036,0.034]	[-0.093,0.006]	[-0.094,0.004]	[-0.093,-0.011]
maternal education:						
upper secondary certificate	0.046**	0.045**	0.050***	0.092***	0.092***	0.091***
	[0.014,0.078]	[0.013,0.077]	[0.023,0.076]	[0.052,0.133]	[0.052,0.133]	[0.056,0.126]
tertiary certificate	0.060***	0.059***	0.073***	0.065**	0.067**	0.068***
	[0.026,0.094]	[0.025,0.093]	[0.046,0.100]	[0.024,0.107]	[0.026,0.108]	[0.031,0.104]
birth month FE	No	Yes	Yes	No	Yes	Yes
school FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2909	2909	4245	2380	2380	3445
R^2	0.591	0.592	0.582	0.394	0.396	0.395

95% confidence intervals in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dependent variable in these regression is a dummy on ever being enrolled at school offering the university entrance diploma and at a university respectively. The standard errors are clustered on school level in all regressions.

Table 3.5: Upper secondary certificate for different ages

	Compliers			Extended sample, IV-regression		
	(1)	(2)	(3)	(4)	(5)	(6)
	At age 19	At age 21	At age 23	At age 19	At age 21	At age 23
school starting age	-0.021***	-0.006**	-0.006*			
	[-0.026,-0.017]	[-0.011,-0.002]	[-0.011,-0.001]			
relative age effect				-0.315***	-0.138	-0.103
				[-0.457,-0.173]	[-0.286,0.010]	[-0.269,0.062]
male	-0.013	-0.065***	-0.055***	-0.008	-0.067***	-0.062***
	[-0.036,0.009]	[-0.095,-0.035]	[-0.087,-0.022]	[-0.026,0.010]	[-0.093,-0.041]	[-0.090,-0.033]
highest track	0.247***	0.534***	0.552***	0.181***	0.492***	0.537***
	[0.187,0.306]	[0.470,0.598]	[0.481,0.623]	[0.138,0.224]	[0.434,0.550]	[0.474,0.601]
not born in CH	0.031	-0.014	-0.020	0.028	-0.007	-0.015
	[-0.003,0.066]	[-0.063,0.035]	[-0.078,0.038]	[-0.001,0.056]	[-0.044,0.029]	[-0.062,0.032]
migration background	-0.004	0.025	0.034	-0.011	0.026	0.040*
	[-0.029,0.021]	[-0.011,0.061]	[-0.006,0.074]	[-0.032,0.009]	[-0.002,0.054]	[0.008,0.072]
other language speaker	-0.006	0.002	0.018	0.008	-0.005	-0.005
	[-0.039,0.027]	[-0.056,0.061]	[-0.044,0.079]	[-0.018,0.033]	[-0.048,0.038]	[-0.051,0.041]
single parents	-0.030*	-0.039*	-0.028	-0.020	-0.036*	-0.022
	[-0.056,-0.004]	[-0.076,-0.003]	[-0.070,0.014]	[-0.041,0.000]	[-0.067,-0.005]	[-0.057,0.013]
maternal education:						
upper secondary certificate	0.028*	0.060***	0.059**	0.008	0.053***	0.058***
	[0.006,0.050]	[0.029,0.092]	[0.024,0.095]	[-0.008,0.024]	[0.028,0.078]	[0.031,0.086]
tertiary certificate	0.012	0.073***	0.085***	0.034**	0.084***	0.098***
	[-0.019,0.043]	[0.036,0.111]	[0.049,0.121]	[0.009,0.059]	[0.053,0.115]	[0.069,0.128]
Observations	3311	3031	2742	4949	4482	4021
R^2	0.431	0.518	0.532	0.353	0.509	0.525

95% confidence intervals in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dependent variable in these regression is a dummy of having obtained the university entrance diploma by the respective birthday. All regressions include fixed effects on schools and the month of birth. The standard errors are clustered on school level in all regressions.

Table 3.6: Tertiary certificate for different ages

	Compliers			Extended sample, IV-regression		
	(1)	(2)	(3)	(4)	(5)	(6)
	At age 23	At age 25	At age 27	At age 23	At age 25	At age 27
school starting age	-0.007***	-0.005	-0.002			
	[-0.011,-0.003]	[-0.011,0.002]	[-0.009,0.005]			
relative age effect				-0.041	0.025	-0.007
				[-0.159,0.076]	[-0.159,0.209]	[-0.224,0.210]
male	-0.023*	-0.033	-0.041	-0.021*	-0.036*	-0.029
	[-0.046,-0.001]	[-0.076,0.010]	[-0.086,0.004]	[-0.039,-0.003]	[-0.070,-0.003]	[-0.064,0.006]
highest track	0.096***	0.335***	0.447***	0.092***	0.341***	0.459***
	[0.053,0.139]	[0.273,0.396]	[0.368,0.527]	[0.061,0.123]	[0.289,0.392]	[0.399,0.518]
not born in CH	-0.017	0.025	0.042	-0.020	-0.010	-0.006
	[-0.056,0.022]	[-0.049,0.100]	[-0.038,0.123]	[-0.046,0.005]	[-0.069,0.050]	[-0.070,0.058]
migration background	0.002	0.009	0.039	-0.004	-0.019	0.030
	[-0.031,0.034]	[-0.040,0.058]	[-0.014,0.092]	[-0.028,0.020]	[-0.059,0.021]	[-0.014,0.073]
other language speaker	-0.009	0.023	-0.009	0.002	0.005	-0.022
	[-0.048,0.030]	[-0.048,0.094]	[-0.088,0.069]	[-0.025,0.028]	[-0.048,0.057]	[-0.082,0.038]
single parents	-0.018	-0.072**	-0.065*	-0.019	-0.078***	-0.079**
	[-0.045,0.010]	[-0.121,-0.023]	[-0.129,-0.002]	[-0.040,0.002]	[-0.116,-0.040]	[-0.126,-0.032]
maternal education:						
upper secondary certificate	0.004	0.059**	0.098***	0.003	0.052***	0.090***
	[-0.021,0.029]	[0.018,0.100]	[0.052,0.145]	[-0.014,0.020]	[0.023,0.082]	[0.054,0.127]
tertiary certificate	-0.012	0.050*	0.063*	0.003	0.055**	0.062**
	[-0.045,0.021]	[0.001,0.099]	[0.011,0.116]	[-0.022,0.028]	[0.013,0.097]	[0.018,0.106]
Observations	2641	2482	2182	3892	3567	3160
R^2	0.129	0.255	0.337	0.114	0.230	0.330

95% confidence intervals in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dependent variable in these regression is a dummy of having obtained a university degree by the respective birthday. All regressions include fixed effects on schools and the month of birth. The standard errors are clustered on school level in all regressions.

Statistically Significant Overlaps in the German Equity Fund Sector: A Network Analysis

Lucie Stoppok*

1 Abstract

How do German equity fund networks perform when markets struggle? The German mutual fund sector and in particular equity funds have received increased attention over the last couple of years because of their growth in size and high level of interconnectedness. However, data limitations have prevented researchers from being able to perform in depth analyses of the sector. With the Deutsche Bundesbank's micro level dataset on German equity funds, I shed light on how fund returns are affected by volatile markets depending on the funds network affiliation. My identification strategy follows a statistical null random network model that allows one to identify statistically important fund networks. Based on these groupings, I run panel regressions on the funds return performance. My results show that German equity funds that are statistically proven to be highly similar in terms of their sector investments suffer negative returns during crisis periods, while their non-significant counterparts experience low but positive returns.

*Lucie received her Master's degree from the University of Bonn in 2017. Afterwards she started her PhD at the Kiel Institute for the World Economy. The present article refers to her Master's thesis under supervision of Prof. Dr. Moritz Schularick which was submitted in March 2017.

2 Introduction

The German equity fund sector has grown fast over the last years with a doubling in value from 137 Billion Euros in January 2010 to 249 Billion Euros in September 2016. Today equity funds represent roughly 15% of the whole German investment funds sector.[1] Because of the sector's size and its interconnectedness, the German equity fund sector has received increased attention from investors, policy makers and regulators. However, because of data limitations, a thorough study that entirely builds on micro level data and which focuses on the return performance of funds with similar sector investments could not yet be conducted. In particular, it is not clear how these funds perform during crisis periods. This paper aims to fill the gap. I follow Gualdi et al. (2016), who employ a Bipartite Configuration Model (BiCM), which looks at any two funds and compares their portfolios in terms of the funds' diversification levels and the number of linkages that form the portfolio overlap. A network structure, in which funds share overlapping portfolios based on common industry sector investments, is created.

While equity fund returns have been studied extensively in the past for both the US and the UK markets (e.g see Fama and French (2010), Berk and Green (2004), Blake and Timmermann (1998), Cuthbertson et al. (2010)), data limitations have kept researchers from examining the German market. My master thesis provides a first step towards an analysis on German equity fund returns by using a novel dataset compiled by the Deutsche Bundesbank's Investment Funds Statistics (IFS) and the ECB's Centralized Security Database (CSDB). It includes micro data on German equity funds and their sector holdings on a monthly basis between January 2010 and September 2016. Panel regressions find that the return performance for funds with non-random overlaps is significantly associated with lower returns and that struggling market situations even amplify the effect.

This paper adds to different strands of literature on the applications of statistical null models, portfolio overlaps and the return performance of (equity) funds. A growing literature has recently started to use network theory to explain economic and financial phenomena. However, while network theory itself is wide-ranging, this thesis contributes to a field of network theory that concentrates on pattern detection in real-world systems in particular based on null (random) network

[1] As measured by the total amount in February 2016 according to the Investment Funds Statistics of the Deutsche Bundesbank.

models. Three papers that employ such models to study patterns in empirical bipartite network structures are Fagiolo et al. (2013), Gualdi et al. (2016) and Saracco et al. (2016). Gualdi et al. (2016) explore systemic risk channels that emerge from a network of financial institutions with overlapping portfolios. Fagiolo et al. (2013) study the binary and weighted directed graph presentation of the World Trade Web over a time period of 50 years from 1950 to 2000 and Saracco et al. (2016) look into the World Trade Web with an additional evaluation on movies and their user ratings. This master thesis contributes to the literature by shedding light on the network of German equity funds using a null (random) network model.

So far, only little work has focused on portfolio overlaps that arise through investments placed in the same industry sectors labeled "common sector investments". One paper that has explored this area is Cai et al. (2011). Cai et al. (2011) study systemic risk in the syndicated loan market using two empirical interconnectedness measures. Their portfolio overlaps build on the industry sector classification "SIC", which stands for Standard Industrial Classification. With the SIC, Cai et al. (2011) are able to classify the destination of bank investments based on the corresponding industry sectors. This master thesis employs the ECB's Nomenclature statistique des activités économiques dans la Communauté européenne (NACE) to study portfolio overlaps through investments made in the same industry sector.

By definition, funds following similar investment strategies generate similar returns (Fricke (2016)). One strand of literature has produced research on the common ownership of stock prices as a source for systemic risk (e.g. see Coval and Stafford (2007), Greenwood et al. (2015)), while another focusses on the relationship between the funds' flows and their return performances (e.g. see Lou (2012)). One paper that studies the effect on return performance from the *liability* side of funds is Anton and Polk (2014). Anton and Polk (2014) investigate the effect the presence of common investors have on stock returns. They show that connected stocks tend to co-move more strongly with each other than stocks of investors that are not connected. This master thesis rests its focus on the *asset* side of funds by showing that similar asset holdings influence fund returns.

The remainder of the paper is structured as follows. Section 3 explains the model approach of the BiCM and its mechanism in detail. Section 4 gives insight into the novel dataset, the model's implementation and the outcomes obtained from it.

Section 5 runs a set of panel regressions on the returns of German equity funds based on the results found in section 4. Section 6 concludes.

3 A Bipartite Network Analysis

In order to disentangle the tightly interconnected German equity fund sector, network theory can be employed. One network structure called bipartite networks has shown to provide a valuable and insightful representation of real-world systems, that allow to identify statistically-significant structural properties (Saracco et al. (2016)). In particular, it is used to understand the interactions that occur between distinct groups of nodes. In my master thesis I employ a statistical null model to obtain statistically-validated projections of a bipartite equity funds network. I follow Gualdi et al. (2016), in the broad sense who apply a Bipartite Configuration Model that looks at any two portfolios of funds and determines whether the funds' investments are too similar to count as random.

3.1 Bipartite Configuration Model

The German equity fund sector is tightly entangled. This interconnectedness makes it difficult to directly spot funds that follow similar investment strategies in terms of the industry sectors they choose. The BiCM helps to overcome this problem. It is a statistical null model that reconstructs a network as a random graph which creates a projection of the bipartite network of funds and sectors by looking at each two funds' portfolio overlap. It sets up a null hypothesis that is discarded if the portfolios are too similar and compares the portfolios based on the funds' diversification levels, the securities' diversification levels and the number of common investments. In doing so, the mechanism allows to separate the network into two groups: In the first group, the funds show non-random overlaps with at least one other fund in the sector. The second group comprises funds which place their industry investments completely randomly.

Projection

The projection of a bipartite graph to a monopartite representation allows for a simplification of the network structure and to encode the actual information of interest by reducing the system of funds and industry sectors to funds and their linkages (Saracco et al. (2016)). The BiCM builds a validated projection of the original bipartite network that only shows overlaps that become statistically significant through the mechanism (Gualdi et al. (2016)).

The transformation of a bipartite network representation to a monopartite structure can be derived by means of matrix notation. Following Gualdi et al. (2016), the dataset at a given time t consists of a set $I(t)$ of German investment funds that invest in a various number of industry sectors $S(t)$. It is described by a $|I(t)| \times |S(t)|$ ownership matrix $\mathscr{W}(t)$ whose generic elements w_{is} show the amount each fund $i \in I(t)$ places in an industry sector $s \in S(t)$. From this, a $|I(t)| \times |S(t)|$ asymmetric incidence matrix $\mathscr{A}(t)$ with binary entries $a_{is}(t)$ is constructed.[2] The elements $a_{is}(t)$ can take the following values:

$$a_{is}(t) = \begin{cases} 1 & \text{if } \; w_{is}(t) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Whenever $w_{is}(t) > 0$, a fund $i \in I(t)$ makes a positive investment in a certain industry sector $s \in S(t)$ such that i and s are linked. Counting the number of investments for each fund i at time t gives the degree vector of funds $d^{Fund}(t)$ of the form $|I(t)| \times 1$ with elements defined by

$$d_i^{Fund}(t) = \sum_{s \in S} a_{is}(t), \;\; i = 1, ..., |I(t)|.$$

$d_i^{Fund}(t)$ describes the sum of all direct links node i shares with every node $s \in S$. In the same fashion, every element in the degree vector of securities $d^{Sec}(t)$ counts the number of funds that invest in a particular industry sector. It is of the form $|S(t)| \times 1$. The elements are given by

$$d_s^{Sec}(t) = \sum_{i \in I} a_{is}(t), \;\; s = 1, ..., |S(t)|.$$

[2] Note that the matrix is not symmetric because the underlying graph is directed.

The number of common investments $o_{ij}(t)$ of two neighbors i and j are described by the sum over the product of i and j in the binary matrix $\mathscr{A}(t)$. More precisely, $o_{ij}(t)$ writes as

$$o_{ij}(t) = \begin{cases} \sum_{s \in S} a_{is}(t) a_{js}(t) & \text{if } i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

Here, every $o_{ij}(t)$ is an element of the portfolio overlap matrix $\mathscr{O}(t)$ with cardinality $|I(t)| \times |I(t)|$. The symmetric matrix $\mathscr{O}(t)$ prescribes the monopartite projection of the bipartite graph. Its generic elements $o_{ij}(t)$ reveal the exact number of links two funds i and j have in common, namely the *overlap* of their portfolios. In doing so, it gives a reduced form of $\mathscr{A}(t)$ down from a complex system of funds and sectors to just one set of funds (Tumminello et al. (2011)).

Based on $\mathscr{O}(t)$ an adjacency matrix $\mathscr{F}(t)$ can be created that replaces positive observation $o_{ij}(t) > 0$ in $\mathscr{O}(t)$ with 1. In contrast to $\mathscr{O}(t)$, which counts the number of links two funds share, $\mathscr{F}(t)$ only indicates whether a portfolio overlap exists. The matrix is given by

$$\mathscr{F}(t) = \begin{cases} 1 & \text{if } \quad o_{ij}(t) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

So far, the naive monopartite projection of the bipartite network has been determined. However, the network of interest is a *validated* projection $\mathscr{V}(t)$ that only shows statistical significant overlaps. To derive the validated projection matrix, a null hypothesis on portfolio overlaps is created. The next subsection will show its constitution.

Null Hypothesis and Threshold Selection

The BiCM creates a validated projection of the bipartite network that consists only of validated portfolio overlaps. The BiCM tests a portfolio overlap for statistical significance using a null hypothesis that assumes each two funds' portfolios to be significantly dissimilar. In order to create such a null hypothesis, a probability distribution $\pi(\cdot|d_i, d_j, t)$ is derived for the overlap between two funds i and j. It gives the possible probabilities of having a certain number of sector investments in common. Based on the probability distribution $\pi(\cdot|d_i,d_j,t)$ the probability of having an overlap $O_{ij}(t)$ larger than the observed one $o_{ij}(t)$ can be derived as

$$P(O_{ij}(t) > o_{ij}(t)) \equiv 1 - \sum_{x=0}^{o_{ij}(t)-1} \pi(x|d_i,d_j,t). \quad (2)$$

This probability is called p-value. The null hypothesis defines this p-value to be larger or equal than a threshold $P^*(t)$:

$$H_0 : P(O_{ij}(t) > o_{ij}(t)) \geq P^*(t).$$

The alternative is given by

$$H_1 : P(O_{ij}(t) > o_{ij}(t)) < P^*(t).$$

In case that the threshold $P^*(t)$ is undercut, the portfolio overlap between i and j becomes statistically significant meaning that the overlap between fund i and fund j becomes *validated*. In case the null hypothesis is retained, a likely realization of the null hypothesis according to the significance level $P^*(t)$ is given (Gualdi et al. (2016)). In such a case the overlap is *not-validated*. A fund itself is defined to be validated at time t if at least one of its overlaps becomes validated. The test for statistical significance is repeated for every portfolio overlap in the German equity fund sector at any point in time.

The benchmark $P^*(t)$ is used for testing the statistical significance of every overlap in the network at time t simultaneously. In order to find a threshold that can solve this multiple comparison problem the so called False Discovery Rate (FDR) procedure is applied. Following Saracco et al. (2016), at time t there are $M \equiv \frac{|I(t)|\times(|I(t)|-1)}{2}$ p-values described by p-value$_1, \ldots,$p-value$_M$. Let p-value$_{(1)} \leq \ldots \leq$ p-value$_{(k)} \leq \ldots \leq$ p-value$_{(M)}$ be the ordered p-values. The authors define the multiple-testing procedure with k as the largest integer that still satisfies the condition

$$\text{p-value}_{(k)} \leq \frac{z \cdot k}{M} \quad (3)$$

with z as the 5% significance level. Next, the procedure rejects every hypothesis with a p-value smaller than p-value$_{(k)}$. The threshold is defined as $P^*(t) \equiv$ p-value$_{(k)}$. If no value fulfils inequality (3), $P^*(t)$ is set to zero. As no p-value is able to undercut this value no overlap will be validated at time t.

Maximization Process

The null model introduced in the last subsection assumes that all randomized variants of the original network have exactly the same degree-sequences as the real-world network but are otherwise random (Fagiolo et al. (2013)). A way to model this is to employ an analytical model introduced in Squartini and Garlaschelli (2011) that is free of assumptions about the structure of the original network other than the diversification levels (Fagiolo et al. (2013)). In this method a probability $P(\mathscr{A})$ is chosen such that $\mathscr{A}$ lies within the set of possible network realizations $\mathscr{G}$ in a way that $\{d_s(\mathscr{A})\}_{s=1}^{|S|}$ and $\{d_i(\mathscr{A})\}_{i=1}^{|I|}$ are, on average, equal to the one in the original network (Fronczak (2014)). The probability writes as

$$P(\mathscr{A}|\vec{\alpha},\vec{\beta}) = \frac{\exp\left(-\vec{\alpha}\cdot\vec{d}^{Fund}(\mathscr{A}) - \vec{\beta}\cdot\vec{d}^{Sec}(\mathscr{A})\right)}{\sum_{\mathscr{A}\in\mathscr{G}}\exp\left(-\vec{\alpha}\cdot\vec{d}^{Fund}(\mathscr{A}) - \vec{\beta}\cdot\vec{d}^{Sec}(\mathscr{A})\right)} = \prod_{i\in I}\prod_{s\in S} Q_{is}^{a_{is}}(1-Q_{is})^{1-a_{is}}, \tag{4}$$

where

$$Q_{is} \equiv \frac{\theta_i\theta_s}{1+\theta_i\theta_s}, \ \text{ with } \theta_i = \exp(-\alpha_i) \ \text{ and } \theta_s = \exp(-\beta_s),$$

is defined as the probability of a connection between node i and node s. In order to satisfy $Q_{is} \geq 0$, it has to hold that $\theta_i \geq 0$ and $\theta_s \geq 0$ for all i and s, respectively. The probability Q_{is} is used to determine the probability distribution from equation (2). However, the maximization of the Shannon entropy only gives the formula for Q_{is} but does not provide values for $\{\theta_i\}_{i=1}^{|I|}$ and $\{\theta_s\}_{s=1}^{|S|}$ to calculate it. The approach used here is to rewrite $P(\mathscr{A}|\vec{\alpha},\vec{\beta})$ only in terms of the observed constraints. In doing so, $\{\theta_i\}_{i=1}^{|I|}$ and $\{\theta_s\}_{s=1}^{|S|}$ can be determined by maximizing the probability of observing $\mathscr{A}$ using a log-likelihood function. The first step yields

$$P(\mathscr{A}|\vec{\theta}) = \prod_{i\in I}\theta_i^{d_i^{Fund}(\mathscr{A})}\prod_{s\in S}\theta_s^{d_s^{Sec}(\mathscr{A})}\prod_{i\in I}\prod_{s\in S}\Big(1+\theta_i\theta_s\Big)^{-1}.$$

In order to obtain estimates for the parameters $\{\theta_i\}_{i=1}^{|I|}$ and $\{\theta_s\}_{s=1}^{|S|}$ the likelihood of observing $\mathscr{A}$ is maximized with a log-likelihood function $\mathscr{L}(\theta_i,\theta_s) = \ln P(\mathscr{A}|\vec{\theta})$:

$$\mathscr{L}(\theta_i,\theta_s) = \sum_{i\in I} d_i^{Fund}(\mathscr{A})\ln(\theta_i) + \sum_{s\in S} d_s^{Sec}(\mathscr{A})\ln(\theta_s) - \sum_{s\in S}\sum_{i\in I}\ln(1+\theta_i\theta_s).$$

The maximization of $\mathscr{L}$ with respect to θ_i and θ_s yields

$$\frac{\partial \mathscr{L}}{\partial \theta_i} = \frac{1}{\theta_i} d_i^{Fund}(\mathscr{A}) - \sum_{s \in S} \frac{\theta_s}{1+\theta_i\theta_s} \stackrel{!}{=} 0 \;\; \forall i \in I, \tag{5}$$

$$\frac{\partial \mathscr{L}}{\partial \theta_s} = \frac{1}{\theta_s} d_s^{Sec}(\mathscr{A}) - \sum_{i \in I} \frac{\theta_i}{1+\theta_i\theta_s} \stackrel{!}{=} 0 \;\; \forall s \in S. \tag{6}$$

Solved for the diversification levels $d_i^{Fund}(\mathscr{A})$ and $d_s^{Sec}(\mathscr{A})$ the two equations (5) and (6) read as

$$d_i^{Fund}(\mathscr{A}) = \sum_{s \in S} \frac{\theta_i\theta_s}{1+\theta_i\theta_s} \equiv \sum_{s \in S} Q_{is} \;\; \forall i \in I,$$

$$d_s^{Sec}(\mathscr{A}) = \sum_{i \in I} \frac{\theta_i\theta_s}{1+\theta_i\theta_s} \equiv \sum_{i \in I} Q_{is} \;\; \forall s \in S.$$

The non-linear system can be solved for these hidden variables which yields unique solutions that underlie the observable degree sequences $\{d_i(\mathscr{A})\}_{i=1}^{|I|}$ and $\{d_s(\mathscr{A})\}_{s=1}^{|S|}$. The probability distribution $\pi(\cdot|d_i, d_j)$ is determined based on these values.

P-Value Derivation

Following Gualdi et al. (2016), the probability distribution $\pi(\cdot|d_i, d_j)$ of the expected overlap under the null hypothesis is given by a binomial distribution that constitutes the sum of S independent Bernoulli trials, where each draw takes place with probability $q_{ij}^s = Q_{is}Q_{js}$. More precisely, q_{ij}^s describes the probability of two funds i and j sharing a link because of their common industry investment in s. As every industry sector s with the same degree possesses the same overlap probability, the number of calculations that have to be performed to obtain the probability for a portfolio overlap can be significantly reduced by eliminating existing doublings. Hence, in the following, the superscript h will be used instead of s where h describes any industry sector in $1, \ldots, S$ with degree $\tilde{d}_h$. The sorted set $\{\tilde{d}_h\}_{h=1}^{d_s^{max}}$ is comprised of the unique occurence of degrees among all possible sectors in the network. How often a sector with degree h occurs is recorded by $\tilde{n}_h$.

Restricted to industry sector(s) h, the expected overlap $\langle o_{ij}^h \rangle_{\mathscr{G}}$ between fund i and fund j follows the binomial mass function

$$\pi_h(x|\tilde{n}_h, q_{ij}^h) = \binom{\tilde{n}_h}{x} [q_{ij}^h]^x [1 - q_{ij}^h]^{\tilde{n}_h - x} \text{ with } x \in [0, \tilde{n}_h]. \tag{7}$$

According to equation (7), there exist $\binom{\tilde{n}_h}{x}$ possibilities to draw x nodes out of $\tilde{n}_h$ vertices. The probability $[q_{ij}^h]^x$ then describes that these vertices are linked. The combination $\binom{\tilde{n}_h}{x}[q_{ij}^h]^x$ gives the probability that a node is connected to these x nodes. However, this means that no edges are allowed to link these nodes to the remaining $\tilde{n}_h - x$ nodes. This is given with probability $[1 - q_{ij}^h]^{\tilde{n}_h - x}$.

Following Butler and Stephens (1993), the probability that x occurs can be written as the product of two discrete binomial random variables:

$$\pi_{\leq h}(x|d_i, d_j) = \sum_{k=0}^{x} \pi_{\leq h-1}(x - k|d_i, d_j) \pi_h(x|\tilde{n}_h, q_{ij}^h)$$

with $\pi_{\leq 0}(x - k|d_i, d_j) = 0$ for all $x \geq k > 0$ such that $\pi_{\leq 1}(x|d_i, d_j) = \pi_1(x|\tilde{n}_h, q_{ij}^h)$. Summing over every possible sector degree yields $\pi(x|d_i, d_j) \equiv \pi_{\leq d_s^{max}}(x|d_i, d_j)$, which gives the probability that the number of links between fund i and fund j is less than or equal to x. It allows to derive the p-value as defined in equation (2):

$$P(O_{ij} > o_{ij}) \equiv 1 - \sum_{x=0}^{o_{ij}-1} \pi(x|d_i, d_j).$$

4 Data

The data employed comes from two sources: the Deutsche Bundesbank and the European Central Bank (ECB). The main dataset constitutes the Bundesbank's Investment Funds Statistics (IFS) to which information on industry sectors are added from the ECB's Centralized Security Database (CSDB). The IFS supplies compulsory registration information on German investment funds and delivers information on both public and specialized funds, the number of securities the funds hold, the number of units outstanding and sales in units, issue and repurchase prices per unit, the sales receipts and the fund repurchases (Bundesbank (2017)).

On a monthly basis the IFS provides information on more than 5700 funds and 122.682 securities.[3]

The CSDB is a database provided by the ECB and which is accessible through the Deutsche Bundesbank. It releases information about more than five million debt securities, equities and mutual fund shares issued by EU residents and about securities held and transacted in by EU residents (European Central Bank (2010)). Moreover, it contains a statistical classification of economic activities in the European community called NACE which assigns securities listed in the CSDB to industry classes such as financial and insurance activities or agriculture, forestry and fishing. Securities listed in the CSDB, for which NACE sectors are provided, can be matched with securities held by German equity funds. The compiled dataset of IFS and CSDB spans a network of German equity funds, their securities and their investments in certain industry sectors from January 2010 to September 2016 with a NACE sector coverage rate of 99.9%. The compiled dataset provides a chain of funds, securities and sectors which can be simplified by removing securities. The reduction gives a dataset on German equity funds and the industry sectors they invest in. The reduced dataset contains the investment strategies of 172 German equity funds over the entire time period of 81 months.

5 A Panel Regression Analysis

Before the question of a difference in return performance between validated and non-validated funds is addressed, the driving factors behind the validation are examined. A closer look at the names of the funds in the validation group reveals that the words “DAX” and “STOXX” clearly stand out. Here, “DAX” refers to DAX tracking funds while “STOXX” stands for an integrated index that covers different market providers such as EURO STOXX 50 or STOXX Europe 600. Funds that have the word “STOXX” in their names refer to funds tracking STOXX Index types. In order to test for a statistical relationship between the validation and a funds’ classification, a panel regression is conducted. A dataset is created that is comprised of the International Security Identification Number (ISIN) that uniquely identifies every fund in the sector, the funds’ names, a dummy variable

[3] According to the IFS in September 2016.

that attributes a 1 to every validated fund in the panel and two dummy variables for "DAX" and "STOXX":

$$D_{i,t}^{Valid} = \alpha + \beta_{i,1} D_i^{DAX} + \beta_{i,2} D_i^{STOXX} + \varepsilon_{i,t}. \tag{8}$$

The regression allows for a mean comparison between "DAX"/"STOXX"-funds and funds that show neither of the two attributes. Table 4.2 provides the regression results of (8). A normal fund has an on average probability of becoming validated of 15.5%. For a fund with the word "STOXX" in its name the probability of validation is on average 30.6%. In case of DAX trackers, the likelihood is even higher: "DAX"-funds have an on average validation probability of 61.1%. The regression output shows, that there is indeed a statistically significant relationship between "DAX"/"STOXX"-funds and the funds' validation.

5.1 Network Affiliation and Return Performance

Next, a panel regression is conducted on the funds' returns, their lagged returns and the dummy variable for validated funds:

$$Return_{i,t} = \alpha + \beta_{i,1} Return_{i,t-1} + \beta_{i,2} D_{i,t}^{Validated} + \varepsilon_{i,t}.$$

Column 1 of Regression Table 4.3 displays the results. Holding the lagged return constant, the monthly average excess return for non-validated German equity fund is 0.309% or 3.8% compounded annually. The next row provides correlational evidence on the negative relationship between the funds' returns and validation. Ceteris paribus, validated funds have a monthly on average return of -0.005% which is -0.314 percentage points lower than the average return of non-validated funds. The question as to whether this result is driven by factors other than the network affiliation of validated funds remains. In order to test for one potential source of omitted variable bias, a regression is conducted on the relationship between the returns, the lagged returns and validation controlling for "DAX" and "STOXX" tracking funds. Column 2 of Table 4.3 shows that neither the regression coefficient for "DAX" nor for "STOXX" is significant at the 1% level. Also the regression coefficient for the dummy variable "Validated" remains highly significant at the 0.1% level. This outcome supports the correlational evidence of column 1 but does not rule out driving factors other than validation.

Return Performance in Crisis-Like Periods

This subsection investigates the correlational effect of validation under stressed market conditions. The German market is under stress when a positive VDAX change reaches its 90th percentile. This paper follows the approach of Dennis et al. (2006) who use the daily change in the standardized implied volatility as a proxy for innovations in the expected volatility of stock returns. The 90th percentile of positive VDAX-changes yields eight observations. In order to evaluate the correlational relationship between fund returns and validation in stressed market situations, a panel regression is conducted that includes the lagged return of the funds, a validation dummy, a time dummy for crisis-like periods and an interaction term between the validation and the crisis dummy:

$$\begin{aligned} Return_{i,t} = \alpha \quad & + \quad \beta_{i,1} Return_{i,t-1} \\ & + \quad \beta_{i,2} D_{i,t}^{Validated} + \beta_{i,3} D_t^{Crisis} + \beta_{i,4} D_{i,t}^{Validated} D_t^{Crisis} + \varepsilon_{i,t}. \end{aligned}$$

Column 3 of Table 4.3 provides the results. Holding the lagged return constant, the average monthly excess return for non-validated funds corresponds to .618%. The reason why the constant in column 3 in Table 4.3 is almost double the size of the constant in column 1 lies in the separation between crisis and non-crisis periods. In normal times funds perform generally better than in times when markets struggle. Outside of crisis periods the negative correlational relationship between validated funds and returns has declined but is still statistically significant at the 1% level. When markets struggle, the crisis dummy clearly amplifies the effect for validated funds: ceteris paribus, the return of a validated fund is on average -5.9% per month. This effect is strong and clearly crisis driven. When controlling for the correlation between fund returns and the DAX return (column 4), the parameter of the interaction term remains stable both in size and in significance. Ceteris paribus, the returns of validated funds are on average still negative even though to a lesser extend.

6 Concluding Remarks

This master thesis has examined the tightly interconnected German equity fund sector by making use of a statistical model, the BiCM and panel regressions.

The mechanism behind the BiCM has detected statistically significant overlaps between funds that follow similar sector investment strategies. Panel regressions have provided evidence on network membership: “STOXX”- and “DAX”-funds show on average validation probabilities of 30.6% and 61.1%, respectively. A correlational relationship were found between the return performance of German equity funds and the network affiliation to validation. The regression output provides a statistical significant negative relationship at the 0.1% level between validation and fund returns. Even after controlling for omitted variable bias, coefficients remain significant. Crisis periods, which were calculated using the 90th percentile of positive VDAX changes, amplify the effect of validation even further. While funds outside the network of validated funds show low returns during crisis periods, funds with statistically significant overlaps face negative returns.

Online Appendix to: Statistically Significant Overlaps in the German Equity Fund Sector: A Network Analysis by Lucie Stoppok

Industry Sector Classification

The broad structure of NACE 1 is presented in Table 4.1. The overview comes from Eurostat (2008) and provides information on the NACE 1 sector classifications according to NACE Rev 2. It allows to construct NACE 2 by adding the division number on the far right to the corresponding NACE section on the left. The exact definitions for all NACE sector classifications 2–4 can be found in Eurostat (2008) which are not listed here due to space limitations.

Incorrect Sector Assignments

The ECB's CSDB attributes whenever possible a NACE sector to a given security. However, in the novel dataset compiled of IFS and CSDB 15 incorrect NACE sectors were detected which do not exist according to NACE Rev. 2. The incorrect NACE sectors are
D34, D341, D343, E40, E401, E4011, E4012, E4013, E402, E4021, E4022, J67, J671, J6712. All of these 15 NACE sectors have been replace by the term "Not Available" or were ruled out by correct NACE sectors attributed to the same security at a later point in time.

Table 4.1: NACE 1 Sector Overview

NACE Section	Title	Division
A	Agriculture, forestry and fishing	01–03
B	Mining and quarrying	05–09
C	Manufacturing	10–33
D	Electricity, gas, steam and air conditioning supply	35
E	Water supply; sewerage, waste management*	36–39
F	Construction	41–43
G	Wholesale and retail trade; repair of motor vehicles*	45–47
H	Transportation and storage	49–53
I	Accommodation and food service activities	55–56
J	Information and communication	58–63
K	Financial and insurance activities	64–66
L	Real estate activities	68
M	Professional, scientific and technical activities	69–75
N	Administrative and support service activities	77–82
O	Public administration and defence*	84
P	Education	85
Q	Human health and social work activities	86–88
R	Arts, entertainment and recreation	90–93
S	Other service activities	94–96
T	Activities of households as employers*	97–98
U	Activities of extraterritorial organisations and bodies	99

*All sector definitions labelled with an asterisk have been shortened to fit the table size.

Regression Outputs

Table 4.2: Validation Driving Factors

	Validated
Constant	0.155***
	(0.00795)
"DAX"	0.456***
	(0.0151)
"STOXX"	0.151***
	(0.0118)
Observations	13742

Standard errors in parentheses
* p <0.05, ** p <0.01, *** p <0.001

Table 4.3: Regression on Returns

	(1)	(2)	(3)	(4)
Constant	0.00309***	0.00324***	0.00618***	0.00715***
	(0.000391)	(0.000404)	(0.000412)	(0.000436)
Validated	-0.00314***	-0.00317***	-0.00275**	-0.00187*
	(0.000918)	(0.000927)	(0.000896)	(0.000927)
"DAX"		0.00156		
		(0.00171)		
"STOXX"		-0.00275*		
		(0.00131)		
Crisis			-0.0267***	-0.0266***
			(0.00131)	(0.00131)
Validated × Crisis			-0.0114***	-0.0115***
			(0.00313)	(0.00313)
Dax Corr.				0.0256***
				(0.00386)
Lagged Return	X	X	X	X
Observations	13644	13644	13644	13644

Standard errors in parentheses
* p <0.05, ** p <0.01, *** p <0.001